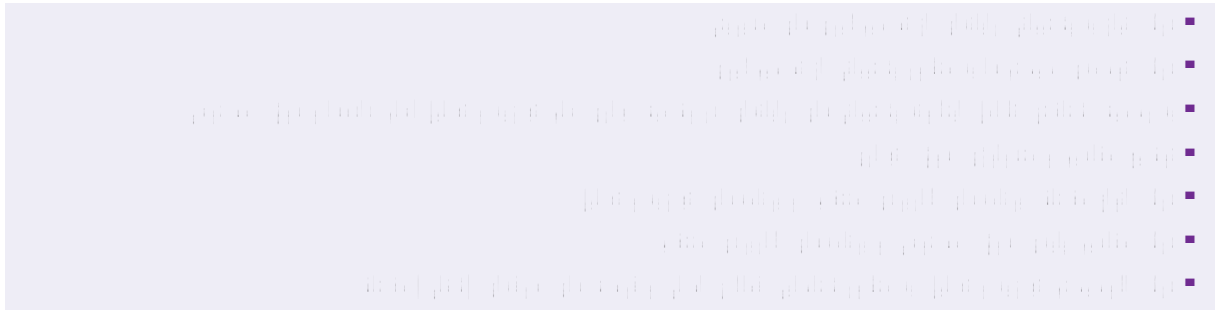


مروری بر هوش تجاری، علم تجزیه و تحلیل، علم داده‌ها و هوش مصنوعی: سیستم‌های پشتیبان تصمیم‌گیری



فضای کسب‌وکار دائماً در حال تغییر است و بر پیچیدگی آن افزوده می‌شود. سازمان‌های خصوصی و دولتی تحت فشارهایی قرار دارند که آنها را مجبور می‌کند تا سریعاً به شرایط متغیر واکنش نشان داده و به شکلی نوآورانه عمل کنند. چنین فعالیت‌هایی نیازمند این هستند که سازمان‌ها به شکلی چابک عمل کرده و تصمیمات استراتژیک، تاکتیکی و عملیاتی که بعضاً بسیار پیچیده هستند را چندین مرتبه و به سرعت اتخاذ کنند. اتخاذ چنین تصمیماتی ممکن است به مقدار زیادی داده، اطلاعات و دانش مربوطه نیاز داشته باشد. اینها در چارچوب تصمیمات لازم باید به سرعت، به دفعات متعدد در زمان واقعی پردازش شوند و معمولاً به پشتیبانی رایانه‌ای نیاز دارد. همزمان با تکامل فناوری‌ها، بسیاری از تصمیمات به صورت خودکار اتخاذ می‌شوند که به طرق مختلف بر کار دانشی و کارگران دانشی تأثیر می‌گذارد.

موضوع این کتاب، استفاده از تجزیه و تحلیل کسب‌وکار و هوش مصنوعی به عنوان یک پورتفولیوی پشتیبانی رایانه‌ای برای تصمیم‌گیری مدیریتی است. تمرکز این کتاب بر مبانی نظری و مفهومی پشتیبانی تصمیم‌گیری و همچنین بر روی ابزارها و تکنیک‌های تجاری موجود می‌باشد. این کتاب به معرفی مبانی تکنیک‌ها و نحوه ساخت و بکارگیری این سیستم‌ها می‌پردازد. ما برای معرفی این موضوعات از یک روش تریپل-ئی (مواجهه، تجربه و اکتشاف)^۱ پیروی کرده‌ایم. این کتاب اساساً به مواجهه با تکنیک‌های مختلف تجزیه و تحلیل / هوش مصنوعی و کاربردهای آنها می‌پردازد. تفکر اصلی این است تا دانشجویان یاد بگیرند که سازمان‌های مختلف چگونه از این فناوری‌ها برای تصمیم‌گیری یا دستیابی به برتری رقابتی استفاده می‌کنند. ما معتقدیم که مواجهه با آنچه از طریق تجزیه و تحلیل بدست می‌آید و نحوه دستیابی به آن کلید اصلی یادگیری در مورد تجزیه و تحلیل است. در توضیح این تکنیک‌ها، ما همچنین مثال‌هایی از ابزارهای نرم‌افزاری خاص که می‌توان برای توسعه چنین برنامه‌هایی استفاده کرد را ارائه می‌کنیم. با این حال، این کتاب به هیچ ابزار نرم‌افزاری خاصی محدود نمی‌شود، بنابراین دانشجویان می‌توانند این تکنیک‌ها را با استفاده از هر تعداد ابزار نرم‌افزاری موجود تجربه کنند. ما امیدواریم که این مواجهه و تجربه بتواند موجب

¹ Exposure, Experience, Exploration (EEE)

انگیزش خوانندگان برای اکتشاف توانمندی‌های بالقوه‌ی این تکنیک‌ها در حوزه‌های تخصصی‌شان شود. به منظور تسهیل چنین اکتشافی، ما تمرین‌هایی را در نظر گرفته‌ایم که خوانندگان را به شبکه دانشگاه ترادیتا (تی یو ان)^۱ و سایت‌های دیگری که شامل تمرین‌های تیمی هستند ارجاع می‌دهد. بر اساس تجربه‌ی آموزشی ما، پروژه‌های کلاسی که دانشجویان پس از مواجهه با حجم انبوهی از مفاهیم و کاربردهای ارائه شده در این کتاب در آنها مشارکت داشته‌اند و نرم‌افزاری خاصی که استاد کلاس معرفی نموده و آن را بکار گرفته‌اند موجب تسهیل اینگونه اکتشاف‌ها شده است.

در این فصل ما به معرفی تجزیه و تحلیل‌ها و هوش مصنوعی و همچنین مرور کلی کتاب می‌پردازیم. ساختار این فصل به شرح زیر است:

- ۱-۱ پرونده‌ی آغازین: چگونه سیستم‌های هوشمند برای آسانسورها و بالابرها شرکت گنه^۲ کار می‌کنند؟
- ۱-۲ محیط‌های کسب‌وکار متغیر و نیازهای روزافزون برای پشتیبان تصمیم و تجزیه و تحلیل.
- ۱-۳ فرایندهای تصمیم‌گیری و چارچوب رایانه‌ای پشتیبان تصمیم
- ۱-۴ رشد و تکامل پشتیبان تصمیم به سمت هوش تجاری/ علم تجزیه و تحلیل/ علم داده‌ها
- ۱-۵ مروری بر علم تجزیه و تحلیل
- ۱-۶ مثال‌هایی از تجزیه و تحلیل در حوزه‌های منتخب
- ۱-۷ مروری بر هوش مصنوعی
- ۱-۸ همگرایی هوش مصنوعی و تحلیل
- ۱-۹ مروری بر اکوسیستم تجزیه و تحلیل
- ۱-۱۰ ساختار کتاب
- ۱-۱۱ منابع، پیوندها و وب سایت شبکه دانشگاه ترادیتا

۱/۱ چگونه سیستم‌های هوشمند برای آسانسورها و بالابرها شرکت گنه کار می‌کنند؟

شرکت گنه یک شرکت صنعتی بین‌المللی در فنلاند است که عمدتاً به تولید آسانسورها و بالابرها پرداخته و همچنین به بیش از ۱.۱ میلیون آسانسور، بالابر و تجهیزات مربوطه در چندین کشور خدمت‌رسانی می‌نماید. این شرکت بیش از ۵۰,۰۰۰ نفر کارمند دارد.

روزانه بیش از ۱ میلیارد نفر از آسانسورها و بالابرها تولیدی شرکت گنه استفاده می‌کنند. اگر این تجهیزات به درستی کار نکنند، مردم ممکن است دیر به محل کارشان برسند، به موقع به منزل نرسند و ممکن است جلسات و رویدادهای مهم را از دست بدهند. بنابراین، هدف شرکت گنه این است که زمان خرابی و دردسرهای کاربران را به حداقل برساند.

¹ Teradata University Network (TUN)

² KONE

این شرکت بیش از ۲۰،۰۰۰ تکنیسین دارد که هر زمان که مشکلی برای آسانسورها رخ می‌دهد برای خدمت‌رسانی اعزام می‌شوند. با افزایش ارتفاع ساختمان‌ها که در بسیاری از مکان‌ها شاهد آن هستیم، تعداد بیشتری از مردم از آسانسورها استفاده می‌کنند و آسانسورها با فشار بیشتری برای مدیریت ترافیک روزافزون مردم مواجه هستند. شرکت کُنه با مسئولیت پاسخگویی ایمن تر و روان تر به استفاده‌کنندگان از آسانسورها روبرو شد.

شرکت کُنه تصمیم گرفت تا از پلتفرم ابری اینترنت اشیا واتسون آی‌بی‌ام^۱ استفاده کند. همانطور که در فصل ۶ خواهیم دید، آی‌بی‌ام قابلیت‌های شناختی خود را در ساختمان‌ها نصب کرده است که امکان تشخیص موقعیت‌ها و رفتارهای انسان‌ها و تجهیزات را فراهم می‌آورد. اینترنت اشیا که در فصل ۱۳ به بررسی آن خواهیم پرداخت، پلتفرمی است که می‌تواند میلیون‌ها «شیء» را به یکدیگر و به یک مرکز فرماندهی که قادر به مدیریت اشیا متصل به یکدیگر است وصل نماید. همچنین، اینترنت اشیا، حسگرهایی که به آسانسورها و بالابرهای شرکت کُنه متصل هستند را به یکدیگر وصل می‌کند. حسگرها، اطلاعات و داده‌های مربوط به آسانسورها (مثلاً سطح نویز) و سایر تجهیزات را به طور آنی و لحظه‌ای گردآوری می‌کنند. سپس اینترنت اشیا، این اطلاعات را از طریق «اَبَر» داده‌های گردآوری شده به مراکز اطلاعاتی ارسال می‌کند. در آنجا، سیستم‌های تحلیل (موتور تحلیل پیشرفته آی‌بی‌ام) و هوش مصنوعی به پردازش داده‌های گردآوری شده پرداخته و چیزهایی از قبیل خرابی‌های بالقوه را پیش‌بینی می‌کنند. این سیستم همچنین به شناسایی علل احتمالی مشکلات پرداخته و راه‌حل‌های بالقوه را توصیه می‌کند. به توان پیش‌بینی «ابزار تحلیلی واتسون آی‌بی‌ام^۲» برای یافتن مشکلات پیش از آن که رخ دهند توجه داشته باشید.

سیستم کُنه حجم زیادی از داده‌ها را گردآوری می‌کند که این داده‌ها برای مقاصد دیگری که موجب بهبود طراحی تجهیزات در آینده می‌شود مورد تجزیه و تحلیل قرار می‌گیرند. این بدان خاطر است که ابزار تحلیلی واتسون محیط راحتی را برای برقراری ارتباط و همکاری در خصوص داده‌ها فراهم می‌کند. علاوه بر این، تحلیل‌ها توصیه‌هایی را به منظور بهینه‌سازی عملکرد ساختمان‌ها و تجهیزات ارائه می‌کنند. در پایان، شرکت کُنه و مشتریان می‌توانند به دیدگاه‌هایی از منظر مالی مدیریت آسانسورها دست یابند. کُنه قابلیت‌های واتسون را با ابزارهای سرویس بخش فروش خود (سرویس کلاود لایت‌نینگ^۳ و فیلد سرویس لایت‌نینگ^۴) ادغام نموده است. این ترکیب به کُنه کمک می‌کند تا بلافاصله در کمترین زمان ممکن با اعزام چندین تکنیسین از مجموعه‌ی ۲۰،۰۰۰ نفری خود به محل وقوع مشکل، به شرایط اضطراری یا خرابی‌های قریب الوقوع پاسخ دهد.

بخش فروش نیز مدیریت ارتباط با مشتریان را به بهترین نحو ممکن انجام می‌دهد. ارتباط بین انسان-ماشین، پرس و جو و همکاری در این سیستم به زبان طبیعی (یکی از قابلیت‌های هوش مصنوعی ابزار تحلیلی واتسون) انجام می‌شود. توجه داشته باشید که ابزار تحلیلی واتسون آی‌بی‌ام شامل دو نوع تحلیل می‌باشد: پیش‌گویانه^۵ که زمان وقوع خرابی را پیش‌بینی می‌کند و تجویزی^۶ که اقداماتی را توصیه می‌کند (مثلاً نگهداری و تعمیرات پیشگیرانه).

¹ IBM Watson IoT Cloud Platform

² IBM Watson Analytics

³ Service Cloud Lightning

⁴ Field Service Lightning

⁵ Predictive

⁶ Prescriptive

شرکت کُنه زمان خرابی خود را به حداقل رسانده و زمان تعمیرات را نیز کوتاهتر کرده است. مشخصاً، کاربران آسانسورها/بالابرها اگر مشکلات خرابی تجهیزات را نداشته باشند خوشحال تر خواهند بود و می‌توانند از این تردهای بین دردسر لذت ببرند. پیش‌بینی « قریب الوقوع » می‌تواند از بروز بسیاری از مشکلات برای مالکان تجهیزات جلوگیری نماید. مالکان همچنین می‌توانند جدول زمانبندی حضور کارمندانشان را بهینه‌سازی کنند (مثلاً کارگران نگهداری و تعمیرات و کارگران خدمات تمیزکاری). از همه اینها گذشته، تصمیم‌گیرندگان هم در شرکت کُنه و هم در ساختمان‌ها می‌توانند تصمیمات بهتر و آگاهانه‌تری را اتخاذ کنند. در آینده‌ای نه چندان دور، ربات‌ها ممکن است نگهداری و تعمیرات آسانسورها و بالابرها را انجام دهند.

[youtube.com/watch?v=6UPJHyjJft0](https://www.youtube.com/watch?v=6UPJHyjJft0) (1:31 min.) (2017)

[youtube.com/watch?v=EVbd3ejEXus](https://www.youtube.com/watch?v=EVbd3ejEXus) (2:49 min.) (2017)

سوالاتی برای پرونده‌ی آغازین

۱. گفته می‌شود که شرکت کُنه از هوشمندی در سرتاسر زنجیره تأمین خود استفاده کرده و ساختمان‌های هوشمندتر را بوجود آورده است. توضیح دهید.
۲. نقش اینترنت اشیاء را در این پرونده توضیح دهید.
۳. ضرورت استفاده از واتسون آی‌بی‌ام در این پرونده چیست؟
۴. ابزار تحلیلی پیشرفته‌ی آی‌بی‌ام را بررسی کنید. چه ابزارهایی وجود دارند که به این پرونده مرتبط می‌شوند؟
۵. ساختمان‌های شناختی آی‌بی‌ام را بررسی کنید. ارتباط آنها با این پرونده چیست؟

از این پرونده چه می‌توانیم بیاموزیم؟

امروزه فناوری‌های هوشمند را به صورت ترکیبی با اینترنت اشیاء می‌توان در پروژه‌های پیچیده و بزرگ بکار گرفت. قابلیت‌های پلتفرم‌های هوشمند یکپارچه از قبیل واتسون آی‌بی‌ام، حل مشکلاتی که تا چند سال پیش از لحاظ اقتصادی و تکنولوژیکی غیرقابل حل بودند را امکان‌پذیر ساخته است. در این پرونده، چندین فناوری شامل تحلیل‌های پیشرفته، حسگرها، اینترنت اشیاء و هوش مصنوعی به خواننده معرفی شده که در این کتاب به تفصیل شرح داده خواهد شد. این پرونده همچنین به کاربرد فناوری « ابر » اشاره می‌کند. فناوری ابر برای پردازش متمرکز حجم زیادی از اطلاعات با استفاده از تحلیل‌ها و الگوریتم‌های هوش مصنوعی من جمله « اشیاء » در مکان‌ها و موقعیت‌های مختلف بکار می‌رود. این پرونده همچنین دو نوع اصلی تحلیل‌ها را به ما معرفی می‌کند: تحلیل‌های پیش‌گویانه (فصل‌های ۴ تا ۶) و تحلیل‌های تجویزی (فصل ۸). چندین فناوری هوش مصنوعی در اینجا بحث شده است: یادگیری ماشین، پردازش زبان طبیعی، بینایی رایانه‌ای و تجزیه و تحلیل تجویزی.

این پرونده، مثالی از هوش افزوده است که در آن، انسان‌ها و ماشین‌ها با یکدیگر کار می‌کنند. این پرونده مزایای هوش افزوده برای فروشندگان، شرکت‌های اجرا کننده و کارکنانشان و همچنین کاربران آسانسورها و بالابرها را به تصویر کشیده است.

۱/۲ محیط‌های کسب‌وکار متغیر و نیازهای روزافزون برای پشتیبانی از تصمیم و تجزیه و تحلیل‌ها

تصمیم‌گیری یکی از مهمترین فعالیتهای و شاید مهمترین فعالیت در تمامی سازمان‌ها می‌باشد. تصمیم‌گیری منجر به موفقیت یا شکست سازمان‌ها و چگونگی عملکرد آنها می‌شود. تصمیم‌گیری به دلیل عوامل داخلی و خارجی دشوارتر شده است. اتخاذ تصمیمات درست می‌تواند پاداش‌های ارزشمندی به همراه داشته باشد در حالیکه اتخاذ تصمیمات نادرست نیز می‌تواند زیان‌های جبران‌ناپذیری داشته باشند.

متأسفانه، تصمیم‌گیری کار آسانی نیست. چندین نوع تصمیم وجود دارد که هر یک نیازمند یک فرایند تصمیم‌گیری متفاوت هستند. به عنوان مثال، دی‌اس‌مت^۱ و همکاران^۱ (۲۰۱۷) از گروه مشاوران مدیریت شرکت مک‌کنزی آند کامپانی^۲ تصمیمات سازمانی را به چهار گروه زیر تقسیم می‌کنند:

- تصمیمات با ریسک بالا و بازده بالا
- تصمیمات متقاطع که تکرارپذیر ولی با ریسک بالا هستند و نیازمند کار گروهی می‌باشند (فصل ۱۱).
- تصمیمات اقتضایی که به صورت دوره‌ای مطرح می‌شوند.
- تصمیمات تفویض‌شده به افراد یا گروه‌های کوچک

بنابراین، ابتدا لازم است تا ماهیت تصمیم‌گیری را درک کنیم. برای بحث جامع در این خصوص به دی‌اس‌مت و همکاران (۲۰۱۷) رجوع نمایید.

کسب‌وکارهای امروزی مملو از تغییرات سریع و عدم قطعیت‌ها هستند. برای حل این مشکل، تصمیم‌گیرندگان سازمانی می‌بایست با داده‌های متغیر که هر روز در حال افزایش هستند دست و پنجه نرم کنند. این کتاب در مورد فناوری‌هایی است که به تصمیم‌گیرندگان در انجام این کار کمک می‌کند.

برای سالیان متمادی، تصمیم‌گیری از نظر مدیران یک هنر محض محسوب می‌شد که نیازمند کسب استعداد از طریق تجربه در یک دوره‌ی طولانی مدت (یعنی یادگیری بر اساس آزمون و خطا) و با استفاده از شهود و الهام بوده است. مدیریت به دلیل تنوع سبک‌های فردی که در بررسی و حل موفقیت‌آمیز مسائل و مشکلات مدیریتی یکسان بکار گرفته می‌شدند یک هنر محسوب می‌شد. این سبک‌ها به جای روش‌های کمی سیستماتیک بر پایه‌ی یک روش علمی، اغلب مبتنی بر خلاقیت، قضاوت، الهام (شهود) و تجربه بودند. با وجود این، پژوهش‌های اخیر بیانگر این هستند که شرکت‌هایی که مدیران ارشد آنها بیشتر بر کار مداوم و مصرانه تمرکز دارند نسبت به رهبرانی که نقاط قوت اصلی آنها مهارت‌های ارتباطی بین فردی است عملکرد بهتری دارند. مدیران معمولاً از طریق یک فرایند چهار مرحله‌ای تصمیم‌گیری می‌کنند:

تعریف مسئله (یعنی وضعیت تصمیم که ممکن است دشواری یا فرصتی را به همراه داشته باشد).

- ایجاد یک مدل که بیانگر مسئله در دنیای واقعی است.
- شناسایی راه‌حل‌های ممکن برای مسئله‌ی مدل‌سازی شده و ارزیابی راه‌حل‌ها.

¹ De Smet et al.

² McKinsey & Company

- مقایسه، انتخاب و توصیه‌ی یک راه‌حل بالقوه برای مسئله.

یک فرایند تفصیلی‌تر با گام‌های زیر توسط کوئین (۲۰۱۸) ارائه شده است:

- تصمیمی که می‌خواهید اتخاذ کنید را درک کنید.
- تمامی اطلاعات را گردآوری کنید.
- راه‌حل‌های جایگزین را شناسایی کنید.
- مزایا و معیاب آنها را شناسایی کنید.
- بهترین راه‌حل را انتخاب کنید.
- تصمیم بگیرید.
- تأثیر تصمیم‌تان را ارزیابی کنید.

ما در بخش ۱.۳ مجدداً به این فرایند رجوع خواهیم کرد.

برای استفاده از این فرایندهای تصمیم‌گیری باید اطمینان حاصل کنیم که راه‌حل‌های جایگزین و به ویژه راه‌حل‌های خوب به تعداد کافی در نظر گرفته شده است تا بتوان پیامدهای استفاده از این جایگزین‌ها را به شکلی منطقی پیش‌بینی کرده و به طور مناسب بتوان آنها را مقایسه نمود. با وجود این، تغییرات سریع در محیط‌های داخلی و خارجی موجب دشوار شدن فرایند ارزیابی به دلایل زیر شده است:

- فناوری، سیستم‌های اطلاعاتی، موتورهای جستجوی پیشرفته و جهانی شدن، گزینه‌های بیشتری را برای انتخاب به وجود می‌آورند.
- مقررات دولتی و نیاز به انطباق، بی‌ثباتی سیاسی و تروریسم، رقابت و تغییر تقاضای مصرف‌کننده عدم اطمینان بیشتری ایجاد می‌کند، پیش‌بینی عواقب و آینده را دشوارتر می‌کند.
- **عوامل سیاسی.** تصمیمات عمده ممکن است تحت تأثیر سیاست خارجی داخلی باشد. به عنوان نمونه می‌توان به جنگ تجاری تعرفه‌ها در سال ۲۰۱۸ اشاره کرد.
- **عوامل اقتصادی.** اینها از رقابت گرفته تا جنس و وضعیت اقتصاد متغیر است. این عوامل چه در کوتاه مدت و چه در طولانی مدت باید مورد توجه قرار گیرند.
- **عوامل جامعه‌شناختی و روان‌شناختی** کارمندان و مشتریان. این موارد باید در هنگام ایجاد تغییرات مورد توجه قرار گیرند.
- **عوامل محیطی.** تأثیر بر محیط فیزیکی باید در بسیاری از موقعیت‌های تصمیم‌گیری مورد ارزیابی قرار گیرد.

از دیگر عوامل می‌توان به نیاز به تصمیم‌گیری سریع، تغییرات مکرر و غیرقابل پیش‌بینی که یادگیری آزمون و خطا را دشوار می‌کند و هزینه‌های احتمالی اشتباهات که ممکن است زیاد باشد اشاره کرد. هر روز بر پیچیدگی این محیط‌ها افزوده می‌شود.

بنابراین، امروزه تصمیم‌گیری واقعاً کار پیچیده‌ای شده است. برای بحث بیشتر، به چارلز^۱ (۲۰۱۸) مراجعه کنید. برای چگونگی تصمیم‌گیری موثر تحت عدم اطمینان و فشار به زین^۲ (۲۰۱۶) مراجعه کنید. به دلیل این روندها و تغییرات، تقریباً غیرممکن است که به یک رویکرد آزمون و خطای صرف در مدیریت اعتماد کنیم. مدیران باید پیچیده‌تر باشند. آنها باید از ابزارها و فنون جدید رشته‌های خود استفاده کنند. بیشتر این ابزارها و فنون در این کتاب بحث شده است. استفاده از آنها برای حمایت از تصمیم‌گیری می‌تواند در تصمیم‌گیری‌های موثر بسیار سودمند باشد. به علاوه، بسیاری از ابزارهایی که در حال تکامل هستند حتی ماهیت چندین وظیفه تصمیم‌گیری که به صورت خودکار انجام می‌شوند را تحت تأثیر قرار می‌دهند. این امر بر تقاضای آینده کارگران دانشی تأثیر گذارده و نیازمند رسیدگی به بسیاری از مسائل حقوقی و اجتماعی تأثیرگذار خواهد بود.

در بخش‌های مختلف این کتاب خواهیم دید که چگونه کل صنعت می‌تواند از تحلیل‌ها به منظور تهیه گزارشات در مورد آنچه در حال رخ دادن است استفاده کرده و سپس برای استفاده‌ی بهینه از موقعیت موجود، تصمیم‌گیری نماید. این گام‌ها نیازمند این است که یک سازمان، انبوهی از داده‌ها را جمع‌آوری و تحلیل نماید. به طور کلی، حجم داده‌ها هر دو سال یکبار دو برابر می‌شود. از کاربردهای سنتی در توزیع حقوق و دستمزد و حسابداری که بگذریم، امروزه از سیستم‌های رایانه‌ای در حوزه‌های مدیریتی پیچیده اعم از طراحی و مدیریت کارخانجات خودکار گرفته تا بکارگیری روش‌های تحلیلی برای ارزیابی ادغام‌ها و اکتساب‌های پیشنهادی استفاده می‌شود. تقریباً تمامی مدیران از نقش حیاتی فناوری اطلاعات در کسب‌وکارهایشان آگاه بوده و این فناوری‌ها را به طور گسترده‌ای بکار می‌گیرند. برنامه‌های رایانه‌ای از فعالیت‌های پایش و پردازش تراکنش‌ها به سمت تجزیه و تحلیل مسئله و بسته‌های نرم‌افزاری جامع حرکت کرده و بیشتر فعالیت‌ها از طریق فناوری‌های مبتنی بر ابر و در بسیاری از موارد از طریق دستگاه‌های سیار انجام می‌پذیرد. ابزارهای تحلیل و هوش تجاری مانند انبارش داده‌ها، داده‌کاوی، پردازش تحلیلی آنلاین^۳، داشبوردها و استفاده از سیستم‌های مبتنی بر ابر برای پشتیبان تصمیم‌گیری، سنگ بنای مدیریت پیشرفته‌ی امروزی هستند. مدیران باید سیستم‌های اطلاعاتی با سرعت بالا و شبکه‌بندی شده‌ای (سیمی یا بی‌سیم) داشته باشند تا به آنها در انجام مهم‌ترین وظیفه‌شان یعنی تصمیم‌گیری یاری رساند. در بسیاری از موارد، اینگونه تصمیمات به طور معمول به صورت کاملاً خودکار در می‌آیند (به فصل ۲ مراجعه کنید) و دیگر نیازی به هیچ‌گونه مداخله مدیریتی نمی‌باشد.

علاوه بر رشد آشکار قابلیت‌های سخت‌افزاری، نرم‌افزاری و شبکه‌ای، برخی از دستاوردهای توسعه‌ای آشکارا موجب تسهیل رشد و شکوفایی فناوری‌های تجزیه و تحلیل و پشتیبان تصمیم به روش‌های متعددی شده‌اند:

- **همکاری و ارتباطات گروهی.** امروزه بسیاری از تصمیمات توسط گروه‌های گرفته می‌شوند که اعضای آنها ممکن است در مکان‌های متفاوتی باشند. گروه‌ها به آسانی و با استفاده از ابزارهای همکاری و همچنین گوشی‌های هوشمند مجهز به فناوری‌های همه‌جای حاضر می‌توانند با یکدیگر ارتباط برقرار کرده و همکاری داشته باشند. موضوع همکاری به ویژه در زنجیره‌های تأمین که شرکا - از تأمین‌کننده گرفته تا مشتریان - می‌بایست اطلاعات را به اشتراک بگذارند اهمیت

¹ Charles

² Zane

³ Online Analytical Processing (OLAP)

پیدا می‌کند. گردهم آوردن گروهی از تصمیم‌گیرندگان و بویژه افراد خبره در یک مکان می‌تواند هزینه‌بر باشد. سیستم‌های اطلاعاتی می‌توانند فرایند همکاری گروهی را بویژه زمانیکه اعضای این گروه در مکان‌های مختلفی هستند بهبود بخشند (صرفه‌جویی در هزینه‌های سفر). مهمتر اینکه، اینگونه همکاری‌ها در زنجیره‌های تأمین به سازندگان اجازه می‌دهد تا از تغییرات در الگوهای تقاضا در زمان واقعی مطلع شده و بدین ترتیب، نسبت به تغییرات بازار سریع‌تر واکنش نشان دهند. برای بحث جامع در خصوص هوش مصنوعی و تأثیر آن به فصل‌های ۲، ۱۰ و ۱۴ رجوع نمایید.

- **بهبود مدیریت داده‌ها.** بسیاری از تصمیمات شامل محاسبات پیچیده هستند. داده‌های مورد نیاز برای اینکار را می‌توان در پایگاه داده‌های مختلف در هر جایی از سازمان و حتی در خارج از سازمان ذخیره نمود. بسیاری مواقع لازم است تا داده‌ها را سریعاً از مکان‌های دور ارسال نمود. سیستم‌های امروزی می‌توانند داده‌های مورد نیاز را با سرعت، امنیت و شفافیت بالا جستجو، ذخیره و ارسال کنند. برای جزئیات بیشتر به فصل‌های ۳ و ۹ و فصل آنلاین رجوع کنید.
- **مدیریت انبارهای داده غول آسا و کلان داده‌ها.** انبارهای داده بزرگ همانند آنهایی که در اختیار وال‌مارت هستند حاوی حجم انبوهی از داده‌ها می‌باشند. برای سازماندهی، جستجو و کاوش این داده‌ها از روش‌های ویژه‌ای من جمله رایانش موازی و هادوپ/اسپارک^۱ استفاده می‌شود. هزینه‌های انبارش و کاوش داده‌ها به سرعت در حال کاهش هستند. فناوری‌هایی که تحت دسته‌بندی گسترده‌ی کلان‌داده‌ها قرار می‌گیرند گردآوری و انبارش حجم انبوه داده‌هایی که از منابع متعدد و به شکل‌های مختلفی می‌آیند را ممکن ساخته‌اند که بدین ترتیب عملکرد سازمانی را می‌توان به شکلی که در گذشته ممکن نبوده است مشاهده کرد. برای جزئیات بیشتر به فصل ۹ مراجعه کنید.
- **پشتیبانی تحلیلی.** با افزایش فناوری‌های داده‌ای و تحلیلی، می‌توان جایگزین‌های بیشتری را مورد ارزیابی قرار داده، پیش‌بینی‌ها را بهبود بخشیده، تحلیل ریسک را سریع‌تر انجام داده و نظرات افراد خبره را (که برخی از آنها ممکن است در مکان‌های دور باشند) با سرعت بیشتر و هزینه کمتری گردآوری نمود. تخصص را حتی می‌توان مستقیماً از سیستم‌های تحلیلی استخراج نمود. با این ابزارها، تصمیم‌گیرندگان می‌توانند شبیه‌سازی‌های پیچیده را انجام داده، تعداد زیادی از سناریوهای ممکن را بررسی کرده و تأثیرات گوناگون آنها را به سرعت و به شکلی اقتصادی مورد ارزیابی قرار دهند. این البته موضوع اصلی بحث ما در فصل‌های مختلف کتاب می‌باشد.
- **غلبه بر محدودیت‌های شناختی در پردازش و انبارش اطلاعات.** ذهن انسان توانایی محدودی برای پردازش و انبارش اطلاعات دارد. به دلیل محدودیت‌های شناختی، انسان‌ها برخی اوقات در یادآوری و بکارگیری بدون خطای اطلاعات دچار مشکل می‌شوند. عبارت محدودیت‌های شناختی^۲ بیانگر این است که زمانیکه به دانش و اطلاعات بسیار متنوعی نیاز باشد، توانایی حل مسئله در فرد محدود می‌شود. سیستم‌های یارانه‌ای به انسان‌ها کمک می‌کنند تا با ارزیابی و پردازش سریع حجم زیادی از اطلاعات انبارش شده، برای محدودیت‌های شناختی‌شان فائق آیند. یک روش برای غلبه بر محدودیت‌های شناختی انسان، استفاده از پشتیبانی هوش مصنوعی است. برای اطلاعات بیشتر در مورد جنبه‌های شناختی به فصل ۶ رجوع کنید.
- **مدیریت دانش.** سازمان‌ها انبارهای عظیمی از اطلاعات مربوط به عملیات، مشتریان، رویه‌های داخلی، تعاملات کارکنان و ... خودشان را از طریق ارتباطات ساخت‌یافته و ساخت‌نیافته‌ای که مابین ذینفعان مختلف رخ می‌دهد ایجاد کرده‌اند. سیستم‌های مدیریت دانش تبدیل به منابع پشتیبانی رسمی و غیررسمی برای تصمیم‌گیری مدیران شده‌اند اگرچه

¹ Hadoop/Spark

² Cognitive limits

برخی اوقات ممکن است نتوانیم آنها را سیستم‌های مدیریت دانش^۱ بنامیم. فناوری‌هایی از قبیل تجزیه و تحلیل متن و واتسون آی بی ام خلق ارزش از چنین ابزارهای دانشی را امکان‌پذیر ساخته‌اند. (برای جزئیات بیشتر به فصل ۶ و ۱۲ مراجعه کنید).

- **پشتیبانی هر وقت و هر کجا.** مدیران با استفاده از فناوری بی‌سیم می‌توانند در هر زمان و از هر مکان به اطلاعات دسترسی یافته، آنها را تحلیل و تفسیر کرده و با آنهایی که این اطلاعات را بکار می‌گیرند ارتباط برقرار کنند. سرعت پردازش این اطلاعات و تبدیل آنها را تصمیمات حقیقتاً انتظارات مشتریان و کسب‌وکارها را دستخوش تغییر کرده است. اینها به همراه سایر قابلیت‌ها عامل محرکی برای استفاده از پشتیبانی تصمیم‌یارانه‌ای از اواخر دهه ۶۰ و بویژه از اواسط دهه ۹۰ شده‌اند. رشد فناوری‌های سیار، پلتفرم‌های رسانه‌های اجتماعی و ابزارهای تحلیلی سطح متفاوتی از سیستم‌های اطلاعاتی را برای پشتیبانی از مدیران فراهم کرده است. این رشد پشتیبانی از داده‌ها برای هرگونه تصمیم تنها محدود به مدیران نبوده بلکه به مصرف‌کنندگان نیز گسترش یافته است. ما در ابتدا به مرور فناوری‌هایی که به طور گسترده‌ای هوش تجاری نامیده می‌شوند می‌پردازیم. از آنها افق دیدمان را گسترده‌تر کرده و به معرفی انواع مختلف تجزیه و تحلیل‌ها خواهیم پرداخت.
- **نوآوری و هوش مصنوعی.** با توجه به پیچیدگی‌های فرایند تصمیم‌گیری که پیش از این بحث شد و محیط پیرامون این فرایند، به رویکرد نوآورانه‌تری نیاز می‌باشد. یکی از تسهیلات مهمی که نوآوری به همراه دارد، به واسطه‌ی هوش مصنوعی فراهم می‌شود. هوش مصنوعی تقریباً بر تمامی گام‌های فرایند تصمیم‌گیری می‌تواند تأثیر بگذارد. هوش مصنوعی همچنین با ابزارهای تجزیه و تحلیل یکپارچه شده و موجب خلق هم‌افزایی در تصمیم‌گیری شده است.

سوالاتی برای مرور بخش ۱.۲

۱. چرا اتخاذ تصمیمات سازمانی کار دشواری است؟
۲. گام‌های اصلی فرایند تصمیم‌گیری را شرح دهید.
۳. محیط‌های خارجی اصلی که می‌توانند بر تصمیم‌گیری تأثیر بگذارند را شرح دهید؟
۴. برخی از روندهای سیستم‌محور مهم که تصمیم‌گیری مبتنی بر سیستم‌های اطلاعاتی را به سطح جدیدی ارتقاء داده‌اند کدام هستند؟
۵. برخی از قابلیت‌های فناوری اطلاعات که می‌توانند تصمیم‌گیری مدیریتی را تسهیل نمایند کدام هستند؟

۱.۳ فرایندهای تصمیم‌گیری و چارچوب پشتیبانی تصمیم رایانه‌ای

در این بخش، تمرکز ما بر برخی از اصول و مبانی کلاسیک تصمیم‌گیری و به طور دقیق‌تر بر فرایند تصمیم‌گیری خواهد بود. این دو مفهوم به ما کمک خواهند کرد تا زیربنای بخش زیادی از آنچه را از منظر تجزیه و تحلیل، علم داده‌ها و هوش مصنوعی یاد خواهیم گرفت را بنا نهیم. تصمیم‌گیری یک فرایند انتخاب از بین دو یا چند اقدام جایگزین برای دستیابی به یک یا چند هدف است. طبق سیمون^۲ (۱۹۷۷)، تصمیم‌گیری مدیریتی مترادف با کل فرایند مدیریت است. کارکرد مدیریتی مهم برنامه‌ریزی را در نظر بگیرید. برنامه‌ریزی شامل مجموعه‌ای از تصمیمات است: چه کاری می‌بایست انجام شود؟ چه زمان؟ کجا؟ چرا؟ چگونه؟

¹ Knowledge Management Systems (KMS)

² Simon

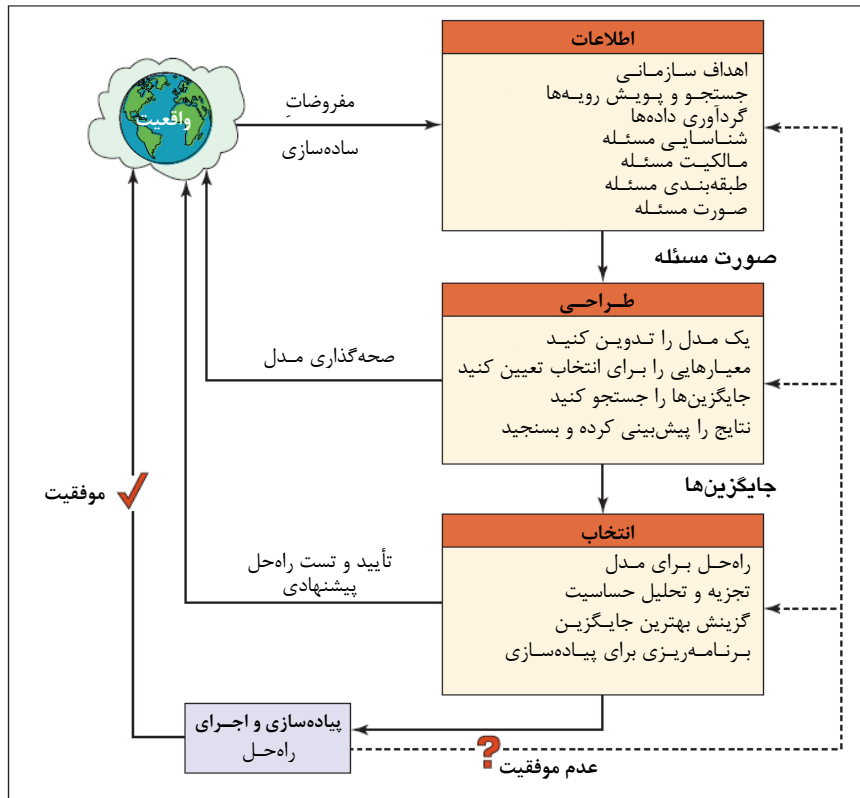
توسط چه کسی؟ مدیران وظیفه‌ی تعیین اهداف یا برنامه را بر عهده دارند؛ از این رو، برنامه‌ریزی به معنای تصمیم‌گیری است. سایر وظایف مدیریتی از قبیل سازماندهی و کنترل نیز شامل تصمیم‌گیری هستند.

توصیه می‌شود تا برای تصمیم‌گیری از یک فرایند سیستماتیک پیروی شود. سیمون (۱۹۷۷) گفته است که این فرایند از سه فاز اصلی تشکیل شده است: اطلاعات، طراحی و انتخاب. او بعدها فاز چهارمی را نیز افزود که پیاده‌سازی نام داشت. پایش را می‌توان به عنوان فاز پنجم در نظر گرفت که شکلی از بازخورد است. با وجود این، ما پایش را به عنوان فاز اطلاعات در نظر می‌گیریم که بر فاز پیاده‌سازی اعمال شده باشد. مدل سیمون، مختصرترین و در عین حال کامل‌ترین مدل تصمیم‌گیری منطقی است. یک تصویر مفهومی از فرایند تصمیم‌گیری در شکل ۱.۱ نشان داده شده است. همچنین به عنوان یک روش پشتیبان تصمیم با استفاده از مدلسازی به تصویر کشیده شده است. جریان مستمر فعالیت از فاز اطلاعات به فاز طراحی و انتخاب وجود دارد (خطوط پررنگ در شکل ۱.۱) ولی در هر فاز، ممکن است به فاز قبل بازگردیم (بازخورد). مدل‌سازی یک بخش ضروری از این فرایند است. ماهیت به ظاهر بی‌نظم پیروی از یک مسیر تصادفی از کشف مسئله تا راه‌حل از طریق تصمیم‌گیری را می‌توان توسط این حلقه‌های بازخورد توضیح داد.

فرایند تصمیم‌گیری با فاز اطلاعات آغاز می‌شود. در این فاز، تصمیم‌گیرنده به بررسی واقعیت پرداخته و مسئله را تعریف می‌کند. در اینجا مالکیت مسئله^۱ نیز تعیین می‌شود. در فاز طراحی، مدلی که بیانگر سیستم از ایجاد می‌شود. اینکار با در نظر گرفتن مفروضاتی که موجب ساده‌سازی واقعیت می‌شوند و با نوشتن روابط بین تمامی متغیرها انجام می‌شود. سپس این مدل اعتبارسنجی شده و معیارهایی به منظور ارزشیابی اقدامات جایگزین که شناسایی شده‌اند تعیین می‌شوند. اغلب، در فرایند توسعه‌ی مدل، راه‌های جایگزین شناسایی می‌شوند و بر عکس.

فاز انتخاب شامل گزینش یک راه‌حل پیشنهادی برای مدل است (نه لزوماً راه‌حلی برای مسئله‌ای که بیانگر آن است). این راه‌حل به منظور تعیین عملی بودن (کارایی) تست می‌شود. زمانیکه راه‌حل پیشنهادی، منطقی به نظر برسد ما آماده‌ی فاز آخر یعنی پیاده‌سازی تصمیم (و نه لزوماً پیاده‌سازی یک سیستم) هستیم. پیاده‌سازی موفق منجر به حل مسئله‌ی واقعی می‌شود. شکست در پیاده‌سازی موجب می‌شود تا ما به یک فاز قبل‌تر در فرایند بازگردیم. در واقع، ما در خلال هر یک از سه فاز آخر فرایند می‌توانیم به یک فاز قبل‌تر بازگردیم. موقعیت‌های تصمیم‌گیری که در پرونده‌ی آغازین شرح داده شدند همانند اکثر موقعیت‌های تصمیم‌گیری دیگر از مدل چهار فازی سیمون پیروی می‌کنند.

¹ Problem Ownership



شکل ۱.۱ فرایند مدل‌سازی/تصمیم‌گیری

فاز اطلاعات با شناسایی اهداف و مقاصد سازمانی مرتبط با یک مشکل مورد نظر (مثلاً مدیریت موجودی، انتخاب شغل، عدم حضور آنلاین) و تعیین اینکه آیا این اهداف برآورده شده‌اند یا خیر آغاز می‌شود. مسائل به دلیل نارضایتی از وضعیت موجود رخ می‌دهند. نارضایتی، نتیجه‌ی تفاوت بین آنچه مردم می‌خواهند (یا انتظار دارند) و آنچه رخ می‌دهد است. در این فاز اول، تصمیم‌گیرنده تلاش می‌کند تا تعیین کند که آیا مشکلی وجود دارد، نشانه‌های آن را شناسایی کند، بزرگی آن را مشخص کند و آن را به طور صریح تعریف کند. اغلب، آنچه به عنوان یک مشکل شرح داده می‌شود (مثلاً هزینه‌های مازاد) ممکن است فقط یک نشانه (یعنی یک سنجه) از مشکل باشد (مثلاً سطح موجودی نامناسب). از آنجائیکه مشکلات دنیای واقعی معمولاً به واسطه‌ی عوامل بهم وابسته، پیچیده می‌شوند، تشخیص نشانه‌ها از مشکل واقعی برخی اوقات دشوار می‌شود. در حین تحقیق در مورد علل نشانه‌ها مطمئناً فرصت‌ها و مشکلات جدیدی کشف خواهند شد.

وجود یک مشکل را می‌توان از طریق پایش و تحلیل سطح بهره‌وری سازمان تعیین نمود. اندازه‌گیری بهره‌وری و ساخت یک مدل بر اساس داده‌های حقیقی انجام می‌شوند. گردآوری داده‌ها و تخمین داده‌های آتی از دشوارترین گام‌ها در تجزیه و تحلیل محسوب می‌شوند.

مسائل گردآوری داده‌ها در ادامه به بررسی برخی از مسائلی که ممکن است در حین گردآوری و برآورد داده‌ها بروز کرده و تصمیم‌گیرندگان را دچار مشکل کند می‌پردازیم:

- در دسترس نبودن داده‌ها. در نتیجه، مدل ما بر اساس تخمین‌های نادقیق ساخته خواهد شد.
- هزینه‌بر بودن دستیابی به داده‌ها.
- دقیق نبودن داده‌ها به اندازه‌ی کافی.
- ذهنی بودن تخمین داده‌ها در اغلب موارد
- بی‌اعتبار بودن داده‌ها.
- داده‌های مهمی که بر نتایج تأثیر می‌گذارند ممکن است ماهیت کیفی داشته باشند.
- حجم زیاد داده‌ها (بار بیش از حد اطلاعات).
- پیامدها یا نتایج ممکن است در یک دوره زمانی طولانی مدت رخ دهند. در نتیجه، درآمدها، هزینه‌ها و سود در نقاط زمانی متفاوتی ثبت خواهد شد. برای رفع این مشکل، در صورتیکه نتایج قابلیت کمی‌سازی داشته باشند می‌توان از یک روش ارزش فعلی استفاده کرد.
- فرض بر این است که داده‌های آینده مشابه داده‌های تاریخی باشد. در غیراینصورت، ماهیت تغییر را می‌بایست در تحلیل‌ها پیش‌بینی کرده و در نظر گرفت.

زمانیکه تحقیقات اولیه به پایان رسید، می‌توان تعیین کرد که آیا با یک مشکل حقیقی روبرو هستیم یا خیر، مشکل کجاست و چقدر اهمیت دارد. یک مسئله‌ی مهم این است که آیا سیستم اطلاعاتی، مشکل را به ما گزارش می‌دهد یا اینکه فقط نشانه‌های یک مشکل را به ما نشان می‌دهد. به عنوان مثال، اگر گزارشات نشان دهند که فروش پایین است، مشکلی وجود دارد ولی این وضعیت، بدون شک نشانه‌ی وجود یک مشکل است. آگاهی و شناخت مشکل حقیقی از اهمیت حیاتی برخوردار است. برخی اوقات ممکن است مشکل ادراکی، تناقض مشوق‌ها یا فرایندهای سازمانی باشد تا اینکه با یک مدل تصمیم‌گیری ضعیف مواجه باشیم. برای این نشان دهیم شناسایی صحیح مشکل تا چه اندازه اهمیت دارد یک مثال کلاسیک در مورد کاربردی ۱.۱ ارائه کرده‌ایم.

سرعت آسانسورها را بیشتر کنیم!

مورد کاربردی ۱.۱

این داستان در مکان‌های متعددی گزارش شده و تقریباً به یک مثال کلاسیک در مورد ضرورت شناسایی مشکل تبدیل شده است. آکاف^۱ (که در لارسون ذکر شده است، ۱۹۸۷) مشکل مدیریت شکایات از کند بودن آسانسورها در یک هتل چندین طبقه را توضیح داده است. پس از بکارگیری راه‌حل‌های متعدد به منظور کاهش شکایات – افزایش سرسام‌آور سرعت آسانسورها برای توقف در طبقات مختلف، افزودن تعداد اپراتورها و غیره مدیریت به این نتیجه رسید که مشکل واقعی ارتباطی به زمان انتظار واقعی نبوده بلکه زمان انتظار/ادراکی بوده است. به همین دلیل، برای حل این مشکل تصمیم گرفته شد تا آینده‌های تمام قد را بر روی درب آسانسورها در هر طبقه نصب کنند.

¹ Ackoff

طبق اظهارات هس و وولسی^۱ (۱۹۷۵)، « خانم‌ها خودشان را در آینه نگاه می‌کردند و خودشان را مرتب می‌کردند در حالیکه آقایان به خانم‌ها نگاه می‌کردند و بدون آنها متوجه زمان شوند، آسانسور سر می‌رسید.» با کاهش زمان انتظار ادراکی، مشکل نیز حل شد. بیکر و کامرون (۱۹۹۶) مثال‌های متعددی از حواس پرتی‌ها من جمله نمایشگرها و نورپردازی‌ها ارائه کرده‌اند که سازمان‌ها برای کاهش زمان انتظار ادراکی بکار می‌گیرند. اگر زمان انتظار ادراکی به عنوان مشکل واقعی شناسایی شود، در راه‌حل‌های پیشنهادی و هزینه‌هایش تفاوت زیادی وجود خواهد داشت. به عنوان مثال، آینه‌های تمام قد شاید هزینه بسیار کمتری از اضافه کردن یک آسانسور به ساختمان داشته باشند!

Sources: Based on J. Baker and M. Cameron. (1996, September). "The Effects of the Service Environment on Affect and Consumer Perception of Waiting Time: An Integrative Review and Research Propositions," *Journal of the Academy of Marketing Science*, 24, pp. 338-349; R. Hesse and G. Woolsey (1975). *Applied Management Science: A Quick and Dirty Approach*. Chicago, IL: SRA Inc; R. C. Larson. (1987, November/December). "Perspectives on Queues: Social Justice and the Psychology of Queuing." *Operations Research*, 35(6), pp. 895-905.

سوالاتی برای مورد کاربردی ۱.۱

۱. این مثال چه ارتباطی به تصمیم‌گیری دارد؟
۲. ارتباط این موقعیت با فاز اطلاعات در تصمیم‌گیری در چیست؟

دسته‌بندی مشکل دسته‌بندی مشکل در واقع به مفهومی‌سازی یک مشکل اشاره دارد تا بتوان آن را در یک گروه تعریف‌پذیر قرار داد که بتوان برای آن یک راه‌حل استاندارد ارائه کرد. یک روش مهم برای دسته‌بندی مشکلات بر اساس میزان ساختارمندی مشهود در آنهاست که از کاملاً ساخت‌یافته (یعنی برنامه‌ریزی شده) تا کاملاً غیرساخت‌یافته (یعنی برنامه‌ریزی نشده) متغیر است.

تجزیه‌ی مشکل بسیاری از مشکلات پیچیده را می‌توان به مشکلات کوچکتر تفکیک کرد. حل مشکلات کوچکتر و ساده‌تر می‌تواند به حل یک مشکل بزرگ و پیچیده کمک کند. همچنین، مشکلاتی که چندان ساخت‌یافته نیستند ممکن است به مشکلات کوچکتری تقسیم شوند که بسیار ساخت‌یافته هستند. زمانیکه برخی از فازهای تصمیم‌گیری، ساخت‌یافته و برخی دیگر غیرساخت‌یافته باشند با یک مشکل نیمه‌ساخت‌یافته مواجه هستیم و زمانیکه برخی از مشکلات کوچکتر یک مسئله‌ی تصمیم‌گیری، ساخت‌یافته و برخی غیرساخت‌یافته باشند، با یک مشکل یا مسئله‌ی نیمه‌ساخت‌یافته مواجه هستیم. با توسعه‌ی سیستم پشتیبانی تصمیم و بالا رفتن سطح آگاهی و اطلاع کارکنان حوزه توسعه سیستم و تصمیم‌گیرندگان در مورد مسئله، به تدریج ساخت‌یافته می‌شود.

مالکیت مشکل در فاز اطلاعات، تعیین مالکیت مشکل بسیار اهمیت دارد. فقط زمانی مشکلی در یک سازمان وجود دارد که فرد یا گروهی از افراد مسئولیت رسیدگی به آن را بر عهده بگیرند و سازمان توانایی حل آن را داشته باشد. تفویض اختیار برای حل یک مسئله یا مشکل را مالکیت مشکل می‌گویند. به عنوان مثال، یک مدیر ممکن است حس کند که به خاطر بالا بودن نرخ بهره مشکلی دارد. از آنجائیکه نرخ بهره‌ها در سطح ملی و بین‌المللی تعیین می‌شوند و اکثر مدیران کاری از دست‌شان بر نمی‌آید، بالا بودن نرخ بهره، مشکل دولت است و کاری از دست یک شرکت خاص برای حل آن بر نمی‌آید. مشکلی که شرکت‌ها در واقع با آن مواجه هستند این است که چگونه در محیطی با نرخ بهره بالا فعالیت کنند. برای یک شرکت، نرخ بهره باید به عنوان یک

¹ Hesse & Woolsey

عامل کنترل‌ناپذیر (محیطی) در نظر گرفته شده و پیش‌بینی شود. زمانیکه مالکیت مشکل تعیین نشده باشد، به این معنی است که یا یک نفر وظیفه‌اش را انجام نمی‌دهد یا اینکه مشکل هنوز شناسایی نشده است. سپس بسیار مهم است که یک نفر دواطلب شده و مالکیت آن را بپذیرد یا اینکه آن را به فرد دیگری واگذار کند. فاز اطلاعات با یک تعریف مسئله‌ی رسمی پایان می‌پذیرد.

فاز طراحی شامل یافتن یا ایجاد و آنالیز اقدامات امکان‌پذیر می‌باشد. اینها شامل درک مسئله یا مشکل و آزمودن راه‌حل‌ها از لحاظ امکان‌سنجی است. مدلی از مسئله‌ی تصمیم‌گیری در اینجا ساخته، آزموده و اعتبارسنجی می‌شود. ابتدا بگذارید مدل را تعریف کنیم.

مدل‌ها یک ویژگی اصلی سیستم پشتیبان رایانه‌ای و بسیاری از ابزارهای هوش تجاری (بویژه تحلیل کسب‌وکار) این است که حداقل از یک مدل تشکیل شده‌اند. فکر اصلی اینکار این است که به جای یک سیستم واقعی، یک مدل از واقعیت را مورد تجزیه و تحلیل قرار دهیم. یک مدل، نمایشی ساده شده یا انتزاعی از واقعیت است. مدل را معمولاً به این دلیل ساده‌سازی می‌کنند چون واقعیت به اندازه‌ای پیچیده است که نمی‌توان آن را دقیقاً توصیف نمود و بیشتر این پیچیدگی‌ها هیچ کمک به حل یک مسئله‌ی خاص نمی‌کند.

مدل‌سازی شامل مفهومی‌سازی یک مسئله یا مشکل و خلاصه‌سازی آن به یک فرم کمی و یا کیفی است. در خصوص یک مدل ریاضی، متغیرها شناسایی شده و ارتباطات بین آنها تعیین می‌شوند. با در نظر گرفتن فرضیات هر زمان که لازم باشد می‌توان ساده‌سازی‌ها را انجام داد. به عنوان مثال، یک ارتباط بین دو متغیر را می‌توان خطی در نظر گرفت اگرچه در واقعیت ممکن است اثرات غیرخطی داشته باشند. برای یک توازن بین هزینه و منفعت می‌بایست به تعادلی مناسب بین میزان ساده‌سازی مدل و نمایش واقعیت دست یافت. مدل‌های ساده‌تر، هزینه‌های توسعه‌ای پایین‌تر و امکان بکارگیری آسان‌تری دارند و سریع‌تر می‌توان به راه‌حل رسید ولی چندان بیانگر مشکل واقعی نیستند و ممکن است به نتایجی نه چندان دقیق منجر شوند. با وجود این، یک مدل ساده به داده‌های کمتری نیاز دارد یا اینکه گردآوری و جمع‌بندی داده‌های مورد نیاز آن راحت‌تر است.

انتخاب، یک اقدام حساس و حیاتی در تصمیم‌گیری است. فاز انتخاب، فازی است که تصمیم واقعی در آن گرفته شده و تعهد به پیروی از یک مجموعه اقدامات خاص ایجاد می‌شود. از آنجائیکه فعالیت‌های خاص را می‌توان در هر دو فاز انجام داد و تصمیم‌گیرنده می‌تواند چندین مرتبه از فعالیت‌های انتخاب به فعالیت‌های طراحی باز گردد (مثلاً ایجاد راه‌حل‌های جایگزین جدید در حین ارزشیابی راه‌حل‌های موجود)، مرز بین فازهای طراحی و انتخاب چندان روشن و شفاف نیستند. فاز انتخاب شامل جستجو، ارزشیابی و توصیه‌ی یک راه‌حل مناسب برای یک مدل است. راه‌حل برای یک مدل عبارتست از مجموعه‌ای مشخص از مقادیر برای متغیرهای تصمیم در یک راه‌حل برگزیده و منتخب. انتخاب‌ها را می‌توان بر اساس سودآوری و دوام‌پذیری ارزیابی نمود.

هر راه‌حل جایگزین را باید ارزیابی نمود. اگر یک راه‌حل چندین هدف داشته باشد، باید تمامی آنها را بررسی نموده و نسبت به یکدیگر به تعادل رساند. از تحلیل حساسیت برای تعیین میزان مقاومت هر راه‌حل جایگزین استفاده کرد. تغییرات جزئی در پارامترها باید در حالت ایده‌آل منجر به تغییرات جزئی در راه‌حل جایگزین انتخابی شوند یا اینکه هیچگونه تغییری در آن ایجاد

نکند. از تحلیل «چه می‌شود اگر؟»^۱ به منظور کشف تغییرات مهم در پارامترها استفاده می‌شود. تکنیک‌های هدف‌یابی^۲ به مدیران در تعیین مقادیر متغیرهای تصمیم به گونه‌ای که هدف مورد نظر را برآورده سازد کمک می‌کند. این عناوین در فصل ۸ بررسی شده‌اند.

ماکیاولی حدود ۵۰۰ سال پیش در کتاب شه‌ریار خود اشاره کرد که «هیچ کاری دشوارتر، یا موفقیت در آن پرابهام‌تر یا مسئولیت آن خطرناک‌تر از آغاز یک نظم جدید در امور نیست.» پیاده‌سازی و اجرای یک راه‌حل پیشنهادی برای یک مشکل در واقع آغاز نظم جدیدی در امور یا ایجاد تغییر است و تغییر باید مدیریت شود. انتظارات کاربران باید به عنوان بخشی از مدیریت تغییر، اداره شود.

تعریف پیاده‌سازی تا حدودی پیچیده است زیرا پیاده‌سازی، فرایندی طولانی با مرزها و محدوده‌های مبهم است. به بیان ساده، فاز پیاده‌سازی شامل اجرا و بکار گرفتن یک راه‌حل توصیه شده و نه لزوماً پیاده‌سازی یک سیستم رایانه‌ای است. بسیاری از مسائل عمومی در پیاده‌سازی از قبیل مقاومت در برابر تغییر، میزان پشتیبانی مدیریت ارشد و آموزش کاربران در بحث تصمیم‌گیری با پشتیبانی سیستم‌های اطلاعاتی نیز اهمیت دارد. علاوه بر این، بسیاری از موج‌های مبتنی بر فناوری قبلی (مثلاً مهندسی مجدد فرایندهای کسب‌وکار و مدیریت دانش) عمدتاً به دلیل مسائل و چالش‌های مدیریت تغییر، نتایج ناهمسانی را به همراه داشته‌اند. مدیریت تغییر تقریباً خودش یک رشته‌ی کامل محسوب می‌شود و به همین دلیل ما خوانندگان را دعوت می‌کنیم تا به طور خاص و مستقل بر این موضوع تمرکز کنند. پیاده‌سازی همچنین شامل درک کامل مدیریت پروژه است. اهمیت مدیریت پروژه فراتر از ابزارهای تحلیل است و به همین دلیل است که طی چندین سال گذشته شاهد رشد چشمگیر برنامه‌های تصدیق و صدور گواهی برای مدیران پروژه بوده‌ایم. یکی از این تصدیق‌های بسیار شایع و محبوب، «مدیر حرفه‌ای پروژه»^۳ می‌باشد. برای جزئیات بیشتر رجوع کنید به pmi.org

پیاده‌سازی باید شامل گردآوری و تحلیل داده‌ها برای یادگیری از تصمیم‌های پیشین و بهبود تصمیم‌های آتی باشد. اگرچه تحلیل داده‌ها معمولاً به منظور شناسایی مشکل و یا راه‌حل انجام می‌شود ولی از تحلیل نیز باید در فرایند بازخورد استفاده کرد. این به ویژه در خصوص هرگونه تصمیمات خط‌مشی عمومی صادق است. ما باید از اعتبار و صحت داده‌هایی که برای شناسایی مشکل بکار می‌روند مطمئن شویم. برخی اوقات، افراد فقط پس از فاز پیاده‌سازی است که به این موضوع پی می‌برند. فرایند تصمیم‌گیری اگرچه توسط انسان انجام می‌شود ولی قابلیت بهبود به کمک پشتیبان رایانه‌ای را دارد که در ادامه به معرفی آن می‌پردازیم.

در تعاریف اولیه از سیستم پشتیبان تصمیم، آن را به عنوان سیستمی به منظور پشتیبانی از مدیران تصمیم‌گیر در موقعیت‌های نیمه‌ساخت‌یافته و غیرساخت‌یافته می‌شناختند. سیستم پشتیبانی تصمیم می‌بایست کمک و دستیار تصمیم‌گیرندگان بوده و قابلیت‌ها و توانمندی‌هایشان را گسترش دهد نه اینکه جایگزین قضاوت تصمیم‌گیرندگان شود. این سیستم برای تصمیماتی که نیازمند قضاوت بودند یا اینکه امکان پشتیبانی کامل با الگوریتم‌ها را نداشتند در نظر گرفته شد. یکی از مفاهیمی که در تعاریف

¹ What-If Analysis

² Goal Seeking

³ Project Management Professional (PMP)

اولیه نه صراحتاً ولی به طور ضمنی به آن اشاره شده بود این بود که این سیستم باید رایانه‌ای باشد، به شکل آنلاین و تعاملی عمل کند و ترجیحاً از قابلیت‌های خروجی گرافیکی برخوردار باشد که امروزه اینکار از طریق مرورگرها و دستگاه‌های سیار ساده‌سازی شده است.

یک چارچوب اولیه برای سیستم‌های پشتیبان تصمیم رایانه‌ای شامل چندین مفهوم اصلی است که در بخش‌ها و فصل‌های بعدی این کتاب به آنها پرداخته‌ایم. گری و اسکات-مورتون^۱ (۱۹۷۱) چارچوبی را در قالب یک ماتریس ۳ * ۳ ارائه کردند که در شکل ۱.۲ نشان داده شده است. دو بُعد این ماتریس عبارتند از درجه یا میزان ساخت‌یافتگی و انواع کنترل.

درجه‌ی ساخت‌یافتگی بخش سمت چپ شکل ۱.۲ بر اساس نظریه‌ی سیمون (۱۹۷۷) می‌باشد که فرایندهای تصمیم‌گیری همگی در یک پیوستار از تصمیمات بسیار ساخت‌یافته (برخی اوقات برنامه‌ریزی شده) تا بسیار ساخت‌نیافته (یعنی برنامه‌ریزی نشده) قرار می‌گیرند. فرایندهای ساخت‌یافته، مشکلاتی روتین و عموماً تکرارشونده هستند که راه‌حل‌های استاندارد برای آنها وجود دارد. فرایندهای ساخت‌نیافته، مشکلات فازی و پیچیده‌ای هستند که روش‌های حل مسئله‌ی مدون و مشخصی برای آنها وجود ندارد.

یک مسئله یا مشکل ساخت‌نیافته به گونه‌ای است که بیان مشکل یا روش حل آن ممکن است فی‌نفسه ساخت‌نیافته باشد. در یک مسئله‌ی ساخت‌یافته، رویه‌های دستیابی به بهترین راه‌حل (یا حداقل، یک راه‌حل خوب)، معلوم هستند. صرف‌نظر از اینکه مسئله‌ی اصلی ما یافتن سطح موجودی مناسب باشد یا انتخاب بهترین استراتژی سرمایه‌گذاری، اهداف به روشنی تعریف شده‌اند. اهداف متداول عبارتند از دستیابی به حداقل هزینه و حداکثر سود.

مسائل نیمه‌ساخت‌یافته در طیف مابین مسائل ساخت‌یافته و ساخت‌نیافته قرار می‌گیرند که برخی از ویژگی‌ها و اجزای مسائل ساخت‌یافته و ساخت‌نیافته را دارند. کین و اسکات-مورتون^۲ (۱۹۷۸)، پیوندهای تجاری، تعیین بودجه بازاریابی برای محصولات مصرفی و انجام تحلیل تحویل سرمایه را به عنوان مسائل نیمه‌ساخت‌یافته ذکر کرده‌اند.

انواع کنترل بخش دوم چارچوب گری و اسکات-مورتون (۱۹۷۱) در شکل ۱.۲ مبتنی بر طبقه‌بندی آنتونی^۳ (۱۹۶۵) است که سه گروه کلی را تعریف می‌کند که شامل کلیه فعالیت‌های مدیریتی می‌باشند: *برنامه‌ریزی/استراتژیک*^۴ که شامل تعریف اهداف بلندمدت و خط‌مشی‌ها برای تخصیص منابع می‌باشد؛ *کنترل مدیریت*^۵ که شامل تحویل و بکارگیری مؤثر منابع در راستای تحقق اهداف سازمانی است و *کنترل عملیاتی*^۶ که اجرای کارا و اثربخش وظایف مشخص است.

¹ Gorry & Scott-Morton

² Keen & Scott-Morton

³ Anthony

⁴ Strategic Planning

⁵ Management Control

⁶ Operational Control

انواع تصمیم	انواع کنترل		
	کنترل عملیاتی	کنترل مدیریتی	برنامه‌ریزی استراتژیک
ساخت یافته	۱ - پایش حساب‌های دریافتی - پایش حساب‌های پرداختی - ثبت سفارشات ورودی	۲ - تحلیل بودجه - پیش‌بینی کوتاه مدت - گزارش‌دهی پرسنل - شناسایی مسئله - تولید یا خرید	۳ - مدیریت تأمین سرمایه - پایش پورتفولیوی سرمایه‌گذاری - جایابی انبارها - پایش سیستم‌های توزیع
نیمه ساخت یافته	۴ - زمانبندی تولید - کنترل موجودی	۵ - ارزشیابی اعتبار - آماده‌سازی بودجه - طرح‌ریزی واحد صنعتی - زمانبندی پروژه - طراحی سیستم پاداش - دسته‌بندی موجودی	۶ - ساخت واحد صنعتی جدید - برنامه‌ریزی ادغام و اکتساب‌ها - برنامه‌ریزی محصولات جدید - برنامه‌ریزی جبران خدمت - تضمین کیفیت - استقرار خط‌مشی‌های منابع انسانی - برنامه‌ریزی موجودی
ساخت نیافته	۷ - خرید نرم‌افزار - تأیید وام‌ها - راه‌اندازی یک میز کمک - انتخاب جلد مجله	۸ - مذاکره - استخدام یک مدیر اجرایی - خرید سخت‌افزار - لابی‌گری	۹ - برنامه‌ریزی تحقیق و توسعه - توسعه‌ی فناوری‌های جدید - برنامه‌ریزی مسئولیت‌های اجتماعی

شکل ۱.۲ چارچوب‌های پشتیبان تصمیم

ماتریس پشتیبان تصمیم طبقه‌بندی‌های آنتونی (۱۹۶۵) و سیمون (۱۹۷۷) در قالب یک ماتریس پشتیبان تصمیم ۹ سلولی در شکل ۱.۲ ترکیب شده‌اند. هدف اولیه از تشکیل این ماتریس، تعریف انواع مختلف پشتیبانی‌های رایانه‌ای برای سلول‌های مختلف در این ماتریس است. گری و اسکات-مورتون (۱۹۷۱) به عنوان مثال توصیه کرده‌اند که برای اتخاذ تصمیمات نیمه‌ساخت یافته و تصمیمات ساخت نیافته، استفاده از سیستم‌های اطلاعات مدیریت^۱ سنتی و ابزارهای علوم مدیریت^۲ کافی نیست. برای اینکار به ذهن انسان و یک روش متفاوت از فناوری‌های رایانه‌ای نیاز است. آنها استفاده از یک سیستم اطلاعات پشتیبان‌کننده را پیشنهاد داده و آن را یک سیستم پشتیبانی تصمیم (دی‌اس‌اس)^۳ نامیدند.

توجه داشته باشید که وظایف کنترل محور ساخت یافته‌تر و عملیاتی (سلول‌های ۱، ۲ و ۴ در شکل ۱.۲) معمولاً توسط مدیران سطح پایین‌تر انجام می‌شوند در حالیکه وظایف سلول‌های ۶، ۸ و ۹ مسئولیت مدیران اجرایی ارشد یا متخصصان آموزش دیده سطح بالا هستند.

¹ Management Information Systems (MIS)

² Management Science (MS)

³ Decision Support System (DSS)

پشتیبانی رایانه‌ای برای تصمیمات ساخت‌یافته از دهه ۱۹۶۰ رایانه‌ها به پشتیبانی تصمیمات ساخت‌یافته و نیمه ساخت‌یافته پرداخته‌اند به ویژه تصمیماتی که شامل کنترل عملیاتی و مدیریتی بوده‌اند. تصمیمات کنترلی عملیاتی و عملیاتی در تمامی حوزه‌های وظیفه‌ای اتخاذ می‌شوند به ویژه در مدیریت تولید (یعنی عملیات) و تأمین سرمایه.

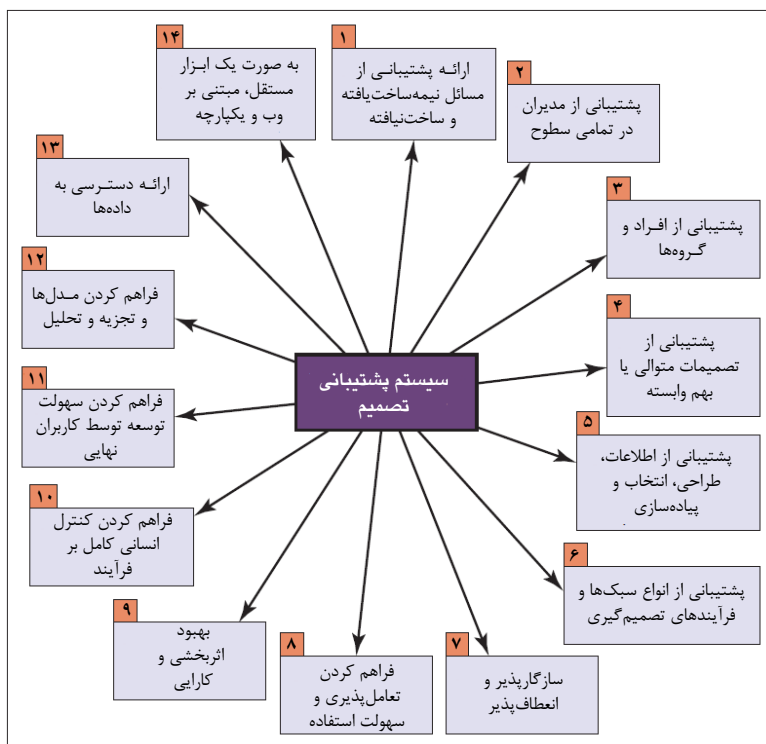
مسائل ساخت‌یافته که مکرراً با آنها مواجه می‌شویم همانطور که از نام‌شان پیداست از سطح ساختار بالایی برخوردارند. از این رو می‌توان آنها را به گروه و دسته‌های مشخص خلاصه، تحلیل و دسته‌بندی نمود. به عنوان مثال، یک تصمیم برای ساخت یا خرید به عنوان یک گروه یا دسته در نظر گرفته می‌شود. مثال‌های دیگر از دسته‌بندی‌ها شامل تصمیمات بودجه‌ریزی سرمایه، تخصیص منابع، توزیع، تدارکات، برنامه‌ریزی و کنترل موجودی می‌باشند. برای هر گروه تصمیم، یک مدل تجویزی ساده و یک رویکرد حل مسئله توسعه داده شده است که معمولاً به شکل فرمول‌های کمی هستند. بنابراین، می‌توان از یک روش علمی برای خودکارسازی بخش‌های تصمیم‌گیری مدیریتی استفاده کرد. راه‌حل‌های مسائل ساخت‌یافته را می‌توان به طور کامل به صورت خودکار در آورد (فصل‌های ۲ و ۱۲).

پشتیبانی رایانه‌ای برای تصمیمات ساخت‌نیافته مشکلات ساخت‌نیافته را فقط به صورت جزئی می‌توان توسط روش‌های کمی رایانه‌ای استاندارد پشتیبانی نمود. معمولاً توسعه‌ی راه‌حل‌های سفارشی ضروری است. با وجود این، این‌گونه راه‌حل‌ها ممکن است داده‌ها و اطلاعات تولید شده از منابع داخلی یا سازمانی نفع ببرند. شهود و قضاوت می‌تواند نقش مهمی در این نوع تصمیمات ایفا کند همانطور که فناوری‌های ارتباطی و همکاری رایانه‌ای و رایانش شناختی (فصل ۶) و یادگیری عمیق (فصل ۵) می‌تواند نقش بسزایی داشته باشد.

پشتیبانی رایانه‌ای برای مشکلات نیمه ساخت‌نیافته حل مسائل و مشکلات نیمه‌ساخت‌یافته ممکن است شامل ترکیبی از رویه‌ها و روش‌های استاندارد و قضاوت انسانی برای حل مسئله باشد. علوم مدیریت می‌توانند مدلی را برای بخش ساخت‌یافته‌ی مسئله‌ی تصمیم‌گیری فراهم کنند. برای بخش ساخت‌نیافته نیز یک سیستم پشتیبانی تصمیم می‌تواند کیفیت اطلاعاتی که تصمیم بر اساس آن اتخاذ می‌شود را نه تنها به واسطه‌ی یک راه‌حل بلکه طیفی از راه‌حل‌های جایگزین به همراه اثرات بالقوه‌شان بهبود بخشد. این قابلیت‌ها به مدیران کمک می‌کنند تا ماهیت مسائل و مشکلات را بهتر درک کرده و بدین ترتیب تصمیمات بهتری بگیرند.

سیستم پشتیبانی تصمیم: قابلیت‌ها در تعاریف اولیه، سیستم پشتیبانی تصمیم یک سیستم با هدف پشتیبانی تصمیم‌گیرندگان مدیریتی در موقعیت‌های نیمه‌ساخت‌یافته و ساخت‌نیافته بود. سیستم پشتیبانی تصمیم می‌بایست کمک و دستیار تصمیم‌گیرندگان بوده و قابلیت‌ها و توانمندی‌هایشان را گسترش دهد نه اینکه جایگزین قضاوت تصمیم‌گیرندگان شود. این سیستم برای تصمیماتی در نظر گرفته شده بود که یا نیازمند قضاوت بودند یا اینکه امکان پشتیبانی کامل با الگوریتم‌ها وجود نداشت. یکی از مفاهیمی که در تعاریف اولیه نه صراحتاً ولی به طور ضمنی به آن اشاره شده بود این بود که این سیستم باید رایانه‌ای باشد، به شکل آنلاین و تعاملی عمل کند و ترجیحاً از قابلیت‌های خروجی گرافیکی برخوردار باشد که امروزه اینکار از طریق مرورگرها و دستگاه‌های سیار ساده‌سازی شده است

یک سیستم پشتیبانی تصمیم معمولاً به منظور پشتیبانی از جواب یا راه حل یک مشکل خاص یا ارزشیابی یک فرصت ساخته می‌شود. این تفاوت اساسی بین کاربردهای سیستم پشتیبانی تصمیم و هوش تجاری است. به مفهومی دقیق، سیستم‌های هوش تجاری وظیفه‌ی پایش موقعیت‌ها و شناسایی مشکلات و یا فرصت‌ها را با استفاده از روش‌های تجزیه و تحلیل دارند. گزارش‌گیری نقش بسیار مهمی در هوش تجاری دارد. کاربر معمولاً باید مشخص کند که آیا یک موقعیت خاص، نیازمند توجه است یا خیر و سپس روش‌های تحلیلی را بکار ببرد. مجدداً اگرچه دسترسی داده‌ها و مدل‌ها (معمولاً از طریق یک انبار داده) در هوش تجاری دیده شده است ولی یک سیستم پشتیبانی تصمیم می‌تواند پایگاه داده خاص خودش را داشته باشد و برای حل یک مسئله‌ی خاص یا مجموعه‌ای از مسائل طراحی و توسعه یافته باشد و از این رو آنها را برنامه‌های کاربردی سیستم پشتیبانی تصمیم^۱ می‌نامند. به لحاظ رسمی، یک سیستم پشتیبانی تصمیم، رویکرد (یا متدولوژی) برای پشتیبانی از تصمیم‌گیری است. این سیستم از سیستم اطلاعات رایانه‌ای^۲ تعاملی، انعطاف‌پذیر و سازگارپذیر استفاده می‌کند که به طور ویژه برای پشتیبانی از راه‌حل یک مسئله‌ی مدیریت ساخت‌نیافته‌ی مشخص توسعه یافته است. علاوه بر این، یک سیستم پشتیبانی تصمیم شامل مدل‌ها است و از طریق یک فرایند تعاملی و تکرارپذیر (احتمالاً توسط کاربران نهایی) توسعه یافته است. این سیستم قادر است از تمامی فازهای تصمیم‌گیری پشتیبانی کند و می‌تواند شامل مؤلفه‌ی دانشی باشد. نهایتاً، یک نفر کاربر می‌تواند از یک سیستم پشتیبانی تصمیم استفاده کرده یا اینکه به شکل مبتنی بر وب برای استفاده‌ی تعداد زیادی از افراد در چندین مکان مورد استفاده قرار گیرد.



شکل ۱.۳ قابلیت‌ها و مشخصه‌های اصلی یک سیستم پشتیبانی تصمیم

¹ DSS Applications

² Computer-Based Information Systems (CBIS)

مشخصه‌ها و قابلیت‌های سیستم پشتیبانی تصمیم با توجه به اینکه اتفاق آرا در خصوص تعریف دقیقی از سیستم پشتیبانی تصمیم وجود ندارد، مشخصاً هیچ توافقی در خصوص قابلیت‌ها و مشخصه‌های استاندارد سیستم پشتیبانی تصمیم نیز وجود ندارد. قابلیت‌های نشان داده شده در شکل ۱.۳ متشکل از یک مجموعه ایده‌آل است که برخی از اجزای آن در تعاریف سیستم پشتیبانی تصمیم شرح داده شده و در موردهای کاربردی به تصویر کشیده شده‌اند. مشخصه‌ها و قابلیت‌های مهم سیستم پشتیبانی تصمیم به شرح زیر می‌باشند (شکل ۱.۳):

۱. از تصمیم‌گیرندگان، عمدتاً در موقعیت‌های نیمه‌ساخت‌یافته و ساخت‌نیافته و با ترکیب قضاوت انسانی و اطلاعات رایانه‌ای پشتیبانی می‌کند. اینگونه مسائل را نمی‌توان توسط سیستم‌های رایانه‌ای یا با بکارگیری روش‌ها یا ابزارهای کمی استاندارد حل نمود. به طور کلی، با توسعه‌ی سیستم پشتیبانی تصمیم این مسائل ساختار پیدا می‌کنند. حتی برخی مسائل ساخت‌یافته توسط سیستم پشتیبانی تصمیم حل شده‌اند.
۲. از کلیه سطوح مدیریتی از مدیران اجرایی ارشد گرفته تا مدیران صف پشتیبانی می‌کند.
۳. از کلیه افراد و گروه‌ها پشتیبانی می‌کند. مسائلی که ساخت‌یافته نیستند اغلب نیازمند مشارکت افراد از دپارتمان‌ها و سطوح سازمانی مختلف یا حتی از سازمان‌های مختلف هستند. سیستم پشتیبانی تصمیم از طریق ابزارهای همکاری وب از تیم‌های مجازی پشتیبانی می‌کند. سیستم پشتیبانی تصمیم برای پشتیبانی از کار فردی و گروهی و همچنین پشتیبانی از تصمیم‌گیری فردی و گروهی که مستقلاً کار می‌کنند توسعه یافته است.
۴. از تصمیمات متوالی و یا بهم وابسته پشتیبانی می‌کند. این تصمیمات ممکن است یک مرتبه، چندین مرتبه یا مکرراً اتخاذ شوند.
۵. از تمامی فازهای فرایند تصمیم‌گیری پشتیبانی می‌کند: اطلاعات، طراحی، انتخاب و پیاده‌سازی.
۶. از انواع سبک‌ها و فرایندهای تصمیم‌گیری پشتیبانی می‌کند.
۷. به دلیل انعطاف‌پذیری، کاربران می‌توانند اجزای اصلی را افزوده، حذف کرده، ترکیب کرده، تغییر داده یا مجدداً مرتب کنند. تصمیم‌گیرنده باید واکنشگر بوده و بتواند سریعاً با شرایط در حال تغییر مقابله کند و بتواند سیستم پشتیبانی تصمیم را با این تغییرات وفق دهد. همچنین از این لحاظ انعطاف‌پذیر است که می‌توان آن را به راحتی برای حل مسائل مشابه دیگر تغییر داد.
۸. کاربردپسند است، دارای قابلیت‌های گرافیکی بالا بوده و یک رابط تعاملی انسان-ماشین به زبان طبیعی دارد که می‌تواند اثربخشی سیستم پشتیبانی تصمیم را بسیار افزایش دهد. اکثر برنامه‌های کاربردی سیستم پشتیبانی تصمیم جدید از رابط‌های مبتنی بر وب یا رابط‌های پلتفرم سیار استفاده می‌کنند.
۹. به جای کارایی (هزینه‌ی تصمیمات گرفته شده)، اثربخشی (یعنی دقت، به موقع بودن، کیفیت) تصمیم‌گیری را بهبود می‌بخشد. زمانیکه سیستم پشتیبانی تصمیم جاری‌سازی شود، زمان تصمیم‌گیری بیشتر می‌شود ولی تصمیمات بهتری گرفته می‌شود.
۱۰. تصمیم‌گیرنده کنترل کامل بر تمامی گام‌های فرایند تصمیم‌گیری در حل مسئله دارد. هدف یک سیستم پشتیبانی تصمیم به طور خاص، پشتیبانی تصمیم‌گیرنده است نه اینکه جایگزین او شود.
۱۱. کاربران نهایی می‌توانند سیستم‌های ساده را خودشان توسعه داده یا اصلاح کنند. سیستم‌های بزرگتر را می‌توان به کمک متخصصان سیستم‌های اطلاعاتی ساخت. بسته‌های نرم‌افزاری صفحه گسترده در توسعه‌ی سیستم‌های ساده‌تر بکار گرفته شده‌اند. کاربران به کمک نرم‌افزارهای داده کاوی و پردازش تحلیلی آنلاین در کنار انبارهای داده می‌توانند سیستم‌های پشتیبانی تصمیم پیچیده و بزرگی را بسازند.

۱۲. مدل‌هایی را ارائه می‌کند که معمولاً برای تحلیل موقعیت‌های تصمیم‌گیری بکار می‌روند. این قابلیت مدل‌سازی امکان آزمایش سیستم را با استراتژی‌های مختلف و تحت پیکربندی‌های مختلف فراهم می‌آورد.

۱۳. دسترسی به انواع مختلف منابع و فرمت‌ها را فراهم می‌کند از جمله داده‌های سیستم‌های اطلاعات جغرافیایی، چندرسانه‌ای و شیء‌گرا.

۱۴. می‌توان آن را به صورت یک ابزار مستقل که یک تصمیم‌گیرنده در یک مکان از آن استفاده می‌کند یا در سرتاسر سازمان توزیع شده است و در چندین سازمان در سرتاسر زنجیره تأمین بکار گرفت. می‌توان آن را با سایر برنامه‌های کاربردی و یا سیستم‌های پشتیبانی تصمیم ادغام نمود و با استفاده از شبکه‌ها و فناوری‌های وب در دخل و خارج سازمان توزیع نمود.

این مشخصه‌ها و قابلیت‌های مهم سیستم پشتیبانی تصمیم به تصمیم‌گیرندگان اجازه می‌دهند تا تصمیمات بهتر و سازگارتری را در زمان مناسب بگیرند و از مولفه‌های اصلی سیستم‌های پشتیبانی تصمیم برخوردارند.

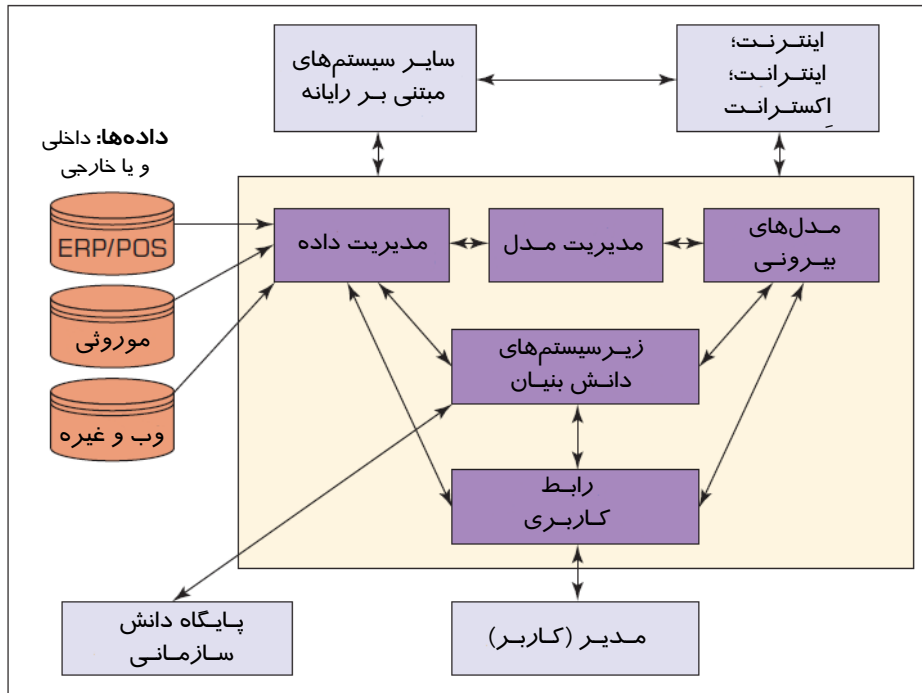
یک برنامه کاربردی سیستم پشتیبانی تصمیم از یک زیرسیستم مدیریت داده‌ها، یک زیرسیستم مدیریت مدل، یک زیرسیستم واسط کاربری و یک زیرسیستم مدیریت مبتنی بر دانش تشکیل شده است. این اجزا در شکل ۱.۴ نشان داده شده‌اند.

زیرسیستم مدیریت داده‌ها شامل یک پایگاه داده است که حاوی داده‌های مربوط به یک موقعیت بوده و توسط نرم‌افزاری به نام سیستم مدیریت پایگاه داده مدیریت می‌شود. سیستم مدیریت پایگاه داده^۱ همانند بسیاری دیگر از سرواژه‌ها در این کتاب هم برای حالت مفرد و هم جمع بکار می‌رود (سیستم و سیستم‌ها). زیرسیستم مدیریت داده را می‌توان به یک انبار داده‌های سازمانی که برای انبارش داده‌های تصمیم‌گیری مرتبط بکار می‌رود متصل نمود. معمولاً، از طریق یک وب سرور پایگاه داده، داده‌ها را ذخیره کرده یا به آنها دسترسی می‌یابند. زیرسیستم مدیریت داده از اجزای زیر تشکیل شده است:

- پایگاه داده‌ی سیستم پشتیبانی تصمیم
- سیستم مدیریت پایگاه داده
- دایرکتوری داده‌ها
- امکانات پرس و جو

نقطه قوت بسیاری از برنامه‌های تحلیلی توصیفی از بخش مدیریت داده‌های زیرسیستم‌ها گرفته شده است.

¹ Database Management System (DBMS)



شکل ۱.۴ نمای شماتیک از یک سیستم پشتیبانی تصمیم

زیرسیستم مدیریت مدل شامل مدل‌های مالی، آماری، علوم مدیریت و سایر مدل‌های کمی است که قابلیت‌های تحلیلی سیستم و مدیریت مناسب نرم‌افزار را امکان‌پذیر می‌کنند. زبان‌های مدل‌سازی برای ایجاد مدل‌های سفارشی نیز شامل می‌شوند. این نرم‌افزار را اغلب یک سیستم مدیریت پایگاه مدل^۱ می‌نامند. این مولفه را می‌توان به یک انبار مدل‌های درون یا برون سازمانی متصل کرد. روش‌های حل مدل و سیستم‌های مدیریت در قالب سیستم‌های توسعه وب (از قبیل جاوا) پیاده‌سازی و بر روی سرورهای کاربردی اجرا می‌شوند. زیرسیستم مدیریت مدل یک سیستم پشتیبانی تصمیم از اجزای زیر تشکیل شده است:

- پایگاه مدل
- سیستم مدیریت پایگاه مدل
- زبان مدل‌سازی
- پردازنده‌ی اجرا، یکپارچه‌سازی و دستوردهی مدل

از آنجائیکه سیستم پشتیبانی تصمیم با مسائل نیمه‌ساخت‌یافته یا ساخت‌نیافته سر و کار دارد، اغلب لازم است تا مدل‌ها را با استفاده از زبان‌ها و ابزارهای برنامه‌نویسی سفارشی‌سازی کنیم. مثال‌هایی از اینها عبارتند از زبان‌های چارچوب دات‌نت، ++C و جاوا. از نرم‌افزار پردازش تحلیلی آنلاین نیز می‌توان برای کار با این مدل‌ها در بخش تحلیل داده استفاده کرد. حتی زبان‌های شبیه‌سازی از قبیل آرنا و بسته‌های آماری از قبیل اسپ‌اس‌اس^۲ نیز ابزارهای مدل‌سازی را ارائه می‌دهند که با استفاده از زبان‌های برنامه‌نویسی انحصاری نوشته شده‌اند. برای سیستم‌های پشتیبانی تصمیم کوچک و متوسط یا برای سیستم‌های ساده‌تر معمولاً

¹ Model Base Management System (MBMS)

² SPSS

از یک صفحه گسترده (مثلاً اکسل^۱) استفاده می‌کنند. ما از اکسل برای مثال‌های متعدد این کتاب استفاده کرده‌ایم. در مورد کاربردی ۱.۲ به توضیح یک سیستم پشتیبانی داده مبتنی بر صفحه گسترده‌ها پرداخته‌ایم.

سیستم پشتیبانی تصمیم‌گیری‌های مخابراتی برای شرکت وان‌نت برای تصمیم‌گیری‌های مخابراتی کمک می‌کند.

مورد کاربردی ۱.۲

ترکیبی از سازمان‌های خصوصی و عمومی هستند که معمولاً به ارائه خدمات شبکه مخابراتی به نهادهای دولتی و مؤسسات آموزشی می‌پردازند. بسیاری از ایالت‌ها در ایالات متحده، یک یا چند آژانس ایالتی دارند که مسئول ارائه خدمات شبکه مخابراتی به مدارس، کالج‌ها و سایر آژانس‌های ایالتی هستند. یک مثال از این آژانس‌ها، وان‌نت^۲ در اکلاهما است. وان‌نت یک انشعاب از هیئت امنای آموزش عالی ایالت اکلاهما^۳ است و با دفتر سرمایه‌گذاری ایالتی^۴ همکاری می‌کند. آژانس‌هایی مثل وان‌نت معمولاً به عنوان یک صندوق سازمانی فعالیت می‌کنند. آنها باید هزینه‌های خود را از طریق صدور صورتحساب برای مشتریان خود و یا توجیه اعتبارات مستقیم از مجلس قانونگذاری ایالتی پوشش دهند. این بازایی هزینه باید از طریق یک مکانیزم قیمت‌گذاری کارآمد، ساده و عادلانه انجام شود. در مدل قیمت‌گذاری باید عوامل بسیاری را در نظر گرفت: همگرایی ترافیک صوتی، ویدئویی و داده‌ها بر روی زیرساخت یکسان؛ تنوع کاربران بر حسب مؤسسات آموزشی و آژانس‌های ایالتی؛ تنوع برنامه‌های کاربردی که کاربران ایالتی استفاده می‌کنند از پست الکترونیک گرفته تا کنفرانس‌های ویدئویی، صدا روی پروتکل اینترنت و یادگیری از راه دور؛ بازایی هزینه‌های جاری و همچنین برنامه‌ریزی برای توسعه‌ها و ارتقاءها در آینده و بهره‌گیری از زیرساخت‌های مشترک برای توسعه اقتصادی و کارهای مشارکتی در سرتاسر ایالت که منجر به کاربردهای نوآورانه وان‌نت می‌شود. این ملاحظات منجر به توسعه‌ی یک مدل مبتنی بر صفحه‌گسترده شد. سیستم پشتیبانی تصمیم‌گیری مبتنی بر کاربرد شبکه سرویس و قیمت‌گذاری^۵ (اس‌ان‌ای‌پی) ما محیط نرم‌افزار مایکروسافت اکسل ۲۰۰۷ توسعه یافته و از زبان برنامه‌نویسی ویژوال بیسیک برای برنامه‌ها^۶ برای توسعه آن استفاده شد. این سیستم با ارائه‌ی یک واسط کاربری گرافیکی و کاربرپسند این امکان را به وان‌نت می‌دهد تا بهترین تعرفه‌های خدماتی را متناسب با استراتژی‌های قیمت‌گذاری مورد نظرشان انتخاب کنند. علاوه بر این، سیستم پشتیبانی تصمیم‌گیری اس‌ان‌ای‌پی نه تنها تأثیر تغییرات عوامل قیمت‌گذاری بر هر تعرفه خدمات را به تصویر می‌کشد بلکه به کاربران اجازه می‌دهد تا گزینه‌های تعرفه‌های خدماتی متنوع را در سناریوهای مختلف و با استفاده از پارامترهای مختلف تجزیه و تحلیل کنند. برنامه‌ریزان مالی وان‌نت از این مدل برای شناخت بیشتری مشتریان و تحلیل سناریوهای متعدد «چه می‌شود اگر؟» برای تعرفه‌های خدماتی مختلف استفاده کرده‌اند.

Source: Based on J. Chongwatpol and R. Sharda. (2010, December). "SNAP: A DSS to Analyze Network Service Pricing for State Networks." *Decision Support Systems*, 50(1), pp. 347–359.

¹ Excel

² OneNet

³ Oklahoma State Regents for Higher Education

⁴ Office of State Finance

⁵ Service Network Application and Pricing (SNAP)-based DSS

⁶ Visual Basic for Applications

کاربر از طریق یک زیرسیستم واسط کاربری با سیستم پشتیبانی تصمیم ارتباط برقرار کرده و فرمان می‌دهد. کاربر در اینجا بخشی از سیستم محسوب می‌شود. پژوهشگران مدعی هستند که برخی از ویژگی‌های منحصر بفرد سیستم پشتیبانی تصمیم از تعامل قوی بین رایانه و تصمیم‌گیرنده گرفته شده است. یکی از علل اصلی که مدیران از رایانه‌ها و تحلیل‌های کمی تا جائیکه می‌توانند استفاده نمی‌کنند، دشوار بودن واسط کاربری است البته با این فرض که این فناوری‌ها در دسترس باشند. در سال‌های ۲۰۰۰، مرورگرهای وب یک ساختار آشنا و سازگار رابط کاربر گرافیکی برای بسیاری از سیستم‌های پشتیبانی تصمیم ارائه می‌کردند. صفحه‌گسترده‌ها برای سیستم‌های پشتیبانی تصمیم محلی یک واسط کاربری آشنا هستند. مرورگرهای وب به خاطر انعطاف‌پذیری، کاربرپسند بودن و اینکه دروازه‌ای به تمامی منابع اطلاعاتی و داده‌های ضروری هستند به عنوان یک رابط کاربر گرافیکی مؤثر برای سیستم‌های پشتیبانی تصمیم شناخته شده‌اند. در اصل، مرورگرهای وب منجر به توسعه‌ی پرتال‌ها و داشبوردها گردیدند که به عنوان قسمت جلویی سیستم پشتیبانی تصمیم عمل می‌کنند.

رشد انفجاری دستگاه‌های قابل حمل یا سیار از جمله تلفن‌های هوشمند و تبلت‌ها، واسط‌های کاربری سیستم‌های پشتیبانی تصمیم را تغییر داده است. این دستگاه‌ها به اجازه می‌دهند تا بتوان مطالب دست‌نویس یا تایپ شده توسط صفحه‌کلیدهای داخلی یا خارجی را وارد سیستم کرد. برخی از واسط‌های کاربر سیستم‌های پشتیبانی تصمیم از زبان طبیعی به عنوان ورودی استفاده می‌کنند (یعنی زبان انسان) که کاربران به راحتی می‌توانند مقصودشان را به شکلی معنادار بیان کنند. در تلفن‌های همراه، وارد کردن اطلاعات از طریق پیامک‌ها و چت‌بات‌ها برای برخی از برنامه‌های کاربردی سیستم پشتیبانی تصمیم متداول شده است. به عنوان مثال، برای درخواست جستجوی هر موضوعی از طریق یک پیامک به (46645) GOOGLE فرستاد. اینگونه قابلیت‌ها بیشترین استفاده را در مکان‌یابی کسب‌وکارهای نزدیک، آدرس‌ها یا شماره تلفن‌ها دارند ولی همچنین می‌توان از آنها در بسیاری دیگر از تصمیم‌گیری‌ها استفاده کرد. به عنوان مثال، کاربران با وارد کردن واژه‌ی «تعریف» پس از هر کلمه‌ای می‌توانند معانی واژگان را بیابند. از قابلیت‌های دیگر می‌توان به موارد زیر اشاره کرد:

- یافتن قیمت: " قیمت گوشی آیفون ۶۴ گیگابایت "
- تبدیل نرخ ارز: " ۱۰ دلار آمریکا به یورو "
- زمان بازی‌ها و نتایج مسابقات: فقط کافی است نام یک تیم را وارد کنید و گوگل، آخرین نتایج مسابقات را به همراه تاریخ و ساعت مسابقه‌ی بعدی برایتان با پیامک ارسال می‌کند.

این نوع قابلیت‌های جستجوی پیامکی در سایر موتورهای جستجو مثل موتور جستجوی شرکت مایکروسافت به نام بینگ^۱ نیز وجود دارد. با ظهور گوشی‌های هوشمند از قبیل آیفون آپل و گوشی‌های هوشمند اندروید با برندهای مختلف، بسیاری از شرکت‌ها در حال توسعه‌ی برنامه‌های کاربردی برای پشتیبانی از تصمیم‌گیری خرید هستند. به عنوان مثال، برنامه‌های کاربردی آمازون به کاربر اجازه می‌دهد تا تصویری از هر قلم کالای موجود در یک فروشگاه را برای آمازون. کام^۲ ارسال کند. آمازون. کام از الگوریتم‌های درک گرافیکی برای تطبیق تصویر با محصول واقعی در پایگاه داده خود استفاده کرده و صفحه‌ای مشابه صفحات معرفی محصول آمازون. کام را برای کاربر ارسال می‌کند و به کاربر اجازه می‌دهد تا قیمت‌ها را در همان لحظه مقایسه کند. میلیون‌ها برنامه کاربردی دیگر نیز توسعه یافته که از مصرف‌کنندگان در تصمیم‌گیری‌ها برای یافتن یا انتخاب

¹ Bing

² Amazon.com

فروشگاه‌ها/رستوران‌ها/ارائه‌دهندگان خدمات بر اساس موقعیت مکانی، توصیه‌های دیگران و به‌ویژه توصیه‌های حلقه‌های دوستان اجتماعی‌تان پشتیبانی می‌کنند. فعالیت‌های جستجو که در پاراگراف قبل ذکر شدند امروزه عمدتاً از طریق برنامه‌های کاربردی ارائه شده توسط شرکت‌های مرتبط انجام می‌شوند.

ورودی صوتی برای این فعالیت‌ها و اسپیکرهای هوشمند جدید مثل آمازون اکو (الکسا)^۱ و گوگل هوم^۲ بسیار متداول بوده و از دقت خوبی برخوردارند (هرچند بی‌عیب نیستند). زمانیکه ورودی صوتی به همراه نرم‌افزار تشخیص گفتار (و نرم‌افزار متن به گفتار که در حال حاضر در دسترس است) بکار می‌رود، دستورالعمل‌های شفاهی به همراه اقدامات و خروجی‌ها را می‌توان بکار برد. این امکانات در حال حاضر برای سیستم‌های پشتیبانی تصمیم موجود بوده و در دستگاه‌های قابل حمل که در بالا توضیح داده شد وجود دارند. یک مثال از ورودی صوتی برای سیستم پشتیبانی تصمیم همه منظوره، برنامه سیری شرکت اپل^۳ و سرویس نو گوگل^۴ هستند. به عنوان مثال، یک کاربر می‌تواند کد پستی خود را وارد کرده و بگوید «تحويل پیتزا». این دستگاه‌ها نتایج جستجو را ارائه کرده و حتی می‌توانند تماس تلفنی برقرار کنند.

بسیاری از توسعه‌های به عمل آمده در رابط‌های کاربری با دو دستاورد جدید مهم در سیستم‌های مبتنی بر دانش همراه بوده‌اند. زیرسیستم مدیریت مبتنی بر دانش می‌تواند از هر زیرسیستم دیگر پشتیبانی کرده یا به صورت یک بخش مستقل عمل کند. این زیرسیستم می‌تواند اطلاعات مورد نیاز برای افزایش اطلاعات تصمیم‌گیرنده یا کمک به درک استعمال کاربر و دادن پاسخ مناسب فراهم کند. این سیستم قابلیت اتصال به مخزن دانشی سازمان (بخشی از سیستم مدیریت دانش) که به آن پایگاه دانش سازمانی^۵ گفته می‌شود یا هزاران منبع دانش برون‌سازمانی را دارد. در نسل جاری سیستم‌های یادگیری از بسیاری از روش‌های هوش مصنوعی استفاده شده و براحتی می‌توان آن را با سایر مؤلفه‌های سیستم پشتیبانی تصمیم ادغام کرد. یکی از محبوب‌ترین سیستم‌های پشتیبانی تصمیم مبتنی بر دانش که تبلیغات زیادی برای آن انجام شده، واتسون شرکت آی‌بی‌ام^۶ است که در پرونده آغازین معرفی شده و در بخش‌های بعدی بیشتر به بررسی آن خواهیم پرداخت. در این بخش به طور خلاصه به توضیح تاریخچه و پیشرفت سیستم‌های پشتیبانی تصمیم پرداختیم. در بخش بعدی، در خصوص تکامل این نوع پشتیبانی در هوش تجاری، تجزیه و تحلیل و علم داده‌ها خواهیم پرداخت.

سوالاتی برای مرور بخش ۱.۳

۱. چهار فاز تصمیم‌گیری سیمون را نامه برده و توضیح دهید.
۲. تفاوت بین یک مشکل و نشانه‌های آن در چیست؟
۳. چرا دسته‌بندی یک مشکل اهمیت دارد؟
۴. پیاده‌سازی را تعریف کنید؟
۵. تصمیمات ساخت‌یافته، نیمه ساخت‌یافته و ساخت‌نیافته چه هستند؟ برای هر کدام دو مثال ارائه کنید.

¹ Amazon Echo (Alexa)

² Google Home

³ Apple's Siri

⁴ Google Now Service

⁵ Organizational Knowledge Base

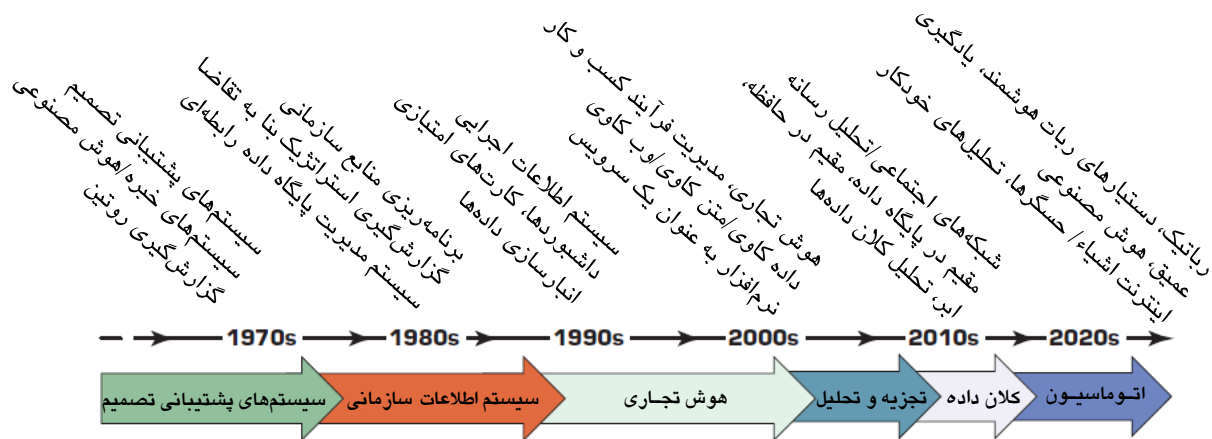
⁶ IBM's Watson

۶. کنترل عملیاتی، کنترل مدیریتی و برنامه‌ریزی استراتژیک را تعریف کنید. برای هر کدام دو مثال ارائه کنید.
۷. نه سلول چارچوب تصمیم چه هستند؟ هر کدام را توضیح دهید.
۸. رایانه‌ها چگونه می‌توانند از تصمیم‌گیری ساخت‌یافته پشتیبانی کنند؟
۹. رایانه‌ها چگونه می‌توانند از تصمیم‌گیری نیمه ساخت‌یافته و ساخت‌نیافته پشتیبانی کنند؟

۱.۴ تکامل پشتیبانی تصمیم رایانه‌ای در هوش تجاری، علم تجزیه و تحلیل و علم داده‌ها

نمودار زمانی نشان داده شده در شکل ۱.۵ مجموعه واژگان تخصصی بکار رفته برای توضیح ابزارهای تجزیه و تحلیل از دهه ۷۰ می‌باشد. در سال‌های ۷۰، تمرکز اصلی سیستم‌های اطلاعاتی پشتیبان برای تصمیم‌گیری بر ارائه‌ی گزارشات ساختارمند و دوره‌ای متمرکز بود تا مدیران بتوانند از آنها برای تصمیم‌گیری استفاده کنند (یا تصمیمات را نادیده بگیرند). کسب‌وکارها شروع به ایجاد گزارشات روتین به منظور مطلع ساختن تصمیم‌گیرندگان (مدیران) در مورد آنچه در دوره قبل اتفاق افتاده است نمودند (مثلاً روز، هفته، ماه یا فصل گذشته). اگرچه این گزارشات برای آگاهی از آنچه در گذشته رخ داده است مفید بودند ولی مدیران به چیزی بیش از این نیاز داشتند: آنها به انواع گزارشات در سطح‌بندی‌های مختلف نیاز داشتن تا نیازهای متغیر و چالش‌های کسب‌وکار را بهتر درک کرده و به آنها رسیدگی کنند. این سیستم‌ها را معمولاً سیستم‌های مدیریت اطلاعات (ام‌آی‌اس)^۱ می‌نامیدند. در اوایل دهه ۷۰، اسکات-مورتون برای اولین بار مفاهیم اصلی سیستم‌های پشتیبانی تصمیم را مطرح کرد. طبق تعریف او سیستم پشتیبانی تصمیم "یک سیستم رایانه‌ای تعاملی است که به تصمیم‌گیرندگان کمک می‌کند تا از داده‌ها و مدل‌ها به منظور حل مسائل ساخت‌نیافته بهره‌برداری کنند". در ادامه تعریف کلاسیک دیگری از سیستم پشتیبانی تصمیم توسط کین و اسکات-مورتون (۱۹۷۸) ارائه شده است:

سیستم‌های پشتیبانی تصمیم، منابع فکری افراد را به منظور بهبود کیفیت تصمیم‌ها با قابلیت‌های رایانه‌ای در هم می‌آمیزند. این یک سیستم پشتیبانی رایانه‌ای است که تصمیم‌گیرندگان مدیریت برای حل و فصل مسائل نیمه ساخت‌یافته از آن استفاده می‌کنند.



شکل ۱.۵ تکامل پشتیبانی تصمیم، هوش تجاری، علم تجزیه و تحلیل و هوش مصنوعی

¹ Information Management Systems (MIS)

توجه داشته باشید که عبارت سیستم پشتیبانی تصمیم همانند سیستم اطلاعات مدیریت و چندین عبارت دیگر در رشته‌های فناوری اطلاعات یک شکل ابراز بدون محتوا است (یعنی برای افراد مختلف، معانی متفاوتی دارد). از این رو، هیچ تعریف پذیرفته شده‌ی جهانی برای سیستم پشتیبانی تصمیم وجود ندارد.

طی روزهای اولیه‌ی پیدایش ابزارهای تجزیه و تحلیل، داده‌ها اغلب از افراد خبره در هر حوزه و با استفاده از فرایندی دستی اخذ می‌شدند (مصاحبه‌ها و نظرسنجی‌ها) تا مدل‌های ریاضی یا مبتنی بر دانش برای حل مسائل بهینه‌سازی دارای محدودیت ایجاد شوند. هدف این بود تا بهترین کار با منابع محدود انجام شود. اینگونه مدل‌های پشتیبانی تصمیم را معمولاً پژوهش عملیاتی می‌نامیدند. مسائلی که بسیار پیچیده بودند و امکان حل بهینه‌ی آنها با استفاده از تکنیک‌های برنامه‌ریزی ریاضی خطی و غیرخطی وجود نداشت با استفاده از روش‌های هیوریستیک مثل مدل‌های شبیه‌سازی حل می‌شدند.

در اواخر دهه ۷۰ و اوایل ۸۰، علاوه بر مدل‌های تکامل یافته‌ی پژوهش عملیاتی که در بسیاری از صنایع و سیستم‌های دولتی استفاده می‌شدند، مدل‌هایی جدید و هیجان‌انگیز ظهور کردند: سیستم‌های خبره مبتنی بر قاعده^۱. این سیستم‌ها قادر به اخذ دانش افراد خبره در قالب قابل پردازش توسط رایانه‌ها بودند (از طریق مجموعه‌ای از قواعد و هیوریستیک‌های شرطی) تا بتوان از آنها برای مشاوره و شناسایی یک مسئله‌ی ساخت‌یافته و تجویز محتمل‌ترین جواب یا راه‌حل استفاده کرد. با استفاده از سیستم‌های خبره مبتنی بر قاعده، تخصص‌های کمیاب و نادر با استفاده از یک سیستم پشتیبانی تصمیم هوشمند در هر زمان و مکان مورد نیاز در دسترس بودند.

دهه ۸۰ شاهد تغییرات مهمی در نحوه‌ی اخذ داده‌های کسب‌وکار توسط سازمان‌ها بود. روش‌های قدیمی اینطور بودند که چندین سیستم اطلاعاتی جدا از هم به شکلی تنظیم می‌شدند تا داده‌های تعاملات را از واحدها یا بخش‌های مختلف سازمان اخذ می‌کردند (مثلاً حسابداری، بازاریابی و فروش، مالی، تولید). در دهه ۸۰، این سیستم‌ها در قالب سیستم‌های اطلاعاتی سازمان یکپارچه‌سازی شدند که اکنون آنها را سیستم‌های برنامه‌ریزی منابع سازمانی^۲ می‌نامیم. الگوهای قدیمی خطی و غیراستاندارد نمایش داده‌ها جای خود را به سیستم‌های مدیریت پایگاه داده رابطه‌ای^۳ دادند. این سیستم‌ها موجب بهبود فرایند گردآوری و انبارش داده‌ها و همچنین روابط بین حوزه‌های داده‌های سازمانی و کاهش بسیار زیاد اطلاعات تکراری شدند. نیاز به سیستم‌های برنامه‌ریزی منابع سازمانی و مدیریت پایگاه‌های داده رابطه‌ای زمانی احساس شد که یکپارچگی و سازگاری داده‌ها تبدیل به یک مشکل شد و اثربخشی فرایندهای کسب‌وکار را شدیداً تحت تأثیر قرار داد. با ظهور برنامه‌ریزی منابع سازمانی، تمامی داده‌ها از هر گوشه‌ی سازمان در قالب یک الگوی منسجم گردآوری و ذخیره می‌شود به گونه‌ای که هر بخشی از سازمان تنها به یک نسخه‌ی یکسان از آنچه حقیقت داشت دسترسی داشتند. علاوه بر پیدایش سیستم‌های برنامه‌ریزی منابع سازمانی یا شاید به دلیل ظهور این سیستم‌ها، گزارش‌گیری کسب‌وکار نیز تبدیل به یک روش بنا به تقاضا شد. تصمیم‌گیرندگان می‌توانستند هر زمان که نیاز به گزارشات تخصصی برای تحقیق و بررسی مسائل یا فرصت‌های سازمانی می‌توانستند تصمیم به ایجاد گزارش بگیرند.

در دهه ۹۰، نیاز به گزارش‌گیری‌های متنوع‌تر منجر به توسعه‌ی سیستم‌های اطلاعات اجرایی گردید (سیستم‌های پشتیبانی تصمیم که به طور خاص برای نیازهای مدیران اجرایی و تصمیم‌گیری آنها طراحی شده و توسعه یافته بودند). این سیستم‌ها به شکل داشبوردها و کارت‌های امتیازی گرافیکی طراحی شدند که می‌شد به صورت نمایشگرهای چشم‌نواز که بر مهم‌ترین عوامل برای تصمیم‌گیرندگان به منظور ردگیری شاخص‌های عملکرد کلیدی تمرکز داشتند استفاده می‌شدند. برای دستیابی به گزارش‌گیری‌های بسیار متنوع و در عین حال، حفظ یکپارچگی سیستم‌های اطلاعات کسب‌وکار، لازم بود تا یک لایه میانی داده‌ها

¹ Rule-based Expert Systems (ES)

² Enterprise Resource Planning (ERP)

³ Relational Database Management Systems (RDBMS)

به نام انبار داده‌ها ایجاد شود که به عنوان یک مخزن خاص برای پشتیبانی تصمیم‌گیری و گزارش‌گیری کسب‌وکار استفاده می‌شد. در مدت زمان بسیار کوتاهی، اکثر سازمان‌ها بزرگ و متوسط از انبارش داده‌ها به عنوان پلتفرمی برای تصمیم‌گیری سازمانی استفاده کردند. داشبوردها و کارت‌های امتیازی داده‌هایشان را از انبار داده می‌گرفتند و با این کار، مانعی برای کارایی سیستم‌های تعاملات کسب‌وکار که عمدتاً سیستم‌های برنامه‌ریزی منابع سازمانی نامیده می‌شدند محسوب نمی‌شدند.

در سال‌های ۲۰۰۰، سیستم‌های پشتیبانی تصمیم مبتنی بر انبار داده‌ها را سیستم‌های هوش تجاری^۱ نامیدند. با توجه به افزایش حجم انبوه داده‌ها در این انبارها، قابلیت‌های سخت‌افزاری و نرم‌افزاری نیز با توجه به تغییرات سریع و نیازهای روزافزون تصمیم‌گیرندگان ارتقاء یافتند. با توجه به بازارهای رقابتی جهانی، تصمیم‌گیرندگان برای رسیدگی به مشکلات کسب‌وکار و بهره‌گیری آنی از فرصت‌های بازار به اطلاعات جاری در یک قالب بسیار قابل فهم نیاز داشتند. چونکه داده‌ها در یک انبار داده به صورت دوره‌ای به‌روزرسانی می‌شدند، آخرین و جدیدترین اطلاعات را ارائه نمی‌کردند. برای رفع این مشکل، تولیدکنندگان انبارهای داده سیستمی را برای به‌روزرسانی مکرر داده‌ها طراحی کرده و توسعه دادند که منجر به *انبارش بلادرنگ داده‌ها*^۲ و به بیان واقع‌بینانه‌تر، *انبارش به موقع داده‌ها*^۳ گردید که تفاوت آنها با نسل قبل خود در این بود که از سیاست‌های به‌روزرسانی داده‌ها بر اساس میزان به‌روز بودن مورد درخواست داده‌ها بوده است (یعنی تمامی داده‌ها نیازی نیست که به طور آنی و بلادرنگ به‌روزرسانی شوند). انبارهای داده، بسیار بزرگ و دارای ویژگی‌های متعددی هستند و کشف دانش جدید و مفید به منظور بهبود فرایندها و رویه‌های کسب‌وکار نیازمند «کاوش» در داده‌های سازمانی می‌باشد. از این رو است که از عبارات *داده‌کاوی*^۴ و *متن‌کاوی*^۵ استفاده می‌شود. با افزایش حجم و تنوع داده‌ها، نیاز به انبارش و توان پردازشی بیشتر نیز بوجود آمد. اگرچه سازمان‌های بزرگ، ابزارهای لازم برای حل این مسائل را داشتند ولی شرکت‌های کوچک و متوسط به مدل‌های کسب‌وکاری نیاز داشتند که از لحاظ مالی برایشان مناسب‌تر باشد. این نیاز منجر به پیدایش نرم‌افزارها و معماری خدمت‌محور و مدل‌های کسب‌وکار تحلیلی زیرساخت به عنوان یک خدمت گردید. بنابراین، شرکت‌های کوچک‌تر بر اساس نیازشان به قابلیت‌های تحلیلی دسترسی یافتند و به جای اینکه بر روی منابع نرم‌افزاری و سخت‌افزاری پرهزینه سرمایه‌گذاری کنند فقط برای آنچه استفاده می‌کردند پول پرداخت می‌کردند.

در سال‌های ۲۰۱۰، ولی شاهد یک تغییر پارادایم در نحوه‌ی اخذ و بکارگیری داده‌ها هستیم که عمدتاً به دلیل استفاده‌ی گسترده از اینترنت، روش‌ها و وسایل جدید تولید داده پدیدار شدند. از بین تمامی این منابع داده (مثلاً برجسب‌های شناسایی فرکانس رادیویی، انرژی مترهای دیجیتالی، وبلاگ‌های جریان کلیک، لوازم خانگی هوشمند، تجهیزات پوشیدنی پایش سلامت)، شاید جالب‌ترین و چالش‌برانگیزترین آنها رسانه‌های اجتماعی/شبکه‌های اجتماعی باشند. این داده‌ها مملو از محتوای اطلاعاتی ساختنیافته هستند ولی تحلیل اینگونه منابع داده‌ای می‌تواند چالش‌های اساسی را از منظر سخت‌افزاری و نرم‌افزاری به سیستم‌های محاسباتی به همراه داشته باشد. اخیراً، عبارت *کلان‌داده*^۶ برای تأکید بر اهمیت چالش‌های جریان‌های داده‌ای جدید برای ما ابداع گردیده است. برای حل و فصل چالش‌های کلان‌داده‌ها، دستاوردهای بسیاری در حوزه سخت‌افزار (مثلاً پردازش‌های موازی حجیم با حافظه‌های محاسباتی بسیار بزرگ و سیستم‌های چندپردازنده رایانش موازی) و نرم‌افزار/الگوریتم‌ها (مثلاً هُدوپ با مَپ‌ردیوس و نواسکیووال^۷، اسپارک^۸) بوجود آمده‌اند.

¹ DW-based DSS

² Real-time Data Warehousing

³ Right-time Data Warehousing

⁴ Data mining

⁵ Text mining

⁶ Big Data

⁷ NoSQL

⁸ Spark

در این چند سال اخیر و دهه‌ی پیش‌رو، شاهد رشد چشمگیری در ابعاد هیجان‌انگیز بسیاری هستیم. به عنوان مثال، فناوری‌های حسگرها و تحلیل‌های جریان داده^۱، موجب شکل‌گیری اینترنت اشیاء شدند. هوش مصنوعی با فراهم کردن روش‌های جدید تحلیل تصاویر از طریق یادگیری عمیق و نه فقط از روش متداول مصورسازی داده‌ها در حال تغییر دادن شکل هوش تجاری است. یادگیری عمیق و هوش مصنوعی نیز به رشد فناوری‌های تشخیص صدا و تلفیق گفتار (تولید گفتار انسان) منجر به ایجاد رابط‌های جدید به منظور تعامل با این فناوری‌ها شده‌اند. تقریباً نیمی از خانوارهای ایالات متحده یک گوینده هوشمند مثل آمازون اکو یا گوگل هوم دارند و شروع به برقراری تعامل با داده‌ها و سیستم‌ها با استفاده از رابط‌های صوتی نموده‌اند. رشد رابط‌های ویدئویی نهایتاً منجر به برقراری تعامل با حرکات و اشارات بدن با سیستم‌ها می‌شود. تمامی این پیشرفت‌ها به واسطه‌ی قابلیت‌های انبارش انبوه داده‌ها در فضای ابری و پردازش‌های فوق‌العاده سریع میسر شده‌اند. ولی هنوز چیزهای زیادی در راه است.

پیش‌بینی اینکه دهه آتی چه چیزهایی برای ما به ارمغان خواهد آورد و شرایط ابزارهای تجزیه و تحلیل جدید چه خواهد بود کار دشواری است. فاصله‌ی زمانی بین تغییرات پارادایمی جدید در سیستم‌های اطلاعاتی و به ویژه در تجزیه و تحلیل‌ها در حال کاهش است و این روند برای آینده‌ی قابل پیش‌بینی ادامه خواهد داشت. اگرچه تجزیه و تحلیل موضوع جدیدی نیست ولی محبوبیت انفجاری آن امر جدیدی است. به لطف انفجار اخیر در حوزه کلان‌داده‌ها، روش‌های گردآوری و انبارش این داده‌ها و ابزارهای نرم‌افزاری شهودی، متخصصان و کارشناسان کسب‌وکارها بیش از گذشته توانسته‌اند تا به درک بهتری از داده‌ها دسترسی یابند. از این رو، در میانه‌ی این رقابت جهانی، فرصت عظیمی برای اتخاذ تصمیمات مدیریتی بهتر با استفاده از داده‌ها و تحلیل‌ها به منظور افزایش درآمد و کاهش هزینه‌ها از طریق ساخت محصولات بهتر، بهبود تجربه‌ی مشتری و پیشگیری از کلاهبرداری‌ها، بهبود مشارکت مشتری از طریق هدف قرار دادن و سفارشی سازی و توسعه‌ی حوزه‌های کسب‌وکار کاملاً جدید وجود دارد. همگی اینها به لطف قدرت داده‌ها و تحلیل‌ها میسر می‌شود. امروزه شرکت‌های بیشتری دانش تخصصی تحلیل‌های کسب‌وکار را در اختیار کارمندان خود قرار می‌دهند تا بتوانند کارایی و اثربخشی را در فرایندهای تصمیم‌گیری روزانه‌شان ارتقاء بخشند. بخش بعدی بر چارچوبی برای هوش تجاری متمرکز است. اگرچه اکثر مردم اتفاق نظر دارند که هوش تجاری به شکل علم داده‌ها و تحلیل رشد و تکامل یافته است ولی بسیاری از شرکت‌ها و پژوهشگران همچنان از آن عبارت استفاده می‌کنند. در چند پاراگراف بعدی ما با تمرکز بر آنچه هوش تجاری خوانده می‌شده است به تاریخچه‌ی آن ادای احترام می‌کنیم. پس از آن، به معرفی تحلیل‌ها و کاربرد آن به عنوان یک برچسب برای دسته‌بندی تمامی مفاهیم مربوطه می‌پردازیم.

مفاهیم پشتیبانی تصمیم که در بخش‌های ۱.۲ و ۱.۳ ارائه شدند به طور تدریجی و تحت نام‌های مختلف توسط بسیاری از شرکت‌های ارائه‌کننده‌ی ابزارها و متدولوژی‌های پشتیبانی تصمیم پیاده‌سازی شده‌اند. همانطور که در بخش ۱.۲ گفته شد، با رشد سیستم‌های سازمانی، مدیران قادر بودند تا به گزارش کاربرپسند دسترسی داشته باشند که به آنها در تصمیم‌گیری سریع کمک می‌کرد. این سیستم‌ها که عموماً سیستم‌های اطلاعات اجرایی نامیده می‌شدند شروع به ارائه قابلیت‌هایی مثل مصورسازی‌های بیشتر، هشدارها و سنجش عملکرد نمودند. تا سال ۲۰۰۶، خدمات و محصولات تجاری مهم تحت عبارت هوش تجاری^۲ ارائه می‌شدند.

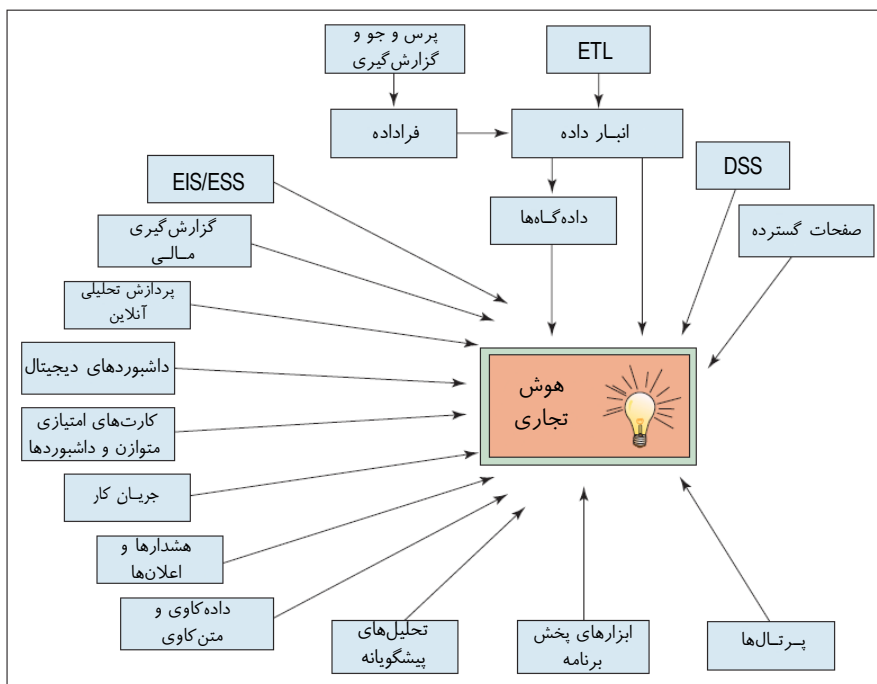
¹ Streaming Analytics

² Business Intelligence (BI)

تعاریف هوش تجاری هوش تجاری یک اصطلاح فراگیر است که معماری‌ها، ابزارها، پایگاه‌های داده، ابزارهای تحلیلی، برنامه‌های کاربردی و متدولوژی‌ها را با یکدیگر ترکیب می‌کند. همانند سیستم پشتیبانی تصمیم، هوش تجاری نیز برای افراد مختلف، معانی متفاوتی دارد. بخشی از سردرگمی‌های مربوط به هوش تجاری ریشه در واژگان اختصاری مرتبط با آن است (مثلاً مدیریت عملکرد کسب‌وکار [پی‌ام]). هدف اصلی هوش تجاری این است که امکان دسترسی تعاملی (برخی اوقات آنی) به داده‌ها، دستکاری داده‌ها را فراهم آورده و توانایی تحلیل‌های مناسب را به مدیران و تحلیل‌گران کسب‌وکارها را بدهد. با تحلیل داده‌های جاری و قبلی و موقعیت‌ها و عملکردها، تصمیم‌گیرندگان می‌توانند به دانش ارزشمندی دست پیدا کنند که آنها را در اتخاذ تصمیمات بهتر و آگاهانه‌تر یاری کند. فرایند هوش تجاری مبتنی بر دگرگون‌سازی و تبدیل داده‌ها به اطلاعات و سپس به تصمیمات و نهایتاً به اقدامات است.

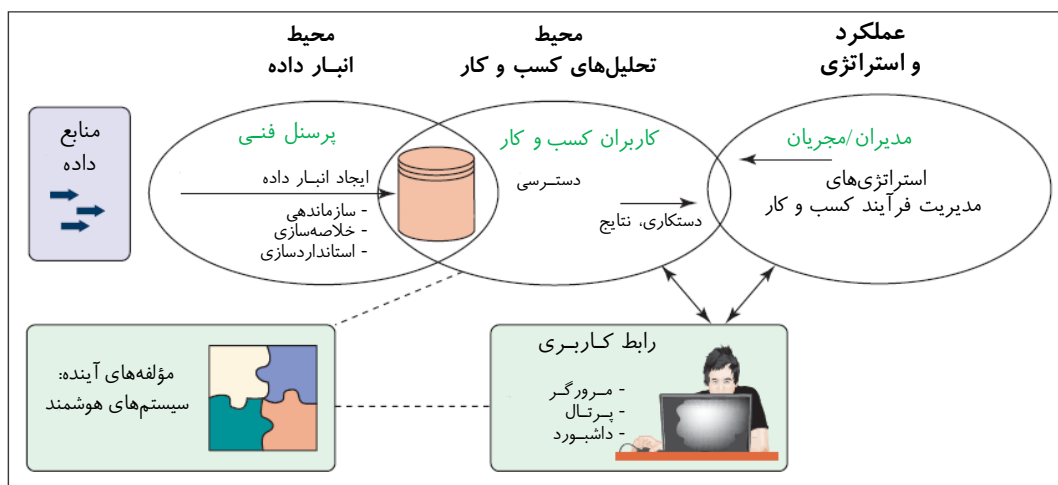
تاریخچه‌ی مختصر هوش تجاری اصطلاح هوش تجاری اولین بار توسط گروه گارتنر در اواسط دهه ۹۰ مطرح شد. با وجود این، همانطور که در بخش تاریخچه اشاره کرد، این مفهوم بسیار قدیمی‌تر از این است و ریشه در سیستم‌های گزارش‌گیری ام‌آی‌اس دهه ۷۰ دارد. در طی آن دوره، سیستم‌های گزارش‌گیری ایستا و دو بُعدی و فاقد هرگونه قابلیت‌های تحلیلی بودند. در اوایل دهه ۸۰، مفهوم سیستم‌های اطلاعات اجرایی ظهور کرد. این مفهوم، پشتیبانی رایانه‌ای را به مدیران اجرایی سطح بالا گسترش می‌داد. برخی از قابلیت‌های این سیستم‌ها شامل گزارش‌های چندبُعدی پویا (اقتضایی یا بر اساس تقاضا)، پیش‌بینی، تحلیل روند، نفوذ به عمق داده‌ها، دسترسی وضعیت و عوامل موفقیت بحرانی بودند. این ویژگی‌ها در دهه‌ها محصول تجاری و تا اواسط دهه ۹۰ وجود داشتند. سپس همین قابلیت‌ها و برخی موارد جدیدتر تحت نام هوش تجاری پدیدار شدند. امروزه، یک سیستم اطلاعات سازمانی خوب مبتنی بر هوش تجاری، حاوی تمامی اطلاعات مورد نیاز مدیران اجرایی است. بنابراین، مفهوم اولیه‌ی سیستم اطلاعات اجرایی به هوش تجاری تبدیل شد. تا سال ۲۰۰۵، سیستم‌های هوش تجاری کم‌کم از قابلیت‌های هوش مصنوعی و همچنین قابلیت‌های تحلیلی قدرتمند بهره‌مند شدند. شکل ۱.۳ تکنیک‌ها و ابزارهای مختلفی که ممکن است در یک سیستم هوش تجاری وجود داشته باشند را به تصویر کشیده است. در این شکل، تکامل هوش تجاری نیز نشان داده شده است. ابزارهای نشان داده شده در شکل ۱.۶ قابلیت‌های هوش تجاری را ارائه می‌کنند. اکثر این قالبیت‌ها در پیچیده‌ترین محصولات هوش تجاری وجود دارند ولی برخی دیگر از آنها جنبه‌ی تخصصی دارند.

یک سیستم هوش تجاری از چهار مؤلفه‌ی اصلی تشکیل شده است: یک *انبار داده* به همراه داده‌های آن؛ *تحلیل‌های کسب‌وکار*، مجموعه‌ای از ابزارها برای دستکاری، کاوش و تحلیل داده‌ها در انبار داده، مدیریت *فرآیند کسب‌وکار* برای پایش و تحلیل عملکرد و یک *رابط کاربری* (مثلاً یک داشبورد). ارتباط بین این مؤلفه‌ها در شکل ۱.۷ نشان داده شده است.



شکل ۱.۶ رشد و تکامل هوش تجاری

روش های پیشرفته‌ی انبار داده و هوش تجاری از کجا آمده‌اند؟ سرمنشأ آنها چیست و چگونه این ریشه‌ها امروزه بر نحوه‌ی مدیریت ابتکار عمل سازمان‌ها تأثیر می‌گذارند؟ بررسی‌های موشکافانه سرمایه‌گذاری‌های امروزی در فناوری اطلاعات از لحاظ تأثیر سود و زبانی افزایش یافته است. همین موضوع در خصوص انبار داده و برنامه‌های کاربردی هوش تجاری که این ابتکار عمل‌ها را امکان‌پذیر می‌سازند صادق است.



شکل ۱.۷ یک معماری سطح بالای هوش تجاری.

سازمان‌ها ناچار به اخذ، درک و تحت کنترل درآوردن داده‌هایشان به منظور پشتیبانی تصمیم‌گیری و بهبود عملیات کسب‌وکار هستند. قوانین و مقررات (مثلاً قانون ساربنز-آکسلی ۲۰۰۲) اکنون رهبران کسب‌وکارها را ملزم به مستندسازی فرایندهای کسب‌وکارشان و تأیید مشروعیت اطلاعاتی که به آنها اتکا کرده و به ذینفعان گزارش می‌کنند می‌نمایند. علاوه بر این، زمان‌های چرخه‌ی کسب‌وکارها امروزه بسیار فشرده‌تر شده و تصمیم‌گیری سریع‌تر، آگاهانه‌تر و بهتر به یک الزام در رقابت بدل شده است. مدیران به *اطلاعات درست در زمان درست* و در مکان درست نیاز دارند. این سرود مقدس برای رویکردهای پیشرفته‌ی هوش تجاری می‌باشد.

سازمان‌ها باید هوشمند کار کنند. توجه کامل به مدیریت راهکارهای هوش تجاری یکی از جنبه‌های ضروری در کسب‌وکارها است. پس جای تعجب نیست که سازمان‌ها هر چه بیشتر در حوزه هوش تجاری و ابزارهای تجزیه و تحلیل توانمند می‌شوند.

سیستم‌های هوش تجاری به انبارهای داده به عنوان منبع اطلاعاتی خود برای پشتیبانی از تصمیمات مدیریتی متکی هستند. حجم انبوهی از داده‌های سازمانی و برون‌سازمانی به منظور پشتیبانی به موقع و دقیق از تصمیمات، اخذ، تبدیل و در یک انبار داده ذخیره می‌شوند. به بیان ساده، یک انبار داده، مخزنی از داده‌های تولید شده به منظور پشتیبانی تصمیم‌گیری است و همچنین مخزنی از داده‌های جاری و قبلی مورد توجه مدیران در سرتاسر سازمان است. ساختار داده‌ها معمولاً به گونه‌ای است که در قالب آماده برای فعالیت‌های پردازش تحلیلی (یعنی پردازش تحلیلی آنلاین، داده‌کاوی، پرس و جو، گزارش‌گیری و سایر کاربردهای پشتیبانی تصمیم) ارائه می‌شوند. یک انبار داده، مجموعه‌ای ذهن‌محور و یکپارچه از داده‌هایی است که در طول زمان تغییر می‌کنند و برای پشتیبانی از فرایند تصمیم‌گیری مدیریت بکار می‌روند. اگرچه یک انبار داده، مخزنی از داده‌ها است ولی انبارش داده به کل فرایند اطلاق می‌شود. انبارش داده رشته‌ای است که منجر به توسعه‌ی برنامه‌های کاربردی با قابلیت پشتیبانی تصمیم شده و دسترسی به اطلاعات کسب‌وکار را فراهم کرده و نسبت به کسب‌وکاری که داریم بینش و ادراکی را ارائه می‌کند. سه نوع اصلی انبارهای داده عبارتند از بازارهای داده^۱، انبارهای داده‌های عملیاتی^۲ و انبارهای داده‌های سازمانی^۳. در حالیکه یک انبار داده، پایگاه‌های داده در سرتاسر سازمان را ادغام می‌کند، داده‌گاه‌ها معمولاً کوچکتر بوده و بر یک موضوع یا دپارتمان خاص تمرکز دارند. یک داده‌گاه، زیرمجموعه‌ای از انبار داده است که معمولاً از یک موضوع تشکیل شده است (مثلاً بازاریابی یا عملیات). یک انبار داده‌ی عملیاتی یک فرم نسبتاً به‌روز از فایل اطلاعات مشتریان را فراهم می‌کند. این نوع پایگاه داده اغلب به عنوان یک فضای آماده‌سازی واسطه برای یک انبار داده عمل می‌کند. برخلاف محتوای ایستای یک انبار داده، محتوای یک انبار داده عملیاتی در سرتاسر مسیر عملیات کسب‌وکار به‌روزآوری می‌شود. یک انبار داده‌ی سازمانی، یک انبار داده‌های بسیار بزرگ است که در سرتاسر سازمان و با هدف پشتیبانی تصمیم‌گیری بکار می‌رود. این ویژگی بزرگ بودن انبارهای داده سازمانی موجب شده تا بتوان به منظور اثربخشی هوش تجاری و برنامه‌های پشتیبانی تصمیم، داده‌ها را از منابع متعددی گرفته و در یک قالب استاندارد درآورد. انبارهای داده‌های عملیاتی قادر هستند تا داده‌ها را برای انواع بسیاری از سیستم‌های پشتیبانی تصمیم از جمله مدیریت ارتباط با مشتریان، مدیریت زنجیره تأمین،

¹ Data Marts (DM)

² Operational Data Stores (ODS)

³ Enterprise Data Warehouse (EDW)

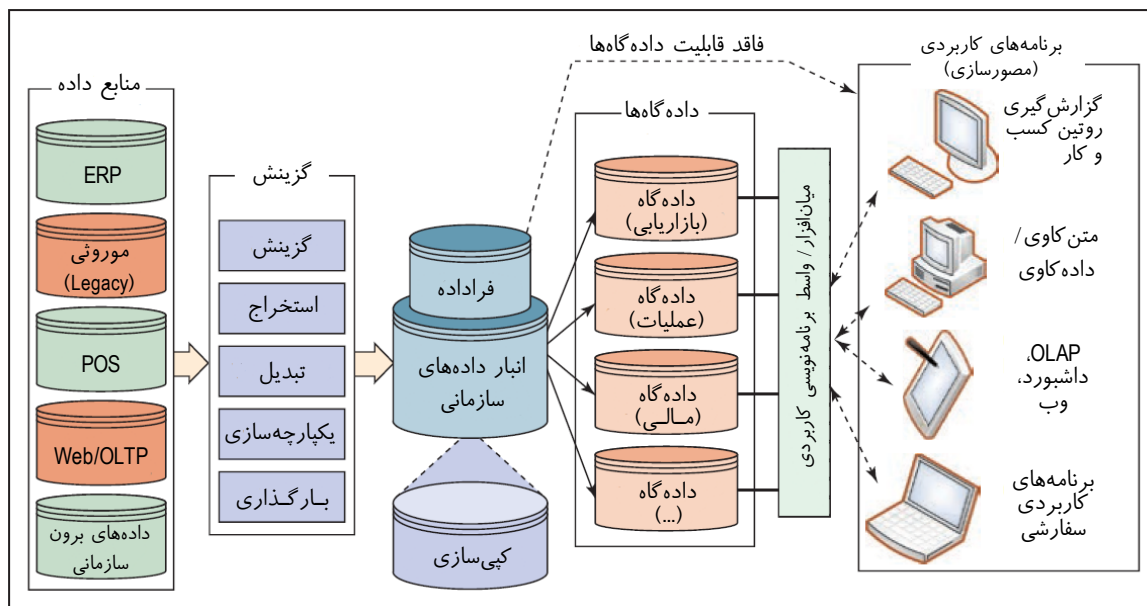
مدیریت فرایند کسب‌وکار، پایش فعالیت‌های کسب‌وکار، مدیریت چرخه‌ی عمر محصول، مدیریت درآمد و برخی اوقات حتی سیستم‌های مدیریت دانش فراهم کنند.

در شکل ۱.۸ مدل مفهومی انبار داده نشان داده شده است. داده‌ها از منابع متعدد و مختلفی استخراج و تبدیل شده و برای دسترسی‌ها و تحلیل‌های آتی به منظور پشتیبانی تصمیم در یک انبار داده ذخیره می‌شوند. جزئیات بیشتر در مورد انبار داده در فصل آنلاین وب سایت کتاب موجود است.

برای اینکه ویژگی‌های اصلی هوش تجاری را به تصویر بکشیم ابتدا باید نشان دهیم که هوش تجاری چه چیزی نیست که منظور در اینجا پردازش تراکنش است. همه ما با سیستم‌های اطلاعاتی از قبیل گرفتن پول از دستگاه‌های خودپرداز، سپرده‌های بانکی و دستگاه‌های کارت‌خوان در فروشگاه‌ها که از تراکنش‌های ما پشتیبانی می‌کنند آشنا هستیم. این سیستم‌های پردازش تراکنش^۱ به طور دائم در حال به‌روزرسانی چیزی هستند که ما آنها را پایگاه داده‌های عملیاتی می‌نامیم. به عنوان مثال، در یک تراکنش گرفتن پول از دستگاه خودپرداز، ما باید تراز سپرده بانکی مان را متناسب با آن کاهش دهیم، سپرده بانکی به یک حساب ما می‌افزاید و خرید از یک فروشگاه احتمالاً در محاسبات فروش کل روزانه آن فروشگاه نمایان می‌شود که متناسب با آن باید موجودی اقلامی که ما از فروشگاه خریده‌ایم نیز کاهش یابد. این سیستم‌های پردازش تراکنش آنلاین^۲ به کسب‌وکار مستمر و روتین شرکت‌ها رسیدگی می‌کنند. بر عکس آن، انبار داده یک سیستم متمایز است که داده‌هایی که برای تجزیه و تحلیل مورد استفاده قرار خواهند گرفت را ذخیره می‌کند. هدف از این تجزیه و تحلیل این است که به مدیریت این امکان را بدهد تا در میان داده‌ها به جستجوی اطلاعات مربوط به کسب‌وکارشان بپردازند و همچنین می‌توان برای ارائه‌ی پشتیبانی تصمیم عملیاتی یا تاکتیکی از آن استفاده کرد. به عنوان مثال، پرسنل صف بتوانند تصمیمات سریع‌تر یا آگاهانه‌تری بگیرند. انبارهای داده با داده‌های اطلاعاتی که در سیستم‌های پردازش تحلیلی آنلاین مورد استفاده قرار می‌گیرند کار می‌کنند.

¹ Transaction Processing

² Online Transaction Processing Systems (OLTP)



شکل ۱.۸ چارچوب و ویوهای یک انبار داده

اکثر داده‌های عملیاتی در سیستم‌های برنامه‌ریزی منابع سازمانی (و در سیستم‌های مکمل آنها مثل مدیریت زنجیره تأمین یا مدیریت ارتباط با مشتریان) در یک سیستم پردازش تراکنش آنلاین ذخیره می‌شوند که نوعی از پردازش رایانه‌ای است که در آن رایانه بلافاصله به درخواست‌های کاربر پاسخ می‌دهد. هر درخواست در اینجا یک تراکنش در نظر گرفته می‌شود که یک رکورد رایانه‌سازی شده از یک رویداد گسسته مثل رسید موجودی یا یک سفارش مشتری است. به عبارت دیگر، در یک تراکنش نیاز است تا دو یا چند پایگاه داده به شکل صفر یا صدی به‌روزروری و تکمیل شوند.

آن ویژگی در طراحی اولیه که موجب اثربخشی یک سیستم OLTP برای پردازش تراکنش‌ها شده، این سیستم را برای تهیه گزارشات، تحلیل‌ها و پرس و جوهای اقتضایی برای کاربران نهایی ناکارآمد ساخته است. در دهه ۸۰، بسیاری از کاربران، از رایانه‌های بزرگ‌شان تحت عنوان «سیاهچاله» یاد می‌کردند چون که تمامی اطلاعات وارد این رایانه‌ها می‌شد ولی هیچ چیزی از آنها بیرون نمی‌آمد. تمامی درخواست‌ها برای تهیه گزارشات می‌بایست توسط پرسنل حوزه فناوری اطلاعات برنامه‌نویسی می‌شد در حالیکه فقط گزارشات «از پیش‌برنامه‌نویسی شده» بر اساس زمانبندی مشخص تولید می‌شدند و امکان پرس و جوی آنی و اقتضایی عملاً وجود نداشت. اگرچه سیستم‌های برنامه‌ریزی منابع سازمانی مبتنی بر سرویس‌دهنده/سرویس‌گیرنده در سال‌های ۹۰ محبوبیت بیشتری داشتند ولی هنوز با آنچه کاربران نهایی به عنوان یک تحلیل تعاملی و گزارش‌گیری عملیاتی نیاز داشتند فاصله زیادی داشتند. برای رفع این مشکلات، مفاهیم انبار داده و هوش تجاری به وجود آمد.

انبارهای داده حاوی انواع بسیار زیادی از داده‌ها هستند که تصویری منسجم و یکپارچه از شرایط کسب‌وکار را در یک نقطه زمانی نمایش می‌دهد. هدف از اینکار، ایجاد یک زیرساخت پایگاه داده بود که همیشه به صورت آنلاین باشد و حاوی تمامی اطلاعات موجود در سیستم‌های OLTP از جمله داده‌های قبلی باشد ولی این داده‌ها می‌بایست به گونه‌ای سازماندهی و ساختاردهی شوند که فعالیت‌های پرس و جو، تحلیل و پشتیبانی تصمیم به سرعت و با کارآمدی بالا انجام‌پذیر شود. تفکیک سیستم OLTP از تحلیل و پشتیبانی تصمیم منجر به مزایای سیستم هوش تجاری می‌شود که قبل‌تر توضیح داده شد.

تی‌یوان^۱ حاوی ویدئوهایی مشابه نمایش تلویزیونی CSI است که مفاهیم تحلیل را در صنایع مختلف به تصویر می‌کشد. این ویدئوها را «تحقیق سناریوهای کسب‌وکار» (بی‌اس‌آی)^۲ می‌نامند. این ویدئوها نه تنها سرگرم‌کننده هستند بلکه سؤالاتی را نیز برای بحث و گفتگو در کلاس مطرح می‌کنند. برای اطلاعات بیشتر می‌تواند به آدرس‌های اینترنتی زیر مراجعه نمایید.

- 1- <https://www.teradatauniversitynetwork.com/Library/Items/BSI-The-Case-of-the-Misconnecting-Passengers/>
- 2- www.youtube.com/watch?v=NXEL5F4_aKA

ویدئوی موجود در سایت یوتیوب را تماشا کنید. شما در اصل باید نقش یک کارشناس مرکز خدمات مشتری را فرض کنید. یک پرواز ورودی تأخیر دارد و چند مسافر ممکن است پروازهای غیرمستقیم خود را از دست بدهند. یک پرواز خروجی دو صندلی خالی دارد و می‌تواند دو نفر از این چهار مسافر را سوار کند. اولویت با کدام مسافران است؟ اطلاعاتی در مورد شرح حال و روابط مسافران با شرکت هواپیمایی با شما داده می‌شود. با مطلع شدن شما از شرح حال مسافران، تصمیمات شما ممکن است تغییر کند.

ویدئو را تماشا کنید. هر جا که نیاز بود توقف کرده و به این سوال که اولویت با کدام مسافران است پاسخ دهید. آنگاه برای اطلاعات بیشتر به تماشای ادامه‌ی ویدئو پردازید. پس از اتمام ویدئو می‌توانید برای مشاهده‌ی اسلایدهای مربوط به این ویدئو و نحوه‌ی تحلیل موضوع به آدرس www.slideshare.net/teradata/bsi-how-we-did-it-the-case-of-the-misconnecting-passengers مراجعه کنید.

این گشت و گذار چندرسانه‌ای، مثالی از اینکه چگونه اطلاعات اضافی موجود در یک انبار داده‌ی سازمانی می‌توانند به تصمیم‌گیری کمک کنند ارائه می‌کند.

اگرچه برخی افراد سیستم پشتیبانی تصمیم را معادل هوش تجاری می‌گیرند ولی این سیستم‌ها در حال حاضر یکسان نیستند. شایان ذکر است که برخی معتقدند که سیستم پشتیبانی تصمیم، بخشی از هوش تجاری و در واقع یکی از ابزارهای تحلیلی آن است. برخی دیگر تصور می‌کنند که هوش تجاری یک حالت خاص از سیستم پشتیبانی تصمیم است که عمدتاً با گزارش‌گیری، ارتباطات و همکاری که شکلی از سیستم پشتیبانی تصمیم داده‌محور است سر و کار دارند. توضیح دیگری که واتسون (۲۰۰۵) ارائه کرده این است که هوش تجاری، ماحصل یک تکامل مستمر است و بر این اساس، سیستم پشتیبانی تصمیم یکی از اجزای اصلی و اولیه‌ی هوش تجاری است. علاوه بر این، همانطور که در بخش بعدی خواهیم گفت، در بسیاری از محافل، هوش تجاری را تحت عناوین جدیدی از قبیل *تحلیل یا علم داده‌ها*^۳ دسته‌بندی کرده‌اند.

برنامه‌ریزی مناسب و همراستایی با استراتژی کسب‌وکار ابتدا و مهم‌تر از هر کاری، علل زیربنایی برای سرمایه‌گذاری در هوش تجاری می‌بایست همراستا با استراتژی کسب‌وکار شرکت باشد. هوش تجاری صرفاً یک رویکرد فنی برای دپارتمان سیستم‌های اطلاعاتی نیست و باید با بهبود فرایندهای کسب‌وکار و تبدیل فرایندهای تصمیم‌گیری به فرایندهای داده‌محور به عنوان روشی برای تغییر نحوه انجام کسب‌وکار شرکت عمل کنند. بسیاری از مشاوران و کارشناسان هوش تجاری که در فعالیتهای موفق هوش تجاری مشارکت داشته‌اند توصیه می‌کنند که استفاده از یک چارچوب برنامه‌ریزی به عنوان پیش‌شرط اینکار ضروری است. یک چارچوب که توسط گارتنر (۲۰۰۴) پیشنهاد شده است فرایند برنامه‌ریزی و اجرا را به مؤلفه‌های کسب‌وکار، سازمان،

¹ TUN

² Business Scenario Investigations (BSI)

³ Data Science

کارکردگرایی^۱ و زیرساخت^۲ تجزیه کرده‌اند. در سطوح کسب‌وکار و سازمانی، اهداف استراتژیک و عملیاتی باید تعریف شده و مهارت‌های سازمانی موجود به منظور تحقق آن اهداف نیز در نظر گرفته شوند. مسائل فرهنگ سازمانی حول راهکارهای هوش تجاری و ایجاد اشتیاق برای آن راهکارها و رویه‌هایی برای به اشتراک‌گذاری بهترین تجارب هوش تجاری در درون سازمان نیز باید توسط مدیریت ارشد مد نظر قرار گرفته و برنامه‌هایی به منظور آماده‌سازی سازمان برای تغییر تعریف شود. یکی از اولین گام‌ها در آن فرایند، ارزیابی سازمان «سیستم‌های اطلاعاتی»، مهارت مجموعه گروه‌های کاربران و تمایل سازمان برای تغییر فرهنگ است. از این ارزیابی و این فرض که توجیه وجود دارد و باید به سمت جلو حرکت کرد، شرکت می‌تواند یک برنامه عملیاتی تفصیلی را تهیه کند. یک مسئله‌ی بحرانی دیگر برای موفقیت در پیاده‌سازی و اجرای هوش تجاری، یکپارچه‌سازی چندین پروژه‌ی هوش تجاری (اکثر سازمان‌ها از چندین پروژه‌ی هوش تجاری استفاده می‌کنند) بین خودشان و سایر سیستم‌های فناوری اطلاعات در سازمان و شرکای تجاری آنهاست.

گارتنر و بسیاری دیگر از سازمان‌های مشاوره در حوزه تحلیل مفهوم یک مرکز شایستگی هوش تجاری را مطرح کردند که برای کارکردهای زیر استفاده می‌شود:

- یک مرکز که نشان می‌دهد هوش تجاری چگونه به استراتژی و اجرای استراتژی مرتبط شده است.
- یک مرکز که می‌تواند برای ترویج تعامل بین کاربران بالقوه کسب‌وکار و سازمان سیستم اطلاعات ایفای نقش کند.
- یک مرکز که می‌تواند به عنوان یک مخزن و منتشرکننده‌ی بهترین تجارب هوش تجاری بین بخش‌های مختلف کسب‌وکار عمل کند.
- سازمان «سیستم‌های اطلاعات» می‌تواند مطالب بسیاری را از طریق تعامل با کاربران یاد بگیرد. مثل دانش در مورد تنوع و انواع تحلیل مورد نیاز.
- کاربران کسب‌وکار و سازمان «سیستم اطلاعات» می‌توانند درک بهتری از این موضوع داشته باشند که چرا لازم است تا پلتفرم انبار داده انعطاف‌پذیر باشد تا بتواند پاسخگوی نیازمندی‌های متغیر کسب‌وکار باشد.
- این مرکز می‌تواند به ذینفعان مهم مثل مدیران اجرایی سطح بالا در فهم نقش مهم هوش تجاری کمک کند.

طی ۱۰ سال گذشته، فکر یک مرکز شایستگی هوش تجاری رها شد چون که بسیاری از فناوری‌های پیشرفته که در این کتاب مطرح شده‌اند نیاز به یک گروه مرکزی برای سازماندهی بسیاری از این کارکردها را کاهش داده‌اند. هوش تجاری از آن حالت اولیه خود اکنون به نقطه‌ای رسیده است که بیشتر کارها توسط کاربران نهایی و در حالت «سلف سرویس» انجام‌پذیر است. به عنوان مثال، کاربران نهایی می‌توانند بسیاری از کارهای مصورسازی داده‌ها را به آسانی با استفاده از جدیدترین بسته‌های نرم‌افزاری مصورسازی انجام دهند (فصل ۳). طبق گفته‌ی دانکن (۲۰۱۶)، امروزه تیم‌های هوش تجاری تمرکز بیشتری را بر تولید مجموعه داده‌های گزینش شده معطوف کرده‌اند تا امکان هوش تجاری سلف سرویس فراهم شود. از آنجائیکه تجزیه و تحلیل‌ها در حال حاضر در کل سازمان نفوذ کرده‌اند، مرکز شایستگی هوش تجاری به منظور ترویج بهترین تجارب و حصول اطمینان از همراستایی راهکارهای تحلیلی با استراتژی سازمانی می‌تواند تبدیل به یک جامعه تعالی تحلیلی شود.

برخی اوقات، نیاز است تا به منظور ایجاد هم‌افزایی، ابزارهای هوش تجاری را با یکدیگر ادغام نمود. این نیاز به یکپارچه‌سازی، شرکت‌های تولید نرم‌افزار را مجبور کرد تا این قابلیت‌ها را به طور مستمر به محصولات خود اضافه کنند. مشتریانی که بسته‌های

¹ Functionality

² Infrastructure

نرم‌افزاری کامل را خریداری می‌کنند فقط با یک شرکت تولیدکننده طرف هستند و با اتصال سیستم کاری ندارند. ولی ممکن است مزیت ایجاد سیستم‌هایی که از بهترین اجزا و مؤلفه‌های تشکیل شده‌اند را از دست بدهند. این منجر به هرج و مرجی اساسی در فضای بازار هوش تجاری شد. بسیاری از ابزارهای نرم‌افزاری که سوار بر موج هوش تجاری بودند (مثلاً ساویون^۱، ویتریا^۲، تیبکو^۳، میکرواستراتژی^۴، هایپریون^۵) به تملک شرکت‌های دیگری درآمده یا اینکه به منظور بهره‌مندی از شش روند اصلی که از موج اول هوش تجاری ظهور کرده‌اند، محصولات و خدمات خود را گسترش دادند:

- کلان داده‌ها
- تمرکز بر تجربه‌ی مشتری بر خلاف تمرکز صرف بر اثربخشی عملیات
- رابط‌های کاربری سیار و حتی جدیدتر (دیداری، صوتی)
- تحلیل‌های پیشگویانه و تجویزی، یادگیری ماشین، هوش مصنوعی
- مهاجرت به فضای ابری
- تمرکز بسیار بیشتر بر امنیت و محافظت از حریم خصوصی

این کتاب بسیاری از این عناوین را با جزئیات فراوان و ارائه مثال‌هایی از نحوه‌ی رشد و تکامل و بکارگیری این فناوری‌ها به همراه نکات مدیریتی پوشش داده است.

سوالاتی برای مرور بخش ۱.۴

۱. سه عبارت که پیشینیان برای تحلیل بکار می‌گرفتند را نام ببرید.
۲. تفاوت اصلی بین سیستم‌های اطلاعات مدیریت، پشتیبانی تصمیم و اطلاعات اجرایی چیست؟
۳. آیا سیستم پشتیبانی تصمیم به هوش تجاری تکامل یافت یا بالعکس؟
۴. هوش تجاری را تعریف کنید.
۵. اجزای اصلی هوش تجاری را نام برده و توضیح دهید.
۶. پردازش تراکنش آنلاین (OLTP) را تعریف کنید.
۷. پردازش تحلیلی آنلاین (OLAP) را تعریف کنید.
۸. برخی از موضوع مربوط به پیاده‌سازی که در گزارش گارتنر به آنها اشاره شده را نام ببرید.
۹. برخی از سایر عوامل موفقیت هوش تجاری را نام ببرید.

¹ Savvion

² Vitria

³ Tibco

⁴ MicroStrategy

⁵ Hyperion

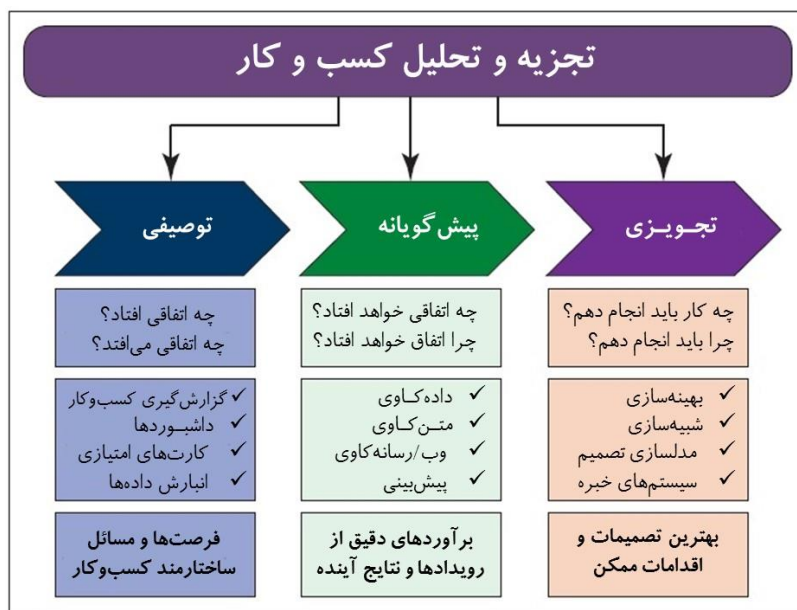
۱.۵ مروری بر علم تجزیه و تحلیل

واژه‌ی تجزیه و تحلیل^۱ تا حدود زیادی جایگزین تک تک اجزای قبلی فناوری‌های پشتیبانی تصمیم رایانه‌ای که تحت عناوین مختلفی در گذشته وجود داشته‌اند شده است. علاوه بر این، بسیاری از متخصصان و دانشگاهیان امروز از واژه‌ی تجزیه و تحلیل به جای هوش تجاری استفاده می‌کنند. اگرچه بسیاری از نویسندگان و مشاوران آن را کمی متفاوت‌تر تعریف کرده‌اند ولی تجزیه و تحلیل را می‌توان به عنوان فرایند ایجاد تصمیمات قابل اقدام یا توصیه‌هایی برای اقدام بر اساس دانش بدست آمده از داده‌های تاریخی دانست. بر اساس تعریف مؤسسه‌ی پژوهش عملیات و علوم مدیریت^۲، تجزیه و تحلیل نماینده‌ی ترکیبی از فناوری رایانه‌ای، تکنیک‌های علم مدیریت و آمار است که برای حل مسائل واقعی بکار می‌روند. البته، بسیاری از سازمان‌ها تفسیرها و انگیزه‌های خاص خودشان از تحلیل را ارائه کرده‌اند. به عنوان مثال، مؤسسه‌ی اس‌ای‌اس^۳ برای تحلیل هشت سطح را در نظر گرفته است که با گزارش استانداردسازی شده از یک سیستم رایانه‌ای شروع می‌شود. این گزارشات در واقع حسی از آنچه در یک سازمان در حال رخ دادن است را به ما می‌دهد. فناوری‌های بیشتر به ما این امکان را می‌دهند تا گزارشات سفارشی‌تری را تولید نمود. سطح بعدی در گزارش‌گیری ما به سمت پرس و جوهای پردازش تحلیلی آنلاین می‌برد که به کاربر اجازه می‌دهد تا عمیق‌تر کاوش نموده و به منابع خاصی از فرصت‌ها یا تهدیدها دست یابند. فناوری‌های امروزی همچنین می‌توانند در صورت نیاز هشدارهایی را به صورت خودکار برای تصمیم‌گیرنده صادر کنند. در سطح مصرف‌کننده اینگونه هشدارها را برای آب و هوا و سایر مسائل می‌بینیم. اما هشدارهای مشابهی را نیز می‌تواند در شرایط و محیط‌های خاص مثل فروش کمتر یا بیشتر از یک سطح مشخص در یک بازه زمانی مشخص یا کاهش سطح موجودی محصولی خاص تولید کرد. تمامی این کاربردها از طریق تجزیه و تحلیل و پرس و جو داده‌های گردآوری شده توسط یک سازمان میسر شده‌اند. سطح بعدی تجزیه و تحلیل شامل تحلیل آماری به منظور درک بهتر الگوها است. اینها را می‌توان برای پیش‌بینی یا ساخت مدل‌های پیش‌بینی نحوه پاسخگویی مشتری به یک طرح تبلیغات بازاریابی خاص یا محصولات/خدمات ارائه شده بکار برد. وقتی یک سازمان دید خوبی نسبت به آنچه در حال اتفاق افتادن است و آنچه احتمال دارد اتفاق بیوفتد داشته باشد می‌تواند از تکنیک‌های دیگری برای اتخاذ بهترین تصمیمات تحت این شرایط استفاده کند.

¹ Analytics

² Institute For Operation Research and Management Science (INFORMS)

³ SAS Inc.



شکل ۱.۹ سه نوع ابزار تجزیه و تحلیل

مؤسسه‌ی اینفورمز^۱ این تفکر نگاه به تمامی داده‌ها برای درک اینکه چه اتفاقی می‌افتد، چه اتفاقی خواهد افتاد و چگونه بهترین بهره‌برداری را از این موضوع کنیم را در قالب ابزار تجزیه و تحلیل سه سطحی پیشنهاد کرده است. این سه سطح به ترتیب عبارتند از توصیفی، پیش‌گویانه و تجویزی. شکل ۱.۹ یک نمای گرافیکی از این سه سطح تحلیل را به تصویر کشیده است. این تصویر به طور ضمنی نشان می‌دهد که این سه سطح تا حدودی مستقل از یکدیگر هستند و یک نوع از کاربردهای تحلیلی منجر به کاربرد بعدی می‌شود. همچنین بیانگر این است که این سه نوع تحلیل تا حدودی همپوشانی دارند. در هر حالت، مشهود است که این کاربردهای متفاوت بهم مرتبط تحلیل می‌باشند. در بخش بعدی به معرفی این سه سطح ابزار تحلیل خواهیم پرداخت.

تجزیه و تحلیل توصیفی (گزارش‌گیری) به این اشاره دارند که بدانیم در سازمان چه اتفاقی می‌افتد و برخی روندهای زیربنایی و علل وقوع آنها را درک کنیم. ابتدا، تثبیت و ادغام منابع داده و در دسترس بودن تمامی داده‌های مرتبط به شکلی است که امکان گزارش‌گیری و تجزیه و تحلیل مناسب را فراهم آورند. معمولاً، توسعه‌ی این زیرساخت داده‌ها، بخشی از انبارهای داده است. از این زیرساخت داده می‌توانیم گزارشات، پرس و جوها، هشدارها و روندهای مناسب را با استفاده از ابزارها و تکنیک‌های گزارش‌های متنوع ایجاد کنیم.

یک فناوری بسیار مهم که به یک فعال اصلی در این حوزه تبدیل شده، مصورسازی می‌باشد. ما اکنون با استفاده از جدیدترین ابزارهای مصورسازی موجود در بازار می‌توانیم درک عمیق و خوبی از عملیات در سازمان مان بدست آوریم. مورد کاربردی ۱.۳ و ۱.۴ به بررسی دقیق‌تر این کاربردها می‌پردازند.

¹ Informs

سیلواریس، کسب و کارش را با استفاده از قابلیت‌های تجزیه و تحلیل دیداری و گزارش‌گیری آنی ارتقاء می‌دهد.

مورد کاربردی ۱.۳

شرکت سیلواریس^۱ در سال ۲۰۰۰ توسط تیمی از متخصصان صنعت جنگل‌داری و به منظور ارائه دستاوردهای تکنولوژیکی در بخش الوار و مصالح ساختمانی تأسیس شد. سیلواریس اولیت پلتفرم تجارت الکترونیک تخصصی محصولات جنگل‌داری در ایالات متحده می‌باشد که دفتر مرکزی آن در سیاتل ایالت واشنگتن واقع شده است. این شرکت یک عمده‌فروش رهبر در محصولات چوبی صنعتی و مصالح ساختمانی مازاد است.

سیلواریس به فروش محصولاتش پرداخته و خدمات لجستیک بین‌المللی را به بیش از ۳,۵۰۰ مشتری ارائه می‌کند. این شرکت به منظور مدیریت فرایندهای مختلفی که در معاملات و تراکنش‌هایش وجود دارند یک پلتفرم تجاری آنلاین انحصاری را به منظور ردگیری جریان اطلاعات مربوط به تراکنش‌ها بین مبادله‌کنندگان، حسابداری، اعتبارات و لجستیک ایجاد کرده است. این به شرکت سیلواریس اجازه می‌دهد تا اطلاعات لحظه‌ای خود را با مشتریان و شرکایش به اشتراک بگذارد. ولی به دلیل تغییرات قیمتی سریع مصالح، سیلواریس مجبور بود تا بدون اینکه داده‌ها را به قالب گزارش‌گیری جداگانه‌ای منتقل کند، دسترسی آنی و بلادرنگ به آن داده‌ها پیدا کند.

سیلواریس شروع به استفاده از نرم‌افزار «تابلو» به دلیل قابلیت اتصال به داده‌ها و مصورسازی زنده‌ی آنها نمود. سیلواریس با توجه به داشبورهای ساده و قابل فهم این نرم‌افزار شروع به استفاده از آن برای مقاصد گزارش‌گیری کرد. اینکار به سیلواریس کمک کرد تا اطلاعات را سریعاً از درون داده‌ها بیرون کشیده و مسائلی که بر کسب و کارشان تأثیر می‌گذاشت را شناسایی کنند. سیلواریس به کمک گزارشات تهیه شده توسط تابلو مفوق شد تا سفارشات آنلاین و آفلاین را مدیریت کند. اکنون، سیلواریس سفارشات آنلاین مشتریان را ردگیری کرده و می‌داند برای ترغیب مشتریان به ادامه خرید آنلاین، چه وقت سفارش‌های جدید را ارسال کند. همچنین، تحلیل‌گران سیلواریس با تولید داشبوردها به کمک تابلو به جای نوشتن گزارشات صد صفحه‌ای می‌توانند در زمان صرفه‌جویی کنند.

Sources: Tableau.com. "Silvaris Augments Proprietary Technology Platform with Tableau's Real-Time Reporting Capabilities."
http://www.tableau.com/sites/default/files/case-studies/silvarisbusiness-dashboards_0.pdf (accessed September 2018); Silvaris.com.
<http://www.silvaris.com> (accessed September 2018).

سوالاتی برای مورد کاربردی ۱.۳

۱. سیلواریس با چه چالشی مواجه بود؟
۲. سیلواریس چگونه این مسئله را با استفاده از مصورسازی داده‌ها در تابلو حل کرد؟

از این مورد کاربردی چه می‌توانیم بیاموزیم

بسیاری از صنایع نیاز دارند تا داده‌ها را به صورت لحظه‌ای و آنی تجزیه و تحلیل کنند. تحلیل‌های بلادرنگ این امکان را به تحلیل‌گران می‌دهد تا مسائل تأثیرگذار بر کسب و کارشان را شناسایی کنند. مصورسازی برخی اوقات بهترین روش برای شروع

¹ Silvaris

تحلیل جریان داده‌های زنده است. نرم‌افزار تابلو یکی از ابزارهای مصورسازی داده‌ها است که قابلیت تجزیه و تحلیل داده‌های زنده را بدون آنکه آنها را وارد یک قالب گزارش‌گیری جداگانه کند دارد.

مورد کاربردی ۱.۴

زیمنس هزینه‌هایش را با استفاده از مصورسازی داده‌ها کاهش می‌دهد.

زیمنس^۱ یک شرکت آلمانی واقع در برلین است. این شرکت یکی از بزرگ‌ترین شرکت‌های دنیا در حوزه تخصصی اتوماسیون، الکتریکی‌سازی و دیجیتال‌سازی می‌باشد. درآمد سالانه این شرکت ۷۶ میلیارد یورو می‌باشد. گروه تحلیل‌های دیداری زیمنس وظیفه‌ی توسعه‌ی راه‌حل‌های گزارش‌گیری انتها به انتها و مشاوره برای تمامی نیازهای هوش تجاری درون سازمانی زیمنس را بر عهده دارد. این گروه با چالش ارائه‌ی راه‌حل‌های گزارش‌گیری به کل سازمان در دپارتمان‌های مختلف و حفظ تعادل بین قابلیت‌های حکمرانی و سلف سرویس روبرو بود. زیمنس به پلتفرمی نیاز داشت تا بتواند نظرسنجی‌های متعدد رضایت مشتری، فرایندهای لجستیک و گزارش‌گیری‌های مالی را تجزیه و تحلیل نماید. این پلتفرم باید برای استفاده‌ی کارکنان آسان باشد تا بتوانند از این داده‌ها برای تجزیه و تحلیل و تصمیم‌گیری استفاده کنند. علاوه بر این، این پلتفرم باید براحتی با سیستم‌های فعلی زیمنس ادغام شده و تجربه‌ی بی‌نظیر و کاملی را برای کاربران فراهم کند. زیمنس شروع به استفاده از هوش تجاری دونداس^۲ نمود که یک رهبر جهانی در ارائه راه‌حل‌های مصورسازی داده‌ها و هوش تجاری است. زیمنس توانست داشبوردهای تعاملی سطح بالایی را ایجاد کند که به شناسایی زود هنگام مشکلات کمک کرده و صرفه‌جویی مالی قابل توجهی را به همراه داشت. داشبوردهایی که هوش تجاری دونداس ارائه می‌کرد به سازمان لجستیک بین‌المللی زیمنس کمک کرد تا به سوالاتی از قبیل اینکه چگونه نرخ تأمین‌های مختلف در مکان‌های مختلف بر عملیات تأثیر می‌گذارد پاسخ داده و بدین ترتیب به شرکت کمک کرد تا زمان سیکل را به میزان ۱۲ درصد و هزینه اسقاط را تا ۲۵ درصد کاهش دهد.

Sources: Tableau.com. "Silvaris Augments Proprietary Technology Platform with Tableau's Real-Time Reporting Capabilities."
http://www.tableau.com/sites/default/files/case-studies/silvarisbusiness-dashboards_0.pdf (accessed September 2018); Silvaris.com.
<http://www.silvaris.com> (accessed September 2018).

هدف تجزیه و تحلیل پیش‌گویانه این است که تعیین کند در آینده چه چیزی احتمالاً رخ خواهد داد. این تجزیه و تحلیل مبتنی بر تکنیک‌های آماری و همچنین تکنیک‌های جدیدتری است که در دسته‌بندی عمومی داده‌کاوی قرار می‌گیرند. هدف از این تکنیک‌ها این است که بتوان پیش‌بینی کرد که آیا احتمال دارد که مشتری جذب رقیب دیگری شود، مشتری احتمالاً در نوبت بعدی چه چیزی را و چگونه خریداری خواهد کرد، مشتری به چه تبلیغاتی پاسخ خواهد داد، آیا مشتری ارزش ریسک کردن برای شرکت را دارد و غیره. چندین تکنیک در توسعه‌ی برنامه‌های تحلیلی پیش‌گویانه از جمله الگوریتم‌های مختلف دسته‌بندی بکار می‌رود. به عنوان مثال، همانطور که در فصل‌های ۴ و ۵ توضیح داده شده است، ما می‌توانیم از تکنیک‌های دسته‌بندی از قبیل رگرسیون لجستیک، مدل‌های درخت تصمیم و شبکه‌های عصبی برای پیش‌بینی فروش یک فیلم در گیشه استفاده کرد. ما همچنین می‌توانیم از الگوریتم‌های خوشه‌بندی برای بخش‌بندی مشتری به خوشه‌های مختلف استفاده کنیم تا بتوانیم تبلیغات مشخصی را برای هر دسته طراحی کنیم. نهایتاً، می‌توان از تکنیک‌های وابستگی کاوی به منظور برآورد روابط بین رفتارهای

¹ Siemens

² Dundas BI

مختلف خرید استفاده کرد. این بدان معناست که اگر یک مشتری، محصولی را خریداری کند، احتمالاً چه محصول دیگری را نیز در کنار آن خواهد خرید؟ اینگونه تجزیه و تحلیل‌ها می‌توانند به یک خرده‌فروش کمک کنند تا محصولات مرتبط با تبلیغ یا توصیه کند. به عنوان مثال، هر جستجو محصول در آمازون، کام، محصولات مشابهی که مشتری ممکن است علاقمند به خرید آنها باشد را نیز نتیجه می‌دهد. ما این تکنیک‌ها و کاربردهایشان را در فصل‌های ۳ الی ۶ مطالعه خواهیم کرد. مورد کاربردی ۱.۵ یکی از این کاربردها را در حوزه ورزش نشان می‌دهد.

تجزیه و تحلیل آسیب‌های ورزشی

مورد کاربردی ۱.۵

در هر فعالیت ورزشی، احتمال وقوع آسیب وجود دارد. اگر به آسیب‌ها به طور مناسب رسیدگی نشود، آنگاه کل تیم متضرر خواهد شد. استفاده از تحلیل به منظور درک آسیب‌ها می‌تواند درک ارزشمندی به مربیان و پزشکان تیم بدهد تا بتوانند ترکیب تیم را مدیریت کرده، شرح حال بازیگران را درک کرده و نهایتاً به تصمیم‌گیری بهتر در مورد در دسترس بودن بازیکن در هر زمان کمک خواهد کرد.

در یک مطالعه‌ی اکتشافی، دانشگاه ایالتی آکلاهما به تجزیه و تحلیل آسیب‌های ورزش فوتبال ایالات متحده با استفاده از ابزارهای پیش‌گویانه و گزارش‌گیری پرداخت. در این پروژه از یک متدولوژی فرایند استاندارد داده‌کاوی بین صنعتی برای درک مسئله‌ی ارائه توصیه به منظور مدیریت آسیب‌ها، درک داده‌های مختلف گردآوری شده در مورد آسیب‌ها، پاک‌سازی داده، مصورسازی‌ها به منظور استنتاج کردن، ایجاد مدل‌های پیش‌گویانه به منظور تجزیه و تحلیل دوره زمانی درمان آسیب و استخراج قواعد توالی به منظور پیش‌بینی روابط بین آسیب‌ها و بخش‌های مختلف بدن که آسیب دیده‌اند استفاده شده است.

مجموعه داده‌های مربوط به آسیب‌ها از بیش از ۵۶۰ رکورد آسیب در ورزش فوتبال تشکیل شده که به تفکیک متغیرهای خاص برای آسیب (عضو بدن/محل، اقدام اتخاذ شده، شدت، نوع آسیب، زمان شروع آسیب و درمان آن) و متغیرهای خاص بازیکن/ورزش (شماره شناسایی بازیکن، پست بازیکن، فعالیت و مکان بازی) دسته‌بندی شده‌اند. زمان درمان برای هر رکورد محاسبه شد که به مجموعه‌های زمانی مختلف ۰-۱ ماه، ۱-۲ ماه، ۲-۴ ماه، ۴-۶ ماه و ۶-۲۴ ماه تقسیم شدند.

مصورسازی‌های مختلفی برای استنتاج دوره زمانی درمان مرتبط با پست هر بازیکن، شدت آسیب‌ها و دوره زمانی درمان، روش درمانی بکار رفته و دوره زمانی درمان، آسیب‌های جدی وارده بر قسمت‌های مختلف بدن و غیره ایجاد شدند.

مدل‌های شبکه‌های عصبی برای پیش‌بینی هر یک از دسته‌بندی‌های درمان با استفاده از مدل‌ساز اس‌پی‌اس‌آی‌بی‌ام^۱ ساخته شدند. برخی از متغیرهای پیش‌بینی شامل وضعیت فعلی آسیب، شدت، قسمت بدن، مکان بدن، نوع آسیب، فعالیت، مکان رویداد، اقدام به عمل آمده و پست بازیکن بودند. دسته‌بندی درمان بر اساس این متغیرها با دقت ۷۹.۶٪ موفقیت‌آمیز بود. توصیه‌های بسیاری بر اساس این تجزیه و تحلیل ارائه شد از قبیل استفاده بیشتر از اطلاعات متخصصان در خصوص آسیب‌ها به جای اینکه به کارکنان اتاق تمرین اجازه دهیم تا بازیکن آسیب دیده را رصد کنند؛ آموزش بازیکنان پست دفاع برای اجتناب از آسیب‌دیدگی و بکارگیری مکانیزم‌های کامل بررسی ایمنی.

Sources: "Sharda, R., Asamoah, D., & Ponna, N. (2013). "Research and Pedagogy in Business Analytics: Opportunities and Illustrative Examples." *Journal of Computing and Information Technology*, 21(3), pp. 171-182.

¹ IBM SPSS Modeler

سوالاتی برای مورد کاربردی ۱.۵

۱. در تجزیه و تحلیل آسیب‌ها از چه نوع تحلیل‌هایی استفاده شده است؟
۲. مصورسازی‌ها چگونه به درک داده‌ها و دستیابی به بینش و آگاهی در خصوص داده‌ها کمک می‌کنند؟
۳. منظور از دسته‌بندی یک مسئله چیست؟
۴. از تجزیه و تحلیل توالی به چه چیزی می‌تواند دست یافت؟

از این مورد کاربردی چه می‌توانیم بیاموزیم

برای هر پروژه‌ی تجزیه و تحلیل همواره مهم است که دامنه کسب‌وکار و وضعیت جاری مسئله را از طریق تجزیه و تحلیل گسترده‌ی تنها منبع موجود یعنی داده‌های تاریخی درک کنیم. مصورسازی‌ها اغلب ابزار بسیار خوبی برای دستیابی به آگاهی و بینش در مورد داده‌ها محسوب می‌شوند که بعداً می‌توان آن را بر اساس عقاید افراد خبره برای شناسایی اهمیت نسبی داده‌های مسئله تصفیه کرد. مصورسازی‌ها همچنین به تولید نظرات در مورد مسائل مبهم کمک می‌کنند که می‌توانند به ساخت PMها که سازمان‌ها را در تصمیم‌گیری‌ها کمک می‌کنند منجر شوند.

دسته‌بندی سوم تحلیل‌ها، **تجزیه و تحلیل تجویزی** نام دارد. هدف از ابزارهای تجویزی این است که ببینیم وضعیت جاری و محتمل در آینده چیست و تصمیماتی را برای دستیابی به بهترین عملکرد ممکن اتخاذ کنیم. این گروه از تکنیک‌ها تا پیش از این تحت علوم مدیریت مطالعه شده و هدف آنها معمولاً بهینه‌سازی عملکرد یک سیستم است. هدف در اینجا این است که برای یک اقدام خاص، یک تصمیم گرفته یا یک توصیه ارائه کنیم. این توصیه‌ها می‌توانند به شکل یک تصمیم بله/خیر برای یک مسئله، یک مقدار عددی مشخص (قیمت یک قلم کالا) یا مجموعه کاملی از طرح‌های تولید باشد. تصمیمات ممکن است در قالب یک گزارش یا مستقیماً در یک سیستم خودکار قواعد تصمیم (مثلاً در سیستم‌های قیمت‌گذاری خطوط هوایی) به تصمیم‌گیرنده نمایش داده شود. بنابراین، این نوع تحلیل‌ها را می‌توان تحلیل تصمیم یا هنجاری^۱ نیز نامید. مورد کاربردی ۱.۶ مثالی از اینگونه کاربردهای تحلیلی تجویزی ارائه می‌کند. ما در فصل ۸ به بررسی برخی جنبه‌های تحلیل تجویزی می‌پردازیم.

کاربرد تحلیل در حوزه‌های مختلف کاربردهای تحلیل در بخش‌های مختلف صنعت موجب شده تا حوزه‌های مرتبط یا عبارات و واژگان باب روز بسیاری بوجود آید. افزودن واژه‌ی **تحلیل** به هر صنعت یا نوع داده‌ی خاص بسیار متداول شده است. علاوه بر دسته‌بندی عمومی تحلیل متن که هدف آن استخراج ارزش از متن است (فصل ۷)، یا تحلیل وب که به تجزیه و تحلیل جریان داده‌های وب می‌پردازد، مشاغل و جریان‌های تحلیلی خاص نیز توسعه یافته‌اند. مثال‌هایی از این حوزه‌ها عبارتند از تحلیل بازاریابی، تحلیل خرده‌فروشی، تحلیل کلابرداری، تحلیل حمل و نقل، تحلیل سلامت، تحلیل ورزش‌ها، ابزارهای تحلیل استعداد، تحلیل‌های رفتاری و غیره. به عنوان مثال، ما چندین کاربرد **تحلیل ورزش** را به زودی خواهیم دید. مورد کاربردی ۱.۵ را می‌توان یک مطالعه موردی در تحلیل سلامت دانست. در بخش بعدی به معرفی کلی تحلیل سلامت و تحلیل بازار خواهیم پرداخت. به لحاظ لفظی، هر گونه تجزیه و تحلیل سیستماتیک داده‌ها در یک بخش خاص را «**تحلیل**» می‌نامند. اگرچه این ممکن است موجب ترویج بیش از حد مفهوم تحلیل شود، مزیت آن این است که افراد بیشتری در صنایع خاص از قدرت و پتانسیل تحلیل‌ها

¹ Normative

آگاه می‌شوند. همچنین به متخصصان و کارشناسان کمک می‌کند تا بر توسعه و بکارگیری مفاهیم تحلیل‌ها در یک بخش عمودی صنعت تمرکز کنند. اگرچه بسیاری از تکنیک‌های توسعه برنامه‌های کاربردی تحلیلی، عمومیت دارند ولی در هر بخش عمودی صنعت، مسائل منحصر بفردی وجود دارد که می‌تواند بر نحوه گردآوری، پردازش، تجزیه و تحلیل و پیاده‌سازی داده‌ها و برنامه‌های کاربردی تأثیر بگذارد. بنابراین، متمایزسازی تحلیل‌ها بر اساس یک تمرکز عمودی برای رشد کلی این رشته بسیار مفید است.

یک شرکت تخصصی تولید شمش فولاد از تحلیل‌ها برای تعیین زمان موجودی قول دادنی استفاده می‌کند

مورد کاربردی ۱.۶

این مورد کاربردی مبتنی بر پروژه‌ای است که یکی از نویسندگان شرکت A که نخواست نامش (یا حتی صنعت مربوطه) فاش شود با مشکل تصمیم‌گیری مواجه بوده که به منظور تأمین نیاز مشتریان از موجودی کدام مواد اولیه استفاده نماید. این شرکت شمش‌های فولادی سفارشی برای مشتریانش تولید می‌کند. این شمش‌ها ممکن است به شکل‌ها یا اندازه‌های خاص بریده شده و مواد اولیه یا پرداخت‌کاری‌های منحصر بفردی نیاز داشته باشند. این شرکت مواد اولیه خود را از سرتاسر جهان تأمین کرده و آنها را در انبار خود ذخیره می‌کند. زمانیکه یک مشتری برای استعلام قیمت شمش‌های تخصصی که مطابق با نیازمندی‌ها و الزامات خاص آن شرکت (ترکیب، منشأ فلز، کیفیت، اندازه، شکل و ...) است با تولیدکننده تماس می‌گیرد، پرسنل فروش معمولاً زمان بسیار کمی برای پاسخ به آن استعلام در اختیار دارند که شامل تاریخ ارسال محصول و البته قیمت و موارد دیگر است. وی باید تصمیمات «موجودی قول دادنی»^۱ را اتخاذ کند که زمان حقیقی تحویل یا ارسال محصولات درخواستی مشتریان در طی مرحله‌ی اخذ قیمت پیشنهادی را تعیین می‌کند. قبلاً، فروشنده می‌بایست اینگونه تصمیمات را از طریق تجزیه و تحلیل گزارشات موجودی در دسترس مواد اولیه اتخاذ می‌کرد. برخی از مواد اولیه‌ی موجود ممکن است به سفارش مشتری دیگری تخصیص داده شده باشد. بنابراین، موجودی انبار ممکن واقعاً بیانگر موجودی در دسترس نباشد. از طرف دیگر، ممکن است مواد اولیه به صورت سفارش در راه داشته باشیم که در آینده‌ی نزدیک وارد انبار شرکت شده و بتوان از آن برای تأمین سفارش مشتری استفاده کرد. نهایتاً، حتی ممکن است فرصت این را داشته باشیم که در صورت تأخیر در انجام یک سفارش نهایی، مجدداً موجودی در نظر گرفته شده برای آن سفارش به سفارش جدیدی تخصیص داده و یک هزینه‌ی اضافی را بابت آن دریافت کرد. البته، اینگونه تصمیمات باید بر اساس تجزیه و تحلیل‌های هزینه-منفعت و برای سفارشی که به تعویق افتاده است اتخاذ شود. بنابراین سیستم باید قادر به استخراج داده‌های بلادرنگ در خصوص موجودی، سفارشات تعهد شده، مواد اولیه‌ی ورودی، محدودیت‌های تولید و غیره باشد.

به منظور پشتیبانی این تصمیمات «موجودی قول دادنی»، یک سیستم پشتیبانی تصمیم‌آنی برای یافتن بهترین حالت تخصیص موجودی در دسترس و پشتیبانی از تجزیه و تحلیل‌های چه-اگر اضافی را توسعه داد. سیستم پشتیبانی تصمیم‌آنی از مجموعه‌ای از مدل‌های برنامه‌نویسی عدد صحیح مختلط که با استفاده از نرم‌افزارهای تجاری حل شده‌اند استفاده می‌کند. این شرکت، برای تسهیل استفاده از تحلیل‌های کسب‌وکار خود، سیستم پشتیبانی تصمیم‌آنی را با سیستم برنامه‌ریزی منابع سازمانی خود ادغام نموده است.

سوالاتی برای مورد کاربردی ۱.۶

۱. چرا تخصیص مجدد موجودی از یک مشتری به مشتری دیگر، مسئله مهمی است؟
۲. یک سیستم پشتیبانی تصمیم‌آنی چگونه به گرفتن این تصمیمات کمک می‌کند؟

¹ Available-to-Promise (ATP)

تحلیل یا علم داده‌ها حتی با وجود اینکه مفهوم تحلیل در صنایع و بین دانشگاهیان بیش از پیش مورد توجه قرار گرفته است ولی عبارت دیگری نیز معرفی شده و در حال معروف شدن است. این عبارت جدید، علم داده‌ها^۱ نام دارد. بنابراین، کارشناسان علم داده‌ها، دانشمندان داده هستند. برخی اوقات، دکتر جی پاتیل در لینکدین را مبدع عبارت علم داده‌ها می‌شناسند. تلاش‌هایی برای توضیح تفاوت‌های بین تحلیلگرهای داده و دانشمندان داده انجام شده است (مثلاً نگاه کنید به "Data Science (Revealed)", 2018) (emc.com/collateral/about/news/emc-data-science-study-wp.pdf). یک دیدگاه این است که تحلیلگر داده فقط عبارت دیگری برای نامیدن متخصصان و کارشناسانی است که هوش تجاری را به شکل کامپایل داده‌ها، پاکسازی، گزارش‌گیری و شاید تا حدودی مصورسازی انجام می‌دهند. مجموعه مهارت‌های آنها شامل استفاده از اکسل، دانش محدود اس‌کیوال و گزارش‌گیری بوده است. این قابلیت‌ها را می‌توان تحت عنوان تحلیل توصیفی یا گزارش‌گیری نامید. برعکس، دانشمندان داده مسئول تجزیه و تحلیل‌های پیش‌گویانه، تجزیه و تحلیل آماری و استفاده از ابزارها و الگوریتم‌های تحلیلی پیشرفته‌تر هستند. آنها ممکن است دانش عمیق‌تری از الگوریتم‌ها داشته باشند و تحت نام‌های مختلفی از قبیل داده‌کاوی، کشف دانش یا یادگیری ماشین شناخته شوند. برخی از این متخصصان ممکن است به دانش برنامه‌نویسی عمیق‌تری نیاز داشته باشند تا بتوانند کدهای مورد نیاز برای تحلیل/پاکسازی داده‌ها را به زبان‌های وب‌گرا مثل جاوا یا پایتون^۲ و زبان‌های آماری مثل « آر^۳ » برنامه‌نویسی کنند. بسیاری از متخصصان تحلیل همچنین نیازمند تخصص بالا در مدلسازی آماری، آزمایش و تجزیه و تحلیل هستند. مجدداً خوانندگان باید توجه داشته باشند که تمامی اینها در مجموعه‌ی تحلیل‌های تجویزی و پیشگویانه قرار می‌گیرند. با وجود این، تحلیل‌های تجویزی شامل تخصص‌های سطح بالاتر در پژوهش عملیاتی من‌جمله بهینه‌سازی، شبیه‌سازی و آنالیز تصمیم نیز می‌شوند. افرادی که در این رشته‌ها تخصص دارند را به جای متخصصان تحلیل، دانشمندان داده می‌نامند.

دیدگاه ما این است که وجه تمایز بین متخصص تحلیل و دانشمند داده بیشتر مربوط به عمق دانش فنی و مجموعه مهارت‌ها است تا کارکرد آنها. همچنین ممکن است مربوط به تمایز بین رشته‌ها باشد. علوم رایانه، آمار و برنامه‌های ریاضیات کاربردی بیشتر، برچسب علم داده‌ها را یدک می‌کشند در حالیکه برچسب تحلیل را بیشتر برای متخصصان کسب‌وکار محور حفظ می‌کنند. مثال دیگری از این موضوع، متخصصان و کارشناسان فیزیک کاربردی است که پیشنهاد کرده‌اند تا برای توصیف تحلیل‌های مربوط به گروهی از افراد از قبیل شبکه‌های اجتماعی، شبکه‌های زنجیره تأمین و غیره از عبارت علم شبکه^۴ استفاده شود. برای اطلاعات بیشتر در این خصوص مراجعه کنید به <http://barabasi.com/networksciencebook>.

گذشته از این وجه تمایز آشکار مجموعه مهارت‌های متخصصان که فقط باید تحلیل‌های توصیفی/گزارش‌گیری در مقایسه با آنهایی که هر سه نوع تحلیل را انجام می‌دهند، تفاوت بین این دو برچسب در بهترین حالت، فازی است. ما شاهد این هستیم که فارغ‌التحصیلان دوره‌های تحلیل تمایل دارند تا وظایفی که بیشتر سازگار با متخصصان علم داده‌ها می‌باشد را انجام دهند تا اینکه صرفاً تحلیل‌های گزارش‌گیری را عهده‌دار شوند. این کتاب مشخصاً قصد دارد تا به معرفی قابلیت‌ها و کارکردهای تمامی تحلیل‌ها (که شامل علم داده‌ها نیز هست) بپردازد. از این به بعد، ما این عبارات را به جای یکدیگر به کار خواهیم برد.

¹ Data Science

² Python

³ R

⁴ Network Science

کلان داده چیست؟ هر کتابی که در مورد تحلیل و علم داده‌ها صحبت می‌کند باید به موضوع آنچه **تحلیل کلان داده** نامیده می‌شود نیز بپردازد. ما در فصل ۹ به بررسی آن خواهیم پرداخت ولی در اینجا معرفی کوتاهی از آن را ارائه می‌کنیم. مغز ما بی‌اندازه سریع و کارآمد کار کرده و در پردازش حجم انبوهی از انواع مختلف داده‌ها از قبیل تصاویر، متن، صدا، بو و ویدئو به صورت چندکاره عمل می‌کند. ما تمامی انواع مختلف داده‌ها را نسبتاً به آسانی پردازش می‌کنیم. رایانه‌ها از طرف دیگر هنوز نتوانسته‌اند خودشان را با سرعت تولید داده‌ها و تجزیه و تحلیل سریع آنها وفق دهند. به همین خاطر است که ما با مشکل کلان داده‌ها مواجه هستیم. پس کلان داده‌ها چه هستند؟ به بیان ساده، کلان داده به داده‌هایی اشاره دارد که نمی‌توان آنها را در یک واحد انبارش ذخیره نمود. کلان داده‌ها معمولاً به داده‌هایی اشاره دارند که به فرم‌های مختلفی دریافت می‌شوند: ساخت یافته، ساخت نیافته، در جریان و غیره. منابع اصلی اینگونه داده‌ها، جریان کلیک در وبسایت‌ها، مطالب منتشر شده در سایت‌های رسانه‌های اجتماعی از قبیل فیسبوک و داده‌های بدست آمده از ترافیک، حسگرها یا آب و هوا است. یک موتور جستجوی وب مثل گوگل برای اینکه بتواند نتایج یک جستجو را در کسری از ثانیه ارائه کند می‌بایست میلیاردها صفحه‌ی وب را جستجو و ارزیابی نمود. اگرچه اینکار به طور آنی انجام نمی‌شود، تولید یک ارزیابی از تمامی صفحات وب در اینترنت کار آسانی نیست. خوشبختانه گوگل توانسته است این مشکل را حل کند. از بین سایر ابزارها، گوگل از تکنیک‌های تحلیلی کلان داده استفاده می‌کند.

مدیریت داده‌ها در این مقیاس دو جنبه دارد: انبارش و پردازش. اگر بتوانیم یک انبار بسیار گران‌قیمت را برای ذخیره‌سازی تمامی اینها در یک مکان و یکجا خریداری کنیم، مقاوم کردن این واحد انبارش داده در برابر خطا بسیار پرهزینه خواهد بود. یک راه حل مبتکرانه این بوده است که این داده‌ها را در تکه‌های بزرگ بر روی دستگاه‌های مختلف که از طریق شبکه بهم متصل شده‌اند ذخیره کنیم یعنی یک یا دو نسخه از این تکه‌ها را در محل‌های متفاوت بر روی شبکه قرار دهیم. این ابتکار ابتدا در گوگل (آن زمان سیستم فایل گوگل^۱ نامیده می‌شد) بکار گرفته شده و سپس در قالب پروژه‌های آپاچی^۲ تحت عنوان سیستم فایل توزیع شده‌ی هُدوپ^۳ توسعه و انتشار یافت.

با وجود این، انبارش این داده‌ها فقط نیمی از مشکل است. داده‌ها در صورتیکه ارزشی برای کسب‌وکار فراهم نکنند بی‌ارزش هستند و برای اینکه ارزشمند باشند باید تجزیه و تحلیل شوند. چگونه چنین حجم انبوهی از داده را می‌توان تجزیه و تحلیل کرد؟ سپردن تمامی محاسبات به یک رایانه‌ی قدرتمند نمی‌تواند مشکل را حل کند و هزینه بالاسری هنگفتی را تحمیل خواهد کرد. یک راه حل مبتکرانه دیگر پیشنهاد شد: تحمیل محاسبات به داده به جای اینکه داده‌ها را به سمت یک ایستگاه محاسباتی روانه کنیم. این چیزی است که امروزه تحت عنوان الگوی برنامه‌نویسی نگاشت‌کاهش^۴ می‌شناسیم که پردازش کلان داده‌ها را به یک واقعیت تبدیل کرد. نگاشت‌کاهش ابتدا در گوگل توسعه یافته و نسخه‌ی بعدی آن توسط پروژه‌ی آپاچی به نام **نگاشت‌کاهش** هُدوپ انتشار یافت.

امروزه، وقتی که در مورد انبارش، پردازش یا تجزیه و تحلیل کلان داده‌ها صحبت می‌کنیم، سیستم فایل توزیع شده هُدوپ و نگاشت‌کاهش تا حدی وجود دارند. استانداردها و نرم‌افزارهای مرتبط دیگری نیز پیشنهاد شده‌اند. اگرچه جعبه ابزار اصلی در صورت

¹ Google File System

² Apache

³ Hadoop Distributed File System (HDFS)

⁴ MapReduce programming

متن باز در دسترس است ولی شرکت‌های متعددی به منظور ارائه‌ی خدمات آموزشی یا خدمات سخت‌افزاری و نرم‌افزاری تحلیلی تخصصی در این فضا راه افتاده‌اند. برخی از آنها عبارتند از هورتون‌وورکز^۱، کلادرا^۲ و ترادیتا آستر^۳. طی چندین سال گذشته، به ظهور برنامه‌های کاربردی کلان‌داده، آنچه کلان‌داده نامیده می‌شد بیش از پیش تغییر کرد. نیاز به پردازش داده‌هایی که با سرعت بالا وارد می‌شدند، مقوله‌ی سرعت را نیز به معامله افزود. یک مثال از پردازش سریع داده‌ها، معاملات الگوریتمی^۴ است که از پلتفرم‌های الکترونیکی مبتنی بر الگوریتم‌ها برای مبادله‌ی سهام در بازار مالی و بر حسب میکروثانیه عمل می‌کند. نیاز به پردازش انواع مختلف داده‌ها موجب شده تا تنوع نیز به معادله افزوده شود. یک مثال دیگر از تنوع گسترده‌ی داده‌ها، تحلیل حساسیت است که از شکل‌های مختلف داده که از پلتفرم‌های رسانه‌های اجتماعی و پاسخ‌های مشتری بدست می‌آیند برای سنجش حساسیت‌ها استفاده می‌شود. امروزه، کلان‌داده‌ها تقریباً با هر نوع داده‌ی بزرگ که ویژگی‌های حجم، سرعت و تنوع را داشته باشد مرتبط هستند. همانطور که قبلاً گفته شد، اینها به سرعت در حال رشد و تکامل هستند تا تحلیل‌های جریان، اینترنت اشیا، رایانش ابری و یادگیری عمیق را شامل شوند. ما در فصل‌های مختلف کتاب به مطالعه و بررسی این موضوعات خواهیم پرداخت.

سوالاتی برای مرور بخش ۱.۵

۱. تحلیل را تعریف کنید.
۲. تحلیل توصیفی چیست؟ ابزارهای مختلف که در تحلیل توصیفی بکار می‌روند چه هستند؟
۳. تفاوت تحلیل توصیفی با گزارش‌گیری سنتی چیست؟
۴. یک انبار داده چیست؟ چگونه فناوری انبار داده می‌تواند به تحلیل کمک کند؟
۵. تحلیل پیشگویانه چیست؟ سازمان‌ها چگونه می‌توانند از تحلیل پیشگویانه استفاده کنند؟
۶. تحلیل تجویزی چیست؟ چه نوع مسائلی را می‌توان با تحلیل تجویزی حل کرد؟
۷. مدلسازی را از منظر تحلیل تعریف کنید.
۸. آیا فکر خوبی است که پیش از بکارگیری تحلیل تجویزی از سلسله مراتب تحلیل توصیفی و پیشگویانه پیروی کنیم؟
۹. تحلیل‌ها چگونه می‌توانند در تصمیم‌گیری عینی کمک کنند؟
۱۰. تحلیل کلان‌داده چیست؟
۱۱. منابع کلان‌داده‌ها چه هستند؟
۱۲. ویژگی‌های کلان‌داده‌ها چه هستند؟
۱۳. برای پردازش کلان‌داده‌ها از چه تکنیک پردازشی استفاده می‌شود؟

¹ HortonWorks

² Coudera

³ Teradata Aster

⁴ Algorithmic Trading

۱.۶ مثال‌هایی از تحلیل‌ها در حوزه‌های منتخب

در سرتاسر فصل‌های مختلف این کتاب مثال‌هایی از کاربردهای تحلیل‌ها را خواهید دید. این یکی از رویکردهای اصلی این کتاب است. در این بخش، تمرکز ما بر سه حوزه کاربردی ورزش، مراقبت‌های بهداشتی و خرده‌فروشی که بیشترین کاربردها و موفقیت‌ها در آنجا گزارش شده خواهد بود.

کاربرد تحلیل‌ها در مسائل کسب‌وکار یک مهارت اساسی است که شما در این کتاب خواهید آموخت. بسیاری از این تکنیک‌ها امروزه به منظور بهبود تصمیم‌گیری در تمامی جنبه‌های ورزشی بکار می‌روند که حوزه‌ی بسیار داغی به نام *تحلیل ورزشی*^۱ نام دارد. این هنر و علم گردآوری داده‌های مربوط به ورزشکاران و تیم‌ها به منظور خلق بینش‌هایی است که منجر به بهبود تصمیمات ورزشی از قبیل تصمیم به استخدام کدام بازیکن‌ها، میزان پرداختی به آنها، نحوه آموزش آنها، نحوه حفظ سلامت آنها و زمان مبادله یا بازنشستگی آنها می‌گردد. در خصوص تیم‌ها، شامل تصمیماتی از قبیل قیمت‌گذاری بلیت‌ها و همچنین تصمیمات لیست زمانبندی، تجزیه و تحلیل نقاط قوت و ضعف هر یک از رقبا و بسیاری از تصمیمات مربوط به روز بازی می‌شود.

علاوه بر این، تحلیل‌های ورزشی در بین تحلیل‌ها تبدیل به یک تخصص شده‌اند. ورزش به دلیل حجم درآمد سالانه‌ی حدوداً ۱۴۵ میلیارد دلاری به علاوه‌ی ۱۰۰ میلیارد دلار از طریق شرط‌بندی‌های قانونی و ۳۰۰ میلیارد دلار از طریق شرط‌بندی‌های غیرقانونی یک حوزه بسیار مهم محسوب می‌شود (پرایس و اترهاوس) ("تغییر بازی: چشم‌انداز بازار جهانی ورزش در سال ۲۰۱۵). در سال ۲۰۱۴ تنها ۱۲۵ میلیون دلار برای تحلیل سرمایه‌گذاری شد (کمتر از ۰.۱ درصد از درآمد). انتظار می‌رود که این رقم تا سال ۲۰۲۱ به ۴.۷ میلیارد دلار افزایش یابد ("Sports Analytics Market Worth 4.7B\$ by 2021").

استفاده از تحلیل‌ها برای ورزش با کتاب *مانی بال*^۲ اثر مایکل لوئیس^۳ در سال ۲۰۰۳ و فیلمی با شرکت برد پیت در سال ۲۰۱۱ محبوبیت یافت. این کتاب به زنگی بیلی بین^۴، مدیرکل آکلند ای و استفاده او از داده‌ها و تحلیل برای تبدیل یک تیم بازنده به یک برنده می‌پردازد. به طور خاص، وی یک تحلیلگر را استخدام می‌کند که از تحلیل‌ها برای انتخاب بازیکنانی که می‌توانستند به بیس برسند در مقابل بازیکنانی که در سنجه‌های سنتی مثل تعداد دفعات به بیس دویده کارکشته هستند استفاده می‌کردند. این بینش‌ها به تیم اجازه می‌دهد تا نامزدهایی که توسط سایر تیم‌ها نادیده گرفته شده‌اند را با دستمزدهای پایه‌ی قابل قبول استخدام کنند. این روش جواب داد و تیم توانست تا در سال‌های ۲۰۰۲ و ۲۰۰۳ به پلی‌آف راه پیدا کند.

امروزه تحلیل‌ها در تمامی بخش‌های ورزش استفاده می‌شوند. تحلیل‌ها را می‌توان بین جلوی صحنه سازمانی و پشت صحنه سازمان تقسیم نمود. در مقاله‌ی تام دَوَنپورت^۵ شرح خوبی به همراه ۳۰ مثال آورده شده است. تحلیل‌های کسب‌وکار جلوی صحنه سازمانی شامل تحلیل رفتار طرفداران از مدل‌های پیشگویانه برای تمدید بلیت‌های فصل و فروش بلیت‌های عادی گرفته تا ارسال توئیت‌ها توسط طرفداران در مورد تیم، ورزشکایان، مربیان و صاحبان می‌شود. این بسیار شبیه مدیریت ارتباط با مشتریان سنتی است. تجزیه و تحلیل مالی نیز یک حوزه مهم است که محدودیت‌های بورس تحصیلی یا سقف حقوق و دستمزد بخشی از معادله هستند.

¹ Sports analytics

² Moneyball

³ Michael Lewis

⁴ Billy Beane

⁵ Tom Davenport

کاربردهای پشت صحنه‌ی سازمان شامل تجزیه و تحلیل بازی ورزشکاران و کل تیم است. در خصوص بازیکنان فردی، تمرکز بر تحلیل کشف و مدل‌های بکارگیری آنها، تحلیل نقاط قوت و ضعف و سطح آمادگی و همچنین توسعه و PMها برای اجتناب از تمرین بیش از حد و آسیب‌دیدگی‌ها است. پژوهش در زمینه ضربه‌های مغزی یکی از حوزه‌های بسیار داغ است. تحلیل تیم شامل استراتژی‌ها و تاکتیک‌ها، ارزیابی‌های رقابتی و انتخاب‌های بهینه لیست وظایف تحت موقعیت‌های مختلف در زمین و حین مسابقه است.

مثال‌های بعدی نشان می‌دهند که چگونه دو سازمان ورزشی از داده‌ها و تحلیل به منظور بهبود عملیات ورزش به همان نحوی که در تصمیم‌گیری صنعتی نتیجه داده‌اند استفاده می‌کنند.

تیم‌های فوتبال

دیو^۱ وارد^۱ به عنوان تحلیل‌گر کسب‌وکار برای یک تیم حرفه‌ای بیسیال و با تخصص موضوع درآمدی مشغول به کار است. او میزان فروش بلیت‌ها را هم از جانب بلیت‌داران فصلی و هم خریداران تک بلیت تجزیه و تحلیل می‌کند. برخی از سوالات نمونه در حیطه مسئولیت‌های وی این هستند که چرا بلیت‌داران فصلی، بلیت‌هایشان را تمدید می‌کنند یا نمی‌کنند و همچنین چه عواملی منجر به خریدهای دقیقه آخری بلیت می‌شود. سوال دیگر این است که چگونه بلیت‌ها را قیمت‌گذاری می‌کنند.

برخی از تکنیک‌های تحلیلی دیو شامل استفاده از آمارهای ساده در خصوص رفتار طرفداران است مثل حضور یا پاسخگویی به سوالات نظرسنجی در مورد احتمال خرید مجدد بلیت. با وجود این، بین آنچه طرفداران می‌گویند و آنچه انجام می‌دهند تفاوت است. دیو یک نظرسنجی از طرفداران تیم بر اساس جایگاه تماشاچیان (لایه) انجام داده و از آنها در مورد احتمال تمدید بلیت‌های فصل سوال می‌کند. اما زمانیکه پاسخ‌ها را با آنچه طرفداران انجام داده‌اند مقایسه می‌کند، اختلاف‌های فاحشی را کشف می‌کند (شکل ۱.۱۰). وی متوجه شد که ۶۹ درصد از طرفداران در سندی‌های لایه ۱ که در نظرسنجی اعلام کرده‌اند که «احتمالاً تمدید نخواهند کرد» واقعاً همین کار را کرده‌اند. این بینش مفید منجر به اقدام شد - مشتریانی که در سلول‌های سبز قرار دارند با احتمال بیشتری بلیت‌هایشان را تمدید خواهند کرد و در مقایسه با مشتریان در سلول‌های آبی تماس‌های بازاریابی کمتری خواهند داشت و دلارهای کمتری را تبدیل خواهند کرد.

لایه ۱	بسیار محتمل	محتمل	شاید	نامحتمل	مطمئناً خیر
1	92	88	75	69	45
2	88	81	70	65	38
3	80	76	68	55	36
4	77	72	65	45	25
5	75	70	60	35	25

شکل ۱.۱۰ تمدید بلیت‌های فصلی - نتایج نظرسنجی

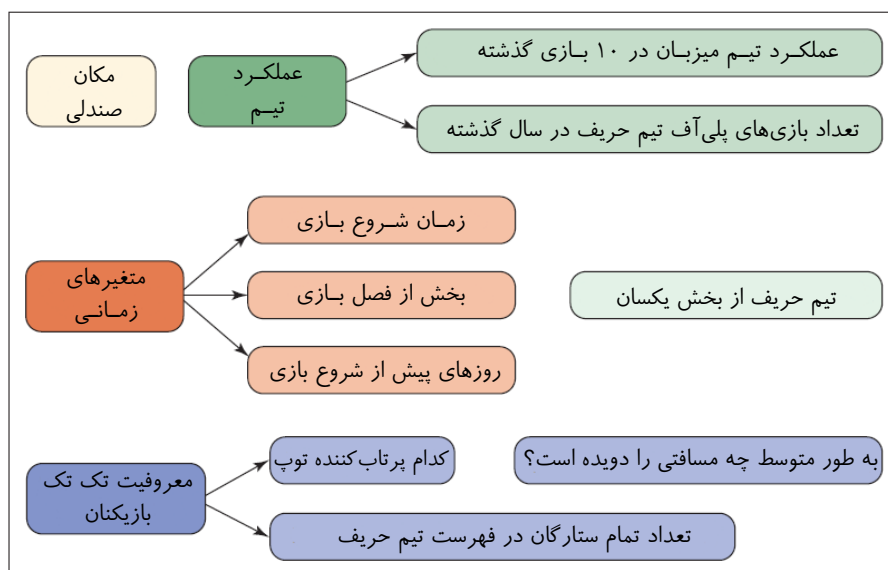
با وجود این، عوامل بسیاری بر رفتار طرفداران برای خرید بلیت تأثیر می‌گذارند به ویژه قیمت که نیازمند تحلیل‌های داده و آمارهای پیچیده‌تر می‌باشد. در هر دو حوزه ولی به طور خاص در مورد بلیت‌های تک بازی، دیو سعی کرده تا از قیمت‌گذاری پویا

¹ Dave Ward

استفاده کند و کسب‌وکار را از رویکرد قیمت‌گذاری ایستا و ساده که بر اساس جایگاه تماشاچی تعیین می‌شود به سمت قیمت‌گذاری روز به روز و بالا به پایین برای تک تک صندلی‌ها سوق دهد. این یک حوزه‌ی پژوهشی غنی برای بسیاری از تیم‌های ورزشی بوده و پتانسیل بسیار بالایی برای افزایش درآمد دارد. به عنوان مثال، وی در روش قیمت‌گذاری خود، سوابق تیم‌ها، طرف مقابل که با آنها بازی می‌کنند، زمان و تاریخ بازی، ستارگانی که برای هر تیم بازی می‌کنند، تاریخچه‌ی تمدید بلیت‌های فصلی طرفداران یا خرید بلیت‌های تکی و عواملی از قبیل جایگاه تماشاچی، تعداد جایگاه‌ها و اطلاعات آنی مثل سابقه‌ی ازدحام ترافیک در زمان برگزاری مسابقه و حتی آب و هوا را نیز در نظر می‌گیرد (شکل ۱.۱۱).

کدامیک از این عوامل مهم هستند تا چه اندازه؟ دیو با توجه به پیش‌زمینه‌ی آماری قوی خود، مدل‌های رگرسیون را برای تعیین عوامل اصلی پیش برنده‌ی رفتارهای قبلی ساخته و PMهایی را برای شناسایی نحوه‌ی هزینه‌کرد منابع بازاریابی و درآمدزایی ایجاد کرده است. او مدل‌های رویگردانی را برای بلیت‌داران فصلی ایجاد کرده تا بخش‌بندی مشتریان را بر اساس آنهایی که تمدید خواهند کرد، تمدید نخواهند کرد یا بی‌تفاوت‌ها انجام دهد و آنگاه بر این اساس طرح‌های تبلیغات و بازاریابی مناسب‌تری را طراحی کند.

علاوه بر این، دیو یک امتیازدهی حساسیت را بر روی نظرات طرفداران مثل توئیت‌ها انجام داده که به او در بخش‌بندی طرفداران بر حسب میزان وفاداری کمک کرده است. مطالعات دیگر در مورد حضور تماشاچیان تک بازی به دپارتمان بازاریابی کمک کرده است تا تأثیر جوایزی از قبیل عروسک‌ها یا تی‌شرت‌ها یا توصیه‌هایی برای خرید آگهی‌های تلویزیونی را بهتر درک کنند. به غیر از درآمد، تیم دیو بر روی حوزه‌های تحلیلی دیگری از جمله بازاریابی، درآمدهای حاصل از پخش رادیویی و تلویزیونی، مذاکرات مدیرکل در خصوص دستمزدها، تحلیل‌های استخدام بویژه با توجه به دستمزدها، اثربخشی تبلیغات از جمله کانال‌های تبلیغات و اطلاع از برند و همچنین تحلیل شرکا نیز کار کرده است.



شکل ۱.۱۱ کار قبلی قیمت‌گذاری بویا - لیگ اصلی بیسبال

باب بریدلاو^۱، مربی فوتبال یکی از تیم‌های مهم دانشگاهی است. همه چیز برای او در برنده شدن در بازی‌ها خلاصه شده است. حوزه تمرکز او شامل بکارگیری بهترین بازیکنان دبیرستان و تربیت و پرورش آنها برای وفق پیدا کردن با سیستم‌های تدافعی و تهاجمی و بهره‌گیری از حداکثر تلاش آنها در روزهای مسابقه است. سوالات نمونه در حوزه مسئولیت او شامل موارد زیر است: چه کسی را استخدام کند؟ از چه مانورهایی برای توسعه‌ی مهارت‌هایشان استفاده کند؟ ورزشکاران‌مان را تا چه اندازه تحت فشار قرار دهیم؟ نقاط قوت و ضعف حریف‌ها کجاست و چگونه می‌توانیم گرایش‌های بازی آنها را متوجه شویم؟

خوشبختانه، تیم او یک خبره‌ی جدید در حوزه عملیات تیمی را به نام «دار برانک»^۲ استخدام کرد که تخصص وی کمک به مربیان برای اتخاذ تصمیمات تاکتیکی بود. او با تیمی از کارآموزان دانشجو کار می‌کند که در حال ایجاد تحلیل‌های تیم حریف هستند. آنها از فیلم بازی مربی برای ایجاد یک مدل درخت تصمیم‌آشاری (شکل ۱.۱۲) برای پیش‌بینی اینکه آیا بازی بعدی یک بازی رانینگ^۳ خواهد بود یا پسینگ^۴. برای هماهنگ‌کننده دفاع، آنها نقشه‌هایی از هر حمله‌ی پسینگ هر حریف تهیه کردند (شکل ۱.۱۳) که گرایش آنها به پرتاب به راست یا چپ و مناطق پوشش دفاعی را به تصویر می‌کشید. نهایتاً، آنها تعدادی تحلیل سری زمانی (شکل ۱.۱۴) را برای بازی‌های انفجاری ساختند (دستیابی به بیش از ۱۶ یارد برای یک بازی پسینگ یا ۱۲ یارد برای یک بازی رانینگ). برای هر بازی، آنها نتیجه را با ترکیب دفاعی خودشان و ترکیب‌های تهاجمی تیم‌های دیگر مقایسه کردند که به بریدلاو کمک کرد تا در حین یک بازی با سرعت بیشتری به تغییر ترکیب تیم واکنش نشان دهد. ما تکنیک‌های تحلیلی که این شکل‌ها تولید کرده‌اند را به طور عمیق‌تر در فصل‌های ۳-۶ و فصل ۹ توضیح خواهیم داد.

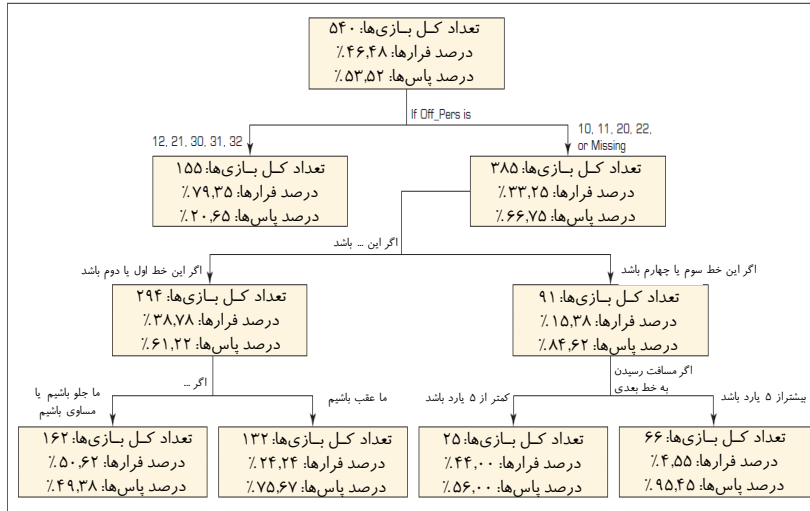
کار جدیدی که «دار» به آن پرداخته است شامل ایجاد مدل‌های بهتری برای استخدام و بکارگیری ورزشکاران دبیرستانی است. به عنوان مثال، این تیم هر ساله سه دانشجو را بورسیه می‌کند. برای «دار»، انتخاب بهترین بازیکنان از سنج‌های ساده‌ای مثل سرعت ورزشکاران، ارتفاع پرش آنها یا طول بازوان آنها به سمت معیارهای جدیدتری مثل اینکه چقدر سریع می‌توانند سرهایشان را برای گرفتن یک عبور بچرخانند، در برابر چندین محرک چه نوع زمان‌های واکنشی را نشان می‌دهد و میزان دقت آنها در دویدن در مسیرها چه میزان است رفته است. برخی از افکار و نظرات او که این مفاهیم را به تصویر می‌کشند در سایت TUN Web موجود هستند. نگاه کنید به «The Case of Precision Football» (2015) Business Scenarion Investigation

¹ Bob Breedlove

² Dar Beranek

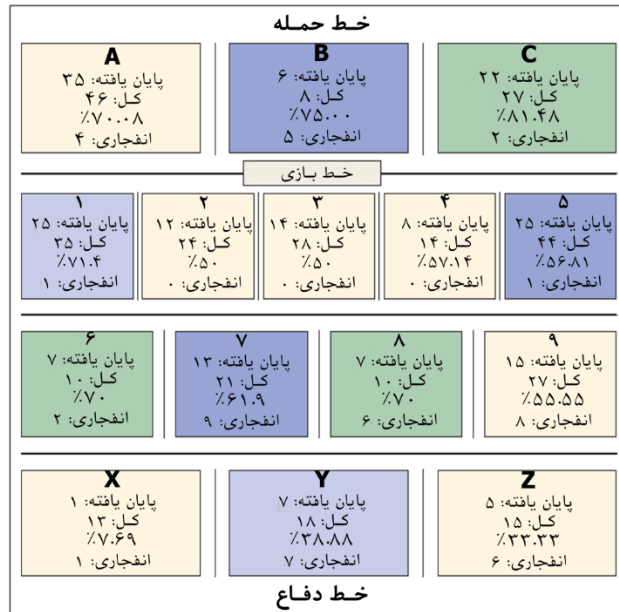
³ Running play

⁴ Passing play



شکل ۱.۱۲ درخت تصمیم آبخاری برای بازی‌های رانینگ یا پسینگ

منبع: اعطا شده توسط دکتر دیو شریدر^۱ که پس از ۲۴ سال فعالیت در بازاریابی و توسعه‌ی پیشرفته در شرکت ترادیتا بازنشسته شد. وی همچنان عضو هیئت مشاورین شبکه دانشگاهی ترادیتا می‌باشد که دوران بازنشستگی خود را برای کمک به دانشجویان و دانشکده برای یادگیری بیشتر در مورد تحلیل ورزشی می‌گذراند. شکل توسط پیتر لینانگ و جاکوب پیرسیون از فارغ التحصیلان دانشگاه ایالتی آکلاهوما و به عنوان بخشی از پروژه‌ی دانشجویی در بهار سال ۲۰۱۶ در کلاس درس پروفسور رامش شاردن^۲ و تحت مربی‌گری دکتر دیو شریدر تهیه شده‌اند.



شکل ۱.۱۳ نقشه‌ی تحلیل منطقه برای بازی‌های پسینگ

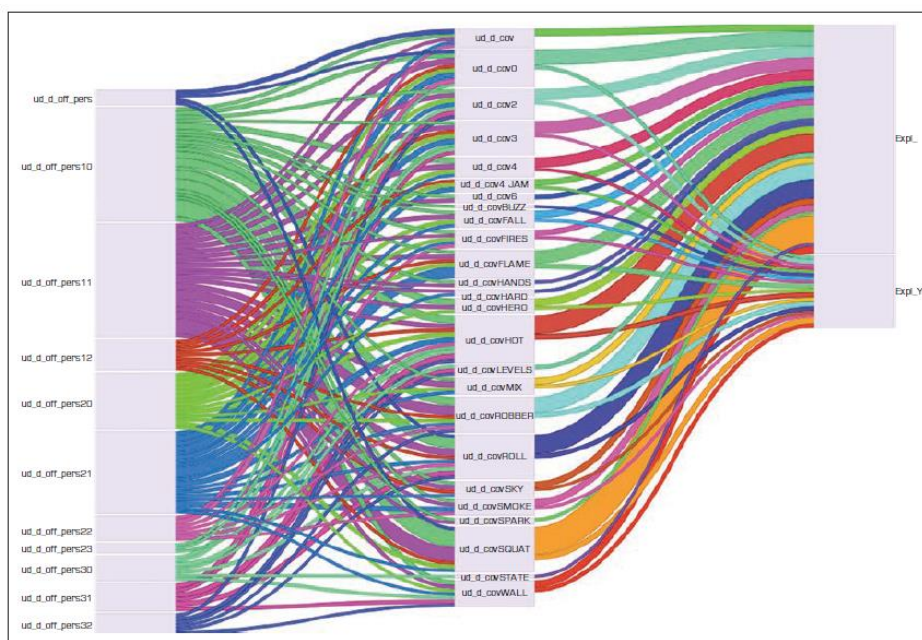
منبع: اعطا شده توسط دکتر دیو شریدر^۳ که پس از ۲۴ سال فعالیت در بازاریابی و توسعه‌ی پیشرفته در شرکت ترادیتا بازنشسته شد. وی همچنان عضو هیئت مشاورین شبکه دانشگاهی ترادیتا می‌باشد که دوران بازنشستگی خود را برای کمک به دانشجویان و دانشکده برای یادگیری بیشتر در مورد تحلیل

¹ Dave Shrader
² Ramesh Sharda
³ Dave Shrader

ورزشی می‌گذراند. شکل توسط پیتر لیانگ و جاکوب پیرسیون از فارغ التحصیلان دانشگاه ایالتی آکلاهما و به عنوان بخشی از پروژه‌ی دانشجویی در بهار سال ۲۰۱۶ در کلاس درس پروفسور رامش شاردا^۱ و تحت مربی‌گری دکتر دیو شریدر تهیه شده‌اند.

از این مثال‌ها چه درس‌هایی می‌توانیم بیاموزیم؟ فراتر از تحلیلگران کسب‌وکار جلوی صحنه، مربیان، آموزش‌دهندگان و خبرگان عملکردی، افراد بسیار زیاد دیگری در حوزه ورزشی از جمله متصدیان نگهداری زمین گلف که شرایط خاک و چمن در تورنمنت‌های پی‌جی‌ای را ارزیابی می‌کنند تا داوران بیسبال و بسکتبال که بر اساس تصمیمات درست و نادرست رتبه‌بندی می‌شوند وجود دارند که از داده‌ها استفاده می‌کنند. در واقع، به سختی می‌توان حوزه‌ای از ورزش را پیدا کرد که تحت تأثیر دسترسی‌پذیری داده‌های بیشتر به‌ویژه داده‌های بدست آمده از حسگرها نباشد.

مهارت‌هایی که شما در این کتاب برای تحلیل‌های کسب‌وکار خواهید آموخت در حوزه ورزشی نیز صدق می‌کنند. اگر می‌خواهید به دانش عمیق‌تری در این حوزه دست پیدا کنید پیشنهاد می‌کنیم تا به بخش تحلیل‌های ورزشی در TUN که یک منبع رایگان برای دانشجویان و دانشکده است رجوع کنید. در این وب‌سایت، شما توضیحاتی در خصوص اینکه چه مطالبی را در مورد تحلیل‌های ورزشی مطالعه کنید، محل‌هایی که مجموعه داده‌های برای تجزیه و تحلیل در دسترس عموم قرار دارند و همچنین مثال‌هایی از پروژه‌های دانشجویی در تحلیل ورزش‌ها و مصاحبه‌هایی با کارشناسان ورزشی که از داده‌ها و تحلیل‌ها برای انجام کارشان استفاده می‌کنند را خواهید یافت. موفق باشید!



شکل ۱.۱۴ نقشه‌ی تحلیل منطقه برای بازی‌های پَسینگ

¹ Ramesh Sharda

اگرچه تحلیل‌های مراقبت بهداشت طیف بسیار گسترده‌ای از کاربردهای را از پیشگیری تا تشخیص و عملیات مؤثر و پیشگیری از کلاهبرداری شامل می‌شوند، ما بر روی برخی از کاربردها که در یک شرکت بزرگ بیمه سلامت و بهداشت در ایالات متحده به نام هومانا توسعه یافته است تمرکز می‌کنیم. بر طبق وبسایت این شرکت، «استراتژی این شرکت، یکپارچه‌سازی ارائه‌ی خدمات مراقبت، تجربه‌ی اعضای شرکت و بینش و درک پرسنل بالینی و مراجعه‌کنندگان به منظور ترغیب مشارکت، تغییر رفتار، اطلاع‌رسانی و سلامتی بالینی فعال... است.»

دستیابی به این اهداف استراتژیک شامل سرمایه‌گذاری‌های هنگفت در فناوری اطلاعات به طور کلی و تحلیل‌ها به طور خاص بود. برایان لُکلر^۱ معاون ارشد و اطلاعات مدیر ارشد شرکت هومانا است. او دارای مدرک دکترا در رشته‌ی سیستم‌های اطلاعات مدیریت از دانشگاه ایالتی آکلاهما است. او از تحلیل به عنوان یک عامل تمایز در رقابت در هومانا استفاده کرده است که شامل اسپانسرینگ مشترک در ایجاد یک مرکزی تعالی تحلیل است. وی پروژه‌های زیر را به عنوان نمونه‌هایی از راهکارهای تحلیل شرکت هومانا برشمرده است که با رهبری ویپین گوپال^۲، مدیر ارشد تحلیل‌های بالینی هومانا انجام شده‌اند.

زمین‌خوردن‌های تصادفی یک مخاطره‌ی بسیار مهم برای سلامت افراد ۶۵ سال و بالاتر است که یک سوم از این جمعیت به طور سالانه دچار چنین حوادثی می‌شوند. زمین‌خوردن، هزینه‌های قابل توجهی را به سیستم مراقبت بهداشت و سلامت ایالات متحده تحمیل می‌کند و هزینه‌های مستقیم این حوادث فقط در سال ۲۰۱۳ معادل ۳۴ میلیارد دلار برآورد شده است. با افزایش درصد جمعیت سالمند در ایالات متحده، پیش‌بینی شده که زمین‌خوردن و هزینه‌های مربوط به آنها نیز افزایش یابد. طبق آمار مراکز کنترل و پیشگیری بیماری^۳، «زمین‌خوردن یک مشکل سلامت عمومی هستند که به میزان زیادی قابل پیشگیری می‌باشند» (www.cdc.gov/homeandcreationalafety/falls/adultfalls.html). زمین‌خوردن همچنین یک عامل اساسی در آسیب‌های مرگ‌بار و غیرمرگ‌بار در بین بزرگسالان هستند که زمین‌خوردن‌های آسیب‌زننده، ریسک ناتوانی را تا ۵۰٪ افزایش می‌دهند (جیل و همکاران ۲۰۱۳). هومانا دومین مرکز بزرگ ارائه‌دهنده‌ی خدمات مدیکر آدوونتیج^۴ با تقریباً ۳.۲ میلیون عضو است که اکثر آنها را افراد مسن تشکیل می‌دهند. هدف اصلی کسب‌وکار این مرکز، کمک به اعضای سالمندش برای یک زندگی خوب و ایمن است که پیشگیری از زمین‌خوردن یکی از مؤلفه‌های مهم آن محسوب می‌شود. با وجود این، هیچ متدولوژی دقیق و محکمی وجود ندارد که بتواند افرادی که بیشترین احتمال زمین‌خوردن را دارند و تلاش‌های پیشگیری از زمین‌خوردن برایشان بسیار مؤثر خواهد بود را شناسایی کنند. برخلاف شرایط پزشکی مزمن از قبیل دیابت و سرطان، زمین‌خوردن یک وضعیت پزشکی واضح و روشن نیست. علاوه بر این، زمین‌خوردن معمولاً در داده‌ها گزارش نشده‌اند زیرا پزشکان معمولاً تمایل دارند تا پیامدها اینگونه زمین‌خوردن را تحت عنوان شکستگی یا دررفتگی طبقه‌بندی کنند. اگرچه ارزیابی‌های بالینی بسیار زیادی برای شناسایی افرادی که زمین‌خوردن‌اند انجام می‌شود ولی جستجوی آنها محدود بوده و فاقد توان پیشگویانه کافی هستند (گیتز و همکاران^۵ ۲۰۰۸). بدین ترتیب، برای شناسایی افرادی که با بیشترین ریسک زمین‌خوردن مواجه هستند به روشی دقیق

¹ Brian LeClaire

² Vipin Gopal

³ Centers for Disease Control and Prevention (CDC)

⁴ Medicare Advantage

⁵ Gates et al.

و آینده‌نگر نیاز داریم تا بتوان به نحوی کنشگرانه از زمین خوردن آنها پیشگیری نمود. تیم تحلیل هومانا در این مقوله، توسعه‌ی یک مدل پیشگویانه زمین خوردن^۱ را در دستور کار خود قرار داد. این اولین PM جامع گزارش شده است که از ادعاهای پزشکی و دارویی اجرایی، داده‌های بالینی، الگوهای بالینی جسمانی، اطلاعات مراجعین و سایر داده‌ها برای شناسایی افرادی که ریسک زمین خوردن بالایی در یک افق زمانی دارند بهره‌برداری کرده است.

امروزه، PM زمین خوردن یکی از توانایی‌های اصلی هومانا برای شناسایی افراد سن بالایی است که اقدامات پیشگیری از زمین خوردن می‌تواند برایشان سودمند باشد. اثبات مفهوم اولیه برای مراجعین هومانا که شامل ۲ درصد اول از جامعه‌ای است که بالاترین ریسک زمین خوردن را دارند نشان می‌دهد که تعداد مراجعینی که از خدمات فیزیوتراپی استفاده می‌کنند افزایش یافته است یعنی آنها برای کاهش ریسک زمین خوردن گام‌هایی را برداشته‌اند. ابتکار عمل دوم از PM زمین خوردن برای شناسایی افراد پرریسک برای برنامه‌های پیش از راه دور استفاده می‌کند. هومانا با استفاده از PM توانست ۲۰,۰۰۰ نفر با ریسک زمین خوردن بالا را شناسایی کرده و تحت پوشش این برنامه قرار دهد. افراد شناسایی شده یک دستگاه پوشیدنی را به همراه داشتند که قادر به تشخیص زمین خوردن بوده و به صورت ۲۴ ساعته و ۷ روز هفته هشدارها را برای کمک‌های فوری ارسال می‌کند. این کار موفق به کسب جایزه‌ی رهبری تحلیل از دانشکده‌ی کسب‌وکار کلی دانشگاه ایندیانا در سال ۲۰۱۵ به خاطر بکارگیری نوآورانه تحلیل در یک محیط کسب‌وکار گردید.

در سال ۲۰۱۴، مؤسسه‌ی هومانا به معرفی « بولد گُل^۲ » سازمان به منظور بهبود سلامت جوامع هدف خود به میزان ۲۰ درصد تا سال ۲۰۲۰ پرداخت. هدف این طرح این بود که به مردم در دستیابی آسان به بهترین وضعیت سلامتی‌شان کمک کند. جوامع هدف هومانا را به شکل‌های مختلفی می‌توان تعریف کرد از جمله بر حسب منطقه جغرافیایی (ایالت، شهر، همسایگی)، محصول (مدیکر آدونتج، طرح‌های کارفرما، خریداری شده به صورت شخصی) یا بر حسب شرح حال بالینی (شرایط اولویت‌دار از جمله دیابتی، فشار خون، نارسایی قلبی، بیماری مزمن انسدادی ریه یا افسردگی). درک سلامت این جوامع و نحوه‌ی ردگیری آنها در طول زمان نه تنها برای ارزیابی هدف، بلکه برای تدوین استراتژی‌های بهبود سلامت کل جامعه حیاتی است.

یک چالش پیش روی سازمان تحلیل، شناسایی سنجه‌ای بود که بتواند کل ماهیت و ذات « بولد گُل » را نشان دهد. سنجه‌های سنتی بیمه سلامت که به صورت عینی اندازه‌گیری می‌شدند مثل پذیرش بیمارستان یا تعداد ملاقاتی اورژانس به ازای هر ۱۰۰۰ نفر نمی‌تواند روح حقیقی این مأموریت را به ما نشان دهد. هدف، شناسایی سنجه‌ای است که بتواند سلامت و بهبود آن را در یک جامعه اندازه‌گیری کند و مرتبط با کسب‌وکار هومانا باشد. هومانا از طریق ارزیابی‌های تحلیلی دقیق نهایتاً پرسشنامه‌ی چهار سوالی کیفیت زندگی « روزهای سالم^۳ » را که در اصل توسط CDC برای ردگیری و سنجش پیشرفت کلی هومانا در دستیابی به « بولد گُل » توسعه یافته بود برگزید.

باید اطمینان حاصل می‌شد که سنجه‌ی انتخابی همبستگی بسیار بالایی با سنجه‌های سلامت و کسب‌وکار دارد به گونه‌ای که هرگونه بهبود در « روزهای سالم » منجر به بهبود سلامت و نتایج کسب‌وکار بهتری می‌شود. مثال‌هایی از نحوه‌ی همبستگی « روزهای سالم » با سنجه‌های مورد نظر به شرح زیر می‌باشند:

¹ Falls Predictive Model

² Bold Goal

³ Healthy Days

- افرادی که تعداد روزهای ناسالم^۱ بیشتری دارند، الگوهای هزینه و بهره‌وری بالاتری را نیز نشان می‌دهند. با یک افزایش پنج روزه در تعداد روزهای ناسالم، (۱) متوسط هزینه‌های ماهانه درمانی و دارویی به میزان ۸۲ دلار افزایش می‌یابد، (۲) پذیرش بستری به ازای هر ۱۰۰۰ بیمار به میزان ۵۲ نفر افزایش می‌یابد و (۳) متوسط زمان اقامت ۰.۲۸ روز افزایش می‌یابد (هیونز، پنیلا، اسلاباق، کوردیه، ردنا و گپال، ۲۰۱۵).
- افرادی که رفتارهای سالم‌تری داشته و شرایط مزمن آنها به خوبی مدیریت شده بود تعداد روزهای ناسالم کمتری داشتند. به عنوان مثال، زمانیکه نگاهی به افراد مبتلا به دیابت می‌اندازیم، اگر معاینه‌ی چشم افراد دیابتی (۲.۳- UHDs) یا غربال LDL (-4.3 UHDs) را گذرانده باشند، تعداد روزهای ناسالم آنها کمتر است. به همین منوال است اگر سطح قند خون تحت کنترل آنها را بر حسب HbA1C (-1.8 UHDs) یا LDL اندازه‌گیری کنیم (هیونز، پنیلا، اسلاباق، کوردیه، ردنا و گپال، ۲۰۱۵).
- افرادی که شرایط مزمن دارند تعداد روزهای ناسالم آنها بیشتر از آنهايي است که مبتلا به بیماری‌های زیر نیستند (۱) نارسایی احتقانی قلبی (16.9 UHDs)، (۲) بیماری سرخرگ کرونری (14.4 UHDs)، (۳) فشار خون (13.3 UHDs)، (۴) دیابت (14.7 UHDs)، (۵) بیماری مزمن انسدادی ریه (17.4 UHDs) یا (۶) افسردگی (22.4 UHDs). (هیونز، پنیلا، اسلاباق و همکاران ۲۰۱۵، چیگولوری، گوتی‌کوندا، اسلاباق، هیونز، پنیلا و کوردیه ۲۰۱۵، کوردیه و همکاران ۲۰۱۵).

هومانا از آن زمان «روزهای سالم» را به عنوان سنجهی مناسب برای اندازه‌گیری میزان پیشرفت در جهت تحقق «بولد گُل» بکار گرفته‌اند.

(Humana, http://populationhealth.humana.com/wpcontent/uploads/2016/05/BoldGoal2016ProgressReport_1.pdf)

قانون پارتو معمولاً در صنعت مراقبت بهداشت و سلامت صادق است یعنی حدود ۲۰ درصد از مشتریان در حدود ۸۰ درصد از کل منابع مراقبت بهداشت و سلامت را به دلیل شرایط مزمن و وخامت سلامتی‌شان به خود اختصاص داده‌اند. بیمه‌گران سلامت مثل هومانا معمولاً پریسک‌ترین ثبت‌نام‌کنندگان را در برنامه‌های مدیریت درمان و بالینی ثبت‌نام کرده‌اند تا به آنها در مدیریت شرایط مزمن کمک کنند.

شناسایی اعضای درست برای این موضوع بسیار مهم است و در سال‌های اخیر، PMها برای شناسایی ثبت‌نام‌کنندگانی که در آینده ریسک بالایی دارند بوجود آمده‌اند. بسیاری از این PMها با اتکای زیاد به داده‌های ادعاهای پزشکی ایجاد شده‌اند که منشأ آنها خدمات پزشکی بوده که ثبت‌نام‌کنندگان استفاده کرده‌اند. با توجه به وقفه‌ای که در ارسال و پردازش داده‌های ادعاهای پزشکی وجود دارد، شناسایی اعضای پریسک برای نام‌نویسی در برنامه‌های بالینی نیز دارای وقفه است. این مسئله به ویژه زمانی مرتبط است که اعضای جدید که تاریخچه‌ی ادعای پزشکی با یک بیمه‌گر ندارند به یک مؤسسه‌ی بیمه‌گر سلامت می‌پیوندند. یک PM مبتنی بر ادعاهای پزشکی ممکن است به طور میانگین بین ۹-۱۲ ماه پس از نام‌نویسی اعضای جدید برای شناسایی آنها جهت ارجاع به برنامه‌های بالینی به طول بینجامد.

¹ Unhealthy Days (UHD)

² Havens, Pena, Slabaugh, Cordier, Redna & Gopal

در اوایل این دهه، هومانا تعداد زیادی از اعضای جدید را جذب محصولات «مدیکر آدونتیج» نمود و به روش بهتری برای مدیریت بالینی این عضویت‌ها نیاز داشت. بدین ترتیب، توسعه‌ی یک رویکرد تحلیل متفاوت برای شناسایی سریع و دقیق اعضای جدید پرریسک برای مدیریت بالینی، حفظ سلامت این گروه و پایین نگه‌داشتن هزینه‌ها بسیار اهمیت پیدا کرد.

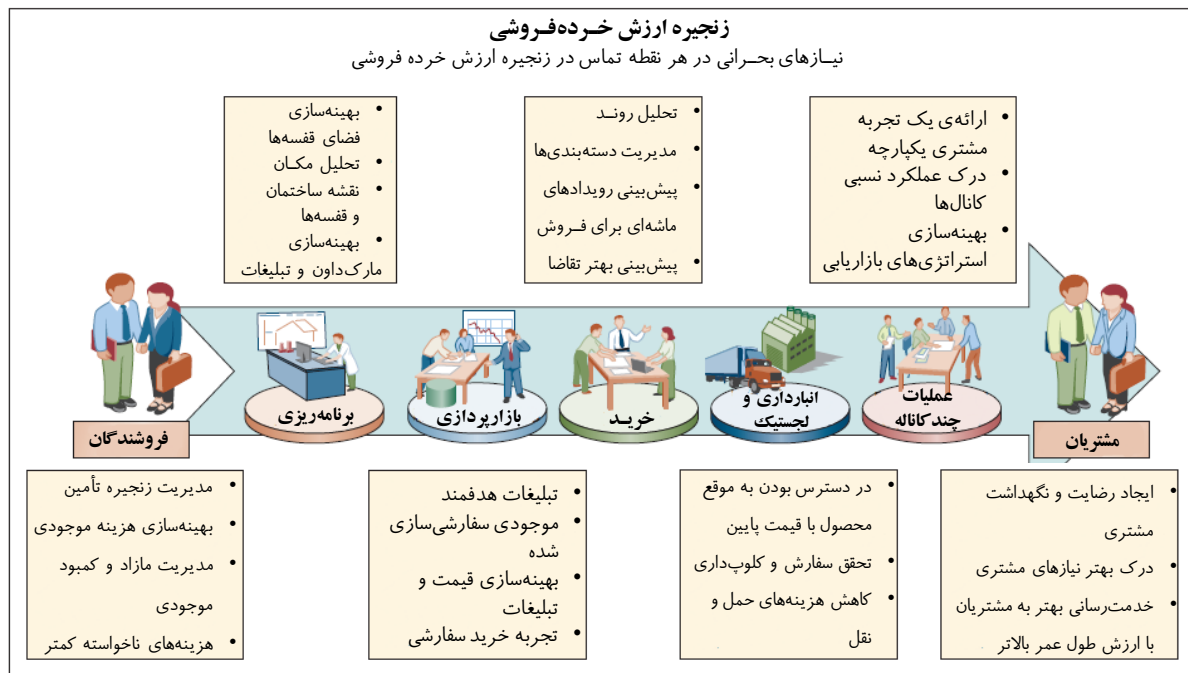
تیم تحلیل بالینی هومانا یک مدل پیشگویانه اعضای جدید را توسعه دادند که به ای اینکه منتظر شوند تا سابقه‌ی ادعاهای پزشکی اعضای جدید برای جمع‌آوری شرح حال‌های بالینی و پیش‌بینی ریسک‌های آینده، سریعاً می‌توانست افراد دارای ریسک را بلافاصله پس از نام‌نویسی آنها در برنامه جدید هومانا شناسایی کند. این مدل که برای رسیدگی به چالش‌های منحصرفرد مربوط به اعضای جدید تدوین شده بود، روش نوینی را توسعه دادند که از مجموعه داده‌های گسترده‌تر و یکپارچه‌تری که فراتر از داده‌های ادعاهای پزشکی از قبیل ارزیابی ریسک‌های سلامت خوداظهاری و شاخص‌های اولیه‌ی داده‌های دارویی بودند بهره می‌بردند. تکنیک‌های پیشرفته‌ی داده‌کاوی را برای کشف الگوها بکار گرفته و به هر مصرف‌کننده‌ی خدمات مدیکر آدونتیج (یک طرح بیمه خاص) به طور روزانه و بر اساس آخرین داده‌هایی که هومانا تا به امروز در اختیار دارد امتیازدهی می‌کنند. این مدل توسط یک تیم چندوظیفه‌ای از حوزه‌های تحلیل، فناوری اطلاعات و عملیات به منظور حصول اطمینان از یکپارچه‌سازی بی‌عیب و نقص عملیاتی و کسب‌وکار اجرایی شد.

از زمان پیاده‌سازی این مدل در ژانویه ۲۰۱۳، به سرعت به شناسایی اعضای جدید پرریسک برای نام‌نویسی در برنامه‌های بالینی هومانا نموده است. نتایج مثبت بدست آمده از این مدل در چندین ابلاغیه مهم رهبران ارشد این مؤسسه بیان شده است. در سه ماهه اول سال ۲۰۱۳، در پوزنت درآمدهای منتشر شده برای سرمایه‌گذاران، بروس بروسارد مدیرعامل هومانا بر اهمیت «بهبود PM اعضای جدید و فرایندهای ارزیابی بالینی» که منجر به نام‌نویسی ۳۱،۰۰۰ عضو جدید در برنامه‌های بالینی شد تأکید کید که این رقم با توجه به تعداد ۴،۰۰۰ نفری دوره مشابه سال گذشته یک افزایش ۶۷۵ درصدی را نشان می‌دهد. علاوه بر افزایش تعداد نام‌نویسی‌ها در برنامه‌های بالینی، مطالعات نتایج نشان داده است که افرادی که توسط NMPM شناسایی شده و به تازگی نام‌نویسی کرده‌اند پیش‌تر نیز از برنامه‌های بالینی استفاده کرده‌اند که بیش از ۵۰ درصد از آنها در سه ماهه ابتدای سال و پس از نام‌نویسی‌های جدید طرح MA بوده‌اند. مشتریان شناسایی شده همچنین با نرخ بالاتری در برنامه‌ها مشارکت کرده و مدت زمان ماندگاری آنها در برنامه‌ها نیز طولانی‌تر بوده است.

این مثال‌ها نشان می‌دهند که یک سازمان چگونه به بررسی دقیق و پیاده‌سازی کاربردهای تحلیل به منظور دستیابی به اهداف استراتژیک خود می‌پردازد. در فصل‌های مختلف این کتاب مثال‌های متعدد دیگری از کاربردهای مراقبت سلامت و بهداشت را مطالعه خواهیم کرد.

تحلیل در زنجیره ارزش خرده‌فروشی بخش خرده‌فروشی، جایی است که شاید بتوانیم بیشترین کاربردهای تحلیل را در آنجا ببینیم. اینجا حوزه‌ای است که با حجم بالای ولی حاشیه سودهای معمولاً پایین سر و کار داریم. اولویت، سلیقه و ذائقه‌ی مشتریان به سرعت تغییر می‌کند. فروشگاه‌های واقعی و مجازی (آنلاین) برای موفقیت با چالش‌های بسیاری مواجه هستند و اینکه زمانی شما بر بازار سلطه داشته باشید تضمین نمی‌کند که موفقیت شما ادامه‌دار باشد. بنابراین سرمایه‌گذاری برای شناخت و یادگیری در مورد تأمین‌کنندگان، مشتریان، کارکنان و تمامی ذینفعانی که به موفقیت یه زنجیره ارزشی خرده‌فروشی کمک می‌کنند و استفاده از آن اطلاعات به منظور اتخاذ تصمیمات بهتر، برای مدت زمان طولانی یکی از اهداف صنعت تحلیل بوده است. حتی خوانندگان عادی شاید با سرمایه‌گذاری‌های هنگفت آمزون در حوزه تحلیل برای پیشبرد زنجیره تأمین‌شان آشنا باشند. به همین منوال، وال‌مارت، تارگت و سایر خرده‌فروشی‌های بزرگ، میلیون‌ها دلار برای تحلیل در زنجیره‌های تأمین خود

سرمایه‌گذاری کرده‌اند. اکثر فناوری‌های تحلیل و ارائه‌دهندگان خدمات در تحلیل‌های خرده‌فروشی حضور پررنگی دارند. پرداختن حتی به بخش کوچکی از این کاربردها برای دستیابی به هدف‌مان می‌تواند یک کتاب کامل را به خود اختصاص دهد. بنابراین در این بخش تنها به ذکر چند کاربرد خواهیم پرداخت. بسیاری از خرده‌فروشی‌ها از این تحلیل‌ها استفاده کرده‌اند و بسیاری از ارائه‌کنندگان فناوری‌ها این تحلیل‌ها را ارائه می‌کنند. به همین دلیل در این بخش به جای بررسی موارد خاص، نگاهی کلی‌تر بر آنها خواهیم داشت. این دیدگاه کلی توسط آبیشک راتی مدیرعامل شرکت **vCreaTek.com** پیشنهاد شده است. شرکت vCreaTek LLC یک شرکت ارائه‌دهنده خدمات و نرم‌افزار تحلیل بوتیک است که دفاتری در هندوستان، ایالات متحده، امارات متحده عربی و بلژیک دارد. این شرکت، برنامه‌های کاربردی را در حوزه‌های مختلف طراحی و ایجاد می‌کند ولی حوزه تخصصی آن، تحلیل‌های خرده‌فروشی است.



شکل ۱.۱۵ یک مثال از کاربرد تحلیل‌ها در زنجیره ارزش خرده‌فروشی

Source: Contributed by Abhishek Rathi, CEO, vCreaTek.com.

در شکل ۱.۱۵ برخی از اجزای منتخب یک زنجیره ارزش خرده‌فروشی نشان داده شده است. این زنجیره از تأمین‌کنندگان شروع شده و به مشتریان ختم می‌شود ولی بسیاری از نقاط تصمیم در فرایند برنامه‌ریزی استراتژیک و عملیاتی را که تحلیل‌ها – توصیفی، پیشگویانه یا تجویزی – می‌توانند در تصمیم‌گیری بهتر بر اساس داده‌ها نقش مهمی ایفا کنند را به تصویر کشیده است. جدول ۱.۱ نیز برخی از حوزه‌های مهم کاربرد تحلیل‌ها به همراه مثال‌هایی از سوالات مهمی که می‌توان از طریق تحلیل‌ها پاسخ داد و همچنین ارزش کسب‌وکار بالقوه ناشی از این تحلیل‌ها را به تصویر کشیده است.

یک سایت خرده‌فروشی آنلاین معمولاً به محض اینکه مشتری در سایت ثبت‌نام کرده و وارد سایت می‌شود، مشتری را شناسایی کرده و می‌تواند صفحات/محصولات سفارشی را برای ارتقاء تجربه‌ی مشتری ارائه کند. برای هر فروشگاه خرده‌فروش، شناخت مشتری در ورودی فروشگاه همچنان موضوع چالش‌برانگیزی است. فروشگاه از طریق ترکیب تحلیل‌های ویدئویی و اطلاعات منتشر

شده از طریق برنامه وفاداری خود می‌تواند مشتری را در زمان ورود به فروشگاه شناسایی کرده و فرصت مضاعفی را برای فروش جانبی^۱ یا بیش‌فروشی^۲ فراهم می‌کند. علاوه بر این، با سفرهای سازی بیشترِ روش‌های مشارکت مشتری در زمان حضورش در فروشگاه می‌تواند تجربه‌ی خرید را شخصی‌سازی نمود.

خرده‌فروشی‌ها، پول زیادی را صرف نمایشگرهای جذاب، رویدادهای تبلیغاتی، تصاویر گرافیکی سفرهای سازی شده، دکوراسیون، آگهی‌های چاپی و بنرها می‌کنند. برای درک اثربخشی این روش‌های بازاریابی، تیم می‌تواند با مشاهده‌ی تصاویر تلویزیون‌های مداربسته و پی بردن به جزئیات جمعیت‌شناختی مشتریان و تحلیل خریدار استفاده کند. تصاویر تلویزیون مدار بسته را می‌توان با استفاده از الگوریتم‌های پیشرفته و پی بردن به جزئیاتی از قبیل سن، جنسیت یا حس و حال شخصی که در داخل فروشگاه گشت و گذار می‌کند تجزیه و تحلیل نمود.

علاوه بر این، داده‌های مربوط به گشت و گذار مشتری در داخل فروشگاه زمانیکه با چیدمان یا نقشه داخل فروشگاه ترکیب شود می‌تواند دیدگاه‌های بیشتر و بهتری را برای شناسایی بخش‌های سودآور/خوش فروش داخل فروشگاه به مدیر فروشگاه ارائه کند. همچنین، مدیر فروشگاه می‌تواند از این اطلاعات برای برنامه‌ریزی تخصیص نیروی کاری خود به بخش‌های مهم در زمان‌های اوج ترافیک استفاده کند.

جدول ۱.۱ مثال‌هایی از کاربرد تحلیل در زنجیره ارزشی خرده‌فروشی

کاربرد تحلیل	سوال کسب‌وکار	ارزش کسب‌وکار
بهینه‌سازی موجودی	۱. کدام محصولات تقاضای بالایی دارند؟ ۲. کدام محصولات گردش کندی دارند یا راکد شده‌اند؟	۱. پیش‌بینی مصرف محصولات پرگردش و سفارش‌گذاری آنها به اندازه کافی برای اجتناب از فقدان موجودی ۲. افزایش گردش محصولات کندگردش از طریق ترکیب آنها با محصولات با تقاضای بالا
کاهش قیمتی	۱. حاشیه سود خالص من روی محصولات چقدر است؟ ۲. بر روی این محصولات چقدر تخفیف می‌توانم بدهم؟	۱. قیمت‌های مارک‌داون برای هر محصول را می‌توان برای کاهش زیان حاشیه، بهینه‌سازی نمود. ۲. بهینه‌سازی قیمت سبد محصولات می‌تواند منجر به صرفه‌جویی حاشیه دلار شود.
تجزیه و تحلیل سبد بازار	۱. برای ارائه سبد پیشنهادی باید چه محصولاتی را ترکیب کنم؟ ۲. آیا محصولات را باید بر اساس ویژگی‌های پرگردش و کندگردش ترکیب کنم؟ ۳. آیا سبد پیشنهادی را باید از محصولات یک گروه یا گروه‌های مختلف ایجاد کنم؟	۱. در تحلیل وابستگی، همبستگی‌های پنهان بین محصولات شناسایی می‌شود که منجر به خلق ارزش‌های زیر می‌شوند: • تدوین استراتژی ایجاد سبد محصولات با تمرکز بر موجودی یا حاشیه سود

¹ Cross-selling

² Upselling

• افزایش فروش جانبی یا بیش‌فروشی با

ایجاد سبد محصولات از گروه‌های مختلف یا یکسان.

۱. از طریق بخش‌بندی مشتریان، صاحب کسب‌وکار می‌تواند پیشنهادات شخصی‌سازی شده ایجاد کرده و منجر به تجربه مشتری بهتر و حفظ و نگهداشت مشتری شود.

۱. کسب‌وکارها می‌توانند روابط بین محصولات و مشتریانی که خوب کار نکرده و ریزش بالایی دارند را شناسایی کنند. بدین ترتیب می‌توانند تمرکز بیشتری بر کیفیت محصول و دلایل ریزش آن داشته باشند.

۲. بر اساس ارزش طول عمر مشتری، کسب‌وکارها می‌توانند بازاریابی هدفمند را در دستور کار قرار داده و به حفظ مشتری کمک کنند.

۱. بودجه‌ی بازاریابی را می‌توان بر اساس پرسونای خریدار و دستیابی به نرخ بازگشت سرمایه‌ی بالاتر بهینه‌سازی نمود.

۱. می‌توان از بهترین تجارب سایر مکان‌ها و کانال‌ها برای شروع کار استفاده کرد.
۲. مقایسه با داده‌های رقبا می‌تواند به ایجاد یک عامل متمایزکننده یا جذب مشتریان جدید کمک کند.

۱. رابطه‌ی محصولات برای تصمیم‌گیری در مورد چیدمان فروشگاه و همراستایی بهتر با نیازهای مشتریان را درک کنید.

۲. استقرار نیروی کار را برای تعامل بهتر با مشتری و رضایت مشتری از تجربه خرید می‌توان برنامه‌ریزی کرد.

پرسونای خریدار

۱. کدام مشتری چه محصولی را از کجا خریداری می‌کند؟

تحلیل عوامل ریزش مشتری

۱. کدام مشتریان شما تمایلی به بازگشت ندارند؟
۲. چه سهمی از کسب‌وکارم را از دست خواهم داد؟
۳. چگونه می‌توانم مشتریانم را حفظ کنم؟
۴. چه بخشی از مشتریان، مشتریان وفادار هستند؟

تجزیه و تحلیل کانال

۱. کدام کانال‌ها برای جذب مشتریان هزینه پایین‌تری دارند؟
۲. کدام کانال‌ها نرخ حفظ و نگهداشت مشتری بالاتری دارند؟
۳. کدام کانال‌ها سودآورتر هستند؟

تجزیه و تحلیل فروشگاه جدید

۱. فروشگاه را در چه مکانی بهتر است افتتاح کنم؟
۲. برای شروع کار چه موجودی و به چه میزان باید داشته باشم؟

چیدمان فروشگاه

۱. چیدمان فروشگاه برای دستیابی به تاپ‌لاین بهتر چیست؟
۲. چگونه می‌توانم تجزیه درون‌فروشگاهی مشتری را ارتقاء بدهم؟

فروش وارد فروشگاه می‌شود؟

می‌توان بر اساس گروه جمعیت‌شناختی

۲. چگونه می‌توانم یک مشتری با ارزش طول

مشتریان ورودی برنامه‌ریزی کرد.

عمر بالا را در بدو ورود به فروشگاه شناسایی

۲. مشارکت هدفمند مشتریان و تخفیفات

کنم تا بتوان تجربه خرید شخصی‌سازی شده

لحظه‌ای می‌توانند تجربه‌ی مشتری و در نتیجه،

بهتری را به او عرضه کنم؟

حفظ و نگهداشت مشتری را ارتقاء دهند.

مدیران گروه محصولات برای افزایش فروش محصولاتی که موجب کندگردش شدن موجودی شده‌اند عمدتاً از تجزیه و تحلیل سبد بازار استفاده کرده‌اند. وابستگی محصول را می‌توان با استفاده از تحلیل‌های پیشرفته‌ی داده‌های موجود در پایین‌ترین سطح واحد نگهداری موجودی بکار برد تا بتوان به نرخ بازگشت سرمایه بالاتری در سبدهای محصول عرضه شده دست پیدا کرد. علاوه بر این، با استفاده از تکنیک‌های کَشش قیمت می‌توان قیمت مارک‌داون یا بهینه سبد محصولات را به گونه‌ای تعیین کرد که ضرر و زیان حاشیه سود کاهش یابد.

بدین ترتیب، یک خرده‌فروش با استفاده از تحلیل‌های داده نه تنها می‌تواند اطلاعاتی را در خصوص عملیاتی جاری خود بدست آورد بلکه می‌تواند به بینش عمیق‌تری برای افزایش درآمد و کاهش هزینه عملیاتی و سود بیشتر دست یابد. یک وبلاگ‌نویس در «مرکز علوم داده»، فهرست نسبتاً جامعی از کاربردهای جاری و بالقوه‌ی تحلیل‌ها در خرده‌فروشی که یک خرده‌فروش بزرگ مثل آمازون می‌توان از آن استفاده کند ارائه کرده است. این لیست در آدرس زیر قابل دسترسی است:

www.datasciencecentral.com/profiles/blogs/20-data-science-systems-used-byamazon-to-operate-its-business همانطور که قبلاً گفته شد، مثال‌های بسیاری از این فرصت‌ها وجود دارد ولی مثال‌های بسیاری از این کاربردها را در سرتاسر کتاب خواهید دید.

تحلیل تصویر همانطور که در این بخش دیدیم، تکنیک‌های تحلیل در صنایع و بر روی داده‌های بسیار متنوعی بکار گرفته می‌شوند. یکی از حوزه‌هایی که رشد چشمگیری داشته است، تجزیه و تحلیل تصاویر دیداری است. پیشرفت‌های به عمل آمده در تصویربرداری از طریق دوربین‌های با تفکیک‌پذیری بالا، قابلیت‌های ذخیره‌سازی و الگوریتم‌های یادگیری عمیق، انجام تجزیه و تحلیل‌های بسیار جالب را امکان‌پذیر کرده‌اند. داده‌های ماهواره‌ای، سودمندی خود را در رشته‌های مختلف اثبات کرده‌اند. مزایای داده‌های ماهواره‌ای با تفکیک‌پذیری بالا و در شکل‌های مختلف از جمله چندطیفی از اهمیت بالایی برای دانشمندی که نیازمند پایش منظم تغییرات جهانی، بهره‌برداری از زمین‌ها و آب و هوا هستند برخوردارند. در واقع، با ترکیب تصاویر ماهواره‌ای و سایر داده‌ها از جمله اطلاعات مربوط به رسانه‌های اجتماعی، فایل‌های دولتی و غیره می‌توان فعالیت‌های برنامه‌ریزی کسب‌وکار، الگوهای ترافیک، تغییرات در پارکینگ خودروها یا فضاهای باز را حدس زد. شرکت‌ها، نهادهای دولتی و سازمان‌های غیردولتی سرمایه‌گذاری هنگفتی در حوزه ماهواره‌ها انجام داده‌اند تا بتوانند از طریق تصویربرداری روزانه از کل کره‌ی زمینو ردگیری تغییرات روزانه در هر مکان از اطلاعات بدست آمده برای پیش‌بینی‌ها استفاده کنند. طی چندین ماه گذشته، مثال‌های بسیار جالبی از پیش‌بینی‌های پیشرفته و جالب گزارش شده است. علاوه بر این، این فعالیت توسط صنایع مختلف در سرتاسر جهان

رهبری شده و عبارت دیگری را به مفهوم کلان داده‌ها افزوده است: *داده‌های جایگزین*^۱. در اینجا چند مثال از تارتار و همکاران (۲۰۱۸) آورده‌ایم. مطالب بیشتر را در فصل ۹ و با مطالعه‌ی موضوع کلان داده‌ها خواهیم دید.

- پژوهشگران بانک جهانی از داده‌های ماهواره‌ای برای ارائه‌ی توصیه‌های استراتژیک به برنامه‌ریزان و مقامات شهری در کشورهای در حال توسعه استفاده کردند. این تجزیه و تحلیل به دلیل وقوع بلایای طبیعی اخیر که منجر به فوت ۴۰۰ نفر در شهر فری‌تاون کشور سیرالئون شد انجام پذیرفت. پژوهشگران به وضوح نشان دادند که فری‌تاون و برخی دیگر از شهرهای در حال توسعه فاقد برنامه‌ریزی سیستماتیک برای زیرساخت‌هایشان بودند که منجر به از دست رفتن زندگی شده است. پژوهشگران بانک از تصویربرداری ماهواره‌ای برای اتخاذ تصمیمات مهم در نواحی شهری ریسک‌پذیر استفاده می‌کنند.
- نرم‌افزار ارت‌کست^۲ اطلاعات دقیق در خصوص شرایط آب و هوایی را برای یک شرکت بزرگ هواپیمایی در ایالات متحده و بر اساس داده‌هایی که از ۶۰ ماهواره‌ی دولتی گرفته است به همراه داده‌های حسگرهای زمینی و هوایی ارائه می‌کند که تقریباً هر چیزی را از رعد و برق گرفته تا توربولانس ردگیری می‌کنند. حتی قابلیت نگاشت شرایط در طول مسیر پرواز را ایجاد کرده و پیش‌بینی‌های سفرهای را برای هر چیزی از بالن‌های هوای داغ تا پهپادها عرضه می‌کند.
- ایمازون^۳ شروع به استفاده از داده‌های ماهواره برای ایجاد تصویری از اطلاعات لحظه‌ای نزدیک از جنگل‌زدایی آمازون نموده است. تصویربرداری آپتیکال و مادون قرمز پیشرفته‌ای را شروع کرده که منجر به شناسایی کارخانه‌های چوب‌بری غیرقانونی شده است. ایمازون اکنون بیشتر بر ارائه‌ی داده‌ها به دولت‌های محلی از طریق برنامه‌ی «شهرداری‌های سبز» متمرکز شده است که به آموزش مقامات برای شناسایی و کنترل جنگل‌زدایی‌ها می‌پردازد.
- دولت اندونزی با همکاری سازمان بین‌المللی غیرانتفاعی دیده‌بان جهانی شیلات اقدام به شناسایی محل و زمان صید غیرقانونی کشتی‌ها با استفاده از اطلاعات ماهواره‌ای استخراج شده بر اساس حرکت و جابجایی کشتی‌ها نموده است (امرت ۲۰۱۸). این ابتکار عمل نتایج آنی را به همراه داشته است: درآمد دولت از صید در سال ۲۰۱۷ در مقایسه با سال ۲۰۱۴ به میزان ۱۲۹ درصد افزایش یافت. انتظار می‌رود که در دهه‌ی پیش‌رو، این سازمان بتواند ۷۵ درصد از کشتی‌هایی که به صید غیرقانونی می‌پردازند را ردیابی نماید.

این مثال‌ها فقط نمونه‌هایی از روش‌های بکارگیری داده‌های ماهواره‌ای را در کنار تحلیل‌ها برای دستیابی به بینش و درک جدید نشان می‌دهند. با پیش‌بینی عصر پیش‌رو که حاوی مشاهدات ماهواره‌ای از کره زمین می‌باشد، دانشمندان و مجامع باید برای شناسایی کاربردهای مهم و مسائل علمی مهم جهت بهبود وضعیت جامعه فکری کنند. اگرچه اینگونه دغدغه‌ها سرانجام توسط سیاست‌گذاران حل خواهند شد ولی آنچه روشن است این است که روش‌های جدید و جالبی از ترکیب داده‌های ماهواره‌ای و بسیاری منابع داده دیگر منجر به شکل‌گیری شرکت‌های تحلیل می‌شوند.

اینگونه تجزیه و تحلیل تصور محدود به تصاویر ماهواره‌ای نیست. دوربین‌های نصب شده بر روی پهپادها و چراغ‌های راهنمایی بر روی هر برجک در ساختمان‌ها و خیابان‌ها توانایی اخذ تصاویر را از ارتفاع چند متری فراهم می‌کند. تجزیه و تحلیل این تصاویر به همراه فناوری‌های تشخیص چهره موجب شده تا کاربردهای جدیدی از شناسایی مشتری تا توانایی دولت برای ردیابی کلیه‌ی

¹ Alternative Data

² EarthCast

³ Imazon

موضوعات مورد توجه امکان‌پذیر شود. کاربردهایی از این نوع منجر به بحث و گفتگوهای بیشتر در خصوص مسائل حریم خصوصی شده‌اند. در مورد کاربردی ۱.۷ ما در مورد کاربردهای خیرخواهانه بیشتری از تحلیل تصاویر خواهیم آموخت که تصاویر توسط یک تلفن همراه گرفته شده و این اپلیکیشن می‌تواند ارزش فوری برای کاربر داشته باشد.

مورد کاربردی ۱.۷

تحلیل تصویر به تخمین پوشش گیاهی کمک می‌کند

تخمین اینکه چه میزان از زمین‌ها دارای پوشش گیاهی می‌باشد در تجزیه و تحلیل یک جنگل یا حتی یک مزرعه بسیار اهمیت دارد. در خصوص جنگل‌ها، اینگونه تجزیه و تحلیل‌ها به کاربران کمک می‌کند تا چگونگی رشد جنگل‌ها، تأثیر آنها بر زمین‌های اطراف و حتی شرایط جوی را درک کنند. در خصوص مزارع، تجزیه و تحلیل‌های مشابه می‌توانند به درک احتمال رشد گیاهان و تخمین میزان باردهی گیاهان در آینده کمک کنند. مشخصاً، اندازه‌گیری پوشش گیاهی جنگل‌ها به صورت دستی کاری غیرممکن و در خصوص مزارع بسیار چالش‌برانگیز است. یک روش متداول این است که تصاویر جنگل/مزرعه را گرفته و سپس آنها را به منظور تخمین میزان پوشش زمین، تجزیه و تحلیل کنیم. انجام اینگونه تجزیه و تحلیل‌ها به صورت چشمی کار بسیار پرهزینه‌ای بوده و همچنین احتمال خطا وجود دارد. کارشناسان مختلف ممکن است درصد پوشش گیاهی زمین را به شکل متفاوتی تخمین بزنند. بنابراین، روش‌های خودکار برای تجزیه و تحلیل این تصاویر و تخمین درصد پوشش گیاهی زمین در حال توسعه است. یکی از این روش‌ها و یک اپلیکیشن برای انجام اینکار از طریق یک دستگاه تلفن همراه توسط پژوهشگران دپارتمان علوم خاک و گیاه دانشگاه آکلاهما با مشارکت مرکز اپلیکیشن دانشگاه و گروه فناوری اطلاعات در بخش علوم کشاورزی و منابع طبیعی عملی شده است.

کانوپو^۱ یک اپلیکیشن رایگان دسکتاپ یا موبایل است که پوشش تاج سبز را از روی تصاویر رفته شده با یک گوشی هوشمند یا یک دوربین دیجیتال تخمین می‌زند. در آزمایشاتی که در مزارع ذرت، گندم، کانولا و سایر محصولات انجام شده، کانوپو درصد پوشش تاج را چندین هزار برابر سریع‌تر از سایر نرم‌افزارهای موجود و بدون کاهش دقت تخمین زده است. به گفته‌ی تیسون آکسنر^۲ استاد خاک‌شناسی دانشگاه ایالتی آکلاهما، این اپلیکیشن برخلاف سایر برنامه‌ها می‌تواند تصاویر ویدئویی را گرفته و تجزیه و تحلیل کند که این ویژگی میزان خطای نمونه‌برداری مربوط به تخمین پوشش تاج را کاهش می‌دهد. به گفته‌ی آکسنر که مدیر توسعه‌ی این اپلیکیشن با همکاری آندرس پاترینیانی^۳ از دانشجویان سابق دکترا و عضو فعلی هیئت علمی دانشگاه ایالتی کانزاس می‌باشد: «ما می‌دانیم که پوشش گیاهی، تاج گیاهان می‌تواند بسیار متغیر باشد». «با استفاده از کانوپو فقط کافی است دستگاه را روشن کرده و در امتداد بخشی از زمین شروع کنید به قدم زدن و نتایج را برای هر فریم از ویدئویی که ضبط می‌کنید دریافت کنید.» با استفاده از دوربین دیجیتال یک گوشی هوشمند یا تبلت، کاربران کانوپو می‌توانند در داخل زمین از گیاهان از جمله محصولات، علفزار و چمن عکس و فیلم گرفته و آنها را وارد اپلیکیشن کرده و با تجزیه و تحلیل هر پیکسل از تصاویر آنها را بر اساس مقادیر عددی قرمز-سبز-آبی دسته‌بندی کنند. کانوپو پیکسل‌ها را بر اساس نسبت پیکسل‌های قرمز به سبز و آبی به سبز و همچنین شاخص سبز ماژاد^۴ تجزیه و تحلیل می‌کند. نتیجه‌ی تجزیه و تحلیل، یک تصویر است که پیکسل‌های رنگی آن به پیکسل‌های سیاه و سفید تبدیل شده‌اند که پیکسل‌های سفید متناظر با گیاهان سبز و پیکسل‌های سیاه بیانگر پس‌زمینه هستند. آزمون‌های مقایسه نشان داده است که کانوپو تصاویر را با سرعت بیشتر و دقت یکسانی با دو بسته‌ی نرم‌افزاری دیگر تجزیه و تحلیل می‌کند.

¹ Canopeo

² Tyson Ochsner

³ Andres Patrigniani

⁴ Excess green index

توسعه‌دهندگان اپلیکیشن کانوپئو انتظار دارند که این برنامه به تولیدکنندگان کمک کند تا تصمیم بگیرند که بهترین زمان برای حذف گندم زمستانه از غذای گله در سیستم‌های « دو منظوره » که گندم برای غلات نیز برداشت می‌شود چه زمان است. پژوهشی که تیم دیگری در OSU انجام داده نشان می‌دهد که خارج کردن گله از زمین در زمانیکه حداقل ۶۰ درصد از پوشش تاج سبز همچنان باقی است می‌تواند باردهی خوبی از لحاظ غلات نیز داشته باشد. به گفته‌ی پاترینیانی: « بنابراین کانوپئو برای این تصمیم‌گیری می‌تواند مفید باشد.» وی و اُکسنر معتقدند که این اپلیکیشن می‌تواند برای مدیریت چمنزارها، ارزیابی آسیب‌دیدگی محصول به خاطر شرایط آب و هوایی یا علف‌کش‌ها، یک جایگزین برای شاخص نرمال شده تفاوت پوشش گیاهی^۱ در توصیه‌های کوددهی و حتی در تصویربرداری‌های هوایی با پرده‌های بدون سرنشین^۲ از جنگل‌ها یا سیستم‌های آبی کاربرد داشته باشد. تجزیه و تحلیل تصاویر، یکی از حوزه‌های کاربردی در حال رشد در یادگیری عمیق و همچنین بسیاری از تکنیک‌های دیگر هوش مصنوعی است. در فصل ۹ چندین مثال از تجزیه و تحلیل تصاویر را که منجر به شکل‌گیری عبارت دیگری تحت عنوان داده‌های جیگزین شده است ارائه خواهیم کرد. در فصل ۶ نیز برخی از کاربردهای مهم آورده شده است.

سوالاتی برای بحث و گفتگو

۱. هدف از اینکه بدانیم چه درصدی از زمین در یک مزرعه یا در یک جنگل با علف سبز پوشیده شده است چیست؟
۲. چرا تحلیل تصویر علفزار از طریق یک اپلیکیشن بهتر از بررسی چشمی است؟
۳. برای درک منطق الگوریتمی زیربنایی تجزیه و تحلیل تصویر به جستجوی مقالات پژوهشی بپردازید.
۴. چه برنامه‌های تجزیه و تحلیل تصویر دیگری وجود دارد؟

Source: Compiled from A. Patrignani and T. E. Ochsner. (2015). "Canopeo: A Powerful New Tool for Measuring Fractional Green Canopy Cover." *Agronomy Journal*, 107(6), pp. 2312–2320; R. Lollato, A. Patrignani, T. E. Ochsner, A. Rocatelli, P. Tomlinson, & J. T. Edwards. (2015). *Improving Grazing Management Using a Smartphone App*. www.bookstore.ksre.ksu.edu/pubs/MF3304.pdf (accessed October 2018); <http://canopeoapp.com/> (accessed October 2018); Oklahoma State University press releases.

ابتکار عمل‌های علم داده‌ها/تحلیل به سرعت در حال رشد بوده و حتی با دستاوردهای جدید در هوش مصنوعی در حال ترکیب شدن هستند. در بخش بعدی مروری کلی بر هوش مصنوعی و بحث و گفتگوی کوتاهی در خصوص همگرایی این دو خواهیم داشت.

سوالاتی برای مرور بخش ۱.۶

۱. سه عامل که می‌توانند بخشی از یک PM برای تمدید بلیت‌های فصل باشند کدام هستند؟
۲. دو تکنیک که تیم‌های فوتبال می‌توانند برای تحلیل رقیب بکار بگیرند چه هستند؟
۳. چه کاربردهای دیگری از تحلیل را می‌توانید در ورزش نام ببرید؟
۴. چرا یک شرکت بیمه سلامت برای تشخیص کلاهبرداری بر روی تحلیل سرمایه‌گذاری می‌کند؟ چرا به نفع شرکت است تا احتمال زمین خوردن بیماران را پیش‌بینی کند؟
۵. چه کاربردهای دیگری مشابه پیش‌بینی زمین خوردن را می‌توانید در نظر بگیرید؟

¹ Normalized Difference Vegetation Index (NDVI)

² UAV based

۶. چگونه یک مشتری جدید مؤسسه بیمه سلامت را می‌توانید متقاعد کنید تا سبک زندگی سالم‌تری را اتخاذ کند؟
۷. حداقل سه فرصت دیگر برای بکارگیری تحلیل‌ها در زنجیره ارزشی خرده‌فروشی را به غیر از آنهایی که در کتاب آورده‌ایم شناسایی کنید.
۸. کدام فروشگاه‌های خرده‌فروشی را می‌شناسید که از برخی از کاربردهای تحلیل که در این بخش معرفی کردیم استفاده می‌کنند؟
۹. نقطه مشترک بین مثال‌هایی که در تحلیل تصویر ذکر کردیم چه چیزی است؟
۱۰. آیا می‌توانید کاربردهای دیگری از بکارگیری داده‌های ماهواره‌ای را در کنار مطالب این بخش تصور کنید؟

اگرچه تحلیل‌های مراقبت بهداشت طیف بسیار گسترده‌ای از کاربردهای را از پیشگیری تا تشخیص و عملیات مؤثر و پیشگیری از کلاهبرداری شامل می‌شوند، ما بر روی برخی از کاربردها که در یک شرکت بزرگ بیمه سلامت و بهداشت در ایالات متحده به نام هوماننا توسعه یافته است تمرکز می‌کنیم. بر طبق وبسایت این شرکت، «استراتژی این شرکت، یکپارچه‌سازی ارائه‌ی خدمات مراقبت، تجربه‌ی اعضای شرکت و بینش و درک پرسنل بالینی و مراجعه‌کنندگان به منظور ترغیب مشارکت، تغییر رفتار، اطلاع‌رسانی و سلامتی بالینی فعال... است.»

۱.۷ مروری بر هوش مصنوعی

در اول سپتامبر ۲۰۱۷، اولین روز شروع مدارس در روسیه، ولادیمیر پوتین، رئیس‌جمهور روسیه در روز ملی درس آزاد برای بیش از ۱,۰۰۰,۰۰۰ نفر سخنرانی کرد. عنوان این سخنرانی تلویزیونی «روسیه‌ی متمرکز بر آینده» بود. در این معرفی، بینندگان شاهد دستاوردهای دانشمندان روسی در رشته‌های مختلف بودند. ولی آنچه همگان از این نمایش به خاطر دارند این یک جمله بود: «کشوری که در حوزه هوش مصنوعی رایانه‌ای پیشگام و رهبر باشد، حکمران آینده‌ی دنیا خواهد شد.» پوتین تنها کسی نیست که از ارزش هوش مصنوعی آگاه است. دولت‌ها و شرکت‌ها میلیاردها دلار هزینه می‌کنند تا تبدیل به رهبر و پیشگام در هوش مصنوعی شوند. به عنوان مثال، در جولای ۲۰۱۷، چین از طرحی برای ایجاد یک صنعت هوش مصنوعی به ارزش ۱۵۰ میلیارد دلار برای اقتصاد چین تا سال ۲۰۲۳ پرده برداشت (متز ۲۰۱۸). شرکت بادو چین امروزه بیش از ۵,۰۰۰ مهندس هوش مصنوعی را در استخدام خود دارد. دولت چین تسهیلات لازم به منظور پژوهش و بکارگیری هوش مصنوعی را به عنوان یک اولویت ملی فراهم کرده است. مؤسسه‌ی حسابداری پرایس واترهاوس کوپر محاسبه کرده که هوش مصنوعی تا سال ۲۰۳۰ ارزشی معادل ۱۵.۷ تریلیون دلار را به اقتصاد جهانی خواهد افزود (حدود ۱۴ درصد، لیبرتو ۲۰۱۷). بدین ترتیب، جای تعجب نیست که در مورد هوش مصنوعی بیش از هر موضوع فناوری دیگری در سال ۲۰۱۸ صحبت شده است.

تعاریف متعددی برای هوش مصنوعی وجود دارد. دلیل آن این است که هوش مصنوعی مبتنی بر نظریه‌هایی است که از چندین رشته‌ی علمی گرفته شده‌اند و شامل مجموعه‌ی گسترده‌ای از فناوری‌ها و کاربردها می‌باشد. بنابراین، شاید بهتر باشد برای درک ماهیت هوش مصنوعی نگاهی به برخی ویژگی‌های آن بیندازیم. هدف اصلی هوش مصنوعی، ایجاد یک ماشین هوشمند است که بتواند وظایفی که در حال حاضر توسط انسان انجام می‌شوند را انجام دهد. به طور ایده‌آل، این وظایف شامل استنتاج، فکر کردن،

یادگیری و حل مسئله است. در هوش مصنوعی به مطالعه‌ی توانایی پردازش فکر انسان برای درک آنچه هوشمند است می‌پردازد تا دانشمندان هوش مصنوعی بتوانند فرایندهای انسانی را بر روی ماشین‌ها کپی‌برداری کنند. (2017) eMarketer فهرست جامعی از تعاریف هوش مصنوعی را ارائه کرده است:

- فناوری که می‌تواند در طول زمان یاد بگیرد که کارها را بهتر انجام دهد.
- فناوری که می‌تواند زبان انسان را بفهمد و درک کند.
- فناوری که می‌تواند به سوالات پاسخ بدهد.

از آنجائیکه هوش مصنوعی به شکل‌های بسیاری وجود دارد، مزایای زیادی نیز دارد. این مزایا در فصل ۲ به طور کامل آورده شده و در اینجا به برخی مزایای مهم آن اشاره می‌کنیم:

- کاهش چشمگیر هزینه انجام کار. این هزینه در طول زمان همچنین کمتر می‌شود در حالیکه هزینه انجام همان کار به طور دستی در طول زمان بیشتر می‌شود.
- کارها را سریع‌تر می‌توان انجام داد.
- به طور کلی کاری که به این شکل انجام می‌شود در مقایسه با کار انسانی منسجم‌تر و سازگارتر است.
- افزایش سودآوری و بهره‌وری و همچنین مزیت رقابتی که محرک‌های اصلی هوش مصنوعی هستند.

چشم‌انداز یا اکوسیستم هوش مصنوعی بخش‌های بسیاری دارد. ما تصمیم گرفتیم تا آن را در قالب پنج گروه مطابق با شکل ۱.۱۶ دسته‌بندی کنیم. چهار گروه در این دسته‌بندی به عنوان پایه و اساس گروه پنجم عمل می‌کنند که همان کاربردهای هوش مصنوعی هستند. گروه‌ها به شرح زیر هستند:

فناوری‌های مهم ما در اینجا فناوری‌های یادگیری ماشین (فصل ۵)، یادگیری عمیق (فصل ۶) و عوامل هوشمند (فصل ۲) را در نظر گرفته‌ایم.

فناوری‌های دانش‌بنیان (فصل ۱۲) عناوین مطرح شده در این گروه شامل سیستم‌های خبره، موتورهای توصیه، چت‌بات‌ها، دستیارهای شخصی مجازی و مشاورهای رباتی می‌باشند.

فناوری‌های مربوط به بیومتریک این گروه شامل پردازش زبان‌های طبیعی (درک و تولید)، دید ماشینی و شناخت صحنه و تصویر و صدا و سایر مؤلفه‌های بیومتریک است (فصل ۶).

پلتفرم‌ها، ابزارها و نظریه‌های پشتیبان رشته‌های دانشگاهی شامل علوم رایانه، علوم شناختی، نظریه کنترل، زبانشناسی، ریاضی، علوم عصب‌شناسی، فلسفه، روانشناسی و آمار.

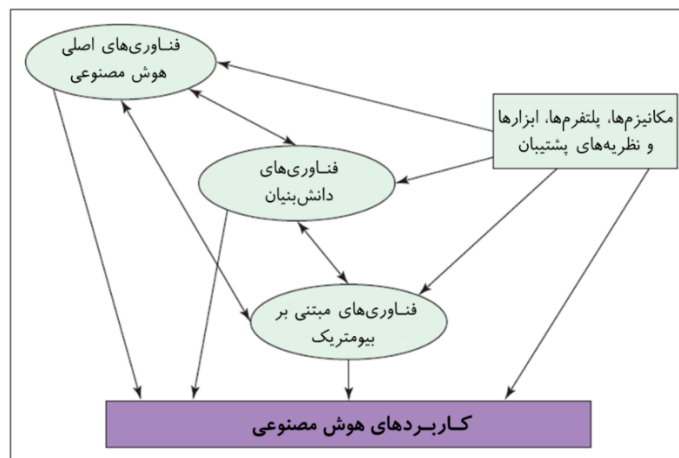
دستگاه‌ها و روش‌ها شامل حسگرها، واقعیت افزوده، آگاهی از زمینه، منطق، رایانش حرکتی، شناخت محتوا، فیلترینگ مشارکتی، شبکه‌های عصبی، داده‌کاوی، نظریه‌های انسان‌نما، استنتاج مبتنی بر مورد، رابط‌های برنامه‌نویسی کاربردی پیشگوبانه، مدیریت دانش، منطق فازی، الگوریتم ژنتیک، بسته‌بندی داده‌ها و غیره می‌باشد.

پلتفرم‌ها و ابزارها اینها توسط شرکت‌های آی‌بی‌ام، مایکروسافت، انویدیا و صدها فروشنده‌ی تخصصی در حوزه‌های مختلف هوش مصنوعی عرضه می‌شوند.

کاربردهای هوش مصنوعی صدها یا هزاران کاربرد برای هوش مصنوعی وجود دارد که فقط یک نمونه را در اینجا ذکر می‌کنیم:

شهرهای هوشمند، خانه‌های هوشمند؛ وسایل نقلیه‌ی خودکار (فصل ۱۳)، تصمیمات خودکار (فصل ۱۲)، ترجمه‌ی زبان، رباتیک (فصل ۱۰)، تشخیص کلاهبرداری، حفظ امنیت، غربالگری محتوا، پیش‌بینی، خدمات شخصی‌سازی شده و غیره. کاربردها در تمامی حوزه‌های کسب‌وکار (فصل ۲) و تقریباً در سایر حوزه‌ها از پزشکی و مراقب سلامت و بهداشت گرفته تا حمل و نقل و آموزش وجود دارد.

نکته: فهرست کامل این کاربردها از فاگلا (۲۰۱۸)^۱ و جاکت (۲۰۱۷)^۲ قابل دسترس می‌باشد. همچنین نگاه کنید به مقاله‌ی « نگاه اجمالی بر هوش مصنوعی^۳ » و فهرستی از «پروژه‌های هوش مصنوعی^۴» در ویکیپدیا. در مورد کاربردی ۱.۸ توضیح خواهیم داد که چگونه چندین فناوری با یکدیگر ترکیب شده و موجب بهبود امنیت و سرعت بخشیدن در رسیدگی به مسافران در فرودگاه‌ها می‌شود.



شکل ۱.۱۶ چشم‌انداز (اکوسیستم) هوش مصنوعی منبع: E. Turban

¹ Faggela

² Jacquet

³ "Outline of Artificial Intelligence", Wikipedia

⁴ AI Projects

انتظار در صف‌های کنترل امنیت در فرودگاه‌ها یا این فکر که ممکن است تروریست‌ها سوار هواپیما شده یا وارد کشور شوند ممکن است برای ما خوشایند نباشد. چندین دستگاه هوش مصنوعی برای به حداقل رساندن این احتمالات طراحی شده است:

۱. تشخیص چهره در فرودگاه. جت بلو^۱ در حال آزمایش فناوری تشخیص چهره است (یک نوع دید ماشینی برای تطبیق چهره‌های مسافران با عکس‌های از پیش ذخیره شده مثل گذرنامه یا گواهینامه رانندگی). بدین ترتیب نیازی به کارت پرواز و افزایش امنیت نخواهد بود. این تطبیق از کیفیت بالایی برخوردار است. این فناوری که بریتیش ایرویز پیشگام آن است توسط دلتا^۲، کی‌اِم^۳ و سایر خطوط هوایی که از فناوری‌های مشابهی برای چک کردن چمدان‌ها استفاده می‌کنند بکار گرفته شده است. فناوری مشابهی توسط آژانس اعمال مهاجرت و گمرک ایالات متحده مورد استفاده قرار گرفته که عکس افراد که در بدو ورود گرفته شده با عکس‌ها و سایر مستندات موجود در پایگاه داده تطبیق داده می‌شود.
۲. سیستم چین. فرودگاه‌های مهم در چین از سیستمی مشابه جت بلو استفاده می‌کنند که از تشخیص چهره برای صحت‌گذاری هویت مسافران استفاده می‌شود. هدف از اینکار، حذف کارت پرواز و افزایش سرعت مسافرگیری می‌باشد. از این سیستم همچنین به منظور شناسایی کارکنان فرودگاه در زمان ورود به بخش‌های با دسترسی محدود استفاده می‌شود.
۳. استفاده از بات‌ها. چندین فرودگاه (نیویورک، پکن) از بات‌های مکالمه‌ای (فصل ۱۲) برای راهنمایی مسافران استفاده می‌کنند. بات‌ها همچنین اطلاعاتی در مورد گمرک و خدمات مهاجرت ارائه می‌کنند.
۴. شناسایی دروغگویان در فرودگاه. این کاربرد نوظهور برای کمک به خدمات مهاجرت به منظور بررسی دقیق مسافران در فرودگاه‌ها و مرزهای ورود زمینی می‌باشد. با افزایش سطح امنیت، ممکن است لازم باشد که پرسنل خدمات مهاجرت و خطوط هوایی از مسافران پرس و جو کنند. این یک راه‌حل مقرون به صرفه برای پرس و جو از تمامی مسافران در کمترین زمان ممکن است که صف‌های طولانی انتظار را کاهش می‌دهد. این سیستم نوظهور، عامل مجازی خودکار برای ارزیابی آنی حقیقت (آواتار)^۴ نام دارد. الزامات این سیستم به شرح زیر می‌باشند:

الف). آواتار یک بات است که ابتدا می‌توانید گذرنامه‌تان را اسکن کنید.

ب). آواتار چند سوال از شما می‌پرسد. چندین فناوری هوش مصنوعی از قبیل هوش مصنوعی، تحلیل کلان‌داده‌ها، رباتیک، فناوری « ابر »، یادگیری ماشین، دید ماشینی و بات‌ها در این پروژه بکار گرفته شده است.

ج). شما به سوالات پاسخ می‌دهید.

د). حسگرها و سایر فناوری‌های هوش مصنوعی، داده‌هایی از قبیل حالت چهره، تغییرات صدا، موقعیت و حرکت چشم‌ها، حرکات دهان و ژست و حرکات بدن را گردآوری می‌کنند. پژوهشگران بر این باور هستند که گفتن حقیقت به مراتب راحت‌تر از مردن است و به همین دلیل پاسخ‌ها را با سوالات روتین مقایسه می‌کنند.

¹ Jet Blue

² Delta

³ KLM

⁴ Automated Virtual Agent for Truth Assessments in Real Time (AVATAR)

سپس ماشین، افراد مشکوک را برای تحقیق و بازجویی بیشتر اعلام می‌کند. این ماشین در حال حاضر توسط عوامل مهاجرت در چندین کشور بکار گرفته شده است.

Sources: Condensed from Thibodeaux, W. (2017, June 29). "This Artificial Intelligence Kiosk Is Designed to Spot Liars at Airports."
Inc.com.; Silk, R. (2017, November). "Biometrics: Facial Recognition Tech Coming to an Airport Near You." *Travel Weekly*, 21.

سوالاتی برای مورد کاربردی ۱.۷

۱. مزایای دستگاه‌های هوش مصنوعی برای مسافران را نام ببرید؟
۲. مزایای هوش مصنوعی برای دولت‌ها و شرکت‌های هواپیمایی را نام ببرید؟
۳. ارتباط بین این مورد کاربردی با دید ماشین و سایر ابزارهای هوش مصنوعی که با بیومتریک انسان‌ها سر و کار دارند را توضیح دهید.

هوش مصنوعی محدود در برابر گسترده رشته‌ی هوش مصنوعی را می‌توان به دو گروه از کاربردهای اصلی تقسیم نمود: محدود (ضعیف)^۱ و گسترده (قوی)^۲.

رشته‌ی هوش مصنوعی محدود بر یک رشته (دامنه) محدود تمرکز دارد. مثال‌های بارز از این نوع عبارتند از سیری^۳ و آلكسا^۴ (فصل ۱۲) که حداقل در اوایل پیدایش در حوزه‌های بسیار محدود و از پیش تعریف شده کاربرد داشتند. با گذشت زمان، کاربرد عمومی‌تری پیدا کرده و دانش بیشتری را کسب کردند. اکثر سیستم‌های خبره (فصل ۱۲) در حوزه‌های کاملاً محدودی فعالیت می‌کردند. اگر توجه کرده باشید، زمانیکه با یک مرکز تماس خودکار تماس می‌گیرید، رایانه (که معمولاً مبتنی بر فناوری هوش مصنوعی است) خیلی هوشمند نیست. اما با گذشت زمان، «هوشمندتر» می‌شوند. تشخیص گفتار به رایانه‌ها اجازه می‌دهد تا صوت را با دقت بالایی به متن تبدیل کنند. به همین منوال، دید رایانه‌ای در حال بهبود است که اشیاء را تشخیص داده و آنها را دسته‌بندی کرده و حتی متوجه جابجایی آنها می‌شود. در مجموع، میلیون‌ها کاربرد محدود برای هوش مصنوعی وجود دارد و این فناوری هر روز در حال بهبود است. با وجود این، هوش مصنوعی به اندازه کافی قوی نشده است چونکه قابلیت‌های حقیقی هوش انسانی را نمی‌تواند نمایش دهد.

هوش مصنوعی گسترده (قوی). ماشین‌ها برای اینکه بتوانند هوش واقعی انسانی را نشان دهند باید طیف کاملی از قابلیت‌های شناختی انسانی را انجام دهند. رایانه‌ها می‌توانند برخی از این قابلیت‌های شناختی را داشته باشند (مثلاً استنتاج و حل مسئله) که در فصل ۶ و بحث رایانش شناختی نشان خواهیم داد. تفاوت بین این دو گروه هوش مصنوعی با هوشمندتر شدن هوش مصنوعی در حال کاهش است. به طور ایده‌آل، هوش مصنوعی گسترده قادر خواهد بود تا انسان را کپی‌سازی کند. ولی هوش حقیقی فقط در حوزه‌های بسیار محدود مثل بازی، تشخیص پزشکی و تشخیص خرابی تجهیزات کاربرد دارد. برخی معتقدند که

¹ Narrow (Weak)

² General (Strong)

³ SIRI

⁴ Alexa

ما هیچگاه قادر نخواهیم بود تا یک ماشین هوش مصنوعی حقیقاً قوی بسازیم. برخی دیگر غیر از این را تصور می‌کنند. در ادامه یک مثال از یک بات هوش مصنوعی گسترده در یک حوزه‌های محدود را بررسی می‌کنیم.

شما اگر در استرالیا یا نیوزیلند زندگی می‌کنید و نزدیک یک دستگاه خودکار فروش کوکاکولا هستید می‌توانید یک بطری نوشیدنی غیرالکلی را با استفاده از تلفن هوشمندتان خریداری کنید. دستگاه‌ها از طریق فناوری ابر به یکدیگر متصل هستند به این معنی که می‌توانید نوشیدنی کوکا را از هر محلی در دنیا و نه فقط برای خودتان بلکه برای هر دوستی که در استرالیا یا نیوزیلند زندگی می‌کند و نزدیک یک دستگاه خودکار فروش کوکا است سفارش دهید (آشانسکی ۲۰۱۷). علاوه بر این، شرکت می‌تواند قیمت‌ها را از راه دور تغییر داده، تبلیغات انجام داده و داده‌های موجودی را گردآوری نماید تا بتواند مجدداً تأمین موجودی کند. تبدیل دستگاه‌های فعلی به دستگاه‌های هوش مصنوعی حدود ۱ ساعت به طول می‌انجامد. یک لحظه صبر کنید! اگر مشکلی پیش آمد آنوقت چه؟ هیچ مشکلی نیست. شما می‌توانید از طریق پیام‌رسان فیسبوک با چت‌بات شرکت کوکاکولا ارتباط برقرار کنید (فصل ۱۲).

استف^۱ (۲۰۱۷) قابلیت‌های سیستم‌های هوش مصنوعی را به سه سطح تقسیم کرده‌اند: همکار، مستقل و افزوده.

هوش مصنوعی همکار که عمدتاً معادل هوش مصنوعی محدود (ضعیف) است و فقط در حوزه‌های محدود کاربرد دارد. این هوش مصنوعی نیازمند ورودی‌ها و خروجی‌هایی است که به طور روشن و شفاف تعریف شده باشند. مثال‌هایی از این سطح شامل سیستم‌های پایش و دستیارهای شخصی مجازی سطح پایین هستند (فصل ۱۲). خودروهای ما مملو از این چنین سیستم‌های پایش هستند که به ما هشدار می‌دهند. به همین منوال، کاربردهای بسیاری در مراقبت بهداشت و سلامت وجود دارد (مانیتورینگ و تشخیص).

این سیستم‌ها در قلمرو هوش مصنوعی گسترده (قوی) ولی در حوزه‌های بسیار محدود هستند. نهایتاً، رایانه‌ها پیروز خواهند شد. ماشین‌ها به عنوان افراد خبره عمل کرده و قدرت مطلق در تصمیم‌گیری خواهند داشت. مشاورهای رباتیک محض، مثال‌هایی از این دستگاه‌ها هستند (فصل ۱۲). وسایل نقلیه‌ی مستقل و ربات‌های مستقل که می‌توانند خودشان را تعمیر کنند نیز مثال‌های خوبی هستند.

هوش مصنوعی افزوده اکثر کاربردهای فعلی هوش مصنوعی چیزی بین حالت همکار و مستقل هستند که آنها را هوش افزوده می‌نامند. این فناوری بر افزودن قابلیت‌های رایانه به قابلیت‌های شناختی انسان (فصل ۶ در مورد رایانش شناختی) و دستیابی به عملکرد بالاتر تمرکز دارد.

¹ Staff

فکر ترکیب کردن عملکرد انسان و ماشین‌ها چیز جدیدی نیست. ما در اینجا قابلیت‌های انسانی را با هوش قدرتمند ماشین ترکیب می‌کنیم (می‌افزاییم). این به معنای جایگزین کردن انسان‌ها با هوش مصنوعی مستقل نیست بلکه می‌خواهیم قابلیت‌های شناختی انسان را توسعه بدهیم. نتیجه‌ی اینکار، توانایی حل مسائل انسانی پیچیده همانند پرونده‌ی آغازین این فصل است. انسان‌ها به کمک رایانه‌ها می‌توانند مسائلی را حل کنند که تا پیش از این حل نشده بودند. پادمانابان^۱ (۲۰۱۸) بین هوش مصنوعی سنتی و افزوده تفاوت‌های زیر را بر شمرده است:

۱. ماشین‌های افزوده به جای اینکه جایگزین تصمیم‌گیری انسان شوند آن را توسعه و بسط می‌دهند و موجب تسهیل خلاقیت می‌شوند.
۲. افزونگی برخلاف هوش مصنوعی گسترده و قوی موجب پیشبرد حل مسائل صنعتی و انسانی پیچیده در حوزه‌های خاص می‌شود.
۳. برخلاف مدل «جعبه سیاه» در برخی تحیل‌ها و هوش مصنوعی، هوش افزوده یک نوع بینش و توصیه‌هایی را از جمله توضیحات فراهم می‌کند.
۴. علاوه بر این، فناوری افزونگی برخلاف هوش مصنوعی همکار با ترکیب اطلاعات فعلی و کشف شده می‌تواند راه‌حل‌های جدیدی را ارائه کند که به شناسایی مشکلات یا نشانه‌ها پرداخته و راه‌حل‌های از پیش تعیین شده را توصیه می‌کند.

پادمانابان (۲۰۱۸) و بسیاری دیگر معتقدند که در حال حاضر، هوش مصنوعی افزوده بهترین گزینه برای حرکت به سمت دگرگونی دنیای هوش مصنوعی است.

برعکس هوش مصنوعی مستقل که ماشین‌ها را به واسطه‌ی طیف گسترده‌ای از قابلیت‌های شناختی تعریف می‌کند (مثلاً وسائل نقلیه‌ی بدون راننده)، هوش افزوده فقط چند قابلیت شناختی محدود دارد.

استف (۲۰۱۷) مثال‌های زیر را آورده است:

- مبارزه با جرائم سایبری. هوش مصنوعی می‌تواند حملات را شناسایی کرده و راه‌حل پیشنهاد دهد.
- تصمیمات تجارت الکترونیک. ابزارهای بازاریابی می‌توانند تست‌ها را ۱۰۰ برابر سریع‌تر انجام دهند و کارکردهای چیدمان و پاسخ یک وب سایت را با کاربرانش تطبیق دهند. ماشین‌ها می‌توانند توصیه کرده و بازاریاب‌ها این توصیه‌ها را می‌توانند قبول یا رد کنند.
- معاملات بازار سهام بسامد بالا^۲. اینکار یا به طور کاملاً مستقل یا در برخی موارد با کنترل و کالیبراسیون توسط انسان انجام می‌شود.

سوالاتی برای بحث و گفتگو

۱. فرضیه‌ی اساسی هوش افزوده چیست؟
۲. تفاوت‌های مهم بین هوش افزوده و هوش مصنوعی سنتی را نام ببرید.
۳. برخی از مزایای هوش افزوده چه هستند؟
۴. این فناوری چه رابطه‌ای با رایانش شناختی دارد؟

اکثر صحبت‌ها در خصوص موضوعاتی از قبیل هوش مصنوعی و بهره‌وری، سرعت و کاهش هزینه است. در یک کنفرانس ملی به میزبانی شرکت معروف مشاوره فناوری اطلاعات، گارتنر، تقریباً نیمی از ۳،۰۰۰ مدیر ارشد فناوری آمریکایی شرکت کننده گزارش

¹ Padmanabhan

² High-Frequency Stock Market Trading

دادند که طرح‌هایی برای جاری‌سازی هوش مصنوعی دارند (ولدون ۲۰۱۸). صنعت نمی‌تواند این مزایای بالقوه‌ی هوش مصنوعی را نادیده بگیرد بویژه مزایای حاصل از افزایش بهره‌وری، کاهش هزینه، افزایش کیفیت و سرعت. شرکت‌کنندگان در کنفرانس در مورد استراتژی و پیاده‌سازی و اجرا صحبت کردند (فصل ۱۴). به نظر می‌رسد که هر شرکتی حداقل در فاز پایلوت و آزمایش هوش مصنوعی درگیر است. با وجود این، نباید تأثیر اجتماعی آن را در میان این هیجانات نادیده بگیریم. بسیاری از این تأثیرات، مثبت و برخی منفی و اکثر آنها ناشناخته هستند. در فصل ۱۴ بحث جامعی در این خصوص ارائه شده است که سه مثال از تأثیرات هوش مصنوعی را در اینجا آورده‌ایم.

تأثیر بر کشاورزی یک تأثیر مهم هوش مصنوعی بر کشاورزی خواهد بود. یک نتیجه‌ی مهم پیش‌بینی شده این است که غذای بیشتری بویژه در کشورهای جهان سوم فراهم شود. چند مثال به شرح زیر است:

- طبق لویز^۱ (۲۰۱۷)، استفاده از هوش مصنوعی و ربات‌ها می‌تواند به کشاورزان کمک کند تا ۷۰ درصد غذای بیشتری را تا سال ۲۰۵۰ تولید کنند. این افزایش، نتیجه‌ی بهره‌وری بالاتر تجهیزات کشاورزی که به واسطه‌ی اینترنت اشیا سرعت گرفته و کاهش هزینه‌ی تولید مواد غذایی است (امروزه فقط ۱۰ درصد از بودجه یک خانواده صرف مواد غذایی می‌شود در حالیکه این رقم در سال ۱۹۶۰ معادل ۱۷.۵ درصد بوده است).
- دید ماشینی به بهبود کاشت و برداشت کمک می‌کند. همچنین، هوش مصنوعی به برداشت مغز دانه‌های خوب کمک می‌کند.
- هوش مصنوعی به بهبود خواص غذایی کمک خواهد کرد.
- هوش مصنوعی هزینه‌ی تولید مواد غذایی را کاهش خواهد داد.
- تراکتورهای بدون راننده در مرحله‌ی آزمایش هستند.
- ربات‌ها می‌دانند چگونه میوه‌ها را چیده و سبزیجات را بکارند تا مشکل کمبود کارگران را در مزرعه جبران کنند.
- باردهی محصول به طور مستمر در هند و سایر کشورها در حال افزایش است.
- کنترل آفت بهبود می‌یابد. به عنوان مثال، هوش مصنوعی می‌تواند حملات آفت‌ها را پیش‌بینی کرده و برنامه‌ریزی برای مقابله با آنها را تسهیل نماید.
- شرایط آب و هوایی توسط ماهواره‌ها رصد می‌شوند. الگوریتم‌های هوش مصنوعی به کشاورزان می‌گویند که چه وقت کاشت یا برداشت کنند.

این فهرست می‌تواند همچنان ادامه داشته باشد. برای کشورهای نظیر هند و بنگلادش، این فعالیت‌ها زندگی کشاورزان را بهبود می‌بخشد. از همه اینها گذشته، هوش مصنوعی به کشاورزان کمک خواهد کرد تا تصمیمات بهتری بگیرند. در خصوص بنگلادش به PE Report (2017) مراجعه کنید. - alsonews.microsoft.com/en-in/features/ai-agriculture-icrisat-upl-india/

نکته: هوش مصنوعی می‌تواند به حیوانات خانگی گرسنه نیز کمک کند. یک دستگاه توزیع‌کننده آب و غذا به نام کتس‌پد^۲ به قیمت حدوداً ۴۷۰ دلار در انگلستان موجود است. شما فقط کافی است شماره شناسه را بر روی حیوان خانگی‌تان بگذارید (فقط گربه‌ها و سگ‌های کوچک). این دستگاه توزیع‌کننده می‌داند که کدام حیوان برای گرفتن غذا می‌آید و غذای مناسب را در حجم مناسب توزیع می‌کند. علاوه بر این، حسگرها (فصل ۱۳) می‌توانند به شما بگویند که هر حیوان چه میزان غذا خورده است.

¹ Lopez

² Catspad

همچنین اگر نیاز به افزودن آب باشد، دستگاه شما را مطلع می‌کند. علاقمند هستید؟ برای جزئیات بیشتر به دِئال^۱ (۲۰۱۸) رجوع کنید.

نقش سیستم‌های هوشمند بر مراقبت بهداشت و سلامت سیستم‌های هوشمند نقش مهمی در مراقبت سلامت و بهداشت ما ایفا می‌کنند. نوآوری‌های جدید تقریباً هر روزه در بخشی از این جهان (دولت‌ها، موسسات پژوهشی و پژوهش‌های هوش مصنوعی در حوزه پزشکی با حمایت سازمان‌ها) معرفی می‌شوند. برخی از نوآوری‌های جالب را در اینجا آورده‌ایم.

- هوش مصنوعی در پیش‌بینی بیماری‌ها پیشرفت کرده است (مثلاً پیش‌بینی وقوع بیماری‌های عفونی یک هفته جلوتر).
- هوش مصنوعی می‌تواند خونریزی‌های مغزی را تشخیص دهد.
- هوش مصنوعی می‌تواند مصرف دارو را پیگیری کرده، هشدارهای پزشکی را ارسال کرده، سفارش داروها را ارسال کرده و تطبیق دارو یا نسخه را بهبود بخشد.
- ربات‌های حضور از راه دور سیار^۲ می‌توانند پزشکان و بیماران را از راه دور به یکدیگر متصل کنند.
- ابررایانه‌ی تصویربرداری پزشکی انویدیا به متخصصان تشخیص بیماری کمک کرده و موجب تسهیل درمان بیماری‌ها می‌شود.
- رباتیک و هوش مصنوعی می‌توانند به طراحی مجدد زنجیره‌های تأمین دارویی کمک کنند.
- هوش مصنوعی می‌تواند ریسک‌های قلبی عروقی را از روی تصاویر شبکه‌ی چشم پیش‌بینی کند.
- یادگیری عمیق به پیش‌بینی سرطان کمک کرده و یادگیری ماشین نیز به تشخیص ملانوما کمک می‌کند.
- یک دستیار شخصی مجازی می‌تواند به ارزیابی خلق و خو و احساس بیمار بپردازد.
- پورتال‌های بسیاری هستند که اطلاعات پزشکی را به بیماران و حتی جراحان ارائه می‌کنند. یک مثال از آن، Adoptive spine IT است.
- مرکز پژوهش هوش مصنوعی در خصوص طول عمر بر روی افراد سالمند در ایالات متحده فعالیت می‌کند. فعالیت‌های مشابهی در ژاپن وجود دارد.
- استفاده از دست و پاهای بیونیک با هوش مصنوعی در حال بهبود است.
- اخبار فناوری اطلاعات مراقبت بهداشت و سلامت^۳ (۲۰۱۷) توضیح می‌دهد که چگونه هوش مصنوعی با استفاده از دستیارهای مجازی، مشکلات مراقبت بهداشت و سلامت را بهبود می‌بخشد (فصل ۱۲).

این فهرست همچنان ادامه دارد. نورمان^۴ (۲۰۱۸) سناریوی جایگزینی دکترها با سیستم‌های هوشمند را توضیح می‌دهد. نکته: هوش مصنوعی در پزشکی به عنوان یک رشته‌ی علمی شناخته شده که کنفرانس‌های سالانه ملی و بین‌المللی برای آن برگزار می‌شود. برای یک کتاب جامع در این خصوص به آگاه^۵ (۲۰۱۷) مراجعه کنید.

¹ Deahl

² Mobile Telepresence robots

³ Healthcare IT News

⁴ Norman

⁵ Agah

سایر کاربردهای اجتماعی هوش مصنوعی کاربردهای بسیاری در حمل و نقل، تأسیسات، آموزش، خدمات اجتماعی و سایر رشته‌ها دارد. برخی از این کاربردها تحت عنوان شهرهای هوشمند مطرح شده‌اند (فصل ۱۳). رسانه‌های اجتماعی و سایرین به منظور کنترل محتوا از جمله اخبار جعلی از هوش مصنوعی استفاده می‌کنند. نهایتاً، چگونه از این فناوری برای ریشه‌کن کردن بردگی کودکان در خاورمیانه استفاده کنیم؟ رجوع کنید به مورد کاربردی ۱.۹.

ربات‌ها برای منفعت اجتماعی، به عنوان آدمک‌ها در مسابقات شترسواری استفاده می‌شوند.

مورد کاربردی ۱.۹

در چندین کشور خاورمیانه، و به خصوص اردن، ابوظبی و سایر کشورهای حوزه خلیج فارس، مسابقات شترسواری برای نسل‌های متعددی بسیار طرفدار داشته است. صاحبان شترهایی که در مسابقات برنده می‌شوند می‌توانند از جوایز بسیار ارزشمندی بهره‌مند شوند (تا ۱,۰۰۰,۰۰۰ دلار برای مقام اول). همچنین، شترسواری یک رویداد فرهنگی و اجتماعی به شمار می‌رود.

مسئله

برای مدت زمان طولانی، شترها در این مسابقات توسط سوارکارها هدایت می‌شدند. هرچه وزن سوارکار کمتر بود، شانس برنده شدن نیز بیشتر بود. به همین خاطر، صاحبان این شترها، کودکان (حتی ۷ ساله) را برای اینکار آموزش می‌دادند. در سودان، هند، بنگلادش و سایر کشورهای فقیر، پسر بچه‌ها از خانواده‌های فقیر خریداری شده (یا حتی دزدیده شده) و به عنوان سوارکارها آموزش داده می‌شدند. در واقع، این روش برای نسل‌ها انجام می‌شد تا زمانیکه بین سال‌های ۲۰۰۵-۲۰۱۰ در تمامی کشورهای خاورمیانه ممنوع گردید. یکی از عوامل مهمی که منجر به ممنوعیت شد استفاده از ربات‌ها بود.

راه‌حل ربات‌ها

مسابقات شترسواری برای چندین نسل جزو آداب و رسوم محسوب شده و به یک ورزش پرمنفعت تبدیل شد. به همین خاطر هیچکس تمایل نداشت تا این ورزش متوقف شود. طبق آپفر (۲۰۱۶)، استفاده از ربات‌ها برای شترسواری یک دلیل بشردوستانه داشت و آن نجات کودکان بود. امروزه، در خاورمیانه فقط از ربات‌ها استفاده می‌شود. ربات‌ها را به پشت شترها می‌بندند که شبیه سوارکارهای کوچک به نظر می‌رسند و از راه دور از خودروهایی که به موازات مسیر حرکت شترها حرکت می‌کنند کنترل می‌شوند. صاحبان شترها می‌توانند شترها را صدا زده و به آنها فرمان بدهند و همچنین می‌توانند از یک شلاق مکانیکی برای زدن شترها و سریعتر دویدن آنها استفاده کنند. توجه داشته باشید که شترها اگر صدای انسان را نشنیده یا چیزی شبیه انسان را بر پشت خودشان نبینند نخواهند دوید.

فناوری

این دوربین ویدئویی وجود دارد که به کسانی که در داخل خودروها به موازات شترها حرکت می‌کنند وضعیت را نشان می‌دهند. صاحبان شترها می‌توانند از داخل خودرو به شترهایشان فرمان بدهند. یک شلاق مکانیکی که به پشت شتر وصل شده رامی‌توان از راه دور بکار گرفت.

نتایج

نتایج، حیرت‌انگیز هستند. نه تنها بردگی کودکان ریشه‌کن شد بلکه رکورد سرعت شترها نیز افزایش یافت. از همه اینها گذشته، ربات‌هایی که برای این مسابقات بکار گرفته می‌شدند فقط ۶ پاوند وزن داشتند و موجب خستگی شترها نمی‌شدند. برای تماشای این موضوع نگاه کنید به [youtube.com/watch?v=GVeVhWXB7sk](https://www.youtube.com/watch?v=GVeVhWXB7sk) (2:47 min). برای تماشای یک مسابقه‌ی کامل نگاه کنید به [youtube.com/watch?v=xFCRhk4GYds](https://www.youtube.com/watch?v=xFCRhk4GYds) (9:08 min.). ممکن است زمانیکه این فیلم را تماشا می‌کنید شانس اینکه اعضای خانواده سلطنتی را نیز ببینید داشته باشید. نهایتاً، برای جزئیات بیشتر می‌توانید به [youtube.com/watch?v=C1uYAXJibYg](https://www.youtube.com/watch?v=C1uYAXJibYg) (8:08 min.) مراجعه کنید.

Sources: Compiled from C. Chung. (2016, April 4). "Dubai Camel Race Ride-Along." [YouTube.com](https://www.youtube.com/watch?v=xFCRhk4GYds). [youtube.com/watch?v=xFCRhk4GYds](https://www.youtube.com/watch?v=xFCRhk4GYds) (accessed September 2018); P. Boddington. (2017, January 3). "Case Study: Robot Camel Jockeys. Yes, really." *Ethics for Artificial Intelligence*; and L. Slade. (2017, December 21). "Meet the Jordanian Camel Races Using Robot Jockeys." Sbs.com.au.

سوالاتی برای بحث و گفتگو

۱. گفته می‌شود که ربات‌ها بردگی کودکان را ریشه‌کن کرده‌اند. توضیح دهید.
۲. چرا صاحبان شترها باید در کنار شترهایشان سوار خودرو باشند؟
۳. چرا نمی‌توان این فناوری را در اسب‌سواری بکار برد؟
۴. جنبه‌های اخلاقی این مورد را جمع‌بندی کنید (بادینگتون^۱ ۲۰۱۷).

سوالاتی برای مرور بخش ۱.۷

۱. ویژگی‌های اصلی هوش مصنوعی چه هستند؟
۲. مزایای اصلی هوش مصنوعی را نام ببرید؟
۳. گروه‌های اصلی در اکوسیستم هوش مصنوعی چه هستند؟ محتوای اصلی هر یک را نام ببرید.
۴. چرا یادگیری ماشین اهمیت دارد؟
۵. تفاوت بین هوش مصنوعی محدود (ضعیف) و گسترده (قوی) را توضیح دهید.
۶. برخی معتقدند که هوش مصنوعی هیچ کاربرد گسترده (قوی) ندارد. چرا؟
۷. هوش همکار، هوش افزوده و هوش مستقل را تعریف کنید.
۸. تفاوت بین هوش مصنوعی سنتی و هوش افزوده در چیست؟
۹. ارتباط بین انواع هوش مصنوعی با رایانش شناختی را توضیح دهید.
۱۰. پنج کاربرد اصلی هوش مصنوعی برای افزایش عرضه‌ی مواد غذایی را نام ببرید.
۱۱. پنج نقش اصلی هوش مصنوعی در مراقبت‌های پزشکی را نام ببرید.

¹ Boddington

۱.۸ همگرایی تحلیل و هوش مصنوعی

تا به اینجا ما به معرفی تحلیل و هوش مصنوعی به صورت دو موجودیت مستقل و مجزا پرداخته‌ایم. اما همانطور که در پرونده‌ی آغازین دیدیم، این فناوری‌ها را می‌توانیم برای حل مسائل پیچیده ترکیب کنیم. در این بخش، ما به بحث و بررسی در خصوص همگرایی این تکنیک‌ها و اینکه چگونه یکدیگر را تکمیل می‌کنند می‌پردازیم. ما همچنین امکان افزودن فناوری‌های دیگر به ویژه اینترنت اشیا را توضیح خواهیم داد که راه‌حل‌هایی برای مسائل بسیار پیچیده ارائه می‌کنند.

همانگونه که از بخش ۱.۴ به یاد دارید، تحلیل‌ها داده‌های تاریخی را با استفاده از ابزارهای آماری، علوم مدیریت و سایر ابزارهای رایانشی و محاسباتی به منظور توضیح موقعیت‌ها (تحلیل توصیفی)، پیش‌بینی نتایج از جمله پیش‌بینی (تحلیل پیشگویانه) و توصیه‌ها برای راه‌حل مسائل (تحلیل پیشگویانه) پردازش می‌کنند. تأکید در اینجا بر علوم مدیریت آماری و سایر ابزارهای محاسباتی است که به تحلیل داده‌ها کمک می‌کنند.

از طرف دیگر، هوش مصنوعی از ابزارهای متفاوتی استفاده می‌کند ولی هدف اصلی آن تقلید از روش فکر کردن، یاد گرفتن، استدلال کردن، تصمیم گرفتن و حل مسائل توسط انسان است. تأکید در اینجا بر دانش و اطلاعات به عنوان ابزارهای اساسی برای حل مسئله به جای تکیه بر محاسبات است. همچنین، هوش مصنوعی نیز با رایانش شناختی سر و کار دارد. در واقعیت، این تفاوت‌ها چندان واضح و روشن نیست زیرا در کاربردهای پیشرفته‌ی تحلیل، موقعیت‌های استفاده از یادگیری ماشین (یک فناوری هوش مصنوعی)، پشتیبانی تحلیل هم در پیش‌بینی و هم تجویز وجود دارد. در این بخش، ما به توضیح همگرایی فناوری‌های هوشمند می‌پردازیم.

هم هوش مصنوعی و هم تحلیل‌ها و فناوری‌های متفاوت آنها زمانیکه به تنهایی بکار گرفته شوند مزایای خاص خودشان را برای بسیاری از سازمان‌ها به همراه دارند. ولی هرکدام از آنها محدودیت‌هایی دارند. طبق مطالعه‌ی گارتنر، شانس اینکه راهکارهای تحلیل کسب‌وکار نتوانند اهداف سازمانی را برآورده کنند بین ۷۰ تا ۸۰ درصد است. حداقل ۷۰ درصد از نیازهای سازمانی محقق نمی‌شوند. به عبارت دیگر، شانس کمی وجود دارد که ابتکار عمل‌های هوش تجاری بتوانند منجر به تعالی سازمانی شوند. چندین علل برای این وجود دارد:

- مدل‌های پیش‌گویانه دارای اثرات ناخواسته‌ای هستند (فصل ۱۴).
- مدل‌ها باید با رعایت اصول اخلاقی، مسئولیت‌پذیرانه و هوشیارانه بکار گرفته شوند.
- نتایج تحلیل‌ها ممکن است برای برخی کاربردهای خیلی خوب و برای برخی دیگر خوب نباشد.
- کیفیت مدل‌ها به کیفیت داده‌های ورودی و مفروضات بستگی دارد.
- داده‌ها ممکن است ناقص باشند. تغییرات محیطی می‌توانند موجب شوند تا داده‌ها به سرعت مهجور شوند. مدل‌ها ممکن است وفق‌پذیر نباشند.
- داده‌هایی که از انسان به دست آمده‌اند ممکن است دقیق نباشند.
- داده‌های گردآوری شده از منابع مختلف ممکن است شکل و کیفیت‌های متفاوتی داشته باشند.

دلایل دیگری که برای ادغام سیستم‌های هوشمند مطرح هستند برای تمامی پروژه‌های فناوری اطلاعات عمومیت دارند و در بخش ۱۴.۲ در مورد آنها صحبت خواهیم کرد.

نرخ شکست ابتکار عمل‌های هوش مصنوعی نیز بسیار بالاست. برخی از دلایل این امر مشابه نرخ شکست تحلیل‌ها می باشد. با این وجود، یک دلیل اصلی این است که برخی فناوری‌های هوش مصنوعی به داده‌های بسیار زیادی و برخی اوقات به کلان داده‌ها نیاز دارند. به عنوان مثال، روزانه میلیون‌ها میلیون داده وارد آکس‌می‌شود تا دانش آن را ارتقاء بخشد. بدون جریان مستمر داده‌ها، هیچگونه یادگیری مثبتی در هوش مصنوعی رخ نخواهد داد.

سوال این است که آیا هوش مصنوعی و تحلیل (و سایر سیستم‌های هوشمند) را می‌توان به گونه‌ای ترکیب کرد که موجب هم‌افزایی و دستیابی به نتایج بهتر شود یا خیر.

طبق ناداو^۱ (۲۰۱۷)، هوش تجاری و تحلیل‌های آنها می‌توانند به اکثر سوالات چ/و چه چیزی با نظر گرفتن کفایت و شایستگی حل مسئله پاسخ دهند. افزودن تحلیل‌های تجویزی، هزینه بیشتری به همراه خواهد داشت ولی لزوماً نتایج بهتری نخواهد داد. از این رو، نسل بعدی پلتفرم‌های هوش تجاری نیز از هوش مصنوعی برای موقعیت‌یابی خودکار، مصورسازی و شرح امور مهم استفاده خواهند کرد. این را می‌توان همچنین برای ایجاد هشدارها و اعلان‌های خودکار بکار برد. علاوه بر این، یادگیری ماشین و یادگیری عمیق می‌توانند از طریق شناخت الگوها و پیش بینی‌های دقیق‌تر از تحلیل‌ها پشتیبانی کنند. هوش مصنوعی به مقایسه‌ی عملکرد واقعی با عملکرد پیش‌بینی شده کمک می‌کند (بخش ۱۴.۶). یادگیری ماشین و سایر فناوری‌های هوش مصنوعی نیز استراتژی بهبود مستمر را فراهم می‌کنند. ناداو همچنین افزودن نظرات خبرگی از طریق هوش جمعی را توصیه کرده است (فصل ۱۱). در بخش باقیمانده ما به بررسی تفصیلی‌تر جنبه‌های همگرایی برخی سیستم‌های هوشمند خواهیم پرداخت.

ویژگی‌های کلان داده‌ها شامل حجم، تنوع و سرعت است که فراتر از دسترس معمول محیط‌های سخت‌افزاری و/یا قابلیت‌های ابزارهای نرم‌افزاری برای پردازش داده‌ها می‌باشد. با وجود این، امروزه روش‌ها و فناوری‌هایی وجود دارند که امکان اخذ، پاکسازی و تجزیه و تحلیل کلان داده‌ها را فراهم می‌کنند. این روش‌ها و فناوری‌ها شرکت‌ها را قادر می‌سازند تا تصمیمات لحظه‌ای بگیرند. همگرایی با هوش مصنوعی و یادگیری ماشین یک عامل مهم و اساسی در این راستا است. در دسترس بودن تحلیل‌های جدید کلان داده، قابلیت‌های جدیدی را در فناوری‌های هوش مصنوعی فراهم آورده که تا همین اواخر امکان‌پذیر نبودند. طبق گفته‌ی بین (۲۰۱۷)، کلان داده‌ها به دلیل زیر می‌توانند هوش مصنوعی را صاحب قدرت و اختیار کنند:

- قابلیت‌های جدید پردازش کلان داده‌ها با هزینه‌ی بسیار کمتر
- در دسترس بودن حجم انبوه داده‌ها به صورت آنلاین
- افزایش الگوریتم‌ها از جمله یادگیری عمیق که قابلیت‌های هوش مصنوعی قدرتمند را میسر کرده است.

¹ Nadav

مت‌لایف^۱ یک شبکه بین‌المللی بیمه واقع در کانادا است که به خاطر استفاده از فناوری اطلاعات به منظور روانسازی عملیات و افزایش رضایت مشتریان شناخته شده است. این شرکت برای اینکه بیشترین بهره‌برداری را از فناوری داشته باشد از هوش مصنوعی که با تحلیل‌های کلان‌داده درهم آمیخته است استفاده می‌کند:

- پیگیری حوادث و نتایج آنها به واسطه‌ی تشخیص گفتار بهبود یافته است.
- یادگیری ماشین، ورشکستگی‌های بلا تکلیف را نشان می‌دهد. علاوه بر این، گزارشاتی که دکترها در مورد زخمی‌ها یا بیماران نوشته‌اند و ادعای خسارت‌های پرداخت شده توسط شرکت بیمه در عرض چند ثانیه توسط سیستم تجزیه و تحلیل می‌شوند.
- سرعت بخشیدن به اجرای سیاست‌های بیمه‌گر در مورد اموال و بیمه تلفات با استفاده از هوش مصنوعی و تحلیل‌ها انجام می‌شود.
- جنبه‌ی پشت صحنه‌ی دفتر برای رسیدگی به ادعاها شامل حجم زیادی از داده‌های ساخت‌نیافته است که داخل ادعاها موجود می‌باشند. بخش تجزیه و تحلیل شامل داده‌های سلامت بیماران می‌باشد. یادگیری ماشین به منظور شناسایی بسیار سریع آنومالی‌ها در گزارشات استفاده می‌شود. برای اطلاعات بیشتر در مورد هوش مصنوعی و کسب‌وکار بیمه به فصل ۲ مراجعه کنید. برای اطلاعات بیشتر در مورد همگرایی کلان‌داده‌ها و هوش مصنوعی به طور کلی و در مت‌لایف به بین^۲ (۲۰۱۷) مراجعه کنید.

در پرونده‌ی آغازین دیدیم که چگونه فناوری‌های هوش مصنوعی زمانیکه با اینترنت اشیا ادغام شوند می‌توانند راه‌حلهایی برای مسائل پیچیده ارائه کنند. اینترنت اشیا، حجم زیادی از داده‌ها را از حسگرها و سایر «اشیا» گردآوری می‌کند. این داده‌ها برای تصمیم‌گیری باید پردازش شوند. در ادامه خواهیم دید که کورتانا مایکروسافت^۳ چگونه اینکار را انجام می‌دهد. باتنر^۴ (۲۰۱۸) توضیح می‌دهد که ترکیب هوش مصنوعی و اینترنت اشیا چگونه می‌تواند منجر به «راه‌حل‌ها و تجربیات سطح بالا» شود. تأکید در چنین ادغام‌هایی بر یادگیری بیشتر در مورد مشتریان و نیازهایشان است. این یکپارچه‌سازی همچنین می‌تواند موجب تسهیل تحلیل رقابتی و عملیات کسب‌وکار شود. جفت ترکیب شده‌ی هوش مصنوعی و اینترنت اشیا به ویژه زمانیکه با کلان‌داده‌ها ترکیب شوند می‌توانند به کشف محصولات جدید، فرایندهای کسب‌وکار و فرصت‌ها کمک کنند. از توان بالقوه‌ی اینترنت اشیا می‌توان در کنار فناوری‌های هوش مصنوعی بهره برد. علاوه بر این، تنها راه معنادار کردن داده‌های بدست آمده از «اشیا» از طریق اینترنت اشیا و دستیابی به یک بینش و ادراک از آنها این است که آنها را تحت تجزیه و تحلیل هوش مصنوعی قرار دهیم. فاگلا^۵ (۲۰۱۷) سه مثال زیر از ترکیب هوش مصنوعی و اینترنت اشیا را ارائه کرده است:

۱. ترموستات هوشمند آزمایشگاه‌های نست^۶ (نگاه کنید به خانه‌های هوشمند در فصل ۱۳).

1 MetLife

2 Bean

3 Microsoft's Cortana

4 Butner

5 Faggela

6 Nest Labs

۲. جاروبرقی‌های خودکار مثل رومبا آی‌ریبات^۱ (نگاه کنید به جاروبرقی‌های هوشمند فصل ۲).
۳. وسایل نقلیه‌ی خودران (فصل ۱۳).

اینترنت اشیاء می‌تواند در صورت ترکیب با آی بی ام واتسون آنالیتیکز که شامل یادگیری ماشین است بسیار هوشمند شود. مثال‌هایی از آن در پرونده‌ی آغازین و فصل ۱۳ ارائه شده است.

چندین کارشناس خبره امکان همگرایی هوش مصنوعی، تحلیل‌ها و زنجیره‌ی بلوکی^۲ را مطرح کرده‌اند (کوره‌آ،^۳ ۲۰۱۷ و کرانز^۴، ۲۰۱۷). در اینجا همگرایی ممکن است منجر به طراحی یا بازطراحی فناوری‌ها و پارادایم‌ها شود. فناوری زنجیره‌ی بلوکی می‌تواند به امنیت داده‌هایی که در یک شبکه توزیع شده بین تمامی طرف‌ها به اشتراک گذاشته شده‌اند و امکان ثبت داده‌های تراکنش وجود دارد بیفزاید. کرانز معتقد است همگرایی با زنجیره‌ی بلوکی می‌تواند راه‌حل‌های جدیدی را برای مسائل پیچیده به ارمغان بیاورد. اینگونه همگرایی‌ها باید شامل اینترنت اشیاء باشد. کارنز همچنین نقشی را برای رایانش مه^۵ در نظر گرفته است (فصل ۹). اینگونه ترکیب‌ها می‌توانند در کاربردهای پیچیده مثل وسایل نقلیه‌ی مستقل و آمازون گو^۶ (مورد کاربردی ۱.۱۰) بسیار مفید باشند.

آمازون گو آماده‌ی کسب‌وکار است

مورد کاربردی ۱.۱۰

در اوایل سال ۲۰۱۸، آمازون کام اولین خواربارفروشی تماماً خودکار خود را در مرکز شهر سیاتل افتتاح کرد. این شرکت در سال ۲۰۱۷ توانسته بود با این نوع فروشگاه موفقیت کسب کند که البته به طور آزمایشی و فقط با کارکنان شرکت انجام شده بود. خریداران وارد فروشگاه شده، محصولات را برداشته و به خانه می‌رفتند. هزینه خرید بعداً از حساب‌شان برداشته می‌شد. جالب به نظر می‌رسد! از صف انتظار برای بسته‌بندی کالاها و پرداخت پول آنها خبری نبود – نه صندوق‌داری و نه بحث و جدلی. از یک جهت، خریداران فرایندی مشابه خرید آنلاین را تجربه می‌کنند – محصولات/خدمات مورد نظرشان را پیدا کرده، آنها را خریداری کرده و منتظر می‌شوند تا هزینه‌ها به صورت الکترونیکی و ماهانه شارژ شوند.

فرایند خرید

شما برای شرکت در این فرایند باید یک اپلیکیشن رایگان ویژه روی گوشی هوشمندتان نصب کنید. سپس از طریق این اپلیکیشن به حساب آمازون کام وصل شوید. سپس اینکار را انجام دهید:

۱. اپلیکیشن را اجرا کنید.

¹ iRobot Roomba

² Blockchain

³ Coreia

⁴ Kranz

⁵ Fog Computing

⁶ Amazon Go

۲. گوشی هوشمند را به سمت درب فروشگاه تکان دهید. در آنجا با کد کیوآر کار خواهد کرد.

۳. وارد فروشگاه شوید.

۴. شروع به خرید کنید. تمامی محصولات، آماده شده‌اند. آنها را داخل سبد خرید بگذارید. لحظه‌ای که جنسی را از قفسه بر می‌دارید در داخل سبد خرید مجازی شما ثبت می‌شود. این فعالیت توسط حسگرها/دوربین‌ها انجام می‌شود. حساب شما به‌هکار می‌شود. اگر نظرتان عوض شد و جنس را به قفسه بازگردانید، سیستم نیز بلافاصله حساب شما را اصلاح می‌کند. حسگرها نیز جابجایی شما را در فروشگاه ردیابی می‌کنند. (این یک مسئله‌ی حریم خصوصی دیجیتال است. فصل ۱۴ بخش ۱۴.۳). حسگرها از نوع آراف‌آی‌دی هستند.

۵. خرید تمام شد؟ کافی است از فروشگاه بیرون بیایید (مطمئن شوید که اپلیکیشن باز است تا بتوانید از درب فروشگاه خارج شوید). سیستم از خروج شما و محصولات که خریداری کرده‌اید آگاه است و سفر خرید شما به پایان رسیده است. سیستم، کل هزینه شما را محاسبه می‌کند که هر زمان که بخواهید می‌توانید بر روی گوشی هوشمندتان مشاهده کنید.

۶. آمازون. کام عادات خرید شما را ثبت می‌کند (بازهم یک مسئله‌ی حریم خصوصی) که به شما در تجربه‌ی خرید آینده‌تان و به آمازون برای تهیه‌ی فهرست توصیه‌ها کمک می‌کند (فصل ۲). هدف «گو» این است که شما را به سمت مواد غذایی سالم راهنمایی کند (آمازون بسته‌های وعده‌های غذایی سالم خود را در این فروشگاه عرضه می‌کند).

نکته: امروزه تعداد بسیار معدودی در این فروشگاه کار می‌کنند. کارکنان فقط مسئولیت پر کردن قفسه‌ها و راهنمایی شما را بر عهده دارند. شرکت آمازون قصد دارد تا چند فروشگاه دیگر را در سال ۲۰۱۸ افتتاح نماید.

فناوری بکار رفته

آمازون برخی از فناوری‌های مورد استفاده خود را افشا کرده است که شامل الگوریتم‌های یادگیری عمیق، دید رایانه‌ای و ترکیب حسگرها می‌باشد. به سایر فناوری‌های مورد استفاده اشاره نشده است. نگاه کنید به [youtube.com/watch?v=NrmMk1Myrxc](https://www.youtube.com/watch?v=NrmMk1Myrxc) (1:50 min.).

Sources: Condensed for C. Jarrett. (2018). "Amazon Set to Open Doors on AI-Powered Grocery Store."

[venturebeat.com.venturebeat.com/2018/01/21/amazon-set-to-open-doorson-ai-powered-grocery-store/](https://venturebeat.com/2018/01/21/amazon-set-to-open-doorson-ai-powered-grocery-store/) (accessed September 2018);

D. Reisinger. (2018, February 22). "Here Are the Next Cities to Get Amazon Go Cashier-Less Stores." *Fortune*.

سوالاتی برای مورد کاربردی ۱.۱۰

۱. ویدئو را تماشا کنید. چه چیزی برایتان خوشایند و چه چیزی ناخوشایند بود؟

۲. این فرایند را با فرایند سلف‌چک که امروزه در بسیاری از سوپرمارکت‌ها و فروشگاه‌های بزرگ وجود دارد مقایسه کنید.

۳. این فروشگاه در مرکز شهر سیاتل افتتاح شد. چرا این مکان در مرکز شهر انتخاب شد؟

۴. مزایای این فروشگاه برای مشتریان چیست؟ برای آمازون چه مزایایی به همراه دارد؟

۵. آیا مشتریان آماده هستند تا در قبال راحتی، از حریم خصوصی خود بگذرند؟ بحث کنید.

برای گزارشی جامع در خصوص همگرایی فناوری‌های هوشمند به reportbuyer.com/product/5023639 مراجعه کنید. علاوه بر زنجیره‌ی بلوکی، همانطور که پیشتر توصیه شد می‌توان اینترنت اشیا و کلان‌داده‌ها و همچنین فناوری‌های هوشمند بیشتری را نیز افزود (مثلاً دید ماشینی، فناوری‌های صدا). اینها ممکن است موجب غنای کاربردها شوند. به طور کلی، هرچه از

فناوری‌های بیشتری استفاده شود، مسائل پیچیده‌تری را می‌توان حل نمود و کارایی عملکرد سیستم‌های همگرا افزایش می‌یابد (سرعت، دقت). برای بحث نگاه کنید به i-scoop.eu/convergence-ai-iot-big-data-analytics/.

شرکت‌های بسیاری هستند که ابزارها یا پلتفرم‌هایی را به منظور پشتیبانی از همگرایی سیستم‌های هوشمند ارائه می‌دهند. دو نمونه از آنها به شرح زیر می‌باشند:

آی‌بی‌ام آی‌بی‌ام برای پشتیبانی از همگرایی هوش مصنوعی و تحلیل‌ها دو پلتفرم خود را ترکیب نموده است. پاورِ آی‌بی‌ام یک پلتفرم توزیع برای هوش مصنوعی و یادگیری ماشینی است. این یک روش برای پشتیبانی از پلتفرم تحلیلی آی‌بی‌ام به نام دیتا ساینس اکسپیرینس^۱ (مجهز به فناوری ابر) می‌باشد. ترکیب این دو موجب بهبودهایی در فرایند تحلیل داده می‌شود. همچنین به دانشمندان داده کمک می‌کند تا آموزش مدل‌های پیچیده‌ی هوش مصنوعی و شبکه‌های عصبی را تسهیل نمایند. پژوهشگران می‌توانند از این سیستم ترکیبی برای پروژه‌های یادگیری عمیق استفاده کنند. از همه اینها گذشته، این ترکیب می‌تواند بینش و ادراک بهتری برای حل مسئله فراهم کند. برای جزئیات بیشتر نگاه کنید به (FinTech Futures (2017). همانطور که از پروژه‌ی آغازین به خاطر دارید، آی‌بی‌ام واتسون نیز هوش مصنوعی، اینترنت اشیا و تحلیل‌ها را در پروژه‌های ساختمانی شناختی ادغام نموده است.

مجموعه کورتانا مایکروسافت مایکروسافت، ترکیبی از تحلیل‌های پیشرفته، هوش تجاری سنتی و تحلیل‌های کلان‌داده خود را تحت پلتفرم ابری خود به نام آژور^۲ (فصل ۱۳) ارائه کرده است. این مجموعه به کاربران اجازه می‌دهد تا داده‌ها را به اقدامات هوشمند تبدیل کنند.

با استفاده از کورتانا می‌توان داده‌های بدست آمده از منابع متعدد از جمله حسگرهای اینترنت اشیا را تبدیل نموده و تحلیل‌های پیشرفته (مثلاً داده‌کاوی) و هوش مصنوعی (مثلاً یادگیری ماشینی) را بر روی آنها اعمال کرده و به توصیه‌های اقدام‌پذیر دست یافت که اینها به تصمیم‌گیرندگان، اپلیکیشن‌ها یا سیستم‌های کاملاً خودکار ارسال می‌شوند. برای جزئیات بیشتر در مورد این سیستم و معماری کورتانا به آدرس زیر مراجعه کنید:

mssqltips.com/sqlservertip/4360/introduction-to-microsoft-cortana-intelligence-suite/

سوالاتی برای مرور بخش ۱.۸

۱. مزایای اصلی همگرایی سیستم‌های هوشمند چه هستند؟
۲. دلیل عدم موفقیت بسیار بالای ابتکار عمل‌های تحلیل در گذشته چه بوده است؟
۳. از ترکیب تحلیل‌ها و هوش مصنوعی به چه نوع هم‌افزایی می‌توان دست یافت؟
۴. چرا تدارک و آماده‌سازی کلان‌داده‌ها برای ابتکار عمل‌های هوش مصنوعی ضروری است؟
۵. مزایای افزودن اینترنت اشیا به کاربردهای فناوری‌های هوشمند چه هستند؟

¹ Data Science Experience

² AZURE

۶. چرا توصیه شده است تا از زنجیره‌ی بلوکی برای پشتیبانی کاربردهای هوشمند استفاده شود؟

۱.۹ مروری بر اکوسیستم تحلیل‌ها

پس شما در خصوص توان بالقوه‌ی تحلیل‌ها، علم داده‌ها و هوش مصنوعی هیجان زده هستید و می‌خواهید به این صنعت در حال رشد بپیوندید. فعالان جاری این صنعت چه کسانی هستند و چه کار می‌کنند؟ جایگاه درست شما در این صنعت کجاست؟ هدف از این بخش، شناسایی بخش‌های مختلف صنعت تحلیل، ارائه‌ی یک طبقه‌بندی از انواع مختلف شرکت‌کنندگان در این صنعت و به تصویر کشیدن انواع فرصت‌های موجود برای کارشناسان تحلیل است. در یک اکوسیستم تحلیل، یازده نوع فعال مختلف شناسایی شده است. خوانندگان با درک این اکوسیستم می‌توانند به دید وسیع‌تری از اینکه فعالان مختلف چگونه کنار یکدیگر قرار می‌گیرند دست پیدا کنند. هدف دوم از درک این اکوسیستم تحلیل برای یک کارشناس این است که از سازمان‌ها و فرصت‌ها و محصولات جدید در بخش‌های تحلیل آگاه باشد.

اگرچه برخی پژوهشگران بین کارشناسان تحلیل کسب‌وکار و دانشمندان داده تمایز قائل شده‌اند (دَوِنپورت و پاتیل^۱، ۲۰۱۲)، همانگونه که قبلاً گفته شد، به منظور درک کلی اکوسیستم تحلیل ما این دو را یکسان در نظر می‌گیریم. مشخصاً، نیازهای مهارتی برای یک ریاضی‌دان، یک برنامه‌نویس، یک مدلساز تا یک ارتباط دهنده متغیر است و ما معتقدیم که این مسئله باید در یک سطح خُرد/فردی حل و فصل شود. ما همچنین عام‌ترین تعریف تحلیل‌ها را در نظر گرفته‌ایم تا هر سه نوع تحلیل ارائه شده توسط اینفورمز یعنی توصیفی/ گزارش‌گیری/مصورسازی، پیش‌گویانه و تجویزی را شامل گردد. ما همچنین هوش مصنوعی را در همین گروه در نظر گرفته‌ایم.

شکل ۱.۱۷ یک منظر از اکوسیستم تحلیل را نشان می‌دهد. مؤلفه‌های این اکوسیستم به شکل گلبُرگ‌ها نشان داده شده‌اند. یازده بخش یا خوشه‌ی اصلی در فضای تحلیل شناسایی شده است. مؤلفه‌های اکوسیستم تحلیل در سه گروه دسته‌بندی شده‌اند که توسط گلبُرگ‌های درونی، گلبُرگ‌های بیرونی و دانه (بخش میانی) گل نشان داده شده‌اند. شش گلبُرگ بیرونی را می‌توان به طور عمومی، ارائه‌دهندگان فناوری^۲ نامید. درآمد اصلی این گروه از ارائه‌ی فناوری، راه‌حل و آموزش به سازمان‌های استفاده کننده از تحلیل‌ها است تا بتوانند این فناوری‌ها را به اثربخش‌ترین و کاراترین شکل بکار بگیرند. گلبُرگ‌های درونی را به طور کلی می‌توان شتاب‌دهندگان تحلیل‌ها^۳ نامید. آنها هم با ارائه‌دهندگان فناوری و هم کاربران کار می‌کنند. نهایتاً، هسته‌ی اکوسیستم که متشکل از سازمان‌های کاربر تحلیل‌ها^۴ است. این مهم‌ترین مؤلفه‌ی اکوسیستم است زیرا سازمان‌های کاربر، عامل محرک هر خوشه‌ی صنعت تحلیل به شمار می‌روند.

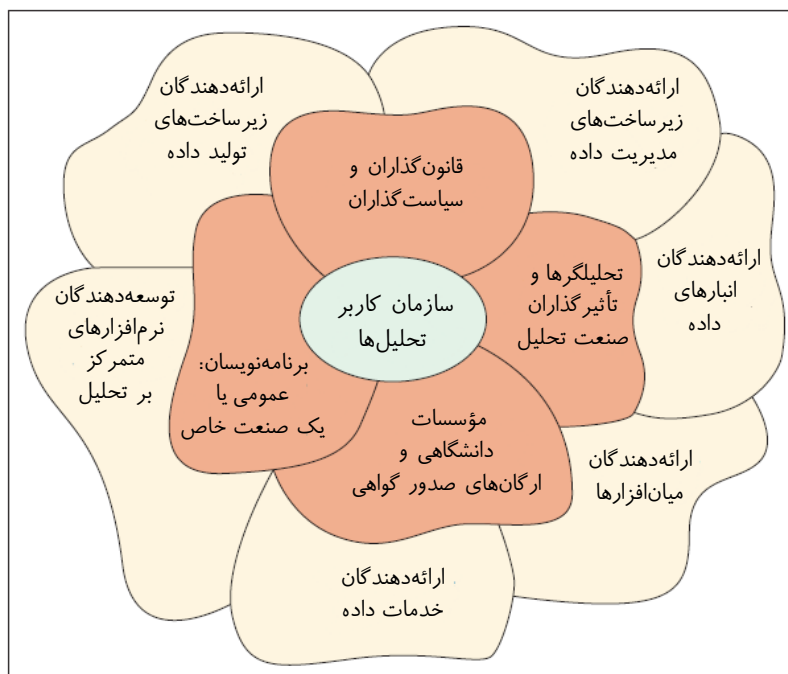
استعاره‌ی گل به خوبی برانزده‌ی اکوسیستم تحلیل است زیرا مؤلفه‌های متعدد با یکدیگر همپوشانی دارند. تمامی این گلبُرگ‌ها همانند یک ارگانیسم زنده با یکدیگر رشد کرده و پژمرده می‌شوند. شرکت‌های بسیاری هستند که در چندین بخش صنعت تحلیل فعالیت کرده و فرصت‌هایی برای جابجایی افقی و عمودی در این حوزه فراهم می‌کنند.

¹ Davenport & Patil

² Technology Providers

³ Analytics accelerators

⁴ Analytics user organizations



شکل ۱.۱۷ اکوسیستم تحلیل

برای جزئیات بیشتر در خصوص اکوسیستم تحلیل می‌توانید به نسخه خلاصه‌تر کتاب ما (شاردا، دِلِن و توربان^۱، ۲۰۱۷) و همچنین شاردا و کال‌گوترا^۲ (۲۰۱۸) رجوع کنید. مَت تارک^۳ یک سرمایه‌دار جسور در شرکت فرست‌مارک^۴ یک اکوسیستم تحلیل را با تمرکز بر کلان‌داده‌ها توسعه داده و به‌روزروری می‌نماید. هدف او ردیابی فعالان جدید و فعلی در بخش‌های مختلف صنعت کلان‌داده است. یک تصویر دیداری بسیار زیاد از تفسیر وی از اکوسیستم و یک فهرست کامل و جامع از شرکت‌ها در وب‌سایت شخصی وی به آدرس <http://mattturck.com/2016/02/01/big-data-landscape/> (accessed September 2018) در دسترس است.

۱.۱۰ ساختار کتاب

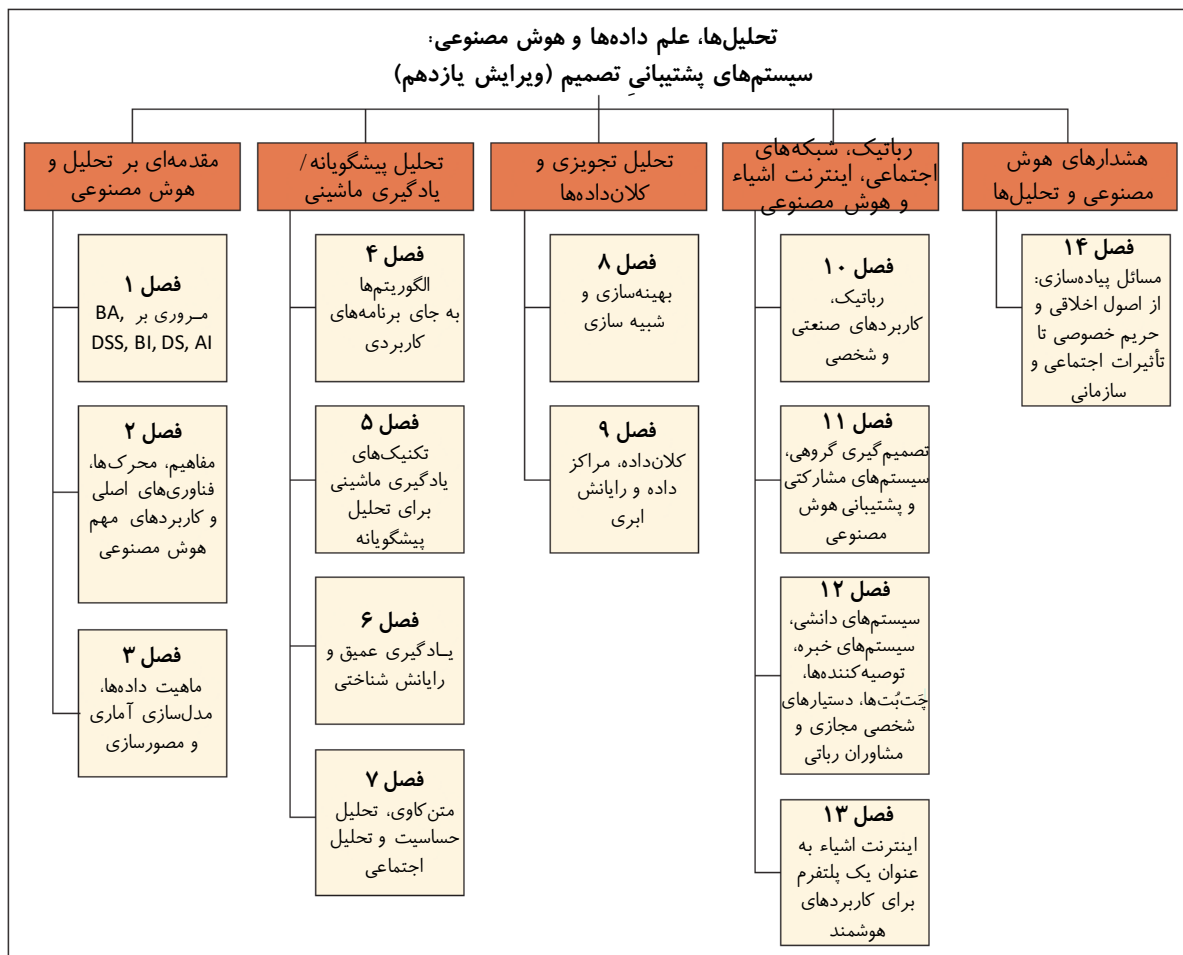
در بخش‌های قبلی به درک و فهمی از نیاز به فناوری اطلاعات در تصمیم‌گیری، رشد و تکامل هوش تجاری، تحلیل، علم داده‌ها و هوش مصنوعی دست یافتیم. در چندین بخش گذشته نیز مروری بر انواع مختلف تحلیل‌ها و کاربردهایشان داشته‌ایم. اکنون آماده هستیم تا گشت و گذار مدیریتی مفصل‌تری در این موضوعات و به همراه تجارب عمیق در برخی موضوعات فنی داشته باشیم. در شکل ۱.۱۸ ساختار ادامه‌ی کتاب نشان داده شده است.

¹ Sharda, Delen, Turban

² Sharda, Kalgotra

³ Matt Turk

⁴ FirstMark



شکل ۱.۱۸ ساختار کتاب

در این فصل، ما به ارائه یک مقدمه، تعاریف و مروری بر سیستم پشتیبانی تصمیم، هوش تجاری و تحلیل، از جمله تحلیل‌های کلان‌داده‌ها و علم داده‌ها پرداختیم. ما همچنین مروری بر اکوسیستم تحلیل داشتیم تا شما به عمق و گستردگی این صنعت پی ببرید. فصل‌های ۲ و ۳ به بررسی تحلیل‌های توصیفی و مسائل داده‌ها می‌پردازند. داده‌ها مشخصاً پایه و زیربنای هر کاربرد تحلیل را شکل می‌دهند. بنابراین، ما مقدمه‌ای را بر مسائل انبارش، کاربردها و فناوری‌های داده‌ها اختصاص می‌دهیم. این فصل همچنین کاربردها و فناوری‌های مصورسازی و گزارش‌گیری کسب‌وکار را پوشش می‌دهد.

ما فصل جاری را با معرفی عمیق‌تری از هوش مصنوعی در فصل ۲ ادامه می‌دهیم. از آنجائیکه داده‌ها، زیربنای هر تجزیه و تحلیلی هستند، فصل ۳ به معرفی مسائل داده‌ها و همچنین تحلیل‌های توصیفی از جمله مفاهیم آماری و مصورسازی می‌پردازد. در یک فصل آنلاین به بررسی عمیق‌تر فرایندها و مبنای انبارش برای علاقمندان پرداخته‌ایم. بخش بعدی کتاب، تحلیل‌های پیشگویانه و یادگیری ماشینی را پوشش می‌دهد. در فصل ۴ به معرفی کاربردهای داده‌کاوی و فرایند داده‌کاوی می‌پردازیم. فصل ۵ به معرفی تعداد زیادی از تکنیک‌های متداول داده‌کاوی می‌پردازد: دسته‌بندی، خوشه‌بندی، رابطه‌کاوی و غیره. فصل ۶ شامل موضوعات یادگیری عمیق و رایانش شناختی می‌باشد. فصل ۷ بر کاربردهای متن‌کاوی و همچنین تحلیل وب از جمله تحلیل رسانه‌های اجتماعی و سایر موضوعات مرتبط می‌پردازد. بخش بعدی نگاه عمیق‌تری به موضوع «علم داده‌ها» دارد. فصل ۸

تحلیل تجویزی را پوشش می‌دهد. فصل ۹ شامل جزئیات بیشتری از تحلیل کلان‌داده‌ها است و همچنین مقدمه‌ای بر تحلیل مبتنی بر ابر و تحلیل مکان را شامل می‌شود. در بخش بعدی به موضوع رباتیک، شبکه‌های اجتماعی، هوش مصنوعی و اینترنت اشیا پرداخته شده است. فصل ۱۰ به معرفی ربات‌ها در کاربردهای کسب‌وکار و شخصی پرداخته و در خصوص تأثیر آتی اینگونه دستگاه‌ها بر جامعه بحث می‌کند. فصل ۱۱ بر سیستم‌های مشارکتی، جمع‌سپاری و شبکه‌های اجتماعی متمرکز است. فصل ۱۲ مروری بر دستیارهای شخصی، چت‌بات‌ها و دستاوردهای هیجان‌انگیز در این فضا دارد. فصل ۱۳ به مطالعه‌ی اینترنت اشیا و توان بالقوه‌ی آن در پشتیبانی تصمیم و جامعه‌ی هوشمندتر می‌پردازد. حضور همه جانبه‌ی دستگاه‌های جی پی اس و بی‌سیم و سایر حسگرها منجر به خلق پایگاه‌های داده جدید و انبوه و کاربردهای همه جا حاضر می‌شود. نسل جدیدی از شرکت‌های تحلیل در حال ظهور هستند که هدف آنها تجزیه و تحلیل این پایگاه‌های داده جدید و دستیابی به درک بهتر و عمیق‌تری از رفتارها و حرکات مشتریان است. این منجر به اتوماسیون تحلیل‌ها شده و حوزه‌ی جدیدی را به نام «اینترنت اشیا» بوجود آورده است. در پایان، فصل ۱۴ کار را با یک بحث کوتاه در خصوص امنیت، حریم خصوصی و ابعاد اجتماعی تحلیل/هوش مصنوعی به پایان می‌برد.

۱.۱۱ منابع، پیوندها و ارتباط شبکه‌ی دانشگاه ترادیتا

استفاده از ابزارهای ارائه شده در بخش‌های بعدی می‌تواند به درک و بهره‌برداری بیشتر از این فصل و اکثر فصل‌های این کتاب کمک کند.

ما منابع و پیوندهای مهم سازمانی زیر را توصیه می‌کنیم:

- مؤسسه‌ی انبارش داده (tdwi.org)
- مرکز علوم داده (datasciencecentral.com)
- منابع سیستم پشتیبانی تصمیم (dssresources.com)
- کنسرسیوم سازمانی مایکروسافت (enterprise.waltoncollege.uark.edu/mec/asp)

اکثر فروشندگان، نسخه‌های آزمایشی نرم‌افزاری محصولات و برنامه‌هایشان را عرضه می‌کنند. اطلاعات مربوط به این محصولات، معماری و نرم‌افزار در آدرس dssresources.com موجود است.

ما مطالعه‌ی نشریه‌های دوره‌ای زیر را توصیه می‌کنیم:

- سیستم‌های پشتیبانی تصمیم^۱ (www.journals.elsevier.com/decision-support-systems)

¹ Decision Support Systems

این کتاب کاملاً با منابع رایگان ارائه شده توسط دانشگاه ترادیتا مرتبط است (www.teradatauniversitynetwork.com). پورتال تی‌یوان به دو بخش اصلی تقسیم شده است: یک بخش برای دانشجویان و بخش دیگر برای اعضای هیئت علمی. این کتاب از طریق یک بخش ویژه در انتهای هر فصل به پورتال تی‌یوان متصل می‌شود. آن بخش شامل پیوندهای مناسب برای هر فصل است که به منابع مرتبط اشاره دارد. علاوه بر این، ما با استفاده از نرم‌افزار و سایر محتوای موجود در تی‌یوان تمرین‌های دست اولی را ارائه می‌کنیم.

وب سایت این کتاب به آدرس pearsonhighered.com/sharda حاوی محتوای متنی مکمل است که در قالب فصل‌های وبی که متناظر با فصل‌های کتاب می‌باشند سازماندهی شده‌اند. موضوعات این فصل‌ها در فهرست عناوین فصل آنلاین آورده شده است.

با ارسال این کتاب برای نشر، ما فعال و معتبر بودن تمامی وبسایت‌های اشاره شده در متن را صحت‌گذاری کردیم. با وجود این، یوآرال‌ها پویا هستند. وبسایت‌هایی که ما در متن به آنها ارجاع داده‌ایم برخی اوقات به دلیل تغییر نام شرکت‌ها یا خرید و فروش یا ادغام شرکت‌ها تغییر کرده یا خارج از دسترس می‌شوند. برخی اوقات، وبسایت‌ها برای تعمیرات، نگهداری یا طراحی مجدد غیرفعال شده‌اند. بسیاری از سازمان‌ها دیگر از “www” در اول آدرس سایت‌شان استفاده نمی‌کنند ولی برخی همچنان از آن استفاده می‌کنند. اگر شما برای ورود به وبسایتی که ما اشاره کرده‌ایم مشکل دارید لطفاً صبور باشید و یک جستجو در وب انجام دهید تا بتوانید وب سایت جدید احتمالی را پیدا کنید. در اکثر مواقع، شما سریعاً می‌تواند وب سایت جدید را از طریق یکی از موتورهای جستجوی معروف بیابید. ما پیشاپیش به خاطر زحمت و ناراحتی پیش آمده پوزش می‌خواهیم.

- پیچیدگی محیط کسب‌وکار افزایش یافته و به سرعت در حال تغییر است که تصمیم‌گیری را دشوارتر می‌سازد.
- کسب‌وکارها باید به سرعت با اتخاذ تصمیمات سریع و بهتر به این محیط متغیر پاسخ داده و خود را با آن وفق دهند.
- یک مدل، نمایش ساده‌سازی شده یا انتزاعی از واقعیت است.
- تصمیم‌گیری از چهار فاز اصلی تشکیل شده است: اطلاعات، طراحی، انتخاب و پیاده‌سازی.
- در فاز اطلاعات، مسئله (فرصت) شناسایی، طبقه‌بندی و تجزیه می‌شود (در صورت نیاز) و مالکیت مسئله مشخص می‌شود.
- در فاز طراحی، یک مدل از سیستم ساخته شده، معیارهای گزینش مورد توافق قرار گرفته، جایگزین‌ها تولید شده، نتایج پیش‌بینی شده و یک متدولوژی تصمیم ایجاد می‌شود.

- در فاز انتخاب، جایگزین‌ها مقایسه شده و جستجو برای یافتن بهترین جواب آغاز می‌شود. تکنیک‌های جستجوی زیادی وجود دارد.
- در پیاده‌سازی و اجرای جایگزین‌ها، یک تصمیم‌گیرنده باید اهداف متعدد و مسائل تحلیل حساسیت را در نظر بگیرد.
- بازه‌ی زمانی برای تصمیم‌گیری در حال محدود شدن است در حالیکه ماهیت جهانی تصمیم‌گیری در حال گسترش بوده و توسعه و استفاده از سیستم پشتیبانی تصمیم رایانه‌ای را ایجاب می‌کند. یک چارچوب اولیه برای پشتیبانی تصمیم، موقعیت‌های تصمیم را بسته به درجه‌ی ساختارمند بودن و فعالیت‌های مدیریتی به ۹ دسته تقسیم می‌کند. هر دسته به طور متفاوتی پشتیبانی می‌شوند.
- تصمیمات تکرارپذیر و ساخت‌یافته توسط روش‌های تحلیل کمی استاندارد مثل ام‌اس، ام‌آی‌اس و پشتیبانی تصمیم خودکار مبتنی بر قواعد پشتیبانی می‌شوند.
- سیستم‌های پشتیبانی تصمیم از داده‌ها، مدل‌ها و برخی اوقات مدیریت دانش برای یافتن راه‌حل‌ها برای مسائل نیمه‌ساخت‌یافته و ساخت‌نیافته استفاده می‌کنند.
- مؤلفه‌های اصلی یک سیستم پشتیبانی تصمیم عبارتند از پایگاه داده و مدیریت آن، یک مدل مینا و مدیریت آن و یک رابط کاربرپسند. همچنین می‌توان یک مؤلفه‌ی هوشمند (دانش‌بنیان) را در این سیستم گنجانند. کاربر نیز به عنوان یک مؤلفه‌ی سیستم پشتیبانی تصمیم در نظر گرفته می‌شود.
- روش‌های هوش تجاری از یک مخزن مرکزی به نام انبار داده استفاده می‌کنند که امکان داده‌کاوی، پردازش تحلیلی آنلاین، مدیریت فرایند کسب‌وکار و مصورسازی داده‌ها را به شکلی اثربخش فراهم می‌کند.
- معماری هوش تجاری شامل یک انبار داده، ابزارهای تحلیل کسب‌وکار مورد استفاده‌ی کاربران نهایی و یک واسط کاربری (مثل داشبورد) هستند.
- بسیاری از سازمان‌ها از تحلیل‌های توصیفی به جای ابزارهای گزارش‌گیری سنتی خود استفاده می‌کنند تا بتوانند به بینش‌ها، روندها و الگوها در داده‌های تراکنش‌ها دست یابند.
- تحلیل‌های پیشگویانه، سازمان‌ها را قادر می‌سازند تا قواعد پیشگویانه را که محرک نتایج کسب‌وکار هستند از طریق تحلیل داده‌های تاریخی رفتارهای فعلی مشتریان برقرار کنند.
- تحلیل‌های تجویزی به ساخت مدل‌هایی که شامل تکنیک‌های پیش‌بینی و بهینه‌سازی مبتنی بر اصول پژوهش عملیاتی و علم مدیریت هستند کمک می‌کنند تا سازمان‌ها بتوانند تصمیمات بهتری را اتخاذ کنند.
- تحلیل کلان‌داده‌ها بر مجموعه داده‌های بزرگ و ساخت‌نیافته تمرکز دارند که ممکن است شامل انواع بسیار متفاوت و گوناگونی از داده‌ها برای تحلیل نیز باشد.
- تحلیل‌ها به عنوان یک رشته در صنایع و حوزه‌های خاص نیز کاربرد دارند مثل تحلیل ورزشی که با نام‌های دیگری از جمله علم داده‌ها یا علم شبکه نیز شناخته می‌شوند.
- مراقبت بهداشت و سلامت و زنجیره‌های خرده‌فروشی، دو حوزه کاربرد تحلیل‌ها هستند که در آینده شاهد کاربردهای بیشتری در آنها خواهیم بود.
- تحلیل تصویر، رشته‌ای است که به سرعت در حال رشد است که به کاربردهای بسیار زیادی از یادگیری عمیق نفوذ کرده است.

- اکوسیستم تحلیل را می‌توان در وهله‌ی اول به صورت مجموعه‌ای از ارائه‌دهندگان، کاربران و تسهیل‌گران در نظر گرفت که به ۱۱ بخش یا دسته تقسیم می‌شود.

پردازش تحلیلی آنلایین (OLAP)	داشبورد	تحلیل
پردازش تراکنش آنلایین (OLTP)	داده‌کاو	اکوسیستم تحلیل
تحلیل پیشگویانه	تحلیل هنجاری یا تصمیم	هوش مصنوعی
تحلیل تجویزی	تحلیل توصیفی (گزارش‌گیری)	هوش افزوده
	فاز طراحی	تحلیل کلان‌داده‌ها
	فاز پیاده‌سازی	هوش تجاری
	فاز اطلاعات	فاز انتخاب

۱.	۲.	۳.	۴.	۵.	۶.	۷.	۸.	۹.	۱۰.
----	----	----	----	----	----	----	----	----	-----

	teradatauniversitynetwork.com	.1
	Explore the Sports Analytics	.2
	"Cases, Projects, and Assignments"	.3
	"Harrah's High Payoff from Customer Information"	
"Data Warehousing	teradatauniversitynetwork.com	.4
	Supports Corporate Strategy at First American Corporation"	
	http://analytics-magazine.org/issues/digitaleditions	.5
"Predictive Analytics	"Special Issue: The Future of Healthcare"	
	– Saving Lives and Lowering Medical Bills"	
	"FICO Medication Adherence Score "	
	FICO Medication Adherence Score HoH	
	FICO Medication Adherence Score	
	http://analytics-magazine.org/issues/digitaleditions	.6
	"Big Data, Analytics and Elections"	
	"Work Social"	

۲

هوش مصنوعی

مفاهیم، محرک‌ها، فناوری‌های مهم و کاربردهای کسب‌وکار

هوش مصنوعی که برای نسل‌های متمادی به صورت یک کنجکاو بوده به سرعت در حال تبدیل شدن به یک فناوری کاربردی مهم با کاربردهای بسیار زیاد در رشته‌های گوناگون است. مأموریت اپن‌ای‌آی^۱ (یک مؤسسه پژوهشی هوش مصنوعی در فصل ۱۴) بیان می‌کند که هوش مصنوعی مهمترین فناوری ساخته شده به دست بشر خواهد بود. هوش مصنوعی به شکل‌های گوناگونی پدیدار شده و تعاریف متعددی دارد. به بیان خیلی ساده، هدف هوش مصنوعی این است که تا حد امکان کاری کند تا ماشین‌ها هوشی نزدیک به انسان و به نفع انسان‌ها از خود نشان دهند. آخرین دستاوردها در فناوری‌های رایانشی، هوش مصنوعی را به سطوح و دستاوردهای جدیدی سوق داده‌اند. به عنوان مثال، راهنمای مخارج آی‌دی‌سی^۲ (۲۲ مارس ۲۰۱۸) پیش‌بینی

کرده است که مخارج جهانی هوش مصنوعی در سال ۲۰۱۸ به ۱۹.۱ میلیارد دلار خواهد رسید. همچنین پیش‌بینی کرده است که مخارج سالانه در آینده‌ی نزدیک به رشد دو رقمی برسد. طبق شارما (۲۰۱۷)، انتظار می‌رود که چین با مخارج ۶۰ میلیارد دلار در سال ۲۰۲۵ به رهبر جهانی در هوش مصنوعی تبدیل شود. در خصوص ارزش کسب‌وکار هوش مصنوعی به گریگ^۳ (۲۰۱۸) نگاه کنید.

در این فصل، ما به معرفی الزامات هوش مصنوعی، فناوری‌های مهم آن، پشتیبانی آن برای تصمیم‌گیری و یک نمونه از کاربردهای آن در حوزه‌های مهم و کارکردی کسب‌وکار می‌پردازیم. بخش‌های مختلف این فصل به شرح زیر است:

- ۲/۱. پرونده‌ی آغازین: این‌ریکس مشکلات حمل و نقل را حمل می‌کند.
- ۲/۲. مقدمه‌ای بر هوش مصنوعی
- ۲/۳. هوش انسانی و رایانه‌ای
- ۲/۴. فناوری‌های مهم هوش مصنوعی و برخی مشتقات آن
- ۲/۵. پشتیبانی هوش مصنوعی از تصمیم‌گیری
- ۲/۶. کاربردهای هوش مصنوعی در حسابداری
- ۲/۷. کاربردهای هوش مصنوعی در خدمات مالی
- ۲/۸. هوش مصنوعی در مدیریت منابع انسانی
- ۲/۹. هوش مصنوعی در بازاریابی، تبلیغات و مدیریت ارتباط با مشتریان

¹ OpenAI

² IDC Spending Guide

³ Greig

ازدحام ترافیک یکی از مشکلات فزاینده در بسیاری از بخش‌های کلانشهرهای بزرگ است. رانندگان ممکن است هر روزه چندین ساعت را در ترافیک بمانند. علاوه بر این، آلودگی هوا افزایش می‌یابد و تصادفات بیشتری اتفاق می‌افتد.

شرکت این‌ریکس^۱ (inrix.com) رانندگان را قادر می‌سازد تا اطلاعات ترافیک را به صورت لحظه‌ای و آنی دریافت کنند. آنها می‌توانند اپلیکیشن این‌ریکس-اکس دی^۲ را برای پلتفرم‌های آی اِس و اندروید دانلود کنند. اطلاعات ارائه شده از طریق یک تجزیه و تحلیل پیشگویانه بر روی انبوهی از داده‌های مشتریان و محیط تولید شده است (مثلاً تصادفات، راه‌سازی). منابع اطلاعاتی شامل موارد زیر هستند:

- داده‌های ترافیک که توسط چرخ‌بال‌ها، پهپادها و غیره گردآوری شده‌اند و شامل جریان آنی و لحظه‌ای ترافیک و اطلاعات تصادفات می‌باشند.
- اطلاعات ارائه شده توسط شرکت‌های حمل و نقل شرکت‌کننده و بالغ بر ۱۰۰ میلیون راننده‌ی دواطلب ناشناس که گوشی‌های هوشمند منجر به جی‌پی‌اس داشته و همگی در لحظه گزارش می‌دهند.
- اطلاعات ارائه شده توسط گزارشات ازدحام ترافیک (مثلاً تأخیرهای ناشی از تعمیرات راه‌ها)

این‌ریکس اطلاعات گردآوری شده را توسط ابزارهای تحلیلی و فرمول‌های اختصاصی که برخی از آنها مبتنی بر هوش مصنوعی هستند پردازش می‌کند. از اطلاعات پردازش شده به منظور تهیه‌ی پیش‌بینی ترافیک استفاده می‌شود. به عنوان مثال، تصویری از جریان‌های ترافیکی و تأخیرهای پیش‌بینی شده برای ۱۵ الی ۲۰ دقیقه، چند ساعت و چندین روز آینده برای بسیاری از مکان‌ها تهیه می‌کند. این پیش‌بینی‌ها رانندگان را قادر می‌سازد تا برای بهترین مسیرها برنامه‌ریزی کنند. در سال ۲۰۱۸، شرکت این‌ریکس در ۴۵ کشور جهان و بسیاری از شهرهای بزرگ خدمت‌رسانی کرده و به تجزیه و تحلیل اطلاعات ترافیکی بیش از ۱۰۰ منبع پرداخته است. این خدمت با نقشه‌های دیجیتالی ادغام شده است. به عنوان مثال در سیاتل، اطلاعات ترافیک از طریق گوشی‌های هوشمند و کدهای رنگی بر روی بیلبوردهای کنار بزرگراه‌ها منتشر می‌شود. گوشی‌های هوشمند همچنین زمان‌های تخمینی

¹ INRIX

² INRIX-XD

ترافیک یا باز بودن راه‌ها را نشان می‌دهند. در سال ۲۰۱۸، این شرکت بیش از ۵,۰۰۰,۰۰۰ مایل از بزرگراه‌ها را در سرتاسر جهان پوشش داده و بهترین مسیرهای پیشنهادی را بر اساس درخواست و به صورت بلادرنگ ارائه داده است. سیستم این‌ریکس، اطلاعات را برای تصمیماتی از قبیل موارد زیر فراهم می‌کند:

- بهترین مسیرها برای وسایل نقلیه‌ی حمل بار و سایر مسافران
- بهترین زمان برای رفتن به سر کار یا محلی دیگر از یک مکان مشخص
- اطلاعات برای مسیردهی مجدد یک سفر به منظور اجتناب از ترافیک که به تازگی ایجاد شده است.
- پرداخت عوارض بزرگراه‌ها که بر اساس شرایط ترافیک و زمان روز محاسبه می‌شود.

فناوری‌های بکار رفته برای گردآوری داده‌ها عبارتند از:

- دوربین‌های تلویزیونی مداربسته و رادارها که شرایط ترافیکی را پایش می‌کنند.
- گزارشات ایمنی عمومی و اطلاعات ترافیک
- اطلاعات مربوط به دسترسی به بزرگراه‌ها و خروجی شهرها
- فناوری‌هایی که صف‌های پرداخت عوارض را می‌سنجند.
- ردیاب‌های حسگر مغناطیسی که در زیر سطح جاده‌ها کار گذاشته شده‌اند.
- گوشی‌های هوشمند و سایر دستگاه‌های گردآوری داده‌ها که داده‌ها را از این‌ریکس اخذ می‌کنند.

این اطلاعات توسط چندین تکنیک هوش مصنوعی از قبیل سیستم‌های خبره و مدل‌های تحلیلی مختلف (مثل شبیه‌سازی) پردازش می‌شوند.

چندین منبع اطلاعاتی از طریق اینترنت اشیاء به شرکت متصل هستند (فصل ۱۳). طبق وبسایت شرکت این‌ریکس، این شرکت به منظور پخش مستقیم و بلادرنگ داده‌های ترافیکی به وسایل نقلیه با کلپیر چنل رادیو^۱ مشارکت نموده و این داده‌ها را از طریق سیستم‌های ناوبری پرتابل، رسانه‌های پخش رادیویی، خدمات بی‌سیم و اینترنتی یا Ln Carr منتشر می‌کند. شبکه ترافیک کلپیر چنل^۲ برای بیش از ۱۲۵ کلانشهر در ۴ کشور در دسترس می‌باشد (inrix.com/press-release/2654). در سال ۲۰۱۸، این سیستم بر روی بیش از ۲۷۵ میلیون خودرو و دستگاه گردآوری داده نصب شد. این سیستم، اطلاعات لحظه‌ای ترافیک را از دستگاه‌ها جمع‌آوری می‌کند.

اطلاعات پردازش شده علاوه بر اینکه رانندگان توسط سازمان‌ها و برنامه‌ریزان شهری نیز جهت اتخاذ تصمیمات برنامه‌ریزی مورد استفاده قرار می‌گیرند. همچنین، ازدحام ترافیک کمتری در شهرهایی شرکت‌کننده ثبت شده که منجر به آلودگی کمتر، تصادفات جاده‌ای کمتر و بهره‌وری بیشتر به لحاظ کارکنان شادتری است که زمان کمتری را در ترافیک می‌گذرانند.

¹ Clear Channel Radio

² Clear Channel Traffic Network

اپلیکیشن ترافیک این‌ریکس (برای دانلود به inrix.com/mobile-app مراجعه کنید) سازگار با تمامی گوشی‌های همراه است و از ۱۰ زبان از جمله انگلیسی، فرانسه و اسپانیایی پشتیبانی می‌کند. برای آگاهی از ویژگی‌های رایگان ترافیکی این‌ریکس به سایت inrixtraffic.com/features مراجعه کنید. برای مطالعات موردی جالب به inrix.com/case-studies مراجعه کنید. در سال ۲۰۱۶، این‌ریکس یک نسخه‌ی بهبود یافته از اپلیکیشن ترافیک خود را منتشر کرد که هم از هوش مصنوعی و هم جمع‌سپاری (فصل ۱۱) برای پشتیبانی تصمیمات رانندگان برای انتخاب بهترین مسیر استفاده می‌کند (کوروسک ۲۰۱۶). فناوری هوش مصنوعی به تجزیه و تحلیل سوابق فعالیت رانندگان به منظور استنتاج فعالیت‌های آینده آنها می‌پردازد. نکته: اپلیکیشن‌های معروف گوشی همراه از قبیل ویزا^۱ و موویت^۲ امکانات راهبردی و گردآوری داده‌های مشابه این‌ریکس را عرضه می‌کنند.

Sources: Based on inrix.com, Gitlin (2016), Korosec (2016), and inrix.com/mobile-apps (all accessed June 2018).

سوالاتی برای مرور پرونده‌ی آغازین

۱. توضیح دهید چرا علیرغم بالا بودن ازدحام، ترافیک ممکن است کم باشد (inrix.com/uk-highway-agency)
۲. ارتباط این مورد با پشتیبانی تصمیم به چه شکل است؟
۳. عناصر هوش مصنوعی در این سیستم را شناسایی کنید؟
۴. دستاوردهای مرتبط با هوش مصنوعی را با مطالعه‌ی جدیدترین شماره‌های اعلامیه‌های رسمی شرکت طی چهار ماه گذشته از inrix.com/press-release شناسایی کنید. یک گزارش بنویسید.
۵. طبق جیتلین^۳ (۲۰۱۶)، اپلیکیشن ترافیک موبایل جدید این‌ریکس، تهدیدی برای ویز محسوب می‌شود. دلیل آن را توضیح دهید.
۶. به سایت sitezeus.com/data/inrix رفته و رابطه‌ی بین این‌ریکس و زئوس^۴ را توضیح دهید. ویدئوی ۲ دقیقه‌ای را در sitezeus.com/data/inrix تماشا کرده و چرا سیستم نامبرده در این ویدئو را "کمک‌کننده برای تصمیم"^۵ می‌نامند؟

مطالعه‌ی موردی این‌ریکس به ما نشان می‌دهد که چگونه گردآوری و تجزیه و تحلیل حجم بسیار زیادی از اطلاعات (کلان‌داده‌ها) می‌تواند موجب بهبود جابجایی و حرکت وسایل نقلیه در شهرهای بزرگ شود. مشخصاً، این‌ریکس با گردآوری اطلاعات از رانندگان و سایر منابع به جای اینکه فقط به حسگرهای گرانشیمت اتکا کند، جابجایی در شهرها را بهینه ساخته است. این مهم با پشتیبانی تصمیماتی که توسط رانندگان گرفته شده و با تحلیل جریان ترافیک محقق شده است. این‌ریکس همچنین از کاربردهای مختلف اینترنت اشیا برای اتصال وسایل نقلیه و دستگاه‌ها به سیستم رایانشی خود استفاده کرده است. این کاربرد یکی از اجزای سازنده

¹ Waze

² Moovit

³ Gitlin

⁴ Zeus

⁵ Decision Helper

و زیربنایی شهرهای هوشمند است (فصل ۱۳). از الگوریتم‌های قدرتمند که برخی از آنها برنامه‌های کاربردی هوش مصنوعی هستند برای تجزیه و تحلیل داده‌های گردآوری شده استفاده می‌شود.

همه ما دوست داریم تا تصمیم‌گیری رایانه‌ای را به شکلی ببینیم که ساده‌تر بوده، استفاده از آن راحت‌تر شده و شهودی‌تر باشد. علاوه بر این، در طول زمان تلاش‌هایی به منظور ساده‌سازی و خودکار سازی فعالیت‌های متعدد در فرایند تصمیم‌گیری به عمل آمده است. فقط به این فکر کنید که روزی یخچال‌ها بتوانند محتوای داخل‌شان را اندازه‌گیری و ارزیابی نموده و برای کالاهایی که در حال اتمام هستند سفارش‌گذاری کنند. تحقق چنین چیزی خیلی دور نیست و هوش مصنوعی از آن پشتیبانی خواهد کرد.

بینش مدیر ارشد اطلاعات پیش‌بینی کرده است که تا سال ۲۰۳۵، فناوری‌های رایانه‌ای هوشمند منجر به خلق ارزش اقتصادی ۵-۸.۳ تریلیون دلاری می‌شود

(نگاه کنید به cioinsight.com/blogs/how-ai-will-impactthe-global-economy.html). از میان فناوری‌هایی که به عنوان فناوری‌های هوشمند شناخته شده‌اند در این کتاب ما اینترنت اشیاء، رباتیک پیشرفته و وسائل نقلیه خودران را توضیح داده‌ایم. گارتنر یکی از شرکت‌های پیشگام مشاوره در حوزه فناوری، موارد زیر را در فهرست چرخه‌ی محبوبیت فناوری‌های نوظهور سال ۲۰۱۶ و ۲۰۱۷ آورده است: مشاوران خبره، سوال و پاسخ زبان طبیعی، پهنادهای تجاری، فضای کارهای هوشمند، پلتفرم‌های اینترنت اشیاء، کشف هوشمند داده‌ها، هوش ماشینی همه منظوره و دستیارهای شخصی مجازی. اکثر این موارد در این کتاب شرح داده شده یا ذکر شده‌اند (گرینگارد ۲۰۱۶)^۱. برای تاریخچه‌ی هوش مصنوعی نگاه کنید به زارکاداکیس^۲ (۲۰۱۶) و en.wikipedia.org/wiki/History_of_artificial_intelligence.

هوش مصنوعی چندین تعریف دارد (مار^۳ ۲۰۱۸)؛ با وجود این؛ بسیاری از خبرگان معتقدند که هوش مصنوعی با دو مفهوم اصلی سر و کار دارد: (۱) مطالعه‌ی فرایندهای ذهنی انسان (درک اینکه هوش به چه چیزی گفته می‌شود) و (۲) نمایش و تکثیر آن فرایندهای ذهنی و فکری در ماشین (رایانه‌ها، ربات‌ها). این بدان معناست که انتظار می‌رود که ماشین‌ها فرایندهای فکری و ذهنی همچون انسان داشته باشند.

یکی از تعاریف بسیار پذیرفته شده از هوش مصنوعی «قابلیت‌های ماشین برای تقلید از رفتار هوشمند انسانی» است (فرهنگ *واژگان مریام-وبستر*). پیش‌زمینه‌ی تاریخی هوش مصنوعی مبتنی بر منطق است که در چندین نوآوری علوم رایانه‌ای نیز بکار رفته است. در خصوص ارتباط بین هوش مصنوعی و منطق نگاه کنید به plato.stanford.edu/entries/logic-ai. یکی از کاربردهای اولیه‌ی شناخته شده برای هوش مصنوعی، برنامه شطرنج ابررایانه‌ی آی‌بی‌ام به نام (دیپ بلو^۴) بود. این سیستم موفق شد تا استاد قهرمان شطرنج معروف جهان گری کاسپاروف را شکست دهد.

¹ Greengard

² Zarkadakis

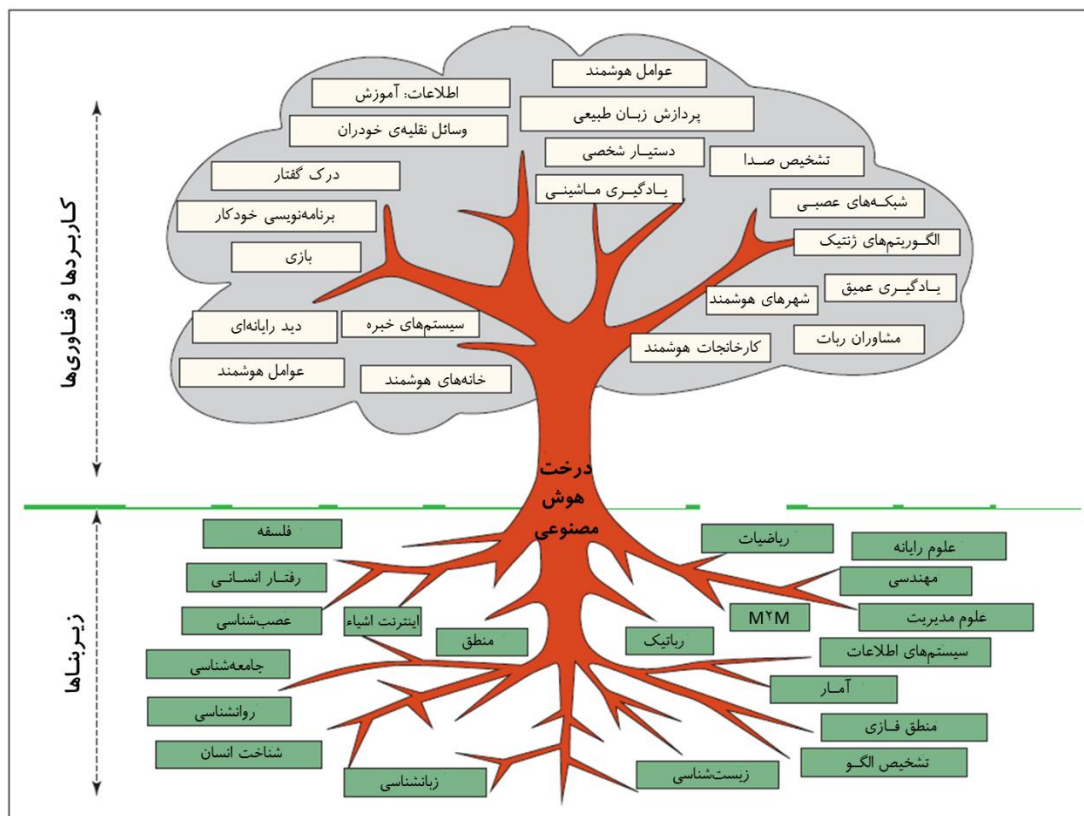
³ Mar

⁴ Deep Blue

هوش مصنوعی یک عبارت چتری از تکنیک‌های متعددی است که مشخصه‌ها و قابلیت‌های مشابه مشترکی دارند. برهی فهرست ۵۰ فناوری منحصر بفرد هوش مصنوعی به استیفی (۲۰۱۷) رجوع کنید. برای ۳۳ نوع فناوری هوش مصنوعی نگاه کنید به www.simpllicable.com/news/types-of-artificial-intelligence.

روند فزاینده‌ای برای «هوشمندتر» کردن رایانه‌ها وجود دارد. به عنوان مثال، وی ۳۰ به دنبال ایجاد سیستم‌های رایانه‌ای است که نسبت به وب ۲۰ سطح هوشمندی بسیار بالاتری را نشان دهند. تا کنون چندین کاربرد بر اساس تکنیک‌های چندگانه هوش مصنوعی ارائه شده است. به عنوان مثال، حوزه‌ی ترجمه‌ی ماشینی زبان‌ها به انسان‌ها کمک می‌کند تا به منظور همکاری و همچنین خرید آنلاین محصولات که به زبان‌های دیگری تبلیغ می‌شوند به زبان‌های متفاوت با هم صحبت کنند. به همین منوال، ترجمه‌ی ماشینی می‌تواند به انسان‌هایی که فقط زبان مادری خود را صحبت می‌کنند کمک کند تا با افرادی که به زبان‌های دیگری صحبت می‌کنند ارتباط برقرار کنند و تصمیمات مشترکی را اتخاذ کنند.

همانطور که در فصل ۱ توضیح داده شد، هوش مصنوعی از چشم‌انداز گسترده‌ای برخوردار است که صدها یا هزاران مؤلفه را شامل می‌شود. ما در شکل ۲.۱ زیربنا و فناوری‌های اصلی آن را به تصویر کشیده‌ایم. توجه داشته باشید که ما آنها را به دو گروه تقسیم کرده‌ایم: «زیربناها» و «فناوری‌ها و کاربردها». فناوری‌های اصلی بعداً در این فصل و در سرتاسر کتاب توضیح داده خواهد شد.



شکل ۲.۱ کارکردها و کاربردهای هوش مصنوعی

فناوری‌های هوش مصنوعی برای خلق کاربردهای بسیار زیادی استفاده می‌شوند. در بخش‌های ۲.۶ الی ۲.۱۰ ما مثال‌هایی از کاربردها را در حوزه‌های مهم کارکردی کسب‌وکار ارائه خواهیم کرد.

کاربردهای هوشمند شامل آنهایی هستند که می‌توانند به ماشین‌ها کمک کنند تا به سوالاتی که مشتریان به زبان طبیعی خود می‌پرسند پاسخ دهند. یک حوزه دیگر از کاربردها، سیستم‌های دانش‌بنیان هستند که می‌توانند به مردم مشاوره داده و در تصمیم‌گیری آنها را یاری نمایند یا حتی خودشان تصمیم‌گیری کنند. به عنوان مثال، اینگونه سیستم‌ها می‌توانند درخواست خرید مشتریان برای خرید آنلاین را تأیید یا رد کنند (اگر خریداران از قبل تأیید نشده‌اند یا خط اعتباری آزاد ندارند). مثال‌های دیگر شامل تولید خودکار سفارشات خرید آنلاین و ترتیب رسیدگی به سفارشات است که به صورت آنلاین ارسال شده‌اند. هم‌گوش هم فیس‌بوک پروژه‌هایی را در فاز آزمایش دارند که سعی می‌کنند نحوه یادگیری و پشتیبانی یا حتی تصمیم‌گیری مستقل را به ماشین‌ها بیاموزند. در خصوص کاربردهای هوشمند در سازمان‌ها به داج^۱ (۲۰۱۶)، فین‌لی^۲ (۲۰۱۷)، مک‌فرسون^۳ (۲۰۱۷) و راینهارتز^۴ (۲۰۱۷) مراجعه کنید. در خصوص نحوه بکارگیری راه‌حل‌های هوش مصنوعی به منظور تسهیل خدمات دولتی به برنداستودیدو (۲۰۱۷) نگاه کنید.

سیستم‌های مبتنی بر هوش مصنوعی همچنین برای نوآوری بسیار اهمیت دارند و مربوط به حوزه‌های تحلیل‌ها و پردازش کلان‌داده‌ها می‌باشند. یکی از پیشرفته‌ترین پروژه‌ها در این حوزه، تحلیل واتسون آی‌بی‌ام است (فصل ۶). برای پوشش جامع موضوع هوش مصنوعی از جمله تعاریف و تاریخچه و آینده‌ی آن نگاه کنید به کاپلان (۲۰۱۶).

نکته: در ژانویه ۲۰۱۶، مارک زوکربرگ مدیرعامل شرکت فیس‌بوک اعلام کرد که هدف این شرکت برای سال ۲۰۱۶، ساختن یک دستیار مبتنی بر هوش مصنوعی برای کمک به فعالیت‌های شخصی و کسب‌وکار و تصمیمات خودش است. زوکربرگ به یک ماشین آموزش می‌داد تا صدای او را متوجه شده و دستورات اصلی را پیروی کرده و چهره‌ی دوستان و شرکای کسب‌وکارش را تشخیص دهد. دستیارهای شخصی امروزه توسط میلیون‌ها نفر استفاده می‌شوند (فصل ۱۲).

پتینی باوز^۵ یک شرکت بین‌المللی ارائه‌دهنده راه‌حل در ایالات متحده است که در حوزه‌هایی از قبیل ارسال کالا، هوش مکانی، مشارکت دادن مشتری و مدیریت اطلاعات مشتری فعالیت دارد. این شرکت، نیروی محرکه‌ی سالانه میلیاردها تراکنش فیزیکی و دیجیتالی در سرتاسر جهان متصل و بدون مرز تجارت است.

امروزه، در پیتنی باوز، هزینه‌ی ارسال به طور خودکار بر اساس ابعاد، وزن و بسته‌بندی هر مرسوله تعیین می‌شود. این محاسبات رایگان، داده‌هایی را تولید می‌کنند که وارد الگوریتم‌های هوش مصنوعی می‌شود. هرچه داده‌های بیشتری پردازش شوند، دقت

¹ Dodge

² Finlay

³ McPherson

⁴ Reinharz

⁵ Pitney Bowes

محاسبات بیشتر خواهد بود (یکی از ویژگی‌های یادگیری ماشینی). این شرکت تخمین زده است که بکارگیری این الگوریتم‌ها یک بهبود ۲۵ درصدی را در محاسبات به همراه داشته است. این سیستم به پیتنی باوز یک مبنای دقیق برای قیمت‌گذاری، افزایش رضایت مشتری و ارتقاء مزیت رقابتی می‌دهد.

هدف کلی هوش مصنوعی، خلق ماشین‌های هوشمندی است که قادر به اجرای وظایف مختلفی باشند که در حال حاضر توسط انسان انجام می‌شود. به لحاظ ایده‌آل، ماشین‌های هوش مصنوعی می‌بایست قادر به استدلال، تفکر انتزاعی، برنامه‌ریزی، حل مسائل و یادگیری باشند.

برخی از اهداف خاص هوش مصنوعی عبارتند از:

- ادراک و واکنش مناسب به تغییرات محیط که بر فرایندها و عملیات خاص کسب‌وکار تأثیر می‌گذارند.
- ایجاد خلاقیت در فرایندهای کسب‌وکار و تصمیم‌گیری

کاربرد هوش مصنوعی تحت تأثیر عوامل محرک زیر بوده است:

- علاقه‌ی مردم به ماشین‌های هوشمند و مغزهای مصنوعی
- هزینه‌ی پایین کاربردهای هوش مصنوعی در برابر هزینه‌ی بالای نیروی کار دستی (برای انجام کار یکسان)
- تمایل شرکت‌های بزرگ فناوری به دستیابی به مزیت رقابتی و سهم بازار هوش مصنوعی و تمایل آنها به سرمایه‌گذاری میلیاردی در زمینه‌ی هوش مصنوعی
- فشار مدیریت به افزایش بهره‌وری و سرعت
- در دسترس بودن داده‌های با کیفیت که منجر به پیشرفت هوش مصنوعی شده است
- افزایش کارکردها و کاهش هزینه‌های رایانه‌ها به طور کلی
- توسعه‌ی فناوری‌های جدید به‌ویژه رایانش ابری

مزایای اصلی هوش مصنوعی به شرح زیر می‌باشند:

- هوش مصنوعی توانایی تکمیل بسیار سریع‌تر وظایف مشخص را نسبت به انسان دارد.
- سازگاری کارهای تکمیل شده با هوش مصنوعی به مراتب بهتر از انسان‌ها است.
- ماشین‌های هوش مصنوعی خطا نمی‌کنند.
- سیستم‌های هوش مصنوعی امکان انجام پروژه‌های بهبود مستمر را فراهم می‌کنند.
- هوش مصنوعی را به واسطه‌ی توانایی تشخیص الگو می‌توان برای تحلیل پیشگویانه بکار برد.
- هوش مصنوعی می‌تواند تأخیرها و موانع را در فرایندهای کسب‌وکار برطرف نماید.
- ماشین‌های هوش مصنوعی متوقف نمی‌شوند، استراحت نمی‌کنند و نمی‌خوابند.
- ماشین‌های هوش مصنوعی می‌توانند به طور مستقل از انسان یا به کمک انسان کار کنند.
- کارکردهای ماشین‌های هوش مصنوعی مرتباً در حال افزایش هستند.

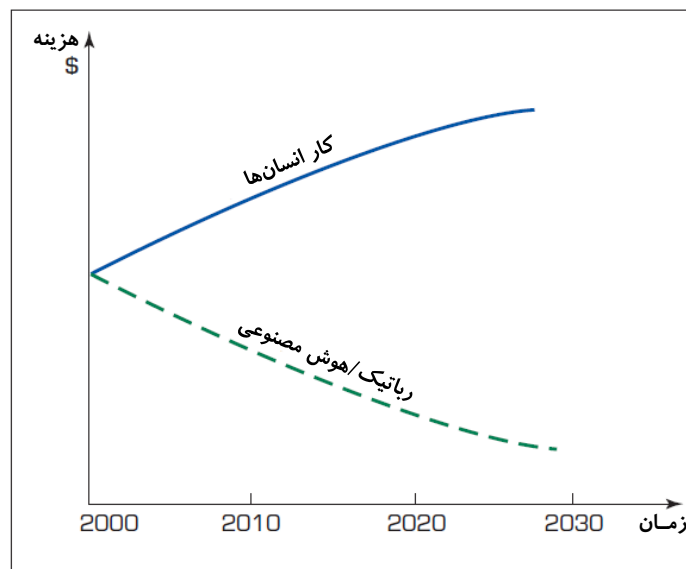
- ماشین‌های هوش مصنوعی می‌توانند در محیط‌هایی که برای انسان‌ها مخاطره‌آمیز هستند کار کنند.
- ماشین‌های هوش مصنوعی می‌توانند نوآوری‌های انسانی را تسهیل نمایند (یعنی پشتیبانی از تحقیق و توسعه)
- هیچگونه موانع عاطفی با کار هوش مصنوعی تداخل پیدا نمی‌کند.
- هوش مصنوعی در تشخیص کلاهبرداری و امنیت سرآمد است.
- هوش مصنوعی موجب بهبود عملیات صنعتی می‌شود.
- هوش مصنوعی کارهای دانشی را بهبود می‌بخشد.
- هوش مصنوعی موجب افزایش سرعت و مقیاس می‌شود.
- هوش مصنوعی به یکپارچه‌سازی و ادغام عملیات کسب‌وکار کمک می‌کند.
- کاربردهای هوش مصنوعی می‌توانند ریسک را کاهش دهند.
- هوش مصنوعی می‌تواند وقت کارکنان را آزاد کند تا بر روی مشاغل پیچیده‌تر و بهره‌ور کار کنند.
- هوش مصنوعی می‌تواند مراقبت از مشتری را بهبود بخشد.
- هوش مصنوعی می‌تواند مسائل دشواری را که قبلاً لاینحل بودند حل کند (کارپال ۲۰۱۷).^۱
- هوش مصنوعی موجب افزایش همکاری و سرعت‌دهی به یادگیری می‌شود.

این مزایا آنگونه که آگراوال^۲ (۲۰۱۸) گزارش داده موجب تسهیل مزیت‌های رقابتی می‌شوند. نکته: تمامی سیستم‌های هوش مصنوعی کلیه این مزایا را به همراه ندارند. سیستم‌های خاص ممکن است فقط برخی از این مزایا را ارائه کنند. قابلیت کاهش هزینه و افزایش بهره‌وری ممکن است منجر به افزایش زیاد سود شود (ویولینو^۳ ۲۰۱۷). علاوه بر منفعت‌رسانی به تک تک شرکت‌ها، هوش مصنوعی می‌تواند موجب افزایش چشمگیر رشد اقتصادی یک کشور شود مثل سنگاپور.

¹ Kharpal

² Agrawal

³ Violino



شکل ۲.۲ هزینه کار انسانی و هزینه کار هوش مصنوعی

مثال‌هایی از مزایای هوش مصنوعی در اینجا مزایای متداول هوش مصنوعی در حوزه‌های کاربردی مختلف ارائه شده است:

۱. انجمن بین‌المللی سواب و مشتقات^۱ از هوش مصنوعی برای حذف فعالیت‌های خسته‌کننده و کسل‌کننده در رویه‌های قراردادها استفاده می‌کند. به عنوان مثال، با استفاده از فناوری تشخیص کاراکتر چشمی به همراه هوش مصنوعی، قراردادها را به صورت دیجیتالی درآورده و سپس آنها را بایگانی می‌کند.
۲. هوش مصنوعی شروع به انقلاب در مشاغل استخدام نموده است (۱) از طریق انجام موثرتر و عادلانه‌تر غربال‌گری نامزدها، (۲) تطبیق‌های بهتر نامزدها با شغل‌ها و (۳) کمک به حفظ استعدادهای آینده‌ی سازمان‌ها. برای جزئیات بیشتر نگاه کنید به (SMBWorld Asia Editors (2017) و بخش ۲.۸.
۳. هوش مصنوعی در حال بازتعریف مدیریت است. طبق کولب‌یورنسرود و همکاران^۲ (۲۰۱۶)، استفاده از هوش مصنوعی منجر به پنج روش زیر می‌شود:

- می‌تواند وظایف اداری و اجرایی روتین را انجام دهد.
- مدیران می‌توانند بر بخشی از کارها که نیازمند قضاوت است تمرکز کنند.
- ماشین‌های هوشمند را به چشم همکاران می‌بینند (یعنی مدیران به توصیه‌ها و نصایحی که هوش مصنوعی ارائه می‌کند اعتماد دارند). علاوه بر این، همکاری بین انسان-ماشین وجود دارد (فصل ۱۱).
- مدیران بر قابلیت‌های خلاقانه‌ای تمرکز دارند که به وسیله‌ی هوش مصنوعی می‌توان از آنها پشتیبانی کرد.
- مدیران در حال توسعه‌ی مهارت‌های اجتماعی هستند که برای همکاری، رهبری و مربی‌گری بهتر ضروری هستند.

¹ International Swaps and Derivatives Association (ISDA)

² Kolbjørnsrud et al

۴. مؤسسه‌ی آکسِنچور^۱ با استفاده از پردازش زبان طبیعی و تشخیص تصویر، راه‌حل‌های هوش مصنوعی را برای کمک به افراد نابینا در هند برای دستیابی به تجربه‌ای بهتر از زندگی در دنیای اطراف خود توسعه داده است. بدین ترتیب آنها می‌توانند زندگی بهتری داشته باشند و آنهایی که شغل دارند می‌توانند بهتر و سریع‌تر کار کنند و شغل‌های چالش‌برانگیزتری را انجام دهند.

۵. مؤسسه‌ی اعتباری موتور فورد^۲ از یادگیری ماشینی برای شناسایی وام‌گیرندگان نادیده گرفته شده استفاده می‌کند. علاوه بر این، از یادگیری ماشینی برای کمک به پذیره‌نویسان برای درک بهتر متقاضیان وام استفاده می‌کند. این برنامه به بهره‌وری پذیره‌نویسان و متقاضیان نادیده گرفته شده کمک می‌کند. نهایتاً، سیستم قادر به پیش‌بینی قابلیت اعتبار وام‌گیرندگان بالقوه است و ضرر و زیان‌های فورد را به حداقل می‌رساند.

۶. آلاستیر کول^۳ از داده‌های گردآوری شده از منابع مختلف با واتسون آی‌بی‌ام برای پیش‌بینی انتظارات مشتریان از شرکت استفاده می‌کند. داده‌های گردآوری شده برای پشتیبانی تصمیمات مؤثرتر کسب‌وکار مورد استفاده قرار می‌گیرند.

۷. شرکت‌ها کسب‌وکارهایشان را بر محوریت هوش مصنوعی می‌سازند. مثال‌های بسیاری از شرکت‌های نوپا یا فعلی وجود دارد که تلاش می‌کنند تا کسب‌وکارهای جدیدی را ایجاد کنند.

دو حوزه که بیشترین منفعت را از هوش مصنوعی برده‌اند تجربه مشتری و لذت مشتری هستند. طبق یک نظرسنجی جهانی گزارش شده توسط «سی‌ام‌اِ ایناویژن ادیتورز^۴» (۲۰۱۷)، ۹۱ درصد از شرکت‌های تراز اول از راه‌حل‌های هوش مصنوعی برای پشتیبانی از تجربه مشتری استفاده کرده‌اند.

برخی از محدودیت‌های ماشین‌های هوش مصنوعی عبارتند از:

- فقدان لمس و احساس انسانی
- عدم توجه به محیط پیرامون
- ممکن است منجر به وابستگی بیش از حد انسان به ماشین‌های هوش مصنوعی شود (مثلاً ممکن است انسان فکر کردن را کنار بگذارد).
- ممکن است با هدف تخریب و نابودی برنامه‌ریزی شود (فصل ۱۴).
- ممکن است موجب از بین رفتن شغل بسیاری از مردم شود (فصل ۱۴).
- ممکن است شروع به فکر کردن کنند و آسیب‌های جدی را بوجود آورند (فصل ۱۴).

برخی از این محدودیت‌ها با گذشت زمان از بین می‌روند. با وجود این، مخاطرات همچنان وجود دارند. از این رو، لازم است تا توسعه‌ی هوش مصنوعی به درستی مدیریت شده و مخاطرات را به حداقل رساند.

¹ Accenture

² Ford Motor Credit

³ Alastair Cole

⁴ CMO Innovation Editors

هوش مصنوعی قادر به انجام چه کارهایی هست و قادر به انجام چه کارهایی نیست؟ محدودیت‌های شناسایی شده فقط قابلیت‌ها و توانمندی‌های تجاری هوش مصنوعی را محدود می‌کنند. به عنوان مثال، استفاده‌ی تجاری آنها بسیار پرهزینه خواهد بود. یک ارزیابی از قابلیت‌های هوش مصنوعی در سال ۲۰۱۶ انجام شد. این موضوع به دو دلیل اهمیت دارد: (۱) مدیران اجرایی نیاز دارند تا بدانند هوش مصنوعی چه کارهایی را می‌توان به شکل اقتصادی انجام دهد و شرکت‌ها چگونه می‌توانند از آن به نفع خودشان بهره‌برداری کنند و (۲) مدیران اجرایی نیاز دارند تا بدانند هوش مصنوعی چه کارهایی را نمی‌تواند به شکل اقتصادی انجام دهد.

هوش مصنوعی موجب دگرگونی جستجوی وب، خدمات بانکداری و خرده‌فروشی، لجستیک، تجارت آنلاین، سرگرمی و غیره شده است. صدها میلیون نفر از هوش مصنوعی بر روی گوشی‌های هوشمند خود یا به روش‌های دیگر استفاده می‌کنند. با وجود این، طبق ان جی (۲۰۱۶)، کاربردها در این حوزه‌ها صرفاً مبتنی بر تبدیل یک ورودی ساده به یک خروجی ساده به صورت پاسخ است. مثلاً، در تأیید خودکار وام، ورودی شرح حال متقاضی است و خروجی به شکل تأیید یا رد درخواست است. کاربردها در این حوزه‌ها طبیعتاً به طور کاملاً خودکار در آمده‌اند. وظایف خودکار معمولاً تکرار شونده بوده و با دوره‌های زمانی کوتاه آموزش توسط انسان انجام می‌شوند. ماشین‌های هوش مصنوعی به داده‌هایی متکی هستند که دستیابی به آنها ممکن است دشوار بوده (مثلاً به فرد دیگری تعلق داشته باشند) یا دقت کافی نداشته باشند. دومین مانع، نیاز به افراد خبره هوش مصنوعی است که یافتن آنها دشوار بوده یا هزینه بالایی دارند. در خصوص سایر موانع به فصل ۱۴ مراجعه کنید.

استف (۲۰۱۷) قابلیت‌های سیستم‌های هوش مصنوعی را به سه سطح تقسیم کرده است: همکار، مستقل (خودکار) و افزوده.

هوش همکار این به میزان زیادی معادل هوش ضعیف مصنوعی (محدود) است که فقط در برخی حوزه‌های خاص کاربرد دارد و نیازمند ورودی‌ها و خروجی‌هایی است که به طور شفاف تعریف شده باشند. مثال‌هایی از آن سیستم‌های پایش و دستیارهای شخصی مجازی سطح پایین هستند (فصل ۱۲). اینگونه سیستم‌ها و دستیارها در وسائل نقلیه ما و برای هشدار دادن استفاده شده‌اند. سیستم‌ای مشابهی را می‌توان در بسیاری از کاربردهای مراقبت سلامت و بهداشت بکار برد (مثلاً مانیتورینگ، تشخیص بیماری).

هوش مستقل این سیستم‌ها در قلمروی هوش مصنوعی قوی (گسترده) قرار دارند ولی دامنه‌های کاربرد آنها بسیار محدود است. سرانجام، رایانه‌ها بسیاری از وظایف و فعالیت‌ها را در اختیار گرفته و آنها را کاملاً خودکار خواهند کرد. ماشین‌ها به عنوان افراد خبره عمل کرده و قدرت تصمیم‌گیری مطلق خواهند داشت. مشاورهای تماماً رباتیک (فصل ۱۲) مثال‌هایی از اینگونه ماشین‌ها هستند. وسائل نقلیه‌ی مستقل و ربات‌هایی که می‌توانند خودشان را تعمیر کنند نیز مثال‌های خوبی هستند.

هوش افزوده اکثر کاربردهای فعلی هوش مصنوعی که مابین هوش همکار و مستقل قرار می‌گیرند را هوش افزوده می‌نامند (یا افزودن هوش). فناوری آنها می‌تواند بر وظایف رایانه‌ها تا حدی بیفزاید که به قابلیت‌های شناختی انسان دست یابند (فصل ۶ رایانش شناختی) و عملکرد سطح بالاتری را نتیجه دهند که در بینش فناوری ۲.۱ توضیح داده شده است.

مغز مصنوعی یک ماشین ساخت دست بشر است که مطلوبیت آن در هوشمند، خلاق و خودآگاه بودن همچون انسان‌ها است. تا به امروز هیچ‌کس نتوانسته است چنین ماشینی را اختراع کند (artificialbrains.com). یکی از پیشگامان در این حوزه، شرکت آی‌بی‌ام است. آی‌بی‌ام و نیروی هوایی ایالات متحده سیستمی معادل ۶۴ میلیون نورون عصبی را ساخته‌اند که قرار است تا سال ۲۰۲ به ۱۰ میلیارد نورون برسد. توجه داشته باشید که مغز انسان حاوی ۱۰۰ میلیارد نورون عصبی است. این سیستم سعی دارد تا از یک مغز بیولوژیکی تقلید کرده و از لحاظ مصرف انرژی نیز اثربخش باشد. این پروژه‌ی آی‌بی‌ام، «ترو نورث^۱» یا «بلو برین^۲» نام دارد و قادر است تا از مغز انسان یاد بگیرد. بسیاری معتقدند که دستیابی به ماشین‌های هوش مصنوعی که به اندازه انسان‌ها خلاق باشند فرایندی طولانی و کند خواهد بود (مثلاً دورمهیل^۳ ۲۰۱۷).

بینش فناوری ۲.۱ هوش افزوده

فکر ترکیب کردن عملکرد انسان و ماشین‌ها چیز جدیدی نیست. ما در اینجا قابلیت‌های انسانی را با هوش قدرتمند ماشین ترکیب می‌کنیم (می‌افزاییم). این به معنای جایگزین کردن انسان‌ها با هوش مصنوعی مستقل نیست بلکه می‌خواهیم قابلیت‌های شناختی انسان را توسعه بدهیم. نتیجه‌ی اینکار، توانایی حل مسائل انسانی پیچیده همانند پرونده‌ی آغازین این فصل است. انسان‌ها به کمک رایانه‌ها می‌توانند مسائلی را حل کنند که تا پیش از این حل نشده بودند. پادمانابان^۴ (۲۰۱۸) بین هوش مصنوعی سنتی و افزود تفاوت‌های زیر را بر شمرده است:

۱. ماشین‌های افزوده به جای اینکه جایگزین تصمیم‌گیری انسان شوند آن را توسعه و بسط می‌دهند و موجب تسهیل خلاقیت می‌شوند.
۲. افزونگی برخلاف هوش مصنوعی گسترده و قوی موجب پیشبرد حل مسائل صنعتی و انسانی پیچیده در حوزه‌های خاص می‌شود.
۳. برخلاف مدل «جعبه سیاه» در برخی تحیل‌ها و هوش مصنوعی، هوش افزوده یک نوع بینش و توصیه‌هایی را از جمله توضیحات فراهم می‌کند.
۴. علاوه بر این، فناوری افزونگی برخلاف هوش مصنوعی همکار با ترکیب اطلاعات فعلی و کشف شده می‌تواند راه‌حل‌های جدیدی را ارائه کند که به شناسایی مشکلات یا نشانه‌ها پرداخته و راه‌حل‌های از پیش تعیین شده را توصیه می‌کند.

پادمانابان (۲۰۱۸) و بسیاری دیگر معتقدند که در حال حاضر، هوش مصنوعی افزوده بهترین گزینه برای حرکت به سمت دگرگونی دنیای هوش مصنوعی است.

برعکس هوش مصنوعی مستقل که ماشین‌ها را به واسطه‌ی طیف گسترده‌ای از قابلیت‌های شناختی تعریف می‌کند (مثلاً وسائل نقلیه‌ی بدون راننده)، هوش افزوده فقط چند قابلیت شناختی محدود دارد.

استف (۲۰۱۷) مثال‌های زیر را آورده است:

- مبارزه با جرائم سایبری. هوش مصنوعی می‌تواند حملات را شناسایی کرده و راه‌حل پیشنهاد دهد.

¹ TrueNorth

² BlueBrain

³ Dormehl

⁴ Padmanabhan

- **تصمیمات تجارت الکترونیک.** ابزارهای بازاریابی می‌توانند تست‌ها را ۱۰۰ برابر سریع‌تر انجام دهند و کارکردهای چیدمان و پاسخ یک وب سایت را با کاربرانش تطبیق دهند. ماشین‌ها می‌توانند توصیه کرده و بازاریاب‌ها این توصیه‌ها را می‌توانند قبول یا رد کنند.
- **معاملات بازار سهام بسامد بالا^۱.** اینکار یا به طور کاملاً مستقل یا در برخی موارد با کنترل و کالیبراسیون توسط انسان انجام می‌شود.

سوالاتی برای بحث و گفتگو

۱. فرضیه‌ی اساسی هوش افزوده چیست؟
۲. تفاوت‌های مهم بین هوش افزوده و هوش مصنوعی سنتی را نام ببرید.
۳. برخی از مزایای هوش افزوده چه هستند؟
۴. این فناوری چه رابطه‌ای با رایانش شناختی دارد؟

سوالاتی برای مرور بخش ۲.۲

۱. هوش مصنوعی را تعریف کنید.
۲. اهداف اصلی هوش مصنوعی چه هستند؟
۳. برخی از مشخصه‌های هوش مصنوعی را نام ببرید.
۴. برخی از محرک‌های هوش مصنوعی را نام ببرید.
۵. برخی از مزیت‌های کاربردهای هوش مصنوعی را نام ببرید.
۶. برخی از محدودیت‌های هوش مصنوعی را نام ببرید.
۷. مغز مصنوعی را توضیح دهید.
۸. سه ویژگی هوش مصنوعی را نام برده و افزونگی را توضیح دهید.

هوش مصنوعی به دلیل افزایش قابلیت‌ها و توانمندی‌هایش به سرعت در حال رشد است. به منظور درک هوش مصنوعی ما ابتدا باید معنای هوش را متوجه شویم.

¹ High-Frequency Stock Market Trading

هوش را می‌توان یک عبارت چتری دانست و معمولاً بر اساس یک آزمون آی کیو سنجیده می‌شود. با وجود این، برخی معتقدند که چندین نوع هوش وجود دارد. به عنوان مثال، دکتر هاوارد گاردنر از دانشگاه هاروارد انواع هوش را بیان کرده است:

- زبانشناختی و شفاهی
- منطقی
- مکانی
- بدن/حرکت
- موسیقایی
- بین شخصی
- بیناشخصی
- طبیعت‌گرا

بنابراین، هوش یک مفهوم ساده نیست.

محتوای هوش هوش از استدلال، یادگیری، منطقی، توانایی حل مسئله، ادراک و توانایی زبانشناختی تشکیل شده است. مشخصاً، هوش مفهوم ساده‌ای نیست.

قابلیت‌های هوش به منظور درک هوش مصنوعی، ابتدا بهتر است قابلیت‌هایی که نشانه‌های هوش انسانی محسوب می‌شوند را ارزیابی کنیم:

- یادگیری از تجربه‌ها یا درک آنها
- کشف معنا و مفهوم از پیام‌ها و اطلاعات مبهم، ناقص یا حتی متناقض
- واکنش سریع و موفق نسبت به یک موقعیت جدید (یعنی استفاده از صحیح‌ترین پاسخ‌ها)
- درک و استنباط به یک روش منطقی، حل مسائل و هدایت مؤثر برای انجام کار
- بکارگیری دانش برای دستکار محیط و موقعیت‌ها
- تشخیص و قضاوت در مورد اهمیت نسبی عناصر مختلف در یک موقعیت

هوش مصنوعی تلاش می‌کند تا برخی یا تمامی این قابلیت‌ها را ارائه دهد ولی به طور کلی فعلاً قادر نیست تا خودش را با هوش انسانی تطبیق دهد.

ماشین‌های هوش مصنوعی در انجام بازی‌های پیچیده مثل شطرنج (شکست دادن قهرمان جهان)، ¹جئوپاردی! (شکست دادن بهترین بازیکنان) و گو² (یک بازی پیچیده‌ی چینی) برتری خود را نسبت به انسان نشان داده‌اند که برترین بازیکنان توسط

¹ Jeopardy!

² Go

برنامه‌های معروف دیپ‌ماینند گوگل^۱ شکست خورده‌اند (هیو ۲۰۱۶)^۲. علیرغم این نمایش‌های خارق‌العاده (که بسیار پرهزینه بوده‌اند)، بسیاری از کاربردهای هوش مصنوعی هنوز پایین‌تر از سطح هوش انسانی هستند.

مقایسه‌ی هوش انسان با هوش مصنوعی تلاش‌های متعددی به منظور مقایسه‌ی هوش انسانی با هوش مصنوعی به عمل آمده است. انجام این مقایسه کار دشواری است چون که یک موقعیت چندبعدی محسوب می‌شود. یک مقایسه در جدول ۲.۱ ارائه شده است.

جدول ۲.۱ هوش انسانی در برابر هوش مصنوعی

حوزه	هوش مصنوعی	انسان
اجرا	بسیار سریع	می‌تواند کند باشد
عواطف	وجود ندارد	مثبت یا منفی
سرعت محاسبه	بسیار سریع	کند است و ممکن است مشکل باشد
تصورات	فقط چیزی که برایش برنامه‌نویسی شده	می‌تواند دانش موجود را گسترش دهد
پاسخ به سوالات	آنچه در برنامه وجود دارد	می‌تواند دارای نوآوری باشد
انعطاف‌پذیری	فاقد انعطاف‌پذیری	انعطاف‌پذیر
زیربنا	یک کد دودویی	حواس پنجگانه
سازگاری	بالا	متغیر، ممکن است ضعیف باشد
پردازش	به شکلی که مدلسازی شده	شناختی
فرم	اعداد	سیگنال‌ها
حافظه	توکار یا دسترسی از فضای ابر	استفاده از حافظه‌ی محتوایی و شمایی
مغز	مستقل	متصل به جسم
خلایقیت	فاقد الهام‌بخشی	واقعاً خلاق
دوام‌پذیری	دائمی ولی در صورتیکه به‌روزآوری نشود ممکن است متروک و مهجور شود.	از بین رفتنی ولی امکان به‌روزآوری دارد
تکثیر، مستندسازی و انتشار	آسان	دشوار
هزینه	معمولاً پایین و در حال کاهش	ممکن است بالا و رو به افزایش باشد
سازگاری	پایدار	در برخی مواقع غیرقابل پیش‌بینی
فرایند استدلال	واضح، مشهود	ردیابی آن در مواقعی دشوار است
ادراک	بر اساس قواعد و داده‌ها	بر اساس الگوها
پی بردن به داده‌ها از دست رفته	معمولاً امکان‌پذیر نیست	اغلب امکان‌پذیر است

¹ Google's DeepMind

² Hughes

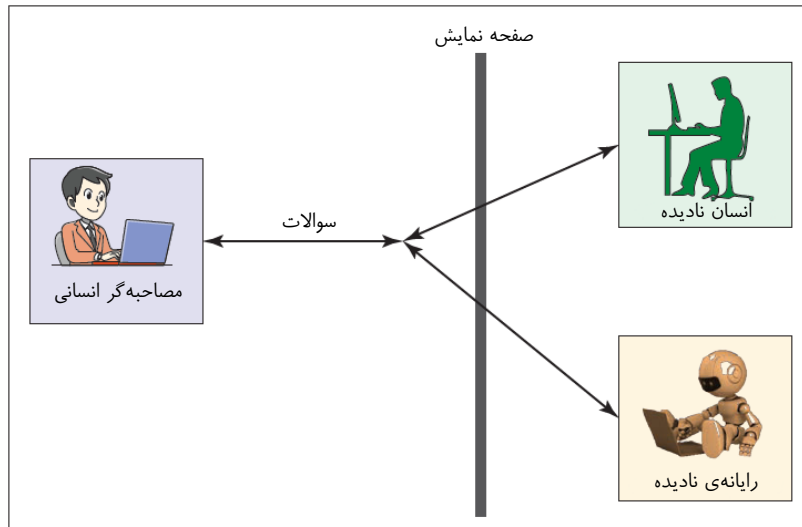
آزمون تورینگ^۱ یک روش شناخته شده به منظور اندازه‌گیری سطح هوش ماشین‌های هوش مصنوعی می‌باشد.

آزمون تورینگ: سنجش کلاسیک هوش ماشین آلن تورینگ آزمونی را تحت عنوان آزمون تورینگ طراحی کرد که تعیین می‌کند آیا یک رایانه رفتار هوشمندی نشان می‌دهد یا خیر. بر اساس این آزمون، یک رایانه را تنها زمانی می‌تواند باهوش دانست که مصاحبه‌گر انسانی که سؤالات یکسانی را هم از یک رایانه و هم انسان دیگری که در معرض دید نیستند می‌پرسد قادر نباشد تشخیص دهد که کدامیک انسان و کدامیک ماشین هستند (شکل ۲.۳). توجه داشته باشید که این آزمون محدود به حالت پرسش و پاسخ است.

برای اینکه یک رایانه بتواند آزمون تورینگ را با موفقیت پشت سر بگذارد باید قادر به درک یک زبان انسانی بوده، از هوش انسانی برخوردار بوده (مثلاً یک پایگاه دانشی داشته باشد)، با استفاده از دانش ذخیره شده بتواند استدلال کند و قادر به یادگیری از تجربیاتش باشد (یادگیری ماشینی).
نکته: جایزه‌ی ۱۰۰,۰۰۰ دلاری از انتظار کسانی است که بتوانند نرم‌افزاری را توسعه دهند که حقیقتاً هوشمند باشد (یعنی آزمون تورینگ را با موفقیت پشت سر بگذارد).

سایر آزمون‌ها طی سال‌ها، پیشنهادات دیگری در خصوص نحوه‌ی سنجش هوش ماشین ارائه شده است. به عنوان مثال، بهبودهای متعددی در آزمون تورینگ به عمل آمد. دانشگاه‌های مهم در ایالات متحده (مثلاً دانشگاه ایلینوی، مؤسسه فناوری ماساچوست، دانشگاه استنفورد) درگیر مطالعه‌ی آی کیو برای هوش مصنوعی هستند. علاوه بر این، چندین آزمون سنجش دیگر نیز وجود دارد. یکی از این آزمون‌ها را در مورد کاربردی ۲.۱ بررسی خواهیم کرد. در مجموع، سنجش سطح هوش انسان‌ها و ماشین‌ها کار دشواری است که به شرایط و سنجه‌های بکار رفته بستگی دارد.

¹ Turing test



شکل ۲.۳ تصویر نمایشی از آزمون تورینگ

یک جاروبرقی تا چه اندازه می‌تواند باهوش باشد؟

مورد کاربردی ۲.۱

محض اطلاع شما، جاروبرقی‌ها می‌توانند باهوش باشند. برخی از شما ممکن است از محصول رومبا^۱ شرکت آی‌ربات^۲ استفاده کرده باشید. این جاروبرقی به تنهایی می‌تواند کف زمین را تمیز کند و تا حدودی هوشمند است. با وجود این، در خانه‌های هوشمند (فصل ۱۳) ما انتظار داریم تا جاروبرقی‌های باهوش‌تری را ببینیم. یکی از اینها، رُبوکینگ توربو پلاس^۳ محصول شرکت ال جی کره است. پژوهشگران در دانشگاه ملی رباتیک^۴ و آزمایشگاه سیستم هوشمند^۵ در سئول کره جنوبی به مطالعه و بررسی رُبوکینگ پرداخته و تأیید کرده‌اند که الگوریتم یادگیری عمیق این محصول به نحوی است که هوش یک کودک شش یا هفت ساله را دارد. اگر خودروهای خودران داریم چرا نتوانیم یک جاروبرقی خودران داشته باشیم که به مراتب ساده‌تر از یک خودرو است. جاروبرقی‌ها فقط کافی است در اطراف اتاق حرکت کنند. برای اینکار، ماشین‌ها باید موقعیت خود را در اتاق « دیده » و موانع سر راه خود را شناسایی کنند. سپس پایگاه دانش جاروبرقی باید ببیند بهترین کاری که باید انجام دهد چیست. این اصولاً چیزی است که بسیاری از حسگرهای ماشین‌ها، پایگاه‌های دانش و قواعد هوش مصنوعی انجام می‌دهند. علاوه بر این، ماشین‌های هوش مصنوعی باید از تجربیات گذشته یاد بگیرند (مثلاً چه کاری را که در گذشته جواب نداده است نباید انجام دهد).

رُبوکینگ مجهز به برنامه‌ی هوش مصنوعی دیپ تین کیوتی‌ام^۶ شرکت ال جی است که جاروبرقی را قادر می‌سازد تا به ماهیت موانعی که با آنها مواجه می‌شود پی ببرد. این برنامه به جاروبرقی می‌گوید که در اطراف مبلمان حرکت کند، منتظر شود تا سگ

¹ Roomba

² iRobot

³ Roboking Turbo Plus

⁴ Seoul National University Robotics

⁵ Intelligent System Lab

⁶ Deep Thin QTM

کنار برود یا توقف کند. بدین ترتیب، این ماشین تا چه اندازه هوشمند است؟ برای پاسخ به این سوال، پژوهشگران کره‌ای ۱۰۰ سنجه را تعریف کرده و جاروبرقی‌ها را آزموده‌اند.

عملکرد جاروبرقی‌های آزموده شده بر اساس این ۱۰۰ سنجه به سه سطح تقسیم شده است. این سطوح عبارتند از هوشمند در سطح یک دلفین، هوشمند در سطح یک میمون و هوشمند در سطح یک کودک ۶ تا ۷ ساله. این مطالعه تأیید می‌کند که ربوکنینگ وظایفش را در سطحی بالاتر از هوش ماشین انجام داده است.

Sources: Compiled from Fuller (2017) and webwire.com/ViewPressRel.asp?aId=211017 news dated July 18, 2017.

سوالاتی برای مورد کاربردی ۲.۱

۵. پژوهشگران کره‌ای چگونه عملکرد جاروبرقی‌ها را تعیین کردند؟
۶. اگر شما یک دستگاه رومبا داشته باشید به نظرتان تا چه اندازه هوشمند است؟
۷. با ویژگی یادگیری عمیق به چه قابلیت‌هایی می‌توان دست یافت؟
۸. جدیدترین اطلاعات به ویژه جدیدترین بهبودها را در مورد ربوکنینگ ال جی پیدا کنید.

صرفنظر از تعیین میزان هوشمند بودن یک ماشین، هوش مصنوعی مزایای بسیاری دارد که قبل تر توضیح داده شد. شایان ذکر است که قابلیت‌های هوش مصنوعی با گذشت زمان افزایش می‌یابد. به عنوان مثال، یک آزمایش در دانشگاه استنفورد (پام ۲۰۱۸)^۱ نشان داد که برنامه‌های هوش مصنوعی در شرکت‌های مایکروسافت و علی‌بابا در آزمون‌های درک مطلب امتیازهای بالاتری را نسبت به صدها نفر شرکت‌کننده کسب کرده‌اند. (البته اینها برنامه‌های هوش مصنوعی بسیار گران قیمتی هستند). برای بحث در خصوص مقایسه‌ی هوش انسانی و هوش مصنوعی به کارنی^۲ (۲۰۱۸) رجوع کنید.

سوالاتی برای مرور بخش ۲.۳

۱. منظور از هوش چیست؟
۲. توانمندی‌های اصلی هوش انسانی چه هستند؟ کدامیک بر ماشین‌های هوش مصنوعی برتری دارند؟
۳. هوش مصنوعی تا چه اندازه هوشمند است؟
۴. چگونه می‌توانیم سطح هوشمندی هوش مصنوعی را بسنجیم؟
۵. آزمون تورینگ چیست و محدودیت‌های آن چه هستند؟
۶. چگونه می‌توانیم سطح هوشمندی یک جاروبرقی را اندازه‌گیری کنیم؟

رشته‌ی هوش مصنوعی بسیار گسترده است و می‌توان کاربردها و فناوری‌های هوش مصنوعی را در صدها رشته از پزشکی گرفته تا ورزش مشاهده کرد. پرس (۲۰۱۷) فهرستی از ۱۰ فناوری برتر هوش مصنوعی را مشابه آنچه ما در این کتاب نوشته‌ایم ارائه کرده است. پرس همچنین وضعیت چرخه زندگی (فاز اکوسیستم) فناوری‌ها را نیز ارائه کرده است. در این بخش ما به معرفی

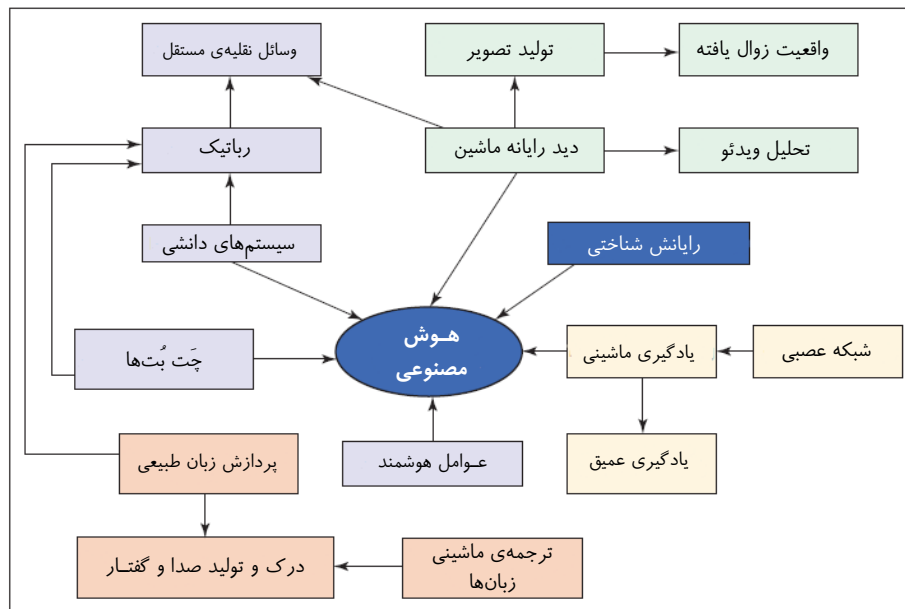
¹ Pham

² Carney

برخی از فناوری‌های اصلی هوش مصنوعی و مشتقات آنها که با کسب‌وکار مرتبط هستند می‌پردازیم. این فهرست منتخب در شکل ۲.۴ به تصویر کشیده شده است.

یک عامل هوشمند^۱ یک برنامه‌ی نرم‌افزاری رایانه‌ای مستقل و نسبتاً کوچک است که با انجام مستقل برخی وظایف و فعالیت‌های خاص می‌تواند محیط پیرامونش را رصد کرده و مطابق با تغییراتی که بوجود می‌آید عمل کند. عامل هوشمند، فعالیت‌های عامل را در جهت دستیابی به اهداف مشخصی که مربوط به تغییرات در محیط پیرامونش است هدایت می‌کند. عوامل هوشمند ممکن است با بکارگیری و گسترش دانشی که در آنها تعبیه شده توانایی یادگیری داشته باشند. عوامل هوشمند، ابزارهای مؤثری برای غلبه بر مهمترین مسئولیت اضافه‌بار اطلاعات اینترنت و تبدیل رایانه‌ها به ابزارهای دوام‌پذیر برای پشتیبانی تصمیم هستند. توجه به استفاده از عوامل هوشمند برای کسب‌وکار و تجارت الکترونیک در فضای دانشگاهی از اواسط دهه ۹۰ آغاز شد. با وجود این، از سال ۲۰۱۴ و زمانیکه توانمندی‌های هوش مصنوعی افزایش قابل توجهی داشت ما شاهد کاربردهای قدرتمند در بسیاری از حوزه‌های کسب‌وکار، اقتصاد، دولت و خدمات بودیم.

در ابتدا، عوامل هوشمند عمدتاً برای پشتیبانی از فعالیت‌های روزمره مثل جستجوی محصولات، توصیه‌ها، تعیین قیمت محصولات، برنامه‌ریزی بازاریابی، افزایش امنیت رایانه‌ها، مدیریت مزایده‌ها و حراجی‌ها، تسهیل پرداخت‌ها و بهبود مدیریت موجودی استفاده می‌شدند. با وجود این، این کاربردها بسیار ساده بودند و از سطح هوشمندی پایینی بهره می‌بردند. مزایای اصلی آنها افزایش سرعت، کاهش هزینه‌ها، کاهش خطاها و بهبود خدمات مشتریان بود. کاربردهای امروزی همانطور که در این فصل خواهیم دید بسیار پیچیده‌تر هستند.



شکل ۲.۴ فناوری‌های هوش مصنوعی اصلی

¹ Intelligent Agent (IA)

یک مثال ساده از یک عامل نرم‌افزاری هوشمند، یک برنامه‌ی تشخیص ویروس است. این برنامه بر روی یک رایانه نصب شده و تمامی داده‌های ورودی را پویش کرده و ویروس‌های یافت شده را به صورت خودکار حذف می‌نماید و در عین حال یاد می‌گیرد تا انواع ویروس‌های جدید و روش‌های تشخیص را نیز تشخیص دهد.

بیمه‌ی کسب‌وکار آل استیت^۱ از یک عامل هوشمند برای کاهش ترافیک مرکز تماس و کمک به عوامل بیمه انسانی در طی فرایند اعلام نرخ به مشتریان استفاده می‌کند. در این موارد، اعلام نرخ می‌تواند کار پیچیده‌ای باشد. عوامل حتی اگر با طور کامل با مسائل مربوطه آشنا نباشند با استفاده از این سیستم می‌توانند سریعاً به سوالات مطرح شده توسط مشتریان حقوقی پاسخ دهند. عوامل هوشمند همچنین در سرویس‌دهنده‌های پست الکترونیک، فیلترینگ و نشر خبر، رسیدگی به قرار ملاقات‌ها و گردآوری خودکار اطلاعات مورد بهره‌برداری قرار می‌گیرند.

در حال حاضر، سیستم‌های هوش مصنوعی فاقد توانمندی‌های یادگیری انسان‌ها هستند و بیشتر روش‌های ساده یادگیری ماشین را دارند که از روی روش‌های یادگیری انسانی مدل‌سازی شده‌اند. دانشمندان حوزه‌ی یادگیری ماشین سعی دارند تا به رایانه‌ها آموزش دهند که *الگوها را شناسایی کنند* و با نشان دادن حجم انبوهی از مثال‌ها و داده‌های مرتبط، ارتباط برقرار کنند. یادگیری ماشین همچنین به سیستم‌های رایانه‌ای اجازه می‌دهد تا فعالیت‌های محیط پیرامون خود را پایش و حس کنند تا ماشین‌ها بتوانند رفتارشان را متناسب با تغییرات بوجود آمده در محیط وفق دهند. از این فناوری همچنین می‌توان برای پیش‌بینی عملکرد، پیکربندی مجدد برنامه‌ها بر اساس شرایط متغیر و بسیاری مسائل دیگر استفاده کرد. از لحاظ فنی، یادگیری ماشین یک رشته‌ی علمی است که با طراحی و توسعه‌ی الگوریتم‌هایی سر و کار دارد که به رایانه‌ها اجازه می‌دهد تا بر اساس داده‌هایی که از حسگرها، پایگاه‌های داده و سایر منابع به دست می‌آیند یاد بگیرند. سپس از این یادگیری برای پیش‌بینی، تشخیص الگو و پشتیبانی تصمیم‌گیرها استفاده می‌شود. برای یک مرور کلی به آلپایدین^۲ (۲۰۱۶) و تئوبالد^۳ (۲۰۱۷) مراجعه کنید.

امروزه بسیاری از شرکت‌ها از الگوریتم‌های یادگیری ماشین (فصل ۵ برای توضیح و بحث) استفاده می‌کنند. برای یک راهنمای اجرایی در خصوص یادگیری ماشین نگاه کنید به پایل^۴ و سان خوزه^۵ (۲۰۱۵).

فرایند یادگیری ماشین شامل برنامه‌های رایانه‌ای است که در مواجهه با موقعیت‌های جدید، یاد می‌گیرند. این برنامه‌ها داده‌ها را گردآوری و تحلیل نموده و سپس خودشان را برای رسیدن به نتایج «آموزش می‌دهند». به عنوان مثال، با نشان دادن مثال‌هایی از موقعیت‌ها به یک برنامه یادگیری ماشین، برنامه می‌تواند اجزایی که بدون آن برنامه به آسانی قابل مشاهده نیستند را پیدا کند. یک مثال معروف از این، رایانه‌هایی هستند که کلاهبرداری کارت اعتباری را تشخیص می‌دهند. مورد کاربردی ۲.۲ نشان می‌دهد که یادگیری ماشین چگونه می‌تواند فرایندهای کسب‌وکار شرکت‌ها را بهبود بخشد.

¹ Allstate Business Insurance

² Alpaydin

³ Theobald

⁴ Pyle

⁵ San Jose

طبق تیلور^۱ (۲۰۱۶)، «افزایش توان محاسباتی به همراه سایر بهبودها از جمله الگوریتم‌های بهتر و شبکه‌های عصبی عمیق برای پردازش تصویر و پایگاه‌های داده حافظه فوق العاده سریع مثل سَپ هانا^۲ موجب شده‌اند تا یادگیری ماشین امروزه تبدیل به داغ‌ترین حوزه دستاوردها در نرم‌افزارهای سازمانی شود.» کاربردهای یادگیری ماشین نیز به دلیل دسترسی پذیری منابع کلان داده‌ها به ویژه آنهایی که توسط اینترنت اشیاء ارائه می‌شوند در حال گسترش است. یادگیری ماشین اصولاً به معنی یادگیری از داده‌ها است.

روش‌های متعددی برای یادگیری ماشین از شبکه‌های عصبی تا استدلال مبتنی بر مورد است. مهمترین آنها در فصل ۵ آورده شده است.

چگونه یادگیری عمیق به بهبود کسب و کار کمک می‌کند؟

مورد کاربردی ۲.۲

ولرز و همکاران^۳ (۲۰۱۷) که گفته‌اند «امروزه سازمان‌های پیشرو از ابزارهای مبتنی بر یادگیری ماشین برای خودکارسازی فرایندهای تصمیم استفاده می‌کنند.» مثال‌های زیر را از بکارگیری یادگیری ماشین ارائه کرده‌اند:

۱. ارتقاء نگهداشت و وفاداری مشتری. شرکت‌ها به منظور پیش‌بینی نگهداشت و وفاداری مشتریان به کاوش در فعالیت‌ها، معاملات و تعاملات اجتماعی و احساسات و تمایلات مشتریان می‌پردازند. شرکت‌ها می‌توانند از یادگیری ماشین به عنوان مثال برای پیش‌بینی تمایل افراد برای تغییر شغل استفاده کنند و کارفرمایان می‌توانند برای حفظ کارکنان یا جذب کارکنان بالقوه که در جای دیگری کار می‌کنند پیشنهادات جذابی ارائه کنند.
۲. استخدام افراد درست. با فرض اینکه به طور متوسط ۲۵۰ نفر متقاضی یک شغل خوب در شرکت‌های خاصی هستند، یک برنامه‌ی هوش مصنوعی می‌تواند رزومه‌های متقاضیان را تحلیل کرده و نامزدهای واجد شرایطی که تقاضا نکرده‌اند ولی رزومه خود را به صورت آنلاین قرار داده‌اند را پیدا کند.
۳. خودکارسازی امور مالی. تراکنش‌های مالی ناتمام که برخی داده‌های آنها تکمیل نیستند (مثلاً شماره سفارش) نیازمند توجه ویژه‌ای هستند. سیستم‌های یادگیری ماشین می‌توانند یاد بگیرند که چگونه اینگونه موقعیت‌ها را تشخیص داده و تصحیح نمایند. برنامه‌ی هوش مصنوعی می‌تواند اقدامات اصلاحی لازم را به طور خودکار انجام دهد.
۴. تشخیص کلاهبرداری. الگوریتم‌های یادگیری ماشین از شناخت الگو برای تشخیص کلاهبرداری استفاده می‌کنند. این برنامه‌ها به دنبال موارد نامتعارف گشته و سپس در خصوص فعالیت‌های شناسایی شده‌ای که باید به دنبال کلاهبرداری در آنها باشد استنتاج می‌کند. مؤسسات مالی، کاربران اصلی این برنامه هستند.
۵. نگهداری و تعمیرات پیشگیرانه. یادگیری ماشین می‌تواند موارد نامتعارف را در عملیات یک دستگاه و پیش از آنکه از کار بیفتد شناسایی کند. بنابراین، اقدامات اصلاحی بلافاصله و با هزینه‌ای بسیار کمتر از زمانی که دستگاه از کار می‌افتد انجام می‌شوند. علاوه بر این، نگهداری پیشگیرانه را به بهینه‌ترین روش می‌توان انجام داد (پرونده‌ی آغازین فصل ۱).

¹ Taylor

² SAP HANA

³ Wellers et al.

۶. تحلیل فضای قفسه‌ی خرده‌فروشی^۱. یادگیری ماشین به همراه بینایی ماشینی می‌تواند صفحات نمایش در فروشگاه‌های فیزیکی را به منظور بررسی اینکه آیا اقلام در مکان مناسب خود در قفسه‌های قرار گرفته‌اند یا خیر، آیا اجناس به خوبی در قفسه‌های چیده شده‌اند و آیا برچسب محصولات (از جمله قیمت) به خوبی نشان داده می‌شود یا خیر تحلیل کنند.

۷. سایر پیش‌بینی‌ها. از یادگیری ماشین برای انواع بسیار زیادی از پیش‌بینی‌ها از پزشکی گرفته تا سرمایه‌گذاری استفاده می‌شود. یک مثال، گوگل فلایتز^۲ است که تأخیرهایی که هنوز توسط خطوط هوایی اعلام نشده‌اند را پیش‌بینی می‌کند.

Source: Compiled from Wellers, et al. (2017) and Theobald (2017).

سوالاتی برای مورد کاربردی ۲.۲

۱. در مورد مزایای ترکیب یادگیری ماشین با سایر فناوری‌های هوش مصنوعی بحث کنید.
۲. یادگیری ماشین چگونه می‌تواند موجب بهبود بازاریابی شود؟
۳. در مورد فرصت‌های بهبود مدیریت منابع انسانی بحث کنید.
۴. در مورد مزایای آن برای خدمات مشتری بحث کنید.

یادگیری عمیق یک زیرمجموعه یا حالت اصلاح یافته از یادگیری ماشین را یادگیری عمیق^۳ می‌نامند. این فناوری که در فصل ۶ بحث شده است سعی دارد تا از نحوه‌ی کارکرد مغز انسان تقلید کند. یادگیری عمیق از فناوری عصبی مصنوعی استفاده کرده و نقش مهمی را در رسیدگی به کاربردهای پیچیده که یادگیری ماشین عادی و سایر فناوری‌های هوش مصنوعی قادر به حل آنها نیستند ایفا می‌کند. یادگیری عمیق، سیستم‌هایی را تحویل می‌دهد که نه تنها فکر می‌کنند بلکه به طور مستمر در حال یادگیری هستند و بدین ترتیب با جریان داده‌های تازه می‌توانند خود را سمت و سو بدهند. یادگیری عمیق می‌تواند مسائلی که تا پیش از این لاینحل بوده‌اند را با استفاده از الگوریتم‌های یادگیری قدرتمند خود حل کند. به عنوان مثال، یادگیری عمیق یک فناوری مهم در وسائل نقلیه‌ی مستقل است که به تفسیر علائم و موانع جاده‌ای کمک می‌کند. یادگیری عمیق همچنین نقش حیاتی را در گوشی‌های هوشمند، رباتیک، تبلت‌ها، خانه‌های هوشمند و شهرهای هوشمند ایفا می‌کند (فصل ۱۳). یادگیری عمیق عمدتاً برای کاربردهای تعاملی بلندرنج در حوزه‌های بینایی ماشینی، تشخیص صحنه، رباتیک و پردازش صدا و گفتار مفید است. کلید این قضیه، یادگیری مستمر است. تا زمانیکه داده‌های جدید وارد می‌شوند، یادگیری ادامه دارد.

شرکت کارجیل ارائه دهنده‌ی تحلیل‌های سنتی و متداول و تحلیل‌های مبتنی بر یادگیری عمیق است که به کشاورزان کمک می‌کند تا کارهای سودآورتری انجام دهند. به عنوان مثال، کشاورزان می‌توانند میگوهای بهتری را با هزینه پایین‌تر پرورش دهند. از یادگیری عمیق به طور گسترده‌ای در تحلیل و پیش‌بینی بازار بورس استفاده می‌شود. برای جزئیات بیشتر به اسمیت (۲۰۱۷) و فصل ۶ مراجعه کنید.

¹ Retail shelf analysis

² Google Flights

³ Deep Learning (DL)

تعاریف بینایی ماشینی به دلیل وجود سیستم‌های بینایی رایانه‌های مختلف از جمله تفاوت در سخت‌افزارها و نرم‌افزارها و همچنین سایر مؤلفه‌ها بسیار متغیر است. به بیان کلی، تعاریف کلاسیک *بینایی ماشینی*¹ این است که شامل « فناوری و روش‌هایی است که به منظور بازرسی و تحلیل خودکار تصویربرداری برای کاربردهایی از قبیل هدایت ربات، کنترل فرایند، وسائل نقلیه‌ی مستقل و بازرسی.» بکار می‌رود. بینایی ماشینی، ابزار مهمی برای بهینه‌سازی تولید و فرایندهای رباتیک است. یک بخش مهم در بینایی ماشینی، دوربین‌های صنعتی هستند که اطلاعات دیداری را گرفته، ذخیره و بایگانی می‌کنند. این اطلاعات سپس برای تحلیل و نهایتاً تصمیم‌گیری خودکار یا پشتیبانی از تصمیم‌گیری انسانی به کاربر یا برنامه رایانه‌ای ارسال می‌شود. بینایی ماشینی ممکن است با بینایی رایانه‌ای اشتباه گرفته شود زیرا برخی اوقات این دو عبارت به عنوان مترادف یکدیگر بکار می‌روند ولی برای کاربران و پژوهشگران این دو را کاملاً متفاوت می‌دانند. بینایی ماشینی بیشتر به عنوان یک رشته‌ی زیرمجموعه مهندسی در نظر گرفته می‌شود در حالیکه بینایی رایانه‌ای به حوزه علوم رایانه مربوط است.

بینایی رایانه‌ای بینایی رایانه‌ای طبق ویکپدیا « یک رشته‌ی میان‌رشته‌ای است که با نحوه‌ی دستیابی رایانه‌ها به درک سطح بالا از تصاویر و ویدئوهای دیجیتالی سر و کار دارد. از منظر مهندسی، بینایی رایانه‌ای به دنبال خودکارسازی وظایفی است که سیستم دیداری انسان قادر به انجام آن هستند.» بینایی رایانه‌ای قادر به اکتساب یا پردازش، تحلیل و تفسیر تصاویر و تولید اطلاعات معنادار برای تصمیم‌گیری می‌باشد. داده‌های تصویری می‌توانند در قالب‌های متعددی از جمله عکس‌ها یا ویدئوها بوده و از منابع چند بُعدی گرفته شوند (مثلاً اسکنرهای پزشکی). تشخیص صحنه و آیتم از اجزای مهم در بینایی رایانه‌ای هستند. رشته‌ی بینایی رایانه‌ای نقش حیاتی را در حوزه‌های ایمنی، امنیت، سلامت و بهداشت و سرگرمی ایفا می‌کند. بینایی رایانه‌ای یکی از فناوری‌های هوش مصنوعی محسوب می‌شود که ربات‌ها و وسائل نقلیه‌ی مستقل را قادر به دیدن می‌کند. هم بینایی رایانه‌ای و هم بینایی ماشینی موجب خودکارسازی بسیاری از فعالیت‌ها و وظایف انسانی می‌شوند (مثل بازرسی). این وظایف می‌توانند با یک تصویر یا مجموعه‌ای از تصاویر مرتبط باشند. مزایای اصلی این دو فناوری، کاهش هزینه‌های انجام کارها به ویژه فعالیت‌های تکرارشونده هستند که موجب خستگی چشم انسان می‌شوند. این دو فناوری با پردازش تصویر² ترکیب شده و موجب تسهیل کاربردهای پیچیده از قبیل کنترل کیفیت بصری (چشمی) می‌شوند. از یک دیدگاه دیگر، این دو فناوری بر اساس پردازش تصویر و به اشتراک‌گذاری انواع رشته‌های تأثیرگذار بر موضوع با یکدیگر مرتبط هستند. یکی از حوزه‌های کاربرد بینایی ماشینی، **تشخیص صحنه** است که توسط بینایی رایانه‌ای انجام می‌شود. تشخیص صحنه امکان تشخیص و تفسیر اشیاء، عکس‌ها و مناظر را فراهم می‌کند.

قاجاق الوار یکی از معضلات اساسی در بسیاری از کشورها می‌باشد. بررسی چوب در محل به منظور تطابق با قوانین ایالات متحده، اروپا و سایر کشورها امری ضروری است. اینکار نیازمند تخصص است. بر اساس وزارت کشاورزی ایالات متحده، «نیاز مبرم به اینگونه تخصص در محل، آموزش و استقرار افراد به منظور شناسایی چوب‌های فراوری شده در محل (یعنی در بنادر، ایستگاه‌های مرزی، وزن ایستگاه‌ها، فرودگاه‌ها یا سایر نقاط تجاری ورودی) بسیار پرهزینه و از لحاظ لجستیکی پیچیده و دشوار است. پروژه‌ی

¹ Machine vision

² Image processing

شناسایی چوب توسط بینایی ماشینی یک سیستم نمونه بینایی ماشینی برای شناسایی چوب طراحی و توسعه داده است.» به همین منوال، بینایی رایانه‌ای هوش مصنوعی به همراه یادگیری عمیق برای شناسایی قاچاقچیان حیوانات بکار می‌رود (یوسی اس ۲۰۱۸).

یک مثال دیگر از این کاربرد، تشخیص چهره^۱ است که در بسیاری از کاربردهای امنیتی از قبیل استفاده پلیس چین از عینک‌های هوشمند برای شناسایی اعتراضات احتمالی استفاده می‌شود. در سال ۲۰۱۸، پلیس چین فرد مشکوکی را که در یک کنسرت موسیقی شرکت کرده بود شناسایی کرد. حدود ۶۰،۰۰۰ نفر جمعیت در این کنسرت حضور داشت. فرد نامبرده در قسمت ورودی که عسکش توسط یک دوربین گرفته شد شناسایی گردید ([youtube.com/watch?v=Fq1SEqNT-7c](https://www.youtube.com/watch?v=Fq1SEqNT-7c)). در سال ۲۰۱۸، خدمات مهاجرت و شهروندی ایالات متحده افرادی که از گذرنامه‌های جعلی استفاده کرده بودند به همین روش شناسایی نمود.

تحلیل ویدئو بکارگیری تکنیک‌های بینایی رایانه‌ای بر روی ویدئوها امکان تشخیص الگوها (مثلاً تشخیص کلاهبرداری) و شناسایی رویدادها را به ما می‌دهد. این یکی از کاربردهای مشتق شده بینایی رایانه‌ای می‌باشد. یک مثال دیگر، با پخش نمایش‌های تلویزیونی برای رایانه‌ها این است که به رایانه‌ها آموزش دهیم تا در خصوص تعاملات انسانی و موفقیت تبلیغات پیش‌بینی کنند.

سیستم‌های حسگر مشابه آنهایی که برای تشخیص صحنه و پردازش سیگنال بکار می‌روند زمانیکه با سایر فناوری‌های هوش مصنوعی ترکیب شوند، گروه وسیع‌تری از سیستم‌های یکپارچه و احتمالاً پیچیده را به نام رباتیک^۲ تشکیل می‌دهند. تعاریف متعددی برای رباتیک وجود دارد که در طول زمان تغییر می‌کنند. یک تعریف کلاسیک این است: «یک ربات، یک دستگاه یا وسیله‌ی الکترومکانیکی است که توسط یک برنامه‌ی رایانه‌ای برای انجام کارهای ذهنی و/یا دستی هدایت می‌شود.» مؤسسه رباتیک آمریکا تعریف رسمی ربات را اینگونه ارائه کرده است: «یک منیپولاتور چند وظیفه‌ای قابل برنامه‌ریزی که برای جابجا نمودن مواد، قطعات، ابزارها یا لوازم و دستگاه‌های تخصصی از طریق حرکات برنامه‌ریزی شده‌ی متغیر برای انواع مختلف فعالیت‌ها طراحی شده است.» در این تعریف، بسیاری از وظایفی که توسط ربات‌های امروزی انجام می‌شود نادیده گرفته شده است. یک ربات «هوشمند» دارای نوعی آپاراتوس حسگر مثل دوربین است که اطلاعات مربوط به محیط پیرامون ربات و عملیات آن را گردآوری می‌کند. داده‌های گردآوری شده توسط مغز ربات پردازش می‌شوند که به او اجازه می‌دهد تا نسبت به تغییرات محیط واکنش نشان دهد.

ربات‌ها می‌توانند به طور کاملاً مستقل (به گونه‌ای برنامه‌ریزی شده‌اند که کارها را تماماً خودشان و به طور مستقل انجام دهند حتی تعمیرات) عمل کرده یا از راه دور توسط یک انسان کنترل شوند. برخی ربات‌ها که با نام اندروئید^۳ شناخته می‌شوند به انسان شباهت دارند ولی تعداد اندکی از ربات‌های صنعتی از این نوع هستند. ربات‌های مستقل مجهز به عوامل هوشمند هوش مصنوعی می‌باشند. ربات‌های هوشمند و پیشرفته‌تر نه تنها مستقل عمل می‌کنند بلکه می‌توانند از محیط پیرامون‌شان یاد گرفته و بر قابلیت‌هایشان بیفزایند. برخی ربات‌های امروزی با مشاهده‌ی انسان‌ها می‌توانند وظایف پیچیده را فرا بگیرند. این منجر به تعامل بهتر بین انسان و ربات می‌شود. اینترآکتیو گروپ در دانشگاه ام آی تی در حال آزمایش این قابلیت از طریق آموزش ربات‌ها

¹ Facial Recognition

² Robotics

³ Androids

برای اتخاذ تصمیمات پیچیده است. برای جزئیات بیشتر نگاه کنید به شاه^۱ (۲۰۱۶). برای مرور کلی بر تکامل ربات‌ها نگاه کنید به واکسیر^۲ (۲۰۱۶).

کارایی فروشگاه‌های وال‌مارت به چینش مناسب قفسه‌های بستگی دارد. استفاده از نیروی انسانی برای بررسی این موضوع بسیار پرهزینه بوده و می‌تواند از دقت لازم برخوردار نباشد. از اواخر ۲۰۱۷، ربات‌ها در تصمیم‌گیری مربوط به انبارش و چینش قفسه‌های شرکت کمک می‌کنند.

در وال‌مارت، ربات‌های نیم متری از یک دوربین/حسگر برای پایش قفسه‌ها و یافتن اقلامی که در جای اشتباه قرار داده شده‌اند، گم شده‌اند یا قیمت آنها اشتباه ثبت شده استفاده می‌کنند. اطلاعات گردآوری شده و تفسیر مسائل و مشکلات توسط ربات‌های خود حرکت‌کننده انجام می‌شوند. نتایج به منظور اتخاذ اقدامات اصلاحی برای انسان‌ها ارسال می‌شود. ربات‌ها وظایف‌شان را سریع‌تر و دقیق‌تر از انسان‌ها انجام می‌دهند. وال‌مارت این موضوع را در ۵۰ فروشگاه خود در سال ۲۰۱۸ آزمایش نمود. نتایج اولیه بسیار مثبت بودند و انتظار نیز می‌رفت که رضایت مشتری را افزایش دهند. ربات‌ها موجب نخواهند شد که کارکنان شغل‌شان را از دست بدهند.

ربات‌ها به طور گسترده‌ای در انبارهای تجارت الکترونیک (مثلاً ده‌ها هزار از آنها در آمازون. کام بکار برده می‌شوند) استفاده می‌شود. از آنها همچنین در ساخت و تولید بر اساس سفارش و تولید انبوه (مثلاً خودروها) و اخیراً وسائل نقلیه‌ی خودران استفاده می‌شود. نسل جدیدی از ربات‌ها برای کار به عنوان مشاور طراحی شده است (فصل ۱۲). این ربات‌ها تا پیش از این در خصوص موضوعاتی از قبیل سرمایه‌گذاری‌ها، سفر، مراقبت سلامت و بهداشت و مسائل حقوقی مشاوره می‌دادند. ربات‌ها می‌توانند به عنوان مسئول پذیرش در پشت میز پذیرش و همچنین معلمان و مربیان خدمت کنند.

ربات‌ها می‌توانند با گردآوری اطلاعات خرید، تطبیق خریدار و محصول و انجام مقایسه‌های قیمت و قابلیت به خرید آنلاین کمک کنند. اینها را شاپ‌بات^۳ می‌نامند (مثلاً نگاه کنید به igi-global.com/dictionary/shopbot/26826). ربات‌ها می‌توانند کالاها را در بازارهای رو باز برای خریداران حمل کنند. وال‌مارت در حال حاضر سبدهای خرید رباتی را آزمایش می‌کند (نایت ۲۰۱۶). برای یک ویدئو نگاه کنید به businessinsider.com/personal-robots-for-shopping-and-e-commerce-2016-9?IR=T. شرکت ژاپنی سافت‌بانک^۴ یک فروشگاه تلفن سلولی را در توکیو افتتاح کرد که تمامی کارکنان آن را ربات‌ها تشکیل می‌دهند و نام همه آنها پپر^۵ است. هر ربات، سیار (چرخ‌دار) بوده و می‌تواند به مشتریان نزدیک شود. در ابتدا، ارتباط با مشتریان از طریق ورود اطلاعات به یک تبلت متصل به هر پپر انجام می‌شد. یکی از مسائل عمده‌ی این ربات‌ها این بود که جایگزین پرسنل انسانی می‌شدند. برای یک بحث در این خصوص به بخش ۱۴.۶ مراجعه کنید.

¹ Shah

² Waxer

³ ShopBot

⁴ SoftBank

⁵ Pepper

پردازش زبان طبیعی (ان‌ال‌پی)^۱ یک فناوری است که به کاربران توانایی برقراری ارتباط با یک رایانه را به زبان مادری‌شان می‌دهد. برقراری ارتباط می‌تواند به صورت متنی و/یا صوتی (گفتار) باشد. این فناوری اجازه برقراری یک نوع رابط کاربری محاوره‌ای را فراهم می‌کند که برخلاف استفاده از یک زبان برنامه‌نویسی که شامل اصطلاحات رایانه‌ای، نحو و دستورات است می‌باشد. ان‌ال‌پی شامل دو زیررشته است:

- درک زبان طبیعی که به بررسی روش‌های توانمندسازی رایانه‌ها برای درک دستورات یا پرس و جوهای طرح شده به زبان انگلیسی معمول یا سایر زبان‌های انسانی است.
- تولید زبان طبیعی که سعی دارد تا رایانه‌ها، زبان محاوره معمول را تولید کنند تا انسان‌ها بتوانند زبان رایانه‌ها را راحت‌تر درک کنند. برای جزئیات بیشتر و تاریخچه‌ی ان‌ال‌پی به en.wikipedia.org/wiki/Natural_language_processing و فصل ۶ مراجعه کنید.

ان‌ال‌پی با داده‌های تولید شده‌ی صوتی و همچنین متنی و سایر شکل‌های ارتباطی مرتبط است. **درک گفتار (صوت)** درک گفتار (صوت) عبارتست از تشخیص و درک زبان‌های گفتاری توسط یک رایانه. کاربردهای این فناوری بسیار گسترش یافته است. به عنوان مثال، بسیاری از شرکت‌ها از این فناوری در مراکز تماس خودکارشان استفاده کرده‌اند. برای یک کاربرد جالب نگاه کنید به cs.cmu.edu/~listen. ترجمه‌ی ماشینی زبان‌ها از موضوعات مرتبط با ان‌ال‌پی است که هم به صورت متن نوشتاری (مثلاً محتوای وب) و هم مکالمه‌ی صوتی انجام می‌شود.

ترجمه‌ی ماشینی زبان‌ها ترجمه‌ی ماشینی از برنامه‌های رایانه‌ای برای ترجمه‌ی واژگان و جملات از یک زبان به زبان دیگر استفاده می‌کند. به عنوان مثال، ترجمه‌ی بابل‌فیش^۲ که از سایت babelfish.com قابل دسترس است بیش از ۲۵ ترکیب مختلف از ترجمه‌ی زبان‌ها را ارائه می‌کند. به طور مشابه، ترجمه‌گر گوگل^۳ (translate.google.com) می‌تواند ده‌ها زبان مختلف را ترجمه کند. نهایتاً، کاربران می‌توانند وضعیت خود را به چندین زبان بر روی فیسبوک قرار دهند.

این شرکت چینی در سال ۲۰۱۸ یک دستگاه مسافرتی پورتابل مجهز به هوش مصنوعی را معرفی نمود. مردم چین امروزه بیش از پیش به سایر کشورها سفر می‌کنند (انتظار می‌رود این تعداد در سال ۲۰۲۰ به ۲۰۰ میلیون نفر در مقابل ۱۲۲ میلیون نفر در سال ۲۰۱۶ برسد). هدف از معرفی این دستگاه این است که گردشگران چینی بتوانند سفرهایشان را برنامه‌ریزی کنند (تا بتوانند وبسایت‌هایی از قبیل تریپ‌آدوایزر^۴ را که به زبان انگلیسی است مطالعه کنند). این دستگاه مسافرتی پورتابل مجهز به هوش مصنوعی به گردشگران امکان خواندن منوها، علائم و نام خیابان‌ها و برقراری ارتباط با صحبت‌کنندگان به زبان بومی را می‌دهد. این دستگاه که از ان‌ال‌پی و تشخیص تصویر استفاده می‌کند به موتور جستجوی سوگو^۵ متصل است. برخلاف فرهنگ واژگان معمولی چینی-انگلیسی، این دستگاه به طور خاص برای مسافران و نیازهای آنها طراحی شده است.

¹ Natural Language Processing (NLP)

² Babel Fish Translation

³ Google's Translate

⁴ Trip Advisor

⁵ Sogou

این سیستم‌ها که در فصل ۱۲ معرفی شده‌اند، برنامه‌هایی رایانه‌ای هستند که دانش را ذخیره می‌کنند که برنامه‌های کاربردی آنها از این دانش برای تولید مشاوره‌های خبره و/یا حل مسئله استفاده می‌کنند. سیستم‌های خبره‌ی دانش‌بنیان همچنین می‌توانند به انسان‌ها در صحنه‌گذاری اطلاعات و اتخاذ انواع خاص تصمیمات روتین و خودکار کمک کنند. سیستم‌های توصیه‌گر (فصل ۱۲)، سیستم‌های دانش‌بنیان هستند که در مورد خرید و سایر مسائل به انسان‌ها توصیه می‌کنند. یک سیستم دانشی دیگر، چت‌بات‌ها هستند (فصل ۱۲).

منابع دانش و اکتساب برای سیستم‌های هوشمند بسیاری از سیستم‌های هوشمند برای اینکه بتوانند کار کنند نیازمند دانش هستند. فرایند کسب این دانش را **اکتساب دانش**^۱ می‌گویند. این فعالیت می‌تواند پیچیده باشد زیرا باید مطمئن شویم که چه دانشی مورد نیاز است. این دانش باید متناسب و سازگار با سیستم مورد نظر باشد. علاوه بر این، منابع دانشی را به منظور حصول اطمینان از امکان‌سنجی اکتساب دانش باید شناسایی نمود. روش‌های خاصی برای اکتساب دانش باید شناسایی گردد و در صورتیکه افراد خبره، منبع دانش هستند، همکاری آنها ضروری است. علاوه بر این، روش بازنمایی دانش و استدلال از دانش گردآوری شده باید مد نظر قرار گیرد و دانش باید صحنه‌گذاری شده و سازگار باشد.

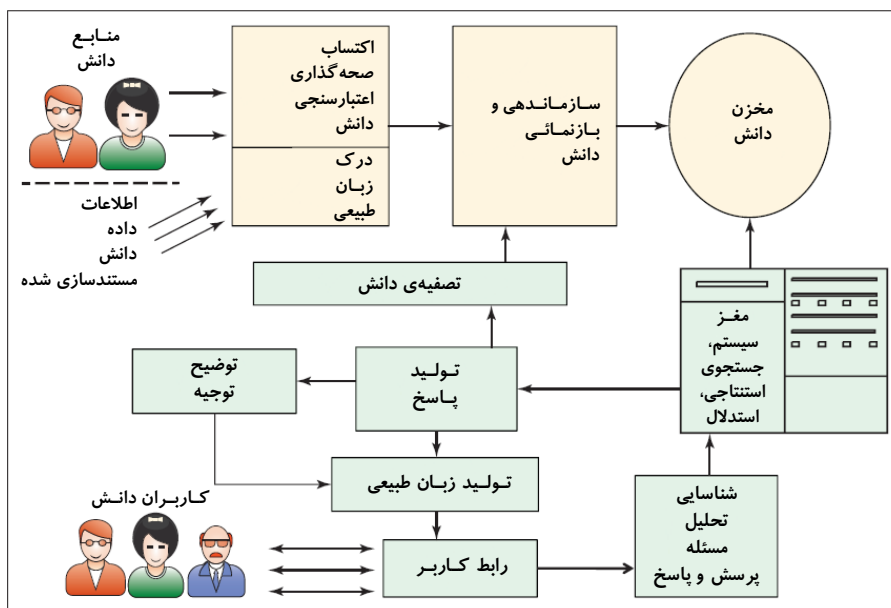
با توجه به این اطلاعات، براحتی می‌توان دید که فرایند اکتساب دانش (شکل ۲.۵) می‌تواند بسیار پیچیده باشد که شامل استخراج و ساختاردهی به دانش است. اینکار چندین روش دارد (مثلاً مشاهده، مصاحبه، سناریونویسی و بحث و گفتگو) که ممکن است برای اکتساب دانش و ایجاد سیستم به مهندس خبره و آموزش دیده دانشی نیاز باشد. در بسیاری از موارد، تیم‌های افراد خبره با مهارت‌های مختلف برای کسب دانش ایجاد می‌شوند. دانش را می‌توان از داده‌ها تولید کرده و سپس از کارشناسان خبره به منظور تأیید آنها استفاده کرد. دانش کسب شده در قالب فعالیتی تحت عنوان **بازنمایی دانش**^۲ سازماندهی می‌شود.

بازنمایی دانش دانش کسب شده باید سازماندهی و ذخیره شود. روش‌های متعددی برای انجام اینکار وجود دارد که به کاربرد دانش، نحوه استدلال از این دانش، نحوه تعامل کاربران با این دانش و ... بستگی دارد. یک راه آسان برای بازنمایی دانش به شکل پرسش و پاسخ است.

استدلال از روی دانش شاید مهم‌ترین مؤلفه در یک سیستم هوشمند، ویژگی استدلال آن باشد. این ویژگی به پردازش درخواست‌های کاربران پرداخته و به کاربران پاسخ می‌دهد (مثلاً راه‌حل، توصیه). تفاوت عمده بین انواع مختلف فناوری‌های هوشمند، نوع استدلالی است که بکار می‌برند.

¹ Knowledge acquisition

² Knowledge representation



شکل ۲.۵ فرایند تصمیم‌گیری خودکار

ربات‌ها به شکل‌ها و انواع مختلفی وجود دارند. یک نوع که در سال‌های اخیر بسیار محبوبیت یافته، چت‌بات می‌باشد. یک چت‌بات که در فصل ۱۲ به معرفی آن خواهیم پرداخت یک ربات محاوره‌ای است که برای گپ زدن با انسان‌ها بکار می‌رود. (بات شکل کوتاه شده‌ی ربات است). بسته به هدفی که از گپ زدن دنبال می‌کنیم که می‌تواند به شکل نوشتاری یا صوتی باشد، بات‌ها می‌توانند به شکل عوامل هوشمند که اطلاعات یا بازیابی می‌کنند یا دستیارهای شخصی که مشاوره می‌دهند وجود داشته باشند. در هر حالت، چت‌بات‌ها معمولاً مجهز به این‌ال‌پی هستند که امکان مکالمه به زبان‌های انسانی طبیعی را به جای زبان رایانه‌ای برنامه‌ریزی شده فراهم می‌کنند. توجه داشته باشید که گوگل شش صدای متفاوت را به دستیار گوگل خود افزوده است.

چندین فناوری هوش مصنوعی جدید در حال ظهور می‌باشند. چند مثال از آنها عبارتند از:

- **رایانش مؤثر.** این فناوری‌ها قادر به تشخیص حالات عاطفی و احساسی افراد بوده و راه‌حلی را برای مسائل و مشکلات شناسایی شده ارائه می‌کنند.
- **تحلیل بیومتریک.** این فناوری‌ها می‌توانند یک هویت را بر اساس ویژگی‌های بیولوژیکی منحصر بفرد که با ویژگی‌های ذخیره شده مقایسه می‌شوند تأیید کنند (مثلاً تشخیص چهره).

رایانش شناختی رایانش شناختی، کاربرد دانش کسب شده از نظریه‌های علوم شناختی (مطالعه‌ی مغز انسان) و علوم رایانه‌ای به منظور شبیه‌سازی فرایندهای تفکر انسانی (یکی از اهداف هوش مصنوعی) است به نحوی که رایانه‌ها بتوانند قابلیت‌های حل مسئله و تصمیم‌گیری را نشان داده و/یا پشتیبانی کنند (فصل ۶). برای این کار، رایانه‌ها باید قادر به استفاده از الگوریتم‌های

خودیادگیرنده^۱، تشخیص الگو، انال پی، بینایی ماشینی و سایر فناوری‌های هوش مصنوعی باشند. آی‌بی‌ام یکی از طرفداران اصلی این مفهوم بوده و فناوری‌هایی را برای پشتیبانی از انسان‌ها در تصمیم‌گیری‌های پیچیده توسعه داده است (مثلاً واتسون). سیستم‌های رایانش شناختی یاد می‌گیرند تا با هدف استدلال کرده و به شکلی طبیعی با دیگران تعامل داشته باشند. برای جزئیات بیشتر به فصل ۵ و مار (۲۰۱۶) مراجعه کنید.

واقعیت افزوده واقعیت افزوده به یکپارچه‌سازی اطلاعات دیجیتالی با محیط پیرامون کاربر (عمدتاً دیداری و صوتی) در لحظه حال اشاره دارد. این فناوری یک تجربه‌ی تعاملی با محیط را برای کاربران میسر می‌کند. از این رو، اطلاعات ممکن است موجب تغییر نحوه کار، یادگیری، بازی، خرید و ارتباطات افراد شود. برنامه‌های پیچیده‌ی واقعیت افزوده ممکن است شامل بینایی ماشینی، تشخیص صحنه و تشخیص حالت باشند. واقعیت افزوده بر روی گوشی‌های آیفون تحت عنوان ای آر کیت^۲ موجود است (متر ۲۰۱۷).

این سیستم‌های واقعیت افزوده از داده‌های اخذ شده توسط حسگرها (مثلاً بینایی، صدا، دما) برای افزودن و ارتقاء به محیط دنیای واقعی استفاده می‌کنند. به عنوان مثال، اگر شما با گوشی همراه‌تان از یک خانه عکس بگیرید بلافاصله می‌توانید اطلاعات عمومی در مورد مالک، ساختمان و بدهی‌های مالیاتی آن را بر روی گوشی‌تان مشاهده کنید.

سوالاتی برای مرور بخش ۲.۴

۱. عوامل هوشمند را تعریف کرده و برخی از قابلیت‌های آنها را فهرست کنید.
۲. فهرستی از کاربردهای عوامل هوشمند تهیه کنید.
۳. یادگیری ماشینی چیست؟ چگونه می‌توان از آن در کسب‌وکار استفاده کرد؟
۴. یادگیری عمیق را تعریف کنید.
۵. رباتیک را تعریف کرده و اهمیت آن در ساخت و تولید و حمل و نقل را توضیح دهید.
۶. انال پی چیست؟ دو قالب اصلی آن کدام هستند؟
۷. ترجمه ماشینی زبان‌ها را توضیح دهید. چرا در کسب‌وکار اهمیت دارد؟
۸. سیستم‌های دانشی چه هستند؟
۹. رایانش شناختی چیست؟
۱۰. واقعیت افزوده چیست؟

تقریباً از زمان شکل‌گیری هوش مصنوعی، پژوهشگران به فرصت‌های استفاده از آن برای پشتیبانی از فرایند تصمیم‌گیری و خودکارسازی کامل تصمیم‌گیری پی برده‌اند. جف بزوس مدیرعامل شرکت آمازون. کام در سال ۲۰۱۷ گفت که هوش مصنوعی در عصر طلایی خود قرار دارد و مشکلاتی را حل می‌کند که زمانی در قلمرو علمی تخیلی بودند (کارپال ۲۰۱۷). بزوس همچنین گفت که آمازون. کام از هوش مصنوعی در تقریباً صدها کاربرد استفاده می‌کند و حقیقتاً کمک شایانی کرده است. آمازون. کام به عنوان مثال بیش از ۲۰ سال است که از هوش مصنوعی برای توصیه‌ی محصولات استفاده می‌کند. این شرکت همچنین از هوش

¹ Self-learning algorithms

² ARKit

مصنوعی برای قیمت‌گذاری محصول و طبق گفته‌ی بزوس برای حل بسیاری از مسائل دشوار و پیچیده استفاده کرده است. علاوه بر این، هوش مصنوعی از زمان شکل‌گیری با حل مسئله و تصمیم‌گیری مرتبط بوده است. فناوری‌های هوش مصنوعی به افراد اجازه می‌دهند تا تصمیمات بهتری بگیرند. واقعیت این است که هوش مصنوعی می‌تواند:

- مسائل پیچیده‌ای که انسان‌ها قادر به حل آنها نیستند را حل کنند (توجه داشته باشید که حل مسئله عمدتاً شامل تصمیم‌گیری می‌باشد).
- تصمیمات سریع‌تری گرفته شود. به عنوان مثال، آمازون میلیون‌ها تصمیم قیمت‌گذاری و توصیه‌ی محصول را در کسری از ثانیه انجام می‌دهد.
- اطلاعات مرتبط را حتی در منابع داده‌ای بزرگ با سرعت بالایی پیدا کند.
- محاسبات پیچیده را سریعاً انجام دهد.
- مقایسه‌ها و ارزیابی‌های پیچیده را در لحظه انجام دهد.

به بیان خلاصه، هوش مصنوعی می‌تواند برخی از تصمیمات را به مراتب سریع‌تر و منسجم‌تر از انسان‌ها اتخاذ کند. برای جزئیات، ویدیوی [youtube.com/watch?v=Dr9jeRy9whQ](https://www.youtube.com/watch?v=Dr9jeRy9whQ) را تماشا کنید. ماهیت تصمیم‌گیری به ویژه تصمیم‌های غیرروتین بسیار پیچیده است. در فصل ۱ بحث کردیم که انواع متعددی تصمیم و سطوح مدیریتی متعددی برای اتخاذ آنها وجود دارد و به یک فرایند معمول تصمیم‌گیری نگاه کردیم. بسیاری از تصمیم‌گیری‌ها که برای حل مسئله بکار می‌روند نیازمند هوش و تخصص هستند. هدف هوش مصنوعی ارائه‌ی هر دوی اینهاست. در نتیجه، روشن است که استفاده از هوش مصنوعی برای تسهیل تصمیم‌گیری، فرصت‌ها، مزایا و واریاسیون‌های بسیاری را در بر می‌گیرد. به عنوان مثال، هوش مصنوعی می‌تواند از انواع تصمیم‌گیری‌ها با موفقیت پشتیبانی کرده و بسیاری دیگر را به طور کامل خودکار کند. در این بخش، ما به بحث و بررسی در خصوص برخی مسائل عمومی پشتیبانی تصمیم هوش مصنوعی می‌پردازیم. این بخش همچنین به تمایز بین پشتیبانی تصمیم‌گیری و خودکارسازی کامل تصمیم‌گیری می‌پردازد.

چندین مسئله وجود دارد که توجیه‌پذیری استفاده از هوش مصنوعی و شانس آن برای موفقیت را تعیین می‌کند. این مسائل عبارتند از:

- ماهیت تصمیم. به عنوان مثال، تصمیمات روتین را به احتمال بیشتری می‌توان خودکارسازی کرد به ویژه اگر تصمیمات ساده‌ای باشند.
- روش پشتیبانی به این معنای که از چه فناوری(هایی) استفاده می‌شود. در ابتدا، پشتیبانی‌های تصمیم خودکار، قاعده محور بودند. در عمل، سیستم‌های خبره به منظور تولید راه‌حل برای موقعیت‌های تصمیم خاص در حوزه‌های تعریف شده ایجاد شدند. یک فناوری محبوب دیگر که قبلاً ذکر شد، «توصیه‌گر» است که همراه با تجارت الکترونیک در دهه ۹۰ پدیدار شد. امروزه، استفاده از یادگیری ماشینی و یادگیری عمیق افزایش یافته است. یک فناوری مرتبط، تشخیص الگو است. امروزه توجه خاصی به انواع تشخیص‌های بیومتریک داده می‌شود.

به عنوان مثال، پژوهش‌ها همچنان به توسعه‌ی یک ماشین هوش مصنوعی ادامه می‌دهند که بتواند با مردم در فرودگاه مصاحبه کرده، یک یا دو سوال پرسیده و سپس تعیین کند که آیا آنها حقیقت را می‌گویند یا خیر. الگوریتم‌های مشابهی را می‌توان برای رسیدگی به امور پناهندگان یا سایر مهاجرها بکار برد.

- **تحلیل‌های ریسک و صرفه-هزینه.** این تحلیل‌ها برای اتخاذ تصمیمات بزرگ ضروری هستند ولی پردازش این مقادیر با استفاده از مدل‌های هوش مصنوعی ممکن است به دلیل دشواری‌های ناشی از مزایا، ریسک‌ها و هزینه‌های اندازه‌گیری کار ساده‌ای نباشد. به عنوان مثال، همانطور که قبلاً گفته شد، پژوهشگران از ۱۰۰ سنجه به منظور اندازه‌گیری سطح هوش جاروبرقی‌ها استفاده کردند.
- **استفاده از قواعد کسب‌وکار.** بسیاری از سیستم‌های هوش مصنوعی مبتنی بر قواعد کسب‌وکار یا سایر انواع قواعد هستند. کیفیت تصمیمات خودکارسازی شده به کیفیت این قواعد بستگی دارد. سیستم‌های هوش مصنوعی پیشرفته قادر به یادگیری و بهبود این قواعد هستند.
- **الگوریتم‌های هوش مصنوعی.** تعداد الگوریتم‌های هوش مصنوعی که پایه و مبنای تصمیمات و پشتیبانی تصمیم خودکار هستند در حال انفجار است. کیفیت این تصمیمات به ورودی الگوریتم‌ها بستگی دارد که ممکن است تحت تأثیر تغییرات محیط کسب‌وکار باشند.
- **سرعت.** خودکارسازی تصمیم نیز به سرعت تصمیم‌گیری بستگی دارد. برخی تصمیمات را نمی‌توان به دلیل زمان زیاد مورد نیاز برای اخذ تمامی داده‌های ورودی به صورت خودکار درآورد. از طرف دیگر، تصمیمات دستی ممکن است برای برخی شرایط خاص بسیار کند باشند.

اکثر پشتیبانی‌های هوش مصنوعی را می‌توان امروزه برای گام‌های مختلف فرایند تصمیم‌گیری بکار گرفت. تصمیمات کاملاً خودکار در موقعیت‌های روتین متداول هستند و در بخش بعدی در مورد آنها بحث خواهیم کرد. ما در اینجا از گام‌های فرایند تصمیم‌گیری که در فصل ۱ توضیح دادیم پیروی می‌کنیم.

شناسایی مسئله سیستم‌های هوش مصنوعی به طور گسترده‌ای در شناسایی مسئله و عمدتاً عیب‌یابی تجهیزات و مسائل پزشکی، یافتن رخنه‌های امنیتی، برآورد کردن سلامت مالی و غیره بکار می‌روند. از چندین فناوری برای این منظور استفاده می‌شود. به عنوان مثال، الگوریتم‌های هوش مصنوعی از داده‌های گردآوری شده توسط حسگرها استفاده می‌کنند. سطح عملکرد ماشین‌آلات با سطح استاندارد مقایسه شده و تحلیل روند می‌تواند به فرصت‌ها یا مشکلات اشاره کند.

ایجاد یا یافتن راه‌حل‌های جایگزین فناوری‌های متعدد هوش مصنوعی با تطبیق دادن مشخصه‌های مسئله با بهترین تجارب یا راه‌حل‌های اثبات شده‌ای که در پایگاه‌های داده ذخیره شده‌اند، راه‌حل‌های جایگزین را ارائه می‌کنند. هم سیستم‌های خبره و هم چت‌بات‌ها از این روش استفاده می‌کنند. آنها می‌توانند راه‌حل‌های توصیه شده را ایجاد کرده یا گزینه‌های متعددی را برای انتخاب ارائه کنند. ابزارهای هوش مصنوعی از قبیل استدلال مبتنی بر مورد و رایانش عصبی برای این مقصود بکار می‌روند.

انتخاب یک راه‌حل از مدل‌های هوش مصنوعی به منظور ارزیابی راه‌حل‌های پیشنهادی، به عنوان مثال، پیش‌بینی تأثیر آنها در آینده (تحلیل پیشگویانه)، ارزیابی احتمال موفقیت آنها یا پیش‌بینی واکنش یک سازمان به اقدام اتخاذ شده توسط رقیب استفاده می‌شود.

پایاده‌سازی راه‌حل‌ها از هوش مصنوعی می‌توان برای پشتیبانی اجرای راه‌حل‌های پیچیده استفاده کرد. به عنوان مثال، می‌توان از آن برای نمایش برتری و ارجحیت پیشنهادها و ارزیابی مقاومت در برابر تغییرات استفاده کرد. شرکت‌ها با بکارگیری هوش مصنوعی بر یک یا چند فرایند و گام تصمیم‌گیری قادر به حل مسائل پیچیده‌ی دنیای واقعی خواهند بود که در مورد کاربردی ۲.۳ نشان داده شده است.

با افزایش توان فناوری‌های هوش مصنوعی، توانایی آن برای خودکارسازی موقعیت‌های تصمیم‌گیری پیچیده‌تر نیز افزایش می‌یابد.

شرکت‌ها چگونه با استفاده از ابزارهای یادگیری ماشینی گوگل مسائل دنیای واقعی خود را حل می‌کنند؟

مورد کاربردی ۲.۳

مثال‌های زیر از فارست (۲۰۱۷) اقتباس شده‌اند:

موتور یادگیری ماشینی ابری گوگل^۱ و تِنسور فلو^۲ امکان دسترسی منحصر بفرده به ابزارهای یادگیری ماشینی را بدون نیاز به دانشمندان درجه دکتراى رشته‌ی داده‌ها فراهم می‌کند.

شرکت‌های زیر از ابزارهای گوگل برای حل مسائل استفاده می‌کنند.

۱. آکسا اینترنشنال^۳. این شرکت بین‌المللی بیمه از یادگیری ماشینی برای پیش‌بینی اینکه کدام رانندگان به احتمال بیشتری موجب بروز سوانح بزرگ خواهند شد استفاده می‌کند. دقت پیش‌بینی ۷۸ درصد است. از این پیش‌بینی برای تعیین حق بیمه‌های مناسب استفاده می‌شود.

۲. ایرباس دیفنس اند اسپیس^۴. تشخیص ابرها در تصاویر ماهواره‌ای برای چندین دهه‌ها به صورت دستی انجام می‌شد. این فرایند با استفاده از یادگیری ماشینی ۴۰ درصد سریع‌تر شده و نرخ خطا نیز از ۱۱ درصد به ۳ درصد کاهش یافته است.

۳. پیشگیری از صید بی‌رویه در جهان. یک ارگان دولتی تا پیش از این فقط مناطق کوچکی را برای یافتن متخلفان صیادی زیر نظر داشت. در حال حاضر با استفاده از موقعیت‌یابی ای‌آی‌اس ماهواره‌ای، این ارگان می‌تواند کل اقیانوس را رصد نماید. این ارگان با استفاده از یادگیری ماشینی قادر است تا تمامی کشتی‌های صیادی را برای یافتن متخلفان ردیابی نماید.

۴. شناسایی کلاهبرداری کارت اعتباری در ژاپن. یک شرکت خدمات مالی ژاپنی به نام اس‌ام‌اف‌جی^۵ از یادگیری ماشینی گوگل (یک کاربرد یادگیری عمیق) برای رصد کلاهبرداری‌های مربوط به کارت‌های اعتباری با دقت تشخیص ۸۰-۹۰ درصد استفاده می‌کند. به محض تشخیص، یک هشدار برای اتخاذ اقدامات لازم صادر می‌شود.

¹ Google's Cloud Machine Learning Engine

² Tensor Flow

³ Axa International

⁴ Airbus Defense & Space

⁵ SMFG

۵. شرکت غذایی کیوپای^۱ ژاپن. این شرکت تکه‌های معیوب سیب‌زمینی را به صورت دستی و با استفاده از یک فرایند پرهزینه و کند شناسایی می‌نمود. اکنون با استفاده از ابزارهای هوش مصنوعی گوگل این شرکت قادر است تا تصاویر ویدئویی را به صورت خودکار نظارت کرده و بازرسان را برای حذف سیب‌زمینی‌های معیوب مطلع سازد.

Source: Condensed and compiled from Forrest (2017).

سوالاتی برای مورد کاربردی ۲.۳

۱. چرا از یادگیری ماشینی برای پیش‌بینی‌ها استفاده می‌شود؟
۲. چرا از یادگیری ماشینی برای شناسایی‌ها استفاده می‌شود؟
۳. در این پنج مورد از چه تصمیمات مشخصی پشتیبانی شده است؟

پشتیبانی تصمیم هوشمند و خودکار از اوایل دهه ۷۰، تلاش‌هایی به منظور خودکارسازی تصمیم‌گیری صورت گرفت. این تلاش‌ها عموماً با استفاده از سیستم‌های خبره‌ی مبتنی بر قاعده انجام می‌شدند که راه‌حلی را برای مسائل و مشکلات تکرارپذیر مدیریتی توصیه می‌کردند. مثال‌هایی از تصمیماتی که به صورت خودکار اتخاذ می‌شدند به شرح زیر می‌باشند:

- تأیید وام‌های کوچک
- غربالگری اولیه‌ی متقاضیان مشاغل
- تأمین مجدد موجودی‌های ساده
- قیمت‌های محصولات و خدمات (زمان و نحوه‌ی تغییر آنها)
- توصیه‌ی محصول (مثلاً در **آمازون.کام**)

فرایند خودکارسازی تصمیم‌گیری در شکل ۲.۵ به تصویر کشیده شده است. این فرایند با اکتساب دانش و ایجاد یک مخزن دانش شروع می‌شود. کاربران، سوالات خود را برای مغز سیستم ارسال می‌کنند که پاسخی را برای آن سوال ایجاد کرده و آن را برای کاربران ارسال می‌کند. علاوه بر این، راه‌حل‌ها مورد ارزشیابی قرار می‌گیرند تا مخزن دانشی و استدلال‌های سیستم را بتوان بهبود بخشید. موقعیت‌های پیچیده برای عوامل انسانی ارسال می‌شوند. این فرایند به ویژه در سیستم‌های دانش‌بنیان استفاده می‌شود. توجه داشته باشید که فرایند شکل ۲.۵ برای اکتساب دانش، تصمیم‌گیری خودکار را نیز نمایش می‌دهد. شرکت‌ها از تصمیم‌گیری خودکار هم برای عملیات برون‌سازمانی (مثلاً فروش) و هم عملیات درون‌سازمانی (مثلاً تخصیص منابع، مدیریت موجودی) خود استفاده می‌کنند.

طی مطالعه‌ای که در یک بیمارستان در تایوان (لیائو و همکاران، ۲۰۱۵) انجام پذیرفته، بکارگیری هوش مصنوعی به منظور تولید تشخیص‌های پرستاران مورد بررسی قرار گرفته و با تشخیص‌های انجام شده توسط انسان‌ها مورد مقایسه قرار گرفته است.

¹ Kewpie Food of Japan

تشخیص‌ها نیازمند دانش جامع و کامل، تجربه‌ی بالینی و حس‌گریزی بودند. پژوهشگران از چندین ابزار هوش مصنوعی از جمله یادگیری ماشینی برای داده‌کاوی و تحلیل به منظور پیش‌بینی احتمال موفقیت تشخیص‌های خودکار بر اساس ویژگی‌های بیمار استفاده کردند. نتایج بدست آمده نشان می‌دهد که بین تصمیمات تشخیص‌ انسانی و هوش مصنوعی ۸۷ درصد تطابق وجود دارد. اینگونه فناوری‌ها را می‌توان در مکان‌هایی که فاقد پرسنل پرستاری انسانی هستند و همچنین پرسنل پرستاری که می‌خواهند دقت پیش‌بینی‌های تشخیصی خود را صحت‌گذاری کنند بکار گرفت. این سیستم همچنین می‌تواند آموزش پرسنل پرستاری را تسهیل نماید. تصمیمات خودکار به شکل‌های مختلفی وجود دارند که در بینش فناوری ۲.۲ به تصویر کشیده شده است.

شک نداریم که هوش مصنوعی قادر است فرایند تصمیم‌گیری را برای کسب‌وکارها تغییر دهد. به عنوان مثال نگاه کنید به سینکاواج^۱ (۲۰۱۷). ماهیت تغییر، بسته به شرایط، متغیر است. ولی به طور کلی انتظار داریم که هوش مصنوعی تأثیر اساسی برای اتخاذ تصمیمات بهتر، سریع‌تر و مؤثرتر داشته باشد. توجه داشته باشید که در برخی موارد، به یک دیده‌بان هوش مصنوعی به منظور تنظیم فرایند نیاز داریم (سمپل ۲۰۱۷)^۲.

سوالاتی برای مرور بخش ۲.۵

۱. وجه تمایز بین تصمیم‌گیری کاملاً خودکار و تصمیم‌گیری تحت پشتیبانی در چیست؟
۲. فهرستی از مزایای هوش مصنوعی برای پشتیبانی تصمیم تهیه کنید.
۳. چه عواملی بر بکارگیری هوش مصنوعی در پشتیبانی تصمیم تأثیر می‌گذارند؟
۴. ارتباط بین هوش مصنوعی و گام‌های فرایند کلاسیک تصمیم‌گیری چیست؟
۵. شرایط لازم برای اینکه هوش مصنوعی قادر به خودکارسازی تصمیم‌گیری باشد چه هستند؟
۶. چهار مدل اسکراج را توضیح دهید.

بینش فناوری ۲.۲ مدل‌های اسکراج برای بکارگیری هوش مصنوعی در تصمیم‌گیری

اسکراج^۳ (۲۰۱۷) از مدرسه‌ی اسلون‌ام‌آی‌تی^۴ چهار مدل زیر را برای بکارگیری هوش مصنوعی در تصمیم‌گیری مستقل کسب‌وکارها پیشنهاد کرده است:
۱. مشاور مستقل^۵. این یک مدل مدیریت داده‌محور است که از الگوریتم‌های هوش مصنوعی به منظور تولید بهترین راهبردها و دستورالعمل‌ها در خصوص اینکه چه کاری انجام شود استفاده کرده و توصیه‌های مشخص و خاصی را ارائه می‌نماید. با وجود این، فقط انسان‌ها می‌توانند این توصیه‌ها را تأیید کنند (مثلاً راه‌حل‌های پیشنهادی).
اسکراج مثالی را ارائه کرده که در آن یک شرکت خرده‌فروش آمریکایی، یک ماشین هوش مصنوعی را جایگزین کل دپارتمان بازاریابی و فروش نموده و به کارکنان دستور داده تا از دستورات آن پیروی کنند. مشخصاً، این کار مقاومت و رنجش به همراه داشته است. شرکت برای اطمینان از انطباق، نرم‌افزارهای مانیتورینگ و ممیزی نصب کرده است.

¹ Sincavage

² Sample

³ Schrage

⁴ MIT's Sloan School

⁵ The Autonomous Advisor

۲. برون‌سپاری مستقل^۱! در اینجا، مدل سنتی برون‌سپاری فرایند کسب‌وکار به یک الگوریتم فرایند کسب‌وکار تغییر یافته است. به منظور خودکارسازی این فرایند، ایجاد قواعد و دستورالعمل‌های روشن و شفاف امری ضروری است. از آنجائیکه این فرایند شامل تخصیص منابع می‌باشد سناریوی پیچیده‌ای محسوب می‌شود. قابلیت پیش‌بینی و قابلیت اطمینان صحیح در اینجا ضرورت دارد.
۳. همکاری انسان-ماشین^۲. با این فرض که در این مدل، الگوریتم‌ها بتوانند بهترین تصمیمات را اخذ کنند، انسان‌ها باید با ماشین‌های هوشمند، تماماً خودکار ولی محدود همکاری کنند. برای حصول اطمینان از این همکاری، لازم است تا انسان‌ها برای کار با ماشین‌های هوش مصنوعی آموزش ببینند. غول‌های فناوری مثل تفلیکس، علی بابا و گوگل از این مدل استفاده می‌کنند.
۴. استقلال کامل ماشین^۳. در این مدل، سازمان‌ها فرایندهایشان را به طور کامل خودکارسازی کرده‌اند. مدیریت می‌بایست کاملاً به مدل‌های هوش مصنوعی اعتماد کند که البته اینکار ممکن است چندین سال به طور بینجامد. اسکراج مثالی از یک صندوق پوشش ریسک ارائه کرده که مکرراً بر اساس توصیه‌های یک ماشین مبادله می‌کند. این شرکت از یادگیری ماشینی برای آموزش دادن الگوریتم‌های مبادله استفاده می‌کند.

پیاده‌سازی و اجرای این چهار مدل نیازمند مشارکت، همکاری و رهبری مدیریت با دانشمندان علم داده‌ها است. برای توصیه در خصوص چگونگی انجام این کار با اسکراج (۲۰۱۷) که چندین کتاب در این رابطه به رشته‌ی تحریر درآورده مشورت نمایید. کایران (۲۰۱۷) در مورد اینکه چرا مدیران باید استفاده از هوش مصنوعی برای پشتیبانی تصمیم را مد نظر قرار دهند بحث می‌کند. یک نکته جالب این است که رقابت بین شرکت‌ها در واقع مابین الگوریتم‌های مستقل داده‌محور و مدل‌های کسب‌وکار مربوطه رخ خواهد داد.

سوالاتی برای مرور بخش ۲.۵

۱. وجه تمایز بین مشاور مستقل و همکاری بین انسان و ماشین چیست؟
۲. در تمامی چهار مدل، تا حدودی تعامل بین انسان و ماشین وجود دارد. بحث کنید.
۳. چرا استفاده از مدل ۴ برای تصمیمات سرمایه‌گذاری آسان‌تر از مثلاً راهبردهای بازاریابی است؟
۴. چرا همکاری مدیریت ارشد با دانشمندان علم داده‌ها در ماشین‌های هوش مصنوعی مستقل اهمیت دارد؟

در سرتاسر این کتاب، ما مثال‌های متعددی از کاربردهای هوش مصنوعی در کسب‌وکار، خدمات و دولت را ارائه کرده‌ایم. در پنج بخش بعدی، ما کاربردهای بیشتری از هوش مصنوعی را در حوزه‌های رایج کسب‌وکار از قبیل حسابداری، تأمین مالی، مدیریت منابع انسانی، بازاریابی، تبلیغات و مدیریت ارتباط با مشتریان و مدیریت تولید-عملیات ارائه خواهیم کرد.

چاندی (۲۰۱۷) مدیرعامل شرکت نرم‌افزارهای حسابداری اسلیک‌پای برای کسب‌وکارهای کوچک متوجه روندهایی در بین حسابداری حرفه‌ای گردید: استفاده از هوش مصنوعی از جمله بات‌ها در کارهای روتین حرفه‌ای آنها افزایش یافته است. چاندی مشاهده کرد که محرک‌های اصلی این روند، صرفه‌جویی در زمان و پول و افزایش دقت و بهره‌وری است. بکارگیری هوش مصنوعی

¹ The Autonomous Outsource

² People-Machine Collaboration

³ Complete Machine Collaboration

بسیار سریع بوده و بهبودهای چشمگیری را به همراه داشته است. یک مثال، اجرای رویه‌های انطباق است که به عنوان مثال، ارنست آند یانگ (ای‌ای‌وای)^۱ از یادگیری ماشین برای تشخیص داده‌های مغایر استفاده می‌کند (مثلاً صورتحساب‌های مشکوک).

کاربران اصلی هوش مصنوعی، شرکت‌های بزرگ حسابداری و مالیاتی هستند که در مورد کاربرد ۲.۴ نشان داده شده‌اند.

مؤسسات حسابداری کوچک نیز از هوش مصنوعی استفاده می‌کنند. به عنوان مثال، کراو هوروات در شیکاگو از هوش مصنوعی برای حل مسائل پیچیده‌ی صدور صورتحساب در صنعت مراقبت از سلامت و بهداشت استفاده می‌کند. این به مراجعین کمک می‌کند تا به به ادعاها و بازپرداخت پول‌هایشان رسیدگی کنند. مؤسسه اکنون می‌تواند مسائل دشواری را که قبلاً راه‌حلی نداشتند حل کند. کاربردهای بسیار دیگری نیز با پشتیبانی هوش مصنوعی استفاده می‌شوند از جمله تحلیل قراردادهای املاک و مستغلات تا تحلیل ریسک. اینکه مؤسسات کوچک نیز به بهره‌برداری از هوش مصنوعی بپردازند صرفاً مستلزم زمان است.

چگونه ای‌وای، دلوثیت و پی‌دبلیو‌سی از هوش مصنوعی استفاده می‌کنند؟

مورد کاربردی ۲.۴

شرکت‌های حسابداری بزرگ از هوش مصنوعی استفاده کرده یا آن را جایگزین فعالیت‌های انسانی پشتیبان در وظایفی از قبیل آماده‌سازی مالیات، حسابرسی، مشاوره استراتژی، خدمات حسابداری نموده‌اند. آنها عمدتاً از ان‌ا‌ل‌پی، اتوماسیون فرایند رباتیک، متن‌کاوی و یادگیری ماشین استفاده می‌کنند. با وجود این، آنها از استراتژی‌های متفاوتی طبق توضیح ژو (۲۰۱۷) استفاده می‌کنند:

- ای‌وای سعی دارد تا نرخ بازگشت سرمایه‌ی سریع و مثبتی را در یک مقیاس کوچک داشته باشد. استراتژی این شرکت متمرکز بر ارزش کسب‌وکار است. ای‌وای از هوش مصنوعی به عنوان مثال برای بازنگری اسناد قانونی مرتبط با لیزینگ استفاده می‌کند (مثلاً برای اطمینان از تطابق با قوانین و مقررات دولتی جدید).
- شرکت پرایس واتر‌هاوس کوپرز (پی‌دبلیو‌سی) از پروژه‌های کوچکی نفع می‌برد که در عرض چهار هفته بتوان آنها را عملیاتی نمود. هدف از اینکار، نشان دادن ارزش هوش مصنوعی به شرکت‌های خدمت‌گیرنده است. به محض نمایش این پروژه‌ها به مشتریان، تصفیه حساب صورت می‌پذیرد. پس دبلیو‌سی سالانه ۷۰ تا ۸۰ پروژه را انجام می‌دهد.
- دلوثیت توشی توماتسو که به اختصار دلوثیت نامیده می‌شود، مواردی را برای هدایت پروژه‌های هوش مصنوعی هم برای کاربردهای درون‌سازمانی و هم برون‌سازمانی ایجاد می‌کند. هدف از اینکار، تسهیل فرایند نوآوری است. یکی از حوزه‌های موفق، بکارگیری ان‌ا‌ل‌پی برای بازنگری قراردادهای بزرگی است که می‌تواند شامل صدها هزار سند قانونی باشد. این شرکت، زمان بازنگری را از شش ماه به کمتر از یک ماه کاهش داد و همچنین تعداد کارمندانی که مسئول انجام اینکار بودند را تا ۷۰ درصد کاهش داد. دلوثیت همانند رقبایش از هوش مصنوعی برای ارزشیابی هم‌افزایی‌های تدارکات بالقوه

¹ Ernst & Young (EY)

برای تصمیمات ادغام و اکتسابها استفاده می‌کند. اینگونه ارزشیابی‌ها فعالیت بسیار زمان‌بری هستند زیرا لازم است تا حجم بسیار زیادی از داده‌ها (برخی اوقات میلیون‌ها خط داده) بررسی شود. در نتیجه، دلوثیت می‌تواند این ارزشیابی را در کمتر از یک هفته به پایان برساند در حالیکه تا پیش از این به چهار تا پنج ماه زمان نیاز بود. دلوثیت گفته است که به کمک هوش مصنوعی، داده‌ها را به نحوی مشاهده می‌کند که تا پیش از این حتی تصور آن را نمی‌کرده است (آپاسکا-فیو ۲۰۱۷).

تمامی شرکت‌های حسابداری بزرگ از هوش مصنوعی برای کمک به تولید گزارشات و انجام بسیاری از وظایف و فعالیت‌های سنگین و روتین استفاده می‌کنند. هوش مصنوعی توانسته کیفیت کارها را افزایش داده و دقت کارها را با مرور زمان بهتر و بهتر کند.

Sources: Compiled from Chandi (2017), Zhou (2017), and Ovaska-Few (2017).

سوالاتی برای مورد کاربردی ۲.۴

۱. مشخصه‌های کارها و وظایفی که برای انجام آنها از هوش مصنوعی استفاده می‌شود چیست؟
۲. چرا شرکت‌های حسابداری بزرگ از استراتژی‌های متفاوتی برای پیاده‌سازی و اجرا استفاده می‌کنند؟
۳. در این پنج مورد از چه تصمیمات مشخصی پشتیبانی شده است؟

مطالعه‌ی جامع کاربرد هوش مصنوعی در حسابداری دانشکده‌ی فناوری اطلاعات آی سی ای ایی دلبلیو مطالعه‌ی جامع رایگانی را

با عنوان «هوش مصنوعی و آینده‌ی حسابداری» ارائه کرده است. مزایای هوش مصنوعی بر اساس این گزارش به شرح زیر است:

- ارائه‌ی داده‌های ارزان‌تر و بهتر به منظور پشتیبانی تصمیم‌گیری و حل مسائل حسابداری
 - ایجاد بینش از طریق تحلیل داده‌ها
 - آزادسازی زمان حسابداران به منظور تمرکز بر حل مسئله و تصمیم‌گیری
- در این گزارش به استفاده از موارد زیر نیز اشاره شده است:
- یادگیری ماشین برای شناسایی کلاهبرداری و تقلب و پیش‌بینی فعالیت‌های مشکوک به کلاهبرداری
 - یادگیری ماشین و سیستم‌های دانش‌بنیان برای صحت‌گذاری وظایف حسابداری
 - یادگیری عمیق برای تحلیل داده‌های ساخت‌نیافته از قبیل داده‌های موجود در قراردادهای و پست‌های الکترونیکی

بسیاری از وظایف روتین امروزی حسابداری توسط هوش مصنوعی و تحلیل‌ها خودکارسازی خواهد شد (فصل ۱۴) و بسیاری از آنها شغل‌های خود را از دست خواهند داد. از طرف دیگر، حسابداری باید سیستم‌های حسابداری مبتنی بر هوش مصنوعی را مدیریت کنند. نهایتاً، حسابداری برای موفقیت یا حتی بقا می‌بایست به سمت نوآوری در هوش مصنوعی بروند (واراوا ۲۰۱۷)^۱.

سوالاتی برای مرور بخش ۲.۶

۱. علل اصلی استفاده از هوش مصنوعی در حسابداری چه هستند؟
۲. برخی از کاربردهای هوش مصنوعی در سازمان‌های بزرگ حسابداری را نام ببرید؟

۳. چرا مؤسسات حسابداری بزرگ، پیشگام در استفاده از هوش مصنوعی هستند؟
۴. برخی مزایای استفاده از هوش مصنوعی که در گزارش آی سی ای ایی دلبلیو آمده چه هستند؟
۵. مشاغل حسابداران چگونه ممکن است تحت تأثیر هوش مصنوعی قرار بگیرد؟

خدمات مالی از تنوع بسیاری بالایی برخوردار بوده و به همین دلیل، کاربرد هوش مصنوعی نیز در این حوزه بسیار متنوع است. یک روش برای سازماندهی فعالیت‌های هوش مصنوعی، بر اساس بخش‌های اصلی خدمات است. در این بخش، ما فقط در مورد دو بخش بحث خواهیم کرد: بانکداری و بیمه.

- سینگ^۱ (۲۰۱۷) متوجه شده است که فعالیت‌های زیر در انواع مختلف خدمات مالی وجود دارند:
- شخصی‌سازی فوق‌العاده زیاد (مثلاً استفاده از چَت‌بات‌ها، دستیارهای شخصی و مشاوران سرمایه‌گذاری رباتی) (فصل ۱۲)
 - تغییر رفتار مشتری هم در شاخه‌های آنلاین و هم فیزیکی
 - تسهیل اعتماد به هویت دیجیتال
 - ایجاد انقلاب در پرداخت‌ها
 - به اشتراک‌گذاری فعالیت‌های اقتصادی (مثلاً وام‌های شخص به شخص)
 - ارائه خدمات مالی به صورت ۲۴ ساعته و ۷ روز هفته و در سطح جهانی

Consultancy.uk (2017) مروری بر نحوه‌ی تغییر شکل و دگرگونی صنعت بانکداری به واسطه‌ی هوش مصنوعی ارائه کرده است. این مؤسسه پی برد که کاربردهای هوش مصنوعی، اکثراً در حوزه‌های فناوری اطلاعات، مالی و حسابداری، بازاریابی و فروش، مدیریت منابع انسانی، خدمات مشتریان و عملیات است. یک نظرسنجی جامع در خصوص بکارگیری هوش مصنوعی در بانکداری در سال ۲۰۱۷ انجام شد و گزارش آن جهت خرید موجود است (تیوان ۲۰۱۷)^۲. یافته‌های اصلی این گزارش به شرح زیر می‌باشند:

- فناوری‌های هوش مصنوعی در بانکداری شامل تمامی موارد فهرست شده در بخش ۲.۷ و چند ابزار تحلیل دیگر (فصل‌های ۳ تا ۱۱ این کتاب).
- این فناوری‌ها به بانک‌ها کمک می‌کنند تا عملیات مدیران و پشت صحنه عملیاتی را بهبود بخشند.
- فعالیت‌های مهم عبارتند از بکارگیری چَت‌بات‌ها برای بهبود خدمات مشتری و برقراری ارتباط با مشتریان (فصل ۱۲) و مشاوره‌های رباتی که در برخی مؤسسات مالی استفاده می‌شوند.
- از تشخیص چهره برای بانکداری آنلاین ایمن‌تر استفاده می‌شود.

¹ Singh

² Tiwan

- تحلیل‌های پیشرفته در تصمیمات سرمایه‌گذاری به مشتریان کمک می‌کنند. برای مثال‌ها به نوردرام^۱ (۲۰۱۷)، ای‌وی استاف^۲ (۲۰۱۷) و آگراوال^۳ (۲۰۱۸) مراجعه کنید.
- الگوریتم‌های هوش مصنوعی به بانک‌ها در شناسایی و مسدود کردن فعالیت‌های مشکوک به کلاهبرداری از جمله پولشویی کمک می‌کنند.
- الگوریتم‌های هوش مصنوعی می‌توانند به ارزیابی خوش‌حسابی متقاضیان وام کمک کنند (برای یک مطالعه‌ی موردی از کاربرد هوش مصنوعی در پایش اعتبار به ai-toolkit.blogspot.com/2017/01/case-study-artificial-intelligence-in.html مراجعه کنید).

مؤسسات بانکداری زیر از هوش مصنوعی استفاده می‌کنند:

- بانک‌ها از ماشین‌های هوش مصنوعی از قبیل واتسون آی‌بی‌ام برای نظارت بر کارکنان استفاده می‌کنند. این موضوع در پیشگیری از فعالیت‌های غیرقانونی از قبیل مواردیکه در شرکت بانکداری و خدمات مالی ولز فارگو رخ داد بسیار حائز اهمیت است. برای جزئیات به information-management.com/articles/banks-using-algothimsto-set-up-employee-surveillance مراجعه کنید.
- بانک‌ها از برنامه‌های کاربردی برای آماده‌سازی مالیات استفاده می‌کنند. ایچ آند آر بلاک^۴ از واتسون آی‌بی‌ام برای مرور و بازنگری اظهارنامه‌های مالیاتی استفاده می‌کنند. این برنامه اطمینان حاصل می‌کند که افراد فقط مبلغی را که بدهکار هستند پرداخت کنند. این ماشین با استفاده از مکالمات تعاملی سعی می‌کند تا صورتحساب‌های مالیاتی افراد کاهش یابد.
- پاسخگویی به تعداد زیادی از اعلام‌ها در آن واحد. به عنوان مثال، شرکت رین‌برد^۵ یک فروشنده‌ی هوش مصنوعی است که به ماشین‌ها آموزش می‌دهد تا به سوالات مشتریان پاسخ دهند. سوالات مشتریان که به میلیون‌ها عدد می‌رسد موجب مشغولیت وقت کارکنان بانک می‌شود. بات‌ها به کارکنان ستاد بانک کمک می‌کنند تا پاسخ‌های مناسب برای اعلام‌ها را سریعاً پیدا کنند. این موضوع به ویژه در بانک‌هایی که نرخ گردش کارکنان آنها بالا است اهمیت دارد. همچنین، به دلیل تغییرات مکرر در سیاست‌ها و مقررات، دانش موجود نیز در طول زمان کاهش می‌یابد.
- رین‌برد به صورت یکپارچه با واتسون آی‌بی‌ام درآمده که از قابلیت‌های هوش مصنوعی و استدلال شناختی برای درک ماهیت اعلام‌ها و ارائه جواب یا راه‌حل استفاده می‌کند. مکالمات بین برنامه و کارمندان از طریق چت‌بات‌های شرکت رین‌برد انجام می‌شود که در کلیه‌ی شعب بانکی سرتاسر انگلستان سرویس‌دهی می‌کنند.
- در بانک کپیتال وان^۶ و چند بانک دیگر، مشتریان می‌توانند برای پرداخت صورتحساب‌ها از طریق کارت اعتباری و چک کردن حساب‌هایشان با آلکسا آمازون صحبت کنند.

¹ Nordrum

² E. V. Staff

³ Agrawal

⁴ H&R Block

⁵ Rainbird Co. (rainbird.ai)

⁶ Capital One

- تی دی بانک^۱ و برخی بانک‌های دیگر (یوکران ۲۰۱۷) برنامه‌ی آکسا را آزمایش می‌کنند که قابلیت‌های یادگیری ماشین و واقعیت افزوده را برای پاسخگویی به استعلام‌ها ارائه می‌کند.
 - بانک دانامون^۲ از یادگیری ماشین برای تشخیص فعالیت‌های کلاهبرداری و تقلب و فعالیت‌های ضد پولشویی استفاده کرده و تجربه‌ی مشتری را بهبود می‌بخشد.
 - مشتریان در اچ‌اس‌بی‌سی^۳ برای یافتن اطلاعات در مورد حساب‌هایشان و اطلاع یافتن از موارد امنیتی می‌توانند با دستیار مجازی بانکداری به نام اولیویا^۴ صحبت کنند. اولیویا می‌تواند از تجارب خود یاد گرفته و مفیدتر شود.
 - بانک سانتاندر^۵ از یک دستیار مجازی به نام نینا^۶ برای انتقال پول، پرداخت صورتحساب‌ها و کارهای دیگر استفاده می‌کند. نینا همچنین می‌تواند مشتریان را از طریق یک سیستم تشخیص صدای هوش مصنوعی احراز هویت کند.
 - لوو^۷ از شرکت آر بی‌اس^۸ یک بات مدیریت خدمات مشتری و ارتباط با مشتریان است که به سوالات مشتریان پاسخ می‌دهد.
 - در آکسینچر^۹، مشاور مجازی رهن و اجاره به نام کولت^{۱۰} به ارائه مشاوره شخصی می‌پردازد.
 - یک ربات به نام نائو^{۱۱} قادر است تا حالت چهره و رفتار مشتریانی که وارد شعب برخی بانک‌های خاص می‌شوند را تحلیل کرده و ملیت آنها را تعیین کند. آنگاه زبان مناسب را برای برقراری ارتباط با آن مشتری انتخاب می‌نماید (ژاپنی، چینی یا انگلیسی).
- واتسون آی‌بی‌ام، خدمات بسیار زیاد دیگری از مبارزه با جرم و جنایت گرفته تا انطباق با قوانین و مقررات را به بانک‌ها ارائه می‌کند.

مقررات دولتی، مسئولیت‌های سنگینی را بر عهده‌ی بانک‌ها و سایر مؤسسات مالی نهاده‌اند. بانک‌ها به منظور انطباق با این قوانین و مقررات باید روزانه زمان بسیار زیادی را صرف ارزیابی و بررسی حجم داده‌های تولید شده نمایند. واتسون آی‌بی‌ام که توسط گروه مالی پرومونتوری (یکی از شرکت‌های زیرمجموعه آی‌بی‌ام) توسعه یافته است مجموعه‌ای از ابزارها را برای حل مشکل انطباق با قوانین و مقررات ارائه کرده است. مجموعه ابزارها با استفاده از دانش بدست آمده از قانون‌گذاران قبلی و بررسی داده‌های بیش از ۲۰۰ منبع مختلف آموزش داده شده‌اند. از همه اینها گذشته، این برنامه مبتنی ۶۰,۰۰۰ حضاربه‌ی نظارتی است. این ابزار شامل سه مجموعه از ابزارهای شناختی است که با انطباق قوانین و مقررات سر و کار دارند. یکی از این

¹ TD Bank

² Danamon Bank

³ HSBC

⁴ Olivia

⁵ Santander

⁶ Nina

⁷ Luvo

⁸ RBS

⁹ Accenture

¹⁰ Collette

¹¹ NaO

ابزارها با جرائم مالی، تشخیص و نشانه‌گذاری تراکنش‌های مشکوک و تقلب و کلاهبرداری‌های ممکن سر و کار دارد. ابزار دوم به پایش عدم انطباق و ابزار سوم با حجم انبوه داده‌ها سر و کار دارد. واتسون به عنوان یک مشاور مالی بانکداری برای این مسائل و سایر مسائل بانکداری عمل می‌کند.

ابزارهای آی‌بی‌ام با هدف کمک به مؤسسات مالی برای توجیه تصمیمات مهم ایجاد شده‌اند. الگوریتم‌های هوش مصنوعی به ارزیابی داده‌های ورودی و خروجی در تصمیم‌گیری مدیریتی می‌پردازند. به عنوان مثال، زمانیکه برنامه‌ای یک فعالیت مشکوک را تشخیص می‌دهد، مدیر مربوطه را مطلع می‌کند تا اقدام لازم را اتخاذ کند. برای جزئیات، نگاه کنید به کلوزل (۲۰۱۷). در مورد کاربردی ۲.۵ استفاده از هوش مصنوعی برای بهبود خدمات مشتری در بانک ایالات متحده به تصویر کشیده شده است.

پیشرفت‌های به عمل آمده در هوش مصنوعی در حوزه‌های متعدد در صنعت بیمه و به ویژه در صدور سیاست‌ها و رسیدگی به کلیم‌ها در حال بهبود است.

طبق هائوآری^۱ (۲۰۱۷)، اهداف اصلی پشتیبانی هوش مصنوعی، بهبود نتایج تحلیل و ارتقاء تجربه مشتری هستند. کلیم‌های ورودی توسط هوش مصنوعی تحلیل می‌شوند و بسته به ماهیت آنها برای تسویه‌کنندگان موجود ارسال می‌شوند. فناوری‌های بکار رفته شامل ان‌ال‌پی و تشخیص متن (فصل ۶ و ۷) می‌باشند. نرم‌افزار هوش مصنوعی می‌تواند به گردآوری و تحلیل داده‌ها و داده‌کاوی کلیم‌های قدیمی کمک کند.

عوامل بیمه، سابقاً زمان بسیار زیادی را صرف پرسیدن سوالات روتین از ارسال‌کنندگان کلیم‌های بیمه می‌کردند. ماشین‌های هوش مصنوعی طبق گفته‌ی بوشان (۲۰۱۶) بر سرعت، دقت و اثربخشی اجرای این فرایند می‌افزایند. سپس هوش مصنوعی می‌تواند موجب تسهیل فرایند تعهد بیمه شود.

به همین منوال، رسیدگی به کلیم‌ها به کمک هوش مصنوعی، روان و آسان شده است و زمان رسیدگی را تا حداکثر ۹۰٪ کاهش داده و موجب افزایش دقت می‌شود. قابلیت‌های یادگیری ماشین و سایر برنامه‌های هوش مصنوعی را می‌توان در عرض چند ثانیه در سطح جهانی به اشتراک گذاشت.

خدمات و قدردانی از مشتری در بانک یو اس

مورد کاربردی ۲.۵

از جولای ۲۰۱۷، بانک یو اس^۲ توانسته است تا زمانیکه اعضای خدمات نظامی و کهنه سربازان با یکی از شعب آن تماس گرفته یا وارد شعبه می‌شوند آنها را به طور خودکار شناسایی کند. این کار ساده‌ای نیست. اعضای خدمات توسط « اینشتین^۳ » که یک سرویس مدیریت ارتباط با مشتریان مبتنی بر هوش مصنوعی محصول شرکت سیلzfورس^۴ است شناخته می‌شوند.

¹ Hauari

² US Bank

³ Einstein

⁴ Salesforce

بانک یو اس سعی دارد تا مشتریان را شناسایی کرده و نیازهایشان را درک کند. « اینشتین » به بانک‌ها کمک می‌کند تا در اینکار به یک مزیت رقابتی دست پیدا کنند. دانشی که بدین طریق ارائه می‌شود نه تنها برای بازاریابی و ارائه خدمات مالی حرفه‌ای هدفمند بلکه برای ارسال تبریک سالروز تولید مشتریان یا تشکر از آنها به خاطر استفاده از خدمات بانک نیز استفاده می‌شود. بانک در حال حاضر اطلاعات بسیار زیادی در مورد مشتریان دارد که می‌تواند در لحظه در اختیار عوامل انسانی خود قرار دهد. اینگونه اطلاعات به مشتریان در زمانی که آنلاین هستند و زمانیکه به یکی از شعب بانک مراجعه کرده‌اند کمک می‌کند. کاربرد هوش مصنوعی هر آنچه لازم است درباره‌ی مشتری بدانیم را به ما می‌گوید و ما می‌توانیم خدمات مناسب را ارائه کنیم. به عنوان مثال، اگر مشتری به بیمه نیاز دارد، هوش مصنوعی این نیاز را تشخیص داده و ما یک راه‌حل مناسب ارائه می‌کنیم. هوش مصنوعی همچنین اطلاعات را به یک مشتری آنلاین عرضه می‌کند: "سلام مری، متوجه شدم که شما پرداخت قسط رهن‌ها را چک می‌کنید. خبرهای خوبی برایتان دارم..."

Source: Compiled from Crosman (2017) and Carey (2017).

سوالاتی برای مورد کاربردی ۲.۵

۱. مزیت‌های « اینشتین » برای بانک یو اس چه هستند؟
۲. مزیت‌های آن برای مشتریان چه هستند؟
۳. مزایای برقراری ارتباط صوتی چه هستند؟

بیمه‌گرها نیز همانند سایر استفاده‌کنندگان خدمات هوش مصنوعی باید دستخوش تحول شده و خودشان را با تغییر وفق دهند. عوامل فردی و سازمانی می‌توانند از پیشگامان اولیه‌ی هوش مصنوعی یاد بگیرند. در خصوص نحوه‌ی انجام اینکار در مت‌لایف نگاه کنید به بلاگ (۲۰۱۷).

مترومایل^۱ یک شرکت نوآور در حوزه‌ی بیمه وسیله‌ی نقلیه است که از مدل پرداخت به ازای هر مایل استفاده می‌کند. این شرکت در هفته ایالت فعالیت می‌کند. در اواسط سال ۲۰۱۷، این شرکت شروع به استفاده از برنامه‌های مبتنی بر هوش مصنوعی برای خودکارسازی داده‌های تصادفات، رسیدگی به ادعای خسارت تصادفات و پرداخت ادعای خسارت به مشتریان نمود. طبق گفته‌ی سانتانا (۲۰۱۷)، پلتفرم خودکار از یک بات ادعای خسارت هوشمند به نام « آوا »^۲ بهره می‌برد که تصاویر ارسالی توسط مشتریان را پردازش نموده و داده‌های تلماتیک مربوطه را استخراج می‌کند. بات هوش مصنوعی، نقاط اصلی تصادف را شبیه‌سازی کرده و بر اساس قواعد تصمیم، صحنه‌گذاری را انجام داده و اجازه‌ی پرداخت‌ها را برای صحنه‌گذاری‌های موفق صادر می‌کند. این فرایند چند دقیقه طول می‌کشد. فقط پرونده‌های پیچیده برای تحقیق و بررسی برای عوامل انسانی ارسال می‌شوند. مشتریان از آنجائیکه مشکلاتشان سریع حل می‌شود خوشحال هستند. اگرچه « آوا » در حال حاضر محدود به چند نوع خاص از ادعاهای خسارت است ولی محدوده‌ی آن با توجه به قابلیت‌های ایدگیری ماشین و پیشرفت‌های به عمل آمده از الگوریتم‌های هوش مصنوعی در حال افزایش است.

¹ Metromile

² AVA

نکته: یک شرکت نوپا به نام لیموناد در سال ۲۰۱۵ یک پلتفرم مبتنی بر هوش مصنوعی را برای بیمه تهیه کرد که شامل بات‌ها و یادگیری ماشین است. برای جزئیات به گاکلیاردی (۲۰۱۷) مراجعه کنید.

سوالاتی برای مرور بخش ۲.۷

۱. روش‌های جدیدی که بانک‌ها برای تعامل با مشتریان با استفاده از هوش مصنوعی بکار می‌گیرد چه هستند؟
۲. گفته می‌شود که خدمات مالی، با پشتیبانی هوش مصنوعی شخصی‌سازی‌تر می‌شوند. توضیح دهید.
۳. کدام فعالیت‌های پشتیبان اداره توسط هوش مصنوعی تسهیل می‌شوند؟
۴. هوش مصنوعی چگونه می‌تواند به ایمنی و امنیت کمک کند؟
۵. نقش چت‌بات‌ها و دستیارهای مجازی در خدمات مالی چیست؟
۶. واتسون آی بی‌ام چگونه می‌تواند به خدمات بانکداری کمک کند؟
۷. ارتباط بین «اینشتین» شرکت سیلزفورس و مدیریت ارتباط با مشتری در خدمات مالی چیست؟
۸. هوش مصنوعی چگونه می‌تواند به رسیدگی به ادعاهای خسارت بیمه کمک کند؟

همانند سایر حوزه‌های وظیفه‌ای کسب‌وکار، استفاده از فناوری‌های هوش مصنوعی در مدیریت منابع انسانی به سرعت در حال گسترش است. و همانند سایر حوزه‌ها، خدمات هوش مصنوعی موجب کاهش هزینه و افزایش بهره‌وری، سازگاری و سرعت اجرا می‌شوند.

ساوار^۱ (۲۰۱۷) دلایل زیر را برای تغییر و تحول در مدیریت منابع انسانی به کمک هوش مصنوعی بر می‌شمرد: (۱) کاهش تبعیض و تعصب انسانی، (۲) افزایش کارایی، بهره‌وری و بینش در ارزیابی نامزدها و (۳) بهبود روابط با کارکنان فعلی. ویزلو^۲ (۲۰۱۷) به استفاده از هوش مصنوعی به عنوان ادامه‌ی راه اتوماسیون نگاه می‌کند که از مدیریت منابع انسانی پشتیبانی کرده و مدام در حال تغییر آن است. ویزلو بیان می‌کند که اینگونه اتوماسیون، چگونه نحوه‌ی کارکرد و مشارکت کارکنان مدیریت منابع انسانی را دستخوش تغییر می‌کند. این تغییر همچنین موجب تحکیم و تقویت کار تیمی می‌شود. ویزلو تأثیرات هوش مصنوعی را به بخش‌های زیر تقسیم کرده است:

استخدام (جذب نیرو) یکی از ارکان اصلی در وظایف مدیریت منابع انسانی به ویژه در سازمان‌های بزرگ، جذب کارمندان جدید است. واقعیت این است که بسیاری از پست‌های سازمانی به دلیل مشکلات در یافتن کارمندان مناسب خالی مانده‌اند. در همین حین، بسیاری از افراد واجد شرایط قادر به یافتن شغل مناسب نیستند. هوش مصنوعی همانطور که در مورد کاربرد ۲.۶ نشان داده شده است موجب بهبود فرایند جذب نیرو می‌شود. استفاده از چت‌بات‌ها برای تسهیل جذب نیرو توسط مایستر^۳ (۲۰۱۷) نیز توضیح داده شده است. شرکت‌هایی که به استخدام کنندگان و

¹ Savar

² Wislow

³ Meister

جویندگان کار هستند کمک می‌کنند، به ویژه لینکدین، از الگوریتم‌های هوش مصنوعی برای توصیه‌ی موارد منطبق با نیاز استخدام‌کنندگان و جویندگان کار استفاده می‌کنند. هاینز^۱ (۲۰۱۷) این فرایند را توضیح می‌دهد، با اشاره به این نکته که یک مزیت اصلی این فرایند، حذف تبعیض‌های ناخواسته و پیش‌داورهای انسانی است.

هوش مصنوعی موجب تسهیل آموزش می‌شود دستاوردهای تکنولوژیکی سریع، آموزش و نگهداشت کارکنان را به یک ضرورت بدل کرده است. از روش‌های هوش مصنوعی می‌توان برای تسهیل یادگیری استفاده کرد. به عنوان مثال، از چَت‌ها می‌توان به عنوان یک منابع دانشی برای پاسخگویی به سوالات فراگیران استفاده کرد. دوره‌های آنلاین برای کارکنان امری بسیار متداول است. از هوش مصنوعی می‌توان برای آزمودن میزان پیشرفت استفاده کرد. علاوه بر این، از هوش مصنوعی می‌توان برای ارائه آموزش‌های آنلاین شخصی‌سازی شده برای افراد و طراحی دوره‌های گروهی استفاده کرد.

هوش مصنوعی از تحلیل عملکرد پشتیبانی می‌کند (ارزیابی) مدیریت منابع انسانی با استفاده از ابزارهای هوش مصنوعی می‌تواند با تفکیک کارها به مؤلفه‌ها و اجزای کوچک‌تر و سنجش عملکرد هر یک از کارکنان و تیم‌ها در هر یک از اجزا، عملکرد را تحلیل کند. هوش مصنوعی همچنین قادر به ردیابی تغییرات و پیشرفت‌ها از طریق ترکیب ابزارهای تحلیل و هوش مصنوعی است.

استفاده از هوش مصنوعی در نگهداشت و ترک کار پرسنل برای نگهداشت پرسنل و جلوگیری از ترک کار آنها، سازمان‌ها باید روش‌های خرسندسازی کارکنان را تحلیل و پیش‌بینی کنند. از یادگیری ماشین می‌توان به روش شناسایی الگوهای تأثیرگذار برای تشخیص علل ترک کارکنان استفاده کرد.

به محض استخدام کارکنان جدید، دپارتمان منابع انسانی باید آنها را با فرهنگ سازمانی و فرایندهای عملیاتی آشنا نماید. برخی کارکنان جدید به توجه بیشتری نیاز دارند. هوش مصنوعی به مدیریت منابع انسانی کمک می‌کند تا مسیرهای سفارشی برای پذیرش سازمانی را که برای افراد جدیدالاستخدام بهترین گزینه هستند تدارک ببینند. نتایج نشان می‌دهد که کارکنانی که توسط طرح‌های مبتنی بر هوش مصنوعی پشتیبانی می‌شوند بیشتر تمای دارند تا در سازمان‌ها باقی بمانند (ویزلو ۲۰۱۷).

چگونه «الکساندر مان سولوشنز» (ای ام اس) از هوش مصنوعی برای پشتیبانی از فرایند جذب نیروی کار استفاده می‌کند.

مورد کاربردی ۲.۶

«الکساندر مان»^۲ یک شرکت در شیکاگو است که راه‌حل‌های هوش مصنوعی را برای پشتیبانی از فرایند جذب نیروی کار ارائه می‌کند. هدف اصلی این شرکت، کمک به شرکت‌ها برای حل چالش‌ها و مسائل مدیریت منابع انسانی آنها است. از هوش مصنوعی برای موارد زیر استفاده می‌شود:

۱. کمک به شرکت‌ها برای ارزیابی متقاضیان و رزومه‌های آنها با استفاده از یادگیری ماشین. نتیجه اینکار، تصمیم‌گیری برای دعوت کردن متقاضیان برای انجام مصاحبه است.

¹ Haines

² Alexander Mann Solutions

۲. کمک به شرکت‌ها برای ارزیابی رزومه‌هایی که بر روی وب منتشر می‌شوند. نرم‌افزار هوش مصنوعی قادر است عبارات کلیدی را برای جستجوی کارکنان با سوابق مورد نظر بکار بگیرد (مثلاً آموزش، مدت تجربه).

۳. ارزیابی رزومه‌های بهترین کارکنانی که در حال حاضر در یک شرکت کار می‌کنند و پروفایل‌هایی از آنها را برای زمانیکه پست خالی ایجاد شده تهیه می‌کنند. سپس این پروفایل‌ها با پروفایل‌های نامزدهای متقاضی مقایسه شده و بهترین‌ها بر اساس تناسب آنها با هر پست یا شغل موجود رتبه‌بندی می‌شوند. علاوه بر رتبه‌بندی، برنامه‌ی هوش مصنوعی میزان تناسب یا برازندگی آنها را با هر یک از معیارهای مطلوب نشان می‌دهد. در این مرحله، فرد استخدام‌کننده می‌تواند تصمیم‌نهایی را برای گزینش بگیرد. بدین ترتیب، فرایند گزینش، سریع‌تر و بسیار دقیق‌تر انجام می‌شود.

دقت این فرایند توانسته تا مشکل حجم نامزدهای متقاضی را حل کند و اطمینان حاصل کند که افراد واجد شرایط از قلم نیوفتاده‌اند و متقاضیان ضعیف، برگزیده نشده‌اند.

«الکساندر مان» همچنین به مشتریان خود کمک می‌کند تا چَت‌بات‌هایی را برای پاسخگویی به سوالات نامزدهای متقاضی در رابطه به شغل‌ها و شرایط کاری در شرکت استخدام‌کننده نصب کنند. (در خصوص چَت‌بات‌های جذب نیروی کار نگاه کنید به دیکسون ۲۰۱۷).

Sources: Compiled from Huang (2017), Dickson (2017), and alexandermansolutions.com, accessed June 2018.

سوالاتی برای مورد کاربردی ۲.۶

۱. چه نوع تصمیماتی پشتیبانی می‌شوند؟
۲. در خصوص همکاری انسان-ماشین اظهارنظر کنید.
۳. مزایای آن برای استخدام‌کنندگان چیست؟ برای متقاضیان کار چیست؟
۴. چه وظایف و فعالیت‌هایی در فرایند جذب نیروی کار به طور کامل خودکارسازی شده‌اند؟
۵. مزایای اینگونه خودکارسازی چیست؟

استفاده از چَت‌بات‌ها برای پشتیبانی مدیریت منابع انسانی استفاده از چَت‌بات‌ها در مدیریت منابع انسانی به سرعت در حال افزایش است. توانایی آنها در ارائه‌ی اطلاعات جاری به کارمندان در هر زمان یکی از دلایل اصلی آن است. دیکسون (۲۰۱۷) به چَت‌بات‌ها زیر اشاره می‌کند: «میا»، یک دستیار جذب نیروی کار و جاب‌بات که برای پشتیبانی از استخدام کارگران ساعتی پشتیبانی می‌کند. این بات همچنین به عنوان یک پلاگین برای کریگزلیست استفاده می‌شود. یک چَت‌بات دیگر که قبلاً به آن اشاره شد اولیویا نام دارد (Olivia.paradox.ai).

معرفی و بکارگیری هوش مصنوعی در مدیریت منابع انسانی همانند معرفی هوش مصنوعی در سایر حوزه‌های کارکردی است. مایستر (۲۰۱۷) فعالیت‌های زیر را توصیه کرده است:

۱. آزمایش کردن انواع مختلف چَت‌بات‌ها
۲. ایجاد یک رویکرد تیمم از سایر حوزه‌های وظیفه‌ای

۳. طرح‌ریزی مناسب یک نقشه راه فناوری برای کوتاه مدت و بلند مدت مشتمل بر چشم‌انداز مشترک با سایر حوزه‌های وظیفه‌ای
۴. شناسایی نقش‌های شغلی جدید و تغییرات در نقش‌های شغلی فعلی در یک محیط تغییرشکل یافته
۵. آموزش و تربیت تیم مدیریت منابع انسانی برای درک هوش مصنوعی و دستیابی به تخصص در آن

سوالاتی برای مرور بخش ۲.۸

۱. فهرستی از فعالیت‌ها را در جذب نیروی کار تهیه کرده و پشتیبانی ارائه شده توسط هوش مصنوعی برای هر یک را توضیح دهید؟
۲. مزایایی که هوش مصنوعی برای استخدام‌کنندگان دارد چه هستند؟
۳. مزایای آن برای جویندگان کار چیست؟
۴. هوش مصنوعی چگونه موجب تسهیل آموزش می‌شود؟
۵. ارزیابی عملکرد کارکنان چگونه با هوش مصنوعی بهبود می‌یابد؟
۶. شرکت‌ها چگونه می‌توانند به کمک هوش مصنوعی، نگهداشت پرسنل را افزایش دهند؟
۷. نقش چَت‌بات‌ها را در پشتیبانی مدیریت منابع انسانی توضیح دهید.

در مقایسه با سایر حوزه‌های کسب‌وکار، هوش مصنوعی احتمالاً کاربردهای بیشتری را در بازاریابی و تبلیغات دارد. به عنوان مثال، آمازون. کام و سایر شرکت‌های تجارت الکترونیک برای بیش از ۲۰ سال از توصیه محصولات مبتنی بر هوش مصنوعی استفاده کرده‌اند. به دلیل کثرت کاربردها ما در اینجا به ذکر چند مثال اکتفا می‌کنیم.

دیویس^۱ (۲۰۱۶) ۱۵ مثال از هوش مصنوعی در بازاریابی ارائه کرده که به همراه توضیحات توسط مولفان این کتاب و مارتین (۲۰۱۷) ارائه شده است. همچنین نگاه کنید به پنینگتون^۲ (۲۰۱۸).

۱. *توصیه محصول و توصیه‌های شخصی.* فناوری‌های مبتنی بر هوش مصنوعی که از توصیه‌ی کتاب آمازون. کام برای فیلم‌های نتفلیکس آغاز شد به طور گسترده‌ای برای توصیه‌های شخصی‌سازی شده مورد استفاده قرار می‌گیرند (مارتین ۲۰۱۷).

۲. *موتورهای جستجوی هوشمند.* گوگل از سیستم هوش مصنوعی رَنک برین^۳ برای تفسیر اعلام‌های کاربران استفاده می‌کند. بکارگیری این ال پی به درک محصولات یا خدماتی که کاربران آنلاین به دنبال آنها هستند کمک می‌کند. این شامل استفاده از ارتباطات صوتی نیز می‌باشد.

¹ Davis

² Pennington

³ RankBrain

۳. شناسایی کلاهبرداری و نقض داده‌ها. کاربرد هوش مصنوعی برای این مورد شامل استفاده از کارت‌های نقدی/اعتباری برای سالیان متمادی، محافظت از صادرکنندگان ویزا کارت و سایر کارت‌ها می‌باشد. فناوری‌های مشابهی برای محافظت از خرده‌فروش‌ها در برابر حملات هکرها وجود دارد (مثلاً تارگت و نیمان مارکوس).
۴. سمانتیک/اجتماعی. خرده‌فروش‌ها با استفاده از فناوری‌های مبتنی بر هوش مصنوعی از قبیل تحلیل حس‌ها و تشخیص صدا و تصویر می‌توانند از نیازهای مشتریان آگاه شده و تبلیغات هدفمند ارائه کرده و محصولات را از طریق پست الکترونیکی و رسانه‌های اجتماعی مستقیماً به مشتریان توصیه کنند.
۵. طراحی وب سایت. بازاریاب‌ها با استفاده از روش‌های هوش مصنوعی می‌توانند وب سایت‌های جذابی را طراحی کنند.
۶. قیمت‌گذاری/تولیدکننده. الگوریتم‌های هوش مصنوعی به خرده‌فروش‌ها در قیمت‌گذاری محصولات و خدمات به شکلی پویا و بر اساس رقابت، نیازمندی‌های مشتری و غیره کمک می‌کند. به عنوان مثال، هوش مصنوعی یک تحلیل پیش‌گویان را برای پیش‌بینی تأثیر سطوح قیمتی مختلف ارائه می‌کند.
۷. خدمات مشتری پیشگویانه. مشابه پیش‌بینی تأثیر قیمت‌گذاری، هوش مصنوعی می‌تواند به پیش‌بینی تأثیر گزینه‌های مختلف خدمات مشتریان کمک کند.
۸. آگهی‌های هدفمند. همانند توصیه‌ی محصول که مبتنی بر شرح حال کاربر هستند، بازاریاب‌ها می‌توانند آگهی‌های تبلیغاتی را متناسب با هر مشتری تهیه کنند. ماشین‌های هوش مصنوعی تلاش می‌کنند تا آگهی‌های مختلف را با مشتریان تطبیق دهند.
۹. تشخیص گفتار. با افزایش روند بکارگیری تعاملات صوتی بین انسان و ماشین، استفاده از بات‌ها توسط بازاریاب‌ها برای ارائه قیمت و اطلاعات محصول نیز شتاب می‌گیرد. مشتریان به جای نوشتن ترجیح می‌دهند تا با بات‌ها صحبت کنند.
۱۰. ترجمه‌ی زبان. به کمک هوش مصنوعی، دو نفر که به زبان‌های متفاوتی صحبت می‌کنند می‌توانند با یکدیگر گفتگو کنند. همچنین، مشتریان می‌توانند با استفاده از مترجم گوگل از وب سایت‌هایی که به زبان‌هایی غیر از زبان خودشان نوشته شده خرید کنند.
۱۱. بخش‌بندی مشتریان. بازاریاب‌ها ابتدا مشتریان را به گروه‌هایی بخش‌بندی کرده و سپس آگهی‌های تبلیغاتی سفارشی را برای هر گروه مشتری ایجاد می‌کنند. اگرچه این روش اثربخشی کمتری نسبت به هدف قراردادن تک تک مشتریان دارد ولی از تبلیغات انبوه مؤثرتر است. هوش مصنوعی می‌تواند از متن کاوی و داده‌کاوی برای کمک به بازاریاب‌ها در شناسایی مشخصه‌های هر بخش خاص از مشتریان (مثلاً از طریق کاوش در فایل‌های قدیمی) و همچنین سفارشی‌سازی بهترین آگهی‌های تبلیغاتی برای هر بخش بازار کمک کند.
۱۲. پیش‌بینی فروش. برنامه‌ریزی و استراتژی بازاریاب‌ها مبتنی بر پیش‌بینی فروش است. اینگونه پیش‌بینی‌ها ممکن است برای برخی محصولات خاص بسیار دشوار باشد. تحلیل‌های پیشگویان و سایر ابزارهای هوش مصنوعی می‌توانند پیش‌بینی بهتری نسبت به ابزارهای آماری سنتی ارائه کنند.
۱۳. تشخیص تصویر. این فناوری می‌تواند در تحقیقات بازار بسیار مفید باشد (مثلاً برای شناسایی اولویت‌ها و ترجیحات مشتری محصولات یک شرکت در برابر رقیب آن شرکت). همچنین می‌توان از آن برای شناسایی عیوب در تولید و/یا بسته‌بندی محصولات استفاده کرد.
۱۴. تولید محتوا. بازاریاب‌ها به طور مستمر در حال ایجاد آگهی‌ها و اطلاعات محصول هستند. هوش مصنوعی می‌تواند به این فعالیت سرعت بخشیده و اطمینان حاصل کند که با قوانین و مقررات منطبق و سازگار هستند. همچنین، هوش مصنوعی می‌تواند به تولید محتوای هدفمند هم برای افراد و هم برای بخش‌های مشتریان کمک کند.

۱۵. استفاده از بات‌ها، دستیارها و مشاورهای ربات. در فصل ۱۲، ما توضیح خواهیم داد که چگونه بات‌ها، دستیارهای شخصی و مشاوران رباتی به مشتریان خدمات و محصولات کمک می‌کنند. همچنین، این ماشین‌های هوش مصنوعی موجب تسهیل تجربه‌ی مشتری و تحکیم مدیریت ارتباط با مشتریان می‌شوند. برخی کارشناسان، بات‌ها و دستیارهای شخصی مجازی را « ویرترین بازاریابی » می‌نامند.

فهرست دیگری از کاربردها در en.wikipedia.org/wiki/Marketing_and_artificial_intelligence موجود است.

به روش‌های بسیاری می‌توان از هوش مصنوعی در بازاریابی استفاده کرد. یک روش در مورد کاربردی ۲.۷ در مورد کرافت فودز به تصویر کشیده شده است.

همانطور که قبلاً توضیح داده شد، یک تأثیر مهم فناوری‌های هوش مصنوعی، تغییر تجربیات مشتری است. یک مثال بارز از آن، استفاده از بات‌های مکالمه‌ای است. بات‌ها (مثلاً آکسا) می‌توانند اطلاعاتی را در مورد محصولات و شرکت‌ها ارائه کرده و مشاوره و راهنمایی بدهند (مثلاً مشاوران رباتی برای سرمایه‌گذاری). گانگوانی^۱ (۲۰۱۶) روش‌های زیر را برای بهبود تجربیات مشتریان فهرست کرده است:

۱. استفاده از این ال پی برای تولید مستندات کاربر. این قابلیت همچنین موجب بهبود گفت‌وگو بین مشتری و ماشین می‌شود.
 ۲. استفاده از دسته‌بندی دیداری برای سازماندهی تصاویر (مثلاً تشخیص دیداری آی بی ام^۲ و کلاریفای^۳).
 ۳. ارائه‌ی خدمات شخصی‌سازی شده و بخش‌بندی شده از طریق تحلیل داده‌های مشتریان. این شامل بهبود تجربه‌ی خرید و مدیریت ارتباط با مشتریان نیز می‌شود.
- یک مثال بارز از هوش مصنوعی در مدیریت ارتباط با مشتریان، « اینشتین » محصول شرکت سیلزفورس است.

« اینشتین » محصول شرکت سیلزفورس یک مجموعه از فناوری‌های هوش مصنوعی است (مثلاً اینشتین ویزن^۴ برای تشخیص تصویر) که برای ارتقای تعاملات با مشتری و پشتیبانی فروش بکار می‌رود. به عنوان مثال، این سیستم، داشبوردهای فروش دینامیک را برای نمایندگان فروش ارائه می‌کند. همچنین با استفاده از تحلیل‌های فروش، کار تیمی را مدیریت کرده و به ردیابی عملکرد می‌پردازد. این محصول هوش مصنوعی همچنین قادر به پیش‌بینی و توصیه است و از پلتفرم « سیلزفورس کاستومر ساکسسفول^۵ » و سایر محصولات سیلزفورس پشتیبانی می‌کند.

¹ Gangwani

² IBM Visual Recognition

³ Clarifai

⁴ Einstein Vision

⁵ Salesforce Customer Successful Platform

اینشتین به طور خودکار به اولویت‌بندی سرنخ‌های فروش پرداخته و موجب بهره‌وری بیشتر نمایندگان فروش می‌شود. نمایندگان فروش همچنین به بینش و درک بهتری از احساس مشتریان، حضور رقبا و سایر اطلاعات دست پیدا می‌کنند.

کرافت فودز از هوش مصنوعی برای بازاریابی و مدیریت ارتباط با مشتریان

استفاده می‌کند

مورد کاربردی ۲.۷

با افزایش تعداد کاربران سیار، تعداد خریداران سیار نیز در حال افزایش است. کرافت فودز^۱ به این نکته پی برده است و به همین دلیل فروش و تبلیغات خود را با این روند سازگار کرده است. مشتریان سیار به دنبال برندهای کرافت و تعامل با آنها می‌باشند. کرافت فودز به دنبال این بود تا تعامل با شرکت را برای مشتریان در هر زمان و مکان ممکن آسان‌تر نماید. برای دستیابی به این هدف، کرافت فودز، «دستیار غذایی^۲» خود را به نام «دستیار غذایی کرافت^۳» ایجاد کرد.

دستیار غذای کرافت

دستیار غذای کرافت یک اپلیکیشن برای گوشی‌های هوشمند است که اجازه‌ی دسترسی به بیش از ۷۰۰ طرز تهیه را به مشتریان می‌دهد. بدین ترتیب، مشتری می‌تواند براحتی به دنبال ایده‌ها بگردد. مشتری وارد یک فروشگاه مجازی شده و «طرز تهیه‌ی روز» را باز می‌کند. دستیار، تمامی مواد مورد نیاز برای آن طرز تهیه یا هر طرز تهیه مورد نظر مشتری را به وی می‌گوید. همچنین تمامی کوپن‌های موجود برای مواد اولیه را بر روی گوشی هوشمند کاربر ارسال می‌کند. مشتریان کافی است گوشی هوشمند خود را به سوپرمارکت برده، کوپن‌ها را اسکن کرده و در خرید مواد اولیه صرفه‌جویی کنند. طرز تهیه‌ها به صورت ویدئویی نیز نمایش داده می‌شوند. ویژگی منحصربفرد این برنامه، بکارگیری یک الگوریتم هوش مصنوعی است که قابلیت یادگیری از سفارشات کاربران و نتیجه‌گیری را دارد (مثلاً اندازه‌ی خانواده‌ی کاربر). هوش مصنوعی هرچه بیشتر در مورد کاربران یاد می‌گیرد، توصیه‌های بیشتری نیز ارائه می‌کند. به عنوان مثال به کاربران می‌گوید که با باقیمنده غذاهایشان چکار کنند. علاوه بر این، دستیار غذایی هرچه بیشتر در مورد کاربران یاد می‌گیرد، توصیه‌های مفیدتری برای طرز تهیه و پخت می‌تواند ارائه کند. این مشابه توصیه‌کننده‌ی نتفلیکس است. هرچه مشتریان، محصولات بیشتری از کرافت را خریداری کنند، مشاوره و راهنمایی‌های بیشتری نیز دریافت خواهند کرد. این برنامه همچنین مشتریان را به نزدیک‌ترین فروشگاه که مواد اولیه‌ی طرز تهیه‌ی مورد نظر آنها را دارد هدایت می‌کند. مشتریان در عرض ۲۰ دقیقه می‌توانند در مورد نحوه‌ی آماده‌سازی غذا و بسیاری از موضوعات دیگر مرتبط با پخ و پز راهنمایی بگیرند. هوش مصنوعی، رفتارهای مشتریان را ردیابی می‌کند. اطلاعات در داخل کارت وفاداری هر مشتری ذخیره می‌شود. سیستم در مورد علایق مشتریان نتیجه‌گیری کرده و تبلیغات هدفمندی را برای آنها ایجاد می‌کند. این فرایند، تشخیص الگوی رفتاری نامیده می‌شود و مبتنی بر فنون هوش مصنوعی از قبیل «فیلترینگ مشارکتی» است (فصل ۱۲).

دستیارهای هوش مصنوعی همچنین می‌توانند پیام‌هایی را برای کاربران ارسال کنند و مطلع شوند که آیا کاربران به موضوعات آنها علاقمند هستند یا خیر. دستیارها همچنین می‌دانند که آیا مشتریان به این تبلیغات پاسخ مثبت می‌دهند یا خیر و آیا به آنها انگیزه داده شده تا محصولات جدید یا محصولاتی که قبلاً خریداری می‌کردند را اکنون بیشتر بخرند؟ دستیار هوش مصنوعی

¹ Kraft Foods

² Food Assistant

³ Kraft Food Assistant

کرافت در واقع قصد دارد تا بر رفتار مشتری تأثیر گذارنده و در برخی موارد آن را اصلاح کند. کرافت نیز همانند سایر فروشندگان از اطلاعات گردآوری شده توسط دستیار هوش مصنوعی برای تدوین و اجرای راهبردهای تجارت سیار و معمول استفاده می‌کند. کرافت و فروشندگان مشابه با استفاده از اطلاعات گردآوری شده می‌توانند برنامه‌های بازاریابی آنلاین و فیزیکی خود را گسترش دهند.

نکته: کاربران می‌توانند از طریق صدا با این سیستم تعامل برقرار کنند (Nuance Communications). این سیستم مبتنی بر پردازش زبان طبیعی است.

Sources: Compiled from Celentano (2016), press releases at nuance.com, and kraftrecipes.com/media/iphoneassistant.aspx/, accessed March 2018.

سوالاتی برای مورد کاربردی ۲.۷

۱. تمامی فناوری‌های هوش مصنوعی بکار رفته در دستیار غذای کرافت را شناسایی کنید.
۲. مزایای آن را برای مشتریان نام ببرید.
۳. مزایای آن را برای کرافت فودز نام ببرید.
۴. تبلیغات به چه شکل انجام می‌شود؟
۵. نقش «تشخیص الگوی رفتاری» در اینجا چیست؟
۶. دستیار غذای کرافت را با سیستم‌های توصیه‌ی آمازون. کام و نتفلیکس مقایسه کنید.

برای اطلاعات بیشتر و یک پیش‌نمایش نگاه کنید به salesforce.com/products/Einstein/overview برای ویژگی‌ها و شرح محصول نگاه کنید به zdnet.com/article/salesforces-einsteinai-platform-what-you-need-to-know برای ویژگی‌های بیشتر نگاه کنید به salesforce.com/products/einstein/features/.

در ادامه به تنوع و گوناگونی فناوری‌های هوش مصنوعی در بازاریابی اشاره می‌کنیم:

- از هوش مصنوعی برای تقلید تخصص پرسنل فروش داخل فروشگاه استفاده می‌شود. در بسیاری از فروشگاه‌های فیزیکی، پرسنل کافی برای کمک و راهنمایی به مشتریانی که تمایل ندارند تا مدت طولانی را منتظر بمانند وجود ندارد. به همین دلیل، خرید با راهنمایی بات‌ها به مراتب آسان‌تر می‌شود. یک فروشگاه در ژاپن تمامی خدمات خود در یک فروشگاه فیزیکی را به کمک ربات‌های سخنگو ارائه می‌کند.
- تولید سرنخ. همانطور که در مورد «اینشتین» دیدیم، هوش مصنوعی می‌تواند با تحلیل داده‌های کاربران به تولید سرنخ‌های فروش کمک کند. این برنامه قادر به انجام پیش‌بینی است. به کمک ابزارهای تحلیل هوشمند می‌توان به درک و بینش بهتری دست یافت.
- می‌تواند از طریق شخصی‌سازی به افزایش وفاداری مشتری کمک کند. به عنوان مثال، برخی از فنون هوش مصنوعی قادر به شناسایی مشتریان دائمی (مثلاً در بانک‌ها) هستند. واتسون آی‌بی‌ام می‌تواند از توثیقه‌هایی که ارسال می‌شود افراد را شناسایی کند.

- سبیلزفورس یک کتاب الکترونیک رایگان به نام « هر آنچه لازم است تا در مورد هوش مصنوعی در مدیریت ارتباط با مشتریان بدانید » منتشر کرده است (salesforce.com/form/pdf/ai-for-crm.jsp).
- می‌تواند موجب بهبود پایپ لاین‌های فروش شود. نارایان (۲۰۱۸)، فرایندی را برای شرکت‌ها طراحی کرده تا بتوانند از هوش مصنوعی و ربات‌ها برای انجام اینکار استفاده کنند. به طور خاص، ربات‌ها می‌توانند بازدیدکنندگان ناشناس را تبدیل به مشتری کنند. ربات‌ها از سه مرحله استفاده می‌کنند: (۱) فهرستی از مشتریان هدف را در پایگاه داده تهیه می‌کنند (۲) اطلاعات، آگهی‌های تبلیغاتی، ویدئوها و سایر محتواها را برای لیست تهیه شده ارسال می‌کنند و (۳) فهرستی از سرخ‌ها که با موفقیت تبدیل به مشتری شده‌اند را برای دپارتمان فروش شرکت ارسال می‌کنند.

سوالاتی برای مرور بخش ۲.۹

۱. پنج مورد از ۱۵ کاربرد دیویس (۲۰۱۶) را نام ببرید. در مورد هر یک اظهار نظر کنید.
۲. کدامیک از ۱۵ کاربرد مربوط به فروش است؟
۳. کدامیک از ۱۵ کاربرد مربوط به تبلیغات است؟
۴. کدامیک از ۱۵ کاربرد مربوط به خدمات مشتری و مدیریت ارتباط با مشتری است؟
۵. از قابلیت‌های پیش‌بینی هوش مصنوعی برای چه چیزهایی استفاده می‌شود؟
۶. « اینشتین » شرکت سبیلزفورس چیست؟
۷. چگونه می‌توان از هوش مصنوعی برای بهبود مدیریت ارتباط با مشتری استفاده کرد؟

رشته‌ی مدیریت تولید و عملیات بسیار متنوع بوده و امروزه شاهد بکارگیری هوش مصنوعی در بسیاری از حوزه‌های آن هستیم. توضیح تمامی آنها در قالب یک کتاب نمی‌گنجد. به همین دلیل در فصل‌های باقیمانده ما مثال‌هایی از کاربردهای هوش مصنوعی در مدیریت تولید-عملیات ارائه خواهیم کرد. در اینجا صرفاً بحث مختصری در مورد دو حوزه کاربردی مربوطه خواهیم داشت: ساخت و تولید و لجستیک.

به منظور مدیریت هزینه‌های همواره فزاینده‌ی نیروی کار، تغییرات در نیازهای مشتریان، افزایش رقابت جهانی و مقررات دولتی، شرکت‌های تولیدی از سطح بالایی از خودکارسازی و دیجیتالی‌سازی استفاده می‌کنند. بر اساس بولارد و همکاران (۲۰۱۷)، یک مدل پنج‌بخشی را برای شرکت‌های تولیدی جهت استفاده از فناوری‌های هوشمند پیشنهاد داده‌اند که شامل بخش‌های زیر است:

- روانسازی فرایندها از جمله کمینه‌سازی ضایعات، بازطراحی فرایندها و استفاده از مدیریت فرایند کسب‌وکار
- برون‌سپاری برخی فرایندهای اصلی کسب‌وکار از جمله تجارت به صورت آفشور
- استفاده از اطلاعات در تصمیم‌گیری با استقرار هوش مصنوعی و تحلیل‌ها
- جایگزینی وظایف انسانی با خودکارسازی هوشمند
- دیجیتالی‌سازی تجربه‌ی مشتریان

شرکت‌ها برای مدت زمان طولانی از این مدل استفاده کرده‌اند. در واقع، رباتیک از حدود ۱۹۶۰ مورد استفاده قرار گرفته است (مثلاً یونی‌مات در جنرال موتورز). با وجود این، ربات‌ها «کودن» بودند و معمولاً یک وظیفه‌ی ساده را انجام می‌دادند. امروزه، شرکت‌ها از ربات‌های هوشمند برای وظایف پیچیده استفاده می‌کنند که امکان تولید سفارشی انبوه و ساخت بر اساس سفارش را فراهم می‌کنند. به عبارت دیگر، بسیاری از وظایف ذهنی و شناختی در حال خودکارسازی هستند. این دستاوردها از جمله هوش مصنوعی و حسگرها امکان پشتیبانی یا حتی خودکارسازی تصمیمات تولید را در لحظه می‌دهند.

زمانیکه یک حسگر، یک محصول معیوب یا یک عملکرد نادرست را تشخیص می‌دهد، داده‌ها توسط یک الگوریتم هوش مصنوعی پردازش می‌شوند. سپس یک اقدام، بلافاصله و به طور خودکار انجام می‌شود. به عنوان مثال، یک قلم کالای معیوب را می‌توان حذف کرده یا جایگزین نمود. هوش مصنوعی حتی می‌تواند خرابی تجهیزات را پیش از آنکه رخ دهند پیش‌بینی کند. این اقدام بلادرنگ موجب صرفه‌جویی زیادی برای تولیدکنندگان می‌شود.

نهایتاً، شرکت‌ها از کارخانجات هوشمند یا باهوش استفاده خواهند کرد (فصل ۱۳). این کارخانجات از حسگرها و نرم‌افزارهای پیچیده استفاده می‌کنند. یک مثال از یک تأمین‌کننده‌ی پیشگام، جنرال الکتریک است که نرم‌افزاری از قبیل آیبی ایی پرفورمنس آنالایزر^۱ و پروداکشن اگزکیوشن سوپروایزر^۲ را عرضه می‌کند. این نرم‌افزار در فضای «ابر» نگهداری شده و به عنوان یک نرم‌افزار به عنوان یک سرویس^۳ عرضه می‌شود. جنرال الکتریک برای ارائه‌ی امنیت، ارتباطات و تحلیل‌های ویژه با سیسکو و اف تی سی شراکت کرده است. علاوه بر جنرال الکتریک، شرکت‌های معروفی از قبیل زیمنس و هیتاچی نیز راه‌حل‌های جامعی ارائه کرده‌اند. به عنوان مثال، نگاه کنید به گزارش فناوری هوش مصنوعی هیتاچی (socialinnovation.hitachi/ph/solutions/ai/pdf/ai_en_170310.pdf). فروشندگان کوچک بسیاری هستند که به طور تخصصی در جنبه‌های مختلف هوش مصنوعی در ساخت و تولید فعالیت می‌کنند. به عنوان مثال، بله‌اوک سیستمز کورپوریشن^۴ که به ارائه خدمات تخصصی در زمینه‌ی ردیابی بلادرنگ عملیات به شرکت‌های کوچک می‌پردازد (گرین ۲۰۱۶).

شرکت‌های بزرگ از قبیل پروکتر آند گمبل^۵ و تویوتا موفقیت‌های زود هنگامی را ثبت کرده‌اند. با وجود این، با گذشت زمان، شرکت‌های کوچک و متوسط نیز می‌توانند از خدمات هوش مصنوعی بهره‌مند شوند. برای اطلاعات بیشتر نگاه کنید به bellhawk.com.

هوش مصنوعی و ربات‌های هوشمند به طور گسترده‌ای در لجستیک سازمانی و حمل و نقل درون و برون سازمانی و همچنین در مدیریت زنجیره‌ی تأمین مورد استفاده قرار می‌گیرند. به عنوان مثال، آمازون کام از بیش از ۵۰,۰۰۰ ربات برای جابجایی اقلام در مراکز توزیع

¹ OEE Performance Analyzer

² Production Execution Supervisor

³ Software as a service

⁴ Bellhawk Systems Corporation

⁵ Procter & Gamble

خود استفاده می‌کند (سایر شرکت‌های تجارت الکترونیک نیز همین کار را می‌کنند). به زودی، شاهد کامیون‌های بدون راننده و سایر وسایل نقلیه‌ی مستقل در سرتاسر جهان خواهیم بود (فصل ۱۳).

دی‌اچ‌آل^۱ یک شرکت حمل و نقل بین‌المللی است (رقیب شرکت‌های فِدکس و یوپی‌اِس). این شرکت یک واحد زنجیره‌ی تأمین دارد که با شرکای کسب‌وکار زیادی کار می‌کند. هوش مصنوعی و اینترنت اشیا روش کار این شرکت با شرکایش و حتی با رقبایش را تغییر داده‌اند. دی‌اچ‌آل در حال توسعه‌ی مدل‌های کسب‌وکار حمل و نقل و لجستیک نوآورانه است که عمدتاً با استفاده از هوش مصنوعی، اینترنت اشیا و یادگیری ماشین انجام می‌شوند. این مدل‌ها همچنین به مشتریان دی‌اچ‌آل کمک می‌کنند تا به یک مزیت رقابتی دست پیدا کنند (و به همین دلیل است که شرکت نتوانسته است جزئیات این کار را در گزارشات خود ارائه کند).

چندین پروژه‌ی اینترنت اشیا به یادگیری ماشین و به ویژه در حوزه‌های حسگرها، ارتباطات، مدیریت دستگاه‌ها، امنیت و تحلیل پیوند خورده‌اند. یادگیری ماشینی در چنین مواردی به سفرهای سازی راه‌حل‌ها متناسب با نیازمندی‌های خاص کمک می‌کند. در مجموع، دی‌اچ‌آل تمرکز خود را بر حوزه‌های زنجیره‌ی تأمین (مثلاً شناسایی موجودی‌ها و کنترل آنها در سرتاسر زنجیره‌ی تأمین) و مدیریت انبارها معطوف کرده است. یادگیری ماشینی و سایر الگوریتم‌های هوش مصنوعی امکان تأمین دقیق‌تر، برنامه‌ریزی تولید و هماهنگی کار را فراهم می‌آورند. تگ‌گذاری و ردیابی اقلام با استفاده از کد پاسخ سریع (کیو آر) و شناسایی با فرکانس رادویی (آر اف آی دی) امکان ردیابی در سرتاسر زنجیره تأمین را فراهم می‌کند. در پایان، هوش مصنوعی موجب تسهیل تحلیل پیشگویانه و برنامه‌ریزی منابع می‌شود. برای جزئیات به کووارد^۲ (۲۰۱۷) نگاه کنید.

سوالاتی برای مرور بخش ۲.۱۰

۱. نقش ربات‌ها در ساخت و تولید را توضیح دهید.
۲. چرا باید از هوش مصنوعی در ساخت و تولید استفاده کرد؟
۳. مدل اجرایی بولارد و همکاران را توضیح دهید.
۴. منظور از یک کارخانه‌ی هوشمند چیست؟
۵. فناوری‌های هوش مصنوعی چگونه از لجستیک داخلی و خارجی پشتیبانی می‌کنند؟

- هدف هوش مصنوعی این است که تا حد امکان کاری کند تا ماشین‌ها، کارهایی که انسان‌ها انجام می‌دهند را به شکل هوشمند انجام دهند.
- یک دلیل اصلی استفاده از هوش مصنوعی، آسان‌تر کارها و تصمیم‌گیری است. هوش مصنوعی توانمندتر (مدل‌های کسب‌وکار و کاربردهای جدید)، شهودی‌تر و کم‌خطرتر از سایر برنامه‌های کاربردی پشتیبان تصمیم است.
- یک دلیل اصلی استفاده از هوش مصنوعی، کاهش هزینه‌ها و/یا افزایش بهره‌وری است.

¹ DHL

² Coward

- سیستم‌های هوش مصنوعی می‌توانند به طور مستقل کار کنند و در زمان هزینه و زمان صرفه‌جویی کرده و به شکلی سازگار کار کنند. آنها همچنین می‌توانند در نواحی روستایی و دور افتاده که تخصص انسانی در دسترس نیست کار کنند.
- از هوش مصنوعی می‌توان برای بهبود تمامی گام‌های تصمیم‌گیری استفاده کرد.
- سیستم‌های مجازی هوشمند می‌توانند به عنوان دستیار انسان‌ها کار کنند.
- سیستم‌های هوش مصنوعی، سیستم‌های رایانه‌ای هستند که از سطح هوش پایین (ولی در حال افزایشی) برخوردارند.
- هوش مصنوعی دارای تعاریف و مشتقات متعددی است و اهمیت آن به سرعت در حال گسترش است. دولت ایالات متحده آمریکا، هوش مصنوعی را به عنوان « یک عامل محرک حیاتی برای اقتصاد ایالات متحده» عنوان کرده است (گودین ۲۰۱۶)¹.
- فناوری‌های اصلی هوش مصنوعی عبارتند از عوامل هوشمند، یادگیری ماشین، سیستم‌های رباتیک، ان‌ال‌پی و تشخیص صدا، بینایی ماشینی و سیستم‌های دانشی.
- چَت‌بات‌ها و مشاوران رباتی، همگی مبتنی بر دانش انتقال یافته به ماشین‌ها هستند.
- محدودیت‌های اصلی هوش مصنوعی عبارتند از فقدان تماس و احساس انسانی، هراس از بیکار شدن انسان‌ها و احتمال مخرب بودن آنها.
- هوش مصنوعی در بسیاری از کارهای شناختی به هیچ عنوان در حد هوش انسانی نیست ولی می‌تواند بسیاری از وظایف دستی را سریع‌تر و با هزینه کمتری انجام دهد.
- چندین نوع هوش وجود دارد و به همین دلیل اندازه‌گیری ظرفیت هوش مصنوعی کار دشواری است.
- به طور کلی، هوش انسانی برتر از هوش ماشینی است. با وجود این، ماشین‌ها قادر هستند تا انسان‌ها را در بازی‌های پیچیده شکست دهند.
- یادگیری ماشین در حال حاضر پرکاربردترین فناوری هوش مصنوعی است و تلاش می‌کند تا از طریق تجربه، یاد گرفته و عملیات را بهبود بخشد.
- فناوری‌های هوش مصنوعی با استفاده از یادگیری عمیق می‌توانند از یکدیگر یاد گرفته و موجب هم‌افزایی در یادگیری شوند.
- عوامل هوشمند موجب تعالی و پیشبرد در انجام وظایف ساده در زمان بسیار کمتر و به شکلی سازگارتر نسبت به انسان‌ها شده‌اند (مثلاً تشخیص ویروس‌های رایانه‌ای).
- قدرت اصلی فناوری یادگیری ماشین، ماحصل توانایی ماشین‌ها در یادگیری از داده‌ها و دستکاری داده‌ها است.
- یادگیری عمیق می‌تواند بسیاری از مسائل دشوار را حل کند.
- بینایی رایانه‌ای قادر به درک و فهم از روی تصاویر و از جمله ویدئوها است.
- ربات‌ها سیستم‌های رایانه‌ای الکترومکانیکی هستند که قادر به انجام وظایف جسمی و ذهنی هستند. در صورتیکه مجهز به دستگاه‌های حسگر شوند، تبدیل به ربات‌های هوشمند خواهند شد.
- رایانه‌ها قادر به درک زبان انسان‌ها بوده و می‌توانند نوشتار یا صوت را به زبان انسانی تولید کنند.
- رایانش شناختی برای حل مسائل و تصمیم‌گیری به شبیه‌سازی فرایند تفکر انسانی می‌پردازد.

¹ Gaudin

- رایانه‌ها را برای انجام وظایف ذهنی و دستی ساده با استفاده از هوش مصنوعی می‌تواند کاملاً خودکارسازی کرد.
- چندین نوع تصمیم‌گیری با استفاده از هوش مصنوعی به طور کامل خودکارسازی شده‌اند و برخی دیگر پشتیبانی می‌شوند.
- هوش مصنوعی به طور گسترده‌ای در تمامی دپارتمان‌های وظیفه‌ای سازمان‌ها بکار گرفته شده و موجب کاهش هزینه و افزایش بهره‌وری، دقت و سازگاری می‌شوند. تمایل به استفاده از چَت‌بات‌ها افزایش یافته و همگی آنها به خوبی از تصمیم‌گیری پشتیبانی می‌کنند.
- هوش مصنوعی به طور گسترده‌ای در حسابداری، خودکارسازی تراکنش‌های ساده، کمک به استفاده از کلان‌داده‌ها، یافتن تراکنش‌ها و معاملات مشکوک به کلاهبرداری، افزایش امنیت و کمک به حسابرسی و انطباق استفاده می‌شوند.
- هوش مصنوعی به طور گسترده‌ای در خدمات مالی برای بهبود خدمات مشتریان، ارائه مشاوره‌های سرمایه‌گذاری، افزایش امنیت و تسهیل پرداخت استفاده می‌شود. کاربردهای مهم آن در بانکداری و بیمه است.
- مدیریت منابع انسانی از هوش مصنوعی برای تسهیل جذب نیروی کار، ارتقاء آموزش، کمک به پذیرش سازمانی و روان‌سازی عملیات استفاده می‌کند.
- هوش مصنوعی کاربردهای گسترده‌ای در بازاریابی، فروش و تبلیغات دارد. از هوش مصنوعی برای پشتیبانی توصیه‌ی محصول، کمک به جستجوی محصولات و خدمات، تسهیل طراحی وب سایت، پشتیبانی تصمیمات قیمت‌گذاری، ترجمه زبان در تجارت بین‌الملل، کمک به پیش‌بینی‌ها و استفاده از چَت‌بات‌ها در بسیاری از فعالیت‌های بازاریابی و خدمات مشتریان استفاده می‌شود.
- هوش مصنوعی برای چندین دهه در ساخت و تولید استفاده شده است. امروزه از هوش مصنوعی برای پشتیبانی برنامه‌ریزی، هماهنگی زنجیره تأمین، لجستیک و حمل و نقل و عملیات کارخانجات هوشمند استفاده می‌شوند.

تشخیص صحنه	عامل هوشمند	مغز مصنوعی
شاپ‌بات	یادگیری ماشین	هوش مصنوعی (ای‌آی)
درک گفتار (صدا)	بینایی ماشین	هوش افزوده
آزمون تورینگ	پردازش زبان طبیعی (ان‌آل‌پی)	چَت‌بات‌ها
	تحلیل توصیفی (گزارش‌گیری)	بینایی رایانه‌ای
	ربات	یادگیری عمیق

	.5
	.6
	.7
	.8
	.9
	.10
	.11
	.12
	.13
	.14
	.15
	.16

itunes.apple.com/us/app/public-transit-appmoovit/id498477945?mt=8	.1
sitezeus.com	.2
Investopedia	.3
	.4
	.5
youtube.com/watch?v=yv0IG1D-OdU	
aaai.org/home.html	.6
csail.mit.edu ai.sri.com	
crosschx.com	.7

¹ “Ask the AI Experts: What Advice Would You Give to Executives About AI?”

² Association for the Advancement of Artificial Intelligence

	moovitapp.com	waze.com	.٨
		sentient.ai	.٩
		artificialbrain.org	.١٠
			.١١
		salesforce.com	.١٢
			.١٣
			.١٤
	nuance.com		.١٥
		cs_netradyn.com	.١٦
		salesforce.com	.١٧
			.١٨
	mckinsey.com/featured-insights/artificial-intelligence/fivefifty-real-world-ai		
		“Real World AI”	.١٩
	strategicsourceror.com/2018/03/giant-scalesupply-chains-can-make.html		.٢٠

فصل ٣

¹ Edge AI

² Dragon

ماهیت داده‌ها، مدل‌سازی آماری و مصورسازی

در عصر کلان‌داده‌ها و تحلیل‌های کسب‌وکار که در آن زندگی می‌کنیم، اهمیت داده‌ها انکارناپذیر است. جملاتی که به تازگی

بیان شده‌اند از قبیل «داده‌ها حکم نفت را دارند»، «داده‌ها واحد پولی جدید هستند» و «داده‌ها پادشاهی می‌کنند» بیش از پیش بر اهمیت داده‌ها تأکید می‌کنند. اما داده‌هایی که ما از آنها صحبت می‌کنیم مشخصاً هر نوع داده‌ای نیست. مفهوم یا اصل «ورودی زیاله، خروجی زیاله» بیش از هر تعریف دیگری که در گذشته در مورد داده‌ها داشته‌ایم امروزه در خصوص پدیده‌ی کلان‌داده‌ها مصداق دارد. داده‌ها برای اینکه بتوانند خلق ارزش کرده و تبدیل به بینش و ادراک شوند باید به دقت ایجاد/شناسایی، گردآوری، یکپارچه، پاکسازی، تغییرشکل داده شده و برای استفاده در تصمیم‌گیری‌های دقیق و به موقع به شکلی مناسب با توجه به زمینه، مفهومی‌سازی شوند.

داده‌ها، مضمون اصلی این فصل هستند. بر همین اساس، این فصل را با توضیح ماهیت داده‌ها آغاز می‌کنیم: داده‌ها چه هستند، انواع و شکل‌های مختلف آنها چیست و چگونه می‌توان آنها را پردازش کرده و برای تحلیل آماده نمود. چند بخش اول این فصل به درک و پردازش عمیق ولی در حین حال ضروری داده‌ها اختصاص دارد. در چند بخش بعدی، روش‌های آماری بکار رفته برای آماده‌سازی داده‌ها به عنوان یک ورودی برای تولید سنج‌های توصیفی و استنتاجی را توضیح خواهیم داد. پس از بخش مربوط به آمار، به گزارش‌گیری و مصورسازی داده‌ها می‌پردازیم. یک گزارش، یک مصنوع ارتباطی است که با هدف خاص تبدیل داده‌ها به اطلاعات و دانش و انتقال این اطلاعات به شکلی که به آسانی قابل درک باشد گفته می‌شود. امروزه، این گزارشات، به صورت دیداری بوده و اغلب با استفاده از آیکون‌های رنگی و گرافیکی که در مجموع به شکل داشبورد درآمده و محتوای اطلاعاتی را غنا می‌بخشند ایجاد می‌شوند. از این رو، بخش آخر این فصل به طراحی، پیاده‌سازی و بهترین تجارب مرتبط با مصورسازی اطلاعات، قصه‌گویی و داشبوردهای اطلاعاتی اختصاص دارد. بخش‌های این فصل به شرح زیر است:

۳/۱. پرونده‌ی آغازین: سیریوس ایکسام از طریق بازاریابی داده‌محور، نسل جدیدی از مخاطبان رادیو را جذب کرده و

مشارکت می‌دهد

۳/۲. ماهیت داده‌ها

۳/۳. طبقه‌بندی ساده‌ی داده‌ها

- ۳/۴. هنر و علم پیش‌پردازش داده‌ها
- ۳/۵. مدل‌سازی آماری تحلیل‌های کسب‌وکار
- ۳/۶. مدل‌سازی رگرسیونی آمارهای استنتاجی
- ۳/۷. گزارش‌گیری کسب‌وکار
- ۳/۸. مصورسازی داده‌ها
- ۳/۹. انواع مختلف نمودارها و گراف‌ها
- ۳/۱۰. پیدایش تحلیل‌های دیداری
- ۳/۱۱. داشبوردهای اطلاعاتی

رادیوی سیریوس ایکس‌ام^۱ یک نیروگاه رادیوی ماهواره‌ای است که بزرگ‌ترین شرکت رادیویی جهان با درآمد سالانه‌ی ۳.۸ میلیارد دلار و طیف گسترده‌ای از ایستگاه‌های موسیقی، ورزشی، اخبار، سرگرمی و گفتگو می‌باشد. این شرکت که پخش برنامه‌های خود را با ۵۰,۰۰۰ مشترک در سال ۲۰۰۱ آغاز کرد در سال ۲۰۰۹ به ۱۸.۸ میلیون مشترک دست یافت و امروزه نزدیک به ۲۹ میلیون نفر مشترک دارد.

قسمت بیشتر رشد سیریوس ایکس‌ام تا به امروز ریشه در ترکیب‌های خلاقانه آن با شرکت‌های خودروسازی دارد. امروزه نزدیک به ۷۰ درصد از خودروهای جدید از سیریوس ایکس‌ام استفاده می‌کنند. اما دامنه‌ی دسترسی این شرکت فراتر از رادیوی خودروها در ایالات متحده رفته و حضوری جهانی در اینترنت، تلفن‌های هوشمند و سایر خدمات و کانال‌های توزیع از جمله سونوس^۲، جت‌بلو^۳ و دیش^۴ پیدا کرده است.

علیرغم این موفقیت‌های چشمگیر، تغییر جمعیت‌شناختی مشتریان، فناوری و یک چشم‌انداز رقابتی طی چند سال گذشته، شرکت سیریوس ایکس‌ام را با مجموعه چالش‌ها و فرصت‌های جدیدی روبرو کرده است. برخی از مهمترین آنها به شرح زیر می‌باشند:

- با افزایش نفوذ بازار شرکت در بین خودروهای نو، جمعیت‌شناسی خریداران نیز به سمت افراد جوان‌تر با درآمد کمتر تغییر کرده است. سیریوس ایکس‌ام چگونه می‌تواند به این گروه جدید از مشتریان دست پیدا کند؟
- با تبدیل خودروهای نو به دست دوم، سیریوس ایکس‌ام چگونه می‌تواند مالکان خودروهای دست دوم را شناسایی کرده و آنها را به مشتریانی برای خرید خودروهای نو تبدیل کند؟

¹ SiriusXM Radio

² SONOS

³ JetBlue

⁴ Dish

- با تملک کسب‌وکار وسایل نقلیه‌ی متصل از شرکت آجرو - یکی از پیشگامان در ارائه‌ی خدمات تلمتیک در بازار خودروی ایالات متحده، سیریوس‌ایکس‌ام قادر به ارائه خدمات خود از طریق شبکه‌های ماهواره‌ای و بی‌سیم شد. شرکت چگونه می‌تواند از این تملک برای کسب جریان‌های درآمدی جدید استفاده کند؟

سیریوس‌ایکس‌ام به این نکته پی برد که برای رسیدگی به این چالش‌ها باید به یک سازمان بازاریابی داده‌محور با عملکرد سطح بالا تبدیل شود. این شرکت با استقرار سه اصل تفکری بنیادی شروع به این تغییر کرد. ابتدا، تعاملات شخصی‌سازی شده به جای بازاریابی انبوه. شرکت به سرعت متوجه شد که برای بازاریابی شخصی‌تر باید بر اساس تاریخچه و تعاملات و همچنین درک مشتاقانه از جایگاه مصرف‌کننده در چرخه‌ی طول عمر نام‌نویسی عمل کند.

دوم، برای دستیابی به آن درک، فناوری اطلاعات و شرکای فناوری برون سازمانی آن باید قادر به ارائه‌ی داده‌های یکپارچه، تحلیل‌های پیشرفته، پلتفرم‌های بازاریابی یکپارچه و سیستم‌های تحویل چندکاناله باشند.

و سوم، شرکت بدون یک دیدگاه مشترک یکپارچه و منسجم نمی‌توانست به اهداف کسب‌وکار خود دست پیدا کند. مهمتر از همه اینکه، جنبه‌های فناوری و کسب‌وکار سیریوس‌ایکس‌ام برای اینکه به بهترین نحو بتواند به چالش‌های تبدیل شدن به یک سازمان بازاریابی سطح بالا بر اساس بینش‌های داده‌محور غلبه کند باید شرکای حقیقی می‌شدند.

آن بینش‌های داده‌محور به عنوان مثال شرکت را قادر می‌سازند تا بین مصرف‌کنندگان، مالکان، رانندگان، شنوندگان و صاحبان حساب تمایز قائل شوند. بینش‌ها به سیریوس‌ایکس‌ام کمک می‌کنند تا درک کنند که کدام وسایل نقلیه و خدمات دیگر نیز بخشی از هر خانوار هستند و فرصت‌های جدیدی را بر آن اساس خلق کنند. علاوه بر این، با ایجاد یک چشم‌انداز ۳۶۰ درجه منسجم و قابل اطمینان از تمامی مصرف‌کنندگان، شرکت می‌تواند اطمینان حاصل کند که تمامی پیام‌رسانی‌ها در کلیه تعاملات و طرح‌های تبلیغاتی به صورت سفارشی، مرتبط و سازگار در تمامی کانال‌ها پخش می‌شوند. نکته مهم این است که هرچه بازاریابی، سفارشی‌تر و مؤثرتر باشد از لحاظ هزینه نیز اقتصادی‌تر است.

شرکت سیریوس‌ایکس‌ام در زمانی که تصمیم گرفت تا تبدیل به یک شرکت بازاریابی با عملکرد سطح بالا شود، با یک پلتفرم بازاریابی شخص ثالث همکاری می‌کرد که ظرفیت لازم برای حمایت از بلندپروازی‌های سیریوس‌ایکس‌ام را نداشت. شرکت سپس تصمیم مهم و آینده‌نگرانه‌ای را گرفت تا فعالیت‌های بازاریابی را در درون سازمان انجام دهد و سپس به دقت برنامه‌ریزی لازم برای انتقال موفقیت‌آمیز این فعالیت را انجام داد.

۱. افزایش داده‌های تمیز از طریق بهبود مدیریت و حکمرانی داده‌های اصلی. اگرچه شرکت به طرز قابل درکی برای عملی‌سازی افکار خود عجله داشت ولی بهداشت داده‌ها اولین گام مهم برای ایجاد پنجره‌ای مطمئن به سوی رفتار مشتریان بود.

۲. آوردن تحلیل‌های بازاریابی به درون سازمان و گسترش انبار داده‌ها برای پشتیبانی کامل از تحلیل‌های بازاریابی یکپارچه.

۳. توسعه‌ی مدل‌های بخش‌بندی و امتیازدهی جدید برای اجرا در پایگاه‌های داده، حذف تأخیرها و کپی و تکثیر داده‌ها.

۴. توسعه‌ی انبار داده‌های یکپارچه برای انبارش داده‌های بازاریابی و امتیازدهی و بکارگیری از تحلیل‌های درون پایگاه داده.

۵. بکارگیری یک پلتفرم بازاریابی برای طراحی و توسعه‌ی طرح‌های تبلیغاتی.

۶. گردهم آوردن تمامی قابلیت‌های برای مدیریت بلادرنگ عرضه‌ی خدمات در سرتاسر کانال‌های بازاریابی اعم از مرکز تماس، موبایل، وب و اپلیکیشن‌ها.

تکمیل این گام‌ها به معنای یافتن شریک فناوری مناسب و درست است. سیریوس ایکس‌ام شرکت ترادیتا را به دلیل نقاط قوتش که گزینه‌های مناسب قدرتمندی برای این پروژه و شرکت بود انتخاب کرد. ترادیتا قابلیت‌های زیر را عرضه کرد:

- تثبیت و ادغام منابع داده در قالب یک انبار داده‌ی یکپارچه، تحلیل‌های پیشرفته و برنامه‌های بازاریابی قدرتمند.
- حل مسائل مربوط به تأخیر داده‌ها
- کاهش چشمگیر جابجایی داده‌ها بین پایگاه‌های داده و برنامه‌های متعدد.
- تعامل یکپارچه با برنامه‌ها و ماژول‌ها در تمامی حوزه‌های بازاریابی.
- عملکرد در سطوح بسیار بالا برای اجرای طرح‌های تبلیغاتی و تحلیل‌های درون پایگاه داده.
- برقراری ارتباطات بلادرنگ با مشتریان.
- ارائه‌ی پشتیبانی عملیاتی چه از طریق فناوری ابر یا سنتی.

این شراکت موجب شد تا سیریوس ایکس‌ام بتواند به آرامی و روانی در نقشه مسیر خود پیش رفته و در حال حاضر این شرکت در میانه‌ی فرایند پنج ساله‌ی تغییرات خود است. سیریوس ایکس‌ام پس از آنکه فرایند قدرتمند حکمرانی داده‌های خود را استقرار داد، شروع به پیاده‌سازی انبار داده‌های خود کرد که به شرکت اجازه می‌داد تا بینش و ادراک‌های کسب شده در سرتاسر سازمان را سریع و مطمئن به مرحله عملیاتی برساند.

سپس، شرکت «مدیر تعاملات مشتریان^۱» را پیاده‌سازی کرد که بخشی از «ابر بازاریابی یکپارچه‌ی ترادیتا^۲» بود که امکان برقراری تعاملات آنی و گفتمانی را در سرتاسر کانال‌های ارتباطی سنتی و دیجیتال فراهم می‌کرد. سیریوس ایکس‌ام همچنین از «مرکز پیام‌رسانی دیجیتال ترادیتا^۳» استفاده خواهد کرد.

تمامی این مجموعه قابلیت‌ها به سیریوس ایکس‌ام اجازه می‌دهد تا ارتباطات مستقیم را در سرتاسر کانال‌های متعدد خود مدیریت کند. این رشد و تکامل، ارائه‌ی خدمات در لحظه، پیام‌های بازاریابی و توصیه‌ها بر اساس رفتارهای قبلی مشتریان را امکان‌پذیر می‌سازد.

علاوه بر روان‌سازی روش اجرا و بهینه‌سازی فعالیت‌های بازاریابی برون‌سازمانی، سیریوس ایکس‌ام کنترل عملیات بازاریابی درون‌سازمانی خود را نیز با پیاده‌سازی «مدیریت منابع بازاریابی^۴» که بخشی از «ابر بازاریابی یکپارچه‌ی ترادیتا» است در

¹ Customer Interaction Manager

² Teradata Integrated Marketing Cloud

³ Teradata Digital Messaging Center

⁴ Marketing Resource Management

اختیار می‌گیرد. این راه حل به شرکت اجازه می‌دهد تا جریان کاری خود را روان کرده، منابع بازاریابی را بهینه‌سازی کرده و حداکثر بهره را از بودجه بازاریابی خود کسب نماید.

همچنان که سیریوس ایکس‌ام در مسیر رشد خود به سوی یک سازمان بازاریابی سطح بالا حرکت می‌کند، از منافع حاصل از اجرای متفکرانه استراتژی خود بهره‌مند شده است. سیریوس ایکس‌ام با بهره‌گیری از بینش کسب شده از مشتریان خود تا سطح خانوار و یک چشم‌انداز کامل از استراتژی بازاریابی تماسی قادر خواهد بود تا خدمات خود را به شکلی هدمندتر در سطح خانوار، مصرف‌کننده و دستگاه ارائه کند. سیریوس ایکس‌ام با آوردن داده‌ها و قابلیت‌های تحلیل بازاریابی به داخل شرکت به موارد زیر دست یافت:

- طرح‌های تبلیغاتی به جای چهار روز به صورت آنی جواب می‌دهند و موجب کاهش انبوه زمان‌های سیکل طرح‌های تبلیغاتی و تحلیلگرهایی که از آنها پشتیبانی می‌کنند می‌شود.
- شفافیت حلقه بسته که به تحلیلگرها اجازه می‌دهد تا گفتگوهای چندمرحله‌ای و اصلاحات در حین برگزاری طرح‌های تبلیغاتی داشته باشند تا بدین ترتیب بر اثربخشی طرح‌های تبلیغاتی بیفزایند.
- مدل‌سازی و امتیازدهی لحظه‌ای برای افزایش اطلاعات بازاریابی و متمرکزسازی طرح‌های تبلیغاتی و پاسخگویی با سرعت بالا.

نهایتاً، تجربه‌ی سیریوس ایکس‌ام این تفکر را تقویت کرده است که بازاریابی با عملکرد بالا به طور دائم و مستمر در حال رشد و تکامل است. این شرکت هم فناوری و هم فرایندهای مورد نیاز برای رشد مستمر و منعطف خود را پیاده‌سازی و اجرا کرده است.

سوالاتی برای مرور پرونده‌ی آغازین

۷. سیریوس ایکس‌ام چه کاری انجام می‌دهد؟ کسب‌وکار این شرکت در چه بازاری است؟
۸. چالش‌های آن چه بوده‌اند؟ در خصوص فناوری و چالش‌های مربوط به داده‌ها اظهار نظر کنید.
۹. راه‌حل‌های پیشنهادی چه بوده‌اند؟
۱۰. شرکت چگونه راه‌حل‌های پیشنهادی را پیاده‌سازی کرد؟ آیا با چالش‌هایی در پیاده‌سازی مواجه شد؟
۱۱. نتایج و منافع اینکار چه بودند؟ چرا اینکار ارزش تلاش/سرمایه‌گذاری داشت؟
۱۲. آیا می‌توانید شرکت‌های دیگری که با چالش‌های مشابه روبرو شده‌اند و توانسته‌اند از راه‌حل‌های بازاریابی داده‌محور مشابه بهره‌برده‌اند را نام ببرید؟

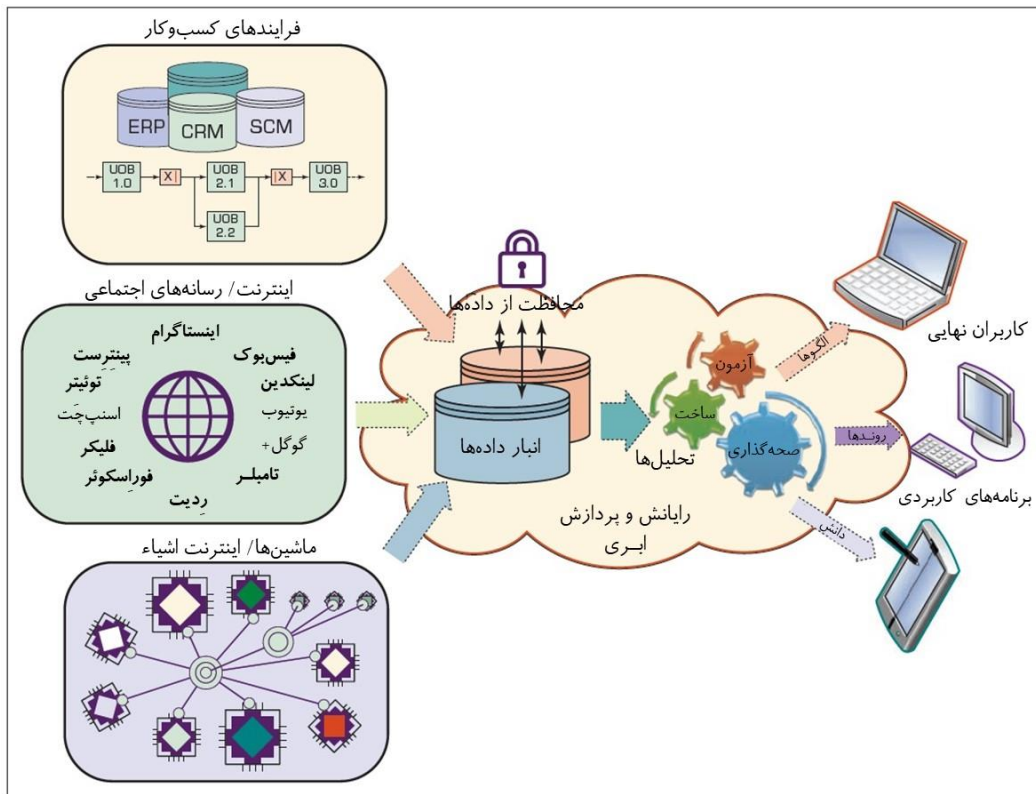
سیریوس ایکس‌ام در تلاش برای توسعه و ترقی در یک صنعت رقابتی که به سرعت در حال تغییر است متوجه شد که برای تسری مؤثر ارزش‌های پیشنهادی خود به مشتریان فعلی و بالقوه به یک زیرساخت بازاریابی جدید و بهبود یافته (که متکی بر داده‌ها و تحلیل‌ها باشد) نیاز دارد. همانند هر صنعتی، موفقیت یا حتی بقا در صنعت سرگرمی به دریافت هوشمندانه‌ی روندهای متغیر

(پسندیدن‌ها و نپسندیدن‌ها) و کنار هم قراردادن پیام‌ها و سیاست‌های درست برای جلب مشتریان جدید و در عین حال، حفظ مشتریان فعلی بستگی دارد. کلید این کار، ایجاد و مدیریت طرح‌های تبلیغاتی بازاریابی موفق است که همسو با جامعه‌ی مشتریان هدف بوده و از یک حلقه‌ی بازخور بسته برای تنظیم و اصلاح پیام‌ها برای دستیابی به بهترین نتایج برخوردار باشد. در پایان، دقت در اجرای کسب‌وکار سیریوس‌ایکس‌ام حرف اول و آخر را می‌زند یعنی اتخاذ رویکرد کنشگرانه نسبت به ماهیت متغیر مشتریان و ایجاد و ارسال به موقع محصولات و خدمات مناسب و درست با استفاده از یک استراتژی بازاریابی کل‌نگر و داده‌محور/واقعیت‌محور. شناسایی منابع، ایجاد منابع، دسترسی و تخصیص، یکپارچه‌سازی، پاکسازی، تغییرشکل، انبارش و پردازش داده‌های مرتبط، نقش بسیار مهمی را در موفقیت سیریوس‌ایکس‌ام در طراحی و اجرای یک استراتژی تحلیلی بازاریابی ایفا کرده است که البته این موضوع برای هر شرکتی صرفنظر از صنعتی که در آن فعالیت دارد صدق می‌کند.

Sources: C. Quinn, "Data-Driven Marketing at SiriusXM," Teradata Articles & News, 2016. <http://bigdata.teradata.com/US/Articles-News/Data-Driven-Marketing-At-SiriusXM/> (accessed August 2016); "SiriusXM Attracts and Engages a New Generation of Radio Consumers." <http://assets.teradata.com/resourceCenter/downloads/CaseStudies/EB8597.pdf?processed=1>.

داده‌ها، جزء تشکیل دهنده‌ی اصلی هر راهکار هوش کسب‌وکار، علم داده‌ها و تحلیل‌های کسب‌وکار می‌باشند. در واقع، داده‌ها را می‌توان به صورت مواد اولیه‌ی فناوری‌های تصمیم‌گیری متداول جهت تولید اطلاعات، بینش و دانش دانست. بدون داده‌ها، هیچیک از این فناوری‌ها وجود نداشته و عمومیت نخواهند یافت. اگرچه ما مدل‌های تحلیلی را با استفاده از دانش و تجربه افراد خبره و بدون استفاده از داده‌ها یا حجم ناچیزی از داده‌ها ساخته‌ایم ولی آن روزها به گذشته تعلق دارد و امروزه داده‌ها، جوهره‌ی هر چیزی هستند. با توجه به چالش عظیم گردآوری، انبارش و مدیریت داده‌ها، امروزه داده‌ها به عنوان ارزشمندترین دارایی یک سازمان شناخته می‌شوند که از توان بالقوه برای ایجاد بینش‌های فوق‌العاده ارزشمند جهت درک بهتر مشتریان، رقبا و فرایندهای کسب‌وکار برخوردارند.

داده‌ها ممکن است کوچک یا خیلی بزرگ باشند. ممکن است ساختیافته (به خوبی برای پردازش توسط رایانه‌ها سازماندهی شده) یا ساختنیافته باشند (مثلاً متونی که مخاطب آن انسان‌ها هستند و رایانه‌ها به راحتی قادر به درک/استفاده از آنها نیستند). داده‌ها ممکن است در قالب بسته‌های کوچک به طور مستمر آمده یا اینکه به یکباره به صورت یک بسته‌ی بزرگ بیایند. اینها برخی از ویژگی‌های معرف ماهیت ذاتی داده‌های امروزی هستند که ما اغلب آنها را کلان‌داده می‌نامیم. اگرچه این ویژگی‌های داده‌ها موجب شده تا پردازش و استفاده از آنها چالش‌برانگیز شود ولی به دلیل غنا بخشیدن به داده‌ها و امکان کشف دانش نوین و جدید، آنها را ارزشمندتر کرده است. روش‌های قدیمی گردآوری دستی داده‌ها (یا از طریق نظرسنجی‌ها یا تراکنش‌های کسب‌وکار که داده‌ها توسط انسان وارد می‌شوند) عمدتاً جای خود را به مکانیزم‌های پیشرفته‌ی امروزی گردآوری داده‌ها که از اینترنت و/یا شبکه‌های رایانه‌ای مبتنی بر شناسایی فرکانس رادیویی (آر‌اف‌آی‌دی)/حسگرها استفاده می‌کنند داده‌اند. این سیستم‌های خودکار گردآوری داده‌ها نه تنها ما را قادر به گردآوری حجم بیشتری از داده‌ها کرده‌اند بلکه موجب غنای کیفیت و یکپارچگی داده‌ها نیز شده‌اند. شکل ۳.۱ یک پیوستار تحلیل متداول را از مرحله‌ی داده‌ها به تحلیل‌ها تا اطلاعات قابل اجرا به تصویر کشیده است.



شکل ۳.۱ پیوستار تبدیل داده‌ها به دانش

اگرچه گزاره ارزشی داده‌ها انکارناپذیر است ولی برای اینکه بتوانند انتظارات را برآورده کنند باید با برخی سنج‌های کیفی و قابلیت استفاده‌ی پایه وفق و سازگاری داشته باشند. مشخصاً، تمامی داده‌ها برای تمامی وظایف مناسب نیستند. این بدان معناست که داده‌ها باید با وظیفه‌ای که قرار است با آن بکار برده شود تطابق داشته باشد. حتی برای یک وظیفه‌ی خاص، داده‌های مرتبط موجود باید با الزامات کیفی و کمی تطابق داشته باشند. داده‌ها لزوماً باید آماده‌ی تحلیل باشند. پس آماده کردن داده‌ها برای تحلیل به چه معناست؟ علاوه بر مرتبط بودن آن با مسئله‌ی موجود و الزامات کمی/کیفی، داده‌ها باید ساختار مشخصی به همراه متغیرها/فیلدهای اصلی و مقادیری که به خوبی نرمال‌سازی شده‌اند داشته باشند.

علاوه بر این، باید تعریف مورد توافقی در سطح سازمان برای متغیرهای متداول و موضوعات (که برخی اوقات مدیریت داده‌های اصلی نامیده می‌شود) از قبیل چگونگی تعریف یک مشتری (از کدام ویژگی‌های مشتری برای تولید بازنمایی کل‌نگر برای تحلیل‌ها استفاده می‌شود) و در کجای فرایند کسب‌وکار است که اطلاعات مرتبط با مشتری اخذ، اعتبارسنجی، ذخیره و به‌روزآوری می‌شود. برخی اوقات، بازنمایی داده‌ها به نوع تحلیلی که بکار می‌رود بستگی دارد. الگوریتم‌های پیشگویانه معمولاً به یک فایل مسطح و یک متغیر هدف نیاز دارند و بدین ترتیب، آماده‌سازی داده‌ها برای پیش‌بینی به این معناست که مجموعه داده‌ها باید به قالب یک فایل مسطح تبدیل شده و آماده‌ی تغذیه به الگوریتم‌های پیشگویانه شوند. انطباق دادن داده‌ها با نیازها و خواسته‌های یک الگوریتم پیشگویانه‌ی خاص و/یا یک ابزار نرم‌افزاری نیز امری الزامی است. به عنوان مثال، در الگوریتم‌های شبکه عصبی، تمامی متغیرهای ورودی باید به صورت عددی بازنمایی شوند (حتی متغیرهای اسمی باید به متغیرهای شبه عددی دودویی تبدیل شوند) در حالیکه

الگوریتم‌های درخت تصمیم به چنین تبدیل عددی نیاز ندارند و به آسانی می‌توانند با ترکیبی از متغیرهای عددی و اسمی کار کنند.

پروژه‌های تحلیلی که وظایف مرتبط با داده‌ها را نادیده می‌گیرند اغلب قادر به یافتن پاسخ درست برای مسئله نیستند و این جواب‌های به ظاهر خوب و درست که به نحوی ناخواسته ایجاد شده‌اند می‌توانند منجر به تصمیمات نادقیق و بی‌موقع شوند. در ادامه، برخی از ویژگی‌های (سنجه‌های) معرف سطح آمادگی داده‌ها برای یک مطالعه‌ی تحلیلی آورده شده است (دیلن^۱ ۲۰۱۵، کوک، مک‌کوین و کورنر^۲ ۱۹۹۷).

- **قابلیت اطمینان منبع داده‌ها.** این عبارت به اصالت و میزان مناسب بودن وسیله‌ی انبارشی که داده‌ها از آنجا به دست آمده است اشاره دارد – پاسخ به این سوال که « آیا ما به این منبع داده اطمینان و اعتماد کافی داریم؟ » در صورت امکان، همواره باید به دنبال منبع اصلی داده‌ها باشیم تا امکان سوء نمایش و تغییرشکل داده‌ها که ناشی از جابجایی آنها از منبع به مقصد در طی یک یا چند گام است را از بین برده یا از آن پیشگیری کنیم. هرگونه جابجایی داده‌ها، این احتمال را به وجود می‌آورد که بخشی از داده‌ها به طور ناخواسته از قلم افتاده یا شکل و قالب آنها تغییر کند که انسجام آنها و شاید دقت حقیقی مجموعه داده‌ها را مخدوش می‌نماید.
- **دقت محتوای داده‌ها.** این بدان معناست که داده‌ها صحیح بوده و به خوبی با مسئله‌ی تحلیل انطباق دارند – پاسخ به این سوال که « آیا ما داده‌های درستی را برای کارمان در اختیار داریم؟ » داده‌ها باید نمایانگر آن چیزی باشند که توسط منبع اصلی داده‌ها تعریف یا تعیین شده است. به عنوان مثال، اطلاعات تماس مشتریان که در داخل یک پایگاه داده ثبت شده باید دقیقاً همان چیزی باشد که مشتری اعلان کرده است. در خصوص دقت داده‌ها به تفصیل در بخش بعدی صحبت شده است.
- **دسترسی‌پذیری داده‌ها.** این عبارت بدین معناست که داده‌ها به راحتی و آسانی قابل دستیابی باشند و پاسخ به این سوال که « آیا به آسانی می‌توانیم داده‌ها را زمانیکه به آنها نیاز داریم به دست بیاوریم؟ » دسترسی به داده‌ها می‌تواند دشوار باشد به ویژه اگر این داده‌ها در بیش از یک محل یا وسیله‌ی انبارش ذخیره شده باشند و برای دسترسی و دستیابی به آنها لازم باشد که آنها را ادغام نموده یا تغییرشکل داد. از آنجائیکه سیستم‌های مدیریت پایگاه‌های داده رابطه‌ای سنتی جای خود را به سیستم‌های جدیدتر می‌دهند (یا به طور همزمان در کنار نسل جدید وسیله‌های انبارش داده‌ها از قبیل دریاچه داده‌ها^۳ یا زیرساخت هَدوپ^۴ حضور دارند)، اهمیت/حساسیت دسترسی‌پذیری به داده‌ها نیز افزایش می‌یابد.
- **امنیت و حریم خصوصی داده‌ها.** امنیت داده‌ها بدین معناست که فقط کسانی که اجازه و نیاز به دسترسی به داده‌ها را دارند از آنها برخوردار شوند و از دسترسی هرکس دیگری به آنها جلوگیری به عمل آید. افزایش محبوبیت مدارک آموزشی و برنامه‌های اعطای گواهی تضمین اطلاعات^۵، مدرکی دال بر حساسیت و افزایش ضرورت استفاده از این سنجه‌ی کیفیت داده‌ها می‌باشد. هر سازمانی که سوابق سلامت و بهداشت بیماران را نگهداری می‌کند باید سیستم‌هایی

¹ Delen

² Kock, McQueen & Corner

³ Data Lake

⁴ Hadoop

⁵ Information Assurance

داشته باشند که نه تنها از داده‌ها در برابر دسترسی غیرمجاز محافظت کنند (که توسط قوانین فدرال از قبیل *قانون انتقال و پاسخ‌گویی الکترونیک بیمه سلامت* [آی‌ای‌پی‌ای] ^۱ الزام می‌شود)، بلکه هر بیمار را برای دسترسی مناسب و به موقع به سوابق دقیقاً شناسایی کنند (آناس ^۲ ۲۰۰۳).

- **غنا‌ی داده‌ها.** این بدان معناست که تمامی عناصر داده‌ای مورد نیاز در مجموعه داده‌ها وجود دارد. در اصل، غنا (یا جامعیت) بدان معناست که متغیرهای موجود، تعداد کافی و غنی از ابعاد موضوع مورد بحث را برای یک مطالعه‌ی تحلیلی دقیق و ارزشمند به تصویر بکشند. همچنین بدین معناست که محتوای اطلاعاتی برای ساخت یک مدل تحلیل پیشگویانه و/یا تجویزی، کامل (یا تقریباً کامل) است.
- **سازگاری داده‌ها.** این بدان معناست که داده‌ها به دقت گردآوری و ترکیب/ادغام شده‌اند. داده‌های سازگار، بیانگر اطلاعات بُعدی (متغیرهای مورد نظر) است که از منابع مجزا ولی با موضوع یکسان می‌آیند. اگر یکپارچه‌سازی/ادغام داده‌ها به درستی انجام نشود، برخی از متغیرها که موضوعات متفاوتی دارند ممکن است در رکورد یکسانی پدیدار شوند یعنی سوابق دو بیمار متفاوت درهم می‌آمیزد. به عنوان مثال، چنین حالتی ممکن است در زمان ادغام سوابق داده‌های نتایج تست بالینی و جمعیت‌شناسی رخ دهد.
- **دقت داده‌ها/به‌هنگامی داده‌ها.** این بدان معناست که داده‌ها باید برای استفاده در یک مدل تحلیل مفروض، به‌روز باشند (یا تا جائیکه لازم است جدید باشند). همچنین بدین معناست که داده‌ها دقیقاً در زمان رویداد یا مشاهده یا نزدیک به آن ثبت می‌شوند تا از یادآوری و رمزنگاری غیرواقعی و نادرست داده‌ها که ناشی از تأخیرهای زمانی است جلوگیری به عمل آید. از آنجائیکه تحلیل‌های دقیق به داده‌های دقیق و به‌هنگام بستگی دارد، یکی از ویژگی‌های اساسی داده‌های آماده برای تحلیل، ایجاد و دسترسی به‌هنگام به عناصر داده‌ها است.
- **دانه‌بندی داده‌ها.** در اینجا متغیرها و مقادیر داده‌ای باید در پایین‌ترین سطح از جزئیات و برای استفاده‌ی مورد نظر از داده‌ها تعریف شوند. اگر داده‌ها تجمیع شده باشند، ممکن است حاوی جزئیات مورد نیاز برای الگوریتم‌های تحلیل نباشند تا این الگوریتم‌ها بتوانند یاد بگیرند که چگونه سوابق/پرونده‌های متفاوت را از یکدیگر تشخیص دهند. به عنوان مثال، در یک محیط پزشکی، مقادیر عددی نتایج آزمایشگاهی باید تا چندین رقم اعشار که برای تفسیر معنادار نتایج تست و بکارگیری آن مقادیر در یک الگوریتم تحلیلی مورد نیاز است ثبت شوند. به همین منوال، در گردآوری داده‌های جمعیتی، عناصر داده‌ها باید در یک سطح دانه‌بندی تعریف شوند تا بتوان تفاوت‌های بین نتایج حاصل از مراقبت‌ها بین زیرگروه‌های جمعیتی مختلف را تعیین نمود. آنچه باید به خاطر داشت این است که داده‌های تجمیع شده را نمی‌توان بدون دسترسی به منبع اصلی به حالت اولیه بازگرداند ولی به آسانی می‌توان آنها را از حالت دانه‌بندی شده مجدداً تجمیع نمود.
- **اعتبار داده‌ها.** این عبارتی است که برای توصیف یک تطبیق/عدم تطبیق بین مقادیر داده‌ای واقعی و مورد انتظار برای یک متغیر خاص بکار می‌رود. مقادیر قابل قبول یا دامنه‌ی مقادیر برای هر جزء داده باید به عنوان بخشی از تعریف داده‌ها تعیین شود. به عنوان مثال، یک تعریف معتبر از داده‌های مربوط به جنسیت باید شامل سه مقدار باشد: مذکر، مؤنث و نامشخص.

¹ Health Insurance Portability and Accountability Act (HIPAA)

² Annas

- **مرتبط بودن داده‌ها.** این بدان معناست که متغیرهای یک مجموعه داده با موضوع مورد مطالعه، مرتبط هستند. مرتبط بودن، یک سنجی دوبرخی نیست (یعنی اینکه یک متغیر، مرتبط است یا مرتبط نیست)، بلکه بیانگر یک طیف از مرتبط بودن از کمترین میزان ارتباط تا بیشترین میزان ارتباط است. بر اساس الگوریتم‌های تحلیلی مورد استفاده می‌توانیم انتخاب کنیم که آیا فقط از مرتبط‌ترین اطلاعات استفاده کنیم (یعنی متغیرها) یا اگر الگوریتم مورد نظر قادر به مرتب‌سازی این داده‌ها می‌باشد، تمامی اطلاعات مرتبط را صرفنظر از سطح مرتبط بودن آنها بکار بگیریم. یکی از مواردی که در مطالعات تحلیلی باید از آن اجتناب کرد این است که داده‌های کاملاً نامرتبط را در ساخت مدل بکار بگیریم چون که اینکار می‌تواند موجب آلودگی اطلاعات به دست آمده از الگوریتم شده و منجر به نتایج نادقیق و گمراه کننده شود.

ویژگی‌هایی که در بالا گفته شد را شاید بتوان متداول‌ترین سنجه‌هایی دانست که باید از آنها مطلع باشیم. کیفیت حقیقی داده‌ها و آمادگی عالی برای تحلیل در یک حوزه‌های کاربردی خاص، نیازمند این هستند که سطوح متفاوتی از تأکیدها را باید بر این ابعاد سنجی در نظر داشت و شاید تأکید بیشتری را به این گروه از ویژگی‌ها داشت. در بخش بعدی به بررسی ماهیت داده‌ها از یک منظر نظام طبقه‌بندی برای فهرست‌بندی و تعریف انواع مختلف داده‌ها با توجه به ارتباط آنها با پروژه‌های تحلیلی مختلف خواهیم پرداخت.

سوالاتی برای مرور بخش ۳.۲

۹. اهمیت داده‌ها در تحلیل را چگونه توضیح می‌دهید؟ آیا می‌توانیم تحلیل را بدون داده‌ها تصور کنیم؟
۱۰. با توجه به تعریف جدید و گسترده‌ی تحلیل کسب‌وکار، ورودی‌ها و خروجی‌های اصلی پیوستار تحلیل چه هستند؟
۱۱. داده‌های مورد نیاز برای تحلیل کسب‌وکار از کجا می‌آیند؟
۱۲. به نظر شما، سه چالش اصلی مربوط به داده‌ها برای تحلیل‌های بهتر کدام‌ها هستند؟
۱۳. متداول‌ترین سنجه‌هایی که داده‌ها را برای تحلیل‌ها آماده می‌کنند کدام‌ها هستند؟

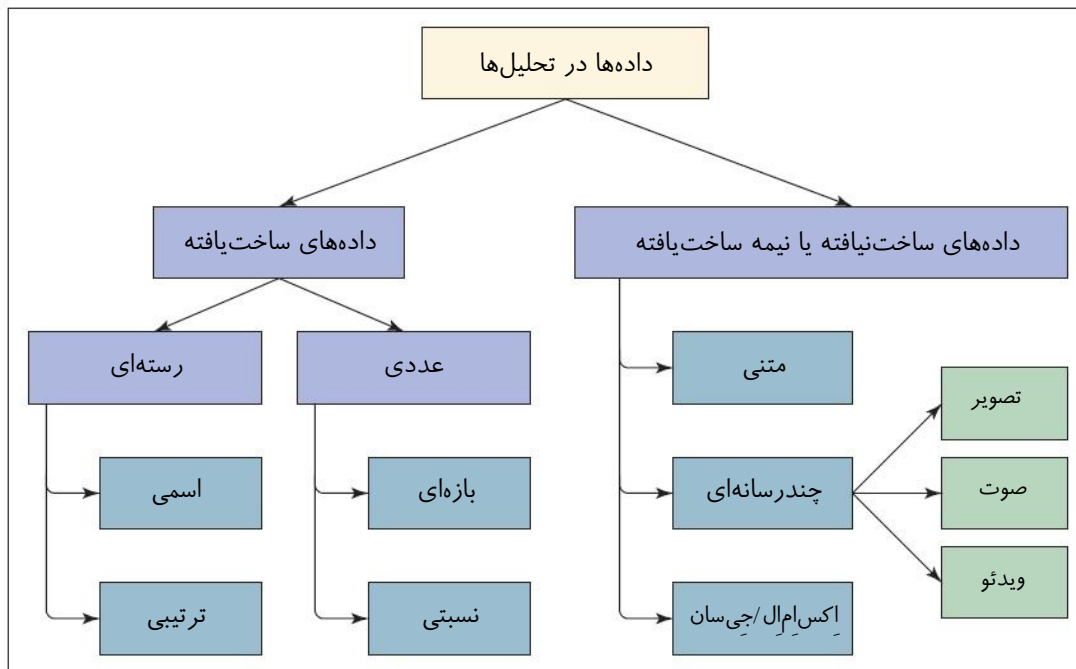
عبارت *داده‌ها* (که حالت مفرد آن را داده می‌گویند) به مجموعه‌ای از حقایق که معمولاً در نتیجه‌ی آزمایشات، مشاهدات، تراکنش‌ها یا تجربیات به دست می‌آیند اشاره دارد. داده‌ها می‌توانند شامل اعداد، حروف، واژگان، تصاویر، صداهای ضبط شده و غیره باشند که به مجموعه‌ای از متغیرها (ویژگی‌های موضوع یا رویدادی که ما علاقمند به مطالعه‌ی آن هستیم) را اندازه‌گیری می‌کنند. داده‌ها اغلب به عنوان پایین‌ترین سطح چکیدگی یا انتزاع شناخته می‌شوند که اطلاعات و سپس دانش از آنها استخراج می‌شود.

در بالاترین سطح انتزاع، داده‌ها را می‌توان به صورت داده‌های ساخت‌یافته و ساخت‌نیافته (یا نیمه‌ساخت‌یافته) طبقه‌بندی کرد. **داده‌های ساخت‌نیافته** / نیمه ساخت‌یافته به هر ترکیبی از محتوای متنی، تصویری، صوتی و وب گفته می‌شوند. در فصل مربوط به متن کاوی و وب کاوی به تفصیل در مورد داده‌های ساخت‌نیافته و نیمه ساخت‌یافته صحبت خواهیم کرد. الگوریتم‌های داده کاوی از **داده‌های ساخت‌یافته** استفاده می‌کنند و می‌توان آنها را بر حسب عددی یا رسته‌ای طبقه‌بندی کرد. **داده‌های رسته‌ای**^۱ را

¹ Categorical data

می‌توان به داده‌های اسمی یا داده‌های ترتیبی^۱ تقسیم کرد در حالیکه داده‌های عددی را می‌توان به بازه‌ها یا نسبت‌ها تقسیم کرد. در شکل ۳.۲ یک طبقه‌بندی ساده از داده‌ها^۲ نشان داده شده است.

- **داده‌های رشته‌ای.** این داده‌ها معرف برچسب‌های چندین کلاس هستند که برای تقسیم یک متغیر به گروه‌های خاص بکار می‌روند. مثال‌هایی از متغیرهای رشته‌ای شامل نژاد، جنسیت، سن، گروه و سطح تحصیلات است. اگرچه دو متغیر آخر را می‌توان با بکارگیری مقادیر دقیق برای سن و بالاترین سطح تحصیلات، جزو داده‌های عددی نیز به حساب آورد ولی در اغلب موارد بهتر است که اینگونه متغیرها را به صورت دسته‌های مرتب شده‌ی نسبتاً کوچکی دسته‌بندی کنیم. داده‌های رشته‌ای را داده‌های گسسته نیز می‌نامند زیرا معرف تعداد محدودی از مقادیری هستند که هیچ پیوستاری بین آنها وجود ندارد. حتی اگر از مقادیر عددی برای متغیرهای رشته‌ای یا گسسته استفاده شده باشد، این اعداد، چیزی بیش از نشانه‌ها نبوده و هیچگونه امکان انجام محاسبات مقادیر کسری بر روی آنها وجود ندارد.
- **داده‌های اسمی.** این داده‌ها شامل اندازه‌گیری کدهای ساده‌ای هستند که به صورت برچسب به اشیائی که اندازه‌گیری نیستند تخصیص داده شده‌اند. به عنوان مثال، متغیر وضعی تأهل را به طور کلی می‌توان به صورت (۱) مجرد، (۲) متأهل و (۳) مطعلقه دسته‌بندی کرد. **داده‌های اسمی** را می‌توان با مقادیر دوجمله‌ای که دارای دو مقدار ممکن هستند (مثلاً بله/خیر، درست/غلط، خوب/بد) یا مقادیر چندجمله‌ای که دارای سه یا چند مقدار ممکن هستند نشان داد (مثلاً قهوه‌ای/سبز/آبی، سفید پوست/سیاه پوست/آمریکای لاتین/زرد پوست، مجرد/متأهل/مطعلقه).



شکل ۳.۲ یک طبقه‌بندی ساده‌ی داده‌ها

¹ Ordinal data

² Data taxonomy

- **داده‌های ترتیبی.** اینها حاوی کدهایی هستند که به عنوان برچسب به اشیاء یا رویدادهایی تخصیص یافته‌اند که بیانگر ترتیب یا رتبه نیز بین آنها می‌باشند. به عنوان مثال، متغیر *امتیاز/اعتبار* را به طور کلی می‌توان به صورت (۱) کم، (۲) متوسط و (۳) زیاد دسته‌بندی کرد. ارتباطات رتبه‌بندی شده مشابهی را می‌توان در متغیرهایی از قبیل گروه سنی (یعنی کودک، جوان، میان‌سال، سالمند) و سطح تحصیلات (یعنی دبیرستان، کالج، دانشگاه) یافت. برخی الگوریتم‌های تحلیلی پیشگویانه از قبیل *رگرسیون لجستیک چندگانه‌ی ترتیبی*^۱، این اطلاعات رتبه-ترتیب را برای ایجاد یک مدل دسته‌بندی بهتر در نظر می‌گیرند.

- **داده‌های عددی.** اینها بیانگر مقادیر عددی برای متغیرهای خاص هستند. مثال‌هایی از متغیرهای با مقادیر عددی شامل سن، تعداد فرزندان، درآمد کل خانوار (بر حسب دلار آمریکا)، مسافت سفر (بر حسب مایل) و دما (بر حسب درجه‌ی فارنهایت) می‌باشد. مقادیر عددی بیانگر یک متغیر هستند که می‌تواند به صورت عدد صحیح (فقط اعداد کامل) یا حقیقی (شامل اعداد کسری نیز می‌شود) باشد. داده‌های عددی، داده‌های پیوسته نیز نامیده می‌شوند که بدین معناست که شامل مقادیر پیوسته در یک مقیاس خاص هستند که می‌توان مقادیر میانی را نیز به آنها افزود. برخلاف متغیرهای گسسته که بیانگر داده‌های محدود و قابل شمارش هستند، یک متغیر پیوسته بیانگر اندازه‌گیری‌های مقیاس پذیر است که یک داده می‌تواند حاوی تعداد نامحدودی مقادیر کسری باشد.

- **داده‌های بازه‌ای.** اینها متغیرهایی هستند که می‌توان آن را در مقیاس‌های بازه‌ای اندازه‌گیری نمود. یک مثال متداول از اندازه‌گیری مقیاس بازه‌ای، دما در مقیاس سلسیوس است. در این مقیاس ویژه، واحد اندازه‌گیری به صورت ۱/۱۰۰ اختلاف بین دمای ذوب و دمای جوش آب در فشار اتمسفر تعریف می‌شود یعنی مقدار صفر مطلق وجود ندارد.

- **داده‌های نسبتی.** اینها شامل متغیرهای اندازه‌گیری در علوم فیزیکی و مهندسی هستند. جرم، طول، زمان، زاویه‌ی سطح، انرژی و بار الکتریکی مثال‌هایی از سنج‌های فیزیکی در مقیاس نسبت هستند. نوع مقیاس، نام خود را از این واقعیت گرفته است که اندازه‌گیری به صورت یک برآورد یا تخمین از نسبت بین بزرگی یک کمیت پیوسته و یک بزرگی واحد از همان نوع است. ویژگی متمایزکننده‌ی یک مقیاس نسبت، داشتن یک مقدار صفر غیرتصادفی است. به عنوان مثال، مقیاس دمای کلونین دارای نقطه صفر غیرتصادفی صفر مطلق است که معادل ۲۷۳.۱۵- درجه سلسیوس است. این نقطه صفر، غیرتصادفی است زیرا ذراتی که در این دما ماده را تشکیل می‌دهند دارای انرژی جنبشی صفر هستند.

سایر انواع داده‌ها از جمله متنی، تصویری، ویدئویی و صوتی پیش از آنکه بتوان آنها را توسط روش‌های تحلیلی پردازش کرد (الگوریتم‌های داده کاوی، دین ۲۰۱۵) باید به شکل رسته‌ای یا عددی تبدیل شوند. داده‌ها را همچنین می‌توان بر حسب ایستا یا پویا دسته‌بندی کرد (یعنی زمانی یا سری‌های زمانی).

برخی روش‌های تحلیل پیشگویانه (یعنی داده کاوی) الگوریتم‌های یادگیری ماشین در خصوص نوع داده‌هایی که می‌توانند استفاده کنند بسیار انتخابی هستند. در صورتیکه داده‌های ناسازگار در این الگوریتم‌ها و روش‌های بکار گرفته شود می‌تواند منجر به مدل‌های غلط شده و فرایند توسعه مدل را متوقف کند. به عنوان مثال، در برخی روش‌های داده کاوی باید تمامی متغیرها (هم ورودی و هم خروجی) به صورت متغیرهای مقادیر عددی تعریف شوند (مثلاً شبکه‌های عصبی، ماشین‌های بردار پشتیبانی، رگرسیون لجستیک). متغیرهای اسمی یا ترتیبی را با استفاده از نوعی شبه متغیرهای یک به چند (مثلاً یک متغیر رسته‌ای با سه

¹ Ordinal multiple logistic regression

مقدار منحصر بفرد را می‌توان به سه شبه‌متغیر با مقادیر دودویی یعنی ۱ و ۰ تبدیل کرد) می‌توان به صورت عددی تبدیل کرد. چونکه این فرایند می‌تواند تعداد متغیرها را افزایش دهد، در خصوص تأثیر اینگونه نمایش‌ها به ویژه برای متغیرهای رسته‌ای که تعداد زیادی مقادیر منحصر بفرد دارند باید محتاطانه عمل کرد.

به همین منوال، برخی روش‌های تحلیل پیشگوانه مثل یک الگوریتم درخت تصمیم کلاسیک به نام آی‌دی ۳ و مجموعه‌های دقیق (یک الگوریتم استنتاج قاعده‌ای نسبتاً جدید) تماماً به متغیرهایی با مقادیر رسته‌ای نیاز دارند. در نسخه‌های اولیه‌ی این روش‌ها، کاربر پیش از آنکه بتواند متغیرها را توسط الگوریتم پردازش کند، باید مقادیر عددی را به صورت رسته‌ای نمایش می‌داد. خبر خوب اینکه در اکثر پیاده‌سازی‌های این الگوریتم‌ها در ابزارهای نرم‌افزاری موجود، ترکیبی از متغیرهای عددی و اسمی را می‌توان استفاده کرد و این نرم‌افزارها می‌توانند پیش از پردازش داده‌ها، مقادیر را تبدیل کنند.

داده‌ها در انواع و طرح‌های نمایشی بسیاری وجود دارند. ابزارهای تحلیل کسب‌وکار برای کمک به دانشمندان علوم داده در تبدیل داده‌ها و نمایش داده‌ها دائماً در حال بهبود هستند تا نیازهای داده‌ای مدل‌ها و الگوریتم‌های پیشگوانه خاص را بتوان به شکلی مناسب اجرا نمود. مورد کاربردی ۳.۱ یک سناریوی کسب‌وکار را به تصویر کشیده است که یکی از بزرگ‌ترین شرکت‌های مخابراتی از انواع مختلف منابع داده‌ای غنی برای تولید بینش مشتریان برای پیشگیری از رویگردانی^۱ و ایجاد منابع درآمدی جدید استفاده کرده است.

وریزون به فراخوانه نوآوری پاسخ می‌دهد: بزرگ‌ترین ارائه دهنده‌ی خدمات شبکه در کشور از تحلیل‌های پیشرفته برای خدمات آینده برای مشتریان استفاده می‌کند.

مورد کاربردی ۳.۱

مسئله

در صنعت مخابرات شدیداً رقابتی، حفظ ارتباط با مشتریان و در عین حال، یافتن منابع درآمدی جدید اهمیت حیاتی دارد به ویژه چونکه منابع درآمدی جای در حال کاهش هستند. برای وریزون که جزو ۱۳ شرکت فورچون است، سلاح مخفی که شرکت را به سمت بزرگ‌ترین و مطمئن‌ترین ارائه دهنده‌ی خدمات شبکه‌ای در کشور پرتاب کرد، راهنمای این شرکت در جهت موفقیت‌های آتی است. سلاح مخفی؟ داده‌ها و تحلیل. چونکه شرکت‌های مخابراتی معمولاً از لحاظ داده‌ها غنی هستند، داشتن یک راه‌حل تحلیلی و پرسنل درست می‌تواند منج به کشف بینش‌های بحرانی که برای تمامی حوزه‌های سازمان سودمند هستند شود.

¹ Churn

Verizon by the Numbers

The top ranked wireless carrier in the U.S. has:

\$131.6B in revenue

177K employees

1,700 retail locations

112.1M retail connections

106.5M postpaid customers

13M TV and Internet subscribers



ستون فقرات شرکت

شرکت وریزون از زمان شکل‌گیری در سال ۲۰۰۰ برای ایجاد معماری داده‌ها و تحلیل‌هایی که محرک نوآوری و تصمیم‌گیری مبتنی بر دانش باشند با شرکت ترادیتا شراکت کرده است. هدف از اینکار، حفظ ارتباط با مشتریان و همچنین شناسایی فرصت‌های کسب‌وکار جدید و ایجاد اصلاحاتی است که موجب اثربخشی هزینه‌ای عملیات می‌شود.

گریس هونانگ، مدیر اجرایی هوش تجری، تحلیل و عملکرد مالی وریزون می‌گوید: «به کمک هوش کسب‌وکار ما به شرکت کمک می‌کنیم تا فرصت‌های کسب‌وکار جدید را شناسایی کرده یا اصلاحاتی را در مسیر سازمان برای فعالیت به شکلی مقرون به صرفه تر انجام دهند. ما با استفاده از مرتبط‌ترین اطلاعات از تصمیم‌گیرندگان برای ارتقای مزیت رقابتی وریزون پشتیبانی می‌کنیم.»

وریزون با بهره‌گیری از داده‌ها و تحلیل‌ها همچنین قادر به ارائه‌ی یک شبکه مطمئن، حصول اطمینان از رضایت مشتریان و توسعه‌ی محصولات و خدماتی باشد که مصرف‌کنندگان، خواستار خرید آنها هستند. به گفته‌ی هونانگ، «انکوباتور محصولات و خدمات جدید ما کمک می‌کند تا آینده را برای مشتریان به ارمغان بیاوریم. ما از شبکه‌مان برای پیشرفت در سرگرمی‌های تعاملی، رسانه‌های دیجیتال، اینترنت اشیا و خدمات باندهن استفاده می‌کنیم.»

بینش داده‌ها در سه واحد کسب‌وکار شرکت

وریزون متکی بر تحلیل‌های پیشرفته‌ای است که در بخش «معماری داده‌های متحدالشکل ترادیتا» برای پشتیبانی از واحدهای کسب‌وکارش انجام می‌شوند. این تحلیل‌ها شرکت را قادر می‌سازند تا به وعده‌ی خود برای کمک به مشتریان در ایجاد نوآوری در سبک زندگی‌شان عمل کرده و بینش‌های مهم و اساسی را برای پشتیبانی از این سه حوزه فراهم کنند:

۱. شناسایی منابع درآمدی جدید. تیم‌های تحقیق و توسعه از داده‌ها، تحلیل‌ها و شراکت‌های استراتژیک برای آزمون و توسعه با کمک اینترنت اشیا استفاده می‌کنند. اینترنت اشیا، خط مقدم جدید در حوزه داده‌ها می‌باشد که منجر به درآمدهای جدید خواهد شد که آنها نیز به نوبه‌ی خود فرصت‌هایی را برای رشد ایجاد خواهند کرد. خودروهای هوشمند، کشاورزهای هوشمند و اینترنت اشیا هوشمند همگی بخشی از این رشد نوین خواهند بود.

۲. پیش‌بینی رویگردانی در کسب‌وکار محوری سیار. وریزون چندین مورد کاربردی دارد که نشان می‌دهند چگونه تحلیل‌های پیشرفته‌اش امکان پیش‌بینی‌های بسیار دقیق **رویگردانی** – در حد حاشیه سود ۱ تا ۲ درصد- را در فضای سیار فراهم می‌کنند. برای یک شرکت ۱۳۱ میلیارد دلاری، پیش‌بینی رویگردانی با چنین دقتی بسیار مهم است. با شناسایی الگوهای خاص در بکارگیری داده‌های تبلت‌ها، وریزون می‌تواند شناسایی کند کدام مشتریان اغلب اوقات به تبلت‌ها دسترسی دارند و سپس آنهایی که از تبلت استفاده نمی‌کنند را درگیر نماید.

۳. پیش‌بینی طرح‌های تلفن سیار. تحلیل‌های رفتار مشتریان اجازه می‌دهد تا دریافتی‌ها را در شرایط بسیار متغیر بازار بهتر پیش‌بینی کنیم. صنعت بی‌سیم در ایالات متحده از روش پرداخت‌های ماهانه برای تلفن و خدمات به سمت پرداخت مستقل برای تلفن حرکت کرده است. بدین ترتیب فرصت جدیدی برای وریزون فراهم می‌شود تا تجارت را در دست بگیرد. محیط تحلیل در پیش‌بینی بهتر رویگردانی طرح‌های جدید و پیش‌بینی تأثیر تغییرات بر روی طرح‌های قیمت‌گذاری به وریزون کمک می‌کند.

تحلیل‌ها، «داده‌های صادق» را برای وریزون فراهم می‌کنند که واحدهای مختلف کسب‌وکار را مطلع و آگاه می‌کند. به گفته‌ی هونانگ، «مأموریت ما این است که صدای صادق و عقیده‌ی مستقل شخص ثالث در خصوص موفقیت یا فرصت‌هایی برای بهبود کسب‌وکار باشیم. به همین دلیل است که واحد من به عنوان منبع طلایی اطلاعات دیده می‌شود و ما به صداهای صادقانه رویرو می‌شویم و بسیاری از تصمیمات تجاری در طی مراحل مختلف اصلاح می‌شوند.»

هونانگ می‌افزاید که اغلب مواقع، آنچه یک شرکت را مجبور به واکنش می‌کند، رقبایی هستند که تغییرات را در بازار بوجود می‌آورند نه اینکه شرکت تصمیمات نادرستی می‌گیرد. او می‌افزاید «بنابراین ما سعی می‌کنیم سازمان را در صورت امکان از طریق بهترین روش اجرایی هدایت کنیم تا بتوانیم نتایج رکوردشکنی را سال به سال شاهد باشیم. من شک ندارم که هوش کسب‌وکار منجر به چنین موفقیت‌هایی در گذشته شده است.»

تحول آفرین و نوآور

وریزون از تحلیل‌های پیشرفته برای بهینه‌سازی بازاریابی از طریق ارسال مرتبط‌ترین پیشنهادات برای مشتریان بهره می‌برد. در همین حین، شرکت برای حصول اطمینان از اینکه تیزبینی مالی لازم برای حفظ مقام اول در بازار موبایل ایالات متحده برخوردار است به تحلیل‌ها متکی است. وریزون با ادامه‌ی تحول آفرینی در صنعت از طریق نوآوری در محصولات و خدمات به عنوان یک استاندارد در صنعت بی‌سیم جایگاه خود را حفظ کرده است.

هونانگ می‌گوید: «ما برای تولید مرتبط‌ترین خدمات و محصولات به مشتریان‌مان به چشم‌انداز بازاریابی و دقت در فروش نیاز داریم و در عین حال نیازمند دقت مالی برای حصول اطمینان در سودآور بودن محصولات و خدمات عرضه شده به مشتریان هستیم تا بتوانیم در قبال سهامداران خود پاسخگو و مسئول باشیم.»

خلاصه – اجرای هفت P بازاریابی مدرن

وریزون، غول مخابراتی از مدل هفت P برای پیشبرد فعالیت‌های بازاریابی پیشرفته‌ی خود استفاده می‌کند. زمانیکه مدل هفت P به طور هماهنگ و متحد بکار گرفته شود به وریزون کمک می‌کند تا به گونه‌ای که پیش‌بینی کرده است در بازار نفوذ کند.

۱. کارکنان. درک مشتریان و نیازهای آنها برای خلق محصول.
۲. محل. جائیکه مشتریان در آنجا خرید می‌کنند.
۳. محصول. اقلامی که ساخته و فروخته می‌شوند.
۴. فرایند. مشتریان برای خرید محصول چگونه به فروشگاه یا محل مورد نظر دسترسی پیدا می‌کنند.
۵. قیمت‌گذاری. همکاری با حوزه تبلیغات برای جلب توجه مشتریان.

۶. تبلیغات. همکاری با قیمت‌گذاری برای جلب توجه مشتریان.
۷. شواهد فیزیکی. هوش کسب‌وکار که منجر به بینش و ادراک می‌شود.

به گفته‌ی هوئانگ، مدیر اجرایی بخش هوش کسب‌وکار، تحلیل و عملکرد مالی شرکت وریزون، «محیط آستر و هَدوپ به ما اجازه می‌دهد چیزهایی را که گمان می‌بریم علل خرابی یا عدم موفقیت در مدل هفت P باشند را مورد کشف و بررسی قرار دهیم. این ریشه در ارائه‌ی ارزش کسب‌وکار به تصمیم‌گیرندگان ما دارد. با هر گام مدل هفت P، باید بتوانیم به آنها بگوییم که فرصت‌هایی برای بهبود وجود دارد.»

Source: Teradata Case Study "Verizon Answers the Call for Innovation" <https://www.teradata.com/Resources/Case-Studies/Verizon-answers-the-call-for-innovation> (accessed July 2018).

سوالاتی برای مورد کاربردی ۳.۱

۹. شرکت وریزون به چه چالشی روبرو بود؟
۱۰. راه‌حل داده‌محور پیشنهادی برای واحدهای کسب‌وکار وریزون چه بودند؟
۱۱. چه نتایجی به دست آمد؟

سوالاتی برای مرور بخش ۳.۳

۱. داده‌ها چه هستند؟ فرق داده‌ها با اطلاعات و دانش در چیست؟
۲. دسته‌بندی‌های اصلی داده‌ها کدام هستند؟ از چه نوع داده‌هایی می‌توانیم برای هوش کسب‌وکار و تحلیل‌ها استفاده کنیم؟
۳. آیا می‌توانیم از نمایش یکسان داده‌ها در تمامی مدل‌های تحلیل استفاده کنیم؟ چرا؟
۴. منظور از نمایش داده‌های یک به چند چیست؟ چرا و در کجای تحلیل‌ها از آن استفاده می‌شود؟

داده‌ها در شکل اصلی خود (یعنی داده‌های دنیای واقعی) معمولاً برای استفاده در تحلیل‌ها آماده نیستند. آنها اغلب آلوده، بیش از اندازه پیچیده، ناآراسته و نادقیق هستند. تبدیل این داده‌های خام دنیای واقعی به شکل مناسب برای الگوریتم‌های تحلیل نیازمند یک فرایند طاقت‌فرسا و زمان‌بر به نام **پیش‌پردازش داده‌ها**^۱ است (کوتسیانتیس، کانلوپولوس و پینتلاس ۲۰۰۶). بسیاری از کارشناسان تحلیل گواهی می‌دهند که زمان صرف شده برای پیش‌پردازش داده‌ها (که شاید ناخوشایندترین فاز در کل فرایند باشد) بسیار طولانی‌تر از زمان صرف شده بر روی سایر فعالیت‌های فرایند تحلیل است (ایجاد و ارزیابی مدل تحلیل). شکل ۳.۳ گام‌های اصلی فرایند پیش‌پردازش داده‌ها را نشان می‌دهد.

در گام اول پیش‌پردازش داده‌ها، داده‌های مربوطه از منابع شناسایی شده گردآوری شده، سوابق و متغیرهای لازم انتخاب شده (بر اساس درک عمیق از داده‌ها، اطلاعات غیرضروری فیلتر و خارج می‌شوند) و سوابقی که از منابع داده‌ای متعدد بدست آمده‌اند

¹ Data Preprocessing

یکپارچه/ادغام می‌شوند (مجدداً با استفاده از درک عمیقی از داده‌ها، مترادف‌ها و همان‌ها را به نحوی مناسب می‌توان مدیریت کرد).

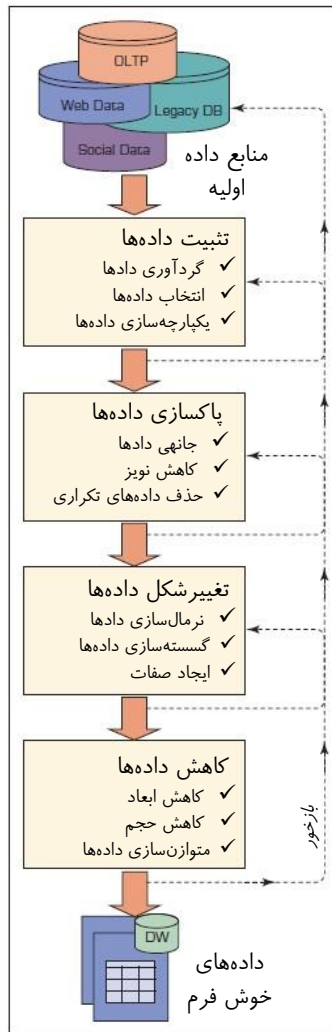
در گام دوم پیش‌پردازش داده‌ها، پاکسازی داده‌ها^۱ انجام می‌شود که آن را شستشوی داده^۲ نیز می‌نامند. داده‌ها در شکل اصلی/دنیای واقعی خود معمولاً ناپاک (آلوده) هستند (هرناندز آند استولفو ۱۹۹۸؛ کیم و همکاران ۲۰۰۳). در این فاز، مقادیر در مجموعه داده‌ها شناسایی شده و به آنها رسیدگی می‌شود. در برخی موارد، مقادیر مفقود به صورت یک ناهنجاری در مجموعه داده نمایان می‌شوند که در آن صورت باید با محتمل‌ترین مقدار پر شده یا نادیده گرفته شوند. در سایر موارد، مقادیر مفقود، بخشی طبیعی از مجموعه داده محسوب می‌شوند (مثلاً، اغلب اوقات افرادی که در لایه‌ی بالایی درآمد قرار دارند فیلد درآمد خانوار را تکمیل نمی‌کنند). در این گام، تحلیلگر باید مقادیر خطادار را در داده‌ها شناسایی کرده و آنها را صاف و هموار کند. علاوه بر این، ناسازگاری‌ها (مقادیر غیرمعمول در یک متغیر) در داده‌ها باید با استفاده از نظر خبرگی و/یا دانش تخصصی مدیریت شود.

در گام سوم پیش‌پردازش داده‌ها، داده‌ها برای پردازش بهتر، تغییرشکل^۳ می‌یابند. به عنوان مثال، در بسیاری موارد، برای پیشگیری از از اینکه یک متغیر با مقادیر عددی بزرگ مثلاً درآمد خانوار) بر متغیرهای دیگری که مقادیر کوچکتری دارند سلطه پیدا کند (مثل تعداد وابستگان یا سال‌های خدمت که ممکن است مهمتر باشند)، داده‌ها باید برای تمامی متغیرها بین یک ماکزیمم و مینیمم مشخص نرمال‌سازی شوند. یک تغییرشکل دیگر، گسسته‌سازی و/یا تجمیع است. در برخی موارد، متغیرهای عددی باید به متغیرهای رسته‌ای تبدیل شوند (مثلاً کم، متوسط؛ زیاد). در سایر موارد، دامنه‌ی مقدار منحصر بفرد یک متغیر با استفاده از سلسله مراتب مفهومی به مجموعه کوچکتری تبدیل می‌شود (مثلاً به جای استفاده از حالات فردی با ۵۰ مقدار مختلف، می‌توان از مناطق مختلف برای یک متغیر که مکان را نشان می‌دهد استفاده کرد) تا مجموعه داده برای پردازش رایانه‌ای آماده‌تر باشد. هنوز در برخی موارد، می‌توان متغیرهای جدیدی را بر اساس متغیرهای موجود ایجاد کرد تا اطلاعات یافت شده در مجموعه‌ی متغیرهای داده‌ها را بزرگنمایی کرد. به عنوان مثال، در یک مجموعه داده‌ی مربوط به پیوند اعضا ممکن است از یک متغیر که نشان دهنده‌ی تطبیق گروه خونی است به جای مقادیر چندجمله‌ای جداگانه برای گروه خونی اهدا کننده و دریافت کننده عضو استفاده کنیم. اینگونه ساده‌سازی‌ها می‌توانند محتوای اطلاعاتی را افزایش داده و در عین حال، پیچیدگی ارتباطات در داده‌ها را کاهش دهند.

¹ Data cleaning

² Data scrubbing

³ Data transformation



شکل ۳.۳ گام‌های پیش‌پردازش داده‌ها

گام نهایی پیش‌پردازش داده‌ها، داده کاهی^۱ یا تقلیل داده‌ها نام دارد. اگرچه دانشمندان علم داده‌ها (یعنی متخصصان تحلیل) تمایل دارند تا از مجموعه داده‌های بزرگ استفاده کنند ولی حجم بیش از اندازه داده‌ها نیز می‌تواند مشکل‌ساز باشد. در ساده‌ترین مفهوم، داده‌های مورد استفاده از پروژه‌های تحلیلی پیشگویانه را می‌توان به صورت یک فایل مسطح که دارای دو بُعد است در نظر گرفت: متغیرها (تعداد ستون‌ها) و رکوردها (تعداد ردیف‌ها). در برخی موارد (مثلاً پروژه‌های گنوم^۲ و تصویرپردازی با داده‌های ریزآرایه‌ی پیچیده)، تعداد متغیرها می‌تواند زیاد باشد و تحلیلگر باید تعداد آنها را به حدی کنترل‌پذیر کاهش دهد. چونکه داده‌ها با ابعاد (دیمانسیون) متفاوتی همراه هستند که پدیده‌ها را از منظرهای متفاوتی توضیح می‌دهند، این فرایند را در تحلیل پیشگویانه

¹ Data reduction

² Genome projects

و در داده کاوی، تقلیل بُعدی^۱ یا انتخاب متغیر^۲ می نامند. اگرچه یک روش بهینه‌ی منحصر بفرد برای اینکار وجود ندارد، می توانیم از یافته‌های ادبیات موضوع که قبلاً منتشر شده است استفاده کرده؛ از نظرات افراد خبره بهره برده، آزمون‌های آماری مناسب را اجرا کرده (مثلاً تحلیل مؤلفه‌ی اصلی^۳ یا تحلیل مؤلفه‌ی مستقل^۴) و یا ترجیحاً از ترکیبی از این تکنیک‌ها برای کاهش ابعاد داده‌ها به یک تعداد کنترل پذیرتر و دستیابی به مرتبط‌ترین مجموعه داده‌ها استفاده کنیم.

با توجه به بُعد دیگر (یعنی تعداد پرونده‌ها)، برخی مجموعه داده‌ها ممکن است شامل میلیون‌ها یا میلیاردها رکورد باشند. اگرچه توان رایانشی به شکلی نمایی در حال افزایش است ولی پردازش این حجم انبوه رکوردها عملی یا امکان پذیر نیست. در برخی موارد، ممکن است به یک نمونه از داده‌ها برای تحلیل نیاز باشد. فرضیه‌ی زیربنایی در نمونه برداری این است که زیرمجموعه‌ی داده‌ها حاوی تمامی الگوهای مربوطه از کل زیرمجموعه داده‌ها است. در یک زیرمجموعه متجانس از داده‌ها، چنین فرضی ممکن است صادق باشد ولی داده‌های دنیای واقعی به ندرت متجانس هستند. تحلیلگر باید در انتخاب زیرمجموعه‌ای از داده‌ها که بیانگر ذات اصلی مجموعه کامل داده‌ها بوده و خاص یک زیرگروه یا زیردسته از داده‌ها نباشد دقت کند. داده‌ها معمولاً بر حسب یک متغیر مرتب‌سازی شده و در نظر گرفتن یک بخش داده‌ها از قسمت بالایی یا پایینی می‌تواند منجر به انحراف مجموعه داده‌ها بر روی مقادیر خاص یک متغیر اندیس دار شود. از این رو، همیشه سعی کنید تا رکوردها را به صورت تصادفی از یک مجموعه داده انتخاب کنید. در خصوص داده‌های دارای چولگی نیز ممکن است نمونه برداری تصادفی ساده کافی نباشد و لازم باشد که از نمونه برداری طبقه بندی شده^۵ استفاده کنیم (یک نمایش نسبی یا متناسب از زیرگروه‌های متفاوت در داده‌ها که در یک زیرمجموعه داده‌ی نمونه نمایش داده می‌شود). زمانیکه بحث داده‌های دارای چولگی پیش می‌آید بهتر است که از طریق بیش نمونه برداری از دسته‌هایی که کمتر ابراز شده‌اند یا کم نمونه برداری کلاس‌هایی که بیشتر ابراز شده‌اند، چولگی بالای داده‌ها را به حالت توازن درآورد. پژوهش‌ها نشان داده است که مجموعه داده‌های متوازن نسبت به مجموعه‌های نامتوازن، مدل‌های پیش‌بینی بهتری را تولید می‌کنند (تاماسیری و همکاران^۶ ۲۰۱۴).

ماهیت پیش‌پردازش داده‌ها به طور خلاصه در جدول ۳.۱ ارائه شده است که فازهای اصلی (به همراه توضیحات مسئله) را در کنار فهرستی که بیانگر وظایف و الگوریتم‌های مربوطه است نشان می‌دهد.

¹ Dimensional reduction

² Variable selection

³ Principal component analysis

⁴ Independent component analysis

⁵ Stratified sampling

⁶ Thammasiri et al

جدول ۳.۱ چکیده‌ای از روش‌های بالقوه و وظایف پیش‌پردازش داده‌ها

وظیفه‌ی اصلی	زیروظایف	روش‌های تداول
تنبیت داده‌ها	دسترسی به داده‌ها و گردآوری آنها انتخاب و فیلترسازی داده‌ها یکپارچه‌سازی و متحدسازی داده‌ها	پرس و جوهای اس‌کیوال، عوامل نرم‌افزاری، خدمات وب. تخصص دامنه، پرس و جوهای اس‌کیوال، آزمون‌های آماری پرس و جوهای اس‌کیوال، تخصص دامنه، نگاشت مبتنی بر آنتولوژی داده‌ها
پاکسازی داده‌ها	مدیریت داده‌های مفقود در داده‌ها	جانهی مقادیر مفقود با مناسب‌ترین مقادیر (میان، میانگین، بیشینه/کمینه، مُد و غیره)؛ جایگزینی مقادیر مفقود با یک مقدار ثابت مثل "ML"؛ حذف سوابق دارای مقادیر مفقود؛ عدم اقدام خاص.
	شناسایی و کاهش نویز در داده‌ها	شناسایی داده‌های پرت با استفاده از تکنیک‌های آماری (مثل میانگین و انحراف استاندارد) یا تحیل خوشه، حذف داده پرت یا هموارسازی آنها با استفاده از رگرسیون، دسته‌بندی یا میانگین‌های ساده.
	یافتن و حذف داده‌های غلط (نادرست)	شناسایی مقادیر غلط یا نادرست در داده‌ها (به غیر از داده‌های پرت)، مثل مقادیر نامتعارف، برچسب کلاس‌های ناسازگار، توزیع‌های نامتعارف؛ استفاده از تخصص دامنه برای تصحیح مقادیر یا حذف رکوردهای دارای مقادیر غلط.
تغییر شکل داده‌ها	نرمال‌سازی داده‌ها	کاهش دامنه‌ی مقادیر در هر متغیر دارای مقدار عددی به یک دامنه‌ی استاندارد (مثلاً ۰ تا ۱ یا -۱ تا +۱) با استفاده از انواع تکنیک‌های نرمال‌سازی یا مقیاس‌دهی.
	گسسته‌سازی یا تجمیع داده‌ها	در صورت نیاز، تبدیل متغیرهای عددی به نمایش‌های گسسته با استفاده از تکنیک‌های طبقه‌بندی مبتنی بر دامنه یا تواتر؛ در خصوص متغیرهای رسته‌ای، کاهش تعداد مقادیر با بکارگیری سلسله‌مراتب مفهومی مناسب.
	ایجاد صفات جدید	استخراج متغیرهای جدید و حاوی اطلاعات مفیدتر از داده‌های موجود با استفاده از طیف گسترده‌ای از توابع ریاضی (مثل جمع یا ضرب یا ترکیب هیبریدی تبدیل‌های لگاریتمی).
تقلیل داده‌ها	کاهش تعداد صفات	استفاده از تحلیل مؤلفه‌ی اصلی، تحلیل مؤلفه‌ی مستقل، آزمون مجذور کا، تحلیل همبستگی و استنتاج درخت تصمیم.

ارزش قابل ارائه‌ی پیش‌پردازش داده‌ها را نمی‌توان دست کم گرفت. این یکی از فعالیت‌های زمان‌بر است که سرمایه‌گذاری زمان و فعالیت‌ها در آن، نرخ بازگشت آن سرمایه‌گذاری هیچ محدودیتی ندارد. این بدان معناست که هرچه منابع بیشتری را در آن سرمایه‌گذاری کنید، در پایان سود بیشتری عایدتان خواهد شد. مورد کاربردی ۳.۲ یک بررسی جالب را با استفاده از داده‌های اولیه‌ی دانشگاهی در یک سازمان آموزشی برای تدوین مدل‌های پیشگویانه جهت درک بهتر علل ریزش دانشجویان و افزایش نگهداشت آنها در یک مؤسسه‌ی تحصیلات تکمیلی بزرگ به تصویر کشیده است. همانطور که در مورد کاربردی به روشنی بیان شده است، تمامی وظایف پیش‌پردازش داده‌ها که در جدول ۳.۱ شرح داده شده‌اند برای اجرای موفق پروژه‌ی تحلیلی زیربنایی دارای اهمیت حیاتی بوده به ویژه وظایف مرتبط با متوازن‌سازی مجموعه داده‌ها.

مورد کاربردی ۳.۲

افزایش نگهداشت دانشجویان با استفاده از تحلیل‌های داده‌محور

نرخ ریزش دانشجویان به یکی از چالش‌برانگیزترین مسائل برای تصمیم‌گیرندگان در موسسات دانشگاهی تبدیل شده است. علیرغم تمامی برنامه‌ها و خدماتی که برای کمک به حفظ دانشجویان انجام شده است، طبق آمار مرکز آمار وزارت آموزش ایالات متحده (nces.ed.gov) فقط نیمی از آنهايي که به دانشگاه راه پیدا می‌کنند موفق به کسب مدرک کارشناسی می‌شوند. مدیریت نام‌نویسی و نگهداشت دانشجویان به اولویت اول مدیران کالج‌ها و دانشگاه‌ها در ایالات متحده و سایر کشورهای سرتاسر جهان تبدیل شده است. نرخ ریزش بالای دانشجویان معمولاً منجر به زیان مالی، کاهش نرخ فارغ‌التحصیلی و کاهش اعتبار دانشگاه از دید تمامی سهامداران می‌شود. قانون‌گذاران و سیاست‌گذارانی که بر آموزش عالی و تخصیص بودجه نظارت دارند، والدینی که هزینه‌ی آموزش فرزندانشان را برای آماده‌سازی آنها برای آینده‌ی بهتر پرداخت می‌کنند و دانشجویانی که از بین دانشگاه‌ها انتخاب می‌کنند به دنبال شواهدی از کیفیت و شهرت هستند تا بتوانند فرایندهای تصمیم‌گیری را بهتر انجام دهند.

راه‌حل پیشنهادی

به منظور بهبود نگهداشت دانشجویان، باید سعی کنیم تا علل مهم در پشت این ریزش را درک کنیم. برای موفقیت، باید بتوانیم دانشجویانی که ریسک اخراج از دانشگاه دارند را به دقت شناسایی کنیم. تا به اکنون، اکثریت پژوهش‌هایی که حول ریزش دانشجویان انجام شده به درک این پدیده‌ی اجتماعی پیچیده ولی حیاتی اختصاص داشته است. اگرچه این مطالعات کیفی، رفتاری و مبتنی بر نظرسنجی از طریق توسعه و آزمون طیف گسترده‌ای از نظریه‌ها، بینش ارزشمندی را در اختیارمان قرار داده‌اند اما ابزارهای چندین مفیدی را برای پیش‌بینی دقیق (و بهبود) ریزش دانشجویان ارائه نمی‌کنند. پروژه‌ای که به طور خلاصه در این مطالعه‌ی موردی معرفی شده یک روش پژوهش کمی را پیشنهاد کرده که در آن می‌توان از داده‌های سازمانی تاریخی موجود در پایگاه داده‌های دانشجویان برای توسعه مدل‌هایی که قادر به پیش‌بینی و همچنین توضیح ماهیت مسئله‌ی ریزش برای هر موسسه یا دانشگاه خاص هستند استفاده کرد. روش تحلیل پیشنهادی در شکل ۳.۴ نشان داده شده است.

اگرچه این مفهوم در بحث آموزش عالی نسبتاً جدید است ولی برای بیش از تقریباً یک دهه، مسائل مشابهی در رشته‌ی مدیریت بازاریابی با استفاده از تکنیک‌های تحلیل پیشگویانه داده‌ها و تحت عنوان «تحلیل رویگردانی» مورد مطالعه و بررسی قرار گرفته که هدف از آنها شناسایی یک نمونه از مشتریان فعلی برای پاسخ به این سوال است که «چه کسی از میان مشتریان جاری بیشتر احتمال دارد که محصولات و خدمات را دیگر خریداری نکند؟» تا نوعی فرایند مداخله یا تسهیلگری را بتوان برای حفظ و نگهداشت آنها انجام داد. همانطور که می‌دانیم و پژوهش‌های مربوطه نیز بارها و بارها نشان داده‌اند، حفظ مشتریان فعلی بسیار حیاتی است زیرا جذب یک مشتری جدید از لحاظ زمان، پول و فعالیت، هزینه‌ای به مراتب بیشتر در مقایسه با حفظ مشتری فعلی طلب می‌کند.

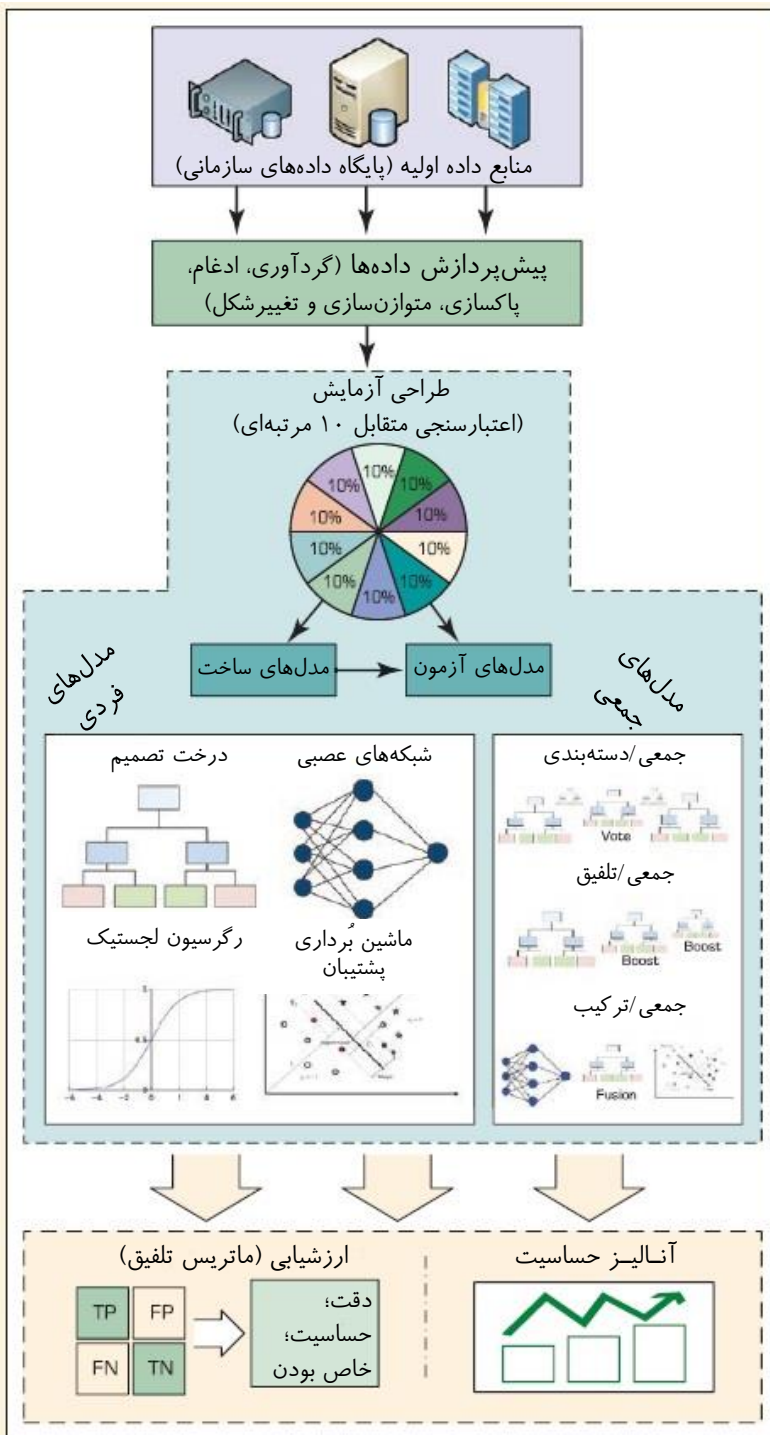
داده‌ها اصل موضوع هستند

داده‌های این پروژه‌ی پژوهشی از یک مؤسسه (یک دانشگاه دولتی جامع واقع در منطقه‌ی میدوست ایالات متحده) به دست آمده است که مشتمل بر نام‌نویسی میانگین ۲۳,۰۰۰ دانشجو است که حدوداً ۸۰ درصد آنها ساکن همان ایالت بوده و حدوداً ۱۹ درصد دانشجویان در دسته‌بندی اقلیت‌ها قرار دارند. هیچ اختلاف معناداری بین جنسیت‌ها در تعداد نام‌نویسی وجود ندارد. میانگین نرخ نگهداشت دانشجویان سال‌های اول حدود ۸۰ درصد بوده و میانگین نرخ فارغ‌التحصیلی در شش سال حدود ۶۰ درصد بوده است. در این مطالعه از داده‌های پنج ساله دانشگاه که شامل اطلاعات بیش از ۱۶,۰۰۰ دانشجوی نام‌نویسی کرده است که از پایگاه‌های داده مختلف و گوناگون دانشگاه گردآوری و جمع‌آوری شده‌اند. داده‌ها حاوی متغیرهای مربوط به مشخصه‌های مالی، جمعیت‌شناختی و آکادمیک دانشجویان است. پس از ادغام و تبدیل داده‌های چندبعدی دانشجویان در قالب یک فایل تخت (یک فایل که ستون‌های آنها بیانگر متغیرها و ردیف‌های آن بیانگر رکوردهای دانشجویان است)، فایل به دست آمده برای شناسایی و رفع ناهنجاری‌ها و مقادیر غیرقابل استفاده مورد ارزیابی و پیش‌پردازش قرار گرفت. به عنوان مثال، در این مطالعه، تمامی رکوردهای دانشجویان بین‌المللی از مجموعه داده به دلیل اینکه فاقد اطلاعات مربوط به برخی از مهمترین عوامل پیش‌بینی بوده است از مجموعه داده‌ها حذف گردیدند (مثلاً GPA دبیرستان، امتیازهای SAT). در فاز تغییرشکل داده‌ها، برخی از متغیرها جمع‌بندی شدند (مثلاً متغیرهای "Major" و "Concentration" برای تفسیر بهتر مدلسازی پیشگویانه در قالب متغیرهای دودویی MajorDeclared و ConcentrationSpecified جمع‌بندی شدند). علاوه بر این، برخی از متغیرها برای اشتقاق متغیرهای جدید استفاده شدند (مثلاً نسبت Earned/Registered و YearsAfterHighSchool).

$$\text{Earned/Registered} = \frac{\text{EarnedHours}}{\text{RegisteredHours}}$$

$$\text{YearsAfterHighSchool} = \text{FreshmenEnrollmentYear} - \text{HighSchoolGraduationYear}$$

نسبت *Earned/Registered* برای نمایش بهتر پایداری و قاطعیت دانشجویان در اولین ترم تحصیلشان ایجاد شد. انتظار می‌رود که این متغیر برای اینکه تأثیر مثبتی بر نگهداشت/استقامت دانشجویان داشته باشد از مقادیر بالایی برخوردار باشد. متغیر *YearsAfterHighSchool* برای سنجش تأثیر فاصله‌ی زمانی بین اتمام دبیرستان و نام‌نویسی اولیه در کالج ایجاد شد. در اینجا نیز انتظار داریم که این متغیر بر پیش‌بینی ریزش دانشجویان تأثیر داشته باشد. این جمع‌بندی‌ها و متغیرهای مشتق شده بر اساس چندین آزمایش انجام شده برای چندین فرضیه‌ی منطقی تعیین شده‌اند. آنهایی که معنادارتر بودند و دقت پیش‌بینی بهتری داشتند در مجموعه نهایی متغیرها نگهداشته شدند. متغیر وابسته (یعنی «Second Fall Registered») که بیانگر ماهیت حقیقی جامعه‌ی نمونه بود (یعنی دانشجویان سال اول) حاوی تعداد رکوردهای آری بیشتری (تقریباً ۸۰٪) نسبت به رکوردهای خیر بود (تقریباً ۲۰٪). در شکل ۳.۵.



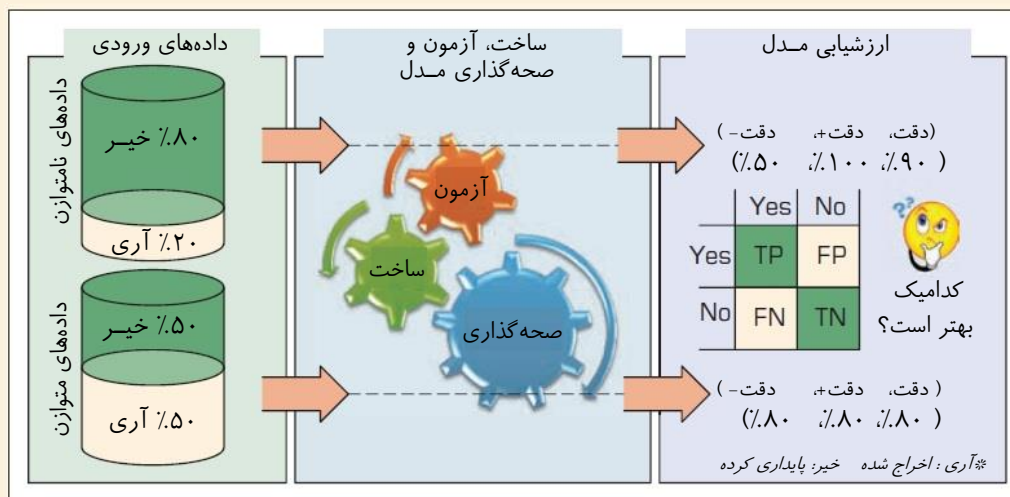
شکل ۳.۴ رویکرد تحلیلی برای پیش‌بینی ریزش دانشجویان

پژوهش نشان می‌دهد که داشتن اینگونه داده‌های نامتوازن، تأثیر منفی بر عملکرد مدل خواهد داشت. از این رو، این مطالعه با گزینه‌های استفاده و مقایسه‌ی نتایج مدل‌های یکسان که با داده‌های نامتوازن اصلی (رکوردهای آری دارای انحراف) و داده‌های متوازن ساخته شده‌اند آزمایش گردید.

مدل‌سازی و ارزشیابی

در این مطالعه از چهار روش دسته‌بندی معروف یعنی شبکه‌های عصبی مصنوعی، درخت تصمیم، ماشین‌های بردار پشتیبان و رگرسیون لجستیک به همراه سه تکنیک تجمیعی مدل‌سازی یعنی دسته‌بندی^۱، تقویتی^۲ و ترکیب اطلاعات^۳ استفاده است. نتایج بدست آمده از تمامی مدل‌ها با استفاده از روش‌های متداول ارزشیابی مدل دسته‌بندی مثل دقت، حساسیت و خاص بودن پیشگویی بر روی نمونه‌های موجود با یکدیگر مقایسه شدند.

در الگوریتم‌های یادگیری ماشین (برخی از آنها در فصل ۴ توضیح داده شده‌اند)، تحلیل حساسیت، روشی برای شناسایی ارتباطات علت و معلولی بین ورودی‌ها و خروجی‌های یک مدل پیش‌بینی مفروض است. نظریه‌ی بنیادی در پشت تحلیل حساسیت این است که اهمیت متغیرهای پیش‌بینی را بر اساس تغییری که در عملکرد مدل‌سازی به واسطه‌ی در نظر نگرفتن یک متغیر پیش‌بینی در مدل رخ می‌دهد می‌سنجد. این روش مدل‌سازی و آزمایش را ارزشیابی یک طرفه^۴ نیز می‌نامند. بنابراین، سنجش حساسیت یک متغیر پیش‌بینی خاص عبارتست از نسبت خطای مدل آموزشی بدون متغیر پیش‌بینی تقسیم بر خطای همان مدل با در نظر گرفتن متغیر پیش‌بینی. هرچه شبکه نسبت به یک متغیر خاص حساس‌تر باشد، کاهش میزان عملکرد آن شبکه در نبود آن متغیر، بیشتر شده و بدین ترتیب، نسبت اهمیت نیز بالا می‌رود. علاوه بر قدرت پیش‌بینی مدل‌ها، در این مطالعه از تحلیل حساسیت برای تعیین اهمیت نسبی متغیرهای ورودی نیز استفاده شده است.



شکل ۳.۵ نمایش گرافیکی از مسئله‌ی توازن دسته

¹ Bagging

² Busting

³ Information Fusion

⁴ Leave-one-out Assessment

نتایج

در اولین مجموعه از آزمایشات، از مجموعه داده‌های نامتوازن اصلی استفاده شد. بر اساس نتایج اعتبارسنجی متقابل ۱۰ برابری، ماشین‌های بردار پشتیبان، بالاترین دقت را با نرخ پیش‌بینی کلی ۸۷.۲۳ درصد نتیجه دادند که پس از آن روش درخت تصمیم با نرخ پیش‌بینی کل ۸۷.۱۶ درصد، شبکه‌های عصبی مصنوعی و نهایتاً رگرسیون لجستیک با نرخ پیش‌بینی‌های کلی ۸۶.۴۵ درصد و ۸۶.۱۲ درصد قرار داشتند (جدول ۳.۲). یک بررسی دقیق این نتایج نشان می‌دهد که دقت پیش‌بینی دسته «آری» به طور معناداری بالاتری از دقت پیش‌بینی دسته «خیر» است. در واقع، هر چهار نوع مدل توانسته بودند احتمال بازگشت دانشجویان برای سال دوم تحصیلی را با دقت بالای ۹۰ درصد پیش‌بینی کنند ولی در پیش‌بینی احتمال اخراج دانشجویان پس از پایان سال اول تحصیل با دقت پیش‌بینی کمتر از ۵۰ درصد عملکرد ضعیفی داشته‌اند. چونکه پیش‌بینی دسته «خیر» هدف اصلی در این مطالعه است، دقت پیش‌بینی کمتر از ۵۰ درصد برای این دسته قابل قبول نیست. چنین اختلافی در دقت پیش‌بینی بین این دو دسته می‌تواند (و باید) مربوط به ماهیت نامتوازن مجموعه داده‌های آموزشی باشد (یعنی حدود ۸۰٪ «آری» و حدود ۲۰٪ «خیر»).

در دور بعدی آزمایشات از مجموعه داده‌های متوازن استفاده شده که در آن دو دسته با تعداد نمونه‌های تقریباً یکسان آورده شده است. برای اجرای این روش، تمامی نمونه‌ها از دسته‌ی اقلیت (در اینجا یعنی دسته‌ی «خیر») و به صورت تصادفی با تعداد نمونه‌های یکسان از دسته‌ی اکثریت (در اینجا یعنی دسته‌ی «آری») انتخاب شدند و این فرایند برای کاهش انحراف یالقه‌ی نمونه‌برداری تصادفی ۱۰ بار تکرار شد. هر یک از این فرایندهای نمونه‌برداری منجر به یک مجموعه داده مشتمل بر بیش از ۷,۰۰۰ رکورد گردید که از هر دو دسته «آری» و «خیر» به طور مساوی در آن حضور داشتند. مجدداً، با استفاده از یک متدولوژی اعتبارسنجی متقابل ۱۰ برابری، مدل‌های پیش‌بینی برای هر چهار نوع مدل ایجاد و آزموده شدند. نتایج این آزمایشات در جدول ۳.۳ نشان داده شده است. بر اساس نتایج نمونه‌ها، ماشین‌های بردار پشتیبانی یکبار دیگر بهترین دقت پیش‌بینی کل را با عدد ۸۱.۱۸ درصد نتیجه داد و پس از آن درخت تصمیم، شبکه‌های عصبی مصنوعی و رگرسیون لجستیک با دقت پیش‌بینی کلی ۸۰.۶۵ درصد، ۷۹.۸۵ درصد و ۷۴.۲۶ درصد بوده‌اند. همانطور که در ارقام دقت پیش از دسته‌بندی مشاهده می‌کنید، دقت پیش‌بینی مدل‌ها برای دسته «خیر» با داده‌های متوازن بسیار بهتر از داده‌های نامتوازن بوده است. در مجموع، سه تکنیک یادگیری ماشین از عملکرد به مراتب بهتری نسبت به همتای آماری خود یعنی رگرسیون لجستیک برخوردار بوده‌اند.

سپس، دور دیگری از آزمایشات به منظور ارزشیابی توان پیش‌بینی سه مدل جمعی انجام شد. بر اساس متدولوژی اعتبارسنجی متقابل ۱۰ برابری، مدل جمعی ترکیب اطلاعات با نرخ پیش‌بینی کلی ۸۲.۱۰ درصد بهترین نتایج را ارائه کرده و پس از آن مدل جمعی دسته‌بندی و تقویتی با نرخ پیش‌بینی کلی ۸۱.۸۰ درصد و ۸۰.۲۱ درصد قرار داشتند (جدول ۳.۴). اگرچه نتایج پیش‌بینی کمی بهتر از مدل‌های فردی هستند ولی در مقایسه با یک مدل پیش‌بینی تکی، مدل‌های جمعی سیستم‌های پیش‌بینی مقاوم‌تری را تولید می‌کنند.

جدول ۳.۲ نتایج پیش‌بینی برای مجموعه داده‌های اصلی/نامتوازن

	شبکه عصبی مصنوعی		درخت تصمیم		ماشین بردار پشتیبان		رگرسیون لجستیک	
	خیر	آری	خیر	آری	خیر	آری	خیر	آری
خیر	۱,۴۹۴	۳۸۴	۱,۵۱۸	۳۰۴	۱,۴۷۸	۲۵۵	۱,۴۳۸	۳۷۶
آری	۱,۵۹۶	۱۱,۱۴۲	۱,۵۷۲	۱۱,۲۲۲	۱,۶۱۲	۱۱,۲۷۱	۱,۶۵۲	۱۱,۱۵۰
جمع	۳,۰۹۰	۱۱,۵۲۶	۳,۰۹۰	۱۱,۵۲۶	۳,۰۹۰	۱۱,۵۲۶	۳,۰۹۰	۱۱,۵۲۶

۹۶.۷۴٪	۴۶.۵۴٪	۹۷.۷۹٪	۴۷.۸۳٪	۹۷.۳۶٪	۴۹.۱۳٪	۹۶.۶۷٪	۴۸.۳۵٪	دقت پیش از دسته‌بندی
۸۶.۱۳٪		۸۷.۲۳٪		۸۷.۱۶٪		۸۶.۴۵٪		دقت کل

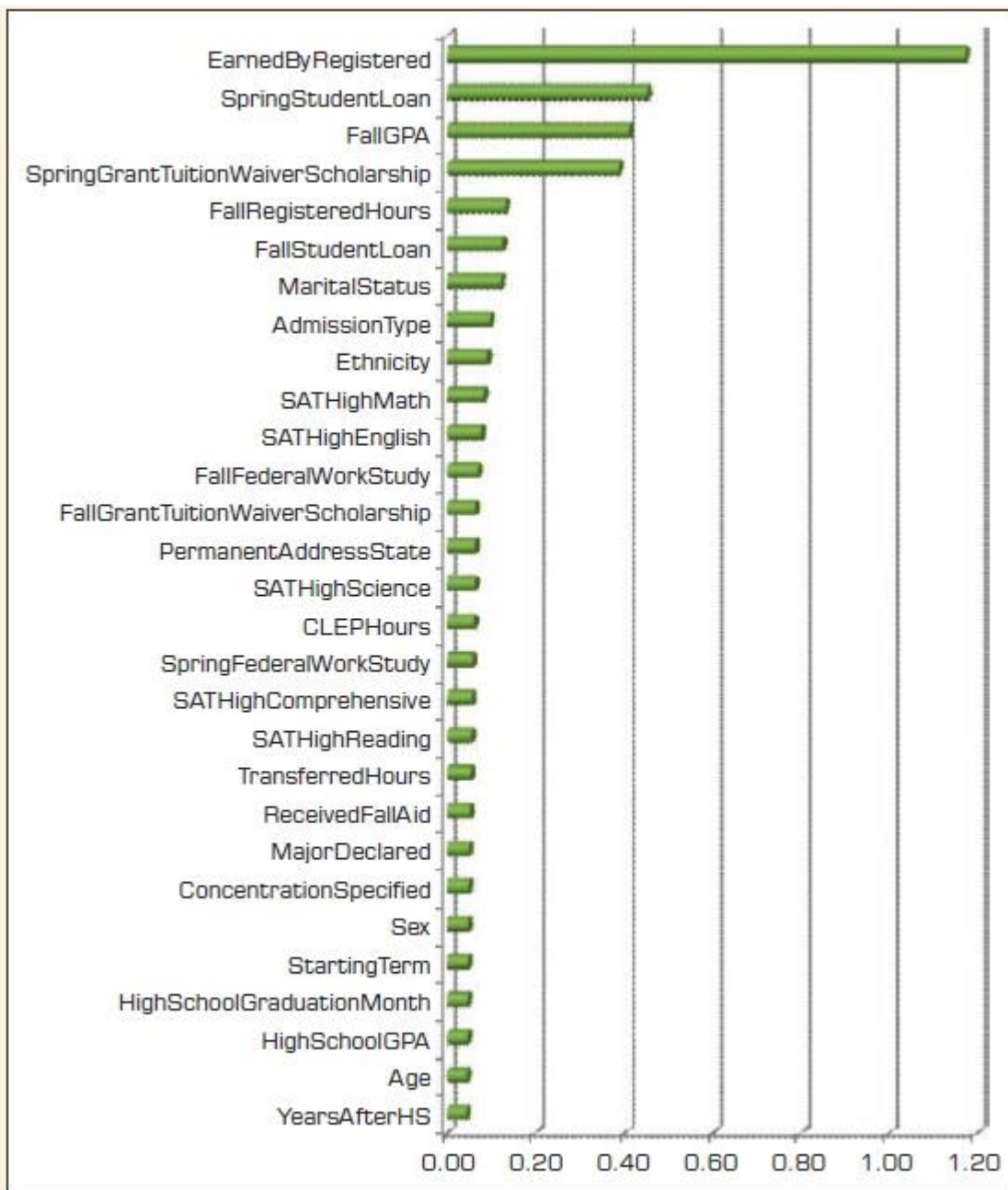
جدول ۳.۳ نتایج پیش‌بینی برای مجموعه داده‌های متوازن

رگرسیون لجستیک		ماشین بردار پشتیبان		درخت تصمیم		شبکه عصبی مصنوعی		
آری	خیر	آری	خیر	آری	خیر	آری	خیر	
۶۲۶	۲,۱۲۵	۳۸۶	۲,۳۱۳	۴۱۷	۲,۳۱۱	۴۶۴	۲,۳۰۹	خیر
۲,۴۶۴	۹۶۵	۲,۷۰۴	۷۷۷	۲,۶۷۳	۷۷۹	۲,۶۲۶	۷۸۱	آری
۳,۰۹۰	۳,۰۹۰	۳,۰۹۰	۳,۰۹۰	۳,۰۹۰	۳,۰۹۰	۳,۰۹۰	۳,۰۹۰	جمع
۷۹.۷۴٪	۶۸.۷۷٪	۸۷.۵۱٪	۷۴.۸۵٪	۸۶.۵۰٪	۷۴.۷۹٪	۸۴.۹۸٪	۷۴.۷۲٪	دقت پیش از دسته‌بندی
۷۴.۲۶٪		۸۱.۱۸٪		۸۰.۶۵٪		۷۹.۸۵٪		دقت کل

علاوه بر ارزشیابی دقت پیش‌بینی هر نوع مدل، یک تحلیل حساسیت نیز با استفاده از مدل‌های پیش‌بینی توسعه یافته برای شناسایی اهمیت نسبی متغیرهای مستقل انجام شد. پس از مشخص شدن نتایج نهایی تحلیل حساسیت، هر یک از چهار نوع مدل، سنج‌های حساسیت خود را تولید کردند و تمامی متغیرهای مستقل به ترتیب اولویت‌بندی شدند. همانگونه که انتظار می‌رفت، رتبه‌بندی‌های متغیرهای مستقل بر حسب حساسیت توسط هر نوع مدل کمی متفاوت بود. پس از گردآوری تمامی چهار مجموعه اعداد حساسیت، این اعداد نرمال‌سازی و تجمیع شده و در یک نمودار میله‌ای افقی ترسیم شدند (شکل ۳.۶).

نتیجه‌گیری

مطالعه نشان داده است که با داشتن داده‌های کافی و متغیرهای مناسب، روش‌های داده‌کاوی قادر به پیش‌بینی ریزش دانشجویان سال اول با دقت تقریباً ۸۰ درصد هستند. نتایج همچنین نشان می‌دهد که صرفنظر از مدل پیش‌بینی بکار رفته، مجموعه داده‌های متوازن (در مقایسه با مجموعه داده‌های اصلی/نامتوازن) مدل‌های پیش‌بینی بهتری را برای شناسایی دانشجویانی که احتمال اخراج آنها از دانشگاه پیش از پایان سال اول وجود دارد تولید می‌کنند. از میان چهار مدل پیش‌بینی فردی بکار رفته در این مطالعه، ماشین‌های بردار پشتیبان از بهترین عملکرد برخوردار بودند و پس از آن، درخت تصمیم، شبکه‌های عصبی و رگرسیون لجستیک قرار دارد. از نقطه نظر قابلیت استفاده، علیرغم این واقعیت که ماشین‌های بردار پشتیبان نتایج پیش‌بینی بهتری را ارائه داده‌اند، ممکن است از درخت تصمیم استفاده شود چون در مقایسه با ماشین‌های بردار پشتیبان و شبکه‌های عصبی، ساختار مدلی شفاف‌تری را دارند. درخت تصمیم، فرایند استنتاج پیش‌بینی‌های مختلف را به صراحت نشان می‌دهد و برای یک نتیجه‌ی خاص، یک توجیه می‌آورد در حالیکه ماشین‌های بردار پشتیبان و شبکه‌های عصبی مصنوعی، مدل‌های ریاضی هستند که چنین دید شفافی از اینکه «کارشان را چگونه انجام می‌دهند» ارائه نمی‌کنند.



شکل ۳.۶ نتایج اهمیت متغیر مبتنی بر تحلیل حساسیت

سوالاتی برای مورد کاربردی ۳.۲

۱. منظور از ریزش دانشجو چیست و چرا مسئله مهمی در آموزش عالی محسوب می‌شود؟
۲. روش‌های قدیمی برای رسیدگی به مسئله‌ی ریزش پرسنل چه بودند؟
۳. چالش‌های موجود مرتبط با داده‌ها در این مطالعه موردی را نام برده و بحث کنید.

۴. راه‌حل پیشنهادی چه بوده است؟ نتایج آن چه بود؟

Sources: D. Thammasiri, D. Delen, P. Meesad, & N. Kasap, "A Critical Assessment of Imbalanced Class Distribution Problem: The Case of Predicting Freshmen Student Attrition," *Expert Systems with Applications*, 41(2), 2014, pp. 321–330; D. Delen, "A Comparative Analysis of Machine Learning Techniques for Student Retention Management," *Decision Support Systems*, 49(4), 2010, pp. 498–506, and "Predicting Student Attrition with Data Mining Methods," *Journal of College Student Retention* 13(1), 2011, pp. 17–35.

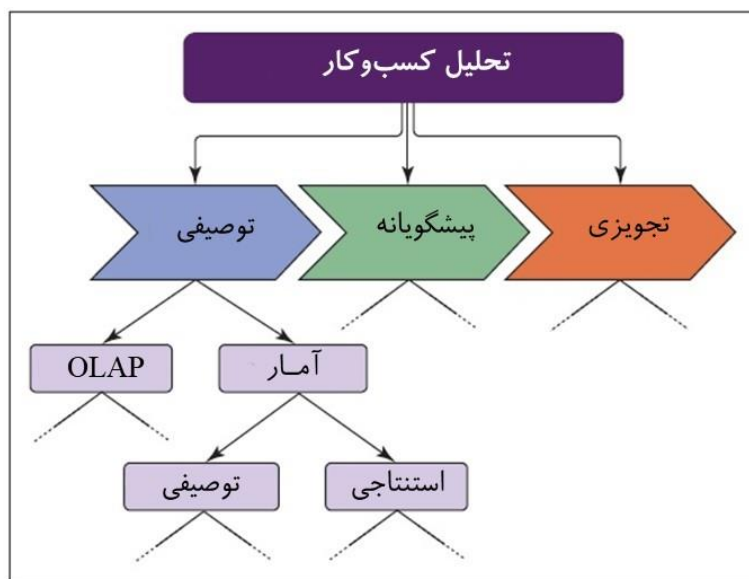
سوالاتی برای مرور بخش ۳.۴

۹. چرا داده‌های اولیه/اصلی به آسانی برای تحلیل‌ها قابل استفاده نیستند؟
۱۰. گام‌های اصلی پیش‌پردازش داده‌ها چه هستند؟
۱۱. پاکسازی/پالایش داده‌ها به چه معناست؟ چه فعالیت‌هایی در این فاز انجام می‌شود؟
۱۲. چرا تغییرشکل داده‌ها لازم است؟ وظایف متداول مورد استفاده برای تغییرشکل داده‌ها کدام هستند؟
۱۳. داده‌گاهی یا تقلیل داده‌ها را می‌توان بر روی سطرها (نمونه‌برداری) و/یا ستون‌ها (انتخاب متغیر) انجام داد. کدامیک چالش بیشتری دارد؟

به دلیل محبوبیت روزافزون تحلیل‌های کسب‌وکار، روش‌های آماری قدیمی و تکنیک‌های زیربنایی آنها نیز با توجه به تأثیری که بر ابزارهای پشتیبان تصمیم‌گیری مدیریتی مبتنی بر شواهد و مدارک دارند مجدداً جذابیت خود را باز می‌یابند. نه تنها بر جذابیت و تحسین این روش‌ها افزوده می‌شود بلکه علاوه بر متخصصان تحلیل و آماردانان، کاربران کسب‌وکار را نیز به خود جذب کرده‌اند. آمار (روش‌های آماری و تکنیک‌های زیربنایی آنها) معمولاً به عنوان بخشی از تحلیل توصیفی شناخته می‌شوند (شکل ۳.۷). برخی از روش‌های آماری را نیز می‌توان به عنوان بخشی از تحلیل پیشگویانه در نظر گرفت مثل تحلیل افتراقی، رگرسیون چندگانه، رگرسیون لجستیک و خوشه‌بندی کی-میانگین. همانطور که در شکل ۳.۷ نشان داده شده است، تحلیل‌های توصیفی دارای دو شاخه‌ی اصلی هستند: آمار و پردازش تحلیلی آنلاین. پردازش تحلیلی آنلاین عبارتست از آنالیز برای تحلیل، مشخصه‌سازی و خلاصه‌سازی داده‌های ساختیافته‌ی ذخیره شده در پایگاه داده‌های سازمانی (اغلب در یک انبار داده یا یک داده‌گاه ذخیره شده‌اند) و با استفاده از مکعب‌ها (یعنی ساختار داده‌های چندبعدی که برای استخراج یک زیرمجموعه از مقادیر داده‌ای برای پاسخگویی به یک پرسش خاص کسب‌وکار ایجاد شده‌اند) بکار می‌رود. شاخه‌ی پردازش تحلیلی آنلاین در تحلیل توصیفی را هوش کسب‌وکار نیز می‌نامند. از طرف دیگر، آمار، به توصیف داده‌ها چه به شکل یک متغیر در هر زمان یا چندمتغیره با استفاده از روش‌های توصیفی یا استنتاجی می‌پردازد.

آمار—مجموعه‌ای از تکنیک‌های ریاضی برای توصیف و تفسیر داده‌ها است که از زمان‌های بسیار دور وجود داشته است. روش‌ها و تکنیک‌های بسیاری برای پاسخگویی به نیازهای کاربران نهایی و مشخصه‌های منحصربفرد داده‌هایی که مورد تحلیل قرار می‌گیرند توسعه یافته است. به بیان کلی، در بالاترین سطح، روش‌های آماری را می‌توان به صورت توصیفی یا استنتاجی دسته‌بندی کرد. تفاوت عمده بین آمار توصیفی و استنتاجی در داده‌های بکار رفته در این روش‌ها است که آمار توصیفی مربوط به توصیف داده‌های نمونه‌ی موجود است در حالیکه آمار استنتاجی مربوط به استنتاج یا نتیجه‌گیری در خصوص مشخصه‌ها و ویژگی‌های

جمعیت است. در این بخش ما به طور مختصر، آمار توصیفی را توضیح می‌دهیم (چون زیرینا و جزء لاینفک تحلیل توصیفی است) و در بخش بعدی به توضیح رگرسیون (رگرسیون خطی و لجستیک) به عنوان بخشی از آمار استنتاجی می‌پردازیم.



شکل ۳.۷ رابطه‌ی بین آمار و تحلیل توصیفی

آمار توصیفی همانگونه که از نامش پیداست به توصیف مشخصه‌های اصلی داده‌های موجود و در اغلب اوقات، یک متغیر در هر زمان می‌پردازد. آمار توصیفی با استفاده از فرمول‌ها و تجمیع‌های عددی به خلاصه‌سازی داده‌ها به گونه‌ای می‌پردازد که الگوهای معنادار و قابل فهمی از آن مطالعه به وجود می‌آید. اگرچه آمار توصیفی در تحلیل داده‌ها بسیار مفید بوده و در بین روش‌های آماری نیز بسیار متداول است ولی امکان نتیجه‌گیری (یا استنتاج) فراتر از نمونه داده‌هایی که تحلیل شده‌اند را به ما نمی‌دهد. این یعنی روشی جذاب برای توصیف داده‌های موجود بدون هرگونه نتیجه‌گیری (استنتاج یا برون‌یابی) در خصوص جمعیت فرضیه‌هایی مرتبطی که در ذهن‌مان داریم است.

در تحلیل کسب و کار، آمار توصیفی نقش حیاتی را ایفا می‌کند و ما به اجازه می‌دهد تا داده‌هایمان را به شکلی معنادار با استفاده از اعداد تجمیعی، جداول داده‌ها یا نمودارها/گراف‌ها توضیح داده یا معرفی کنیم. آمار توصیفی در اصل به ما کمک می‌کند تا اعداد و نشانه‌های خود را به بازنمودهایی معنادار برای هر کس برای فهم و استفاده تبدیل کنیم. چنین درکی نه تنها به کاربران کسب و کار در فرایندهای تصمیم‌گیری‌شان کمک می‌کند بلکه به متخصصان تحلیل و دانشمندان علم داده نیز برای تعریف و ارزشیابی داده‌ها برای تحلیل‌های پیچیده‌تر کمک می‌کند. آمار توصیفی به تحلیلگرها اجازه می‌دهد تا تمرکز داده‌ها، مقادیر بیش از اندازه بزرگ یا کوچک غیرمعمول (مثلاً مقادیر پرت) و مقادیر داده‌های با توزیع غیرمنتظره برای متغیرهای عددی را شناسایی کنند. از این رو، روش‌های آمار توصیفی را می‌توان بر حسب سنجه‌های گرایش به مرکز یا سنجه‌های پراکندگی دسته‌بندی کرد. در بخش بعدی، ما از یک توضیح ساده و فرملاسیون/بازنمایی ریاضی این سنجه‌ها استفاده خواهیم کرد. در بازنمایی ریاضی ما از x_1, x_2, \dots, x_n برای نمایش تک تک مقادیر (مشاهدات) متغیر (سنجه) مورد نظرمان استفاده می‌کنیم.

سنجه‌های مرکزیت^۱، روش‌های ریاضی هستند که به وسیله‌ی آنها می‌توانیم موقعیت مرکزی یک متغیر مورد نظر را توضیح داده یا تخمین بزنیم. یک سنجه‌ی گرایش به مرکز، یک مقدار عددی تکی است که هدف آن توضیح دادن یک مجموعه از داده‌ها با شناسایی یا تخمین موقعیت مرکزی در داخل داده‌ها می‌باشد. میانگین (که اغلب میانگین حسابی^۲ یا متوسط ساده^۳ نامیده می‌شود) پرکاربردترین سنجه‌ی گرایش به مرکز است. علاوه بر میانگین، از میانه یا مد نیز برای توضیح مرکزیت یک متغیر استفاده می‌شود. اگرچه میانگین، میانه و مد همگی سنجه‌های معتبری برای گرایش به مرکز هستند ولی تحت شرایط مختلف، یکی از سنجه‌های مرکزیت از سایر سنجه‌ها مناسب‌تر خواهد بود. در ادامه، توضیحات مختصری از این سنجه‌ها را از جمله نحوه‌ی محاسبه‌ی ریاضی آنها و اشاره‌گرها به شرایطی که کدامیک از سنجه‌ها برای استفاده مناسب‌تر هستند ارائه می‌کنیم.

میانگین حسابی (یا فقط متوسط یا میانگین) عبارتست از جمع تمامی مقادیر/مشاهدات تقسیم بر تعداد مشاهدات در یک مجموعه داده. میانگین حسابی متداول‌ترین و پرکاربردترین سنجه‌ی گرایش به مرکز است. از آن با داده‌های عددی گسسته یا پیوسته استفاده می‌شود. برای یک متغیر مفروش x ، اگر n مقدار/مشاهده داشته باشیم (x_1, x_2, \dots, x_n) ، آنگاه می‌توانیم میانگین حسابی نمونه داده‌ها را به شکل زیر بنویسیم (\bar{x}) که به صورت ایکس-بار تلفظ می‌شود):

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

یا

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

میانگین دارای چندین ویژگی منحصربفرد است. به عنوان مثال، جمع انحرافات مطلق (اختلاف بین میانگین و مشاهدات) بالاتر از میانگین معادل است با جمع انحرافات کمتر از میانگین که بدین ترتیب مقادیر در هر سمت آن را متوازن می‌کند. با وجود این، بدان معنا نیست که نیمی از مشاهدات، بالاتر از میانگین و نیمی دیگر پایین‌تر از میانگین هستند (یک سوء تعبیر متداول بین آنهایی که آمار پایه را نمی‌دانند). همچنین، میانگین برای هر مجموعه داده، منحصربفرد است و برای هم داده‌های عددی نسبی و هم بازه‌ای قابل محاسبه و معنادار است. یکی از نقاط ضعف میانگین این است که می‌تواند تحت تأثیر مقادیر پرت قرار بگیرد (مشاهداتی که بسیار بزرگتر یا کوچکتر از سایر نقاط داده هستند). مقادیر پرت می‌توانند میانگین را به سمت خودشان بکشند و بدین ترتیب انحراف از مرکزیت به وجود می‌آید. از این رو، در صورت وجود مقادیر پرت یا داده‌های چوله‌دار یا بیش از اندازه پراکنده، یا نباید از میانگین به عنوان یک سنجه‌ی مرکزیت استفاده کرد یا اینکه باید آن را با دیگر سنجه‌های گرایش به مرکز مثل میانه و مد مقایسه کرد.

¹ Measures of Centrality Tendency

² Arithmetic mean

³ Simple average

میانۀ^۱ یک سنجۀی مقدار مرکزی در یک مجموعه داده‌ی مفروض است. میانۀ عددی است که در وسط مجموعه داده‌هایی که به ترتیب صعودی یا نزولی مرتب شده‌اند قرار دارد. اگر تعداد مشاهدات ما عددی فرد باشد آنگاه شناسایی میانۀ کار بسیار آسانی است و فقط کافی است مشاهدات را بر اساس مقدار آنها مرتب کرده و عددی که در وسط است را انتخاب کنید. ولی اگر تعداد مشاهدات ما عددی زوج باشد آنگاه باید دو عدد وسطی را شناسایی کرده و سپس میانگین ساده‌ی آن دو را محاسبه کنید. میانۀ برای داده‌های ترتیبی، بازه‌ای و نسبی، قابل محاسبه و معنادار است. پس از تعیین میانۀ، نیمی از نقاط داده‌ها در مجموعه ما در بالا و نیمی دیگر در پایین آن قرار می‌گیرند. برعکس میانگین، میانۀ تحت تأثیر مقادیر پرت یا داده‌های چوله‌دار قرار نمی‌گیرد.

مُد^۲ مشاهده‌ای است که بیش از سایر مقادیر رخ می‌دهد (پرتکرارترین مقدار در مجموعه داده). بر روی یک هیستوگرام، بلندترین میله در یک نمودار میله‌ای بیانگر مُد است و به همین خاطر می‌توان آن را شایع‌ترین مقدار/گزینه دانست. مُد بیش از همه در مجموعه داده‌هایی که حاوی تعداد نسبتاً کمی مقادیر منحصربفرد است کاربرد دارد. یعنی اگر داده‌ها ما حاوی تعداد بسیار زیادی مقادیر منحصربفرد باشد، استفاده از مُد بی‌فایده است (بسیاری از اندازه‌گیری‌های مهندسی که دقت بالا با تعداد زیادی رقم اعشار دارند) چون تعداد تکرارهای هر مقدار به عدد یک یا کمتر از آن می‌رسد. مُد اگرچه سنجۀی مفیدی است (به ویژه برای داده‌های اسمی) ولی بازنمایش خوبی برای مرکزیت نیست و به همین خاطر نباید به عنوان تنها سنجۀی گرایش به مرکز برای یک مجموعه داده خاص بکار رود.

در مجموع، بهترین سنجۀی گرایش به مرکز کدام است؟ اگرچه پاسخ روشنی برای این سوال وجود ندارد، اما نکاتی چند را می‌توان بیان کرد. زمانیکه داده‌ها دارای مقادیر پرت نبوده و چولگی زیادی وجود ندارد از میانگین استفاده کنید؛ زمانیکه داده‌های پرت داریم و/یا داده‌ها ماهیت ترتیبی دارند از میانۀ استفاده کنید و زمانیکه داده‌ها اسمی هستند از مُد. شاید بهترین روش این است که از هر سه سنجۀ استفاده کنیم تا گرایش به مرکز مجموعه داده‌ها را بتوانیم از سه منظر محاسبه کرده و نمایش دهیم. عمدتاً چون "متوسط" بیش از سایر سنجۀها در فعالیت‌های روزانه تقریباً همه شناخته شده و استفاده می‌شود، مدیران (همچنین روزنامه‌نگاران و دانشمندان) اغلب سنجۀهای مرکزیت را (به ویژه میانگین) به نادرستی بکار می‌برند در صورتیکه سایر اطلاعات آماری را باید همراه با مرکزیت در نظر گرفت. بهتر است که آمار توصیفی را به جای یک سنجۀ تکی مثل میانگین به عنوان یک بسته یعنی ترکیبی از سنجۀهای مرکزیت و افتراق یا پراکندگی در نظر بگیریم.

سنجۀهای پراکندگی^۳، روش‌های ریاضی هستند که برای تخمین یا توضیح درجه‌ی واریاسیون یک متغیر مورد نظر بکار می‌روند. آنها بیانگر گستردگی عددی (فشردگی یا عدم فشردگی) یک مجموعه داده مفروض هستند. برای توضیح این پراکندگی، چندین سنجۀ آماری تعریف شده است که معروف‌ترین آنها دامنه، واریانس و انحراف استاندارد (همچنین چارک‌ها و انحراف مطلق) می‌باشند. یکی از علل اصلی اهمیت سنجۀهای پراکندگی/گستردگی مقادیر داده‌ها این است که چارچوبی را برای قضاوت در مورد

¹ Median

² Mode

³ Measures of Dispersion/Spread/Decentrality

گرایش به مرکزیت داده‌ها به ما می‌دهند و در واقع نشان می‌دهند که میانگین (یا سایر سنجه‌های مرکزی) تا چه اندازه نمایانگر داده‌های نمونه هستند. اگر پراکندگی مقادیر در مجموعه داده بزرگ باشد آنگاه میانگین نمی‌تواند بازنمایی خوبی از داده‌ها باشد. این بدان خاطر است که یک سنجه با پراکندگی بالا نشان دهنده‌ی وجود اختلاف زیاد بین تک تک مقادیر است. همچنین مشاهده‌ی واریاسیون کوچک در هر نمونه داده می‌تواند نشان دهنده‌ی تجانس، تشابه یا توانمندی میان داده‌های گردآوری شده باشد که نشانه‌ی مثبتی در پژوهش‌ها می‌باشد.

دامنه^۱ شاید ساده‌ترین سنجه‌ی پراکندگی باشد که عبارتست از اختلاف بین بزرگترین و کوچکترین مقدار در یک مجموعه داده مفروض (یعنی متغیرها). بنابراین ما دامنه را به سادگی با شناسایی بزرگترین (ماکزیمم) و کوچکترین مقدار (مینیمم) در مجموعه داده و محاسبه‌ی اختلاف بین آن دو تعیین می‌کنیم (دامنه = ماکزیمم - مینیمم).

جامع‌ترین و پیچیده‌ترین سنجه‌ی پراکندگی، **واریانس** است. این روشی است که برای محاسبه‌ی انحراف تمامی نقاط داده در یک مجموعه داده از میانگین بکار می‌رود. هرچه واریانس بزرگتر باشد، پراکندگی داده‌ها از میانگین نیز بیشتر بوده و نوسانات بیشتری در نمونه داده‌ها مشاهده می‌شود. برای پیشگیری از متوازن شدن اختلاف‌های مثبت و منفی، مجذور فواصل از میانگین در واریانس در نظر گرفته می‌شود. فرمول واریانس برای یک نمونه داده به شکل زیر نوشته می‌شود:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

که n عبارتست از تعداد نمونه‌ها، \bar{x} میانگین نمونه و x_i که i آمین مقدار در مجموعه داده است. هرچه مقدار واریانس بزرگتر باشد نشان دهنده‌ی پراکندگی بیشتر است در حالیکه واریانس کوچکتر نشان دهنده‌ی فشردگی در کل مجموعه داده‌ها است. با دلیل مجذور بودن اختلاف‌ها، انحرافات بزرگ عمدتاً منجر به واریانس بالا می‌شوند. مجدداً به دلیل مجذور شدن اختلاف‌ها، اعدادی که بیانگر انحراف/واریانس هستند تا حدودی بی‌معنا می‌شوند (بر خلاف یک اختلاف دلار که اختلاف دلار مجذور را نتیجه می‌دهد). از این رو، به جای واریانس، در بسیاری از کاربردهای کسب‌وکار ما از یک سنجه‌ی پراکندگی معنادارتر به نام **انحراف/استاندارد** استفاده می‌کنیم.

انحراف استاندارد^۲ نیز یک سنجه‌ی پراکندگی مقادیر در یک مجموعه از داده‌ها است. انحراف استاندارد به سادگی با گرفتن ریشه‌ی دوم واریاسیون‌ها محاسبه می‌شود. فرمول زیر محاسبه‌ی انحراف استاندارد از یک نمونه نقاط داده‌ها را نشان می‌دهد.

¹ Range

² Standard Deviation

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

علاوه بر واریانس و انحراف استاندارد، برخی اوقات ما از **میانگین قدرمطلق انحرافات**^۱ نیز برای سنجش پراکندگی در یک مجموعه داده استفاده می‌کنیم. این روش ساده‌تری برای محاسبه‌ی انحراف کلی از میانگین است. مشخصاً، میانگین قدرمطلق انحرافات با اندازه‌گیری مقادیر مطلق اختلاف‌های بین هر نقطه داده و میانگین و سپس جمع کردن آنها محاسبه می‌شود. این فرایند، یک سنجش پراکندگی بدون تأثیرپذیری از نقاط داده بالاتر یا پایین‌تر از میانگین را به ما می‌دهد. فرمول زیر محاسبه‌ی میانگین قدرمطلق انحرافات را نشان می‌دهد:

$$MAD = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

چارک‌ها به شما در شناسایی پراکندگی داخل یک زیرمجموعه از داده‌ها کمک می‌کنند. یک **چارک**^۲ عبارتست از یک چهارم تعداد نقاط داده‌ای در یک مجموعه از داده‌ها. چارک‌ها ابتدا با مرتب‌سازی داده‌ها و سپس تقسیم داده‌های مرتب شده به چهار مجموعه داده‌ی منفصل کوچکتر تعیین می‌شوند. چارک‌ها سنجش مفیدی برای پراکندگی هستند زیرا نسبت به سنجش‌های معادل در کل مجموعه داده‌ها، کمتر تحت تأثیر مقادیر پرت یا چولگی در مجموعه داده‌ها قرار می‌گیرند. چارک‌ها اغلب همراه با میانه به عنوان بهترین انتخاب در سنجش پراکندگی و گرایش به مرکز در صورت وجود داده‌های چوله‌دار و/یا داده‌های پرت گزارش می‌شوند. یک روش متداول برای بیان چارک‌ها، یک دامنه‌ی بین چارکی است که اختلاف بین چارک سوم (Q3) و چارک اول (Q1) را نشان داده و در مورد دامنه‌ی نیمه‌ی میانی امتیازها در یک توزیع به ما می‌گوید. سنجش‌های توصیفی چارکی (چه مرکزی و چه پراکندگی) را به بهترین نحو می‌توان توسط نمودار جعبه‌ای توضیح داد.

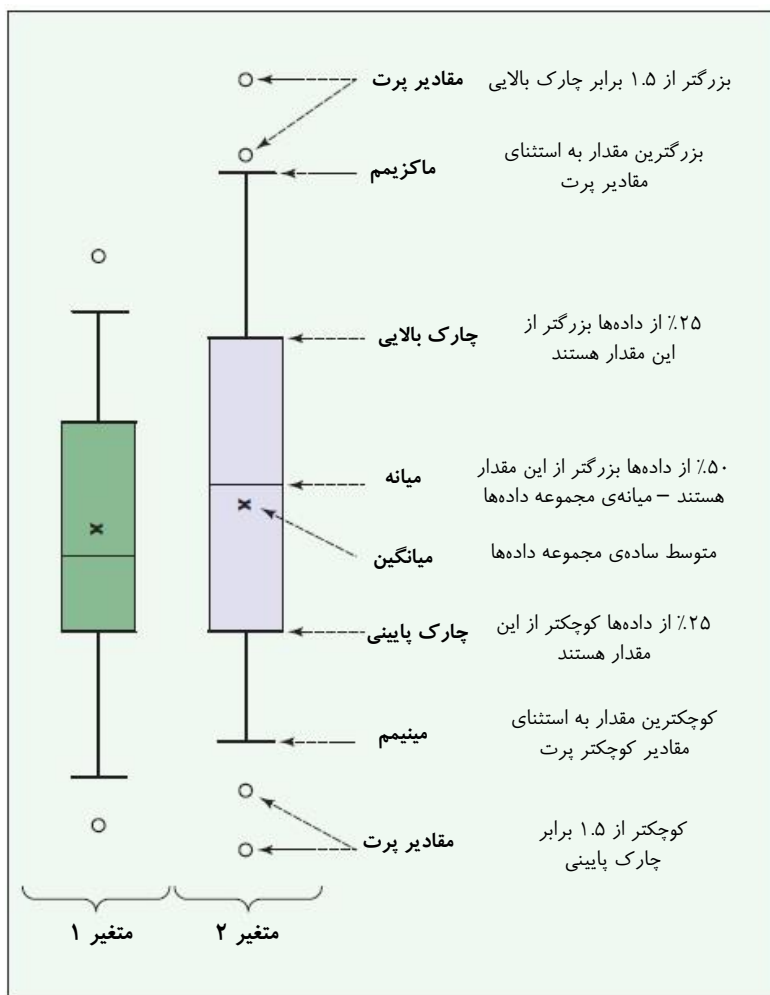
نمودار جعبه‌ای^۳ یک نمایش گرافیکی از چندین آمار توصیفی برای یک مجموعه داده‌ی مفروض است. این نمودارها به صورت افقی یا عمودی هستند ولی شکل عمودی آنها به ویژه در محصولات نرم‌افزاری تحلیلی امروزی متداول‌تر است. این نمودار اولین بار توسط جان دابلوی تیوکی در سال ۱۹۶۹ ایجاد و ارائه شد. نمودار جعبه‌ای اغلب برای نمایش مرکزیت و پراکندگی یک مجموعه داده (یعنی توزیع داده‌های نمونه) در یک قالب گرافیکی که درک آن آسان باشد استفاده می‌شود. شکل ۳.۸ دو نمودار جعبه‌ای را کنار یکدیگر نشان می‌دهد که دارای محور عمودی مشترک هستند. همانطور که مشاهده می‌کنید، برای مقایسه دیداری می‌توان در یک نمودار تکی، یک یا چند نمودار جعبه‌ای را ترسیم کرد. در چنین مواردی، محور عمودی به عنوان یک سنجش متداول بزرگی اندازه (متغیر با مقدار عددی) و محور افقی نشان دهنده‌ی دسته‌ها/زیرمجموعه‌های متفاوت مثل ابعاد زمانی مختلف (مثلاً

¹ Mean absolute deviation

² Quartile

³ Box-and-Whiskers Plot (Box Plot)

آمارهای توصیفی برای مخارج سالانه درمانی در سال ۲۰۱۵ و ۲۰۱۶) یا دسته‌های مختلف (مثلاً آمارهای توصیفی برای هزینه‌های بازاریابی در برابر فروش کل).



شکل ۳.۸ درک جزئیات نمودار جعبه‌ای

اگرچه از لحاظ تاریخی نمودار جعبه‌ای به طور گسترده و اغلب کافی مورد استفاده قرار نگرفته (به ویژه در حوزه‌های غیر از آمار)، با افزایش محبوبیت تحلیل‌های کسب‌وکار، محبوبیت آن در حوزه‌های غیرفنی‌تر در دنیای کسب‌وکار در حال افزایش است. غنای اطلاعاتی و سهولت درک و فهم آن دلیل اصلی محبوبیت اخیر این نمودار بوده است.

نمودار جعبه‌ای قادر به نمایش مرکزیت (میانه و برخی اوقات نیز میانگین) و همچنین پراکندگی (چگالی داده‌ها در نیمه میانی که به صورت یک جعبه بین چارک‌های اول و سوم ترسیم شده است)، دامنه‌های بیشینه و کمینه (به صورت خطوطی که از جعبه بیرون آمده‌اند که ویسکر نام دارند و به صورت ۱.۵ برابر انتهای بالایی یا پایینی جعبه‌ی چارک محاسبه می‌شوند) و مقادیر پرتی که بزرگتر از حدود این

خطوط باشند است. یک نمودار جعبه‌ای همچنین نشان می‌دهد که آیا داده‌ها دارای توزیع متقارن نسبت به میانگین هستند یا خیر. موقعیت نسبی میانه در مقایسه با میانگین و طول خطوط ویسکر در هر دو طرف جعبه، نشانه‌ی خوبی از چولگی بالقوه در داده‌ها است.

اگرچه شکل توزیع داده‌ها به اندازه‌ی مرکزیت و پراکندگی متداول نیست ولی سنجه‌ی مفیدی در آمار توصیفی محسوب می‌شود. پیش از آنکه وارد بحث شکل توزیع شویم ابتدا لازم است خودِ توزیع را تعریف کنیم. به بیان ساده، توزیع^۱ عبارتست از تکرار یا فراوانی نقاط داده‌ای که بر روی تعداد محدودی از دامنه‌های عددی یا برچسب‌های دسته‌ای شمارش شده و ترسیم شده باشند. در یک تصویر گرافیکی از توزیع، محور عمودی (y) نشان دهنده‌ی فراوانی (شمارش یا درصد)، محور افقی (x) نشان دهنده‌ی تک تک دسته‌ها به شکل مرتب شده است. یک توزیع بسیار شناخته‌شده، توزیع نرمال^۲ است که در هر دو طرف میانگین دارای تقارن کامل بوده و چندین ویژگی ریاضی بسیار خوب نیز دارد که آن را به ابزاری بسیار مفید برای پژوهش و کاربرد تبدیل کرده است. با افزایش پراکندگی یک مجموعه داده، انحراف استاندارد نیز افزایش یافته و شکل توزیع، گسترده‌تر می‌شود. یک نمایش گرافیکی از ارتباط بین پراکندگی و شکل توزیع (در خصوص توزیع نرمال) در شکل ۳.۹ نشان داده شده است.

برای محاسبه‌ی مشخصه‌های شکلی یک توزیع، عمدتاً از دو سنجه استفاده می‌شود: چولگی و کشیدگی. اغلب از یک هیستوگرام (نمودار فراوانی) برای نمایش چولگی و کشیدگی استفاده می‌شود.

چولگی^۳ یک سنجه‌ی عدم تقارن در یک توزیع از داده‌ها است که یک ساختار تک نمایی^۴ را به تصویر می‌کشد یعنی در توزیع داده‌ها فقط یک نقطه‌ی اوج وجود دارد. از آنجائیکه توزیع نرمال یک توزیع تک نمایی کاملاً متقارن است، فاقد چولگی است. این بدان معناست که سنجه‌ی چولگی آن (یعنی مقدار ضریب چولگی) معادل صفر است. سنجه/مقدار چولگی می‌تواند مثبت یا منفی باشد. در صورتیکه چولگی توزیع به سمت چپ باشد (یعنی دم در سمت راست بوده و میانگین کوچکتر از میانه باشد)، آنگاه چولگی مثبت خواهد بود و در صورتیکه چولگی توزیع به سمت راست باشد (یعنی کشیدگی دم در سمت چپ بوده و میانگین بزرگتر از میانه باشد) آنگاه چولگی منفی خواهد بود. در شکل ۳.۹، حرف (c) معرف یک توزیع با چولگی مثبت است در حالیکه (d) بیانگر یک توزیع با چولگی منفی است. در همین شکل، هم (a) و هم (b) معرف تقارن کامل بوده و بدین ترتیب مقدار چولگی آنها صفر است.

$$\text{چولگی} = S = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)s^3}$$

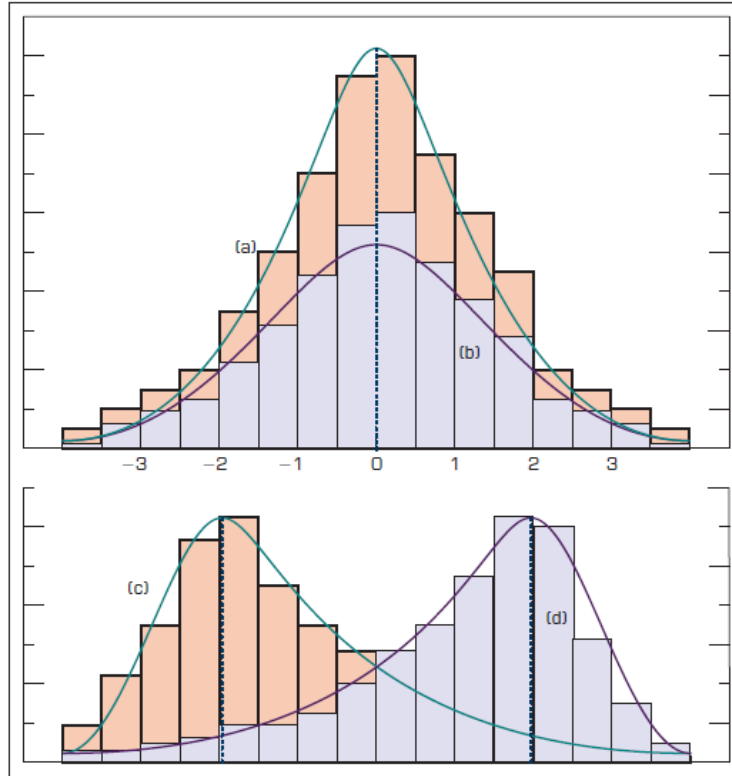
که S انحراف استاندارد و n تعداد نمونه است.

¹ Distribution

² Normal Distribution

³ Skewness

⁴ Unimodal



شکل ۳.۹ ارتباط بین ویژگی‌های شکلی پراکندگی و توزیع

کشیدگی^۱ یک سنج‌هی دیگر برای تعریف شکل یک توزیع تک‌نمایی است. برخلاف چولگی، کشیدگی بیشتر بر تعریف ماهیت قله/بلندی/لاغری توزیع تمرکز دارد. مشخصاً، کشیدگی به اندازه‌گیری درجه‌ی قله‌ای بودن یک توزیع نسبت به توزیع نرمال می‌پردازد. کشیدگی مثبت نشان‌دهنده‌ی یک توزیع نسبتاً بلند/قله‌ای است در حالیکه کشیدگی منفی بیانگر یک توزیع نسبتاً کوتاه/مسطح است. به عنوان یک نقطه‌ی مرجع، یک توزیع نرمال دارای کشیدگی ۳ است. فرمول کشیدگی را به شکل زیر می‌توان نوشت:

$$\text{کشیدگی} = K = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{ns^3} - 3$$

آمار توصیفی (و همچنین آمار استنتاجی) را به آسانی می‌توان با استفاده از بسته‌های نرم‌افزار آماری موجود در بازار (مثلاً اس‌ای‌اس^۲، اس‌پی‌اس‌اس^۳، مینی‌تب^۴، جی‌ام‌پی^۵، استتیسیتیکا^۶) یا ابزارهای منبع باز/رایگان (مثل آر^۷) محاسبه کرد. شاید راحت‌ترین روش برای

¹ Kurtosis

² SAS

³ SPSS

⁴ Minitab

⁵ JMP

⁶ Statistica

⁷ R

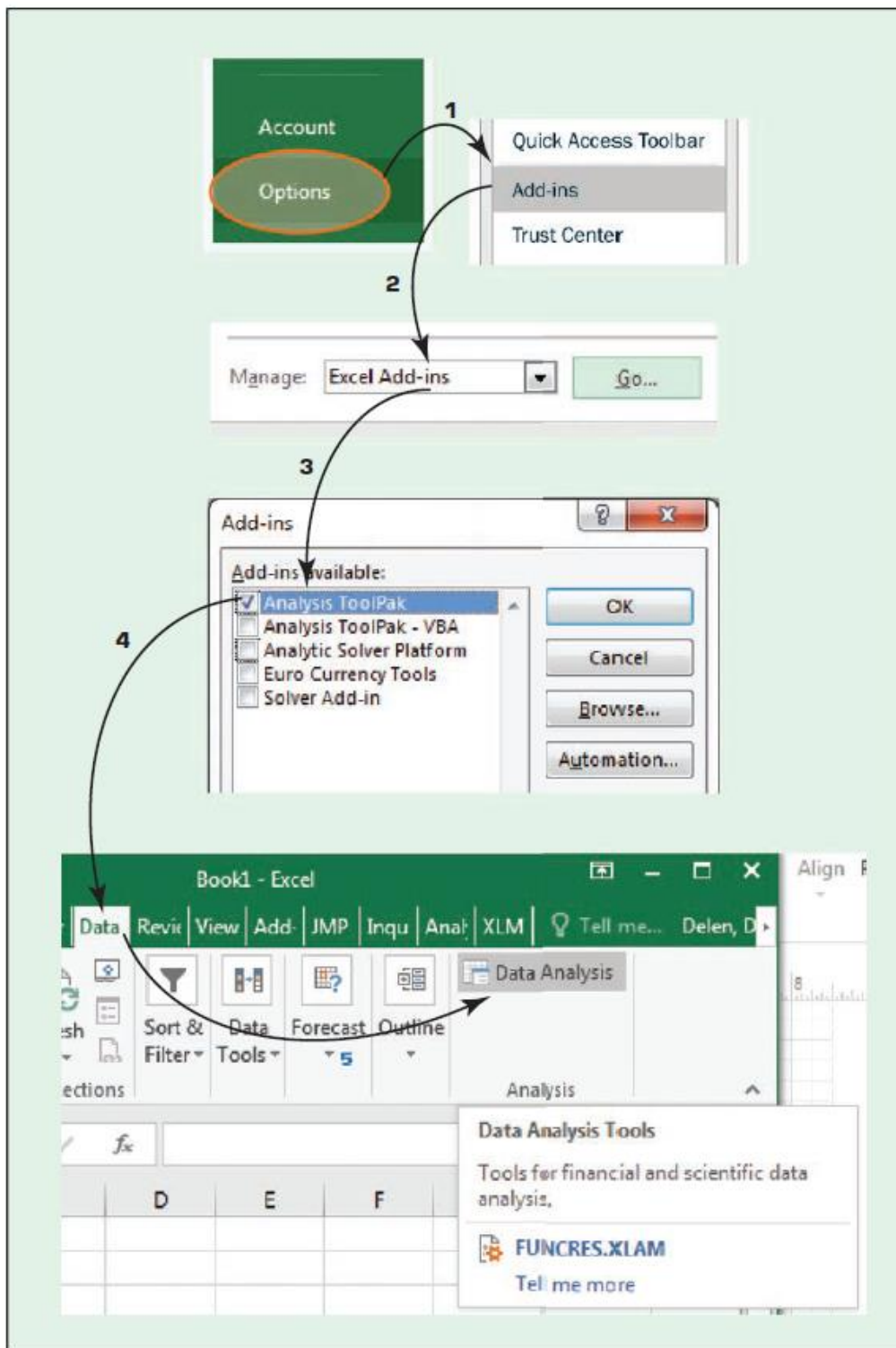
محاسبه‌ی آمارهای توصیفی و برخی از آمارهای استنتاجی، استفاده از اکسل باشد. در بینش فناوری ۳.۱ به تفصیل در مورد نحوه استفاده از مایکروسافت اکسل برای محاسبه‌ی آمارهای توصیفی توضیح داده شده است.

بینش فناوری ۳.۱ چگونه آمارهای توصیفی را در مایکروسافت اکسل محاسبه کنیم

اکسل را شاید بتوان محبوب‌ترین ابزار تحلیل داده در جهان دانست که به آسانی می‌توان از آن برای محاسبات آمار توصیفی استفاده کرد. اگرچه پیکربندی و مبنای اکسل به نظر نمی‌رسد که کارکردهای آماری را به صورت آماده برای کاربران نهایی ارائه کند ولی این توابع و کارکردهای در زمان نصب اکسل در دسترس قرار گرفته و تنها با چند کلیک ماوس می‌توان آنها را فعال نمود. شکل ۳.۱۰ نحوه‌ی فعال‌سازی این توابع آماری (به عنوان بخشی از Analysis ToolPak) را در مایکروسافت اکسل ۲۰۱۶ نشان می‌دهد.

به محض فعال‌سازی، Analysis Toolpak در گزینه‌ی منوی Data و تحت عنوان Data Analysis ظاهر می‌شود. زمانیکه در گروه Analysis و زیر برگه‌ی Data بر روی Data Analysis کلیک کنید، آمار توصیفی به صورت یکی از گزینه‌های موجود رد فهرست منو بار اکسل پدیدار خواهد شد (شکل ۳.۱۱، گام‌های ۱ و ۲). بر روی OK کلیک کنید و جعبه پیغام Descriptive Statistics ظاهر خواهد شد (بخش میانی شکل ۳.۱۱). در این جعبه پیام، شما باید دامنه یا محدوده‌ی داده‌ها که می‌تواند یک یا چند ستون عددی باشد را به همراه چک بکس‌های اولویت انتخاب کرده و بر روی OK کلیک کنید (شکل ۳.۱۱، گام‌های ۳ و ۴). اگر محدوده انتخاب شما شامل بیش از یک ستون عددی باشد، هر ستون به صورت یک مجموعه داده‌ی مجزا در نظر گرفته شده و آمارهای توصیفی برای هر ستون به صورت جداگانه محاسبه خواهد شد.

برای یک مثال ساده، ما دو ستون (با عنوان‌های Expense و Demand) را انتخاب کرده و گزینه‌ی Descriptive Statistics را اجرا کردیم. بخش پایینی شکل ۳.۱۱ خروجی تهیه شده توسط اکسل را نشان می‌دهد. همانطور که مشاهده می‌کنید، اکسل تمامی آمارهای توصیفی بحث شده در بخش قبل را به همراه چند آمار بیشتر به فهرست افزوده است. در اکسل ۲۰۱۶، به آسانی (تنها با چند کلیک ماوس) می‌توانید نمودارهای جعبه‌ای را ترسیم کنید. شکل ۳.۱۲ یک فرایند سه مرحله‌ای ساده را برای ترسیم نمودار جعبه‌ای در اکسل نشان می‌دهد.



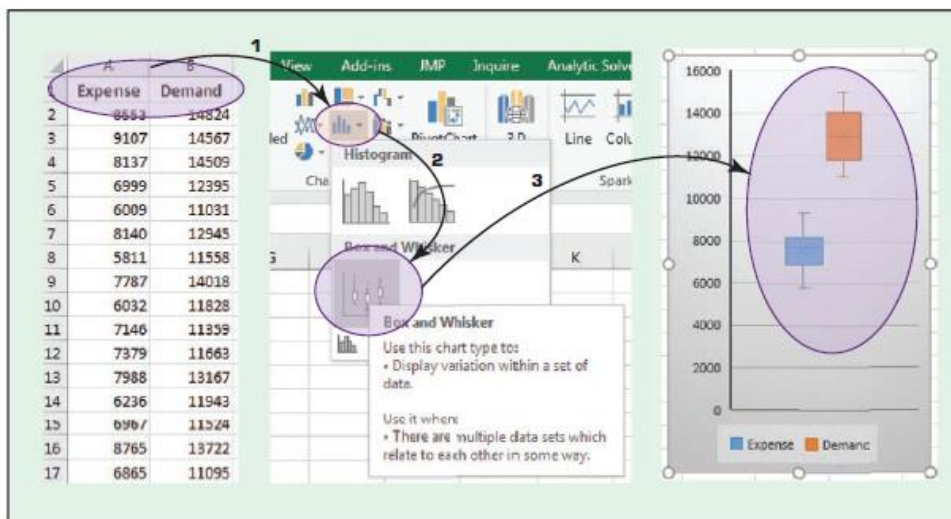
شکل ۳.۱۰ فعال‌سازی تابع آماری در اکسل ۲۰۱۶

The image shows the SAS Data Analysis and Descriptive Statistics dialog boxes. In the 'Data Analysis' dialog, 'Descriptive Statistics' is selected (1). An arrow points to the 'Descriptive Statistics' dialog where the 'Input Range' is set to 'SAS1:SB\$1:21' (2). Another arrow points to the 'Output options' section where 'Summary statistics' and 'Confidence Level for Means' (95%) are checked (3). A final arrow points to the resulting output table (4).

	A	B
1	Expense	Demand
2	8553	14824
3	9107	14567
4	8137	14509
5	6999	12395
6	6009	11031
7	8140	12945
8	5811	11558
9	7787	14018
10	6032	11828
11	7146	11359
12	7379	11663
13	7988	13167
14	6236	11943
15	6967	11524
16	8765	13722
17	6865	11095

	A	B	C	D
1	Expense		Demand	
2				
3	Mean	7595.808333	Mean	12976.53333
4	Standard Error	78.07347263	Standard Error	109.7357916
5	Median	7679	Median	12945
6	Mode	8140	Mode	14567
7	Standard Deviation	855.252042	Standard Deviation	1202.095369
8	Sample Variance	731456.0554	Sample Variance	1445033.276
9	Kurtosis	-0.82138214	Kurtosis	-1.290541404
10	Skewness	-0.235479407	Skewness	-0.002281156
11	Range	3526	Range	3944
12	Minimum	5811	Minimum	11019
13	Maximum	9337	Maximum	14963
14	Sum	91149	Sum	1557184
15	Count	120	Count	120
16	Confidence Level(95.0%)	154.5932735	Confidence Level(95.0%)	217.2878275

شکل ۳.۱۱ دستیابی به آمارهای توصیفی در اکسل



شکل ۳.۱۲ ایجاد یک نمودار جعبه‌ای در اکسل ۲۰۱۶

اگرچه Analysis ToolPak ابزار بسیار مفیدی در اکسل است ولی باید از یک نکته‌ی مهم در خصوص نتایجی که تولید می‌کند آگاه باشید و رفتاری متفاوت نسبت به سایر توابع معمول اکسل دارد: اگرچه توابع اکسل به صورت پویا با تغییر داده‌های زیربنایی در صفحه‌گسترده، تغییر می‌کنند ولی نتایج تولید شده توسط Analysis ToolPak تغییر نمی‌کنند. به عنوان مثال، اگر مقادیر یک یا هر دو ستون را تغییر دهید، نتایج آمار توصیفی تولید شده توسط Analysis ToolPak تغییر نخواهند کرد. با وجود این، این موضوع برای توابع معمول اکسل صادق نیست. اگر قرار باشد مقدار میانگین یک مجموعه داده را محاسبه کنید (با استفاده از $=AVERAGE(A1:A121)$) و سپس مقادیر داخل مجموعه داده‌ها را تغییر دهید آنگاه مقدار میانگین به طور خودکار تغییر خواهد کرد. به طور خلاصه، نتایج تولید شده توسط Analysis ToolPak دارای پیوند پویا به داده‌های زیربنایی نیست و اگر داده‌ها تغییر کنند، تحلیل را باید مجدداً با استفاده از جعبه پیام انجام داد.

کاربردهای موفق تحلیل داده‌ها طیف گسترده‌ای از کسب‌وکارها و محیط‌های سازمانی را پوشش داده و به مسائلی که غیرقابل حل بودند رسیدگی می‌کنند. مورد کاربردی ۳.۳ یک مثال عالی از داستان‌های موفقیت‌آمیزی است که یک مدیریت شهرداری کوچک از رویکرد تحلیل داده‌ها برای تشخیص هوشمندانه و حل مسائل از طریق تحلیل مستمر الگوهای تقاضا و مصرف بکار گرفته است.

سوالاتی برای مرور بخش ۳.۵

۸. ارتباط بین تحلیل آمار و تحلیل کسب‌وکار چیست؟
۹. تفاوت‌های اصلی بین آمار توصیفی و استنتاجی چیست؟
۱۰. سنجه‌های گرایش به مرکز آمارهای توصیفی را نام برده و به طور خلاصه تعریف کنید.
۱۱. سنجه‌های پراکندگی آمارهای توصیفی را نام برده و به طور خلاصه تعریف کنید.
۱۲. نمودار جعبه‌ای چیست؟ چه نوع اطلاعات آماری را نمایش می‌دهد؟
۱۳. دو مشخصه‌ی شکل توزیع که به طور متداول برای توضیح یک توزیع داده استفاده می‌شود کدام‌ها هستند؟

یک شیر آب نشستی دار، یک ماشین ظرفشویی خراب، یک سِری خراب آبپاش. تعمیر اینها چیزی بیش از یک در دسر برای صاحب منزل یا صاحب مغازه است. متأسفانه اینها می‌توانند هزینه‌بر و غیرقابل پیش‌بینی بوده و به سختی قابل تشخیص باشند. شهرداری کاری^۱ در کارولینای شمالی از طریق ترکیبی از دستگاه‌های بی‌سیم و یک پرتال مبتنی بر تحلیل داده که توسط مشتری قابل دسترس است، یافتن و رفع مشکلات هدررفت آب را بسیار آسان کرده است. در این فرایند، شهرداری به تصویر بزرگی از مصرف آب که برای برنامه‌ریزی گسترش نیروگاه‌های آبی در آینده و ترویج فعالیت‌های هدفمند حفظ آب لازم است دست یافته است.

زمانیکه شهر کاری کنتورهای بی‌سیم را برای ۶۰,۰۰۰ مشتری در ۲۰۱۰ نصب نمود می‌دانست که فناوری جدیدش با حذف قرائت دستی کنتورها نمی‌تواند موجب صرفه‌جویی پولی شود. شهرداری نیز متوجه شد که به اطلاعات دقیق‌تر و به‌موقع‌تری در مورد مصرف آب دست پیدا می‌کند. سیستم بی‌سیم آکوآ استار^۲ قادر به قرائت کنتورها در هر ساعت یکبار است - یعنی ۸,۶۷۰ نقطه داده به ازای هر مشتری در هر سال به جای قرائت ۱۲ ماهانه در سال. داده‌ها در صورتیکه به آسانی استفاده شوند ارزش بالقوه‌ی عظیمی دارند.

به گفته‌ی کارن میلیز^۳ مدیر مالی شهرداری کاری، "قرائت ماهانه کنتورها مثل این است که یک گالن داده با ارزش یک گالن آب داشته باشیم. قرائت ساعتی کنتورها شبیه یک استخر به اندازه‌ی استخر مسابقات المپیک پر از داده‌ها است. اس‌ای‌اس به ما کمک می‌کند تا حجم داده‌ها را به خوبی مدیریت کنیم." در واقع، این راه‌حل، شهرداری را قادر می‌کند تا نیم میلیارد نقطه داده در خصوص مصرف آب را تحلیل کرده و سپس آنها را در اختیار تمامی مشتریان قرار دهد.

اینکه صاحب خانه یا مشتری تجاری بتواند سر ساعت نگاهی به داده‌ها داشته باشد منجر به برخی کاربردهای بسیار عملی شده است:

- شهرداری می‌تواند شهروندان را در خصوص نشستی آب در عرض چند روز مطلع کند.
- مشتریان در صورت پیک مصرف آب، هشدارهایی را برای مطلع شدن در عرض چند ساعت دریافت می‌کنند.
- مشتریان می‌توانند میزان مصرف آب را به صورت آنلاین پیگیری کنند که به آنها کمک می‌کند تا در مصرف آب محتاط‌تر باشند.

یک مغازه از طریق پرتال آنلاین متوجه یک پیک در مصرف آب در اواخر هفته و زمانیکه کارکنان حضور نداشتند شد. این به نظر عجیب رسید و قرائت‌های نامعمول به شرکت کمک کرد تا متوجه شود که یک ماشین ظرفشویی تجاری دچار خرابی بوده و تمامی آخر هفته‌ها آب مصرف می‌کرده است. بدون داده‌های کنتور آب‌های بی‌سیم و پرتال مشتریان، شرکت ممکن بود متوجه این مسئله نشده و آب و پول را هدر می‌داد.

شهر اکنون تصویر بسیار دقیق‌تری از سرانه مصرف روزانه آب دارد که برای برنامه‌ریزی گسترش نیروگاه‌های آب در آینده بسیار حیاتی است. شاید جالب‌ترین مزیت این است که شهر قادر بود تا *verify a hunch that has far-reaching cost*

¹ Town of Cary

² Aquastar

³ Karen Mills

ramifications. شهروندان کاری در مصرف آب بسیاری اقتصادی هستند. مدیر منابع آبی شهر، لیلا گودوین^۱ می‌گوید « ما محاسبه کرده‌ایم که با لوازم خانگی مدرن و کارآمد، مصرف آب خانگی می‌تواند تا ۳۵ گالن در روز به ازای هر نفر کاهش یابد.» این چرا مهم است؟ شهر، پول زیادی را صرف ترویج کارایی مصرف آب می‌کرد مثل تخفیف بابت نصب دستشویی‌های کم جریان یا بشکه باران. اکنون شهرداری می‌تواند رویکرد هدفمندتری را برای کمک به مصرف‌کنندگان جهت درک و مدیریت مصرف آب خانگی و تجاری اتخاذ کند.

اس‌ای‌اس فقط برای اینکه به شهروندان کمک کرد تا میزان مصرف آب خود را متوجه شوند حیاتی نبود بلکه در پشت صحنه سعی می‌کرد تا دو پایگاه داده‌ی مجزا به یکدیگر پیوند دهد. به گفته‌ی میلز، «ما یک پایگاه داده‌ی صدور صورتحساب و یک پایگاه داده‌ی قرائت کنتور داریم که باید آنها را کنار یکدیگر بیاوریم تا قابل نمایش و ارائه باشد.»

شهر تخمین می‌زند که سیستم آکوا استار فقط با حذف قرائت‌های دستی بیش از ۱۰ میلیون دلار در هزینه‌های پروژه صرفه‌جویی خواهد کرد. اما مؤلفه‌ی تحلیل می‌تواند صرفه‌جویی‌های بیشتری به همراه داشته باشد. تا اینجا، هم شهرداری و هم شهروندان با تشخیص زود هنگام نشتی آب توانسته‌اند در هزینه‌هایشان صرفه‌جویی کنند. همچنان که شهر کاری به برنامه‌ریزی نیازهای زیرساختی آینده‌ی خود ادامه می‌دهد، داشتن اطلاعات دقیق در خصوص مصرف آب کمک خواهد کرد تا سرمایه‌گذاری مناسب را در زمان مناسب بر روی زیرساخت‌ها انجام دهد. علاوه بر این، درک میزان مصرف آب به شهر کمک خواهد کرد تا اگر در مقابل شرایط سخت همانند خشکسالی برنامه‌ریزی کنند.

به گفته‌ی گودوین، «ما در سال ۲۰۰۷ با خشکسالی مواجه شدیم. اگر یکبار دیگر چنین چیزی را تجربه کنیم، طرحی داریم تا از داده‌های آکوا استار برای کنترل و مشاهده‌ی مصرف دقیق روزانه‌ی آب و اطلاع‌رسانی آن به مشتریان استفاده کنیم. ما در اینجا می‌توانیم نشان بدهیم که چه اتفاقی می‌افتد و به خاطر کم بودن میزان ذخائر، چقدر مجاز به استفاده هستیم. امیدواریم که هیچگاه از این طرح استفاده نکنیم، ولی ما آماده هستیم.»

Source: "Municipality Puts Wireless Water Meter-Reading Data To Work (SAS Analytics)—The Town of Cary, North Carolina Uses SAS Analytics to Analyze Data from Wireless Water Meters, Assess Demand, Detect Problems and Engage Customers." Copyright © 2016 SAS Institute Inc., Cary, NC, USA. Reprinted with permission. All rights reserved.

سوالاتی برای مورد کاربردی ۳.۳

۷. شهر کاری با چه چالش‌هایی مواجه بود؟
۸. راه‌حل پیشنهادی چه بود؟
۹. نتایج آن چه بودند؟
۱۰. چه مسائل و راه‌حل‌های تحلیل داده را برای شهرهایی مثل کاری پیش‌بینی می‌کنید؟

رگرسیون^۲ و به ویژه رگرسیون خطی شاید شناخته شده‌ترین و پرکاربردترین تکنیک تحلیلی در آمار باشد. از لحاظ تاریخی، ریشه‌ی رگرسیون به سال‌های ۱۹۲۰ و ۱۹۳۰ و کارهای اولیه‌ی انجام شده بر روی ویژگی‌های وراثتی نخودفرنگی‌ها که توسط سر

¹ Leila Goodwin

² Regression

فرانسیس گالتون^۱ و سپس کارل پیرسون^۲ انجام شده بود باز می‌گردد. پس از آن، رگرسیون تبدیل به یک تکنیک آماری برای تعریف ارتباطات بین متغیر توصیفی یا مستقل (ورودی)^۳ و متغیر پاسخ یا وابسته (خروجی)^۴ گردید. محبوبیت زیاد رگرسیون بدان خاطر است که یک تکنیک آماری نسبتاً ساده برای مدل‌سازی وابستگی بین یک متغیر (متغیر پاسخ یا خروجی) به یک (یا چند) متغیر توصیفی (ورودی) است. این ارتباط بین متغیرها را به محض شناسایی می‌توان رسماً به صورت یک معادله/تابع خطی/جمع‌پذیر^۵ نمایش داد. همانند بسیاری از تکنیک‌های مدل‌سازی دیگر، هدف رگرسیون این است که ارتباط تابعی بین و مابین ویژگی‌های دنیای واقعی را گرفته و این ارتباط را با یک مدل ریاضی توضیح دهید که سپس می‌توان از آن برای کشف و درک پیچیدگی‌های دنیای واقعی استفاده کرد و به کشف و توضیح ارتباطات یا پیش‌بینی وقوع رخدادها در آینده پرداخت.

از رگرسیون می‌توان برای یکی از دو هدف زیر استفاده کرد: آزمون فرض یعنی تحقیق و بررسی در خصوص ارتباطات بالقوه بین متغیرهای مختلف و پیش‌بینی/پیشگویی یعنی برآورد مقادیر یک متغیر وابسته (پاسخ) بر اساس یک یا چند متغیر توصیفی (مستقل). این دو کاربرد به طور متقابل انحصاری نیستند. توان توصیفی رگرسیون نیز زیربنای توانایی پیش‌بینی آن است. در آزمون فرض، (نظریه‌سازی)، تحلیل رگرسیون می‌تواند وجود/قدرت و جهت‌گیری ارتباطات بین تعدادی از متغیرهای توصیفی (که اغلب با X_i نمایش داده می‌شوند) و متغیر پاسخ (که اغلب با Y نمایش داده می‌شوند) را آشکار کند. در بحث پیشگویی، رگرسیون می‌تواند ارتباطات ریاضی بیشتری را (به شکل یک معادله) بین یک یا چند متغیر توصیفی و یک متغیر پاسخ شناسایی کند. پس از تعیین این معادله می‌توان از آن برای پیش‌بینی مقادیر متغیر پاسخ برای مجموعه مشخصی از مقادیر متغیرهای توصیفی استفاده کرد.

همبستگی و رگرسیون چونکه تحلیل رگرسیون از مطالعات همبستگی نشأت گرفته است و چون هر دو روش سعی دارند تا رابطه‌ی بین دو (یا چند) متغیر را توضیح دهند، متخصصان و حتی دانشمندان اغلب این دو عبارت را با یکدیگر اشتباه می‌گیرند. در همبستگی، هیچ پیش فرضی در مورد اینکه آیا یک متغیر به متغیر(های) دیگر وابسته است یا خیر وجود ندارد و ارتباط بین متغیرها نیز اهمیتی ندارد؛ در عوض، درجه‌ی ارتباط بین متغیرها را تخمین می‌زند. از طرف دیگر، رگرسیون سعی دارد تا وابستگی یک متغیر پاسخ به یک یا چند متغیر توصیفی را توضیح دهد که به طور ضمنی فرض می‌گیرد که یک رابطه‌ی علی یک طرفه از طرف متغیر(های) توصیفی به متغیر پاسخ وجود دارد که مستقیم یا غیرمستقیم بودن این مسیر اهمیتی ندارد. همچنین، اگرچه همبستگی به ارتباطات سطح پایین بین دو متغیر توجه دارد، رگرسیون با ارتباطات بین تمامی متغیرهای توصیفی و متغیر پاسخ سر و کار دارد.

رگرسیون ساده و چندگانه اگر معادله‌ی رگرسیون بین یک متغیر پاسخ و یک متغیر توصیفی برقرار شود آنگاه آن را رگرسیون ساده می‌نامند. به عنوان مثال، معادله‌ی رگرسیون برای پیش‌بینی/توضیح ارتباط بین قد یک شخص (متغیر توصیفی) و وزن یک شخص (متغیر پاسخ)، مثال خوبی از رگرسیون ساده است. رگرسیون چندگانه را می‌توان گسترش رگرسیون ساده با در نظر

¹ Sir Francis Galton

² Karl Pearson

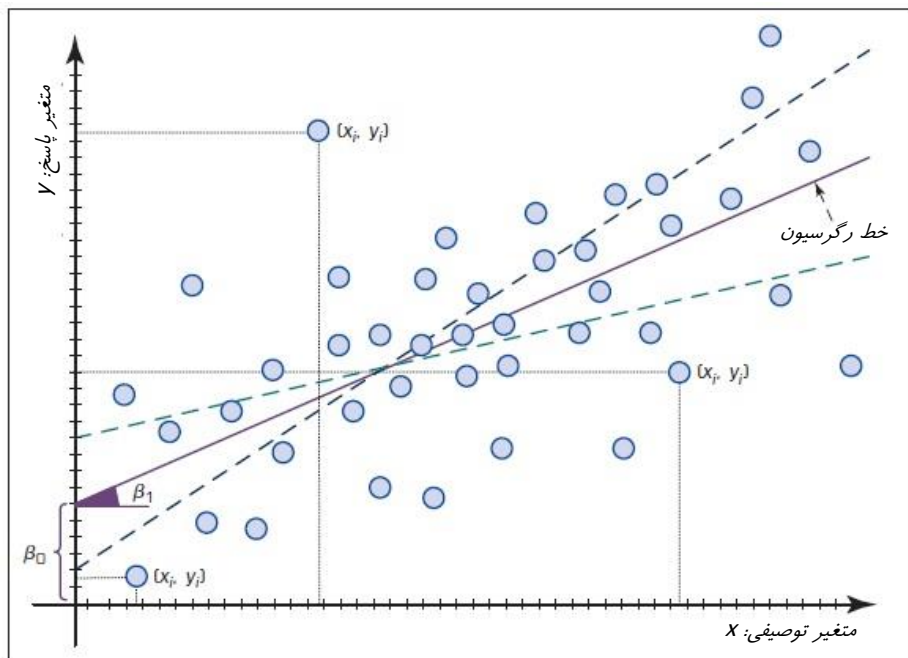
³ Explanatory variable

⁴ Response Variable

⁵ Additive/Linear

گرفتن بیش از یک متغیر توصیفی دانست. به عنوان مثال، در مثال قبل، اگر ما سایر ویژگی‌های شخصی از قبیل جنسیت، نژاد و .. را برای پیش‌بینی وزن شخص در نظر بگیریم آنگاه تحلیل رگرسیون چندگانه انجام داده‌ایم. در هر دو حالت، ارتباط بین متغیر پاسخ و متغیر(های) توصیفی دارای ماهیت خطی و جمع‌پذیر است. اگر ارتباطات به صورت خطی نباشند آنگاه باید از یکی از روش‌های دیگر رگرسیون غیرخطی برای نمایش بهتر ارتباطات بین متغیرهای ورودی و خروجی استفاده کنیم.

برای درک ارتباط بین دو متغیر، ساده‌ترین کار این است که یک نمودار پراکندگی رسم کنیم که محور y آن بیانگر مقادیر متغیر پاسخ و محور x آن بیانگر مقادیر متغیر توصیفی باشد (شکل ۳.۱۳). یک نمودار پراکندگی، تغییرات متغیر پاسخ را به صورت تابعی از تغییرات متغیر توصیفی نمایش می‌دهد. در شکل ۳.۱۳، به نظر می‌رسد ارتباط مثبتی بین این دو برقرار است و با افزایش مقادیر متغیر توصیفی، متغیر پاسخ نیز افزایش می‌یابد.



شکل ۳.۱۳ نمودار پراکندگی و خط رگرسیون خطی

هدف از تحلیل رگرسیون ساده، یافتن یک بازنمایی ریاضی برای این ارتباط است. در واقعیت، سعی می‌کند تا خط مستقیمی را که درست از بین نقاط ترسیم شده (بیانگر داده‌های تاریخی/مشاهدات) به گونه‌ای پیدا کند که فاصله‌ی بین نقاط و خط به حداقل مقدار برسد (مقادیر پیش‌بینی شده بر روی خط رگرسیون نظری). اگرچه چندین روش/الگوریتم پیشنهادی برای شناسایی خط رگرسیون وجود دارد ولی پرکاربردترین آنها، روش حداقل مربعات معمولی (اِل‌اِس)^۱ نام دارد. هدف روش حداقل مربعات معمولی این است که مجموع مربعات باقیمانده (مربعات فواصل عمودی بین مشاهده و نقطه رگرسیون) را به حداقل رسانده و یک عبارت

¹ Ordinary Least Squares (OLS)

ریاضی برای مقدار تخمین شده‌ی خط رگرسیون ارائه دهد (که با نام پارامترهای β شناخته می‌شوند). برای یک رگرسیون خطی ساده، ارتباط مذکور بین متغیر پاسخ (y) و متغیر(های) توصیفی (x) را می‌توان به صورت یک معادله‌ی ساده زیر نوشت:

$$y = \beta_0 + \beta_1 x$$

در این معادله، β_0 را عرض از مبدأ و β_1 را شیب می‌گویند. پس از تعیین مقادیر این دو ضریب، از این معادله‌ی ساده می‌توان برای پیش‌بینی مقادیر y برای مقادیر مفروض x استفاده کرد. علامت و مقدار β_1 نیز نشان دهنده‌ی جهت‌گیری و قوت ارتباط بین این دو متغیر می‌باشد.

اگر این مدل از نوع رگرسیون خطی چندگانه باشد آنگاه ضریب‌های بیشتری را برای هر کدام از متغیرهای توصیفی اضافی باید تعیین نمود. همانطور که در فرمول زیر مشاهده می‌کنید، متغیرهای توصیفی اضافی در ضرایب β_i جدید ضرب شده و برای تشکیل یک نمایش جمع‌پذیر خطی از متغیر پاسخ با یکدیگر جمع می‌شوند.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n$$

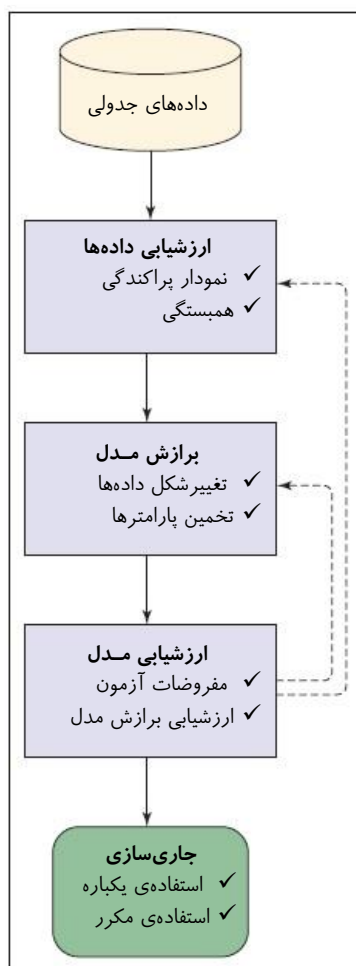
به دلایل متعدد، برخی اوقات مدل‌ها به عنوان بازنمودهایی از واقعیت، عملکرد خوبی ندارند. صرفنظر از تعداد متغیرهای توصیفی موجود، همیشه این امکان وجود دارد که مدل خوبی نداشته باشیم و بنابراین، برازش مدل رگرسیون خطی باید مورد ارزشیابی قرار بگیرد (این مدل تا چه اندازه بیانگر متغیر پاسخ است). به ساده‌ترین مفهوم، یک مدل رگرسیون با برازش مناسب، مقادیری را پیش‌بینی می‌کند که نزدیک به مقادیر داده‌های مشاهده شده هستند. برای ارزیابی عددی، اغلب از سه سنجهی آماری برای ارزشیابی برازش یک مدل رگرسیون استفاده می‌شود: R^2 ، آزمون اف^۱ و خطای میانگین مربعات ریشه (آرام‌اس‌ای)^۲. هر سه این سنجه‌ها مبتنی بر مجموع خطای مربعات هستند (داده‌ها به چه میزان از میانگین و از مقادیر پیش‌بینی شده‌ی مدل فاصله دارند). ترکیبات مختلف این دو مقدار، اطلاعات متفاوتی را در خصوص نحوه‌ی مقایسه‌ی مدل رگرسیون با مدل میانگین به ما می‌دهد.

از بین این سه روش، R^2 مفیدترین و قابل فهم‌ترین معنا را به دلیل مقیاس شهودی خود دارد. مقدار R^2 بین ۰ تا ۱ متغیر است (متناظر با میزان تغییرپذیری بر حسب درصد) که نشان دهنده‌ی این است که ارتباط و توان پیش‌بینی مدل پیشنهادی خوب نیست در حالیکه ۱ نشان می‌دهد که مدل پیشنهادی از برازش کامل برخوردار است که پیش‌بینی‌های دقیقی را ارائه می‌کند (که چنین موردی تقریباً محال است). مقادیر خوب R^2 معمولاً نزدیک به ۱ هستند و در اینجا نزدیک بودن در خصوص پدیده‌ای که مدل‌سازی می‌شود بسیار اهمیت دارد در حالیکه مقدار ۰.۳ برای یک مدل رگرسیون ساده را در علوم اجتماعی می‌توان خوب دانست و مقدار ۰.۷ در مهندسی ممکن است برازش خوبی محسوب نشود. برای بهبود مدل رگرسیون می‌توان متغیرهای توصیفی بیشتری را افزوده یا از تکنیک‌های متفاوتی برای تبدیل داده‌ها استفاده کرد که منجر به افزایش نسبتاً زیاد مقادیر R^2 می‌شود. شکل ۳.۱۴ جریان فرایند توسعه‌ی مدل‌های رگرسیون را نشان می‌دهد. همانگونه که در جریان فرایندی مشاهده می‌کنید، پس

¹ Overall F-Test

² Root Mean Square Error (RSME)

از توسعه‌ی مدل، وظیفه‌ی ارزیابی مدل انجام می‌شود که در آن نه تنها برازش مدل مورد ارزیابی قرار می‌گیرد بلکه به دلیل مفروضات محدود کننده‌ای که مدل‌های خطی با آنها روبرو هستند، اعتبارسنجی مدل نیز باید به دقت انجام شود.



شکل ۳.۱۴ جریان فرایند توسعه‌ی مدل‌های رگرسیون

اگرچه مدل‌های رگرسیون خطی هنوز اولین انتخاب از بین بسیاری از مدل‌ها برای تحلیل داده‌ها هستند (هم اهداف مدل‌سازی پیشگویانه و هم توصیفی)، ولی مدل‌های رگرسیون خطی به لحاظ چندین فرض بسیار محدود کننده دارای ضعف هستند. اعتبارسنجی مدل خطی ساخته شده به توانایی آن برای انطباق با این مفروضات بستگی دارد. متداول‌ترین مفروضات در اینجا بررسی می‌شود:

۱. **خطی بودن.** این فرض بیان می‌کند که یک ارتباط خطی بین متغیر پاسخ و متغیرهای توصیفی وجود دارد. این بدان معناست که مقدار مورد انتظار متغیر پاسخ، تابعی مستقیم (خط مستقیم) از هر متغیر توصیفی است در حالیکه مقادیر

سایر متغیرهای توصیفی ثابت در نظر گرفته شده است. همچنین، شیب خط نیز به مقادیر سایر متغیرها بستگی ندارد. همچنین اشاره ضمنی به اینکه نکته دارد که اثرات متغیرهای توصیفی مختلف بر مقدار مورد انتظار متغیر پاسخ دارای ماهیت جمع‌پذیر است.

۲. **وابستگی (خطاها).** این فرض بیان می‌کند که خطاهای متغیر پاسخ هیچگونه همبستگی با یکدیگر ندارند. این

وابستگی خطاها ضعیف‌تر از وابستگی آماری واقعی است که اغلب برای تحلیل رگرسیون خطی مورد نیاز است.

۳. **نرمال بودن (خطاها).** این فرض بیان می‌کند که خطاهای متغیر پاسخ دارای توزیع نرمال هستند. این بدان معناست

که فرض می‌کنیم اینها کاملاً تصادفی هستند و هیچگونه الگوی غیر تصادفی را نشان نمی‌دهند.

۴. **واریانس ثابت (خطاها).** این فرض که واریانس همسانی نیز نامیده می‌شود بیان می‌کند که خطاهای متغیرهای پاسخ

صرفنظر از مقدار متغیرهای توصیفی دارای واریانس یکسانی هستند. در عمل، اگر متغیر پاسخ در طول یک

طیف/مقیاس وسیع تغییر کند این فرض عملاً اعتبار ندارد.

۵. **هم خطی چندگانه.** این فرض بیان می‌کند که متغیرهای توصیفی دارای همبستگی نیستند (یعنی دیدگاه‌های متفاوتی

از اطلاعات مورد نیاز برای مدل را تولید می‌کنند). هم خطی چندگانه را با استفاده از دو یا چند متغیر توصیفی که

همبستگی کاملی دارند می‌توان ایجاد کرد (مثلاً اگر متغیر توصیفی یکسانی به اشتباه دو بار در مدل آورده شود و یکی

از آنها نسبت به متغیر دیگر کمی تغییر شکل یافته باشد). از طریق یک ارزشیابی همبستگی داده‌ها می‌توان این خطا

را شناسایی کرد.

تکنیک‌های آماری متعددی برای شناسایی تخطی از این مفروضات و پیشگیری از آنها ایجاد شده است. مهمترین بخش وظیفه‌ی

یک مدل‌ساز این است که از وجود آنها مطلع بوده و وسایل ارزشیابی مدل‌ها را برای حصول اطمینان از انطباق آنها با مفروضات

اصلی مهیا کند.

رگرسیون لجستیک^۱ یک الگوریتم دسته‌بندی آماری بسیار محبوب و مبتنی بر احتمالات است که از **یادگیری نظارتی** استفاده

می‌کند. این روش در سال‌های ۱۹۴۰ به عنوان مکمل روش‌های رگرسیون خطی و تحلیل افتراقی (تشخیصی) خطی توسعه

یافت. از این روش به طور گسترده‌ای در رشته‌های متعدد از جمله علوم پزشکی و اجتماعی استفاده شده است. رگرسیون

لجستیک از این لحاظ که به دنبال یافتن یک تابع ریاضی است که ارتباط بین متغیر پاس و متغیرهای توصیفی را با استفاده از

یک نمونه از مشاهدات قبلی توضیح دهد مشابه رگرسیون خطی است. تفاوت رگرسیون لجستیک با رگرسیون خطی در این است

که خروجی آن به جای یک متغیر عددی، یک دسته یا گروه است. این یعنی اینکه در حالیکه از رگرسیون خطی برای تخمین

یک متغیر عددی پیوسته استفاده می‌شود ولی از رگرسیون لجستیک برای دسته‌بندی یک متغیر رسته‌ای استفاده می‌شود.

اگرچه شکل اولیه‌ی رگرسیون لجستیک برای یک متغیر خروجی دودویی طراحی شد (مثلاً ۱ و ۰، بله/خیر، قبول/مردود) ولی

نسخه‌های اصلاح شده‌ی امروزی قادر به پیش‌بینی چندین دسته متغیر خروجی هستند (یعنی رگرسیون لجستیک چندجمله‌ای).

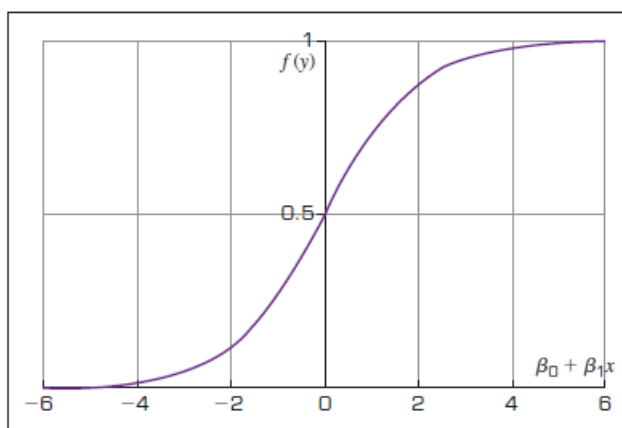
اگر فقط یک متغیر پیشگو و یک متغیر پیش‌بینی شده داشته باشیم، این روش را **رگرسیون لجستیک ساده^۲** می‌گویند (مشابه

مدل‌های رگرسیون خطی تنها با یک متغیر وابسته که رگرسیون خطی ساده نامیده می‌شوند).

¹ Logistic Regression

² Simple Logistic Regression

در تحلیل پیشگویانه، از مدل‌های رگرسیون لجستیک برای ایجاد مدل‌های احتمالی بین یک یا چند متغیر توصیفی/پیشگو (که می‌تواند ترکیبی از متغیرهای پیوسته و رسته‌ای باشد) و یک متغیر پاسخ/دسته (که می‌تواند دوجمله‌ای/دودویی یا چندجمله‌ای/چنددسته‌ای باشد) استفاده می‌شود. برخلاف رگرسیون خطی معمولی، از رگرسیون لجستیک برای پیش‌بینی نتایج رسته‌ای (اغلب دودویی) متغیر پاسخ استفاده می‌شود یعنی با متغیر پاسخ همانند نتیجه‌ی یک آزمون برنولی برخورد می‌شود. بنابراین، رگرسیون لجستیک از لگاریتم طبیعی شانس متغیر پاسخ برای ایجاد یک معیار پیوسته به شکل نسخه‌ی تغییرشکل یافته‌ی متغیر پاسخ استفاده می‌کند. بنابراین، تغییرشکل لاجیت را در رگرسیون لجستیک، تابع پیوند می‌نامند اگرچه متغیر پاسخ در رگرسیون لجستیک دارای ماهیت رسته‌ای یا تک‌جمله‌ است ولی لاجیت^۱ یک معیار پیوسته است که رگرسیون خطی بر روی آن انجام می‌شود. شکل ۳.۱۵ یک تابع رگرسیون لجستیک را نشان می‌دهد که احتمالات بر روی محور x (یک تابع خطی از متغیرهای وابسته) و نتایج احتمالات بر روی محور y (یعنی مقادیر متغیر پاسخ بین ۰ و ۱ تغییر می‌کنند) نشان داده شده است.



شکل ۳.۱۵ تابع لجستیک

تابع لجستیک که با $f(y)$ در شکل ۳.۱۵ نشان داده شد هسته‌ی اصلی رگرسیون لجستیک است که می‌تواند مقادیر بین ۰ و ۱ به خود بگیرد. معادله‌ی زیر یک نمایش ریاضی ساده از این تابع است:

$$f(y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

ضرایب رگرسیون لجستیک (β ها) معمولاً با استفاده از روش تخمین بیشینه احتمال تخمین شده می‌شود. برخلاف رگرسیون خطی که باقیمانده‌های آن دارای توزیع نرمال هستند، نمی‌توان یک عبارت بسته برای مقادیر ضرایبی که بیشینه احتمال تابع را نتیجه می‌دهند پیدا کرد و به همین دلیل به جای آن باید از یک فرایند تکرارشونده استفاده کرد. این فرایند با یک خیر آغازین آزمایشی شروع شده و سپس پارامترها را به صورت جزئی بازبینی می‌کند تا ببیند که آیا جواب، بهبود یافته است یا خیر

^۱ Logit

ساختار و نتایج رویدادهای ورزشی به شکل فاکتورهای عددی یا نشانه‌ای متنوع که فرض می‌شود بر نتایج رویدادها تأثیر داشته باشند در دسترس عموم قرار دارد.

بازی‌های پایان فصل چه از لحاظ مالی (درآمد میلیون دلاری) و چه از لحاظ شهرت (بکارگیری دانشجویان با کیفیت و ورزشکاران تراز اول دبیرستانی برای برنامه‌های ورزشی) برای کالج‌ها بسیار حائز اهمیت هستند (فریمن آند برویر ۲۰۱۶).^۱ تیم‌هایی که برای رقابت در یک بازی سوپر بول انتخاب می‌شوند مبلغ جایزه را تقسیم می‌کنند که اندازه‌ی آن به خودِ مسابقه (برخی مسابقات از اعتبار بالاتری برخوردارند و هر دو تیم باید سهم بیشتری را پرداخت کنند) بستگی داشته و حصول اطمینان از دعوت به بازی سوپر بول هدف اصلی هر برنامه فوتبال کالج بخش I-A می‌باشد. تصمیم‌گیرندگان بازی‌های سوپر بول این حق انتخاب را دارند تا تیم‌های موفق‌تری را که بازی‌های هیجان‌انگیز و رقابتی را برگزار خواهند کرد و واجد شرایط مسابقه هستند (تیمی که شش بُرد در مقابل رقبای بخش I-A در آن فصل داشته) برگزیده و دعوت کنند، طرفداران هر دو طرف را جذب کرده و مابقی طرفداران را از طریق تبلیغات و رسانه‌های مختلف مطلع نگهدارند.

در یک مطالعه‌ی داده‌کاوی اخیر، دیلن و همکاران (۲۰۱۲) از داده‌های هشت سال بازی سوپر بول از سه تکنیک معروف داده‌کاوی (درخت تصمیم، شبکه‌های عصبی و ماشین‌های بردار پشتیبان) برای پیش‌بینی نتایج بازی از نوع دسته‌بندی (برد یا باخت) و رگرسیون (اختلاف بین امتیازهای دو حریف) استفاده کرده‌اند. در ادامه شرح مختصری از مطالعه آنها ارائه شده است.

متدولوژی

در این پژوهش، دیلن و همکارانش از یک متدولوژی داده‌کاوی معروف به نام کریسپ (فرایند استاندارد صنعتی متقاطع برای داده‌کاوی)^۲ که یک فرایند شش مرحله‌ای است استفاده کرده‌اند. این متدولوژی محبوب که در فصل ۴ به آن پرداخته‌ایم روشی سیستماتیک و ساختارمند را برای انجام مطالعه‌ی داده‌کاوی اصلی در اختیار آنها قرار داده و احتمال دستیابی به نتایج قابل اطمینان و دقیق را افزایش بخشیده است. برای ارزشیابی عینی قدرت پیش‌بینی مدل‌های مختلف، آنها از یک متدولوژی اعتبارسنجی متقاطع k برابری استفاده کردند. جزئیات اعتبارسنجی متقاطع k برابری در فصل ۴ ارائه شده است. شکل ۳.۱۶ نمایش گرافیکی متدولوژی بکار رفته توسط پژوهشگران را نشان می‌دهد.

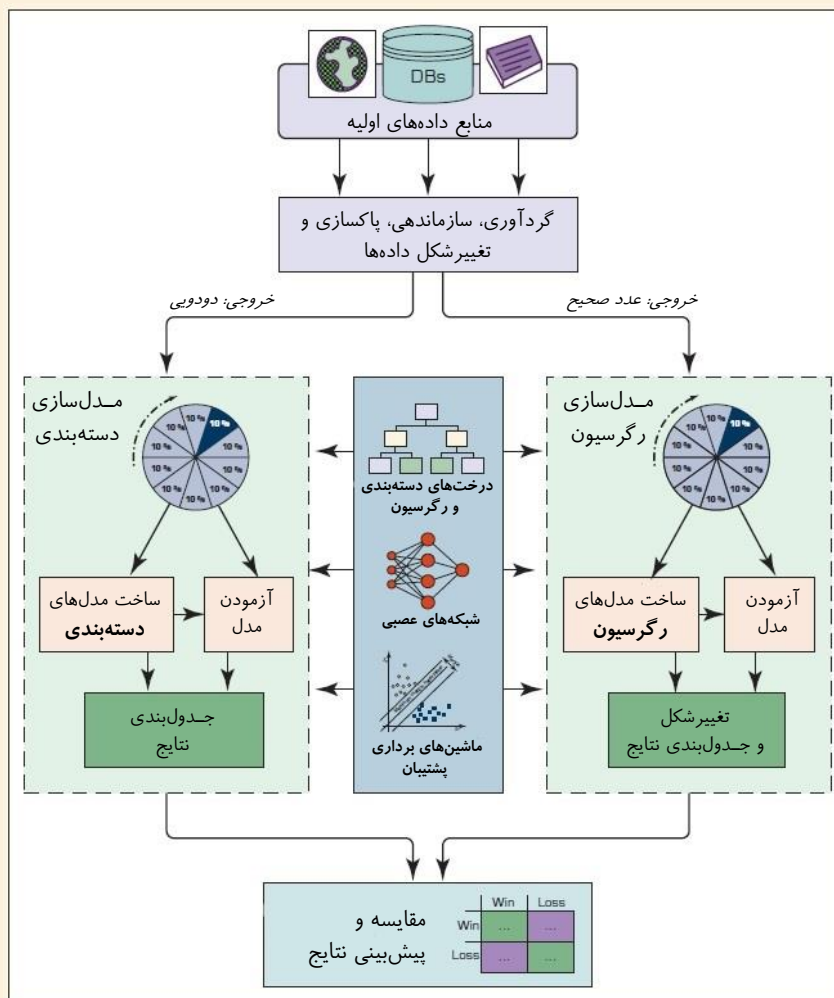
اخذ داده‌ها و پیش‌پردازش داده‌ها

داده‌های نمونه برای این مطالعه از پایگاه‌های داده ورزشی مختلف بر روی وب گردآوری شده‌اند از جمله **jhowel.net, ESPN.com, Covers.com, ncaa.org** و **rauzulusstreet.com**. مجموعه داده‌ها شامل ۲۴۴ بازی بود که هشت فصل کامل از بازی‌های سوپر بول فوتبال بین ۲۰۰۲ و ۲۰۰۹ را شامل می‌شد. دیلن و همکاران همچنین یک مجموعه داده خارج از نمونه را برای اعتبارسنجی بیشتر در نظر گرفتند (بازی‌های سال‌های ۲۰۱۰ و ۲۰۱۱). آنها با یک داده‌کاوی سرانگشتی، مرتبط‌ترین اطلاعات را تا حد امکان در مدل گنجانده‌اند. بنابراین، پس از یک شناسایی عمیق متغیر و فرایند گردآوری، یک مجموعه داده مشتمل بر ۳۶ متغیر بدست آمد که ۶ متغیر اول، متغیرهای شناسایی بودند (یعنی نام و سال بازی، نام تیم‌های میزبان و مهمان و کنفرانس‌های ورزشی آنها که در جدول ۳.۵ آمده است) و پس از آن ۲۸ متغیر ورودی (شامل متغیرهای تعیین آمارهای فصلی تیم در حمله و دفاع، نتایج بازی، ویژگی‌های ترکیب تیم، ویژگی‌های کنفرانس ورزشی و نحوه‌ی مواجهه‌ی آنها با احتمالات است - متغیرهای ۷ تا ۳۴ در جدول ۳.۵) و نهایتاً دو

¹ Freeman & Brewer

² CRISP-DM (Cross-Industry Standard Process for Data Mining)

متغیر آخر که متغیرهای خروجی هستند (یعنی ScoreDiff که اختلاف امتیاز بین تیم میزبان و تیم مهمان بر حسب یک عدد صحیح است و WinLoss که نشان می‌دهد آیا تیم مهمان بازی سوپر بول را برده یا باخته است که به صورت یک برچسب اسمی بیان می‌شود).



شکل ۳.۱۶ نمایش گرافیکی متدولوژی بکار رفته در مطالعه

جدول ۳.۵ شرح متغیرهای بکار رفته در مطالعه

ردیف	رسته	نام متغیر	شرح
۱	ID	YEAR	سالی که بازی سوپر بول در آن انجام شده است.
۲	ID	BOWLGAME	نام بازی سوپر بول
۳	ID	HOMETEAM	تیم میزبان (بر اساس فهرست‌بندی سازمان‌دهندگان مسابقه)
۴	ID	AWAYTEAM	تیم مهمان (بر اساس فهرست‌بندی سازمان‌دهندگان مسابقه)

کنفرانس تیم میزبان	HOMECONFERENCE	ID	۵
کنفرانس تیم مهمان	AWAYCONFERENCE	ID	۶
نقاط دفاعی به ازای هر بازی	DEFPTPGM	I1	۷
یارد راش‌های دفاعی به ازای هر بازی	DEFRYDPGM	I1	۸
یاردهای دفاعی به ازای هر بازی	DEFYDPGM	I1	۹
متوسط تعداد امتیازهای تیم در یک بازی	PPG	I1	۱۰
میانگین کل یاردهای پاس شده به ازای هر بازی	PYDPGM	I1	۱۱
متوسط کل یاردهای راش شده به ازای هر بازی	RYDPGM	I1	۱۲
میانگین کل یاردهای دفاعی هر بازی	YRDPGM	I1	۱۳
درصد برنده شدن در خانه	HMWIN%	I2	۱۴
تیم از بین ۷ بازی آخر خود چه تعداد را برنده شده است	LAST7	I2	۱۵
حاشیه میانگین پیروز شدن	MARGOVIC	I2	۱۶
درصد برنده شده تیم‌هایی که کنفرانس نداشته‌اند.	NCTW	I2	۱۷
آیا تیم در سال گذشته در مسابقات سوپر بول شرکت داشته	PREVAPP	I2	۱۸
درصد برنده شدن در بیرون از خانه	RDWIN%	I2	۱۹
درصد برنده شدن در سال	SEASTW	I2	۲۰
درصد برنده شدن در برابر ۲۵ تیم برتر در سال	TOP25	I2	۲۱
دشواری جداول زمانی مسابقات در سال	TSOS	I3	۲۲
درصد بازی‌های انجام شده توسط بازیکنان گروه سال اولی	FR%	I3	۲۳
درصد بازی‌های انجام شده توسط بازیکنان گروه سال دوم	SO%	I3	۲۴
درصد بازی‌های انجام شده توسط بازیکنان گروه نوجوانان	JR%	I3	۲۵
درصد بازی‌های انجام شده توسط بازیکنان گروه جوانان	SR%	I4	۲۶
درصد دفعاتی که یک تیم در فصل جاری از O/U^3 گذشته	SEASOVUn%	I4	۲۷
درصد پوشش‌دهی در برابر گسترش تیم در بازی‌های قبل	ATSCOV%	I4	۲۸
درصد دفعاتی که یک تیم در بازی‌های قبلی سقوط کرده	UNDER%	I4	۲۹
درصد دفعاتی که یک تیم در بازی‌های قبلی صعود کرده	OVER%	I4	۳۰
درصد پوش در برابر گسترش در فصل جاری	SEASATS%	I4	۳۱
آیا تیم در بازی قهرمانی کنفرانس مربوطه‌اش برنده شده	CONCH	I5	۳۲
دشواری جداول زمانبندی کنفرانس‌ها	CONFOS	I5	۳۳
درصد برنده شدن کنفرانس	CONFWIN%	I5	۳۴
اختلاف امتیاز (HomeTeamScore – AwayTeamScore)	ScoreDiff	O1	۳۵
آیا تیم میزبان در بازی برنده شده یا خیر	WinLoss	O2	۳۶

ID: متغیرهای شناساگر؛ O1: متغیر خروجی مدل‌های رگرسیون؛ O2: متغیر خروجی مدل‌های دسته‌بندی.

I1: دفاع/حمله. I2: نتیجه‌ی بازی؛ I3: ترکیب تیم؛ I4: در عین ناباوری؛ I5: آمار کنفرانس

بالا/زیر - آیا تیم توانسته است از اختلاف امتیاز مورد انتظار بالاتر برود یا خیر.
متغیرهای خروجی - ScoreDiff برای مدل‌های رگرسیونی و WinLoss برای مدل‌های دسته‌بندی دودویی.

در فرمولاسیون مجموعه داده‌ها هر ردیف بیانگر یک بازی سوپر باول است و هر ستون بیانگر یک متغیر (یعنی شناساگر/ورودی یا خروجی). برای نمایش مشخصه‌های مقایسه‌ای دو تیم حریف در متغیرهای ورودی، دیلن و همکاران اختلاف بین سنجه‌های تیم‌های میزبان و مهمان را محاسبه کرده و بکار گرفتند. تمامی این مقادیر متغیرها از منظر تیم میزبان محاسبه شده‌اند. به عنوان مثال، متغیر PPG (میانگین امتیازهایی که یک تیم در هر بازی بدست آورده است) بیانگر اختلاف بین PPG تیم میزبان و PPG تیم مهمان است. متغیرهای خروجی بیانگر این هستند که آیا تیم میزبان در مسابقه‌ی سوپر باول برنده شده است یا خیر. این یعنی اینکه اگر متغیر ScoreDiff یک عدد صحیح مثبت باشد آنگاه انتظار می‌رود که تیم میزبان در مسابقه برنده شود. در غیراینصورت، (اگر متغیر ScoreDiff یک عدد صحیح منفی باشد)، تیم میزبان انتظار می‌رود که در بازی بازنده باشد. در خصوص متغیر WinLoss، مقدار متغیر خروجی یک برچسب دودویی است، "Win" یا "Loss" نشان دهنده‌ی نتیجه‌ی بازی برای تیم میزبان است.

نتایج و ارزیابی

در این مطالعه، از تکنیک‌های متداول پیش‌بینی برای ساخت مدل‌ها استفاده می‌شود (و مقایسه‌ی آنها با یکدیگر): شبکه‌های عصبی مصنوعی، درخت تصمیم و ماشین‌های بردار پشتیبان. این تکنیک‌های پیش‌بینی بر اساس قابلیت مدل‌سازی مسائل پیش‌بینی دسته‌بندی و رگرسیون و محبوبیت آنها در ادبیات جدید منتشر در خصوص داده‌کاوی برگزیده می‌شوند. برای جزئیات بیشتر در مورد این روش‌های متداول داده‌کاوی به فصل ۴ مراجعه کنید.

برای مقایسه‌ی دقت پیش‌بینی تمامی این مدل‌ها با یکدیگر، پژوهشگران از یک متدولوژی اعتبارسنجی متقاطع k برابری طبقه‌بندی شده استفاده کرده‌اند. در نسخه‌ی طبقه‌بندی شده از این متدولوژی، دفعات تکرار به گونه‌ای انجام می‌شوند که حاوی تقریباً نسبت یکسانی از برچسب‌های پیشگو (یعنی دسته‌ها) که در مجموعه داده اصلی است باشند. در این مطالعه، مقدار k را برابر ۱۰ گرفته‌ایم (یعنی کل مجموعه‌ی ۲۴۴ نمونه به ۱۰ زیرمجموعه با تقریباً ۲۵ نمونه تقسیم می‌شود) که روشی متداول در کاربردهای داده‌کاوی پیشگویانه است. یک نمایش گرافیکی از اعتبارسنجی متقاطع ۱۰ برابری قبلاً در این فصل نشان داده شد. برای مقایسه‌ی مدل‌های پیش‌بینی که با استفاده از سه تکنیک داده‌کاوی مذکور ساخته شده‌اند، پژوهشگران تصمیم گرفتند تا از سه معیار عملکردی متداول استفاده کنند: دقت، حساسیت و خاص بودن. فرمول ساده‌ی این سنجه‌ها نیز قبلاً در این فصل ارائه شد.

جدول ۳.۶ نتایج پیش‌بینی برای متدولوژی دسته‌بندی مستقیم

روش پیش‌بینی (دسته‌بندی)	ماتریس تلفیق		دقت (%)	حساسیت (%)	خاص بودن (%)
	Win	Loss			
ANN (MPL)	92	42	75.00	68.66	82.73
	19	91			
SVM (RBF)	105	29	79.51	78.36	80.91
	21	89			
DT (C&RT)	113	21	86.48	84.33	89.09

متغیر خروجی یک متغیر رسته‌ای دودویی است.
تفاوت‌ها معنادار بوده‌اند.

نتایج پیش‌بینی این سه تکنیک مدل‌سازی در جداول ۳.۶ و ۳.۷ ارائه شده‌اند. جدول ۳.۶ نتایج اعتبارسنجی متقابل ۱۰ برابری متدولوژی دسته‌بندی را نشان می‌دهد که در آن، سه تکنیک داده‌کاوی به گونه‌ای تدوین شده‌اند که یک متغیر خروجی دودویی-اسمی داشته باشند (یعنی *WinLoss*). جدول ۳.۷ نتایج اعتبارسنجی متقاطع ۱۰ برابری متدولوژی دسته‌بندی رگرسیونی را که در آن سه تکنیک داده‌کاوی به نحوی فرموله شده‌اند که یک متغیر خروجی عددی داشته باشند (یعنی *ScoreDiff*). در پیش‌بینی دسته‌بندی رگرسیونی، خروجی عددی مدل‌ها از طریق برچسب‌گذاری اعداد مثبت *WinLoss* با "Win" و اعداد منفی *WinLoss* با "Loss" به نوع رسته‌بندی تبدیل شده و سپس در قالب ماتریس‌های تلفیق به صورت جدول‌بندی درآمده‌اند. با استفاده از ماتریس‌های تلفیق، دقت، حساسیت و خاص بودن پیش‌بینی برای هر نوع مدل محاسبه و در جداول ۳.۶ و ۳.۷ آورده شده است. همانطور که نتایج نشان می‌دهد، روش‌های پیش‌بینی از نوع دسته‌بندی عملکرد بهتری نسبت به متدولوژی پیش‌بینی دسته‌بندی رگرسیونی دارند. از میان سه متدولوژی داده‌کاوی، درخت‌های دسته‌بندی و رگرسیون از دقت پیش‌بینی بهتری در هر دو متدولوژی پیش‌بینی برخوردار بودند. در مجموع، دقت اعتبارسنجی ۱۰ برابری برای مدل‌های دسته‌بندی درخت رگرسیون و دسته‌بندی معادل ۸۶.۴۸ درصد بوده است که پس از آن ماشین‌های بردار پشتیبان (۷۹.۵۱ درصد) و شبکه‌های عصبی (۷۵.۰۰) قرار داشته‌اند. پژوهشگران با استفاده از یک آزمون تی پی بردند که این مقادیر دقت با آلفای ۰.۰۵ تفاوت معناداری با یکدیگر دارند بدین معنی که درخت تصمیم نسبت به شبکه عصبی و ماشین برداری پشتیبان از عملکرد به مراتب بهتری برخوردار است و ماشین بردار پشتیبان نسبت به شبکه‌های عصبی عملکرد بسیار بهتری دارد.

جدول ۳.۷ نتایج پیش‌بینی برای متدولوژی دسته‌بندی رگرسیونی

روش پیش‌بینی (دسته‌بندی)	ماتریس تلفیق		دقت (%)	حساسیت (%)	خاص بودن (%)
	Win	Loss			
ANN (MPL)	94	40	72.54	70.15	74.45
	27	83			
SVM (RBF)	100	34	74.59	74.63	74.55
	28	82			
DT (C&RT)	106	28	77.87	76.36	79.10
	26	84			

متغیر خروجی یک متغیر عددی/عدد صحیح است.
تفاوت‌ها معنادار بوده‌اند $p < 0.01$.

نتایج این مطالعه نشان داده است که مدل‌های از نوع دسته‌بندی بهتر از مدل‌های دسته‌بندی رگرسیونی قادر به پیش‌بینی نتایج بازی هستند. اگرچه این نتایج، خاص حوزه کاربردی و داده‌های بکار رفته در این مطالعه هستند و به همین دلیل نباید آنها را به غیر از

دامنه‌ی این مطالعه تعمیم داد ولی بسیار هیجان‌انگیز هستند زیرا درخت تصمیم، نه تنها بهترین پیشگوها بلکه بهترین روش‌ها برای درک و اجرا در مقایسه با دو تکنیک دیگر یادگیری ماشین هستند که در این مطالعه بکار گرفته شده‌اند. برای جزئیات بیشتر در مورد این مطالعه به دیلن و همکاران (۲۰۱۲) مراجعه کنید.

Sources: D. Delen, D. Cogdell, and N. Kasap, "A Comparative Analysis of Data Mining Methods in Predicting NCAA Bowl Outcomes," *International Journal of Forecasting*, 28, 2012, pp. 543-552; K. M. Freeman, and R. M. Brewer, "The Politics of American College Football," *Journal of Applied Business and Economics*, 18(2), 2016, pp. 97-101.

سوالاتی برای مورد کاربرد ۳.۴

۱. چالش‌های محتمل در پیش‌بینی نتایج رویدادهای ورزشی چه هستند (مثلاً بازی‌های سوپر باول دانشگاهی)؟
۲. پژوهشگران چگونه مسئله‌ی پیش‌بینی را تدوین/طراحی کردند (یعنی ورودی‌ها و خروجی‌ها چه بودند و نمایش یک نمونه‌ی ساده - ردیف داده‌ها به چه شکل بوده است)؟
۳. نتایج پیش‌بینی تا چه اندازه موفق بوده‌اند؟ برای بهبود دقت آنها چه کار دیگری می‌توان انجام داد؟

پیش‌بینی سری‌های زمانی یعنی استفاده از مدل‌سازی ریاضی برای پیش‌بینی مقادیر آینده برای یک متغیر مورد نظر بر اساس مقادیر مشاهدات قبلی. نمودارهای سری‌های زمانی بسیار شبیه رگرسیون خطی ساده هستند زیرا همانند آنها ما دو متغیر داریم: متغیر پاسخ و متغیر زمان که در نمودار پراکندگی نمایش داده می‌شود. به جز این تشابه ظاهری به سختی می‌توان تشابه یا ارتباط دیگری بین این دو پیدا کرد. اگرچه تحلیل رگرسیون اغلب در نظریه‌های آزمون و بررسی اینکه آیا مقادیر جاری یک یا چند متغیر توصیفی، متغیر پاسخ را توضیح می‌دهند یا خیر بکار می‌روند، مدل‌های سری زمانی بر برون‌یابی مبتنی بر رفتار تغییرات زمانی برای تخمین مقادیر آینده تمرکز دارند.

در پیش‌بینی بر اساس سری‌های زمانی فرض بر این است که تمامی متغیرهای توصیفی به عنوان یک رفتار متغیر زمانی در قالب یک متغیر پاسخ جمع شده‌اند. از این رو، ثبت رفتار متغیر زمانی، روشی برای پیش‌بینی مقادیر آینده‌ی متغیر پاسخ است. برای اینکار، این الگو تحلیل شده و به مؤلفه‌های اصلی خود تجزیه می‌شود: نوسانات و تغییرات تصادفی، روندهای زمانی و چرخه‌های فصلی. یک مثال از سری‌های زمانی در شکل ۳.۱۷ تمامی این الگوهای متمایز را به تصویر کشیده است.

تکنیک‌های بکار رفته برای پیش‌بینی‌های سری‌های زمانی از تکنیک‌های بسیار ساده (امروز مثل دیروز است) تا بسیار پیچیده مثل آریما^۱ (روشی که الگوهای میانگین متحرک و خودهمبسته را در داده‌ها ترکیب می‌کند) متغیر هستند. متداول‌ترین روش‌ها شاید روش‌های میانگین‌گیری باشند که شامل میانگین ساده، میانگین متحرک، میانگین متحرک موزون و هموارسازی نمایی هستند. بسیاری از این تکنیک‌ها دارای نسخه‌های پیشرفته‌تر نیز هستند که برای در نظر گرفتن ویژگی‌های فصلی بودن و روندها می‌توانند پیش‌بینی‌های دقیق‌تری را ارائه دهند. دقت یک روش معمولاً بر اساس محاسبه‌ی خطای آن روش ارزیابی می‌شود (انحراف محاسبه شده بین مقادیر واقعی و پیش‌بینی شده برای مشاهدات گذشته) که اینکار از طریق میانگین خطای مطلق، خطای میانگین مربعات یا درصد میانگین مطلق خطا انجام می‌شود. اگرچه تمامی آنها از یک سنج‌های خطای اصلی استفاده می‌کنند ولی این سه روش ارزیابی بر جنبه‌های متفاوتی از خطا تأکید دارند که در برخی از آنها جریمه‌ی بیشتری به خطاهای بزرگ اختصاص می‌یابد.

¹ ARIMA



شکل ۳.۱۷ یک نمونه سری زمانی از داده‌ها برای حجم فروش فصلی

سوالاتی برای مرور بخش ۳.۶

۱. رگرسیون چیست و به لحاظ آماری چه هدفی را دنبال می‌کند؟
۲. تشابهات و تفاوت‌های بین رگرسیون و همبستگی چه هستند؟
۳. حداقل مربعات معمولی چیست و چگونه خط رگرسیون خطی را تعیین می‌کند؟
۴. گام‌های اصلی ایجاد یک مدل رگرسیون خطی را نام برده و توضیح دهید.
۵. متداول‌ترین مفروضات برای رگرسیون خطی چه هستند؟
۶. رگرسیون لجستیک را تعریف کنید. تفاوت آن با رگرسیون خطی در چیست؟
۷. سری‌های زمانی چه هستند؟ تکنیک‌های اصلی پیش‌بینی برای داده‌های سری‌های زمانی چه هستند؟

تصمیم‌گیرندگان برای اتخاذ تصمیمات دقیق و به موقع به اطلاعات نیاز دارند. اطلاعات در اصل همان داده‌ها هستند که در یک بافتار قرار می‌گیرند. علاوه بر روش‌های آماری که در بخش قبل توضیح داده شدند، اطلاعات (تحلیل توصیفی) را می‌توان با استفاده از سیستم‌های اِل تی پی^۱ نیز بدست آورد. اطلاعات معمولاً به شکل یک گزارش مکتوب (کاغذی یا دیجیتالی) در اختیار تصمیم‌گیرندگان قرار می‌گیرد. اگرچه ممکن است به صورت شفاهی نیز بیان شود. به بیان ساده، یک گزارش^۲ به هر نوع مصنوع مکاتباتی که با هدف خاص ارائه اطلاعات به شکلی قابل هضم برای هر کسی که نیازمند آن است در هر زمان و هر کجا

¹ OLTP

² Report

گفته می‌شود. گزارش معمولاً به شکل یک مستند است که حاوی اطلاعات (معمولاً گرفته شده از داده‌ها) سازماندهی شده به شکل گفتار، گرافیک و/یا جدولی است که به صورت دوره‌ای (تکرار شونده) یا در صورت نیاز تهیه شده و به دوره‌های زمانی، رویدادها یا موضوعات خاص اشاره دارد. گزارشات کسب‌وکار می‌توانند کارکردهای مختلفی (اما مرتبط) داشته باشند که در اینجا به ذکر متداول‌ترین آنها می‌پردازیم:

حصول اطمینان از اینکه تمامی واحدهای سازمان به خوبی عمل می‌کنند.

- ارائه‌ی اطلاعات
- ارائه‌ی نتایج یک تحلیل
- متقاعد کردن دیگران برای اقدام کردن.
- ایجاد یک حافظه‌ی سازمانی (به عنوان بخشی از سیستم مدیریت دانش)

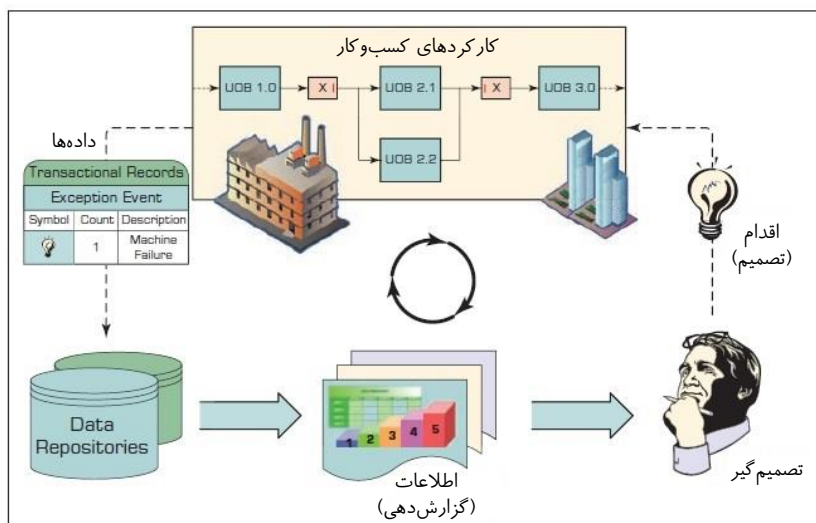
گزارش‌دهی کسب‌وکار (که با نام OLAP یا هوش کسب‌وکار نیز شناخته می‌شود)، بخشی ضروری از یک محرک بزرگتر در جهت تصمیم‌گیری مدیریتی بهینه و مبتنی بر مدرک است. مبنای این گزارشات کسب‌وکار^۱، منابع مختلف داده‌هایی است که از داخل و خارج سازمان می‌آیند. ایجاد این گزارشات شامل استخراج، تغییرشکل و بارگذاری رویه‌ها هماهنگ با یک انبار داده و سپس استفاده از یک یا چند ابزار گزارش‌دهی است.

به دلیل گسترش سریع فناوری اطلاعات به همراه نیاز به بهبود رقابت‌پذیری در کسب‌وکار، استفاده از توان رایانشی برای تولید گزارشات متحدالشکل که دیدگاه‌های مختلف سازمان را در یکجا گرد هم آورند نیز افزایش یافته است. معمولاً، این فرایند گزارش‌گیری شامل پرس و جو از منابع داده ساختیافته که اکثر آنها با استفاده از مدل‌های داده‌ای منطقی مختلف و دیکشنری‌های داده‌ها ایجاد شده‌اند گزارشات قابل خواندن توسط انسان را تولید می‌کنند. این نوع گزارشات کسب‌وکار به مدیران و همکاران اجازه می‌دهد تا مطلع و آگاه بوده، گزینه‌های مختلف را بازبینی کرده و تصمیمات آگاهانه بگیرند. شکل ۳.۱۸ یک چرخه‌ی پیوسته از اخذ/تولید/تصمیم‌گیری/مدیریت فرایند کسب‌وکار را نشان می‌دهد. شاید بحرانی‌ترین وظیفه در این فرایند چرخه‌ای، گزارش‌دهی باشد (یعنی تولید اطلاعات). تبدیل داده‌ها از منابع مختلف به شکل اطلاعات قابل اقدام.

کلید هرگونه گزارش موفق شامل شفافیت، کامل بودن و صحیح بودن است. ماهیت گزارش و سطح اهمیت این عوامل موفقیت به میزان زیادی بسته به شخصی که این گزارش برای وی تهیه می‌شود تغییر می‌کند. اکثر پژوهش‌ها در گزارش‌دهی مؤثر مختص گزارشات داخلی برای ذینفعان و تصمیم‌گیرندگان درون سازمانی است. همچنین گزارشات برون‌سازمانی بین سازمان‌ها و دولت وجود دارد (مثلاً برای موضوعات مالیاتی یا فایلینگ منظم بورس اوراق بهادار). اگرچه گزارشات کسب‌وکار بسیار متنوعی وجود دارد، آنهایی که اغلب برای مقاصد مدیریتی بکار می‌روند را می‌توان در سه گروه اصلی دسته‌بندی کرد (هیل ۲۰۱۶)^۲.

¹ Business Reports

² Hill



شکل ۳.۱۸ نقش گزارش‌دهی اطلاعات در تصمیم‌گیری مدیریتی

مصورسازی داده‌ها^۱ (یا به بیان بهتر، مصورسازی اطلاعات^۲) به این شکل تعریف شده است: «استفاده از بازنمودهای دیداری برای اکتشاف، معنادار کردن و تسری داده‌ها». (فیو ۲۰۰۷). اگرچه عمدتاً از عبارت مصورسازی داده‌ها استفاده می‌شود ولی معمولاً منظور از این کار، مصورسازی اطلاعات می‌باشد. از آنجائیکه اطلاعات، حالت تجمیع، خلاصه و بافتی کردن داده‌ها می‌باشد، آنچه در مصورسازی به نمایش گذاشته می‌شود، اطلاعات هستند و نه داده‌ها. با وجود این، چونکه دو عبارت مصورسازی داده‌ها و مصورسازی اطلاعات به جای یکدیگر و به عنوان مترادف یکدیگر استفاده می‌شوند، ما نیز در این فصل به شکل زیر ادامه خواهیم داد.

مصورسازی داده‌ها رابطه‌ی نزدیکی با نمایش گرافیکی اطلاعات، مصورسازی اطلاعات، مصورسازی علمی و گرافیک‌های آماری دارد. تا همین اواخر، شکل‌های اصلی مصورسازی داده‌ها که در هر دو کاربرد هوش کسب و کار وجود داشته‌اند شامل نمودارها و گراف‌ها و همچنین سایر انواع اجزای دیداری بودند که برای ایجاد داشبوردها و کارت‌های امتیازی بکار می‌رفتند. برای درک بهتر روندهای جاری و آینده در رشته‌ی مصورسازی داده‌ها بهتر است تا با بررسی بافتار تاریخی آنها شروع کنیم.

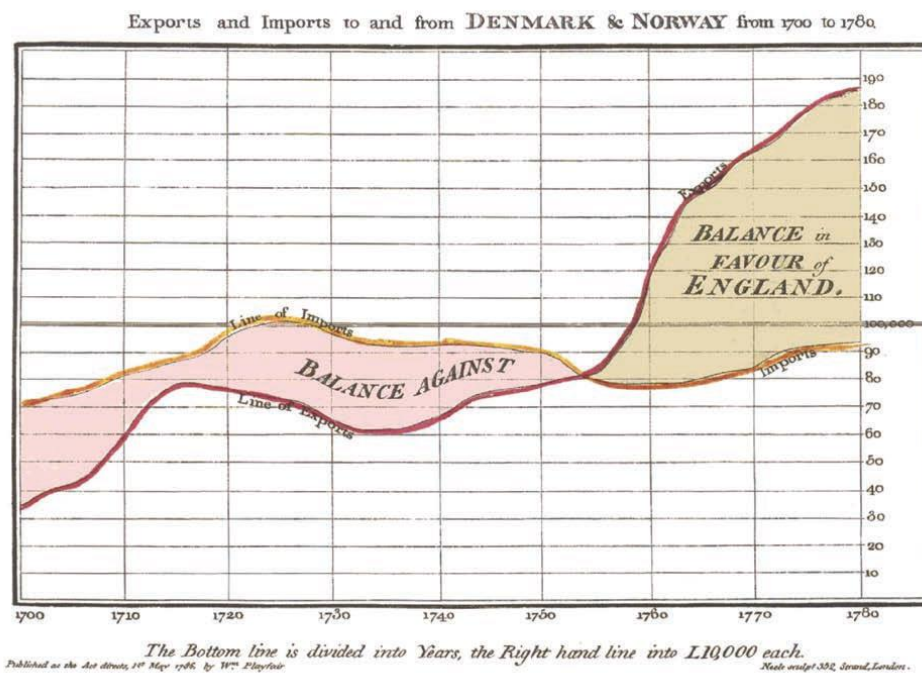
علیرغم این واقعیت که سابقه‌ی مصورسازی داده‌ها به قرن دوم پیش از میلاد می‌رسد، اکثر دستاوردها در این حوزه در ۲۵۰ سال آخر و عمدتاً ۳۰ سال گذشته رخ داده‌اند (فیو ۲۰۰۷). اگرچه مصورسازی تا این اواخر به طور گسترده‌ای به عنوان یک رشته شناخته نشده است ولی محبوب‌ترین شکل‌های دیداری امروزی دارای قدمت چند صد ساله هستند. اکتشافات جغرافیایی، ریاضیات و رواج تاریخ باعث خلق نقشه‌ها، گراف‌ها و گاه‌شمارها (خطوط زمانی) از قرن ۱۷ به بعد شده است ولی ویلیام پلی‌فر^۳

¹ Data Visualization

² Information Visualization

³ William Playfair

را عمدتاً به عنوان مخترع نمودارهای پیشرفته و امروزی می‌شناسند که نمودارهای میله‌ای و خطی توزیعی را برای اولین بار در *اطلس سیاسی و تجاری سال ۱۷۸۶*^۱ خود منتشر نمود. همچنین اولین نمودار خطی سری‌های زمانی را در کتاب *خلاصه آماری*^۲ خود در سال ۱۸۰۱ منتشر نمود (شکل ۳.۱۹). شاید بتوان شارل ژوزف مینار^۳ را برجسته‌ترین مبدع گرافیک‌های اطلاعاتی در این دوره دانست که خسارت‌های وارده بر ارتش ناپلئون در جبهه روسیه در سال ۱۸۱۲ را به صورت گرافیکی به تصویر کشید (شکل ۳.۲۰). با شروع از مرز لهستان و روسیه، اندازه‌ی ارتش در هر موقعیت با یک نوار ضخیم نشان داده شده است. مسیر عقب‌نشینی ارتش ناپلئون از مسکو در زمستان سرد و طاقت‌فرسا با یک نوار تیره در قسمت پایین‌تر این نوار نمایش داده شده که وابسته به مقیاس‌های زمانی و دما است. ادوارد تافت^۴، منتقد، نویسنده و کارشناس معروف مصورسازی می‌گویند که این «می‌تواند بهترین گرافیک آماری باشد که تا به حال ترسیم شده است.» در این گرافیک، مینار توانسته است تا چندین بُعد داده‌ای از جمله اندازه ارتش، جهت حرکت، موقعیت‌های جغرافیایی، دمای بیرون و ... را به طور همزمان به نحوی هنرمندانه و آموزنده نمایش دهد. تعداد بسیار زیادی از مصورسازی‌های علی در قرن ۱۹ ایجاد شدند و اکثر آنها در وب‌سایت تافت (edwardtufte.com) و کتاب‌های مصورسازی وی آورده شده‌اند.



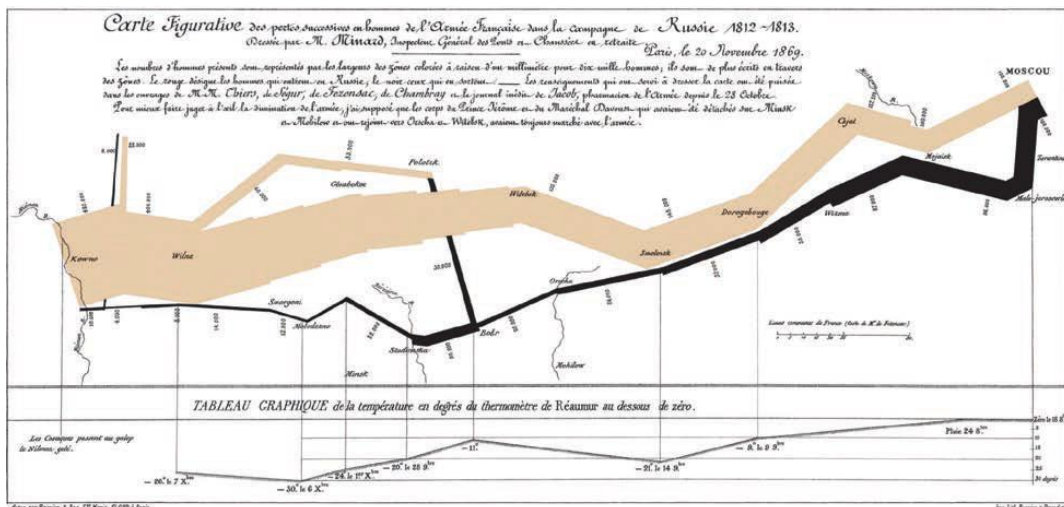
شکل ۳.۱۹ اولین نمودار خطی سری زمانی تهیه شده توسط ویلیام پلی‌فر

¹ Commercial and Political Atlas 1786

² Statistical Breviary

³ Charles Joseph Minard

⁴ Edward Tufte



شکل ۳۰۲۰ قلع و قمع ارتش نپلئون در جبهه روسیه در سال ۱۸۱۲

قرن ۲۰ شاهد ظهور نگرشی رسمی تر و تجربی‌ها نسبت به مصورسازی بود که بیشتر بر جنبه‌هایی از قبیل رنگ، مقادیر ارزشی و نام‌گذاری متمرکز بودند. در اواسط قرن ۲۰، ژاک برتن^۱، نقشه‌نگار و نظریه‌پرداز، نشانه‌شناسی گرافیکی^۲ خود را منتشر کرد که به اعتقاد برخی، زیربنای نظری مصورسازی اطلاعات امروزی است. اگرچه اکثر الگوهایش تاکنون منسوخ شده‌اند یا اینکه در رسانه‌های دیجیتالی به هیچ عنوان قابل اجرا نیستند ولی بسیاری از آنها همچنان کاربرد دارند.

در قرن ۲۱، اینترنت به عنوان وسیله‌ای جدید برای مصورسازی پدیدار شد و قابلیت‌ها و ترفندهای جدیدی را با خود به همراه آورد. توزیع دیجیتالی مصورسازی داده‌ها و اطلاعات نه تنها موجب شده تا مخاطبان گسترده‌تری در سرتاسر جهان به آنها دسترسی یابند (افزایش سطح سواد دیداری در کنار آن) بلکه طراحی شکل‌های جدیدی را رواج داده است که شامل تعامل، پویانمایی و فناوری پردازش گرافیکی منحصربفرد صفحات نمایش و داده‌های بلادرنگ منجر به خلق محیط‌هایی سه بعدی برای ارسال و استفاده از داده‌ها می‌باشد.

افراد و شرکت‌ها به نظر می‌رسد که به یکباره به داده‌ها علاقمند شده‌اند که این علاقه به نوبه‌ی خود نیاز به ابزارهای دیداری که به درک و فهم آنها کمک می‌کنند را بوجود آورده است. حسگرهای سخت‌افزاری ارزان قیمت، چارچوب‌های خودت انجام بده برای ساخت سیستم‌های خودمان موجب کاهش هزینه‌های گردآوری و پردازش داده‌ها شده‌اند. برنامه‌های کاربردی، ابزارهای نرم‌افزاری و کتابخانه‌های کد سطح پایین بیشماری برای کمک به گردآوری، سازماندهی، دستکاری، مصورسازی و درک داده‌ها از هر نوع منبع در حال پدیدار شدن هستند. اینترنت نیز به عنوان یک کانال توزیع خارق‌العاده برای مصورسازی‌ها عمل کرده است و جامعه‌ی متنوعی از طراحان، برنامه‌نویسان، نقشه‌نگاران، متفکران و کارشناسان داده‌ها گردهم آمده‌اند تا انواع افکار و ابزارهای لازم برای کار با داده‌ها چه به شکل دیداری و چه غیردیداری را منتشر کنند.

گوگل مپس نیز قراردادهای رابط کاربری (برای بزرگنمایی دو مرتبه کلیک کنید) و فناوری (کاشی‌های مربعی ۲۵۶ پیکسلی نقشه‌ها با اسامی قابل پیش‌بینی فایل‌ها) برای نمایش آنلاین نقشه‌ی جغرافیایی تعاملی تا جایی در دسترس عموم قرار داد که اکثر

¹ Jacques Bertin

² Semiologie Graphique

مردم اکنون می‌دانند که چگونه باید با یک نقشه‌ی آنلاین کار کنند. فلش^۱ نیز به عنوان یک پلتفرم سازگار با تمامی مرورگرها توانسته به خوبی به طراحی و توسعه‌ی برنامه‌های کاربردی اینترنتی زیبا و غنی کمک کرده و نقشه‌ها و مصورسازی‌های تعاملی داده‌ها را در خود بگنجانند. امروزه، فناوری‌های نوین مرورگرها از قبیل کانواس^۲ و اس‌وی‌جی^۳ (که برخی اوقات تحت اچ‌تی‌ام‌ال^۴ آورده می‌شود) به شکلی پدیدار شده‌اند که سلطه و گستردگی رابط‌های کاربری مصورسازی پویا به دستگاه‌های موبایل را به چالش کشیده‌اند.

پیش‌بینی آینده‌ی مصورسازی داده‌ها/اطلاعات کار بسیار دشواری است. ما فقط از روی آنچه که قبلاً اختراع شده است می‌توانیم نتیجه‌گیری‌هایی داشته باشیم: افزایش مصورسازی‌های سه بُعدی، تجربه‌ی سه بُعدی داده‌های چندبعدی در محیط‌های واقعیت مجازی و مصورسازی هالوگرافیکی اطلاعات. تا پیش از پایان این دهه احتمال زیادی وجود دارد که در قلمروی مصورسازی اطلاعات مشاهده چیزی باشیم که تا کنون ابداع نشده است. مورد کاربردی ۳.۶ نشان می‌دهد که ابزارهای گزارش‌دهی/تحلیلی دیداری مثل تابلو^۵ چگونه می‌توانند از طریق ایجاد/به اشتراک گذاری بینش و اطلاعات موجب تسهیل تصمیم‌گیری کارا و اثربخش شوند.

اسمیت مک‌فارلن بینش بهره‌وری عملیاتی را با تابلو آنلاین بهبود می‌بخشد

مورد کاربردی ۳.۶



تاریخچه

اسمیت مک‌فارلن^۶ جایگاه ویژه‌ای در تاریخ پزشکی به دست آورده است. این شرکت یک قرارملاقات سلطنتی را برای تهیه‌ی داروی علیاحضرت ملکه ویکتوریا ترتیب داده و داروی کلروفورم مورد نیاز سر جیمز سیمپسون^۷، متخصص سرشناس زنان و زایمان را برای کاهش درد در حین انجام آزمایشات عمل و زایمان تأمین نمود. امروزه، اسمیت مک‌فارلن یک زیرمجموعه از بخش فاین کیمیکال آند کاتالیستس^۸ در شرکت جانسون متی^۹ محسوب می‌شود. این شرکت داروسازی، رهبر بازار جهان در زمینه‌ی تولید داروهای افیونی از قبیل کدئین و مورفین می‌باشد.

اسمیت مک‌فارلن هر روزه بر اساس این داده‌ها تصمیم‌گیری می‌کند. این شرکت، داده‌های عملیاتی ساخت و تولید را برای تحقق اهداف بهبود مستمر خود، گردآوری و تحلیل می‌کند. فروش، بازاریابی و تأمین مالی شرکت برای شناسایی فرصت‌های کسب‌وکار داروسازی، رشد درآمد و برآوردن نیازهای مشتریان به این داده‌ها متکی هستند. علاوه بر این، تأسیسات تولیدی شرکت در ادینبورگ برای حصول اطمینان از هویت، کیفیت و خلوص مواد تشکیل دهنده‌ی داروها برای مشتریان، مقامات قانون‌گذار مثل

¹ Flash

² Canvas

³ SVG

⁴ HTML5

⁵ Tableau

⁶ Smith McFarlan

⁷ Sir James Simpson

⁸ Fine Chemical and Catalysts

⁹ Johnson Matthey

اف‌دی‌ای ایالات متحده^۱ و دیگران به عنوان بخشی از روش جاری تولید محصولات خود به پایش، روندگیری و گزارش‌دهی داده‌های کیفیت می‌پردازد.

چالش‌ها: چندین منبع حقیقت و فرایندهای گزارش‌دهی کند و سنگین

فرایند گردآوری داده‌ها، تصمیم‌گیری و گزارش‌دهی کار ساده‌ای نبود. داده‌ها در سرتاسر سازمان از جمله پلتفرم برنامه‌ریزی منابع سازمانی (ای‌آرپی)، داخل پایگاه‌های داده دپارتمانی قدیمی مثل اس‌کی‌و‌ال، پایگاه‌های داده‌ی اکسس و صفحه گسترده‌های مستقل پراکنده بودند. زمانیکه برای تصمیم‌گیری به آن داده‌ها نیاز بود، زمان و منابع بیش از اندازه‌ای صرف استخراج، یکپارچه‌سازی و نمایش آنها در یک صفحه گسترده یا شکل دیگری از نمایش می‌شد.

کیفیت داده‌ها، نگرانی دیگری بود. چونکه تیم‌ها به منبع داده‌های خودشان متکی بودند، شاهد چندین نسخه از حقایق و تعارضات بین داده‌ها بودیم و برخی اوقات به سختی می‌شد تشخیص داد که کدام نسخه از داده‌ها صحیح و کدام نسخه غلط است. قضیه به همینجا ختم نشد. حتی پس از اینکه داده‌ها گردآوری و نمایش داده شدند، اعمال تغییرات بر روی داده‌ها بسیار آهسته و دشوار بود. در واقع، هر زمان که یکی از اعضای تیم اسمیت مک‌فارلن قرار بود روندگیری یا تحلیل دیگری انجام دهد، تغییرات انجام شده بر روی داده‌ها باید تأیید می‌شد. نتیجه‌ی نهایی این بود که داده‌ها تا زمانیکه قرار بود برای تصمیم‌گیری مورد استفاده قرار بگیرند، چندین مرتبه منسوخ شده بودند.

لیام میلز، رئیس بخش بهبود مستمر اسمیت مک‌فارلن به ذکر اهمیت یک سناریوی گزارش‌دهی معمول می‌پردازد:

یکی از فرایندهای گزارش‌دهی اصلی، « اقدام اصلاحی و اقدام پیشگیرانه^۲ » یا « کاپا » نام دارد که تحلیلی است که بر روی فرایندی تولیدی اسمیت مک‌فارلن انجام می‌شود تا عدم انطباق‌ها یا هرگونه وضعیت‌های نامطلوب دیگر را حذف کند. هر ماهه صدها ساعت صرف گردآوری و جمع‌آوری داده‌ها برای گزارش می‌شد که تهیه گزارش نهایی ممکن بود چندین روز به طول بینجامد. تحلیل روند نیز به دلیل ایستا بودن داده‌ها، فوت و فن‌ها بیش از حدی نیاز داشت. در سایر سناریوهای گزارش‌دهی، ما اغلب باید منتظر یک تحلیل جدول محوری صفحه‌گسترده^۳ باشیم که بر روی یک گراف نمایش داده شده و چاپ می‌شد و برای اطلاع همگان بر روی دیوار نصب می‌شد.

فرایندهای گزارش‌دهی کند و سنگین، وجود نسخه‌های مختلف از حقایق و داده‌های ایستا، همگی دلایلی بودند که به تغییر سرعت بخشیدند. به گفته‌ی میلز، « بسیاری از افراد به این خاطر که تصویر کاملی از سازمان در اختیار نداشتند مستأصل شده بودند. ما در مورد مسائلی که با آنها مواجه می‌شدیم بحث‌های بیشتر و بیشتری داشتیم – زمانیکه باید در مورد گزارش‌دهی هوش کسب‌وکار صحبت می‌کردیم.»

راه‌حل: مصورسازی تعاملی داده‌ها

یکی از تیم‌های اسمیت مک‌فارلن قبلاً تجربه‌ی استفاده از تابلو را داشته و به میلز توصیه کرد که در خصوص این راه‌حل نرم‌افزاری بیشتر تحقیق کند. میلز با بررسی یک نسخه‌ی آزمایشی تابلو آنلاین سریعاً متقاعد شد که این راه‌حل مصورسازی تعاملی داده‌ها می‌تواند به مشکلاتی که شرکت در خصوص داده‌ها با آنها مواجه است خاتمه بخشد.

¹ US FDA

² Corrective Action and Preventive Action (CAPA)

³ Spreadsheet Pivot Table

او می‌گوید: « من بلافاصله متقاعد شدم. سهولت استفاده، کارکردها و گستردگی امکانات مصورسازی داده‌ها بسیار چشمگیر بوده است و البته که با توجه به اینکه یک نرم‌افزار به عنوان خدمت بوده است نیازی به سرمایه‌گذاری در زیرساخت فناوری نبوده و این انعطاف‌پذیری وجود دارد که هر زمان که بخواهیم کاربران جدیدی را به آن بیفزاییم.»

موضوع امنیت داده‌های آنلاین یکی از سؤالات اساسی که باید پاسخ داده شود. « شرکت مادر ما، جانسون مَتی یک استراتژی فناوری ابری داشت ولی باید مطمئن میشد که راه‌حل‌های بکار گرفته شده از امنیت کامل برخوردارند. ویژگی‌هایی از قبیل شناسایی یگانه و دسترسی افراد مجاز برای تعامل با داده‌ها در تابلو آنلاین موجب شده تا سطح ایمنی و اطمینان بالایی داشته باشد. »

یک مسئله‌ی امنیتی دیگر که اسمیت مک‌فارلن و جانسون مَتی خواستار پاسخگویی به آن بودند این بود: « داده‌ها از لحاظ فیزیکی کجا ذخیره می‌شوند؟ میلز مجدداً می‌گوید: « ما از عملکرد تابلو آنلاین در برآورده کردن معیارهای امنیت داده‌ها و حریم خصوصی رضایت داریم. داده‌ها و کاربرگ‌ها همگی در مرکز داده‌های جدید تابلو در دوبلین ذخیره می‌شوند و بدین ترتیب داده‌ها از اروپا خارج نمی‌شود. »

مدیر فروش شرکت تابلو طی شش هفته آزمایشی با همکاری میلز و تیمش، یک مورد کسب‌وکار را برای تابلو آنلاین ایجاد کردند. تیم مدیریت تقریباً بلافاصله آن را تأیید کرد و یک برنامه پایلوت متشکل از ۱۰ کاربر شروع بکار کرد. این برنامه پایلوت شامل یک راهکار بهبود کیفیت تولید بود: نگاه به انحرافات از نرم مثل زمانیکه یک دستگاه گرمایش که در تولید مواد دارویی افیونی استفاده می‌شود فراتر از درجه حرارت آستانه کار کند. از اینجا یک داشبورد « عملیات کیفیت » برای ردیابی و اندازه‌گیری انحرافات و تنظیم درجه حرارت برای بهبود کیفیت عملیات و بهره‌وری ایجاد شد.

میلز می‌گوید: « این داشبورد بلافاصله محل وقوع انحرافات را نشان می‌دهد. ما مجبور نبودیم از بین انبوه ردیف داده‌ها جستجو و جو کنیم و جواب‌ها بلافاصله به دست ما می‌رسید.»

در سرتاسر این فاز آزمایش اولیه و پایلوت، تیم از کمک‌های آموزشی تابلو مثل ویدئوهای آموزشی رایگان، راهنمای گام به گام محصول و آموزش‌های آنلاین زنده استفاده کردند. آنها همچنین در یک رویداد « آموزش مبانی » در لندن شرکت کردند. طبق اظهارات میلز، « این آموزش در سطح خبره و بسیار دقیق و کاملاً متناسب با نیازها بود. این دوره به همه نشان داد که تابلو آنلاین تا چه اندازه شهودی است. ما می‌توانیم داده‌های ۱۰ سال را تنها با چند کلیک مصورسازی کنیم.» این شرکت اکنون پنج کاربر تابلو دسکتاپ و حدود ۲۰۰ کاربر مجوزدار تابلو آنلاین دارد.

میلز و تیم وی به ویژه علاقمند ویژگی تابلو یونیون^۱ در نسخه‌ی ۹.۳ شدند که به آنها اجازه می‌داد تا داده‌ی تفکیک شده به فایل‌های کوچک را بهم متصل کنند. وی می‌گوید: « برخی اوقات گردهم آوردن داده‌هایی که برای تحلیل نیاز داریم کار دشواری است. ویژگی یونیون به ما اجازه می‌دهد تا با داده‌هایی که در چندین فایل یا برگه پخش شده‌اند کار کنیم و بدین ترتیب زمان صرف شده برای آماده‌سازی داده‌ها را کاهش می‌دهد. »

نتایج: استفاده از تحلیل ابری برای تغییر شکل تصمیم‌گیری و گزارش‌دهی

اسمیت مک‌فارلن با استانداردسازی تابلو آنلاین موجب تغییر سرعت و دقت تصمیم‌گیری و گزارش‌دهی کسب‌وکارش شد:

¹ Tableau Union

- داشبوردهای تعاملی جدید را می‌توان در عرض یک ساعت ایجاد کرد. سابق بر این، یکپارچه‌سازی و ارائه‌ی داده‌ها در یک صفحه‌گسترده‌ی ایستا چند روز طول می‌کشید.
- تولید گزارش فرایند ساخت و تولید کاپا که نیازمند صدها نفر ساعت در هر ماه و چندین روز برای تولید بود اکنون در عرض چند دقیقه قابل انجام است (با توجه به بینش‌های اشتراک‌گذاری شده در فضای ابر).
- گزارشات را می‌توان به سرعت و به آسانی و بدون نیاز به مداخله فنی در حالت on-the-fly تغییر داد. اسمیت مک‌فارلن از انعطاف‌پذیری لازم برای انتشار داشبوردها با تابلو دسکتاپ و به اشتراک‌گذاری آنها با همکاران، شرکا یا مشتریان برخوردار است.
- شرکت از یک نسخه‌ی تکی و مورد اطمینان از حقایق برخوردار است.
- اسمیت مک‌فارلن اکنون در مورد داده‌هایش بحث و گفتگو می‌کند نه در خصوص مسائل حول یکپارچگی و کیفیت داده‌ها.
- کاربران جدید را بلافاصله می‌توان به صورت آنلاین درآورد و هیچ نیازی به مدیریت زیرساخت‌های فنی نیست.

پس از این موفقیت اولیه، اسمیت مک‌فارلن اکنون در حال گسترش تابلو آنلاین به گزارش‌دهی مالی، تحلیل زنجیره تأمین و پیش‌بینی فروش است. میلز اینگونه نتیجه‌گیری می‌کند: «استراتژی کسب‌وکار ما اکنون به جای عقیده و نظرات، مبتنی بر داده‌ها است. ما به کمک مصورسازی تعاملی می‌توانیم روندها را بلافاصله تشخیص داده، بهبودهای فرایندی را شناسایی کرده و هوش کسب‌وکار را به سطح بعدی ببریم. تابلو معرف حرفه‌ی من است.»

سوالاتی برای مورد کاربردی ۳.۶

۱۲. اسمیت مک‌فارلن در خصوص داده‌ها و گزارش‌دهی با چه چالش‌هایی روبرو بود؟
۱۳. چه راه‌حل/نتایج (منافعی) به دست آمد؟

Source: Tableau Customer Case Study, "Macfarlan Smith improves operational performance insight with Tableau Online,"

<http://www.tableau.com/stories/customer/macfarlan-smith-improvesoperational-performance-insight-tableau-online> (accessed June 2018).

Used with permission from Tableau Software, Inc.

سوالاتی برای مرور بخش ۳.۸

۱۴. مصورسازی داده‌ها را تعریف کنید. چرا به چنین چیزی نیاز داریم؟
۱۵. ریشه‌های تاریخی مصورسازی داده‌ها چه هستند؟
۱۶. نمایش گرافیکی شارل ژوزف مینار از لشگرکشی ناپلئون را به دقت تحلیل کنید. تمامی ابعاد اطلاعاتی در این نمودار تاریخی را شناسایی و در مورد آنها اظهارنظر کنید.
۱۷. ادوارد تافت که بود؟ چرا فکر می‌کنید که باید از کار وی مطلع باشیم؟
۱۸. به نظر شما «چیز بزرگ بعدی» که در مصورسازی داده‌ها باید منتظر آن باشیم چیست؟

کاربران نهایی سیستم‌های تحلیل کسب‌وکار اغلب مطمئن نیستند که برای یک هدف خاص از چه نوع نمودار یا گرافی استفاده کنند. برخی گراف‌ها و نمودارها برای پاسخگویی به انواع خاصی از سوالات مناسب هستند. برخی دیگر بهتر از سایر نمودارها و گراف‌ها به نظر می‌رسند. برخی ساده و برخی پیچیده‌تر و شلوغ‌تر هستند. در ادامه به شرح کوتاهی از انواع نمودارها و/یا گراف‌هایی که در اکثر ابزارهای تحلیل کسب‌وکار یافت می‌شوند و انواع سوالاتی که این نمودارها بهتر از سایر گزینه‌ها قادر به پاسخگویی/تحلیل آنها هستند می‌پردازیم. این مطالب از چندین مقاله‌ی منتشر شده و ادبیات موضوع‌های دیگر گردآوری شده است (آبلا^۱؛ هاردین و همکاران ۲۰۱۲؛ اس‌ای‌اس ۲۰۱۴).

در ادامه به معرفی نمودارها و گراف‌های پایه که عموماً برای مصورسازی اطلاعات بکار می‌روند می‌پردازیم.

نمودار خطی نمودار خطی را شاید بتوان پرکاربردترین نمودار گرافیکی مورد استفاده برای نمایش داده‌های سری زمانی دانست. نمودارهای خطی (یا گراف‌های خطی) ارتباط بین دو متغیر را نشان می‌دهند و اغلب برای ردیابی تغییرات یا روندها در طول زمان بکار می‌روند (یکی از این متغیرها، زمان است که بر روی محور x قرار می‌گیرد). نمودارهای خطی تک تک نقاط داده را به صورت متوالی به یکدیگر متصل کرده و به استنتاج روندهای متغیر در طول یک دوره زمانی کمک می‌کنند. از نمودارهای خطی عمدتاً برای نمایش تغییر مقادیر زمانی مثل تغییر قیمت سهام در طی یک دوره زمانی پنج ساله یا تغییرات تعداد تماس‌های روزانه‌ی خدمات مشتریان در یک ماه است.

نمودار میله‌ای نمودار میله‌ای جزو پایه‌ای‌ترین نمایش‌های دیداری مورد استفاده از نمایش داده‌ها است. این نمودارها برای نمایش داده‌های اسمی یا عددی که به خوبی قابل تفکیک به گروه و دسته‌های مختلف هستند مناسب می‌باشند که به سرعت می‌توان نتایج و روندهای مقایسه‌ای داخل داده‌ها را به سرعت مشاهده کرد. از نمودارهای میله‌ای اغلب برای مقایسه‌ی داده‌ها در گروه و دسته‌های متعدد از قبیل درصد هزینه‌های تبلیغات به تفکیک دپارتمان‌ها یا گروه محصولات استفاده می‌شود. نمودارها میله‌ای به شکل افقی یا عمودی ترسیم می‌شوند. همچنین ممکن است برای نمایش ابعاد متعدد در یک نمودار به صورت تجمعی بر روی یکدیگر ترسیم شوند.

نمودار کلوچه‌ای نمودار کلوچه‌ای همانطور که از نامش پیداست از لحاظ دیداری شبیه کلوچه است. با توجه به جذابی ظاهری این نمودارها اغلب به شکلی نادرست از آنها استفاده می‌شود. نمودارهای کلوچه‌ای را فقط باید برای نمایش سهم نسبی یک سنجه‌ی خاص بکار برد. به عنوان مثال، از این نمودارها می‌توان برای نمایش درصد نسبی بودجه‌ی هزینه‌ی تبلیغات برای خطوط محصولات مختلف استفاده کرده یا می‌توانند سهم نسبی دانشجویان سال دوم دانشگاه را نشان دهند. اگر تعداد گروه یا دسته‌هایی که می‌خواهیم نمایش دهیم زیاد باشد (مثلاً بیش از چهار گروه)، آنگاه به جای نمودار کلوچه‌ای حتماً باید از نمودار میله‌ای استفاده کنیم.

¹ Abela

نمودار پراکندگی از نمودار پراکندگی اغلب برای یافتن ارتباط بین دو یا سه متغیر (در حالت ۲ بعدی یا ۳ بعدی) استفاده می‌شود. چونکه نمودارهای پراکندگی، ابزارهای کشف و بررسی دیداری هستند، به آسانی نمی‌توان بیش از سه متغیر را به بیش از سه بُعد ترجمه کرد. نمودارهای پراکندگی روش مؤثری برای بررسی وجود روندها، جمع‌شدگی‌ها و مقادیر پرت می‌باشند. به عنوان مثال، در یک گراف دو متغیره (دو محوره)، از یک نمودار پراکندگی می‌توان برای نمایش ارتباطات بین سن و وزن بیماران قلبی یا ارتباط بین تعداد نمایندگی‌های مراقبت از مشتریان و تعداد ادعای خسارات باز خدمات مشتریان استفاده کرد. اغلب یک خط روند برای نمایش ماهیت و جنس این ارتباط بر روی نمودار پراکندگی دو بُعدی قرار داده می‌شود.

نمودار حبابی نمودار حبابی در واقع یک نسخه‌ی ارتقاء یافته از نمودارهای پراکندگی است. نمودارهای حبابی یک نوع جدید از مصورسازی محسوب نمی‌شوند و باید آنها را یک تکنیک برای غنی‌سازی داده‌های ترسیم شده در نمودارهای پراکندگی (یا حتی نقشه‌های جغرافیایی) در نظر گرفت. با تغییر اندازه و/یا رنگ دایره‌ها، می‌توانیم ابعاد جدیدی از داده‌ها را نمودار بیفزاییم و بدین ترتیب برای غنای مفهومی داده‌ها اضافه کنیم. به عنوان مثال، از یک نمودار حبابی می‌توان برای نمایش حضور دانشجویان در کلاس‌های دانشگاه بر حسب زمان در روز یا برای نمایش حاشیه سود بر حسب نوع محصول و منطقه‌ی جغرافیایی استفاده کرد.

نمودارها و گراف‌های تخصصی که در این بخش بررسی خواهیم کرد یا حالت‌های ویژه‌ای از نمودارهای پایه هستند یا اینکه نسبتاً جدید بوده و برای یک نوع مسئله یا یک حوزه کاربردی خاص طراحی شده‌اند.

هیستوگرام به لحاظ گرافیکی، هیستوگرام بسیار شبیه نمودار میله‌ای است. تفاوت بین هیستوگرام و نمودارهای میله‌ای معمولی در اطلاعاتی است که نمایش داده می‌شود. از هیستوگرام‌ها برای نمایش توزیع فراوانی یک یا چندین متغیر استفاده می‌شود. در یک هیستوگرام، محور x برای نمایش گروه یا دامنه‌ها و محور y برای نمایش سنجه‌ها/مقادیر/فراوانی‌ها استفاده می‌شود. هیستوگرام‌ها در واقع شکل توزیعی داده‌ها را نشان می‌دهند. بدین ترتیب، به صورت دیداری می‌توان بررسی کرد که آیا داده‌ها دارای توزیع نرمال هستند یا نه. به عنوان مثال، می‌توانیم از یک هیستوگرام برای نمایش نتایج آزمون یک کلاس، نمایش توزیع نمرات و مقایسه‌ها یا توزیع سنی مشتریان استفاده کنیم.

گانت چارت گانت چارت یک حالت خاص از نمودار میله‌ای افقی است که برای ترسیم خطوط زمانی پروژه، مدت زمان فعالیت/وظایف پروژه و همپوشانی بین فعالیت‌ها/وظایف استفاده می‌شود. گانت چارت‌ها با نمایش زمان‌های شروع و پایان فعالیت‌ها/وظایف و ارتباطات همپوشان، کمک‌های ارزشمندی برای مدیریت و کنترل پروژه‌ها فراهم می‌کنند. به عنوان مثال، از گانت چارت‌ها اغلب برای نمایش خطوط زمانی پروژه، همپوشانی وظایف، زمان نسبی تکمیل وظایف (یک میله‌ی ناقص که درصد تکمیل را در داخل یک میله‌ی کامل قرار دارد نشان دهنده‌ی میزان نسبی تکمیل آن وظیفه یا فعالیت است)، منابع تخصیص یافته به هر وظیفه، مایلستون‌ها و اقلام قابل تحویل مورد استفاده قرار می‌گیرند.

نمودار پرت نمودار پرت (یا دیاگرام شبکه) اصولاً برای ساده‌سازی برنامه‌ریزی و زمانبندی پروژه‌های بزرگ و پیچیده کاربرد دارد. یک نمودار پرت، ارتباطات تقدم بین وظایف/فعالیت‌های پروژه را نمایش می‌دهد. این نمودار از گره‌ها (به صورت دایره یا مربع) و خطوط (فلش‌های جهت‌دار) تشکیل شده‌اند. بسته به حالت انتخاب شده می‌توان از گره‌ها یا خطوط برای نمایش وظایف/فعالیت‌های پروژه استفاده کرد (حالت نمایش فعالیت-بر-روی-گره و فعالیت-بر-روی-فلش).

نقشه جغرافیایی زمانیکه مجموعه داده‌ها شامل هر نوعی از داده‌های مکانی باشند (مثلاً آدرس فیزیکی، کد پستی، نام ایالت، نام کشور، طول و عرض جغرافیایی) بهتر است که داده‌ها بر روی نقشه نمایش داده شوند. از نقشه‌ها معمولاً به همراه سایر نمودارها و گراف‌ها استفاده می‌شود. به عنوان مثال، می‌توان از نقشه‌ها برای نمایش توزیع درخواست‌های خدمات مشتریان به تفکیک محصول (به صورت نمودار کلوچه‌ای) بر حسب مکان‌های جغرافیایی استفاده کرد. به کمک نقشه‌های جغرافیایی اغلب می‌توان اطلاعات بسیار متنوع و زیادی را برای کمک به تصمیم‌گیری برای افتتاح یک رستوران یا یک ایستگاه خدماتی جدید نمایش داد (مثلاً توزیع سنی، توزیع درآمد، آموزش، رشد اقتصادی، تغییرات جمعیتی). این نوع سیستم‌ها را *سیستم‌های اطلاعات جغرافیایی* (جی‌آی‌اس)^۱ می‌نامند.

بولت یک نمودار بولت اغلب برای نمایش میزان پیشرفت در راستای یک هدف بکار می‌رود. این گراف در اصل یک گونه از نمودار میله‌ای است که اغلب برای اینکه بتوانند مفاهیم را در یک فضای کوچکتر به نحوی شهودی‌تر ارائه کنند به جای درجه‌های اندازه‌گیری، کنتورها و حرارت‌سنج‌ها در داشبوردها استفاده می‌شوند. بولت گراف‌ها، یک سنجی اصلی (مثلاً درآمد سال) را با یک یا چند سنجی دیگر (مثلاً هدف درآمد سالانه) مقایسه کرده و آن را در بافتار سنجی‌های عملکردی تعریف شده نمایش می‌دهند (مثلاً سهمیه‌ی فروش). یک بولت گراف می‌تواند عملکرد سنجی اصلی نسبت به اهداف کلی را نمایش دهد (مثلاً یک نماینده‌ی فروش تا چه اندازه به تحقق سهمیه‌ی فروش سالانه خود نزدیک شده است).

نقشه حراری نقشه حرارتی برای نمایش مقایسه‌ی مقادیر پیوسته برای دو دسته یا گروه با استفاده از رنگ بکار می‌رود. هدف از اینکار کمک به کاربران است تا سریعاً محل تقاطع قوی‌ترین و ضعیف‌ترین مقادیر عددی سنجی مذکور را پیدا کنند. به عنوان مثال، می‌توان از نقشه حرارتی برای نمایش تحلیل بخش‌بندی بازارهای هدف استفاده کرد که سنجی (گرادیان رنگی) که میزان خرید را نشان می‌دهد و ابعاد به صورت توزیع سنی و درآمدی نمایش داده می‌شوند.

جداول رنگ‌بندی شده جداول رنگ‌بندی شده در واقع همان نقشه‌های حرارتی هستند که یک قدم پیشرفته‌تر شده‌اند و علاوه بر نمایش مقطع داده‌ها با استفاده از رنگ، یک عدد نیز به بالای جدول افزوده می‌شود تا جزئیات بیشتری را ارائه دهد. این بدان معناست که جداول دو بعدی با سلول‌هایی که عملکرد را بر حسب نوع محصول و حجم فروش نمایش می‌دهند.

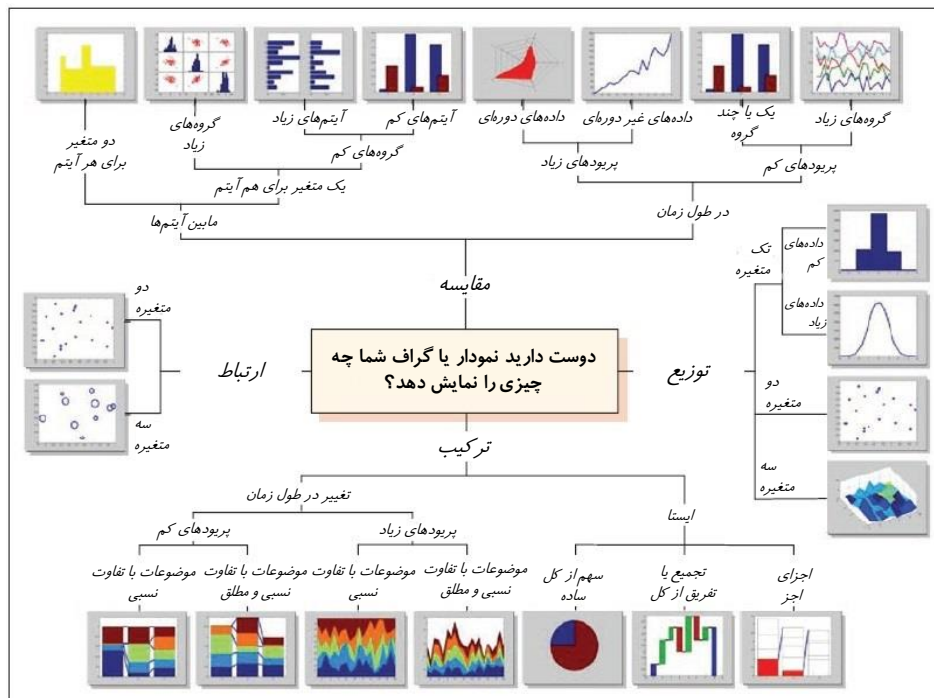
نقشه درختی یک نقشه‌ی درختی، ساختار سلسله‌مراتبی داده‌ها را به صورت مجموعه‌ای از مستطیل‌های تو در تو نمایش می‌دهد. هر شاخه‌ی این درخت یک مستطیل است که به شاخه‌های کوچکتر که با مستطیل‌های کوچکتری نمایش داده می‌شوند تقسیم می‌شود. مساحت برگ‌ها در هر شاخه متناسب با ابعاد خاص داده‌ها است. این برگ‌ها اغلب برای نمایش یک بعد جداگانه

¹ Geographical Information Systems (GIS)

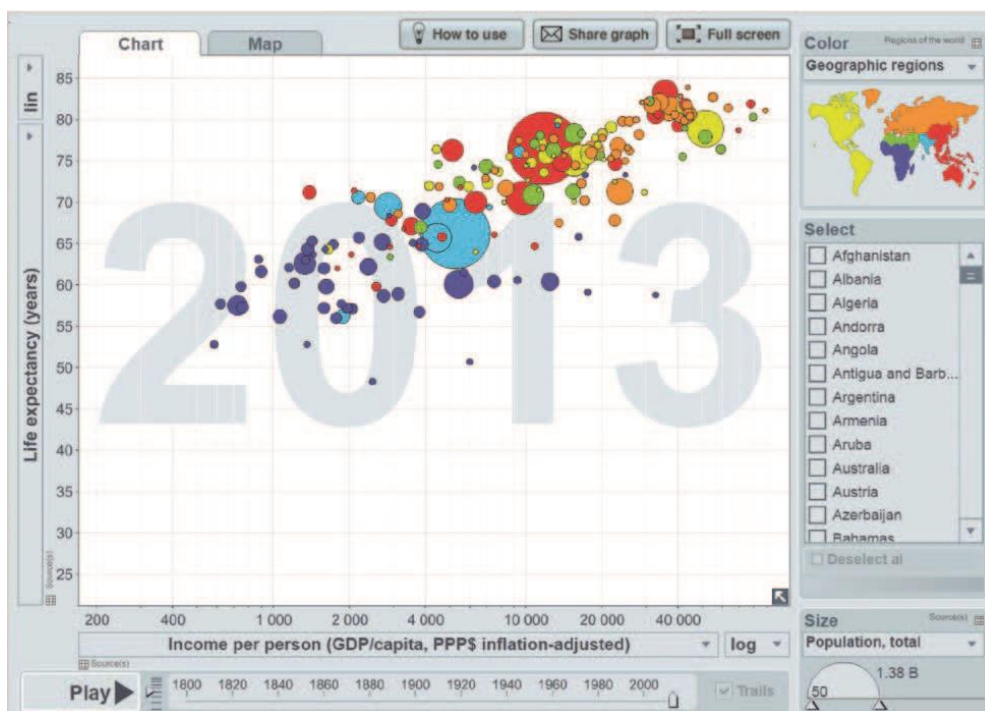
از داده‌ها با رنگ متفاوتی نمایش داده می‌شود. زمانیکه ابعاد رنگ و اندازه به نحوی با ساختار درختی همبستگی داشته باشند، براحتی می‌توانیم الگوهای را مشاهده کنیم که شناسایی آنها به هر روش دیگری بسیار دشوار است. مزیت دوم نقشه‌های درختی این است که از فضا استفاده‌ی مؤثری می‌کنند. در نتیجه، می‌توان هزاران آیتم را به طور همزمان بر روی صفحه نمایش داد.

بهترین نمودار یا گراف که در بخش قبل توضیح دادیم کدام‌ها هستند؟ پاسخ تقریباً ساده است: هیچ نمودار یا گرافی بهترین نیست چونکه اگر اینگونه بود، این تعداد نمودار یا گراف وجود نداشت. همه‌ی آنها مهارت‌های نمایش داده نسبتاً متفاوتی دارند. از این رو، سوال درست این است که « کدام نمودار یا گراف برای یک کار مشخص بهترین گزینه است؟ » قابلیت‌های نمودارهای معرفی شده در بخش قبل می‌تواند به انتخاب و استفاده از نمودار/گراف مناسب برای هر کار خاص کمک کند اما همچنان کار چندان ساده‌ای نیست. از چندین نوع نمودار/گراف می‌توان برای یک مصورسازی استفاده کرد. قانون سرانگشتی این است که ساده‌ترین شکل را انتخاب کنیم تا به درک و هضم آسان موضوع برای مخاطبان کمک کند.

اگرچه یک الگوریتم انتخاب گراف/نمودار یا طبقه‌بندی گراف/نمودار که مورد پذیرش همه باشد وجود ندارد ولی شکل ۳.۲۱ یک سازماندهی جامع و بسیار منطقی از انواع گراف/نمودارها را در قالب یک ساختار طبقه‌بندی شده نشان می‌دهد (نسخه اصلی در آپلا ۲۰۰۸ منتشر شد). ساختار طبقه‌بندی، حول سوالات « دوست دارید نمودار یا گراف شما چه چیزی را نمایش دهد؟ » - یعنی هدف و مقصد بکارگیری نمودار یا چارت سازماندهی شده است. در آن سطح، هدف از بکارگیری به چهار نوع اصلی تقسیم می‌شود - ارتباط، مقایسه، توزیع و ترکیب که سپس آنها بر اساس تعداد متغیرهای درگیر و وابستگی زمانی مصورسازی به زیرگروه‌های بیشتری تقسیم می‌شوند.



اگرچه این نمودارها و گرافها بخش عمده‌ی آنچه را که عمدتاً در مصورسازی اطلاعات بکار می‌رود پوشش می‌دهند ولی همه‌ی آنها را در بر نمی‌گیرند. امروزه، نمودارها و گراف‌های تخصصی زیادی را می‌توان یافت که برای هدف خاصی بکار می‌روند. علاوه بر این، روند جاری به این شکل است که این نمودارها را برای ظاهر جذاب‌تر و مصورسازی شهودی‌تر منابع داده‌ی پیچیده‌ی امروزی با یکدیگر ترکیب و متحرک‌سازی کنند. به عنوان مثال، نمودارهای حبابی تعالی و متحرک موجود در وب سایت گپ‌مایندر^۱ (www.gapminder.org) که روشی خاص برای اکتشاف در داده‌های بهداشت، رفاه و جمعیت جهانی از یک منظر چندبعدی ارائه کرده است. شکل ۳.۲۲ انواع نمایش‌های موجود در این سایت را نشان می‌دهد. در این گراف، حجم جمعیت، امید به زندگی و سرانه درآمد در سطح قاره نشان داده شده و یک پویانمایی متغیر زمانی که نحوه‌ی تغییر این متغیرها را در طول زمان نشان می‌دهد.



- ۱.
- ۲.
- ۳.
- ۴.

¹ Gapminder

همانگونه که سِتِ گریمز^۱ (۲۰۰۹ a, b) گفته است، ابزارها و تکنیک‌های مصورسازی داده‌ها که به کاربران تحلیل‌های کسب‌وکار و سیستم‌های هوش کسب‌وکار در کشف همبستگی‌های پنهان، اطلاع‌رسانی ارتباطات و بافتارهای تاریخی و نقل داستان‌های متقاعدکننده و همگان را به عمل ترغیب می‌کند به شدت در حال افزایش است. آخرین شماره‌ی «مربع جادویی پلتفرم‌های هوش کسب‌وکار و تحلیل^۲» که در فوریه‌ی ۲۰۱۶ توسط گارتنر منتشر شده تأکید بیشتری بر اهمیت مصورسازی داده‌ها در هوش کسب‌وکار و تحلیل‌ها دارد. همانگونه که در نمودار شکل ۳.۲۳ نشان داده شده،



¹ Seth Grims

² Magic Quadrant for Business Intelligence and Analytics Platforms

³ Tableau Software

⁴ QlikTech

⁵ MicroStrategy

⁶ Alteryx

تحلیل‌های دیداری، عبارتی است که اخیراً ابداع شده و اغلب به همان معنای مصورسازی اطلاعات است. منظور از تحلیل دیداری، ترکیبی از مصورسازی و تحلیل‌های پیشگویانه است. درحالی‌که هدف مصورسازی اطلاعات، پاسخگویی به این سوال است که «چه اتفاقی افتاده است؟» و «چه اتفاقی در حال رخ دادن است؟»، تحلیل دیداری به دنبال پاسخگویی به این سوالات است که «چرا رخ می‌دهد؟» و «چه چیزی به احتمال رخ خواهد داد؟» و معمولاً با تحلیل کسب‌وکار مرتبط است (پیش‌بینی، بخش‌بندی، تحلیل همبستگی). بسیاری از ارائه‌دهندگان برنامه‌های مصورسازی اطلاعات، قابلیت‌هایی را افزوده‌اند تا خود را ارائه‌دهندگان راه‌حل‌های تحلیل دیداری بنامند. یکی از برترین و قدیمی‌ترین ارائه‌دهندگان تحلیل‌ها، مؤسسه اس‌ای‌اس می‌باشد که از منظر دیگری به قضیه نگاه کرده است. این شرکت قابلیت‌های تحلیل خود را در داخل یک محیط مصورسازی داده‌ی سطح بالا به نام *تحلیل دیداری*^۱ قرار داده است.

گزارش‌دهی کسب‌وکار چه دیداری چه غیردیداری، چه خودکار و چه دستی، آنلاین یا کاغذی، چندان تفاوتی با یک داستان‌سرایی ندارد. در بینش فناوری ۳.۳ دیدگاه نامتعارفی برای گزارش‌دهی بهتر کسب‌وکار ارائه شده است.

بینش فناوری ۳.۲ مربع جادویی پلتفرم‌های هوش کسب‌وکار و تحلیل

هرکسی که داده‌هایی برای تحلیل داشته باشد، داستانی برای گفتن دارد چه این داده‌ها برای تشخیص علل عیوب ساخت و تولید، فروش یک فکر جدید به نحوی که بتواند ذهنیت و قوه‌ی تخیل مخاطبان هدف را جذب کند یا مطلع ساختن همکاران در مورد یک برنامه‌ی ویژه‌ی بهبود خدمات مشتریان. و زمانیکه داستان را در پس یک انتخاب استراتژیک بزرگ نقل می‌کند که شما و تیم مدیریت ارشد شما بتوانند تصمیم قوی و محکمی را اتخاذ کنند، ارائه‌ی یک داستان مبتنی بر واقعیت می‌تواند چالش‌برانگیز باشد. در تمامی موارد، این وظیفه‌ی خطیری است (کار بزرگی است). شما می‌خواهید جالب و به‌یادماندنی باشید پس کار را برای مدیران و همکاران پرمشغله‌تان باید راحت کنید. شما همچنین می‌دانید که مخصوصاً در دنیای سنجه‌محور امروز باید واقع‌نگر، جزء‌نگر و داده‌محور باشید.

ارائه‌ی صرف داده‌ها و وقایع می‌تواند وسوسه‌کننده باشد ولی زمانیکه همکاران و مدیریت ارشد به واسطه‌ی داده‌ها و وقایع فاقد بافتار مستأصل شده‌اند آنوقت شما بازنده هستید. همه‌ی ما تجربه‌ی پرزنت‌های حجیم و سنگین را داریم که در پایان متوجه می‌شویم مخاطبان به قدری توسط این داده‌ها مستأصل شده‌اند که نمی‌دانند چه فکری بکنند یا اینکه حواسشان کاملاً پرت شده و فقط بخش کوچکی از نکات کلیدی را متوجه شده‌اند.

¹ Visual Analytics

کار را با درگیر کردن تیم مدیریت اجرایی و توضیح قدرتمندانه نتایج و استراتژی‌هایتان در قالب یک داستان شروع کنید. شما به «چه چیز» داستان (داده‌ها و حقایق) و همچنین «چه کسی؟»، «چگونه؟»، «چرا؟» و «حُب که چی؟» نیاز دارید. این عناصر داستانی هستند که داده‌هایتان را برای مخاطبان، مرتبط و مشهود می‌کند. خلق یک داستان خوب می‌تواند به شما و مدیریت ارشد در تمرکز بر آنچه مهم است کمک کند.

داستان‌ها به داده‌ها و حقایق زندگی می‌بخشند. آنها به شما کمک می‌کنند تا به صحبت‌هایتان معنا و مفهوم بدهد و مجموعه حقایق پراکنده را نظم ببخشید. داستان‌ها باعث می‌شوند تا نکات کلیدی را راحت‌تر به خاطر بیاورید و تصویر رنگارنگی از آینده‌ی احتمالی ترسیم کنید. داستان‌ها همچنین موجب برقراری تعاملات بین انسان‌ها و داستان‌ها و ارتباط بین این دو می‌شوند.

از دیرباز فرهنگ‌ها از داستان‌سرایی برای انتقال دانش و محتوا به نسل بعد استفاده می‌کرده‌اند. در برخی فرهنگ‌ها، داستان‌سرایی جزو لاینفک و حیاتی هویت آنها است. به عنوان مثال، در زلاندنو، برخی از مردم مائوری، واژه‌ی *mokus* را بر روی چهره‌شان خالکوبی می‌کنند. *Mokus* یک خالکوبی صورت است که حاوی داستانی در مورد پیشینیان (یک قبیله‌ی خانوادگی) این مردم است. ممکن است طرح یک خالکوبی را بر روی چهره‌ی مردی ببینید که برخی از یک کوسه‌ماهی سرچکشی را برای نشان دادن صفات منحصری‌فرد اجدادش نشان می‌دهد. طرحی که این فرد برای خالکوبی انتخاب می‌کند تعیین می‌کند که «خود حقیقی» او و مأوای اجدادی او را چه چیزی تشکیل می‌دهد.

به همین منوال، زمانیکه سعی می‌کنیم تا داستانی را درک کنیم، داستان‌سرا سعی می‌کند تا در مسیر یافتن «شمال حقیقی» حرکت کند. اگر مدیریت ارشد به دنبال بحث در خصوص نحوه‌ی پاسخگویی به یک تغییر رقابتی است، یک داستان خوب می‌تواند کمکش کند. به عنوان مثال، شما ممکن است داده‌ها و حقایقی از دو مطالعه در اختیار داشته باشید، یکی شامل نتایج بدست آمده از یک مطالعه‌ی تبلیغاتی و دیگری از یک مطالعه‌ی رضایت از محصول. طراحی یک داستانی برای آنچه در هر دو مطالعه می‌سنجید می‌تواند به افراد کمک کند که تصویر جامع و کل از اجزا و بخش‌های پراکنده بدست آورند. برای اینکه بتوانید توزیع کنندگان را برای فروش یک محصول جدید ترغیب کنید می‌توانید از طریق یک داستان، چشم‌اندازی از آینده مورد نظر را به آنها نشان دهید. مهمترین نکته اینک، داستان‌سرایی یک کار تعاملی است و معمولاً داستان‌سرا از واژگان و تصاویری استفاده می‌کند که مخاطبان بتوانند خودشان را در فضای داستان تصور کنند. در نتیجه، آنها بیشتر درگیر موضوع شده و اطلاعات را بهتر درک می‌کنند.

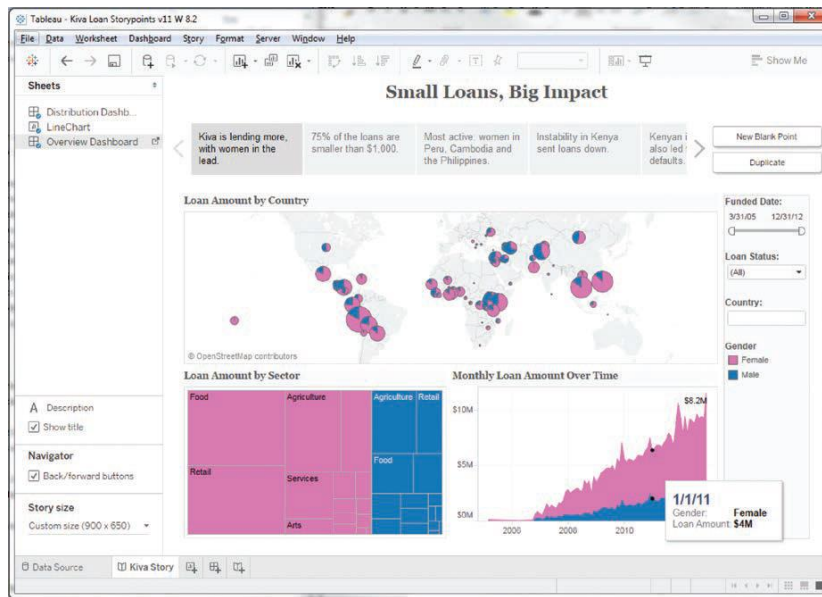
اکثر مردم

زمانیکه می‌خواهید داستانی مملو از داده‌ها طرح کنید، هدف اول این است که یک داستان پیدا کنید. شخصیت‌های این داستان چه کسانی هستند؟ چالش مورد نظر در این داستان چیست؟ بر چه موانعی باید غلبه کرد؟ و در پایان داستان می‌خواهید مخاطبان شما به چه نتیجه‌ای برسند؟

۱. **Small Loans, Big Impact** - A Tableau dashboard showing Kiva loan data. The dashboard includes a world map titled "Loan Amount by Country" and a treemap titled "Loan Amount by Sector". The treemap shows sectors like Food, Agriculture, Retail, Services, and Arts. A line chart titled "Monthly Loan Amount Over Time" shows the growth of loan amounts from 2000 to 2010. The dashboard also features a navigation pane on the left and a filter pane on the right.

- ۱.
- ۲.
- ۳.
- ۴.
- ۵.

Source: Fink, E., & Moore, S. J. (2012). "Five Best Practices for Telling Great Stories with Data." White paper by Tableau Software, Inc., www.tableau.com/whitepapers/telling-data-stories (accessed May 2016).



به دلیل افزایش تقاضا برای تحلیل‌های دیداری به همراه افزایش سریع حجم داده‌ها، سرمایه‌گذاری بر روی سیستم‌های مصورسازی بسیار کارآمد از یک رشد نمایی برخوردار شده است. شرکت غول نرم‌افزارهای آماری ساس با آخرین حرکت به سمت تحلیل‌های دیداری اکنون

در بین پیشگامان این موج فرار گرفته است. محصول جدید این شرکت به نام ساس ویژوال آنالیتیکز^۱ یک راه‌حل رایانشی بسیار توانمند

ساس

•

•



ساس

•

•

¹ SAS Visual Analytics

² End-to-end

سائس

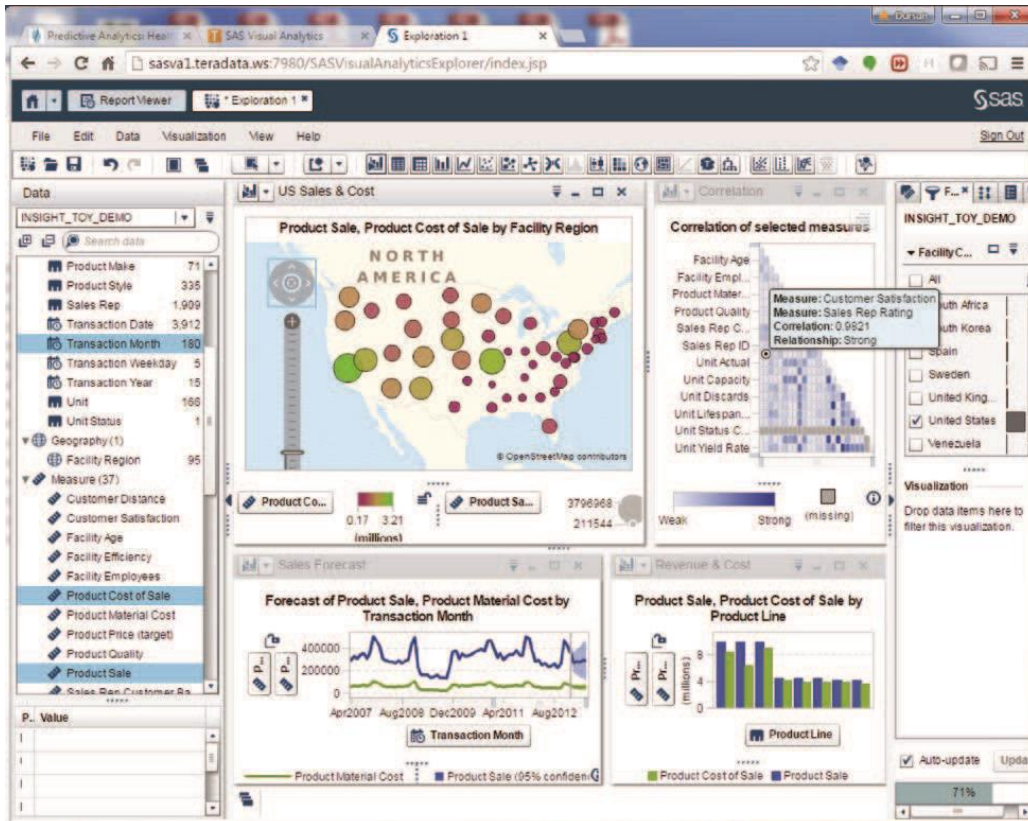
•

سائس

سائس

•

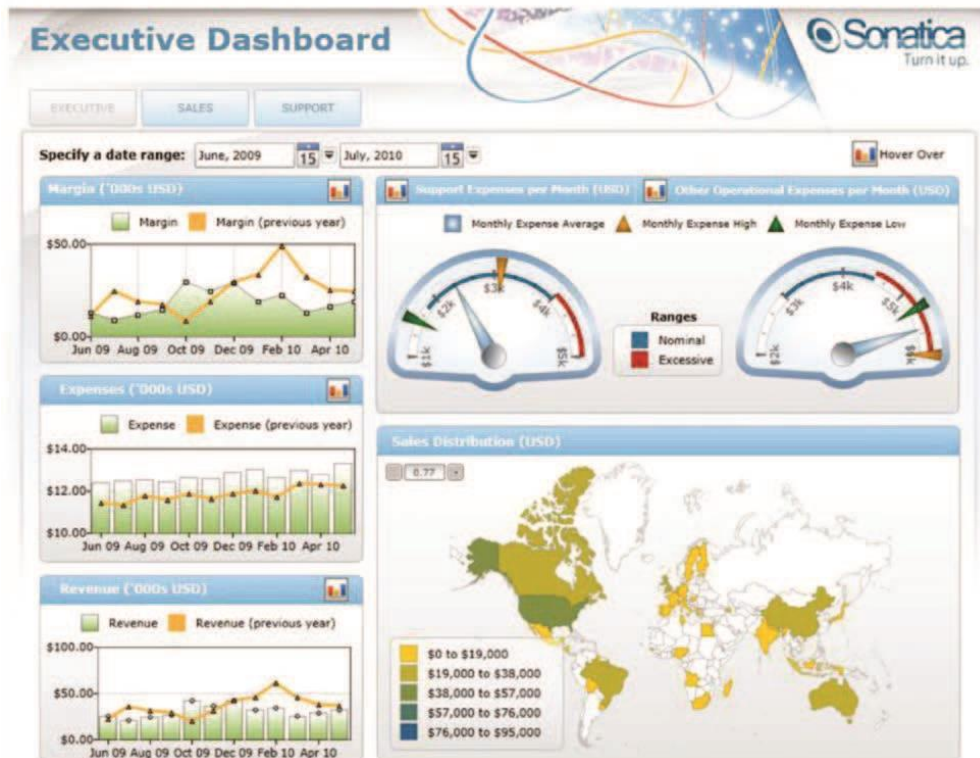
سائس



۱. ...
۲. مربع جادویی پلتفرم‌های هوش کسب‌وکار و تحلیل نگاه کنید
۳. ...
۴. ...
۵. ...

داشبوردهای اطلاعاتی، مؤلفه‌های مشترک اگر نگوییم تمام، ولی اکثر پلتفرم‌های تحلیل کسب‌وکار یا هوش کسب‌وکار، سیستم‌های مدیریت عملکرد کسب‌وکار و مجموعه‌های نرم‌افزاری سنجش عملکرد محسوب می‌شوند. داشبوردها، نمایش‌های دیداری از مهمترین اطلاعات را در یک قالب متمرکز و مرتب شده بر روی یک صفحه نمایش نشان می‌دهند که بدین شکل، اطلاعات را به آسانی و در یک نگاه می‌توان مشاهده و درک نمود و در صورت نیاز به اطلاعات بیشتر وارد سطوح پایین‌تر شد. یک داشبورد معمول در شکل ۳.۲۷ نشان داده شده است. این داشبورد اجرایی خاص، انواع مختلف شاخص‌های کلیدی عملکرد را برای یک شرکت نرم‌افزاری فرضی به نام سوناتیکا^۱ نشان می‌دهد.

¹ Sonatica



Source: A Sample Executive Dashboard from Dundas Data Visualization, Inc., www.dundas.com, reprinted with permission.

در سمت چپ داشبورد، تغییرات فصلی درآمد، مخارج و حاشیه‌ها و همچنین مقایسه‌ی آنها با اعداد و ارقام ماهانه‌ی سال گذشته را مشاهده می‌کنید. در قسمت بالا سمت راست، دو گیج اندازه‌گیری با مناطق رنگی وجود دارد که هزینه‌ها یا مخارج ماهانه‌ی خدمات پشتیبان (سمت چپ) و سایر هزینه‌ها و مخارج (سمت راست) را نشان می‌دهد. بر اساس کدگذاری رنگ‌ها، اگرچه هزینه‌های ماهانه‌ی خدمات پشتیبان در محدوده‌ی نرمال قرار دارند ولی سایر مخارج و هزینه‌ها در منطقه‌ی قرمز هستند که نشان می‌دهد از حد نرمال فراتر رفته است. نقشه‌ی جغرافیایی در قسمت پایین سمت راست، توزیع فروش در سطح کشور در سرتاسر جهان را نشان می‌دهد. در پشت این آیکن‌های گرافیکی، توابع ریاضی متعددی وجود دارد که نقاط داده متعدد را تا بالاترین سطح آمار و ارقام معنادار تجمیع کرده‌اند. با لکیک بر روی این آیکن‌های گرافیکی، استفاده کننده‌ی این اطلاعات می‌تواند تا جزءترین سطوح داده‌ها و اطلاعات پیش برود.

از داشبوردها در انواع کسب‌وکارها و برای دلایل بسیار مختلف و گسترده‌ای استفاده می‌شود. به عنوان مثال، در مورد کاربردی ۳.۷، چکیده‌ای از اجراهای موفق داشبوردهای کسب‌وکار را توسط تیم فوتبال دالاس کابویز^۱ خواهید یافت.

دالاس کابویز با تابلو و تکنیون به امتیاز بالایی دست یافت

مورد کاربردی ۳.۷

¹ Dallas Cowboys

دالاس کابویز که در سال ۱۹۶۰ تأسیس شد یک تیم حرفه‌ای فوتبال آمریکایی مستقر در ایروینگ ایالت تگزاس است. این تیم سابقه‌ی ملی درخشانی دارد که بهترین آن را شاید بتواند رکورد ان‌اف‌ال برای تعداد بازی‌های متوالی در استادیوم‌هایی دانست که تمام بلیت آنها فروخته شده است.

چالش

بیل پریاکوز^۱، مدیر ارشد اجرایی بخش بازاریابی دالاس کابویز و تیمش برای اینکه بتوانند به سودآوری بیشتری دست پیدا کنند به شفافیت بیشتری برای مشاهده داده‌هایشان نیاز داشتند. آنها میکروسافت را به همراه تعدادی از برنامه‌های کاربردی دیگر در حوزه فروش، لجستیک و تجارت الکترونیک (مثلاً ام‌دبلیو^۲) به عنوان پلتفرم مبنای خود برای این منظور انتخاب کردند. آنها انتظار داشتند که این معماری اطلاعاتی جدید بتواند تحلیل‌ها و گزارش‌دهی مورد نیازشان را برآورده کند. متأسفانه، چنین اتفاقی رخ نداد و جستجوها برای یک ابزار گزارش‌دهی، تحلیل و داشبوردسازی درست و محکم آغاز شد.

راه‌حل و نتایج

تابلو و تکنیون^۳ قابلیت‌های گزارش‌دهی و داشبورد لحظه‌ای را ارائه کرده‌اند که فراتر از نیازمندی‌های کابویز بوده است. تیم تکنیون به لحاظ سیستماتیک و متدیک دوش به دوش مالکان و کاربران داده‌ها در دالاس کابویز برای تحویل کارکردهای مورد نظر و تحت بودجه تعیین شده همکاری کردند. به گفته‌ی بیل لوییزی، معاون مدیرعامل شرکت تکنیون، «در اوایل فرایند، ما توانستیم به درک روشنی از آنچه برای عملیات سودآورتر کابویز باید انجام دهیم دست پیدا کنیم.» وی افزود: «این گام فرایند، یک گام مهم در رویکرد تکنیون با هر مشتری است و همواره با پیشرفت طرح پیاده‌سازی، نتایج بسیار خوبی را به همراه داشته است. البته، تابلو نیز در سرتاسر پروژه همکاری بسیار نزدیکی با ما و کابویز داشته است. ما به کمک یکدیگر اطمینان حاصل کردیم که کابویز در زمان تعیین شده بتواند به اهداف گزارش‌دهی و تحلیلی خود برسد. اکنون برای اولین بار، دالاس کابویز قادر است تا کل فعالیت‌های بازاریابی خود را از سازنده تا مشتری نهایی پایش کند و نه تنها شاهد این باشد که چه اتفاقاتی در طول چرخه‌ی عمر محصول اتفاق می‌افتد بلکه برای مشاهده‌ی جزئیات به سطوح پایین‌تر نیز برود. در حال حاضر، از این راه‌حل هوش کسب‌وکار برای گزارش‌دهی و تحلیل فعالیت‌های دپارتمان بازاریابی استفاده می‌شود که مسئولیت تمامی فروش‌های برند دالاس کابویز را بر عهده دارد. بر اساس برآوردهای صنعت، کابویز ۲۰ درصد از کل فروش اقلام ان‌اف‌ال را به خود اختصاص داده است که بیانگر این واقعیت است که شناخته‌شده‌ترین برند فرانشیز ورزشی در جهان هستند. طبق اظهارات گزارشگر دنیای کامپیوتر^۴، اریک لی^۵، اگرچه ممکن است تونی رومو^۶ و سایر اعضای تیم دالاس کابویز طی چندین سال گذشته در حوزه‌ی فوتبال عملکرد متوسطی داشته‌اند ولی در حوزه‌ی بازاریابی آنها تیم شماره ۱ آمریکا محسوب می‌شوند.

¹ Bill Priakos

² MW

³ Teknion

⁴ ComputerWorld

⁵ Eric Lai

⁶ Tony Romo

سوالات برای مورد کاربردی ۳.۷

۱. دالاس کابویز چگونه از مصورسازی اطلاعات استفاده کرد؟
۲. چالش‌ها، راه‌حل پیشنهادی و نتایج بدست آمده چه بودند؟

Sources: Lai, E. (2009, October 8). "BI Visualization Tool Helps Dallas Cowboys Sell More Tony Romo Jerseys," *ComputerWorld*.
Tableau case study. tableau.com/blog/computerworld-dallascowboys-business-intelligence (accessed July 2018).

داشبوردها مفهوم جدیدی نیستند و سابقه‌ی آنها حداقل به دوران شکل‌گیری سیستم‌های اطلاعات مدیریت در دهه ۸۰ باز می‌گردد. امروزه، داشبوردها همه‌جا حاضر و فراگیر هستند. به عنوان مثال، تا چند سال پیش، مؤسسه پژوهشی فورستر برآورد کرده بود که بیش از ۴۰ درصد از ۲,۰۰۰ شرکت بزرگ دنیا از این فناوری استفاده کرده‌اند (آنته آند مک‌گرگور ۲۰۰۶). از آن زمان، به طور قطع می‌توانیم بگوییم که این رقم به طور چشمگیری افزایش یافته است. در واقع، امروزه نمی‌توانیم شرکت بزرگی را پیدا کنیم که از یک سیستم هوش کسب‌وکار استفاده می‌کند ولی از داشبوردهای عملکردی استفاده نکند. سایت داشبورد اسپای وب (dashboardspy.com/about) شواهد بیشتری را در این خصوص ارائه می‌کند. این سایت حاوی توضیحات و تصویری از هزاران داشبورد هوش کسب‌وکار، کارت‌های امتیازی و رابط‌های کاربری هوش کسب‌وکار است که سازمان‌هایی از صنایع و ابعاد مختلف، سازمان‌های غیرانتفاعی و نهادهای دولتی از آنها استفاده می‌کنند می‌باشد.

طبق اظهارات اِکرسون (۲۰۰۶)، یکی از کارشناسان خبره و سرشناس در حوزه‌ی هوش کسب‌وکار و به طور تخصصی در حوزه‌ی داشبوردها، متمایزترین ویژگی یک داشبورد این است که دارای سه لایه اطلاعاتی می‌باشد:

۱. پایش: استفاده از داده‌های گرافیکی و چکیده برای پایش سنجه‌های کلیدی عملکرد.
۲. تحلیل: داده‌های خلاصه‌سازی شده چندبُعدی برای تحلیل علل اصلی مسائل و مشکلات.
۳. مدیریت: داده‌های عملیاتی با جزئیات زیاد که برای شناسایی اقدامات لازم برای حل مسئله استفاده می‌شوند.

با توجه به این لایه‌ها، داشبوردها حاوی حجم انبوهی از اطلاعات هستند که تنها در یک صفحه به طور فشرده نمایش داده می‌شوند. طبق گفته‌ی فیو^۱ (۲۰۰۵)، «چالش اساسی در طراحی داشبوردها این است که چگونه تمامی اطلاعات مورد نیاز را به طور شفاف و بدون پرت کردن حواس و به نحوی که سریعاً بتوانیم به اطلاعات مورد نظرمان دست پیدا کنیم در یک صفحه نمایش دهیم.» برای سرعت‌دهی به ادغام اعداد و ارقام، باید آنها را در بافتار خودشان قرار دهیم. اینکار را می‌توان با مقایسه‌ی اعداد مورد نظر با سایر اهداف کمی یا مبنایها، با نشان دادن اینکه آیا اعداد خوب هستند یا نیستند، با نشان دادن اینکه آیا روند رو به بهبود بوده یا خیر و با استفاده از ویجت‌های نمایشی تخصصی برای ایجاد بافتارهای مقایسه‌ای و ارزشیاب انجام داد. برخی از مقایسه‌هاست متداول که عموماً در سیستم‌های هوش کسب‌وکار انجام می‌شوند شامل مقایسه با مقادیر گذشته، مقادیر پیش‌بینی شده، مقادیر هدف، الگو یا مقادیر میانگین، نمونه‌های متعدد از سنج‌های یکسان و مقادیر سایر سنج‌ها (مثلاً درآمد در برابر هزینه) می‌باشند.

¹ Few

حتی با داشتن سنج‌های مقایسه‌ای، بازهم بسیار مهم است که مشخص کنیم آیا یک عدد، خوب است یا بد است و آیا روند آن در مسیر مناسب حرکت می‌کند یا خیر. بدون این انتخاب‌های ارزشیابانه، تعیین وضعیت یک عدد یا نتیجه کار زمان‌بری خواهد بود. معمولاً، اشیاء دیداری تخصصی (مثل چراغ‌های راهنمایی، گیج‌ها) یا ویژگی‌های دیداری (مثل رنگ‌بندی‌ها) برای تنظیم بافتار ارزشیاب بکار می‌روند. یک راه‌حل جذاب برای گزارش‌دهی داده‌ها در قالب داشبوردها توسط یک شرکت فعال در حوزه انرژی ارائه شده که در مورد کاربردی ۳.۸ توضیح داده شده است.

مورد کاربردی ۳.۸

تحلیل‌های دیداری به عرضه‌کننده‌ی انرژی برای برقراری ارتباطات بهتر کمک می‌کنند

بازارهای انرژی در سرتاسر جهان دستخوش تغییرات و دگرگونی‌های شدیدی شده‌اند و فرصت‌های بسیاری را به همراه این چالش‌ها فراهم کرده‌اند. همانند هر صنعت دیگری، فعالان بیشتری در بازار جذب این فرصت‌ها شده‌اند که موجب افزایش رقابت و کاهش تولرانس‌های تصمیم‌گیری کسب‌وکار شده است. موفقیت در این کار نیازمند ایجاد و انتشار دقیق و به موقع اطلاعات برای هر کسی که بدان نیاز دارد است. به عنوان مثال، اگر لازم باشد که بودجه بازاریابی را به راحتی ردیابی کنید، میزان کار کارمندان را متعادل کنید و مشتریان را با پیام‌های بازاریابی سفارشی شده هدف قرار دهید آنگاه به سه راه‌حل گزارش‌دهی متفاوت نیاز خواهید داشت. الکترا بل جی‌دی‌اف سوئز^۱ تمامی اینکار را با استفاده از پلتفرم تحلیل دیداری اس‌ای‌اس برای واحد بازاریابی و فروش انجام می‌دهد. این رویکرد، در وقت و زمان کارشناسان بازاریابی در هر صنعتی که دستخوش تغییرات عمده است صرفه‌جویی می‌کند. به گفته‌ی دنی نوپ^۲، مدیر توسعه و معماری گزارش‌دهی در واحد بازاریابی و فروش الکترا بل، « تثبیت موقعیت ما در بازار انرژی کار چالش‌برانگیزی است. اینکار شامل حجم، قیمت‌ها و حاشیه‌ها هم برای مشتریان جزء و هم مشتریان سازمانی است.» این شرکت، بزرگترین تأمین‌کننده‌ی برق در بلژیک و بزرگ‌ترین تولیدکننده‌ی برق برای بلژیک و هلند است. به گفته‌ی نوپ، افزایش کارایی در ارتباط با مشتریان با توجه به کانال‌های دیجیتالی جدید و توسعه‌ی خدمات جدید مرتبط با انرژی اهمیت بسیار بالایی دارد. « هرچه مشتری را بهتر بشناسیم، احتمال موفقیت ما بیشتر خواهد بود. به همین خاطر است که ما اطلاعات بدست آمده از منابع مختلف را با یکدیگر ترکیب می‌کنیم - ترافیک تلفن با مشتریان، سوالات آنلاین، پیام‌های متنی و طرح‌های تبلیغاتی پستی. این موجب افزایش دانش ما از مشتریان شد که نهایتاً مزیت رقابتی بیشتری را برای ما به وجود آورد.»

نسخه‌ی یکتا از حقیقت

الکترا بل از پلتفرم‌ها و ابزارهای متنوعی برای گزارش‌دهی استفاده می‌کرد. اینکار برخی اوقات منجر به ابهام در ارقام گزارش‌شده می‌شد. در پردازش حجم بالای داده‌ها نیز مشکلاتی در عملکرد و بهره‌وری وجود داشت. ساس ویژوال آنالیتیکز با بهره‌گیری از فناوری درون حافظه‌ای توانست تا مشکلات ابهام و عملکرد را از بین ببرد. نوپ می‌گوید: « ما برای پاسخگویی به نیاز مصورسازی داده‌ها و اطلاعات مشتریان در داخل سازمان از استقلال عمل و انعطاف‌پذیری برخورداریم. از همه اینها گذشته، گزارش‌دهی سریع یک نیاز ضروری برای دپارتمان‌های عملیاتی مثل بازاریاب و فروش است.»

کارایی بیشتر با هزینه‌ی کمتر

¹ Electrabel GDF SUEZ

² Danny Nope

اس‌ای‌اس و ویژوال آنالیتیکز به خودکارسازی فرایند به‌روزآوری اطلاعات در گزارش می‌پردازد. به جای اینکه یک گزارش منسوخ تهیه کنید، داده‌های تمامی گزارشات یکبار در هفته به‌روزآوری شده و از طریق داشبوردها در دسترس است. الکترا بل در استقرار این راه‌حل از یک رویکرد فزیندی شده استفاده کرد که از گزارشات ساده شروع کرده و به سمت گزارشات پیچیده‌تر رفتند. ایجاد گزارش اول چندین هفته به طول انجامید ولی بقیه‌ی گزارش‌ها سریع‌تر تهیه شدند. موفقیت‌های زیر حاصل شده است:

- کاهش زمان آماده‌سازی داده‌ها از دو روز به دو ساعت.
- بینش گرافیکی شفاف در خصوص صدور صورتحساب و ترکیب صورتحساب‌ها برای مشتریان کسب‌وکار به کسب‌وکار.
- یک گزارش مدیریت بار کاری توسط تیم‌های عملیاتی. مدیران می‌توانند بار کاری تیم را به صورت هفتگی یا بلندمدت ارزیابی کرده و تعدیلات و اصلاحات لازم را به عمل آورند.

نوپ می‌گوید: «ما کارایی‌مان را به میزان چشمگیری بهبود داده‌ایم و می‌توانیم داده‌ها و گزارشات با کیفیت را با هزینه‌ی به مراتب کمتر و تواتر بیشتری تهیه و ارائه کنیم. و اگر شرکت بخواهد داده‌های بدست آمده از منابع مختلف را ترکیب کند، اینکار را براحتی می‌تواند انجام دهد. گزارشات دیداری بر اساس این داده‌گاه‌ها را می‌توان در عرض چند روز یا حتی چند ساعت ایجاد کرد.»
نوپ می‌گوید شرکت قصد دارد تا فعالیت خود را تا سطح رفتارهای دیجیتالی مشتریان گسترش داده و داده‌های بدست آمده از تحلیل‌های وب، نامه‌های الکترونیک و رسانه‌های اجتماعی را با داده‌های بدست آمده از سیستم‌های سمت سرور ترکیب کند. «نهایتاً، ما می‌خواهیم ساس و ویژوال آنالیتیکز را جایگزین تمامی فعالیت‌های گزارش‌دهی که به نیروی کار زیادی نیاز دارند کنیم.»
وی می‌گوید افزودن این انعطاف‌پذیری ساس و ویژوال آنالیتیکز برای دپارتمان‌های امری ضروری و حیاتی است. «اینکار به ما زمان بیشتری برای رسیدگی به سایر چالش‌های موجود می‌دهد. ما همچنین می‌خواهیم این ابزار را بر روی دستگاه‌های سیار (موبایل) نیز ارائه کنیم. مدیران ما به این ترتیب می‌توانند در زمان بازدید مشتریان از گزارشات به‌روز و وفق‌پذیر استفاده کنند. ما یک پلتفرم گزارش‌دهی آینده‌نگر داریم تا هر آنچه می‌خواهیم را انجام دهیم.»

سوالات برای مورد کاربردی ۳.۸

۱. به نظر شما چرا شرکت‌های تأمین انرژی از کاربران اصلی ابزارهای مصورسازی اطلاعات هستند؟
۲. شرکت الکترا بل چگونه از مصورسازی اطلاعات برای دستیابی به نسخه‌ی یکتا از حقیقت استفاده کرد؟
۳. چالش‌ها، راه‌حل پیشنهادی و نتایج بدست آمده چه بودند؟

Source: SAS Customer Story, "Visual Analytics Helps Energy Supplier Make Better Connections."

http://www.sas.com/en_us/customers/electrabel-be.html (accessed July 2018). Copyright © 2018 SAS Institute Inc., Cary, NC, United States. Reprinted with permission. All rights reserved.

اگرچه داشبوردهای عملکردی و سایر چارچوب‌های مصورسازی اطلاعات با یکدیگر تفاوت‌هایی دارند ولی تمامی آنها دارای ویژگی‌های طراحی مشترکی هستند. اول اینکه همه‌ی آنها زیرمجموعه‌ی سیستم اندازه‌گیری عملکرد و/یا هوش کسب‌وکار بزرگتری هستند بدین معنا که معماری زیربنایی آنها از معماری مدیریت عملکرد یا هوش کسب‌وکار یک سیستم بزرگتر است. دوم اینکه، کلیه‌ی داشبوردها و سایر مصورسازی‌های اطلاعات از ویژگی‌های زیر برخوردارند (ناول ۲۰۰۹):

- آنها از مؤلفه‌های دیداری (مثلاً نمودارها، گیج‌ها، میله‌های عملکردی، کنتورها، چراغ‌های راهنمایی) برای برجسته‌سازی داده‌ها در یک نگاه و موارد استثنایی که نیازمند اقدام است استفاده می‌کنند.
- برای کاربران، شفاف هستند یعنی اینکه به کمترین آموزش نیاز داشته و استفاده از آنها بی‌اندازه آسان است.
- داده‌های بدست آمده از سیستم‌های مختلف را در قالب یک منظر چکیده و متحد ترکیب می‌کنند.
- امکان مشاهده گزارشات یا منابع داده زیربنایی تا سطوح پایین و جزئیات بالا را به منظور مقایسه و ارزیابی فراهم می‌کنند.
- یک دید پویا و واقعی با به‌روزرسانی زمانبندی شده‌ی داده‌ها را فراهم کرده و به کاربر نهایی اجازه می‌دهند تا از آخرین تغییرات بوجود آمده در کسب‌وکار مطلع و آگاه باشد.
- پیاده‌سازی، جاری‌سازی و نگهداری آنها به کدنویسی بسیار کمی نیاز دارد.

این ضرب‌المثل معروف بنگاه املاک و مستغلات که می‌گویند « موقعیت، موقعیت، موقعیت » نشان می‌دهد که مهمترین ویژگی یک قطعه زمین یا ملک، موقعیت یا مکانی است که در آن واقع شده است. در خصوص داشبوردها باید بگوییم « داده‌ها، داده‌ها، داده‌ها ». داده‌ها که اغلب نادیده گرفته می‌شوند یکی از مهمترین چیزهایی هستند که در طراحی داشبوردها باید بر آنها تمرکز کرد (کاروتنوتو ۲۰۰۷). حتی اگر یک داشبورد، ظاهر حرفه‌ای و زیبایی داشته باشد و شامل جداول و گراف‌هایی باشد که بر اساس استانداردهای طراحی دیداری ایجاد شده‌اند، همچنان مهم است که در مورد داده‌ها بپرسیم: آیا داده‌ها قابل اطمینان هستند؟ آیا به‌روز هستند؟ آیا داده‌ها کامل هستند؟ آیا بین تمامی داشبوردها سازگار هستند؟ برخی از بهترین تجارب مبتنی بر تجربه در طراحی داشبوردها در اینجا ارائه شده است (رادها ۲۰۰۸).

بسیاری از مشتریان در یک مقطع زمانی می‌خواهند بدانند آیا سنجه‌هایی که اندازه‌گیری می‌کنند، سنجه‌های درست و مناسبی هستند یا خیر. مشتریان برخی اوقات متوجه شده‌اند که سنجه‌هایی که ردیابی می‌کنند سنجه‌های درستی نیستند. انجام یک ارزیابی شکاف با الگوهای صنعت می‌تواند شما را با بهترین تجارب صنعت همراه کند.

اغلب اوقات زمانیکه یک گزارش یا یک کارت امتیازی/داشبورد دیداری به مشتریان سازمان‌ها نمایش داده می‌شود، سوالاتی هستند که بی‌پاسخ می‌مانند. برخی از آنها عبارتند از :

- این داده‌ها را از چه منابعی بدست آورده‌اید؟
- در حین بارگذاری انبار داده‌ها، چه درصدی از داده‌ها مردود شدند یا اینکه با مشکلات کیفی داده‌ها مواجه شدند؟

- آیا این داشبورد اطلاعات « تازه » را نمایش می‌دهد یا خیر؟
- آخرین تاریخ به‌روزرسانی انبار داده چه زمان بوده است؟
- زمان بعدی که انبار داده به‌روزرسانی خواهد شد چه وقت است؟
- آیا هیچ تراکنش ارزش بالایی که روندهای کلی را دچار انحراف کند به عنوان بخشی از فرایند بارگذاری مردود شده است؟

در اکثر محیط‌های داشبورد‌ها، داشبوردها توسط یک متخصص ابزار و بدون توجه به اصول کاربردپذیری طراحی می‌شوند. اگرچه داشبوردها یک انبار داده با طراحی مهندسی هستند که به خوبی عمل می‌کنند ولی بسیاری از کاربران سازمانی به این خاطر که داشبوردها کاربرپسند نیستند از آنها استفاده نمی‌کنند که منجر به عدم بکارگیری زیرساخت‌ها و مشکلات مدیریت تغییر می‌شود. صحت‌گذاری طراحی داشبورد با استفاده از یک متخصص کاربردپذیری می‌تواند از این ریسک پیشگیری کند.

با دلیل وجود هزاران تَن داده‌ها خام، تعریف یک مکانیزم برای ارسال کنشگرانه‌ی رفتارها/استثنائات مهم به استفاده‌کنندگان اطلاعات بسیار اهمیت دارد. یک قاعده‌ی کسب‌وکار را می‌توان برای شناسایی الگوی هشدار مورد نظر کدنویسی کرد. این قاعده را می‌توان با استفاده از رویه‌های ذخیره شده در پایگاه داده در قالب یک برنامه کدنویسی کرد که در جداول داده‌ها جستجو کرده و الگوهایی که نیازمند توجه و رسیدگی فوری هستند را تشخیص دهد. بدین ترتیب، اینجا اطلاعات است که کاربر را پیدا می‌کند برخلاف روش‌های قبلی که کاربر در بین جداول داده‌ها به دنبال وقوع الگوهای بحرانی می‌گردد.

زمانیکه اطلاعات یکسانی توسط داشبوردها برای چندین کاربر ارائه می‌شود، یک جعبه متن کوچک را می‌توان در نظر گرفت که نظرات و دیدگاه‌های کاربران نهایی را اخذ کند. این جعبه متن را می‌توان اغلب بر روی داشبورد تعبیه کرد تا اطلاعات را اخذ کرده و دیدگاه‌ها را به شاخص‌های کلیدی عملکردی ساخته شده اضافه کند.

اطلاعات را می‌توان بسته به دانه‌بندی اطلاعات در سه سطح نمایش داد: سطح داشبورد دیداری، سطح گزارش ایستا و سطح مکعب سلف سرویس. زمانیکه یک کاربر در داشبورد به گشت و گذار می‌پردازد، مجموعه ساده‌ای از ۸ تا ۱۲ شاخص کلیدی عملکرد را می‌توان به وی ارائه کرد که نسبت به اینکه چه چیزهایی به خوبی کار می‌کنند و نمی‌کنند دید پیدا کند.

در ارائه و نمایش اطلاعات در یک داشبورد، بهترین روش نمایش برخی اطلاعات با استفاده از نمودارهای میله‌ای و برخی دیگر با گراف‌های خطی سری‌های زمانی و برای نمایش همبستگی‌ها با استفاده از یک نمودار پراکندگی است. برخی اوقات فقط کافی است که اطلاعات را در قالب جداول ساده نمایش دهیم. پس از اینکه اصول طراحی داشبورد به طور صریح مستندسازی شد، تمامی برنامه‌نویسانی که در قسمت سرویس گیرنده فعالیت می‌کنند می‌توانند به همین اصول پیابند بوده و گزارشات و داشبوردها را پردازش کنند.

در یک سازمان معمولی، کاربران ممکن است از لحاظ بلوغ تحلیلی در سطوح مختلفی قرار داشته باشند. از قابلیت داشبوردها می‌توان برای راهنمایی کاربران «متوسط» برای دسترسی به همان مسیری که سایر کاربران متخصص و خبره در حوزه تحلیل دسترسی دارند استفاده کرد.

۱. منظور از یک داشبورد اطلاعاتی چیست؟ چرا تا این اندازه محبوبیت دارد؟
۲. ویجت‌های گرافیکی که عمدتاً در داشبوردها استفاده می‌شوند کدام‌ها هستند؟ چرا؟
۳. سه لایه‌ی اطلاعاتی که در داشبوردها ارائه می‌شود را نام برده و توضیح دهید.
۴. مشخصه‌های مشترک داشبوردها و سایر تحلیل‌های دیداری اطلاعاتی چه هستند؟
۵. بهترین تجارب در طراحی داشبوردها چه هستند؟

- داده‌ها تبدیل به ارزشمندترین دارایی سازمان‌های امروزی شده‌اند.
- داده‌ها، جزء تشکیل دهنده‌ی اصلی هر راهکار هوش کسب‌وکار، علم داده‌ها و تحلیل‌های کسب‌وکار هستند.
- اگرچه ارزش قابل ارائه‌ی داده‌ها انکارناپذیر است ولی باید با برخی سنجه‌های کیفی و کاربردی‌پذیری پایه سازگار باشد.
- عبارت داده به مجموعه‌ای از حقایق اشاره دارد که معمولاً در نتیجه‌ی آزمایشات، مشاهدات، تراکنش‌ها یا تجارب بدست می‌آید.
- در بالاترین سطح چکیدی، داده‌ها را می‌توان به صورت ساختنیافته و ساختنیافته دسته‌بندی نمود.
- داده‌ها به شکل اصلی/خام معمولاً آماده‌ی این نیستند که در وظایف تحلیلی بکار برده شوند.
- پیش‌پردازش داده‌ها کار بسیار زمان‌بر و طاقت‌فرسا ولی حیاتی در تحلیل کسب‌وکار است.
- آمارها مجموعه‌ای از تکنیک‌های ریاضی برای توضیح و تفسیر داده‌ها هستند.
- روش‌های آماری را می‌توان در قالب توصیفی یا استنتاجی دسته‌بندی کرد.

- آمارها به طور کلی و آمارهای توصیفی به طور خاص، بخشی بحرانی از تحلیل کسب و کار و هوش کسب و کار است.
- از روش‌های آماری توصیفی می‌توان برای سنجش گرایش به مرکز، پراکندگی یا شکل مجموعه داده‌های مفروض استفاده کرد.
- رگرسیون و به ویژه رگرسیون خطی را شاید بتوان شناخته‌شده‌ترین و پرکاربردترین تکنیک تحلیل آماری دانست.
- رگرسیون خطی و رگرسیون لجستیک دو نوع رگرسیون اصلی در آمار هستند.
- رگرسیون لجستیک یک الگوریتم دسته‌بندی مبتنی بر احتمالات است.
- سری‌های زمانی، یک توالی از نقاط داده‌ی یک متغیر هستند که در نقاط متوالی زمانی با بازه‌های زمانی یکسان، اندازه‌گیری و ثبت می‌شوند.
- یک گزارش، یک مصنوع ارتباطی است که با هدف تحویل اطلاعات به شکلی قابل ارائه تهیه می‌شود.
- یک گزارش کسب و کار، یک مستند مکتوب است که حاوی اطلاعاتی در خصوص مسائل کسب و کار است.
- کلید هر گزارش کسب و کار موفق، شفافیت، موجز بودن، کامل بودن و صحیح بودن است.
- مصورسازی داده‌ها عبارتست از استفاده از نمایش دیداری داده‌ها برای اکتشاف، معنا دادن و انتقال دادن داده‌ها.
- شاید برجسته‌ترین گرافیک اطلاعاتی در گذشته توسط شارل ژئی مینار تهیه شده که به لحاظ گرافیکی خسارت‌های وارده بر ارتش ناپلئون در جبهه‌ی روسیه در سال ۱۸۱۲ را به تصویر کشید.
- انواع نمودارهای پایه شامل خطی، میله‌ای و کلوچه‌ای هستند.
- نمودارهای تخصصی اغلب از نمودارهای پایه و به عنوان حالت‌های استثنایی مشتق می‌شوند.
- تکنیک‌ها و ابزارهای مصورسازی داده‌ها، کاربران تحلیل‌های کسب و کار و سیستم‌های هوش کسب و کار را به مصرف‌کنندگان بهتر اطلاعات تبدیل می‌کنند.
- تحلیل‌های دیداری ترکیبی از مصورسازی و تحلیل‌های پیشگویانه هستند.
- تقاضای روزافزون تحلیل‌های دیداری به همراه حجم فزاینده‌ی داده‌ها منجر به رشد نمایی سرمایه‌گذاری‌های انجام شده در سیستم‌های مصورسازی بسیار کارآمد شده است.
- داشبوردها، نمایش دیداری از اطلاعات مهم را به صورت ادغام و تجمیع شده در یک صفحه نمایش می‌دهند تا بتوان اطلاعات را در یک نگاه درک نموده و برای جزئیات بیشتر در آنها به اکتشاف پرداخت.

آماده برای تحلیل	پراکندگی	نمودار کلوچه‌ای
میانگین حسابی	رایانش با عملکرد بالا	چارک
نمودار جعبه‌ای	هیستوگرام	دامنه
نمودار حبابی	آمار استنتاجی	داده‌های نسبی
گزارش کسب و کار	شاخص‌های کلیدی عملکرد	رگرسیون
داده‌های رسته‌ای	دانش	گزارش
مرکزیت	کشیدگی	نمودار پراکندگی

همبستگی	یادگیری	چولگی
داشبوردها	رگرسیون خطی	انحراف استاندارد
پیش‌پردازش داده‌ها	رگرسیون لجستیک	آمار
کیفیت داده‌ها	میانگین قدرمطلق انحرافات	داستان‌سرایی
امنیت داده‌ها	میانه	داده‌های ساخت‌یافته
طبقه‌بندی داده‌ها	مد	پیش‌بینی سری‌های زمانی
مصورسازی داده‌ها	داده‌های اسمی	داده‌های ساخت‌نیافته
داده	پردازش تحلیلی آنلاین	انتخاب متغیر
آمارهای توصیفی	داده‌های ترتیبی	واریانس
تقلیل ابعاد	مربعات حداقلی معمولی	تحلیل دیداری

۱. اهمیت تحلیل داده‌ها را چگونه توضیح می‌دهید؟ آیا می‌توانیم تحلیل‌ها را بدون داده‌ها تصور کنیم؟ توضیح دهید.
۲. با در نظر گرفتن تعریف جدید و وسیع تحلیل کسب‌وکار، ورودی‌ها و خروجی‌های اصلی تحلیل‌ها چه می‌باشند؟
۳. داده‌های مورد استفاده برای تحلیل کسب‌وکار از کجا می‌آیند؟ منابع و ماهیت داده‌های ورودی چه هستند؟
۴. متداول‌ترین سنجه‌ها برای آماده‌سازی داده‌ها جهت تحلیل چه هستند؟
۵. دسته‌بندی‌های اصلی داده‌ها کدام‌ها هستند؟ از چه نوع داده‌هایی می‌توانیم برای هوش کسب‌وکار و تحلیل‌ها استفاده کنیم؟
۶. آیا می‌توانیم از نمایش یکسان داده‌ها برای تمامی مدل‌های تحلیل استفاده کنیم (یعنی آیا مدل‌های تحلیلی مختلف، الگوها و طرح‌های نمایش داده‌های متفاوتی دارند؟) چرا؟
۷. چرا داده‌های اولیه/خام را نمی‌توان بلافاصله در تحلیل‌ها بکار برد؟
۸. گام‌های اصلی پیش‌پردازش داده‌ها چه هستند؟ آنها را نام برده و اهمیت آنها در تحلیل را توضیح دهید.
۹. پاکسازی داده‌ها به چه معناست؟ در این مرحله چه فعالیت‌هایی انجام می‌شود؟
۱۰. تقلیل داده‌ها را می‌توان بر روی ردیف‌ها (نمونه‌برداری) و/یا ستون‌ها (انتخاب متغیر) انجام داد. کدامیک چالش‌برانگیزتر است؟
۱۱. چه ارتباطی بین آمار و تحلیل کسب‌وکار وجود دارد؟ (جایگاه آمار در یک طبقه‌بندی تحلیل کسب‌وکار را در نظر بگیرید).
۱۲. تفاوت‌های اصلی بین آمار توصیفی و استنتاجی چه هستند؟
۱۳. منظور از یک نمودار جعبه‌ای چیست؟ چه نوع اطلاعات آماری را نمایش می‌دهد؟
۱۴. دو مشخصه‌ی شکلی پرکاربرد برای توضیح یک توزیع داده کدامها هستند؟
۱۵. سنجه‌های گرانش به مرکز آمارهای توصیفی را نام برده و مختصراً تعریف کنید.
۱۶. تشابهات و تفاوت‌های بین رگرسیون و همبستگی چه هستند؟
۱۷. گام‌های اصلی توسعه‌ی یک مدل رگرسیون خطی را نام برده و توضیح دهید.

۱۸. متداول‌ترین مفروضات در رگرسیون خطی چه هستند؟ موارد حیاتی در مدل‌های رگرسیون چه می‌باشند؟
۱۹. تشابهات و تفاوت‌های بین رگرسیون خطی و رگرسیون لجستیک چه هستند؟
۲۰. سری‌های زمانی چه هستند؟ تکنیک‌های اصلی پیش‌بینی برای داده‌های سری‌های زمانی چه هستند؟
۲۱. منظور از یک گزارش کسب‌وکار چیست؟ چرا به آن نیاز داریم؟
۲۲. بهترین تجارب در گزارش‌دهی کسب‌وکار چه هستند؟ چگونه می‌توانیم گزارشات خودمان را تهیه کنیم؟
۲۳. فرایند چرخه‌ای مدیریت را توضیح داده و در خصوص نقش گزارشات کسب‌وکار اظهار نظر کنید.
۲۴. سه گروه اصلی گزارشات کسب‌وکار را نام برده و توضیح دهید.
۲۵. چرا مصورسازی اطلاعات به یک بخش مرکزی در هوش کسب‌وکار و تحلیل کسب‌وکار تبدیل شده است؟ آیا بین مصورسازی اطلاعات و تحلیل دیداری تفاوت است؟
۲۶. انواع اصلی نمودارها/گراف‌ها چه هستند؟ چرا تعداد زیادی از آنها وجود دارد؟
۲۷. چگونه می‌توانید نمودار درست را برای یک کار انتخاب کنید؟ دلایل خودتان را توضیح داده و از آنها دفاع کنید.
۲۸. تفاوت بین مصورسازی اطلاعات و تحلیل دیداری در چیست؟
۲۹. چرا داستان‌سرایی باید بخشی از فرایند گزارش‌دهی و مصورسازی داده‌ها باشد؟
۳۰. منظور از یک داشبورد اطلاعاتی چیست؟ معرف چه چیزی است؟
۳۱. بهترین تجارب در طراحی داشبوردهای با کیفیت و سطح بالا چه هستند؟
۳۲. آیا به نظر شما داشبوردهای اطلاعاتی/عملکردی همواره وجود خواهند داشت؟ یا اینکه روزی منسوخ خواهند شد؟ موج بزرگ بعدی در حوزه‌ی تحلیل کسب‌وکار و هوش کسب‌وکار از منظر مصورسازی داده‌ها/اطلاعات چه خواهد بود؟

تمرینات

تمرینات دانشگاه ترادیتا

۱. داده‌های « Voting Behavior » و شرح مختصر داده‌ها را از وب سایت کتاب دانلود کنید. این مجموعه‌ای از داده‌ها است که از تمامی کشورهای اطراف ایالات متحده گردآوری شده است. داده‌ها به صورت جزئی پردازش شده‌اند به این معنی که متغیرهای مشتق شده‌ای ایجاد شده‌اند. وظیفه‌ی شما این است که داده‌ها را از طریق شناسایی خطاها و ناهنجاری‌ها به طور کامل شناسایی کرده و راه‌حل ارائه کنید. در پایان، شما باید نسخه‌ی آماده‌ی تحلیل این داده‌ها را داشته باشید. پس از تکمیل پیش‌پردازش، این داده‌ها را وارد نرم‌افزار تابلو (یا ابزار نرم‌افزاری مصورسازی داده‌های دیگر) کنید تا اطلاعات دیداری مفید را از آن استخراج کنید. برای اینکار، سوالات و مفروضات مرتبط را تدوین کنید (حداقل سه مورد) و مصورسازی‌های مناسب را بتوانند به آن سوالات آزمون فرض پاسخ دهند را انجام دهید.
۲. نرم‌افزار تابلو را از tableau.com دانلود کنید. با استفاده از مجموعه داده‌های نمونه Visualisation MFG به سوالات زیر پاسخ بدهید:

- ا. ارتباط بین درآمد ناخالص گیشه و سایر پارامترهای مرتبط با فیلم در این مجموعه داده چیست؟
- ب. این ارتباط در سال‌های مختلف چگونه تغییر می‌کند؟
- یک گزارش مکتوب حرفه‌ای تهیه کنید که اسکرین‌شات‌هایی از یافته‌های گرافیکی شما در آن آورده شده باشد.
۳. به سایت teradatauniversitynetwork.com بروید. به دنبال مقاله‌ای باشید که با ماهیت داده‌ها، مدیریت داده‌ها و/یا حکمرانی داده‌ها که مرتبط با هوش کسب‌وکار و تحلیل سر و کار داشته باشد و محتوای مقاله را تحلیل کنید.
۴. به مخزن داده‌ی UCI (archive.ics.uci.edu/ml/datasets.html) رفته و مجموعه داده‌های بزرگی که حاوی مقادیر عددی و اسمی باشد را شناسایی کنید. با استفاده از میکروسافت اکسل یا هر نرم‌افزار آماری دیگر:
- ا. سنجه‌های گرایش به مرکز هر متغیر را محاسبه و تفسیر کنید.
- ب. سنجه‌های گستردگی/پراکندگی هر متغیر را محاسبه و تفسیر کنید.
۵. به مخزن داده‌های UCI رفته (archive.ics.uci.edu/ml/datasets.html) و دو مجموعه داده را شناسایی کنید، یکی برای تخمین/رگرسیون و دیگری برای دسته‌بندی. با استفاده از میکروسافت اکسل یا هر نرم‌افزار آماری دیگر:
- ا. یک مدل رگرسیون خطی را توسعه داده و تفسیر کنید.
- ب. یک مدل رگرسیون لجستیک را توسعه داده و تفسیر کنید.
۶. به سایت KDnugget.com رفته و با طیف منابع تحلیل موجود در این پورتال آشنا شوید. سپس یک مقاله، یا یک متن مصاحبه مرتبط با ماهیت داده‌ها، مدیریت داده‌ها و/یا حکمرانی داده‌ها که مرتبط با هوش کسب‌وکار و تحلیل سر و کار داشته باشد و محتوای مقاله را تحلیل کنید.
۷. به وبلاگ استفان فیو به نام (perceptualedge.com) "The Perceptual Edge" بروید. به بخش "Examples" بروید. در این بخش، او نقدهایی در مورد داشبوردهای مختلف می‌نویسد. برخی از این مثال‌ها را مطالعه کنید. اکنون به Dundas.com بروید. بخش "Gallery" سایت را انتخاب کنید. سپس بر روی "Digital Dashboard" کلیک کنید. دموهای بسیار متنوع و مختلفی از داشبوردها برای شما نمایش داده می‌شود. برخی از آنها را اجرا کنید.
- ا. چه نوع سنجه‌ها و اطلاعاتی در این دموها نمایش داده می‌شود؟ چه اقداماتی می‌توانید انجام دهید؟
- ب. با استفاده از برخی از مفاهیم پایه از نقدهای فیو، برخی از نکات خوب و نکات بد در طراحی داشبوردها را ذکر کنید.
۸. یک ابزار مصورسازی اطلاعات مثل تابلو، کلیک‌ویوو یا اسپات‌فایر را دانلود کنید. اگر دانشگاه شما یک توافقنامه‌ی آموزشی با این شرکت‌ها دارد، یک نسخه‌ی آزمایشی برای این تمرین کافی خواهد بود. از داده‌های خودتان یا مجموعه داده‌هایی که به همراه این ابزار نرم‌افزاری ارائه می‌شود استفاده کنید. داده‌ها را بررسی کرده، چندین مسئله‌ی کسب‌وکار را مطرح کرده و از مصورسازی داده‌ها برای تحلیل، مصورسازی و حل این مسائل استفاده کنید.
۹. به سایت teradatauniversitynetwork.com بروید. بخش "Tableau Software Project" را پیدا کنید. توضیحات را خوانده و وظایف را انجام دهید و به سوالات پاسخ دهید.
۱۰. به سایت teradatauniversitynetwork.com بروید. تمرینات مربوط به ais.teradata.com و aiis.teradata.com را پیدا کنید. با استفاده از اطلاعات و دستورالعمل‌های گام به گام که در این تمرین آمده؛ تحلیل را بر روی ابزار ais.teradata.com و aiis.teradata.com آنالیتیکز اجرا کنید (یک سیستم وبی که نیاز به نصب ندارد). به سوالات مطرح شده در این تمرین پاسخ دهید.

۱۱. حداقل دو مقاله پیدا کنید که در مورد داستان‌سرایی و به ویژه در مقوله‌ی تحلیل‌ها صحبت می‌کند. مقالات را مطالعه کرده و تحلیل کنید و گزارشی از یافته‌هایتان و نظرات‌تان در مورد اهمیت داستان‌سرایی در هوش کسب‌وکار و تحلیل کسب‌وکار بنویسید.

۱۲. به سایت **data.gov** که یک پورتال داده‌های دولتی است که تعداد زیادی مجموعه داده در مورد عناوین و موضوعات مختلف از بهداشت و درمان یا آموزش، آب و هوا و ایمنی عمومی دارد بروید. موضوعی که برایتان بسیار جذاب را انتخاب کنید. در اطلاعات و توضیحات ارائه شده در سایت به گشت و گذار بپردازید. در مورد امکانات دانلود داده‌ها بررسی کرده و از ابزار مصورسازی داده‌های محبوب‌تان برای مصورسازی اطلاعات استفاده کنید.

تمرینات تیمی و پروژه‌های ایفای نقش

۱. تحلیل‌ها با داده‌ها شروع می‌شوند. شناسایی، دسترسی، دستیابی و پردازش داده‌های مرتبط، ضروری‌ترین وظیفه در هر مطالعه‌ی تحلیل می‌باشد. به عنوان یک تیم، شما موظف به یافتن حجم انبوهی از داده‌های واقعی (ترجیحاً از سازمان‌تان یا از اینترنت که می‌تواند از یک جستجوی ساده یا پیوندهای داده که بر روی **KDnuggets.com** منتشر شده است) است که ده‌ها هزار ردیف و بیش از ۲۰ متغیر دارد و یک پروژه‌ی پیش‌پردازش کامل از داده‌ها را مستند کنید. در پردازش داده‌ها، ناهنجاری‌ها و پراکندگی‌ها را با استفاده از روش‌ها و سنج‌های آماری یافته و تحلیل داده‌ها را آماده کنید. گام‌ها و تصمیمات پیش‌پردازش خودتان را در یک گزارش جامع نام برده و توضیح دهید.

۲. به وب سایت شرکت معروف عرضه‌کننده‌ی داشبوردهای اطلاعاتی بروید (**dundas.com, idashboards.com, enterprisedashboard.com**). این سایت‌ها مثال‌هایی از داشبوردهای اجرایی ارائه می‌کنند. به عنوان تیم، یک صنعت خاص (مثلاً سلامت و درمان، بانکداری، خطوط هوایی) را انتخاب کنید. داشبوردهای مرتبط با آن صنعت را پیدا کنید. انواع سنج‌های موجود در داشبوردها را شناسایی کنید. از چه نوع ویژگی‌های نمایشی و دیداری برای ارائه اطلاعات استفاده شده است؟ از آنچه در مورد طراحی داشبوردها می‌دانید یک نمونه‌ی کاغذی از داشبوردها را برای این اطلاعات ایجاد کنید.

۳. به سایت **teradatauniversitynetwork.com** بروید. از آنجا به منابع داده‌ی دانشگاه آرکانزاس بروید. یکی از مجموعه داده‌های بزرگ را انتخاب کرده و تعداد زیادی از رکوردها را دانلود کنید (ممکن است اینکار نیازمند این باشد که دستورهای اس‌کیوال را برای ایجاد متغیرهایی که شامل این مجموعه داده‌ها باشند را بنویسید). حداقل ۱۰ سوال را طرح کنید که بتوانید با این مصورسازی اطلاعات به آنها پاسخ دهید. با استفاده از ابزار محبوب مصورسازی داده‌هایتان (مثلاً تابلو)، داده‌ها را تحلیل کرده و یک گزارش مبسوط که شامل اسکرین‌شات و سایر ویژگی‌های دیداری باشد تهیه کنید.

۴

الگوریتم‌ها، روش‌ها و فرایند داده‌کاوی

به بیان کلی، داده‌کاوی روشی برای توسعه‌ی هوش (دانش یا اطلاعات قابل اقدام) از داده‌هایی است که یک سازمان گردآوری، سازماندهی و ذخیره کرده است. سازمان‌ها از طیف گسترده‌ای تکنیک‌های داده‌کاوی برای دستیابی به درک بهتری از مشتریان و عملیات خود و حل مسائل سازمانی پیچیده استفاده می‌کنند. در این فصل، ما به مطالعه‌ی داده‌کاوی به عنوان یک فناوری توانمندساز برای تحلیل کسب‌وکار و تحلیل پیشگویانه می‌پردازیم؛ در مورد فرایندهای استاندارد انجام پروژه‌های داده‌کاوی یاد می‌گیریم؛ در بکارگیری تکنیک‌های اصلی داده‌کاوی به تخصص دست پیدا می‌کنیم؛ در خصوص ابزارهای نرم‌افزاری موجود به آگاهی دست پیدا می‌کنیم و به بحث و بررسی در

- [داده‌کاوی چیست؟](#)
- [داده‌کاوی چگونه کار می‌کند؟](#)
- [داده‌کاوی چه مزایایی دارد؟](#)
- [داده‌کاوی چه چالش‌هایی دارد؟](#)

خصوص مسائل حریم خصوصی، افسانه‌های معمول و موانع و ضعف‌های مرتبط با داده‌کاوی می‌پردازیم.

۴.۱. پرونده‌ی آغازین: دپارتمان پلیس میامی-دید از تحلیل پیشگویانه برای پیش‌بینی و مبارزه با جرائم استفاده می‌کند.

۴.۲. مفاهیم داده‌کاوی

۴.۳. کاربردهای داده‌کاوی

۴.۴. فرایند داده‌کاوی

۴.۵. روش‌های داده‌کاوی

۴.۶. ابزارهای نرم‌افزاری داده‌کاوی

۴.۷. مسائل حریم خصوصی، افسانه‌ها و اشتباهات بزرگ داده‌کاوی

تحلیل‌های پیشگویانه و داده‌کاوی تبدیل به یک جزء لاینفک از هر نهاد اعمال قانون از جمله دپارتمان پلیس میامی-دید^۱ شده‌اند که مأموریت آنها نه تنها محافظت از امنیت بزرگترین بخش ایالت فلوریدا با ۲.۵ میلیون شهروند (هفتمین بخش بزرگ در ایالات متحده) بلکه برقراری فضایی امن و دلپذیر برای میلیون‌ها گردشگری است که از سرتاسر جهان برای استفاده از زیبایی‌های طبیعت، آب و هوای گرم و سواحل چشم‌نواز به آنجا می‌آیند است. با هزینه‌های نزدیک به ۲۰ میلیارد دلاری که گردشگران هرساله در آنجا دارند و تقریباً یک سوم مالیات فروش فلوریدا را تولید می‌کند، نمی‌توان اهمیت گردشگری بر اقتصاد منطقه را کتمان کرد. اگرچه تعداد کمی از افسران پلیس هستند که توسعه اقتصادی را در شرح وظایف خود ذکر می‌کنند تقریباً همه‌ی آنها به پیوند حیاتی بین خیابان‌های امن و رونق گردشگری منطقه اذعان دارند.

این ارتباط برای ستوان آرنلود پالمرا^۲ که در حال حاضر، سرپرست بخش تحقیقات سرقت و سرپرست سابق مداخله سرقت‌ها است بسیار حیاتی است. این تیم متخصص از کارآگاهان تمرکزشان بر نقاط حساس و مستعد سرقت و مجرمان باسابقه است. او و تیمش در در دفاتر طبقه‌ی دوم یک ساختمان بتنی مدرن در کرانه‌ی غربی میامی اقامت دارند. او در ۱۰ سال حضورش در این واحد و ۲۳ تیم شاهد تغییرات بسیاری بوده است که فقط مربوط به شیوه‌های حراست مثل روشی که تیم او برای مشخص کردن نقاط حساس و جرم‌خیز خیابانی از دکمه‌های رنگی بر روی نقشه استفاده می‌کردند نبوده است.

پالمرا و تیمش نیز در گشت‌زنی‌های خیابانی شاهد تأثیر جمعیت روزافزون، تغییرات جمعیت‌شناختی و تغییر اقتصاد بوده‌اند. همانند هر افسر خوب نیروی پلیس، آنها نیز روش‌هایشان را برای مواجهه با چالش‌های حراست که از لحاظ دامنه و پیچیدگی بر آنها افزوده شده است به طور مستمر وفق داده‌اند. اما همانند تقریباً تمامی شاخص‌های دولتی، شدت گرفتن فشارهای بودجه موجب شده تا این دپارتمان بین افزایش تقاضا و کاهش منابع تحت فشار قرار بگیرد.

پالمرا که از دیدگاه او، کارآگاهان به منزله‌ی رزمندگان خط مقدم در برابر موج فزاینده‌ی جرائم خیابانی و کاهش مداوم و همیشگی منابع هستند اینگونه می‌گویند: «چالش اصلی ما این بود که چگونه با توجه به کاهش منابع و تعداد پلیس‌ها در خیابان‌ها، جرائم خیابانی را کاهش دهیم. طی سال‌های گذشته، این تیم سعی کرده است تا ابزارهای جدیدی را که بارزترین آنها برنامه‌ای به نام «اعمال قانون بر اساس تحلیل» بوده است را بکار بگیرد که از داده‌های تاریخی جرائم به عنوان مبنای تعیین موقعیت و جانمایی تیم‌های کارآگاهان استفاده می‌کرده است. «توانایی ما برای پیش‌بینی محل احتمال وقوع سرقت‌ها هم از طریق تحلیل و هم تجربه‌ی جمعی ما از آن زمان تا کنون ارتقاء یافته است.»

چالش عجیب‌تر پالمرا و تیم بازپرسان او که با نیروی پلیس تمامی نواحی روستایی مهم نیز مشترک بوده است، بستن بغرنج‌ترین پرونده‌ها بوده که هیچگونه سرنخ، شاهد، ویدئو – هرگونه مدارک یا حقایق که بتوانند به حل پرونده کمک کنند – وجود ندارد. تعجب‌برانگیز نیست که چون از روش‌های استاندارد مثل صحبت با خبرچین‌ها یا افسران گشت‌زنی یا مردم جامعه برای ایجاد سرنخ استفاده می‌شد، چندان تغییری در نتایج بوجود نیامده است. این رویکردها درست است که جوابگو هستند ولی بر تجربه‌ی

¹ Miami-Dade

² Arnold Palmer

کارآگاهان متکی هستند که تمامی اینها در ذهن آنها است. زمانیکه یک کارآگاه بازنشسته شده یا جابجا می‌شود، آن تجربه نیز به همراه وی از دست می‌رود.»

مسئله‌ی بفرنج پالمیر این بود که نرخ گردش پرسنل ناشی از بازنشستگی بسیاری از مجرب‌ترین بازرپرس‌ها و کارآگاهانش روندی صعودی پیدا کرده بود. درست است که نیروهای تازه نفس وارد نیروی پلیس می‌شدند و این نکته مثبتی بود مخصوصاً این گروه از افراد راحت‌تر می‌توانستند با انواع جدید اطلاعات از قبیل نامه‌های الکترونیک، رسانه‌های اجتماعی و دوربین‌های کنترل ترافیک کار کنند. اما همانگونه که پالمیر تعریف می‌کند، مشکل زمانی به وجود آمد که تعداد بسیار زیادی از کارآگاهان جدید به این واحد پیوستند که به دنبال راهنمایی افسران ارشد خود بودند ولی این افسران، دیگر حضور نداشتند. « ما در آن مقطع زمانی می‌دانستیم که برای پر کردن شکاف تجربه و حرکت به جلو به روش متفاوتی نیاز داشتیم. »

تلاش‌های موقت وی منجر به راه‌حلی شد که بر اساس آن کارآگاهان جدید که وارد گروه شده بودند همان سوالاتی که از افسران ارشد و با تجربه می‌پرسیدند را از یک پایگاه داده‌ی رایانه‌ای طرح می‌کردند. این تفکر، بذری را در ذهن پالمیر کاشت که دیگر از ذهنش بیرون نرفت.

آنچه در درون واحد ضد سرقت در حال شکل‌گیری بود نشان می‌داد که چگونه افکار و نظرات بزرگ می‌توانند از مکان‌های کوچکی آغاز شوند. اما مهمتر اینکه، نشان داد که برای اینکه این افکار و نظرات به بار بنشینند، شرایط «درست» باید همسو با یکدیگر و در زمان درست فراهم شوند. در سطح رهبری، این به معنای یک پیشبرنده‌ی سازمان است که حمایت بالا به پایین و همچنین پذیرش پایین به بالا از رده‌های مختلف سازمان در عین توجه به پرسنل دپارتمان فناوری اطلاعات چه تلاش و توجهی را می‌طلبد و این مسئولیت بر عهده‌ی پالمیر بود. در سطح سازمانی، واحد مبارزه با سرقت با توجه به شیوع مجرمان باسابقه در بین مرتکبین جرم و جنایت، عنوان یک نقطه‌ی شروع خوب برای پیشبرد مدلسازی عمل کرد. نهایتاً، توانایی دپارتمان برای استفاده از توان متحول‌ساز بالقوه‌ی بیشتر تا حد زیادی به توانایی تیم برای ارائه نتایج در مقیاس کوچک بستگی دارد.

با توجه به اینکه آزمون‌ها و دموهای اولیه دلگرم‌کننده بودند - زمانیکه جزئیات پرونده‌های حل شده به مدل داده شد، نتایج دقیق‌تری را ارائه کرد - این تیم توجه‌ها را به سمت خود جلب کردند. زمانیکه سرگرد و سرهنگ واحد مبارزه با سرقت حمایت خود را از راهبری پروژه اعلام کردند، بر سرعت پیشرفت کار افزوده شد و به پالمیر می‌گفتند که اگر بتوانید آن را راه بیندازید می‌توانید با آن کار کنید.» پالمیر توضیح می‌دهد که ولی مهمتر از تشویق‌ها، تمایل آنها جلب حمایت افراد رده بالای دپارتمان از پروژه بود. پالمیر می‌گوید: « اگر مافوق‌ها موضوع را نمی‌پذیرفتند من نمی‌توانستم کار را پیش ببرم. به همین خاطر حمایت آنها اهمیت حیاتی داشت. »

predictive analytics software. The software is designed to help businesses make better decisions by analyzing data and identifying trends. It can be used in a variety of industries, including marketing, sales, and customer service. The software is easy to use and can be integrated with existing systems. It provides a comprehensive view of your data and helps you understand your customers better. The software is also scalable and can handle large amounts of data. It is a powerful tool for businesses looking to improve their performance and make data-driven decisions.

The software is designed to help businesses make better decisions by analyzing data and identifying trends. It can be used in a variety of industries, including marketing, sales, and customer service. The software is easy to use and can be integrated with existing systems. It provides a comprehensive view of your data and helps you understand your customers better. The software is also scalable and can handle large amounts of data. It is a powerful tool for businesses looking to improve their performance and make data-driven decisions.

The software is designed to help businesses make better decisions by analyzing data and identifying trends. It can be used in a variety of industries, including marketing, sales, and customer service. The software is easy to use and can be integrated with existing systems. It provides a comprehensive view of your data and helps you understand your customers better. The software is also scalable and can handle large amounts of data. It is a powerful tool for businesses looking to improve their performance and make data-driven decisions.

¹ Predictive Analytics Lead Modelling Software

سوالاتی برای مرور پرونده‌ی آغازین

۱۳. چرا نهادها و دپارتمان‌های اعمال قانون مثل دپارتمان پلیس میامی-دید از داده‌کاوی و تحلیل‌های هوشمند استقبال می‌کنند؟
۱۴. چالش‌های اساسی دپارتمان‌ها و نهادهای اعمال قانون همانند دپارتمان پلیس میامی-دید چیست؟ آیا می‌توانید چالش‌های دیگری را (که در این پرونده ذکر نشده است) نیز ذکر کنید که به کمک داده‌کاوی بتوان آنها را رفع نمود؟
۱۵. منابع داده‌ای که دپارتمان‌ها و نهادهای اعمال قانون مثل دپارتمان پلیس میامی-دید برای مدل‌سازی پیشگویانه و پروژه‌های داده‌کاوی خود بکار می‌گیرند کدام‌ها هستند؟
۱۶. دپارتمان‌ها و نهادهای اعمال قانون مثل دپارتمان پلیس میامی-دید از چه نوع تحلیل‌هایی برای مبارزه با جرم و جنایت استفاده می‌کنند؟
۱۷. در این پرونده، «تصویر بزرگ از مکان‌های کوچک آغاز می‌شود» به چه معناست؟ توضیح دهید.

دپارتمان‌ها و نهادهای اعمال قانون، با توجه به محدودیت منابع برای اجرای مأموریت حراست و حفاظت از مردم شدیداً تحت فشار هستند. محیطی که در آن به انجام وظایف خود می‌پردازند روز به روز چالش‌برانگیزتر می‌شود به نحویکه باید برای پیشگیری از بروز فجایع، خود را به طور مداوم وفق داده یا شاید حتی چند قدم جلوتر باشند. درک تغییرات ماهوی جرائم و مجرمان، یک چالش مستمر و همیشگی است. در میان این چالش‌ها، آنچه به نفع این نهادها کار می‌کند، در دسترس بودن داده‌ها و فناوری‌های تحلیلی برای تحلیل بهتر رخدادهای گذشته و پیش‌بینی رویدادهای آینده است. در حال حاضر نسبت به گذشته داده‌های بسیار بیشتری در دسترس ما قرار دارد. بکارگیری تحلیل‌های پیشرفته و ابزارهای داده‌کاوی (یعنی تکنیک‌های کشف دانش) برای این منبع داده‌های غنی و بزرگ، بینش و ادراک لازم برای تدارک بهتر و اقدام بر اساس وظایف‌شان را در اختیار آنها قرار داده است. از این رو، نهادهای اعمال قانون تبدیل به کاربران پیشگام این چهره‌ی جدید تحلیل‌ها شده‌اند. داده‌کاوی، یک گزینه‌ی اصلی برای مدیریت و درک بهتر این وظایف حیاتی با سطح دقت بسیار بالا و به موقع است. مطالعه‌ی ارائه شده در اینجا، توان تحلیل‌ها و داده‌کاوی برای ایجاد یک دید کل‌نگر از دنیای جرم و جنایت برای مدیریت و واکنش بهتر و سریع‌تر را به روشنی به تصویر کشیده است. در این فصل، انواع مختلف کاربردهای داده‌کاوی برای حل مسائل پیچیده در صنایع و سازمان‌های مختلف را خواهید دید که از داده‌ها برای کشف بینش‌های اقدام‌پذیر برای بهبود کارایی عملیاتی، آمادگی مأموریت و مزیت رقابتی استفاده می‌کنند.

Sources: "Miami-Dade Police Department: Predictive modeling pinpoints likely suspects based on common crime signatures of previous crimes," IBM Customer Case Studies. www-03.ibm.com/software/businesscasestudies/om/en/corp?synkey=C894638H25952N07; "Law Enforcement Analytics: Intelligence-Led and Predictive Policing by Information Builder." www.informationbuilders.com/solutions/gov-lea.

داده‌کاوی که یک فناوری نسبتاً جدید و جالب است تبدیل به روشی متداول برای اکثریت قریب به اتفاق سازمان‌ها شده است. در مصاحبه‌ای با مجله‌ی دنیای رایانه در ژانویه ۱۹۹۹، دکتر آرنو پنزیاس^۱ (دارنده جایزه نوبل و دانشمند سرپرست سابق آزمایشگاه‌های بل)، داده‌کاوی را به عنوان یک کاربرد کلیدی پایگاه‌های داده سازمانی در آینده‌ی نزدیک شناسایی کرد. در پاسخ به سوال بسیار قدیمی « برنامه‌های کاربردی کُشده در سازمان چه خواهند بود؟ » دکتر پنزیاس پاسخ داد، « داده‌کاوی » و سپس افزود « داده‌کاوی اهمیت بسیار بیشتری خواهد یافت و شرکت‌ها هیچ اطلاعاتی از مشتریان خود را دور نخواهند انداخت زیرا بسیار ارزشمند خواهند بود. اگر شما اینکار را انجام ندهید از دنیای کسب‌وکار خارج می‌شوید. » به همین منوال، توماس داوینپورت^۲ (۲۰۰۶) در مقاله‌ای در هاروارد بیزینس ریویو بیان می‌کند که آخرین سلاح استراتژیک شرکت‌ها، تصمیم‌گیری تحلیلی است و مثال‌هایی از شرکت‌هایی از قبیل **آمازون. کام**، **کپیتال وان**^۳، **ماربوت اینترنشنال**^۴ و سایرین که از تحلیل‌ها برای درک بهتر مشتریان و بهینه‌سازی زنجیره‌های تأمین خود برای دستیابی به حد اکثر بازگشت سرمایه و ارائه بهترین خدمات به مشتریان استفاده کرده‌اند ارائه می‌کند. این سطح از موفقیت به میزان زیادی به درک کامل شرکت از مشتریان، تأمین‌کنندگان، فرایندهای کسب‌وکار و زنجیره تأمین گسترده آن بستگی دارد.

بخش بزرگی از « درک مشتری » می‌تواند از تحلیل حجم انبوه داده‌هایی که شرکت جمع‌آوری می‌کند بدست آید. هزینه‌های انبارش و پردازش داده‌ها طی چند سال گذشته شدیداً کاهش یافته است و در نتیجه، حجم داده‌های ذخیره شده به شکل الکترونیک نیز با نرخ انفجاری افزایش یافته است. با ایجاد پایگاه داده‌های بزرگ، امکان تحلیل داده‌های ذخیره شده در آنها نیز میسر شده است. عبارت *داده‌کاوی* در اصل برای توضیح فرایند کشف الگوهای ناشناخته در داده‌ها بکار رفت. از آن زمان تاکنون، این تعریف فراتر از این محدوده‌ها رفته و برخی فروشندگان نرم‌افزارها به منظور افزایش فروش محصولات خود اکثر شکل‌های تحلیل داده را لحاظ کرده‌اند.

اگرچه عبارت داده‌کاوی نسبتاً جدید است ولی مفاهیم و عقاید پشت آن چندان جدید نیستند. بسیاری از تکنیک‌های بکار رفته در داده‌کاوی ریشه در تحلیل‌های آماری قدیمی و کارهای هوش مصنوعی دارند که از اوایل دهه ۸۰ شکل گرفته‌اند. پس چرا تا این اندازه توجه دنیای کسب‌وکار متوجه آن شده است؟ برخی از مهمترین دلایل آن عبارتند از:

- تشدید رقابت در مقیاس جهانی به دلیل تغییرات مداوم نیازها و خواسته‌های مشتریان در بازارهای اشباع شده.
- پی بردن به ارزش پنهان بهره‌برداری نشده‌ی منابع بزرگ داده‌ها.
- ادغام و یکپارچه‌سازی رکوردهای پایگاه‌های داده که یک دید جامع و تکی از مشتریان، تأمین‌کنندگان، تراکنش‌ها و غیره را می‌دهد.
- ادغام پایگاه‌های داده و سایر مخازن داده در یک مکان به شکل یک انبار داده.
- افزایش نمایی فناوری‌های پردازش و انبارش داده‌ها.
- کاهش چشمگیر هزینه‌ی سخت‌افزاری و نرم‌افزاری پردازش و انبارش داده‌ها.
- حرکت به سمت انبوه‌دابی (تبدیل منابع اطلاعاتی به شکل غیرفیزیکی) روش‌های کسب‌وکار

¹ Dr. Arno Penzias

² Thomas Davenport

³ Capital One

⁴ Marriott International

حجم و پیچیدگی داده‌های تولید شده توسط اینترنت به سرعت در حال افزایش است. حجم انبوهی از داده‌های ژنومیک در سرتاسر جهان تولید و تجمیع می‌شوند. رشته‌هایی از قبیل نجوم و فیزیک هسته‌ای، مقادیر انبوهی از داده‌ها را به طور منظم تولید می‌کنند. پژوهشگران پزشکی و داروسازی نیز به طور مستمر در حال تولید و ذخیره‌سازی داده‌هایی هستند که بعداً بتوان از آنها در کاربردهای داده‌کاوی برای شناسایی روش‌های بهتری جهت تشخیص و درمان دقیق بیماری‌ها و کشف داروهای جدید و ارتقاء یافته استفاده کرد.

از منظر تجاری، شاید متداول‌ترین کاربرد داده‌کاوی را بتوان در بخش‌های مالی، خرده‌فروشی و مراقبت سلامت دانست. از داده‌کاوی به منظور تشخیص و کاهش فعالیت‌های مشکوک به کلاهبرداری به ویژه در ادعاهای بیمه و استفاده از کارت‌های اعتباری (چان و همکاران ۱۹۹۹)؛ شناسایی الگوهای خرید مشتریان (هافمن ۱۹۹۹)؛ بازپس‌گیری مشتریان سودآور (هافمن ۱۹۹۹)؛ شناسایی قواعد تجارت از روی داده‌های تاریخی و کمک به افزایش سودآوری از طریق تحلیل‌های سبد خرید استفاده می‌شود. از داده‌کاوی تا پیش از این نیز به طور گسترده‌ای برای هدف‌گیری بهتر مشتریان استفاده می‌شده است و با توسعه‌ی گسترده‌ی تجارت الکترونیک، با گذشت زمان این امر ضروری‌تر شده است. برای اطلاعات در خصوص نحوه بکارگیری تحلیل‌های پیشگویانه و داده‌کاوی برای بهبود خدمات مشتریان، مبارزه با کلاهبرداری و افزایش سود در اینفینیتی پی آند سی^۱ به مورد کاربردی ۴.۱ مراجعه کنید.

ویزا با کاهش کلاهبرداری‌ها از طریق تجزیه و تحلیل پیشگویانه و داده‌کاوی، تجربه مشتری را ارتقاء می‌بخشد

مورد کاربردی ۴.۱

زمانیکه صادرکنندگان کارت‌های اعتباری برای اولین بار شروع به استفاده از نرم‌افزار قواعد کسب‌وکار خودکار برای مقابله با کلاهبرداری در کارت‌های نقدی و اعتباری کردند، محدودیت‌های تکنولوژیکی در آن حوزه به سرعت خودشان را نشان دادند: مشتریان که در تعطیلات رویایی یا سفرهای کاری مهم به سر می‌بردند از عدم پذیرش پرداخت‌های آزادنده شکایت می‌کردند. ویزا^۲ با مشتریان برای بهبود تجربه مشتری از طریق ارائه ابزارهای پیشرفته‌ی ریسک کلاهبرداری و خدمات مشاوره همکاری می‌کند که بر اثربخشی استراتژی‌هایش می‌افزاید. ویزا با این رویکرد توانست تجربه مشتری را بهبود بخشیده و عدم پذیرش تراکنش‌های نامعتبر را به حداقل برساند.

شبکه جهانی این شرکت، روزانه هزاران مؤسسه مالی را به میلیون‌ها بازرگان، تاجر و دارنده‌ی کارت متصل می‌کند. این شرکت به مدت بیش از ۵۰ سال به عنوان یک پیشگام در عرصه‌ی پرداخت‌های غیرنقدی شناخته شده است. ویزا با استفاده از اس‌ای‌اس آنالیتیکز از مؤسسات مالی برای کاهش کلاهبرداری‌ها و بدون ایجاد نارضایتی مشتریان از عدم پذیرش پرداخت‌های غیرضرور پشتیبانی می‌کند. ویزا هر زمان که تراکنشی را پردازش می‌کند، حداکثر ۵۰۰ متغیر منحصر بفرد را به صورت لحظه‌ای برای ارزیابی ریسک آن تراکنش تحلیل می‌کند. شرکت با استفاده از مجموعه داده‌های گسترده از جمله نقاط حساس جهانی و الگوهای تراکنش‌ها می‌تواند به دقت بررسی و ارزیابی کند که آیا مشتری در حال خرید حلزون در پاریس است یا اینکه شخصی کارت اعتباری او را به سرقت برده است.

¹ Infinity P&C

² Visa

ناتان فالکن‌بورگ^۱، رئیس بخش ویزا پرفورمنس سولوشنز^۲ منطقه‌ی شمال آسیا می‌گوید: «این بدان معناست که اگر قرار باشد به مسافرت بروید ما مطلع خواهیم شد و به مؤسسه مالی‌تان خواهیم گفت تا در پایانه‌ی فروش، درخواست شما رد نشود. ما همچنین به بانک شما کمک می‌کنیم تا استراتژی‌های مناسب برای استفاده از ابزارهای ویزا و سیستم‌های امتیازدهی را تدوین کنند.» ویزا برآورد کرده است که کارهای تحلیل کلان‌داده‌ها، مدل‌ها و سیستم‌های امتیازدهی پیشرفته و به‌روز، توان بالقوه‌ای برای پیشگیری از کلاهبرداری سالانه مبلغی حدود ۲ میلیارد دلار را دارند.

ویزا که برندی شناخته شده در جهان است، انتقال الکترونیکی پول از طریق محصولات صاحب برندی که توسط هزاران شریک مالی صادر می‌شوند را تسهیل می‌کند. این شرکت ۶۴.۹ میلیارد تراکنش را در سال ۲۰۱۴ پردازش کرده و ۴.۷ تریلیون دلار خرید از طریق کارت‌های ویزا در همان سال انجام شده است.

ویزا توانایی پردازش ۵۶,۰۰۰ پیام تراکنش را در ثانیه دارد که بیش از چهار برابر نرخ اوج حقیقی تراکنش‌ها تا به امروز است. ویزا فقط به پردازش و رایانش نمی‌پردازد بلکه به طور مستمر از تحلیل‌ها برای اشتراک‌گذاری بینش‌های عملیاتی و استراتژیک با شرکای مالی خود و کمک به بهبود عملکرد آنها استفاده می‌کند. ویزا از یک سیستم مدیریت داده‌های قوی برای پشتیبانی این هدف کسب‌وکار خویش استفاده می‌کند. ویزا همچنین با توسعه و ارائه‌ی بینش‌های تحلیلی عمیق به بهبود عملکرد مشتریان خود کمک می‌کند.

به گفته‌ی فالکن‌بورگ، «ما با خوشه‌بندی و بخش‌بندی در یک سطح دانه‌بندی شده به درک الگوهای رفتاری می‌پردازیم و این بینش و ادراک را در اختیار شرکای مالی خود نیز قرار می‌دهیم. این روشی مؤثر برای کمک به برقراری ارتباط بهتر با مشتریان و دستیابی به درک عمیق‌تر از آنان است.» به عنوان یک مثال از پشتیبانی بازاریابی، ویزا به مشتریان خود در سرتاسر جهان در شناسایی بخش‌هایی از مشتریان که نیازمند دریافت محصول متفاوتی از ویزا هستند کمک کرده است. فالکن‌بورگ می‌گوید: «درک چرخه‌ی طول عمر مشتری بی‌اندازه مهم است و ویزا اطلاعاتی را در اختیار مشتریان قرار می‌دهد که به آنها در اقدام کردن کمک کرده و محصول درست و مناسب را به مشتری درست و پیش از آنکه گزاره‌ی ارزشی منسوخ شود عرضه می‌کند.»

استفاده از تحلیل‌های درون حافظه‌ای چگونه می‌تواند تفاوت ایجاد کند؟

در اثبات مفهوم اخیر، ویزا از یک راه‌حل با عملکرد سطح بالا از شرکت اس‌ای‌اس که متکی بر رایانش درون حافظه‌ای است برای توانمندسازی مدل‌های آماری و الگوریتم‌های یادگیری ماشین استفاده کرده و سپس اطلاعات را به صورت دیداری نمایش داده است. تحلیل‌های درون حافظه‌ای، نیازه‌ی جایجایی داده‌ها را کاهش داده و امکان اجرای چندین باره‌ی مدل‌ها را فراهم می‌کنند که بدین ترتیب سریع‌تر و دقیق‌تر خواهد شد.

فالکن‌بورگ این راه‌حل را اینگونه مقایسه می‌کند: اطلاعات را حفظ باشید یا اینکه مجبور شوید برای دریافت اطلاعات از جای خود بلند شده و به سراغ قفسه‌های فایل‌ها بروید. «تحلیل‌های درون حافظه‌ای، مغز شما را گرفته و آن را بزرگتر می‌کند. به

¹ Nathan Falkenberg

² Visa Performance Solutions

هر چیزی در لحظه دسترسی دارید. « سرانجام، تحلیل‌های یکپارچه و قوی به شرکت کمک می‌کنند تا چیزی فراتر از پردازش پرداخت‌ها را انجام دهد. فالکن‌بورگ می‌گوید: « ما می‌توانیم گفتگوهای عمیق‌تری با مشتری داشته باشیم و با استفاده از مجموعه کلان‌داده‌های فوق‌العاده و تخصص در کاوش داده‌های تراکنش‌ها می‌توانیم خدمت بهتری به مشتریان بدهیم. ما از توانمندی‌های تحلیلی و مشاوره‌ای برای کمک به مشتریان در حل و فصل چالش‌ها و محافظت از اکوسیستم پرداخت‌ها استفاده می‌کنیم. و این کاری است که با تحلیل‌های عملکرد سطح بالا انجام می‌دهیم. »

چالشی که شرکت ما همانند هر شرکت دیگری که مجموعه داده‌های انبوه خود را مدیریت کرده و بکار می‌گیرد این است که چگونه از تمامی اطلاعات ضروری برای حل یک چالش کسب‌وکار استفاده کنیم – صرفنظر از اینکه برای اینکار مجبور به ارتقاء مدل‌های کلاهدراری باشیم یا به یک مشتری کمک کنیم تا ارتباطات بهتری با مشتریان خود برقرار کند. تحلیل‌های درون حافظه‌ای ما را قادر می‌سازند تا چالاک‌تر باشیم؛ با یک بهبود ۱۰۰ درصدی در سرعت پردازش سیستم تحلیلی، دانشمندان علوم داده و تصمیم‌گیری ما می‌توانند سریع‌تر تکرار کنند.

تحلیل‌های پیشگویانه‌ی سریع و دقیق به ویزا اجازه می‌دهد تا از طریق ارائه خدمات مشاوره‌ی سفارشی، خدمات بهتری به مشتریان ارائه کرده و به آنها کمک کنند تا در صنعت پرداخت که امروزه به سرعت در حال تغییر است موفق شوند.

Source: "Enhancing the Customer Experience While Reducing Fraud (SAS® Analytics) High-Performance Analytics Empowers Visa to Enhance Customer Experience While Reducing Debit and Credit Card Fraud." Copyright © 2018 SAS Institute Inc., Cary, NC, USA. Reprinted with permission. All rights reserved.

سوالاتی برای مورد کاربردی ۴.۱

۱۴. ویزا و سایر شرکت‌های فعال در صنعت کارت اعتباری با چه چالش‌هایی مواجه بودند؟
۱۵. ویزا چگونه توانست خدمات مشتریان و همچنین مفاهیم مربوط به **retention of fraud** را بهبود ببخشد؟
۱۶. تحلیل درون حافظه‌ای چیست و چرا ضرورت دارد؟

داده‌کاوی به بیان ساده عبارتی است که برای توضیح کشف یا « کاوش » دانش از بین حجم انبوهی از داده‌ها بکار می‌رود. در مقام قیاس به آسانی می‌توان متوجه شد که عبارت *داده‌کاوی*، نام غلطی است؛ به این معنا که کاوش طلا از میان سنگ‌ها و گل و لای را به جای « سنگ » کاوی یا « گل و لای » کاوی، کاوش « طلا » می‌گویند. از این رو، شاید بهتر بود که داده‌کاوی را نیز « دانش کاوی » یا « کشف دانش » می‌نامیدند. علیرغم این عدم تطابق بین این عبارت و معنای آن، جامعه‌ی علمی عبارت داده‌کاوی را انتخاب کرده است. بسیاری از اسامی دیگر که با داده‌کاوی مرتبط هستند عبارتند از: *استخراج دانش*^۱، *تحلیل الگو*^۲، *دیرینه‌شناسی داده‌ها*^۳، *درو کردن اطلاعات*^۴، *جستجوی الگو*^۵ و *لایروبی داده‌ها*^۶.

¹ Knowledge extraction

² Pattern analysis

³ Data archeology

⁴ Information harvesting

⁵ Pattern searching

⁶ Data dredging

به لحاظ فنی، داده‌کاوی فرایندی است که از تکنیک‌های هوش مصنوعی، ریاضی و آمار برای استخراج و شناسایی اطلاعات مفید و دانش (یا الگوهای) متعاقب آن از مجموعه داده‌های بزرگ استفاده می‌کند. این الگوها می‌توانند به شکل قواعد کسب‌وکار، وابستگی‌ها، همبستگی‌ها، روندها یا مدل‌های پیش‌بینی باشند (نماتی آند بارکو^۱ ۲۰۰۱). داده‌کاوی در اکثر ادبیات موضوع به صورت «فرایندی غیرمستقیم برای شناسایی الگوهای معتبر، نوین، بالقوه مفید و نهایتاً قابل درک در داده‌های ذخیره شده در پایگاه داده‌های ساخت‌یافته» تعریف می‌شود که داده‌ها در رکوردهایی با ساختارهای رسته‌ای، ترتیبی یا متغیرهای پیوسته سازماندهی شده‌اند. (فایاد، پیاتتسکی-شاپیرو و اسمیث^۲ ۱۹۹۶). در این تعریف، مفاهیم عبارات کلیدی به شرح زیر می‌باشد:

- فرایند به این نکته اشاره دارد که داده‌کاوی از گام‌های تکرار شونده‌ی بسیاری تشکیل شده است.
- غیرمستقیم^۳ به این معناست که نیاز به نوعی جستجو از جنس آزمایش یا استنتاج می‌باشد و مثل محاسبه‌ی مقادیر کمی و از پیش تعریف شده، راحت و بی‌دردسر نیست.
- معتبر به این معنی که الگوهای کشف شده باید با درجه اطمینان مشخصی بر روی داده‌های جدید نیز صدق کنند.
- نوین به این معنا که کاربر قبلاً در بافتار سیستمی که تحلیل می‌کند با این الگوها آشنا نشده است.
- بالقوه مفید به این معنی که الگوهای کشف شده باید منجر به سود و منفعت کاربر شوند.
- نهایتاً قابل درک به این معنا که الگو باید در حوزه کسب‌وکار معنا داشته باشد که حداقل کمی پس از پس‌پردازش داده‌ها موجب شود تا کاربر بگوید «منطقی است! چرا به فکر من نرسید؟»

داده‌کاوی رشته‌ی جدیدی نیست بلکه تعریف جدیدی برای استفاده‌ی تعداد زیادی از رشته‌ها است. داده‌کاوی محل تقاطع بسیاری از رشته‌ها از جمله آمار، هوش مصنوعی، یادگیری ماشین، علم مدیریت، سیستم‌های اطلاعاتی و پایگاه‌های داده است (شکل ۴.۱). داده‌کاوی با بهره‌گیری از تمامی پیشرفت‌های بدست آمده در این رشته‌ها سعی دارد تا در استخراج اطلاعات مفید و دانش از پایگاه داده‌های بزرگ پیشرفت کند. داده‌کاوی رشته‌ای نوظهور است که در مدت زمانی کوتاه، توجه بسیاری را به خود جلب کرده است.

برخی از اهداف و ویژگی‌های اصلی داده‌کاوی در ادامه آورده شده است:

- داده‌ها اغلب در میان پایگاه داده‌های بسیار بزرگ مدفون شده‌اند که برخی اوقات حاوی اطلاعات چندین سال هستند. در بسیاری از موارد، داده‌ها پاکسازی شده و در داخل یک انبار داده ادغام می‌شوند. داده‌ها را می‌توان به شکل‌ها و قالب‌های مختلفی نمایش داد.
- محیط داده‌کاوی معمولاً دارای یک معماری سرویس‌گیرنده/سرویس‌دهنده یا سیستم‌های اطلاعاتی مبتنی بر وب است.
- ابزارهای جدید پیچیده از جمله ابزارهای مصورسازی پیشرفته به حذف اطلاعات مدفون شده در فایل‌های سازمانی یا رکوردهای عمومی بایگانی شده کمک می‌کنند. یافتن این اطلاعات شامل ماساژ دادن و همزمان‌سازی داده‌ها برای رسیدن به نتایج درست است. داده‌کاوها بسیار پیشرفته نیز به اکتشاف سودمندی داده‌های نرم می‌پردازند (یعنی متن

¹ Nemati & Barko

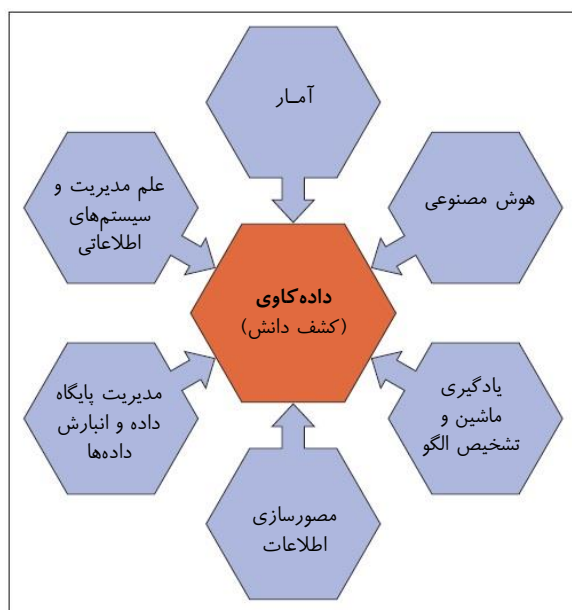
² Fayyad, Piatetsky-Shapiro & Smyth

³ Nontrivial

ساختنیافته ذخیره شده در مکان‌هایی مثل پایگاه‌های داده لوتوس^۱، فایل‌های متنی در اینترنت یا اینترنت‌های سازمانی).

- داده‌کاوی اغلب یک کاربر نهایی است که امکان استفاده از ابزارهای پرس و جوی قدرتمند برای طرح سوالات فی‌البداهه و موردی و دستیابی سریع به پاسخ‌ها را بدون برخورداری از مهارت‌های برنامه‌نویسی دارد.
- «یک شبه پولدار شدن» اغلب شامل یافتن نتایج غیرمنتظره بوده و نیازمند این است که کاربران در سرتاسر فرایند از جمله تفسیر یافته‌ها، تفکر خلاقانه را حفظ کنند.
- ابزارهای داده‌کاوی به آسانی با صفحات گسترده و سایر ابزارهای توسعه نرم‌افزار ادغام می‌شوند. بنابراین، داده‌های کاوش شده را می‌توان به سرعت و آسانی تحلیل کرد.
- به دلیل حجم انبوه داده‌ها و تلاش‌های جستجوی گسترده، برخی اوقات لازم است که از پردازش‌های موازی برای داده‌کاوی استفاده شود.

یک شرکت که از ابزارها و فناوری‌های داده‌کاوی بهره‌برداری موثری می‌کند می‌تواند به یک مزیت رقابتی استراتژیک دست یافته و آن را حفظ نماید. داده‌کاوی از طریق تبدیل داده‌ها به یک سلاح استراتژیک یک محیط تصمیم‌گیری اجتناب‌ناپذیر را برای بهره‌برداری از فرصت‌ها برای سازمان‌ها فراهم می‌کنند. برای بحث بیشتر در خصوص مزایای استراتژیک داده‌کاوی نگاه کنید به نِماتی آند بارکو (۲۰۰۱).



شکل ۴.۱ داده‌کاوی، آمیزه‌ای از رشته‌های مختلف است

داده‌کاوی با استفاده از داده‌های موجود و مرتبط بدست آمده از داخل و خارج سازمان مدلی را برای کشف الگوها بین صفات ارائه شده در مجموعه داده‌ها ایجاد می‌کند. مدل‌ها، بازنمودهای ریاضی (روابط/وابستگی‌های خطی ساده و/یا پیچیده و ارتباطات کاملاً غیرخطی) که قادر به شناسایی الگوها بین صفات اشیاء (مشتریان، رویدادها) در داخل یک مجموعه داده هستند. برخی از

این الگوها حالت توضیحی دارند (ارتباطات درونی و وابستگی‌های بین صفات را توضیح می‌دهند) در حالیکه برخی دیگر ماهیت پیشگویانه دارند (پیش‌بینی مقادیر آینده برای صفات خاص). به طور کلی، داده‌کاوی به دنبال شناسایی چهار نوع الگوی اصلی است:

۱. **وابستگی‌ها** که به دنبال یافتن گروه‌بندی مشترک اشیاء مثل ماء‌الشعیر و پوشک بچه که در تحلیل سبد خرید کنار یکدیگر قرار می‌گیرند.
۲. **پیش‌بینی‌ها** که در مورد ماهیت وقوع برخی رخدادهای خاص در آینده بر اساس آنچه در گذشته اتفاق افتاده به ما می‌گویند مثل پیش‌بینی برنده‌ی مسابقه‌ی سوپر باول یا پیش‌بینی دمای مطلق در یک روز خاص.
۳. **خوشه‌ها** که به شناسایی گروه‌بندی‌های طبیعی اشیاء بر اساس مشخصه‌های شناخته شده آنها می‌پردازند مثل تخصیص مشتریان به بخش‌های مختلف بر اساس رفتارهای خرید گذشته یا جمعیت‌شناختی آنها.
۴. **روابط متوالی** که به کشف رویدادهای با ترتیب زمانی می‌پردازند مثل پیش‌بینی اینکه یک مشتری فعلی بانک که یک حساب جاری دارد در عرض یکسال پس از افتتاح یک حساب سرمایه‌گذاری، یک حساب قرض‌الحسنه افتتاح می‌کند.

در مورد کاربردی ۴.۲ نشان می‌دهیم که چگونه شرکت هوندا در آمریکا از داده‌کاوی (یکی از مؤلفه‌های حیاتی ابزارهای پیشرفته‌ی تحلیل) برای افزایش درک ادعاهای پرداخت گارانتی، پیش‌بینی نیازهای منبع و قطعه و درک بهتر نظرات، خواسته‌ها و نیازهای مشتریان استفاده می‌کند.

آمریکن هوندا از تحلیل‌های پیشرفته برای بهبود پرداخت ادعاهای گارانتی استفاده می‌کند

مورد کاربردی ۴.۲

تاریخچه

زمانیکه یک مالک خودرو یا کامیون، وسیله‌ی نقلیه خود را به یک نمایندگی آکورا یا هوندا در ایالات متحده می‌آورد، کارهایی بیش از یک یک بازدید چک سرویس یا تعمیر انجام می‌شود. در طی هر بازدید، تکنیسین‌های خدماتی، داده‌هایی را در خصوص تعمیرات تولید می‌کنند از جمله ادعاهای گارانتی به شرکت آمریکایی هوندا موتورز^۱ که این داده‌ها مستقیماً وارد پایگاه داده شرکت می‌شوند. این اطلاعات شامل نوع کار انجام شده، هزینه‌ی پرداخت شده توسط مشتری، نظرات مشاور خدمات و بسیاری دیگر از داده‌ها می‌باشد.

امروزه، این فرایند را ضربدر ده‌ها بازدید روزانه در ۱،۲۰۰ نمایندگی سراسر کشور کنید و روشن است که شرکت امریکن هوندا دارای کلان‌داده‌ها است. این به افرادی مثل کندی‌ریک کائو، مدیر دستیار گروه آنالیتیکز گروپ امریکن هوندا بستگی دارد که از این داده‌ها، بینش استخراج کرده و آنها را به یک دارایی مفید تبدیل کند.

بررسی داده‌های گارانتی برای نگهداری و تعمیرات مؤثرتر

امریکن هوندا همانند هر خودروساز بزرگ دیگری با شبکه‌ای از نمایندگی‌ها کار می‌کند که به ارائه خدمات تعمیرات گارانتی برای وسایل نقلیه آنها می‌پردازند. این موضوع می‌تواند هزینه‌های هنگفتی را برای شرکت به همراه داشته باشد و به همین دلیل امریکن هوندا از تحلیل‌ها برای حصول اطمینان از ارسال ادعاهای گارانتی کامل و دقیق استفاده می‌کند.

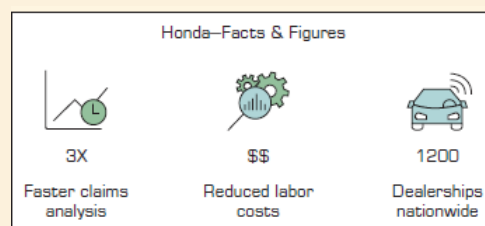
¹ American Honda Motors

در خصوص ادعاهای گارانتی، تیم کائو با ارائه‌ی اطلاعات مفید از طریق یک گزارش آنلاین به نمایندگی‌ها کمک می‌کند فرایندهای گارانتی مناسب را درک کنند. به منظور پشتیبانی از هدف کاهش هزینه‌های نامناسب گارانتی، کائو و تیم وی باید اطلاعات مربوط به تعمیرات، قطعات، مشتریان و سایر جزئیات را غربال کنند. آنها از یک روش دیداری هوش و تحلیل کسب‌وکار شرکت ساس به منظور شناسایی فرصت‌های کاهش هزینه‌ها استفاده کردند. برای کاهش هزینه‌های گارانتی، تیم تحلیل‌های پیشرفته از اس‌ای‌اس آنالیتیکز برای ایجاد یک فرایند انحصاری به منظور بررسی ادعاهای گارانتی مشکوک به صورت روزانه و حصول اطمینان از انطباق آنها با دستورالعمل‌های موجود استفاده کرد. تلاش برای شناسایی و موشکافی ادعاها زمانی به صورت کاملاً دستی انجام می‌شد که بسیار زمان‌بر و ملال‌آور بود.

کائو می‌گوید: « تا پیش از ساس، هر ماه یک هفته طول می‌کشید تا یکی از پرسنل ما داده‌های گارانتی را در صفحات گسترده‌ی اکسل تجمیع و گزارش‌دهی کند. اکنون با استفاده از ساس، همان گزارشات را به صورت خودکار بر روی داشبوردهای آنلاین با دسترسی آسان منتشر می‌کنیم که یک هفته زمان کار نیروی انسانی ما را آزاد کرده است تا به پروژه‌های دیگری تخصیص بدهیم. »

با بکارگیری تحلیل‌های ساس آنالیتیکز بر روی داده‌های گارانتی، گروه تحلیل‌های پیشرفته توانایی شناسایی سریع و دقیق ادعاهای گارانتی ناقص، نادقیق یا نامنطبق را برای پرسنل حوزه و گروه « ادعاها » فراهم کرد. نتایج بسیار چشمگیر بودند. کائو می‌گوید: « در ابتدا، به طور متوسط سه دقیقه طول می‌کشید تا بازرسان ما یک ادعای بالقوه نامنطبق را شناسایی کنند و حتی پس از آن نیز در ۳۵ درصد مواقع موفق به شناسایی یک ادعای کاملاً نامنطبق می‌شدند. اکنون به کمک ساس، در کمتر از یک دقیقه یک ادعای مشکوک شناسایی می‌شود. و در آن زمان، ادعاهای نامنطبق در ۷۶ درصد مواقع شناسایی می‌شوند. »

تلاش برای افزایش انطباق گارانتی‌ها به نفع امریکن هوندا تمام شد. امریکن هوندا از طریق تحلیل کامل‌تر ادعاهای گارانتی و آموزش‌های بیشتر در نمایندگی‌ها توانست هزینه‌های نیروی کار خود را به میزان ۵۲٪ کاهش دهد.



استفاده از داده‌های خدمات برای پیش‌بینی نیازهای آینده

تیم امریکن هوندا آدونس آنالیتیکز نیز از داده‌های قطعات و خدمات برای برقراری ارتباطات قوی‌تر با مشتریان استفاده کرد که برای این منظور اطمینان حاصل کرد که قطعات مورد نیاز برای تعمیرات خودروهای مشتریان در نمایندگی‌ها موجود می‌باشد. موجود بودن قطعات درست و در زمان درست اهمیت بسیار بالایی دارد و به همین دلیل داده‌های مربوط به تعمیرات به عنوان ورودی مستقیم فعالیت‌های حفظ و نگهداشت مشتری و بازاریابی در شرکت امریکن هوندا محسوب می‌شوند.

کائو می‌گوید: « برای تیم بازاریابی، ما بینش‌های استراتژیکی را برای تدوین برنامه‌هایی ارائه می‌کنیم که با هدف جلب مشتریان به سمت نمایندگی‌ها و حفظ وفاداری آنها به برند ما طراحی شده‌اند. هدف هوندا، وفاداری مادام‌العمر مالک خودرو است. ما می‌خواهیم که مشتریان ما تجربه خوبی داشته باشند و یک روش برای اینکار، ارائه خدمات استثنایی است. »

امریکن هوندا از ساس فورگست سرور^۱ برای کمک به برنامه‌ریزی کسب‌وکار و حصول اطمینان از دسترسی به منابع مورد نیاز جهت برآورده کردن نیازهای خدماتی آینده استفاده می‌کند. آنها با استفاده از داده‌های تاریخی درخواست‌های تعمیراتی و گواهی‌ها، سری‌های زمانی از تعمیرات سال‌های گذشته را ایجاد کردند و تیم کائو از طریق ترکیب اطلاعات سری‌های زمانی با داده‌های فروش توانستند بهترین فرصت‌های پیش رو در سال‌های آینده را پیش‌بینی کنند.

کائو می‌گوید: «هدف ما این است که تعداد وسائل نقلیه‌ی در حال خدمت را به منظور پیش‌بینی حجم مشتریانی است که به نمایندگی‌های ما مراجعه خواهند کرد پیش‌بینی کنیم.»

«و این نشان دهنده این است که ما باید چه تعداد قطعه یدکی را در اختیار داشته باشیم و به ما در برنامه‌ریزی نیروی انسانی مورد نیاز جهت برآورده کردن تقاضای مشتریان کمک می‌کند. با نگاه سال به سال گذشته ما در محدوده‌ی ۱ درصدی از آن چیزی که پیش‌بینی می‌کنیم باشیم قرار داریم. این پیش‌بینی بسیار خوبی است و قسمت بیشتر آن را مدیون توانایی‌های نرم‌افزار ساس هستیم.»

بازخورد مشتری، عامل محرک کسب‌وکار

روشی دیگری که امریکن هوندا از تحلیل‌ها استفاده می‌کند برای ارزیابی سریع داده‌های نظرسنجی مشتریان است. تیم آدونس آنالیتیکز با استفاده از نرم‌افزار ساس برای کاوش در داده‌های نظرسنجی مشتریان برای دستیابی به درک و بینش در خصوص نحوه استفاده از وسائل نقلیه و شناسایی تغییرات طراحی که موجب افزایش رضایت مشتری خواهند شد استفاده می‌کند.

تیم تحلیل به صورت هفتگی داده‌های نظرسنجی مشتریان را ارزیابی می‌کند. تیم کائو از اس‌ای‌اس برای مشخص کردن روندهای نوظهور که ممکن است نیازمند توجه گروه‌های طراحی، ساخت و تولید، مهندسی و ... باشند استفاده می‌کند. کاربران با استفاده از فناوری ساس می‌توانند از مسائل سطح کلان و بالا به پاسخ‌های مشخص‌تر و جزء‌تر برای درک علل ریشه‌ای بروند.

ما می‌توانیم به داده‌ها نگاه کرده و متوجه شویم که مشتریان چه می‌گویند. و این منجر به مجموعه‌ای از سوالات می‌شود که باید به آنها پاسخ دهیم. آیا قطعه به بهترین شکل ممکن طراحی شده است؟ آیا آموزش مشتری، مسئله محسوب می‌شود؟ آیا به این موضوع باید در فرایند ساخت و تولید رسیدگی کنیم؟ به لطف ساس ما می‌توانیم این سوالات بحرانی را با استفاده از داده‌هایمان شناسایی کنیم.»

Source: SAS Case Study "American Honda Motor Co., Inc. uses SAS advanced analytics to improve warranty claims"

https://www.sas.com/en_us/customers/american-honda.html (accessed June 2018)

سوالاتی برای مورد کاربردی ۴.۲

۱. امریکن هوندا چگونه از تجزیه و تحلیل‌ها برای بهبود ادعاهای گارانتی استفاده می‌کند؟
۲. علاوه بر ادعاهای گارانتی، امریکن هوندا از روش‌های تجزیه و تحلیل پیشرفته برای چه مقاصدی استفاده می‌کند؟
۳. آیا می‌توانید کاربردهای دیگر تجزیه و تحلیل‌های پیشرفته را در صنعت خودرو نام ببرید؟ برای یافتن جواب‌های این سوال می‌توانید در وب به جستجو بپردازید.

این نوع الگوها قرن‌ها توسط انسان و به صورت دستی استخراج می‌شدند ولی حجم فزاینده‌ی داده‌ها در عصر مدرن موجب شد تا به روش‌های خودکارتری نیاز داشته باشیم. با افزوده شدن بر حجم و پیچیدگی مجموعه داده‌ها، تحلیل دستی و مستقیم داده‌ها

¹ SAS Forecast Server

به میزان زیادی به ابزارهای غیرمستقیم و خودکار پردازش داده‌ها مجهز شد که از روش‌ها، الگوریتم‌ها و متدولوژی‌های پیچیده بهره می‌برند. تجلی تکامل اینگونه روش‌های خودکار و نیمه خودکار پردازش مجموعه داده‌های بزرگ را امروزه داده‌کاوی می‌نامند. به بیان کلی، وظایف داده‌کاوی را می‌توان به سه گروه اصلی تقسیم کرد: پیش‌بینی، برقراری ارتباط و خوشه‌بندی. الگوریتم‌های یادگیری روش‌های داده‌کاوی را بر اساس روش استخراج الگوها از داده‌های تاریخی می‌توان تحت عناوین با نظارت یا بدون نظارت دسته‌بندی کرد. در خصوص الگوریتم‌های یادگیری با نظارت، داده‌های آموزشی شامل صفات توصیفی (یعنی متغیرهای وابسته یا متغیرهای تصمیم) و صفات کلاس (یعنی متغیر خروجی یا متغیر نتیجه) می‌باشند. برعکس، در خصوص یادگیری بدون نظارت، آموزش‌ها فقط شامل صفات توصیفی هستند. شکل ۴.۲ یک دسته‌بندی ساده از وظایف داده‌کاوی را به همراه روش‌های یادگیری و الگوریتم‌های متداول برای هر یک از وظایف داده‌کاوی نشان می‌دهد.

پیش‌گویی پیش‌گویی^۱ معمولاً به فعل صحبت کردن در مورد آینده اشاره دارد. پیش‌گویی از آن جهت با حدس و گمان زدن تفاوت دارد که تجربیات، نظرات و سایر اطلاعات مرتبط برای آینده‌بینی را شامل می‌شود. یکی از عبارات متداول و مرتبط با پیش‌گویی، پیش‌بینی نام دارد. اگرچه بسیاری معتقدند که این دو عبارت مترادف یکدیگر هستند ولی یک تفاوت جزئی ولی بسیار مهم بین این دو عبارت وجود دارد. پیش‌گویی عمدتاً بر اساس تجربه و نظر و عقیده است در حالیکه پیش‌بینی مبتنی بر داده و مدل است. این بدان معناست که برای افزایش قابلیت اطمینان، عبارات مرتبط حدس زدن، پیش‌گویی و پیش‌بینی به ترتیب در یک سطح قرار می‌گیرند. در اصطلاحات واژگان داده‌کاوی، پیش‌گویی و پیش‌بینی به صورت مترادف یکدیگر بکار می‌روند و عبارت پیش‌گویی، بازنمود معمول‌تر این فعل است. بسته به ماهیت آنچه که پیش‌گویی می‌شود، پیش‌گویی را می‌توان به طور مشخص‌تر، طبقه‌بندی (جائیکه موضوعی که پیش‌گویی می‌شود مثلاً پیش‌بینی وضع آب و هوای فردا یک برجسب رسته‌ای باشد مثل «آفتابی» یا «بارانی») یا رگرسیون (که موضوعی که پیش‌گویی می‌شود مثل دمای هوا یک مقدار عددی حقیقی مثل ۶۵ درجه فارنهایت) نامید.

طبقه‌بندی طبقه‌بندی^۲ یا القای با نظارت را شاید بتوان متداول‌ترین وظیفه‌ی داده‌کاوی دانست. هدف از طبقه‌بندی، تحلیل داده‌های تاریخی ذخیره شده در یک پایگاه داده و تولید خودکار یک مدل است که بتواند رفتار آینده را پیش‌گویی کند. این مدل القایی شامل تعمیم رکوردهای یک مجموعه داده آموزشی است که به متمایزسازی دسته‌های از پیش تعیین شده کمک می‌کند. امید است که بتوان از این مدل برای پیش‌گویی دسته‌های سایر رکوردهای طبقه‌بندی نشده و مهمتر از آن، پیش‌گویی دقیق رویدادهای واقعی در آینده استفاده کرد.

ابزارهای متداول طبقه‌بندی شامل شبکه‌های عصبی و درخت تصمیم (از یادگیری ماشین)، رگرسیون لجستیک و تحلیل تشخیصی (از آمار) و ابزارهای نوظهور از قبیل مجموعه‌های راف^۳، ماشین‌های برداری پشتیبان و الگوریتم‌های ژنتیک اشاره کرد. تکنیک‌های طبقه‌بندی آماری (مثلاً رگرسیون لجستیک و تحلیل تشخیصی) به دلیل مفروضات غیرواقع‌بینانه در مورد داده‌ها از قبیل عدم وابستگی و نرمال بودن مورد انتقاداتی قرار گرفته‌اند که کاربرد آنها در پروژه‌های داده‌کاوی طبقه‌بندی را محدود کرده است. شبکه‌های عصبی شامل توسعه‌ی ساختارهای ریاضی (تا حدودی مشابه شبکه‌های عصبی بیولوژیکی در مغز انسان) که قابلیت یادگیری از تجارب گذشته به شکل مجموعه داده‌های ساختارمند را دارند. اثربخشی این شبکه‌های عصبی در شرایطی که تعداد

¹ Prediction

² Classification

³ Rough sets

متغیرهای مورد نظر بسیار زیاد باشد و ارتباطات بین آنها پیچیده و نادقیق باشد بیشتر است. شبکه‌های عصبی معایبی نیز دارند. به عنوان مثال، ارائه‌ی یک منطق خوب برای پیشگویی‌های شبکه‌های عصبی معمولاً بسیار دشوار است. علاوه بر این، آموزش شبکه‌های عصبی بسیار زمان‌بر است. متأسفانه، زمان مورد نیاز برای آموزش آنها با افزایش حجم داده‌ها به صورت نمایی افزایش می‌یابد و به طور کلی، شبکه‌های عصبی را نمی‌توان برای پایگاه داده‌های بسیار بزرگ آموزش داد. این عوامل موجب شده تا کاربرد شبکه‌های عصبی در حوزه‌هایی که داده‌ها بسیار حجیم هستند محدود شود.

درخت‌های تصمیم، داده‌ها را بر اساس مقادیر متغیرهای ورودی به تعداد محدودی از دسته‌ها تقسیم می‌کنند. درخت تصمیم در واقع یک سلسله مراتب از گزاره‌های شرطی if-then است که به همین دلیل بسیار سریع‌تر از شبکه‌های عصبی می‌باشد. این روش مناسب‌ترین روش برای دسته‌بندی **داده‌های رسته‌ای** و **داده‌های بازه‌ای** است. از این رو، استفاده از متغیرهای پیوسته در درخت تصمیم نیازمند **گسسته‌سازی** است یعنی تبدیل متغیرهای عددی پیوسته به دامنه‌ها و رسته‌ها.

یک گروه از ابزارهای طبقه‌بندی، **القاء قانون** است. برخلاف درخت تصمیم، بیان‌های if-then در این روش مستقیماً از داده‌های آموزشی القا می‌شوند و نیازی به ساختار سلسله‌مراتبی نمی‌باشد. یکی دیگر از تکنیک‌های جدیدتر مثل ماشین‌های برداری پشتیبان، مجموعه‌های راف و الگوریتم‌های ژنتیک به تدریج راه خود را به زرادخانه‌ی الگوریتم‌های طبقه‌بندی پیدا کرده‌اند.

نوع یادگیری	الگوریتم‌های داده کاوی	روش‌ها و وظایف داده کاوی
		پیشگویی
نظارت شده	درخت تصمیم، شبکه‌های عصبی، ماشین‌های برداری پشتیبان، الگوریتم کی نزدیک‌ترین همسایه، دسته‌بندی کننده بایز ساده، الگوریتم ژنتیک	دسته‌بندی
نظارت شده	رگرسیون خطی/غیرخطی، شبکه عصبی مصنوعی، درخت‌های رگرسیون، ماشین‌های بردار پشتیبان، الگوریتم کی نزدیک‌ترین همسایه، ژنتیک الگوریتم	رگرسیون
نظارت شده	روش‌های خودهمبستگی، روش‌های میانگین‌گیری هموارسازی نمایی، آرپا	سری‌های زمانی
		وابستگی
نظارت نشده	اِپروری، Eclat، ZeroR، OneR، الگوریتم ژنتیک	سبد-بازار
نظارت نشده	بیشینه‌سازی انتظار، الگوریتم اِپروری، انطباق مبتنی بر گراف	تحلیل پیوند
نظارت نشده	الگوریتم اِپروری، FP-Growth، انطباق مبتنی بر گراف	تحلیل توالی
		بخش‌بندی
نظارت نشده	خوشه‌بندی کی میانگین، بیشینه‌سازی انتظارات	خوشه‌بندی
نظارت نشده	خوشه‌بندی کی میانگین، بیشینه‌سازی انتظارات	تحلیل مقادیر پرت

شکل ۲.۱ طبقه‌بندی ساده وظایف، روش‌ها و الگوریتم‌های داده کاوی

خوشه‌بندی در روش خوشه‌بندی^۱، مجموعه‌ای از اشیاء (مثلاً رویدادها یا اشیاء ارائه شده در یک مجموعه داده‌ی ساختیافته) به بخش‌هایی (یا گروه‌بندی‌های طبیعی) تقسیم می‌شوند که اعضای این بخش‌ها دارای ویژگی‌های مشابهی هستند. بر خلاف طبقه‌بندی، در روش خوشه‌بندی، برچسب دسته‌ها نامشخص است. همچنان که مجموعه داده‌ها در الگوریتم منتخب وارد می‌شود و وجوه مشترک اشیاء بر اساس ویژگی‌های آنها شناسایی می‌شود، خوشه‌ها مشخص می‌شوند. از آنجائیکه خوشه‌ها با استفاده از یک الگوریتم هیوریستیک تعیین می‌شوند و از آنجائیکه الگوریتم‌های مختلف ممکن است مجموعه خوشه‌های متفاوتی را برای مجموعه داده‌های یکسان ارائه کنند، پیش از آنکه نتایج تکنیک‌های خوشه‌بندی را بتوانیم بکار بگیریم، لازم است تا یک فرد خیره، خوشه‌ها را تفسیر و در صورت امکان اصلاح نماید. پس از شناسایی منطقی خوشه‌ها، می‌توان آنها را برای طبقه‌بندی و تفسیر داده‌های خام بکار برد.

¹ Clustering

تعجب برانگیز نیست که تکنیک‌های خوشه‌بندی شامل بهینه‌سازی هستند. هدف از خوشه‌بندی، ایجاد گروه‌هایی است که اعضای درون هر گروه دارای بیشترین تشابه و اعضای بین گروه‌ها دارای کمترین تشابه باشند. متداول‌ترین تکنیک‌های خوشه‌بندی مورد استفاده شامل میانگین‌گیری^۱ (از آمار) و نقشه‌های خودسازمانده^۲ (از یادگیری ماشین) هستند که یک معماری شبکه عصبی منحصرفرد هستند که توسط کوهونن^۳ (۱۹۸۲) توسعه یافته‌اند.

شرکت‌ها اغلب از سیستم‌های داده‌کاوی برای انجام بخش‌بندی بازار با تحلیل خوشه استفاده می‌کنند. تحلیل خوشه، روشی برای شناسایی دسته‌های اقلامی است که اقلام در داخل یک خوشه بیش از اقلام در داخل سایر خوشه‌ها با یکدیگر وجه مشترک دارند. از تحلیل خوشه می‌توان در بخش‌بندی مشتریان و جهت‌دهی محصولات بازاریابی به بخش‌ها در زمان مناسب، به شکل مناسب و با قیمت مناسب استفاده کرد. از تحلیل خوشه همچنین برای شناسایی گروه‌بندی طبیعی رویدادها یا اشیاء استفاده می‌شود که مجموعه مشترکی از صفات و ویژگی‌های این گروه‌ها را می‌توان برای توصیف آنها شناسایی نمود.

وابستگی و وابستگی‌ها^۴ یا یادگیری قاعده وابستگی در داده‌کاوی^۵ یک تکنیک محبوب برای کشف ارتباطات جالب بین متغیرها در پایگاه داده‌های بزرگ است. به لطف فناوری‌های خودکار گردآوری داده‌ها از قبیل اسکنرهای بارکدها، استفاده از قواعد وابستگی برای کشف وجوه مشترک بین محصولات در تراکنش‌های مقیاس بزرگ که توسط سیستم‌های نقطه فروش در سوپرمارکت‌ها استفاده می‌شود، به یک وظیفه‌ی کشف دانش متداول در صنعت خرده‌فروشی تبدیل شده است. در مقوله‌ی صنعت خرده‌فروشی، کاوش قواعد وابستگی را اغلب تحلیل سبد خرید^۶ می‌نامند.

دو روش متداول کاوش قواعد وابستگی که از آن مشتق شده‌اند عبارتند از تحلیل پیوند^۷ و کاوش توالی^۸. در خصوص تحلیل پیوند، پیوندها و ارتباطات بین تعداد بسیار زیادی از اشیاء مورد نظر به صورت خودکار کشف می‌شود مثل پیوند بین صفحات وب و ارتباطات استنتاجی بین گروه مؤلفان نشریات دانشگاهی. در خصوص کاوش توالی، ارتباطات بر اساس ترتیب وقوع آنها در طول زمان برای شناسایی وابستگی‌ها بررسی می‌شوند. الگوریتم‌های بکار رفته در کاوش قواعد وابستگی شامل اپریوری^۹ و FP-Growth، OneR، ZeroR و Eclat می‌باشد.

مصورسازی و پیش‌بینی سری‌های زمانی دو تکنیک که اغلب با داده‌کاوی مرتبط هستند مصورسازی و پیش‌بینی سری‌های زمانی هستند. از مصورسازی می‌توان به همراه سایر تکنیک‌های داده‌کاوی برای رسیدن به درک بهتری از ارتباطات زیربنایی استفاده کرد. با توجه به افزایش اهمیت مصورسازی در سال‌های اخیر، عبارت جدیدی تحت عنوان تحلیل دیداری^{۱۰} پدیدار شده است. هدف از اینکار، ادغام مصورسازی و تحلیل‌ها در یک محیط برای تولید سریع‌تر و آسان‌تر دانش است. در فصل ۳ به تفصیل در مورد تحلیل‌های دیداری صحبت خواهیم کرد. در پیش‌بینی سری‌های زمانی، داده‌ها شامل مقادیری هستند که به متغیر

¹ k-means

² Self-organizer Maps

³ Kohonen

⁴ Association

⁵ Association rule learning in data mining

⁶ Market-Basket Analysis

⁷ Link Analysis

⁸ Sequence Mining

⁹ Apriori

¹⁰ Visual Analytics

یکسانی تعلق دارند و در طی بازه‌های زمانی منظم اخذ و ثبت می‌شوند. سپس از این داده‌ها برای تدوین مدل‌های پیش‌بینی برای برون‌یابی مقادیر آینده برای یک متغیر یکسان استفاده می‌شود.

داده‌کاوی و آمار وجوه اشتراک بسیاری دارند. هر دوی آنها به دنبال یافتن ارتباطات در درون داده‌ها هستند. اکثر مردم آمار را « زیربنای داده‌کاوی » می‌نامند. تفاوت اصلی بین این دو در این است که آمار با یک فرضیه و گزاره که به خوبی تعریف شده آغاز می‌شود در حالیکه داده‌کاوی با یک بیانیه‌ی کشف که به طور مبهم و کلی تعریف شده آغاز می‌شود. در آمار ما داده‌های نمونه را (یعنی داده‌های اصلی) به منظور آزمودن فرضیه‌ها گردآوری می‌کنیم در حالیکه در داده‌کاوی و تحلیل‌ها از تمامی داده‌های موجود (یعنی داده‌های دست دوم یا مشاهداتی) برای کشف ارتباطات و الگوهای نوین استفاده می‌کنیم. یک تفاوت اصلی دیگر مربوط به حجم داده‌های مورد استفاده است. در داده‌کاوی به دنبال مجموعه داده‌های تا حد ممکن « بزرگ » هستیم در حالیکه آمار به دنبال حجم مناسبی از داده‌ها است (اگر داده‌های بزرگ از میزان مورد نیاز برای تحلیل آماری باشند، از آنها نمونه‌برداری می‌شود). مفهوم « داده‌های بزرگ » در آمار و داده‌کاوی تا حدودی متفاوت است. برای یک کارشناس آمار، چند صد تا هزار داده می‌تواند بسیار زیاد باشد در حالیکه چند میلیون تا چند میلیارد داده‌ها برای مطالعات داده‌کاوی بزرگ محسوب می‌شود.

سوالاتی برای مرور بخش ۴.۲

۷. داده‌کاوی را تعریف کنید؟ چرا نام‌ها و تعاریف متعددی برای داده‌کاوی وجود دارد؟
۸. چه عواملی اخیراً موجب شده تا محبوبیت داده‌کاوی افزایش یابد؟
۹. آیا داده‌کاوی رشته‌ی جدیدی است؟ توضیح دهید.
۱۰. برخی از روش‌ها و الگوریتم‌های اصلی داده‌کاوی چه هستند؟
۱۱. تفاوت‌های اساسی بین وظایف اصلی داده‌کاوی چه هستند؟

داده‌کاوی به ابزاری محبوب برای حل و فصل بسیاری از مسائل پیچیده‌ی و بهره‌برداری از فرصت‌های کسب‌وکار تبدیل شده است. مفید و موفق بودن آن در بسیاری از حوزه‌های اثبات شده که در ادامه با ذکر مثال‌هایی برجسته به برخی از آنها اشاره خواهیم کرد. هدف بسیاری از این کاربردهای داده‌کاوی در کسب‌وکار، حل یک مسئله‌ی بغرنج یا جستجوی یک فرصت نوظهور کسب‌وکار برای ایجاد یک مزیت رقابتی پایدار است.

- **مدیریت ارتباط با مشتری.** مدیریت ارتباط با مشتری یک انشعاب از بازاریابی سنتی است. هدف از مدیریت ارتباط با مشتری، برقراری ارتباطات یک به یک با مشتریان از طریق توسعه‌ی درک نزدیک از نیازها و خواسته‌های آنان است. کسب‌وکارها همچنان که در طول زمان از طریق معاملات مختلف با مشتریان خود ارتباطاتی را برقرار می‌کنند (مثلاً استعلام محصول، فروش، درخواست خدمات، فراخوان گارانتی، اظهارنظرات در خصوص محصول، ارتباطات رسانه‌های اجتماعی)؛ حجم انبوهی از داده‌ها را جمع‌آوری می‌کنند. زمانیکه این داده‌ها با صفات و ویژگی‌های جمعیت‌شناختی و اجتماعی/اقتصادی در هم آمیخته شوند، این داده‌های غنی از اطلاعات را می‌توان برای (۱) شناسایی محتمل‌ترین

خریداران محصولات/خدمات جدید (یعنی تهیه پروفایل مشتریان)، (۲) درک علل ریشه‌ای ریزش مشتریان به منظور افزایش نرخ نگهداشت آنان، (۳) کشف وابستگی‌های متغیر زمانی بین محصولات و خدمات به منظور دستیابی به حداکثر فروش و ارزش مشتری و (۴) شناسایی سودآورترین مشتریان و نیازهای اولویت‌دار آنها به منظور تقویت ارتباطات و بهینه‌سازی فروش استفاده کرد.

- **بانک‌داری.** داده‌کاوی می‌تواند در موارد زیر به بانک‌ها کمک کند: (۱) خودکارسازی فرایند درخواست وام از طریق پیش‌بینی دقیق افرادی که به احتمال بسیار زیاد بدهی خود را در زمان سر رسید پرداخت نخواهند کرد، (۲) شناسایی کارت‌های اعتباری و تراکنش‌های بانک‌داری آنلاین مشکوک، (۳) شناسایی روش‌هایی برای خلق ارزش حداکثری برای مشتریان از طریق فروش محصولات و خدماتی که بیش از همه مایل به خرید آنها هستند، (۴) بهینه‌سازی نرخ بازده جریان نقدی از طریق پیش‌بینی دقیق جریان وجود نقد از طریق موجودیت‌های بانک‌داری از قبیل دستگاه‌های خودپرداز، شعب بانکی.
- **خرده‌فروشی و لجستیک.** در صنعت خرده‌فروشی، از داده‌کاوی می‌توان برای (۱) پیش‌بینی دقیق حجم فروش در محل‌های مشخص به منظور تعیین سطوح موجودی درست، (۲) شناسایی ارتباطات فروش بین محصولات مختلف (از طریق تحلیل سبد خرید) برای بهبود چیدمان فروشگاه و ارائه بهترین طرح‌های تبلیغاتی فروش، (۳) پیش‌بینی میزان مصرف انواع مختلف محصولات (بر اساس شرایط فصلی و محیطی) به منظور بهینه‌سازی لجستیک و دستیابی به حداکثر فروش و (۴) کشف الگوهای جالب در جابجایی محصولات (به ویژه محصولاتی که به دلیل انقضاء، آلودگی یا فسادپذیری دارای طول عمر نگهداری کوتاه هستند) در زنجیره تأمین از طریق تحلیل داده‌های بدست آمده از آر اف آی دی.
- **ساخت و تولید.** تولیدکنندگان می‌توانند از داده‌کاوی برای (۱) پیش‌بینی خرابی ماشین‌آلات با استفاده از داده‌های حسگرها پیش از وقوع آنها (نگهداری و تعمیرات شرطی)، (۲) شناسایی ناهنجاری‌ها و وجوه اشتراک در سیستم‌های تولید برای بهینه‌سازی ظرفیت تولید و (۳) کشف الگوهای نوین برای شناسایی و بهبود کیفیت محصول.
- **دلالی و مبادله‌ی اوراق قرضه.** دلال‌ها و تجار از داده‌کاوی برای (۱) پیش‌بینی زمان و میزان تغییر قیمت اوراق قرضه‌های خاص، (۲) پیش‌بینی دامنه و راستای نوسانات سهام، (۳) ارزیابی تأثیر رویدادها و اتفاقات خاص بر تحرکات کلی بازار و (۴) شناسایی و پیشگیری از فعالیت‌های مشکوک در مبادلات اوراق قرضه استفاده می‌کنند.
- **بیمه.** صنعت بیمه از تکنیک‌های داده‌کاوی برای (۱) پیش‌بینی مبلغ ادعاهای هزینه‌های پوشش پزشکی و ملکی برای برنامه‌ریزی بهتر کسب‌وکار، (۲) پیش‌بینی اینکه کدام مشتریان بیمه‌نامه‌های جدید را با شرایط ویژه خریداری خواهند کرد و (۴) شناسایی و پیشگیری پرداخت ادعاهای نادرست و فعالیت‌های مشکوک استفاده می‌کنند.
- **سخت‌افزار و نرم‌افزار رایانه.** از داده‌کاوی می‌توان برای (۱) پیش‌بینی خرابی‌های هارد دیسک‌ها پیش از آنکه اتفاق بیوفتند، (۲) شناسایی و فیلتر کردن پیام‌های نامه‌های الکترونیکی و محتوای وب ناخواسته، (۳) تشخیص و پیشگیری از نفوذهای امنیتی شبکه‌های رایانه‌ای و (۴) شناسایی محصولات نرم‌افزاری ناامن استفاده کرد.
- **دولت و دفاع.** داده‌کاوی دارای کاربردهای نظامی متعددی نیز می‌باشد و می‌توان از آن برای (۱) پیش‌بینی هزینه‌ی جابجایی تجهیزات و پرسنل نظامی، (۲) پیش‌بینی تحرکات دشمن و تدوین استراتژی‌های موفق‌تر برای درگیری‌های نظامی، (۳) پیش‌بینی مصرف منابع برای برنامه‌ریزی و بودجه‌ریزی بهتر و (۴) شناسایی گروه تجارب، استراتژی‌ها و دروس آموخته شده از عملیات نظامی برای به اشتراک‌گذاری بهتر دانش در سرتاسر سازمان استفاده کرد.

- **صنعت گردشگری (شرکت‌های هواپیمایی، هتل‌ها، شرکت‌های اجاره خودرو).** داده‌کاوی دارای کاربردهای متنوعی در صنعت گردشگری می‌باشد و می‌توان از آن برای (۱) پیش‌بینی فروش خدمات مختلف (انواع صندلی‌ها در هواپیماها، انواع اتاق‌ها در هتل‌ها، انواع خودرو در شرکت‌های اجاره‌ی خودرو) برای قیمت‌گذاری بهینه‌ی خدمات برای دستیابی به حداکثر درآمد به عنوان تابعی از تراکنش‌های متغیر زمانی (مدیریت بازده)، (۲) پیش‌بینی تقاضا در مکان‌های مختلف برای تخصیص بهتر منابع سازمانی محدود، (۳) شناسایی سودآورترین مشتریان و ارائه خدمات شخصی‌سازی شده به آنها برای تکرار خرید آنها و (۴) نگهداشت کارکنان ارزشمند از طریق شناسایی و اقدام بر روی علل اصلی ترک خدمت آنها استفاده کرد.
- **مراقبت سلامت.** داده‌کاوی کاربردهای متعددی در مراقبت سلامت دارد که می‌توان از آن برای (۱) شناسایی افراد فاقد بیمه و عوامل زیربنایی عدم استفاده آنها از بیمه، (۲) شناسایی ارتباطات جدید هزینه-منفعت بین درمان‌های مختلف به منظور تدوین استراتژی‌های مؤثرتر، (۳) پیش‌بینی سطح و زمان تقاضا در مکان‌های خدمت‌رسانی مختلف برای تخصیص بهینه‌سازی منابع سازمانی و (۴) درک علل زیربنایی ریزش مشتریان و کارکنان استفاده کرد.
- **پزشکی.** استفاده از داده‌کاوی در پزشکی به عنوان یک مکمل ارزشمند برای پژوهش‌های پزشکی سنتی محسوب می‌شود که عمدتاً ماهیت بالینی و بیولوژیکی دارد. تحلیل‌های داده‌کاوی می‌توانند (۱) الگوهای جدید را به منظور بهبود دوام‌پذیری بیماران مبتلا به سرطان شناسایی کنند، (۲) نرخ موفقیت بیماران پیوند اعضا را برای تدوین سیاست‌نامه‌های بهتر تطبیق اهداکننده‌ی عضو پیش‌بینی کنند، (۳) کارکردهای ژن‌های مختلف در کروموزوم‌های انسان را شناسایی کنند (ژنومیک)، و (۴) ارتباطات بین نشانه‌ها و بیماری‌ها را برای کمک به متخصصان و پزشکان برای اتخاذ تصمیمات به موقع، صحیح و آگاهانه کشف کنند.
- **صنعت سرگرمی.** از داده‌کاوی با موفقیت در صنعت سرگرمی برای (۱) تحلیل داده‌های بیننده برای تصمیم‌گیری در خصوص اینکه چه برنامه‌هایی در زمان اوج نمایش داده شوند و چگونه نرخ بازگشت بینندگان را با نمایش تبلیغات در مقاطع مناسب به حداکثر رساند، (۲) پیش‌بینی موفقیت مالی فیلم‌ها پیش از تولید برای اتخاذ تصمیمات سرمایه‌گذاری و بهینه‌سازی نرخ بازگشت سرمایه؛ (۳) پیش‌بینی تقاضا در مکان‌ها و زمان‌های مختلف برای زمانبندی بهتر رویدادهای سرگرمی و تخصیص بهینه‌سازی منابع و (۴) تدوین سیاست‌های قیمت‌گذاری بهینه برای دستیابی به حداکثر درآمد استفاده می‌شود.
- **امنیت ملی و اعمال قانون.** داده‌کاوی کاربردهای متعددی در امنیت ملی و اعمال قانون دارد. اغلب از آن برای (۱) شناسایی الگوی رفتارهای تروریستی (مورد کاربردی ۴.۳ به عنوان یک مثال از کاربرد داده‌کاوی برای ردیابی فعالیت‌های تأمین مالی تروریست‌ها)، (۲) کشف الگوی جرم و جنایت (مثلاً مکان‌ها، زمانبندی‌ها، رفتارهای جنایی و سایر ویژگی‌های مربوطه) برای حل به موقع پرونده‌های جنایی، (۳) پیش‌بینی و حذف حملات شیمیایی و بیولوژیکی بالقوه به زیرساخت‌های حیاتی کشورها از طریق تحلیل داده‌های بدست آمده از حسگرها و (۴) شناسایی و متوقف‌سازی حملات شرورانه به زیرساخت‌های اطلاعاتی حیاتی (که اغلب جنگ‌افزار اطلاعاتی^۱ نامیده می‌شوند).

¹ Information warfare

حمله تروریستی به مرکز تجارت جهانی در ۱۱ سپتامبر ۲۰۰۱ بر اهمیت اطلاعات منبع باز تأکید کرد. لایحه‌ی میهن دوستی ایالات متحده و تشکیل وزارت امنیت ملی ایالات متحده پیشبرنده‌ی کاربرد بالقوه‌ی فناوری اطلاعات و تکنیک‌های داده‌کاوی برای شناسایی پولشویی و شکل‌های دیگر تأمین مالی تروریست‌ها بودند. نهادهای اعمال قانون نیز تمرکزشان بر فعالیت‌های پولشویی از طریق تراکنش‌های عادی بانک‌ها و سایر سازمان‌های خدماتی مالی بوده است.

نهادهای اعمال قانون در حال حاضر تمرکز خود را بر قیمت‌گذاری تجارت بین‌الملل به عنوان یک ابزار تأمین مالی تروریسم معطوف کرده‌اند. پولشوها از تجارت بین‌الملل به منظور جابجایی مخفیانه پول‌ها به خارج از کشور و بدون جلب توجه دولت‌ها استفاده کرده‌اند. آنها از طریق بیش‌ارزش‌گذاری واردات و کم‌ارزش‌گذاری صادرات موفق به اینکار شده‌اند. به عنوان مثال، یک وارد کننده داخلی و صادرکننده خارجی می‌توانند با یکدیگر شراکت کرده و واردات را بیش‌ارزش‌گذاری کننده و بدین ترتیب پول را از کشور مبدأ انتقال دهند که منجر به جرائم کلاهبرداری گمرکی، تخطی از مالیات بر درآمد و پولشویی می‌شود. صادرکننده خارجی ممکن است عضوی از یک سازمان تروریستی باشد.

تکنیک‌های داده‌کاوی بر تحلیل داده‌های تراکنش‌های واردات و صادرات بدست آمده از وزارت بازرگانی ایالات متحده و نهادهای مربوطه تمرکز دارند. قیمت‌های وارداتی که فراتر از چارک بالایی قیمت واردات باشند و قیمت‌های صادرات که پایین‌تر از چارک پایینی قیمت صادرات باشند ردیابی می‌شوند. تمرکز بر قیمت‌های انتقالی غیرمعمول بین شرکت‌ها است که ممکن است منجر به انتقال درآمدهای مشمول مالیات و مالیات‌ها به خارج از ایالات متحده شود. هرگونه انحراف مشاهده شده در قیمت‌ها می‌تواند مربوط به اقدامات فرار مالیاتی (مالیات بر درآمد)، پولشویی یا تأمین مالی تروریسم باشد. انحراف قیمتی مشاهده شده نیز می‌تواند مربوط به بروز خطا در پایگاه داده تجارت ایالات متحده باشد.

داده‌کاوی منجر به ارزشیابی مؤثر داده‌ها می‌شود که آن نیز به نوبه‌ی خود به مبارزه با تروریسم کمک خواهد کرد. بکارگیری فناوری اطلاعات و تکنیک‌های داده‌کاوی بر تراکنش‌های مالی می‌تواند منجر به بهبود اطلاعات می‌شود.

Sources: J. S. Zdanowic, "Detecting Money Laundering and Terrorist Financing via Data Mining," *Communications of the ACM*, 47(5), May 2004, p. 53; R. J. Bolton, "Statistical Fraud Detection: A Review," *Statistical Science*, 17(3), January 2002, p. 235.

سوالاتی برای مورد کاربردی ۴.۳

۱. از داده‌کاوی چگونه می‌توان برای مبارزه با تروریسم استفاده کرد؟
۲. آیا به نظر شما داده‌کاوی اگرچه برای مبارزه با تروریسم ضروری است می‌تواند حقوق حریم خصوصی افراد را به مخاطره بیندازد؟

- **ورزش.** از داده‌کاوی برای بهبود عملکرد تیم‌های انجمن ملی بسکتبال (ان بی ای) در ایالات متحده استفاده شده است. تیم‌های لیگ برتر بیسیال نیز از تحلیل‌های پوشگویانه و داده‌کاوی برای بهره‌برداری بهینه از منابع محدودشان در یک فصل برنده استفاده می‌کنند. در واقع، اگر نگوییم تمامی ولی اکثر ورزش‌های حرفه‌ای امروزه از داده‌کاوی برای افزایش

شانس برنده شدن استفاده می‌کنند. در مقاله‌ای که توسط دیلن و همکاران (۲۰۱۲) منتشر شده، مدل‌های داده‌کاوی برای پیش‌بینی نتایج سوپر بول اتحادیه ملی ورزش دانشگاهی (ان سی ای ای) با استفاده از طیف گسترده‌ای از متغیرهای بازی‌های قبلی دو تیم رقیب ارائه شده است (جزئیات بیشتر در خصوص این مطالعه موردی در فصل ۳ ارائه شده است). رایت (۲۰۱۲) از انواع تحلیل‌های پوشش‌گویانه برای ارزیابی قهرمانی تیم بسکتبال مردان ان سی ای ای استفاده کرده است.

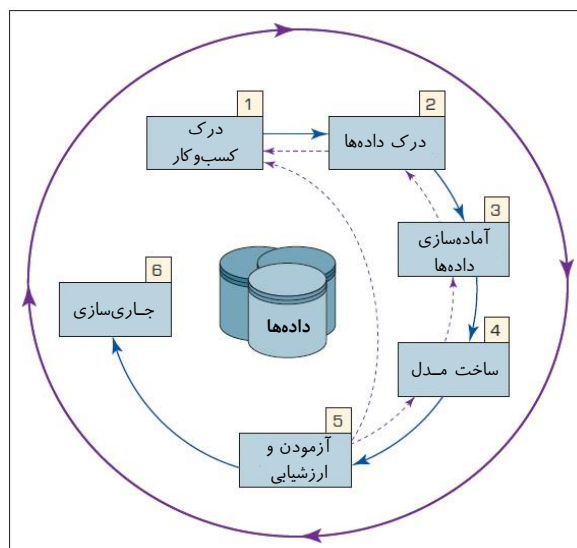
سوالاتی برای مرور بخش ۴.۳

۱. حوزه‌های کاربردی اصلی داده‌کاوی چه هستند؟
۲. حداقل پنج کاربرد خاص داده‌کاوی را شناسایی کرده و پنج ویژگی متداول این کاربردها را نام ببرید؟
۳. به نظر شما مهمترین حوزه کاربردی داده‌کاوی در حال حاضر چیست؟ چرا؟
۴. آیا می‌توانید حوزه‌های کاربردی دیگری از داده‌کاوی را نام ببرید که در این بخش به آنها اشاره نشده است؟ توضیح دهید.

برای اجرای سیستماتیک پروژه‌های داده‌کاوی، معمولاً از یک فرایند عمومی پیروی می‌شود. پژوهشگران و کارشناسان داده‌کاوی بر اساس بهترین تجارب، فرایندهای متعددی را برای بیشینه‌کردن احتمال موفقیت پروژه‌های داده‌کاوی پیشنهاد کرده‌اند (جریان کار یا رویکردهای گام به گام ساده). این تلاش‌ها منجر به چندین فرایند استانداردسازی شده گردیده است که برخی از آنها در این بخش توضیح داده شده‌اند.

یکی از این فرایندهای استانداردسازی شده، که شاید معروف‌ترین آنها نیز باشد، فرایند استاندارد صنعتی متقاطع برای داده‌کاوی یا **کریسپ دی‌ام^۱** است که در اواسط دهه ۹۰ توسط کنسرسیوم شرکت‌های اروپایی برای فعالیت به عنوان یک متدولوژی استاندارد غیراختصاصی داده‌کاوی پیشنهاد شد. شکل ۴.۳ این فرایند پیشنهادی را به تصویر می‌شکد که یک فرایند متوالی شش مرحله‌ای است که با درک مناسب و خوب کسب‌وکار و نیاز به پروژه‌های داده‌کاوی شروع شده (یعنی دامنه‌ی کاربرد) و با جاری‌سازی راه‌حلی که بتواند پاسخگوی آن نیازهای خاص باشد پایان می‌یابد. اگرچه این گام‌ها دارای ماهیت متوالی هستند ولی معمولاً رفت و برگشت‌های زیادی در آنها وجود دارد. با توجه به اینکه داده‌کاوی مبتنی بر تجربه و آزمایش است، بسته به وضعیت مسئله و تجربه/دانش تحلیلگر، کل فرایند ممکن است تکرارشونده (یعنی شخص باید انتظار داشته باشد که در بین گام‌ها به دفعات مکرر به جلو و عقب حرکت کند) و زمان‌بر باشد. از آنجائیکه گام‌های آخر بر اساس نتایج گام‌های اول انجام می‌شوند، باید توجه ویژه‌ای را به گام‌های اولیه داشت تا کل فرایند مطالعه داده‌کاوی در مسیر نادرستی قرار نگیرد.

¹ CRISP-DM



شکل ۴.۳ فرایند داده‌کاوی شش مرحله‌ای کریسپ دی‌ای

یک عنصر اساسی در هر مطالعه‌ی داده‌کاوی این است که بدانیم هدف از داده‌کاوی چیست. اینکار با درک عمیق و کامل نیازهای مدیریت به دانش جدید و مشخص‌سازی صریح اهداف کسب‌وکار در رابطه با مطالعه‌ای که قرار است انجام شود آغاز می‌شود. ما به اهداف مشخصی نیاز داریم که به سؤالاتی از قبیل «ویژگی‌های مشترک مشتریانی که اخیراً به رقبایمان باخته‌ایم چه هستند؟» یا «شرح حال‌های معمول مشتریان چه هستند و هر کدام از آنها چه میزان ارزش برای ما به همراه دارند؟» پاسخ دهند. سپس، طرح پروژه برای یافتن این دانش تدوین می‌شود که افراد مسئول برای گردآوری داده‌ها، تحلیل داده‌ها و گزارش‌دهی یافته‌ها در آن مشخص می‌شوند. در این مرحله‌ی ابتدایی، بودجه‌ی حمایتی مطالعه نیز باید حداقل در یک سطح بالا و با اعداد کلی تعیین شود.

یک مطالعه‌ی داده‌کاوی به طور خاص برای رسیدگی به یک وظیفه‌ی کسب‌وکار تعریف شده انجام می‌شود و وظایف کسب‌وکار مختلف به مجموعه داده‌های متفاوتی نیاز دارند. پس از گام درک کسب‌وکار، فعالیت اصلی فرایند داده‌کاوی شامل شناسایی داده‌های مرتبط از تعداد بسیار زیادی از پایگاه‌های داده می‌باشد. در فاز شناسایی و انتخاب داده‌ها باید برخی نکات مهم را در نظر گرفت. اول و مهمتر از همه، تحلیلگر باید در مورد شرح وظیفه‌ی داده‌کاوی واضح و موجز عمل کند تا بتواند مرتبط‌ترین داده‌ها را شناسایی کند. به عنوان مثال، در یک پروژه‌ی داده‌کاوی برای خرده‌فروشی‌ها می‌توان به دنبال شناسایی رفتارهای هزینه‌کرد مشتریان خانم پرداخت که لباس‌های فصلی خود را بر اساس جمعیت‌شناختی، تراکنش‌های کارت اعتباری و ویژگی‌های اقتصادی/اجتماعی خریداری می‌کنند. علاوه بر این، تحلیلگر باید درک عمیق و نزدیکی از منبع داده‌ها (مثلاً داده‌های مرتبط در کجا و به چه شکل ذخیره می‌شوند؛ از چه نوع فرایند گردآوری داده‌ها استفاده می‌شود - خودکار یا دستی؛ چه کسی داده‌ها را گردآوری می‌کند و داده‌ها در چه بازه‌های زمانی به‌روزآوری می‌شوند) و متغیرها (مثلاً مرتبط‌ترین متغیرها کدام هستند؟ آیا

متغیرهای مترادف و/یا همانم وجود دارند؟ آیا متغیرها مستقل از یکدیگر هستند؟- آیا آنها به عنوان یک منبع اطلاعاتی کامل و بدون همپوشانی یا تعارض اطلاعاتی هستند؟) داشته باشد.

برای درک بهتر داده‌ها، تحلیلگر اغلب از تکنیک‌های مختلف آماری و گرافیکی مثل چکیده‌های آماری ساده‌ی هر متغیر (مثلاً مقادیر میانگین، ماکزیمم/مینیمم، میانه و انحراف استاندارد برای متغیرهای عددی و جداول مُد و تواتر برای متغیرهای رسته‌ای) و تحلیل همبستگی، نمودارهای پراکندگی، هیستوگرام‌ها و نمودارهای جعبه‌ای استفاده می‌کند. شناسایی و انتخاب دقیق منبع داده‌ها و مرتبط‌ترین متغیرها می‌تواند به کشف سریع الگوهای دانشی مفید توسط الگوریتم‌های داده‌کاوی کمک کند.

منابع داده مورد نیاز برای انتخاب داده‌ها می‌توانند بسیار متنوع و متغیر باشند. منابع داده سنتی برای کسب‌وکارها شامل داده‌های جمعیت‌شناختی (مثل درآمد، آموزش، تعداد خانوار و سن)، داده‌های جامعه‌شناختی (مثل سرگرمی‌ها، عضویت در کلوپ‌ها و ...)، داده‌های تراکنش‌ها (رکورد فروش، هزینه‌کردهای کارت‌های اعتباری، چک‌های صادره) و غیره هستند. امروزه منابع داده از مخازن داده برون‌سازمانی (باز یا تجاری)، رسانه‌های اجتماعی و داده‌های تولید شده ماشینی نیز استفاده می‌کنند.

داده‌ها را می‌توان به گروه‌های کمی و کیفی تقسیم‌بندی کرد. داده‌های کمی با استفاده از مقادیر عددی یا داده‌های عددی اندازه‌گیری می‌شوند. این داده‌ها می‌توانند گسسته (اعداد صحیح) یا پیوسته (اعداد حقیقی) باشند. داده‌های کیفی که با نام داده‌های رسته‌ای نیز شناخته می‌شوند حاوی داده‌های اسمی و ترتیبی هستند. داده‌های اسمی دارای مقادیر نامرتب محدود (مثلاً داده‌های مربوط به جنسیت که دارای دو مقدار مرد و زن هستند) هستند. داده‌های ترتیبی دارای مقادیر مرتب محدود هستند. به عنوان مثال، رتبه‌بندی اعتبار مشتری یک داده‌های ترتیبی است چونکه رتبه‌بندی‌ها به صورت عالی، خوب و بد می‌باشد. یک طبقه‌بندی ساده‌ی داده‌ها (یعنی ماهیت داده‌ها) در فصل ۳ ارائه شده است.

داده‌های کمی را می‌توان به آسانی از طریق یک توزیع احتمال نمایش داد.

یک توزیع احتمال نحوه‌ی پراکندگی و شکل‌گیری داده‌ها را توضیح می‌دهد. به عنوان مثال، داده‌هایی که به صورت نرمال توزیع شده‌اند، متقارن بوده و عمدتاً به شکل یک منحنی زنگوله‌ای هستند. داده‌های کیفی را می‌توان با اعداد کدگذاری کرده و سپس با توزیع‌های تواتر توضیح داد. پس از انتخاب داده‌های مرتبط بر اساس هدف کسب‌وکار داده‌کاوی، فرایند پیش‌پردازش داده‌ها را باید شروع کرد.

هدف از آماده‌سازی داده‌ها (یا به بیان بهتر، پیش‌پردازش داده‌ها) این است که داده‌های شناسایی شده در گام قبل را گرفته و آن را برای تحلیل توسط روش‌های داده‌کاوی آماده کنیم. پیش‌پردازش داده‌ها در مقایسه با سایر گام‌های اشاره شده در کریسپ دی‌ام بیشترین زمان و انرژی را طلب می‌کند و اکثر افراد معتقدند که این گام ۸۰ درصد از کل زمان صرف شده در پروژه‌ی داده‌کاوی را به خود اختصاص می‌دهد. دلیل چنین زمان و انرژی عظیمی که صرف این گام می‌شود این واقعیت است که داده‌های دنیای واقعی معمولاً ناقص (فاقد مقادیر صفات، فاقد برخی صفات مورد نظر بوده یا فقط حاوی داده‌های تجمیعی هستند) و ناسازگار (کدها یا نام‌های آنها دارای مغایرت هستند) هستند. ماهیت داده‌ها و مسائل مربوط به پیش‌پردازش داده‌ها برای تحلیل‌ها به تفصیل در فصل ۳ توضیح داده شده است.

در این گام، تکنیک‌های مختلف مدل‌سازی انتخاب شده و برای رسیدگی به نیاز خاص کسب‌وکار بر روی مجموعه داده‌های آماده شده بکار گرفته می‌شوند. گام مدل‌سازی همچنین شامل ارزیابی و تحلیل مقایسه‌ای مدل‌های مختلفی است که ساخته شده‌اند.

از آنجائیکه به صورت کلی، بهترین روش یا الگوریتم برای داده‌کاوی وجود ندارد، باید از مجموعه‌ای از مدل‌های معتبر به همراه یک استراتژی ارزیابی و آزمایش برای شناسایی «بهترین» روش برای یک هدف خاص استفاده کرد. حتی برای یک روش یا الگوریتم، برای دستیابی به نتایج بهینه باید مجموعه‌ای از پارامترها را کالیبره نمود. برخی روش‌ها ممکن است الزاماتی در مورد روش قالب‌بندی داده‌ها داشته باشند که بدین ترتیب بازگشت به گام آماده‌سازی داده‌ها اغلب ضروری است. مورد کاربردی ۴.۴ یک مطالعه‌ی پژوهشی را ارائه می‌کند که در آن چندین نوع مدل ایجاد و با یکدیگر مقایسه شده‌اند.

مورد کاربردی ۴.۴

داده‌کاوی به پژوهش در مورد سرطان کمک می‌کند

طبق آمار انجمن سرطان آمریکا، نیمی از مردان و یک سوم از تمامی زنان در ایالات متحده در طول عمرشان مبتلا به سرطان خواهند شد که انتظار می‌رود که تقریباً ۱.۵ میلیون مورد سرطانی جدید در سال ۲۰۱۳ شناسایی گردد. سرطان، پس از بیماری‌های قلبی و عروقی دومین علت اصلی مرگ و میر در ایالات متحده و در جهان است. امسال، انتظار می‌رود بیش از ۵۰۰,۰۰۰ آمریکایی در اثر سرطان فوت کنند - بیش از ۱,۳۰۰ نفر در روز - که تقریباً یک چهارم از مرگ‌ها را به خود اختصاص داده است.

سرطان به گروهی از بیماری‌ها اطلاق می‌شود که معمولاً مربوط به رشد و گسترش کنترل‌ناپذیر سلول‌های نابهنجار است. اگر رشد و/یا گسترش این سلول‌ها کنترل نشود، سرطان می‌تواند منجر به مرگ شود. اگرچه علل اصلی سرطان ناشناخته است ولی هم عوامل خارجی (مثلاً توتون و تنباکو، ارگانوسم‌های عفونت‌زا، مواد شیمیایی و تشعشعات) و هم عوامل داخلی (جهش‌های موروثی، هورمون‌ها، شرایط ایمنی بدن و جهش‌هایی که در اثر متابولیسم رخ می‌دهند) علل آن دانسته شده‌اند. این عوامل می‌توانند همزمان یا پشت سرهم رخ داده و موجب شکل‌گیری یا گسترش سرطان‌زایی شوند. سرطان به روش‌های جراحی، پرتوافکنی، شیمی‌درمانی، هورمون‌تراپی، درمان زیستی و درمان هدفمند درمان می‌شود. آمارهای درمان‌های موفق بسته به نوع سرطان و مرحله‌ی تشخیص بسیار متغیر است.

بقای نسبی پنج ساله برای تمامی سرطان‌ها در حال افزایش است و نرخ کاهش مرگ و میرهای ناشی از سرطان نیز به ۲۰٪ در سال ۲۰۱۳ رسیده است که بدین معناست که مرگ و میرهای ناشی از سرطان از سال ۱۹۹۱ به میزان ۱.۲ میلیون نفر کاهش داشته است. یعنی روزانه زندگی بیش از ۴۰۰ نفر نجات می‌یابد. این بهبود در نرخ بقا بیانگر پیشرفت در تشخیص سرطان‌های خاص در مراحل اولیه و بهبود در روش‌های درمان است. برای پیشگیری و درمان سرطان به بهبودهای بیشتری نیاز داریم.

اگرچه پژوهش در حوزه سرطان ذاتاً بالینی و بیولوژیکی بوده است ولی در سال‌های اخیر، مطالعات تحلیلی مبتنی بر داده‌ها نیز به آنها افزوده شده است. در حوزه‌های پزشکی که پژوهش‌های مبتنی بر داده و مبتنی بر تحلیل‌ها با موفقیت بکار گرفته شده‌اند، جهت‌گیری‌های پژوهشی نوینی به منظور پیشبرد مطالعات بالینی و بیولوژیکی شناسایی شده است. استفاده از انواع مختلف داده‌ها من جمله داده‌های مولکولی، بالینی، کارآزمایی بالینی و ادبیات موضوع به همراه ابزارها و تکنیک‌های مناسب داده‌کاوی موجب شده تا پژوهشگران قادر به شناسایی الگوهای جدید و هموار کردن مسیر به سمت یک جامعه‌ی عاری از سرطان باشند.

در مطالعه‌ای که توسط دِلِن (۲۰۰۹) انجام شده، از سه تکنیک متداول داده‌کاوی یعنی درخت تصمیم، شبکه‌های عصبی مصنوعی و ماشین‌های برداری پشتیبان به همراه رگرسیون لجستیک برای توسعه‌ی مدل‌های پیش‌بینی احتمال نجات از سرطان پروستات استفاده شده است. این مجموعه داده حاوی ۱۲۰,۰۰۰ رکورد و ۷۷ متغیر است. از یک متدولوژی اعتبارسنجی متقاطع k مرتبه‌ای در ساخت، ارزیابی و مقایسه مدل استفاده شده است. نتایج نشان می‌دهد که مدل‌های برداری پشتیبان، دقیق‌ترین مدل برای پیشگویی در این زمینه هستند (دقت آزمون ۹۲.۸۵٪) که پس از آن شبکه‌های عصبی مصنوعی و درخت تصمیم قرار می‌گیرند. علاوه بر این، در این مطالعه با استفاده از یک روش ارزیابی مبتنی بر تحلیل حساسیت، الگوهای نوین مربوط به عوامل اطلاع قبلی از سرطان پروستات شناسایی شده است.

دِلِن، واکر و کادام^۱ (۲۰۰۵) در مطالعه‌ای از دو الگوریتم داده‌کاوی (شبکه‌های عصبی مصنوعی و درخت تصمیم) و رگرسیون لجستیک برای توسعه‌ی مدل‌های پیش‌بینی بقاء سرطان پستان با استفاده از یک مجموعه داده‌ی بزرگ (بیش از ۲۰۰,۰۰۰ پرونده) استفاده کردند. پژوهشگران با استفاده از یک روش اعتبارسنجی متقاطع ۱۰ مرتبه‌ای، میزان انحراف تخمینی مدل‌های پیش‌بینی در مقایسه عملکردها را اندازه‌گیری کردند که نتایج نشان می‌دهد که درخت تصمیم (الگوریتم C5) با دقت ۹۳.۶ درصد بر روی نمونه، بهترین پیش‌بینی کننده بوده است و پس از آن شبکه‌های عصبی مصنوعی با ۹۱.۲ درصد دقت و رگرسیون لجستیک با ۸۹.۲ درصد دقت قرار دارند. تحلیل‌های بیشتر بر روی مدل‌های پیش‌بینی، اهمیت اولویت‌بندی شده‌اند عوامل اطلاع قبلی را مشخص کرده است که می‌توان از آن به عنوان مبنای مطالعات پژوهشی بالینی و بیولوژیکی استفاده کرد.

ذوالبانین و همکاران^۲ (۲۰۱۵) در جدیدترین مطالعه خود به بررسی تأثیر اختلال همراه بر امکان نجات از سرطان پرداخته‌اند. اگرچه پژوهش‌های پیشین نشان داده است که توصیه‌های تشخیص و درمان را ممکن است بتوان بر اساس شدت اختلال‌های همراه تغییر داد ولی بیماری‌های عروقی و قلبی در اکثر موارد همچنان به طور جداگانه از یکدیگر بررسی می‌شوند. آنها به منظور نشان دادن اهمیت بیماری‌های قلبی و عروقی همزمان در مسیر درمان از داده‌های برنامه‌ی « نظارت، اپیدمیولوژی و نتایج نهایی^۳ » برای ایجاد دو مجموعه داده‌های اختلال همراه استفاده کردند: یکی برای سرطان پستان و آلت تناسلی خانم‌ها و دیگری برای سرطان پروستات و مجاری ادراری. سپس از چندین تکنیک یادگیری ماشین برای ایجاد مدل‌های پیشگویانه استفاده شد (شکل ۴.۴). مقایسه‌ی نتایج نشان می‌دهند که داشتن اطلاعات بیشتر در مورد شرایط اختلال همراه بیماران می‌تواند قدرت پیشگویانه‌ی مدل‌ها را بهبود بخشد که آن نیز به نوبه‌ی خود می‌تواند به متخصصان در تصمیمات تشخیص و درمان بهتر کمک کند. از این رو، این مطالعه بیان می‌کند که شناسایی، ثبت و بکارگیری مناسب وضعیت اختلال همراه بیماران می‌تواند هزینه‌های درمان را کاهش داده و موجب آسان‌تر شدن چالش‌های اقتصادی مرتب با مراقبت بهداشت شود.

این مثال‌ها (و بسیاری دیگر در ادبیات پزشکی) نشان می‌دهند که از تکنیک‌های پیشرفته‌ی داده‌کاوی می‌توان برای توسعه‌ی مدل‌هایی که درجه‌ی پیشگویی و قدرت توضیحی بالایی دارند استفاده کرد. اگرچه روش‌های داده‌کاوی قادر به استخراج الگوها و ارتباطات پنهان در اعماق پایگاه‌های داده‌ی پزشکی بزرگ و پیچیده هستند ولی بدون همکاری و بازخورد از جانب خبرگان پزشکی، نتایج آنها چندان مفید نخواهند بود. الگوهای یافت شده از طریق روش‌های داده‌کاوی باید توسط متخصصان پزشکی که

¹ Delen, Walker and Kadam

² Zolbanin et al.

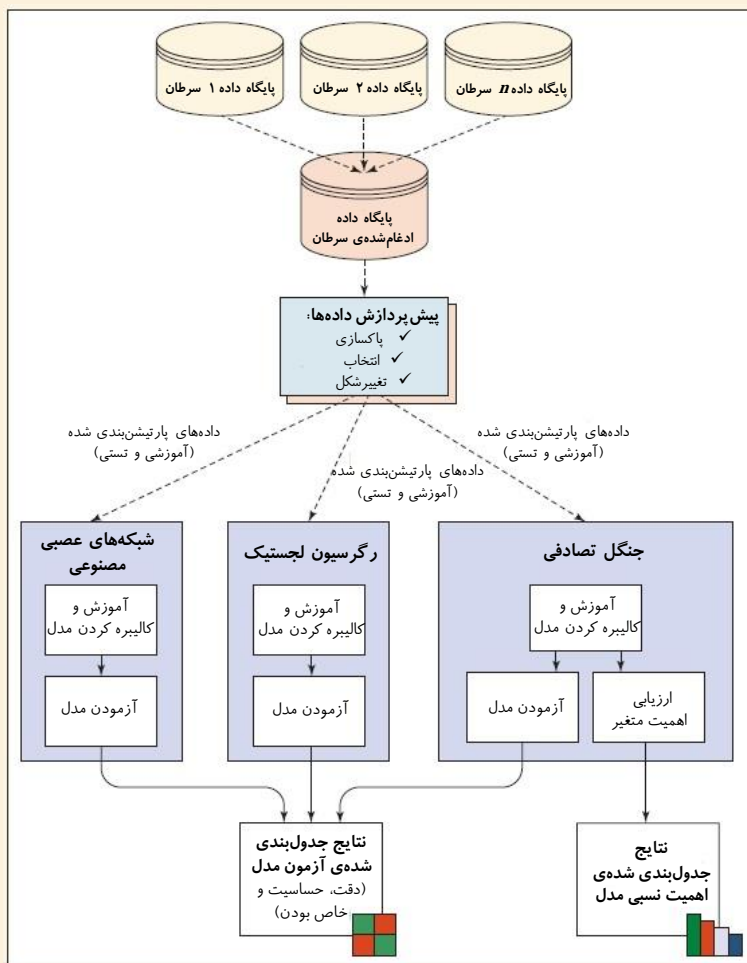
³ Surveillance, Epidemiology, and End Results (SEER)

دارای چندین سال تجربه در حوزه مسئله هستند مورد ارزشیابی قرار بگیرند تا بتوانند تصمیم بگیرند که آیا این روش‌ها به اندازه کافی منطقی، عملی و نوین هستند تا بتوانند جهت‌گیری‌های پژوهشی جدید را تضمین کنند یا خیر. به طور خلاصه، داده‌کاوی جایگزین پژوهشگران و متخصصان پزشکی نخواهد شد بلکه مکمل تلاش‌های ارزشمند آنها برای ارائه‌ی جهت‌گیری‌های پژوهشی نوین مبتنی بر داده‌ها و نهایتاً نجات جان انسان‌ها است.

Sources: H. M. Zolbanin, D. Delen, & A. H. Zadeh, "Predicting Overall Survivability in Comorbidity of Cancers: A Data Mining Approach," *Decision Support Systems*, 74, 2015, pp. 150–161; D. Delen, "Analysis of Cancer Data: A Data Mining Approach," *Expert Systems*, 26(1), 2009, pp. 100–112; J. Thongkam, G. Xu, Y. Zhang, & F. Huang, "Toward Breast Cancer Survivability Prediction Models Through Improving Training Space," *Expert Systems with Applications*, 36(10), 2009, pp. 12200–12209; D. Delen, G. Walker, & A. Kadam, "Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods," *Artificial Intelligence in Medicine*, 34(2), 2005, pp. 113–127.

سوالاتی برای مورد کاربردی ۴.۴

۵. از داده‌کاوی چگونه می‌توان برای درمان بیماری‌هایی مثل سرطان استفاده کرد؟
۶. به نظر شما وعده‌ها و چالش‌های اساسی داده‌کاوا در پژوهش‌های پزشکی و بیولوژیکی چیست؟



شکل ۴.۴ متدولوژی داده‌کاوی برای تحقیق اختلال همراه در قابلیت بقاء سرطان

بسته به نیاز کسب‌وکار، داده‌کاوی را می‌توان برای یک پیش‌بینی (طبقه‌بندی یا رگرسیون)، یک وابستگی یا یک خوشه‌بندی بکار برد. هر یک از این وظایف داده‌کاوی از انواع مختلف الگوریتم‌ها و روش‌های داده‌کاوی استفاده می‌کنند. برخی از این روش‌های داده‌کاوی در این فصل توضیح داده شدند و برخی از محبوب‌ترین الگوریتم‌ها من جمله درخت تصمیم برای طبقه‌بندی، میانگین‌گیری برای خوشه‌بندی و الگوریتم آپریوری برای کاوش قواعد وابستگی را در ادامه‌ی این فصل توضیح خواهیم داد.

در گام ۵، مدل‌های توسعه یافته از لحاظ دقت و عمومیت مورد ارزیابی و ارزشیابی قرار می‌گیرند. در این گام، ارزیابی می‌شود که مدل (یا مدل‌های) انتخاب شده کدام اهداف کسب‌وکار را برآورده می‌کنند و تا چه اندازه (یعنی اینکه آیا لازم است مدل‌های بیشتری توسعه داده شده و ارزیابی شوند؟) یک گزینه‌ی دیگر این است که مدل‌های توسعه یافته را در صورتیکه محدودیت‌های زمانی و بودجه اجازه می‌دهد در قالب سناریوی دنیای واقعی تست کنیم. اگرچه انتظار می‌رود که نتیجه‌ی مدل‌های توسعه یافته با اهداف اصلی کسب‌وکار مرتبط باشد ولی یافته‌های دیگری که لزوماً مرتبط با اهداف اصلی کسب‌وکار نیستند ولی ممکن است اطلاعات یا نکات بیشتری را در مورد جهت‌گیری‌های آینده آشکار کنند اغلب کشف می‌شوند. گام آزمون و ارزشیابی یک وظیفه‌ی حیاتی و چالش‌برانگیز است. داده‌کاوی هیچ ارزش افزوده‌ای به همراه نخواهد داشت تا زمانیکه الگوهای دانشی کشف شده بتوانند بر ارزش کسب‌وکار بیفزایند. تعیین ارزش کسب‌وکار مربوط به الگوهای دانشی کشف شده مشابه حل کردن پازل است. الگوهای دانشی استخراج شده، تکه‌هایی از پازل هستند که باید کنار یکدیگر قرار بگیرند. موفقیت در اینکار به تعامل بین تحلیلگران داده‌ها، تحلیلگران کسب‌وکار و تصمیم‌گیرندگان (مثل مدیران کسب‌وکار) بستگی دارد. از آنجائیکه تحلیلگران داده ممکن است درک کاملی از اهداف داده‌کاوی و مفهوم آن برای تحلیلگران کسب‌وکارها نداشته باشند و تصمیم‌گیرندگان ممکن است از دانش فنی برای تفسیر نتایج راه‌حل‌های ریاضی پیچیده برخوردار نباشند، تعامل بین آنها ضروری است. برای تفسیر مناسب الگوهای دانشی، اغلب لازم است تا از انواع تکنیک‌های مصورسازی و جدول‌بندی استفاده شود (مثلاً نمودارهای کلوچه‌ای، هیستوگرام‌ها، نمودارهای جعبه‌ای، نمودارهای پراکندگی، جداول محوری، جدول‌بندی متقاطع یافته‌ها).

توسعه و ارزیابی مدل‌ها، پایان یک پروژه‌ی داده‌کاوی نیست. حتی اگر هدف از یک مدل این باشد که صرفاً به اکتشاف ساده در داده‌ها بپردازیم، دانش کسب شده از این اکتشاف را باید به گونه‌ای سازماندهی و ارائه کرد که کاربر نهایی قادر به درک و بهره‌گیری از آن باشد. بسته به الزامات و نیازمندی‌ها، فاز جاری‌سازی می‌تواند تا حد یک تولید گزارش، ساده بوده یا اینکه تا پیاده‌سازی بیک فرایند داده‌کاوی تکرارپذیر در سطح سازمان، پیچیده باشد. در بسیاری از موارد، اجرای گام‌های جاری‌سازی بر عهده‌ی مشتری است نه تحلیلگر داده. با وجود این، حتی اگر تحلیلگر داده فعالیت جاری‌سازی را انجام ندهد، برای مشتری بسیار مهم است که اقداماتی که لازم است انجام شود تا مدل‌های ایجاد شده قابل استفاده شوند را از قبل درک کرده باشد.

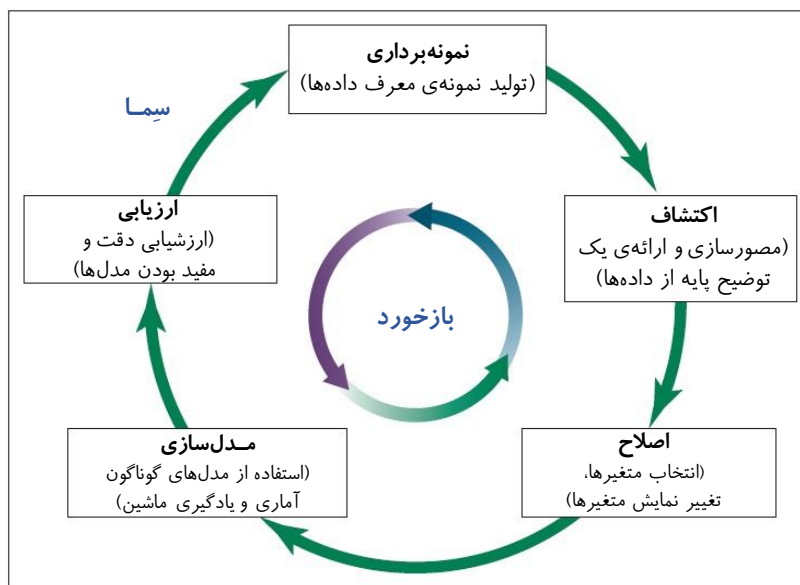
گام جاری‌سازی همچنین می‌تواند شامل فعالیت‌های نگهداری مدل‌های جاری‌سازی شده باشد. چونکه همه چیز در مورد کسب‌وکار به طور مداوم در حال تغییر است، داده‌هایی که بیانگر فعالیت‌های کسب‌وکار هستند نیز دستخوش تغییر می‌شوند. در طول زمان، مدل‌های (و الگوهای درون آنها) ساخته شده بر اساس داده‌های قدیمی ممکن است منسوخ، نامربوط یا گمراه کننده شوند. از این رو، در صورتیکه قرار باشد تا نتایج داده‌کاوی به عنوان بخشی از محیط و کسب‌وکار روزمره ما درآیند، پایش و نگهداری مدل‌ها بسیار اهمیت خواهد یافت. آماده‌سازی دقیق یک استراتژی نگهداری به ما کمک می‌کند تا از بازه‌های طولانی

استفاده‌ی نادرست از نتایج داده‌کاوی اجتناب کنیم. به منظور پایش جای‌سازی نتایج داده‌کاوی، نیاز به یک طرح مفصل در خصوص فرایند پایش می‌باشد که ممکن است در خصوص مدل‌های داده‌کاوی پیچیده کار چندان ساده‌ای نباشد.

یک مطالعه‌ی داده‌کاوی برای اینکه موفقیت‌آمیز باشد باید به صورت یک فرایند در نظر گرفته شود که از یک متدولوژی استانداردسازی شده پیروی می‌کند نه اینکه مجموعه‌ای از تکنیک‌ها و ابزارهای نرم‌افزاری خودکار باشد. علاوه بر کریسپ دی‌ام، متدولوژی معروف دیگری نیز توسط مؤسسه‌ی اس‌ای‌اس به نام سِما^۱ (۲۰۰۹) توسعه یافته است. که مخفف «نمونه‌برداری، اکتشاف، اصلاح، مدل‌سازی و ارزیابی» است.

بر اساس سِما که کارش را با یک نمونه‌ی معرف آماری از داده‌ها آغاز می‌کند، تکنیک‌های مصورسازی و آماری اکتشافی را به آسانی می‌توان بکار گرفت، مهمترین متغیرهای پیشگویانه را انتخاب کرده و تبدیل نمود، متغیرها را برای پیش‌بینی نتایج مدل‌سازی کرد و دقت مدل را تأیید نمود. یک نمایش تصویری از فرایند سِما در شکل ۴.۵ ارائه شده است.

با ارزیابی نتایج هر مرحله در فرایند سِما، توسعه‌دهنده‌ی مدل می‌تواند تعیین کند که چگونه سوالات مطرح شده به واسطه‌ی نتایج قبلی را مدل‌سازی کند و بدین ترتیب مجدداً برای تصفیه‌ی بیشتر داده‌ها به فاز اکتشاف باز گردد. این بدان معنی است که سِما همانند کریسپ دی‌ام دارای یک چرخه‌ی آزمایش شدیداً تکرارشونده است. تفاوت اصلی بین کریسپ دی‌ام و سِما در این است که کریسپ دی‌ام رویکردی جامع‌تر نسبت به پروژه‌های داده‌کاوی دارد - درک کسب‌وکار و داده‌های مربوط به آن - در حالیکه سِما به طور ضمنی فرض می‌کند که اهداف و مقاصد پروژه‌های داده‌کاوی به همراه منابع داده مناسب از قبل شناسایی و درک شده‌اند.

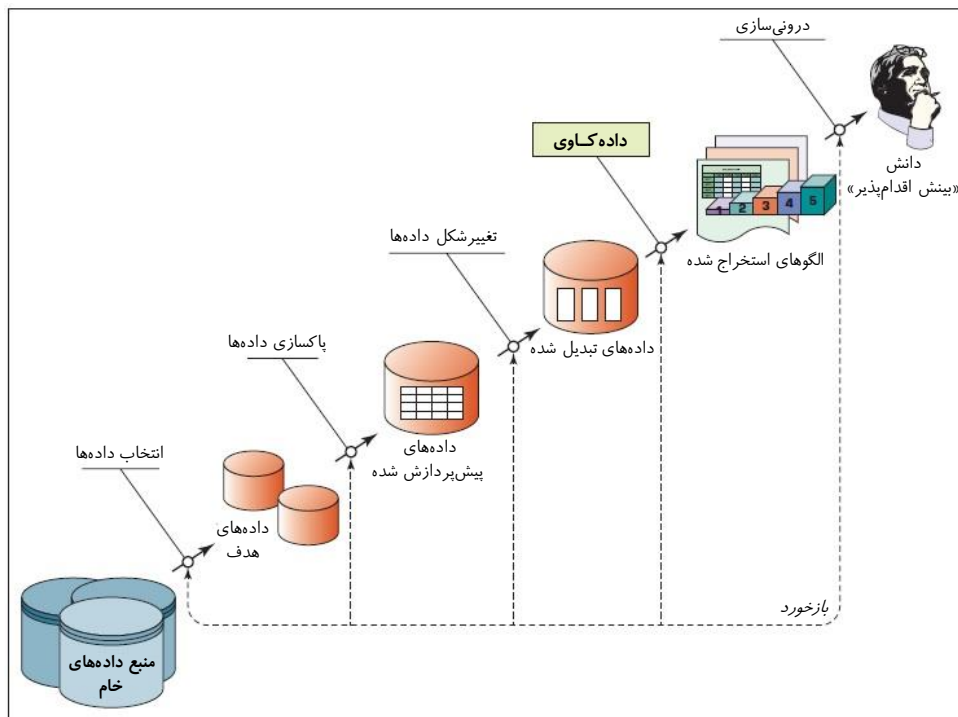


شکل ۴.۵ فرایند داده‌کاوی سِما

¹ SEMMA

برخی متخصصان عمدتاً عبارت **کشف دانش در پایگاه‌های داده**^۱ را مترادف با داده‌کاوی بکار می‌برند. فیاد و همکاران (۱۹۹۶) کشف دانش در پایگاه‌های داده را به صورت فرایند استفاده از روش‌های داده‌کاوی برای یافتن اطلاعات و الگوهای مفید در داده‌ها می‌دانند که نقطه مقابل داده‌کاوی استفاده از الگوریتم‌ها برای شناسایی الگوها در داده‌های بدست آمده از طریق فرایند داده را مترادف با داده‌کاوی بکار می‌برند. فیاد و همکاران^۲ (۱۹۹۶) کشف دانش در پایگاه‌های داده است (شکل ۴.۶). کشف دانش در پایگاه‌های داده فرایند جامعی است که داده‌کاوی را نیز در بر می‌گیرد. ورودی فرایند کشف دانش در پایگاه‌های داده شامل داده‌های سازمانی می‌باشد. با استفاده از انبار داده‌های سازمانی می‌توان این فرایند را به طور موثری پیاده‌سازی نمود زیرا تنها یک منبع داده برای کاوش وجود دارد. دانهام (۲۰۰۳) فرایند کشف دانش در پایگاه‌های داده را طی گام‌های زیر خلاصه کرده است: انتخاب داده‌ها، پیش‌پردازش داده‌ها، تغییر شکل داده‌ها، کاوش داده‌ها و تفسیر/ارزشیابی داده‌ها.

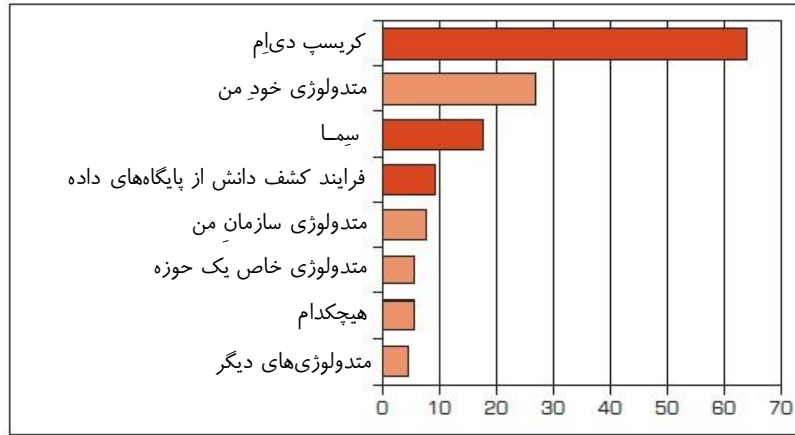
شکل ۴.۷ نتایج رأی‌گیری به این سوال را نشان می‌دهد: «متدولوژی اصلی مورد استفاده‌ی شما برای داده‌کاوی چیست؟» (توسط **KDnuggets.com** در آگوست سال ۲۰۰۷ انجام شده است).



شکل ۴.۶ فرایند کشف دانش در پایگاه‌های داده (کی دی دی)

¹ Knowledge Discovery in Databases (KDD)

² Fayyad et al.



شکل ۴.۷ رتبه‌بندی متدولوژی‌ها/فرایندهای داده‌کاوی

Source: Used with permission from KDnuggets.com.

سوالاتی برای مرور بخش ۴.۴

۱. فرایندهای اصلی داده‌کاوی کدام‌ها هستند؟
۲. فکر می‌کنید چرا فازهای ابتدایی (درک کسب‌وکار و درک داده‌ها) بیشترین زمان را در پروژه‌ی داده‌کاوی به خود اختصاص داده‌اند؟
۳. فازهای فرایند کریسپ دی‌ام را نام برده و به اختصار توضیح دهید.
۴. گام‌های اصلی پیش‌پردازش داده‌ها چه هستند؟ هر گام را به اختصار توضیح داده و مثال بزنید.
۵. تفاوت کریسپ دی‌ام با سیما چیست؟

روش‌های مختلفی برای انجام مطالعات داده‌کاوی وجود دارد که از آن جمله می‌توان به طبقه‌بندی، رگرسیون، خوشه‌بندی و وابستگی اشاره کرد. اکثر ابزارهای نرم‌افزاری داده‌کاوی از بیش از یک تکنیک (یا الگوریتم) برای هر یک از این روش‌ها استفاده می‌کنند. در این بخش به توضیح محبوب‌ترین و متداول‌ترین روش‌های داده‌کاوی خواهیم پرداخت و تکنیک‌های بارز آنها را توضیح خواهیم داد.

طبقه‌بندی را شاید بتوان پرکاربردترین روش داده‌کاوی برای مسائل دنیای واقعی دانست. طبقه‌بندی که عضو معروفی از تکنیک‌های خانواده‌ی یادگیری ماشین است الگوها را از داده‌های گذشته می‌آموزد (مجموعه‌ای از اطلاعات - صفات، رفتارها، متغیرها و ویژگی‌های - مربوط به رویدادها یا اشیاء) و نمونه‌های جدید (با برجسب‌های ناشناس) را در گروه‌ها یا دسته‌های متناظر آنها قرار می‌دهد. به عنوان مثال، می‌توان از طبقه‌بندی برای پیش‌بینی آب و هوا در یک روز خاص که «آفتابی»، «بارانی» یا «بری» خواهد بود استفاده کرد. وظایف متداول طبقه‌بندی شامل تأیید اعتبار (یعنی ریسک اعتبار خوب یا بد)، محل ذخیره‌سازی

(مثلاً خوب، متوسط، بد)، بازاریابی هدف (مثلاً مشتری احتمالی، مشتری نامحتمل)، تشخیص کلاهبرداری (یعنی بله/خیر) و مخابرات (مثلاً احتمال استفاده از یک شرکت مخابراتی دیگر، بله/خیر) می‌باشند. اگر آنچه که قرار است پیش‌بینی شود یک برچسب دسته باشد (مثلاً «آفتابی»، «بارانی»، «ابری»)، مسئله‌ی پیش‌بینی را طبقه‌بندی می‌نامند و اگر یک مقدار عددی باشد (مثلاً درمای ۶۸ درجه‌ی فارنهایت)، مسئله‌ی پیش‌بینی را **رگرسیون** می‌نامند.

اگرچه از خوشه‌بندی (یک روش متداول دیگر برای داده‌کاوی) نیز می‌توان برای تعیین گروه اشیاء (یا عضویت دسته) استفاده کرد، یک تفاوت اساسی بین این دو وجود دارد. دسته‌بندی مشخصه‌های بین اشیاء (یعنی متغیرهای مستقل) و عضویت آنها (یعنی متغیر خروجی) را از طریق یک فرایند یادگیری با نظارت می‌آموزد که در آن، هر دو نوع متغیر (ورودی و خروجی) وارد الگوریتم می‌شوند ولی در خوشه‌بندی، عضویت اشیاء از طریق یک فرایند یادگیری بدون نظارت آموخته می‌شود که فقط متغیرهای ورودی به الگوریتم داده می‌شوند. برخلاف طبقه‌بندی، تکنیک خوشه‌بندی فاقد مکانیزم نظارت (یا کنترل) است در حالیکه الگوریتم‌های خوشه‌بندی از یک یا چند روش هیوریستیک (مثلاً سنج‌های فاصله‌ی چندبعدی) برای کشف گروه‌بندی‌های طبیعی اشیاء استفاده می‌کند.

متداول‌ترین متدولوژی دو مرحله‌ای برای پیش‌بینی از طریق طبقه‌بندی شامل توسعه/آموزش مدل و آزمون/جاری‌سازی مدل است. در فاز توسعه‌ی مدل، مجموعه‌ای از داده‌های ورودی از جمله برچسب‌های واقعی دسته‌ها استفاده می‌شوند. پس از اینکه مدل آموزش داده شد، به منظور ارزیابی دقت توسط داده‌های نمونه آزموده شده و نهایتاً برای پیش‌بینی دسته‌های جدیدی از نمونه داده‌ها (که برچسب دسته مشخص نیست) به طور واقعی جاری‌سازی می‌شود. در ارزیابی مدل عوامل متعددی را باید در نظر گرفت:

- **دقت پیش‌بینی.** توانایی مدل برای پیش‌بینی درست برچسب دسته‌ی جدید یا داده‌هایی که قبلاً نادیده گرفته شده‌اند. دقت پیش‌بینی پرکاربردترین عامل ارزیابی مدل‌های طبقه‌بندی است. برای محاسبه‌ی این سنجه، برچسب‌های دسته‌ی حقیقی مجموعه داده‌های تستی با برچسب دسته‌های پیش‌بینی شده توسط مدل تطبیق داده می‌شوند. سپس می‌توان دقت را به صورت نرخ دقت محاسبه کرد که درصدی از نمونه داده‌های تستی هستند که به درستی توسط مدل دسته‌بندی شده‌اند.
- **سرعت.** هزینه‌های محاسباتی برای ایجاد و بکارگیری مدل‌ها که هرچه سریع‌تر باشند بهتر هستند.
- **استقامت.** توانایی مدل برای انجام پیش‌بینی تا حدی معقول دقیق با توجه به داده‌های نویز دارد یا داده‌هایی که مقادیر آنها اشتباه بوده یا فاقد مقدار هستند.
- **مقیاس‌پذیری.** توانایی ساخت یک مدل پیش‌بینی به شکل مؤثر با حجم انبوهی از داده‌ها.
- **قابلیت تفسیرپذیری.** سطح درک و بینش ارائه شده توسط مدل (مثلاً مدل چگونه در خصوص پیش‌بینی‌های خاص نتیجه‌گیری می‌کند).

در مسائل طبقه‌بندی، منبع اصلی برای تخمین دقت، ماتریس درهم ریختگی^۱ (ماتریس طبقه‌بندی^۲ یا جدول اقتضایی^۳) نام دارد. شکل ۴.۸ یک ماتریس درهم ریختگی را برای یک مسئله‌ی طبقه‌بندی دو دسته‌ای نشان می‌دهد. اعداد نوشته بر روی قطر از گوشه بالا سمت چپ تا پایین سمت راست بیانگر تصمیمات صحیح و اعداد خارج از قطر بیانگر تصمیمات اشتباه هستند. در جدول ۴.۱ معادلات مربوط به سنج‌های متداول برآورد دقت مدل‌های طبقه‌بندی ارائه شده است. زمانیکه مسئله‌ی طبقه‌بندی از نوع دودویی نباشد، ماتریس درهم ریختگی بزرگتر شده (یک ماتریس چهارگوش با تعداد منحصری‌فرد برچسب دسته‌ها) و سنج‌های دقت محدود به دو سنج می‌شوند: نرخ دقت به ازای هر دسته^۴ و دقت کلی طبقه‌بند^۵.

$$(True\ Classification\ Rate)_i = \frac{(True\ Classification)_i}{\sum_{i=1}^n (False\ Classification)_i}$$

$$(Overall\ Classifier\ Accuracy)_i = \frac{\sum_{i=1}^n (True\ Classification)_i}{Total\ Number\ of\ Cases}$$

برآورد کردن یا تخمین دقت یک مدل طبقه‌بندی (یا طبقه‌بند) که توسط یک الگوریتم یادگیری با نظارت القا شده باشد به دو دلیل بسیار اهمیت دارد: اول اینکه می‌توان از آن برای تخمین دقت پیش‌بینی آینده آن مدل استفاده کرد که به معنای سطح اطمینانی است که می‌توان به خروجی طبقه‌بند در یک سیستم پیش‌بینی داشت. دوم اینکه می‌توان از آن برای انتخاب یک طبقه‌بند از بین یک مجموعه استفاده کرد (شناسایی «بهترین» مدل طبقه‌بندی از بین تعدادی زیاد). در ادامه به معرفی متداول‌ترین متدولوژی‌های تخمین برای مدل‌های داده‌کاوی از نوع طبقه‌بندی می‌پردازیم.

		دسته‌ی مشاهده شده / صحیح	
		مثبت	منفی
دسته‌ی پیش‌بینی شده	مثبت	شمارش مثبت صحیح	شمارش مثبت کاذب
	منفی	شمارش منفی کاذب	شمارش منفی صحیح

شکل ۴.۸ ماتریس درهم ریختگی ساده برای جدول‌بندی نتایج طبقه‌بندی دو دسته‌ای

- 1 Confusion Matrix
- 2 Classification Matrix
- 3 Contingency Table
- 4 Per class accuracy rate
- 5 Overall classifier accuracy

جدول ۴.۱ سنجه‌های متداول دقت برای مدل‌های طبقه‌بندی

سنجه	شرح
$Accuracy = \frac{TP+TN}{TP+TN+FN+FP}$	نسبت دسته‌ها یا نمونه‌هایی که به درستی طبقه‌بندی شده‌اند (مثبت و منفی) تقسیم بر کل تعداد نمونه‌ها یا دسته‌ها
$TruePositiveRate = \frac{TP}{TP+FN}$	(یا حساسیت)، نسبت نمونه‌های مثبت که به درستی طبقه‌بندی شده‌اند تقسیم بر تعداد کل نمونه‌های مثبت شمارش شده
$TrueNegativeRate = \frac{TN}{TN+FP}$	(یا ویژگی)، نسبت نمونه‌های منفی که به درستی طبقه‌بندی شده‌اند تقسیم بر تعداد کل نمونه‌های منفی شمارش شده
$Precision = \frac{TP}{TP+FP}$	نسبت نمونه‌های مثبت که به درستی طبقه‌بندی شده‌اند تقسیم بر جمع نمونه‌های مثبت که به درستی طبقه‌بندی شده‌اند و نمونه‌های مثبت که به اشتباه طبقه‌بندی شده‌اند.
$Recall = \frac{TN}{TN+FN}$	نسبت نمونه‌های مثبت که به درستی طبقه‌بندی شده‌اند تقسیم بر جمع نمونه‌های مثبت که به درستی طبقه‌بندی شده‌اند و نمونه‌های منفی که به اشتباه طبقه‌بندی شده‌اند.

تفکیک ساده تفکیک ساده^۱ (یا تخمین نمونه تست^۲ یا نمونه‌ی جدا نگهداشته شده^۳)، داده‌ها را به دو زیرمجموعه کاملاً ناسازگار یا جدا^۴ به نام‌های مجموعه آموزش^۵ و مجموعه تست^۶ (یا مجموعه‌ی داده‌های جدا نگهداشته شده) تقسیم می‌کند. معمولاً دو سوم داده‌ها به عنوان مجموعه آموزشی و یک سوم باقیمانده به عنوان داده‌های تستی در نظر گرفته می‌شوند. الفاکر (سازنده‌ی مدل) از مجموعه آموزشی استفاده کرده و سپس طبقه‌بند را بر روی مجموعه تستی می‌آزمایند. یک استثنا برای این قاعده زمانی است که طبقه‌بند، یک شبکه عصبی مصنوعی باشد. در این حالت، داده‌ها به سه زیرمجموعه‌ی کاملاً ناسازگار و جدا تقسیم می‌شوند: آموزشی، اعتبارسنجی و تستی. مجموعه‌ی اعتبارسنجی در طی فعالیت مدل‌سازی و برای پیشگیری از بیش‌برازش^۷ بکار می‌رود. شکل ۴.۹ یک متدولوژی تفکیک ساده را نشان می‌دهد.

انتقاد اصلی که به این روش وارد است این است که فرض می‌کند داده‌های موجود در دو زیرمجموعه از یک نوع هستند (یعنی ویژگی‌های کاملاً یکسانی دارند). از آنجائیکه این یک روش تقسیم‌بندی تصادفی ساده است، در مجموعه داده‌های واقع‌بینانه‌تر که داده‌ها بر روی متغیر طبقه‌بندی دچار چولگی می‌شوند این فرض ممکن است صادق نباشد.

¹ Simple Split

² Test sample estimation

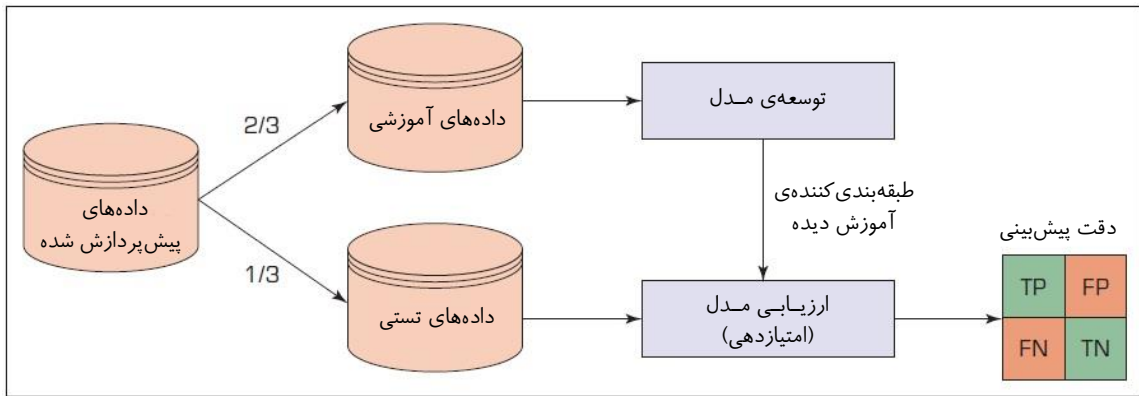
³ Holdout

⁴ Mutually exclusive

⁵ Training set

⁶ Test set

⁷ Overfitting



شکل ۴.۹ تفکیک تصادفی ساده‌ی داده‌ها

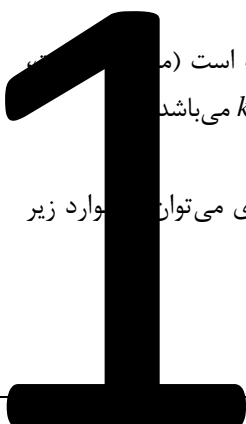
برای بهبود این وضعیت، توصیه می‌شود از روش نمونه‌برداری طبقه‌بندی شده استفاده شود که طبقات تبدیل به متغیر خروجی می‌شوند. اگرچه این روش نسبت به تفکیک ساده یک بهبود محسوب می‌شود ولی همچنان دارای انحرافات ناشی از تفکیک تصادفی ساده است.

اعتبارسنجی متقاطع k مرتبه‌ای برای به حداقل رساندن انحراف ناشی از نمونه‌برداری تصادفی نمونه داده‌های آموزشی و تستی برای مقایسه‌ی دقت دو یا چند روش، می‌توان از یک متدولوژی به نام اعتبارسنجی متقاطع k مرتبه‌ای استفاده کرد. در این روش که تخمین چرخشی^۱ نیز نامیده می‌شود، کل مجموعه داده به شکل تصادفی به k زیرمجموعه‌ی کاملاً ناسازگار و جدا با اندازه تقریباً یکسان تقسیم می‌شود. سپس مدل، k مرتبه آموزش داده شده و تست می‌شود. این روش هر بار برای آموزش تمامی داده‌ها و یکبار انجام شده و سپس یکبار برای روی مابقی داده‌ها تست می‌شود. تخمین اعتبارسنجی متقاطع دقت کلی یک مدل را طبق معادله‌ی زیر به سادگی می‌توان با محاسبه‌ی میانگین مقادیر k سنجه‌ی دقت اندازه‌گیری شده بدست آورد.

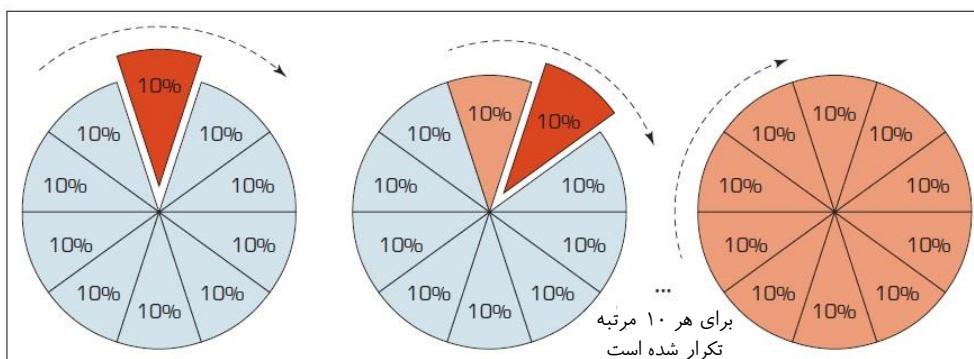
$$CVA = \frac{1}{k} \sum_{i=1}^k A_i$$

که CVA مخفف دقت اعتبارسنجی متقاطع، k تعداد دفعات بکار رفته و A سنجه‌ی دقت در هر مرتبه است (میانگین سنجه‌ی دقت در هر مرتبه). شکل ۴.۱۰ یک نمایش گرافیکی از اعتبارسنجی متقاطع k مرتبه‌ای را نشان می‌دهد که $k = 10$ می‌باشد.

متدولوژی‌های دیگر برای ارزیابی طبقه‌بندی از دیگر متدولوژی‌های متداول برای ارزیابی طبقه‌بندی می‌توان به موارد زیر اشاره کرد:



- **اعتبارسنجی یک طرفه^۱**. روش اعتبارسنجی یک طرفه مشابه اعتبارسنجی متقاطع k مرتبه‌ای است که مقدار k در آن ۱ در نظر گرفته می‌شود. این بدان معناست که هر نقطه داده فقط یکبار برای تست هر تعداد مدل که ایجاد شده است بکار می‌رود. این متدولوژی زمان بر است ولی برخی اوقات برای مجموعه داده‌های کوچک گزینه مناسبی است.
- **بوت‌استرپینگ^۲**. در روش بوت‌استرپینگ تعداد ثابتی از نمونه‌ها از داده‌های اصلی برای آموزش نمونه‌برداری شده و مابقی مجموعه داده‌ها برای تست بکار می‌روند. این فرایند را به تعداد دفعات دلخواه می‌توان تکرار کرد.
- **جک‌نایفینگ^۳**. این متدولوژی اگرچه مشابه اعتبارسنجی یک طرفه است ولی برای محاسبه‌ی دقت مدل، در هر مرتبه تکرار فرایند تخمین، یک نمونه کنار گذاشته می‌شود.
- **مساحت زیر منحنی مشخصه‌ی عملکرد سیستم^۴**. مساحت زیر منحنی مشخصه‌ی عملکرد سیستم یک تکنیک ارزیابی گرافیکی است که مقادیر نرخ مثبت صحیح را بر روی محور عمودی (Y) و مقادیر نرخ مثبت کاذب را بر روی محور افقی (X) ترسیم می‌کند. مساحت زیر منحنی آر‌سی تعیین‌کننده‌ی دقت یک طبقه‌بند است: مقدار ۱ بیانگر یک طبقه‌بند بی‌عیب و نقص و مقدار ۰.۵ نشان می‌دهد که چیزی در حد یک شانس و تصادف است. در واقعیت، مقادیر بین این دو قرار می‌گیرند. به عنوان مثال، در شکل ۴.۱۱، دارای عملکرد طبقه‌بندی بهتری نسبت به B است در حالیکه C یک شانس تصادفی در حد شیر یا خط انداختن است.



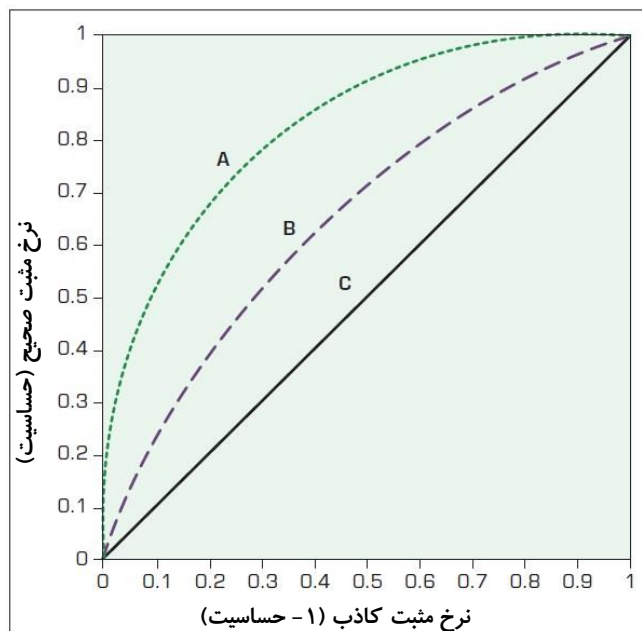
شکل ۴.۱۰ نمایش گرافیکی اعتبارسنجی متقاطع k مرتبه‌ای

¹ Leave-one-out

² Bootstrapping

³ Jackknifing

⁴ Area under ROC Curve



شکل ۴.۱۱. یک نمونه منحنی مشخصه‌ی عملکرد سیستم

روش‌های داده‌کاوی (یعنی الگوریتم‌های یادگیری ماشین) در یافتن ارتباطات پیچیده بین متغیرهای ورودی و خروجی بسیار خوب عمل می‌کنند (مدل‌های پیش‌بینی بسیار دقیقی را ارائه می‌کنند) ولی در توضیح عملکرد و نحوه انجام آنها چندان خوب نیستند (یعنی شفافیت مدل). به منظور پیشگیری از این مشکل (که به آن سندروم جعبه سیاه نیز گفته می‌شود)، جامعه‌ی یادگیری ماشین، چندین روش را پیشنهاد داده‌اند که اکثر آنها در مجموعه‌ی **تحلیل حساسیت** قرار می‌گیرند. در مقوله‌ی مدل‌سازی پیشگویانه، تحلیل حساسیت به یک فرایند آزمایش انحصاری اشاره دارد که هدف آن کشف روابط علت و معلولی بین متغیرهای ورودی و متغیر خروجی است. برخی از روش‌های اهمیت متغیر، خاص هر الگوریتم هستند (یعنی بر روی درخت تصمیم بکار گرفته می‌شوند) و برخی دیگر نیز هیچ ارتباطی با الگوریتم ندارند. در اینجا به معرفی پرکاربردترین روش‌های اهمیت متغیر در مدل‌سازی پیشگویانه و یادگیری ماشین می‌پردازیم:

۱. توسعه و مشاهده‌ی یک مدل درخت تصمیم که به خوبی آموزش دیده برای دیدن قابلیت تشخیص نسبی متغیرهای ورودی - هرچه متغیر بکار رفته برای تفکیک داده‌ها به ریشه‌ی درخت نزدیک‌تر باشد، اهمیت و سهم نسبی آن در مدل پیش‌بینی بیشتر خواهد بود.
۲. توسعه و مشاهده‌ی یک مدل جنگلی تصادفی بزرگ و غنی و ارزیابی آمارهای مربوط به تفکیک متغیر. اگر نسبت انتخاب یک متغیر مفروض به تعداد متغیرهای کاندید بزرگ باشد (یعنی تعداد دفعاتی که یک متغیر در سطح صفر انتخاب می‌شود تقسیم بر تعداد دفعاتی که به طور تصادفی به عنوان یک کاندید تفکیک انتخاب شده است)، آنگاه اهمیت/سهم نسبی آن نیز بزرگتر خواهد بود.

۳. تحلیل حساسیت بر اساس انحراف یا آشفتگی مقدار ورودی که بر اساس آن، متغیرهای ورودی یک به یک به تدریج تغییر کرده/بهم ریخته و تغییر نسبی متغیر خروجی مورد پایش قرار می‌گیرد یعنی هرچه تغییر متغیر خروجی بیشتر باشد، میزان اهمیت متغیر بهم‌ریخته نیز بیشتر خواهد بود. از این روش اغلب در مدل‌سازی شبکه عصبی پیشخور و زمانی استفاده می‌شود که تمامی متغیرها از نوع عددی بوده و استاندارد/نرمال شده باشند. این روش در فصل ۶ در حوزه‌ی یادگیری عمیق و شبکه‌های عصبی عمیق به طور کامل توضیح داده شده است.
۴. تحلیل حساسیت بر اساس متدولوژی اعتبارسنجی یک طرفه. از این روش می‌توان برای هر نوع روش تحلیل پیشگویانه استفاده کرد و از این روش شرح مفصل آن به شکل زیر است:

تحلیل حساسیت (بر اساس متدولوژی اعتبارسنجی یک طرفه) متکی بر فرایند آزمایشی حذف سیستماتیک و یک به یک متغیرهای ورودی از درون مجموعه متغیر ورودی، توسعه‌ی و تست یک مدل و مشاهده‌ی تأثیر عدم حضور این متغیر بر عملکرد پیشگویانه‌ی مدل یادگیری ماشین است. سپس این مدل برای اندازه‌گیری میزان اهمیت/سهم، برای هر متغیر ورودی، آموزش دیده و تست می‌شود (اغلب با استفاده از یک روش اعتبارسنجی متقاطع k مرتبه‌ای). یک نمایش گرافیکی از این فرایند در شکل ۴.۱۲ نشان داده شده است.

این روش اغلب برای ماشین‌های برداری پشتیبان، درخت تصمیم، رگرسیون لجستیک و شبکه‌های عصبی مصنوعی بکار می‌رود. سالتلی^۱ (۲۰۰۲) در کتاب تحلیل حساسیت خود فرمول جبری این فرایند اندازه‌گیری را به شکل زیر ارائه کرده است:

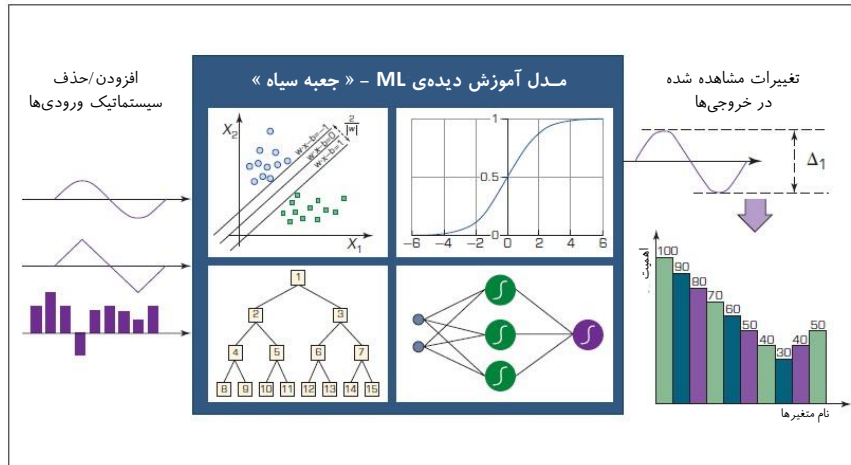
$$S_i = \frac{V_i}{V(F_t)} = \frac{V(E(F_t|X_i))}{V(F_t)}$$

در مخرج کسر، $V(F_t)$ به واریانس متغیر خروجی اشاره دارد. در صورت کسر، $V(E(F_t|X_i))$ که عملگر انتظار فراخوانی برای یک انتگرال بر روی پارامتر X_i است یعنی تمامی متغیرهای ورودی به استثنای X_i و V عبارتست از عملگر واریانس که یک انتگرال اضافی را بر روی X_i انجام می‌دهد. اهمیت متغیر که با S_i برای i امین متغیر نمایش داده می‌شود به صورت سنج‌های حساسیت نرمال شده محاسبه می‌شود. سالتلی و همکاران (۲۰۰۴) در مطالعه‌ای جدیدتر ثابت کردند که این معادله، محتمل‌ترین سنج برای حساسیت مدل است که قادر به رتبه‌بندی متغیرهای ورودی (پیش‌گوها) به ترتیب اهمیت هر ترکیبی از تکرارها از جمله ارتباطات غیرمتعارف بین متغیرهای ورودی می‌باشد. برای اینکه بتوانیم نتایج تحلیل حساسیت بدست آمده از چندین روش پیش‌بینی را به شکلی مناسب ترکیب کنیم، می‌توانیم از یک متدولوژی تلفیق اطلاعات استفاده کنیم و معادله قبل را به شکلی تغییر دهیم که سنج‌های حساسیت متغیر ورودی n بر اساس اطلاعات ترکیب شده (تلفیق شده) از m مدل پیش‌بینی بدست آید. معادله‌ی زیر، این تابع جمع‌بندی موزون را نشان می‌دهد:

$$S_n(\text{fused}) = \sum_{i=1}^m \omega_i S_{in} = \omega_1 S_{1n} + \omega_2 S_{2n} + \dots + \omega_m S_{mn}$$

در این معادله، ω_i بیانگر وزن/سهم نرمال شده برای هر مدل پیش‌گویی است که در آن، سطح وزن/سهم یک مدل به صورت تابعی از توان پیشگویانه نسبی آن محاسبه می‌شود - هرچه توان پیشگویایی بیشتر باشد (یعنی دقت)، مقدار ω نیز بزرگتر خواهد بود.

¹ Salteli



شکل ۴.۱۲ نمایش گرافیکی فرایند تحلیل حساسیت

تکنیک‌های طبقه‌بندی از تکنیک‌ها یا الگوریتم‌های متعددی برای مدل‌سازی طبقه‌بندی استفاده می‌شود:

- **تحلیل درخت تصمیم^۱**: تحلیل درخت تصمیم (یکی از تکنیک‌های یادگیری ماشین) را شاید بتوان محبوب‌ترین تکنیک طبقه‌بندی در حوزه‌ی داده‌کاوی دانست. در بخش بعدی این تکنیک به طور کامل توضیح داده شده است.
- **تحلیل آماری**: تکنیک‌های آماری برای سالیان متمادی و تا زمان پیدایش تکنیک‌های یادگیری ماشین، اصلی‌ترین الگوریتم طبقه‌بندی محسوب می‌شدند. تکنیک‌های طبقه‌بندی یا دسته‌بندی آماری شامل رگرسیون لجستیک و تحلیل تشخیصی است که در هر دو تکنیک، روابط بین متغیرهای ورودی و خروجی به صورت خطی فرض می‌شوند، داده‌ها دارای توزیع نرمال هستند و متغیرها فاقد همبستگی بوده و مستقل از یکدیگر هستند. ماهیت سوال برانگیز این مفروضات موجب شد تا تکنیک‌های یادگیری ماشین جایگزین آنها شوند.
- **شبکه‌های عصبی**: اینها محبوب‌ترین و متداول‌ترین تکنیک‌های یادگیری ماشین هستند که برای مسائل طبقه‌بندی مورد استفاده قرار می‌گیرند.
- **استدلال مبتنی بر مورد^۲**: در این روش از موردهای گذشته برای شناسایی وجوه اشتراک استفاده کرده و یک مورد جدید را به محتمل‌ترین و نزدیک‌ترین گروه تخصیص می‌دهد.
- **دسته‌بندی‌کننده‌های بیز^۳**: این رویکرد از نظریه احتمالات برای ساخت مدل‌های طبقه‌بندی و بر اساس رویدادهای گذشته استفاده می‌کند که قادر است تا نمونه‌های جدید را به محتمل‌ترین گروه یا دسته تخصیص دهد.
- **الگوریتم‌های ژنتیک**: در این روش از تشابه و همسانی تکامل طبیعی برای ساخت مکانیزم‌های جهت‌دار جستجو محور برای طبقه‌بندی و دسته‌بندی نمونه داده‌ها استفاده می‌شود.

¹ Decision Tree Analysis

² Case-based Reasoning

³ Bayesian Classifier

- مجموعه‌های راف. در این روش، عضویت یا تعلق جزئی برچسب دسته یا گروه برای گروه‌های از پیش تعریف شده در ساخت مدل‌های طبقه‌بندی در نظر گرفته می‌شود.

شرح کامل تمامی این تکنیک‌های طبقه‌بندی فراتر از حوصله و دامنه‌ی این کتاب است و در اینجا صرفاً به ارائه محبوب‌ترین آنها پرداخته‌ایم.

درخت تصمیم پیش از آنکه به تشریح جزئیات درخت تصمیم بپردازیم، باید در مورد برخی واژگان و اصطلاحات ساده صحبت کنیم. اول اینکه درخت تصمیم شامل متغیرهای ورودی بسیاری است که ممکن است بر طبقه‌بندی الگوهای مختلف تأثیر بگذارند. این متغیرهای ورودی را معمولاً **خصیصه**^۱ می‌نامند. به عنوان مثال، اگر قرار باشد مدلی را برای طبقه‌بندی ریسک وام و بر اساس دو مشخصه‌ی درآمد و امتیاز اعتبار ایجاد کنیم، این دو مشخصه را **خصیصه** و متغیر خروجی منتج شده را **برچسب دسته یا گروه** (مثلاً **ریسک کم، متوسط یا زیاد**) می‌نامیم. دوم، یک درخت شامل شاخه و گره است. یک **شاخه**^۲ بیانگر نتیجه‌ی یک آزمون برای طبقه‌بندی یک الگو با استفاده از یکی از صفات است. نهایتاً، یک **گره**^۳ بیانگر دسته یا گروه انتخابی نهایی برای یک الگو است (زنجیره‌ای از شاخه‌ها از ریشه تا برگ که به صورت گزاره‌های **if-then** می‌توان آنها را بیان کرد).

ایده‌ی اصلی درخت تصمیم این است که به طور برگشتی، یک مجموعه آموزشی را تقسیم می‌کند تا زمانی که هر بخش به طور کامل از مثال‌های یک دسته یا گروه تشکیل شده باشد. هر گره‌ی غیربرگ این درخت حاوی یک **نقطه‌ی تفکیک**^۴ است که نقطه‌ی تست یک یا چند خصیصه بوده و تعیین می‌کند که داده‌ها چگونه باید تقسیم شوند. الگوریتم‌های درخت تصمیم به طور کلی یک درخت اولیه را از روی داده‌های آموزشی می‌سازند به نحوی که هر گره‌ی برگ خالص باشد و سپس درخت را هرس می‌کنند تا قدرت تعمیم و دقت پیش‌بینی داده‌های تستی افزایش یابد.

در فاز رشد، این درخت از طریق تقسیم برگشتی تا زمانی که هر بخش آن خالص (یعنی حاوی اعضای از دسته یا گروه یکسان باشد) یا نسبتاً کوچک شود ساخته می‌شود. ایده‌ی اصلی آن است که سوالاتی پرسیده شود که پاسخ آنها مشابه وقتی « بیست سوالی » بازی می‌کردیم بیشترین اطلاعات را به ما بدهد.

روش تفکیک بکار رفته برای تقسیم داده‌ها به نوع خصیصه‌ی بکار رفته بستگی دارد. برای یک خصیصه‌ی پیوسته‌ی A ، تفکیک به فرم $\text{Value}(A) < x$ می‌باشد که x یک مقدار تفکیک « بهینه » از A است. به عنوان مثال، تفکیک بر اساس خصیصه‌ی درآمد می‌تواند به شکل « درآمد $> 50,000$ » باشد. در خصوص خصیصه‌ی رسته‌ای A ، تفکیک‌ها به فرم $\text{Value}(A) \in x$ که x زیرمجموعه‌ای از A است. به عنوان مثال، یک تفکیک می‌تواند بر اساس جنسیت باشد: « مذکر – مؤنث ». یک الگوریتم عمومی برای ساخت یک درخت تصمیم به شرح زیر است:

۱. یک گره‌ی ریشه ایجاد کرده و تمامی داده‌های آموزشی را به آن اختصاص بدهید.
۲. بهترین خصیصه‌ی تفکیک را انتخاب کنید.
۳. یک شاخه را به ازای هر مقدار تفکیک به گره‌ی ریشه بیفزایید. داده‌ها را در زیرمجموعه‌های کاملاً جدا در امتداد خطوط تفکیک تقسیم کرده و به سمت شاخه‌ها حرکت کنید.

¹ Attribute
² Class label
³ Branch
⁴ Leaf node
⁵ Split point

۴. گام‌های ۲ و ۳ را برای هر گره‌ی برگ تکرار کنید تا زمانیکه به معیار توقف برسید (یعنی فقط یک برچسب دسته یا گروه در گره باقی بماند).

الگوریتم‌های بسیار متفاوتی برای ایجاد درخت تصمیم پیشنهاد شده است. این الگوریتم‌ها عمدتاً از لحاظ روش تعیین خصیصه‌ی تفکیک (و مقدار آن)، ترتیب تفکیک خصیصه‌ها (تفکیک یک خصیصه فقط یکبار یا چندین مرتبه)، تعداد تفکیک‌ها در هر گروه (دودویی و سه‌تایی)، معیارهای توقف و خالص‌سازی درخت با یکدیگر تفاوت دارند. برخی از معروف‌ترین الگوریتم‌ها شامل ID3 (نسخه‌های بهبود یافته‌ی C4.5 و C5) از حوزه‌ی یادگیری ماشین، درخت‌های رگرسیون و رده‌بندی^۱ از حوزه‌ی آمار و شناسایی اثرات متقابل خودکار کای^۲ از حوزه‌ی تشخیص الگو می‌باشند.

در زمان ساخت یک درخت تصمیم، هدف در هر گره این است که خصیصه و نقطه‌ی تفکیک آن خصیصه را که رکوردهای آموزشی را بهترین شکل برای خالص‌سازی نمایش دسته یا گروه در هر گره تقسیم کند تعیین کنیم. برای ارزیابی برآزش تفکیک، چندین شاخص تفکیک پیشنهاد شده است. دو شاخص بسیار مهم عبارتند از شاخص جینی^۳ و بهره‌ی اطلاعات^۴. از شاخص جینی در الگوریتم‌های درخت‌های رگرسیون و رده‌بندی و اسپرینت^۵ استفاده می‌شود. نسخه‌هایی از شاخص بهره‌ی اطلاعاتی در ID3 (و نسخه‌های جدیدتر C4.5 و C5) استفاده می‌شود.

شاخص جینی در علم اقتصاد برای اندازه‌گیری گوناگونی جمعیت بکار رفته است. همین مفهوم را می‌توان برای تعیین میزان خلوص یک دسته یا گروه خاص که نتیجه‌ی تصمیم برای یک شاخه در امتداد یک خصیصه یا متغیر خاص بکار برد. بهترین تفکیک زمانی اتفاق می‌افتد که خلوص حاصل از تفکیک پیشنهادی را افزایش دهد. در اینجا نگاه مختصری به محاسبه‌ی ساده‌ی شاخص جینی می‌اندازیم.

اگر مجموعه داده‌ی S حاوی مثال‌هایی از n دسته یا گروه باشد، شاخص جینی به شکل زیر تعریف می‌شود:

$$gini(S) = 1 - \sum_{j=1}^n p_j^2$$

که p_j تواتر نسبی دسته یا گروه j در مجموعه داده‌های S است. اگر مجموعه داده‌ی S را به دو زیرمجموعه‌ی S_1 و S_2 با اندازه‌های N_1 و N_2 تفکیک کنیم، شاخص جینی تفکیک داده‌ها حاوی مثال‌هایی از n دسته یا گروه خواهد بود و شاخص جینی به شکل زیر تعریف می‌شود:

$$gini_{split}(S) = \frac{N_1}{N} gini(S_1) + \frac{N_2}{N} gini(S_2)$$

ترکیب خصیصه/تفکیک که کوچکترین مقدار را برای $gini_{split}(S)$ نتیجه دهد به عنوان گره‌ی تفکیک انتخاب می‌شود. برای این منظور باید تمامی نقاط تفکیک امکان‌پذیر را برای هر خصیصه برشمرد.

¹ Classification and Regression Trees (CART)

² Chi-squared automatic interaction detector (CHAID)

³ Gini Index

⁴ Information Gain

⁵ Scalable PaRallelizable INduction of Decision Trees

بهره‌ی اطلاعاتی یک مکانیزم تفکیک است که در ID3 که شاید شناخته‌شده‌ترین الگوریتم درخت تصمیم باشد. این الگوریتم توسط راس کوئینلان^۱ در سال ۱۹۸۶ ابداع شد و پس از آن الگوریتم‌های C4.5 و C5 تکامل یافت. ایده‌ی اصلی ID3 (و نسخه‌های دیگر آن) استفاده از مفهومی به نام **آنترپوی**^۲ به جای شاخص جینی است. آنترپوی میزان عدم قطعیت یا تصادفی بودن مجموعه داده‌ها را اندازه‌گیری می‌کند. اگر تمامی داده‌ها در یک زیرمجموعه فقط به یک دسته یا گروه تعلق داشته باشند، عدم قطعیت یا میزان تصادفی بودن در آن مجموعه داده وجود ندارد و برابر صفر است. هدف از این روش این است که یک درخت زیرمجموعه ساخته شود که آنترپوی هر زیرمجموعه نهایی معادل صفر (یا نزدیک به صفر) باشد. حال نگاهی بیندازیم به محاسبه‌ی بهره‌ی اطلاعاتی.

فرض کنید دو دسته یا گروه داشته باشیم: P (مثبت) و N (منفی). فرض کنید مجموعه مثال‌های S حاوی p شمارش از دسته‌ی P و n شمارش از دسته‌ی N باشد. حجم اطلاعات مورد نیاز برای تصمیم‌گیری در خصوص اینکه یک مثال تصادفی از S به P تعلق دارد یا به N به شکل زیر تعریف می‌شود:

$$I(p, n) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

فرض کنید با استفاده از خصیصه‌ی A ، مجموعه‌ی S را به زیرمجموعه‌های $\{S_1, S_2, \dots, g\}$ تقسیم کنیم. اگر S_i حاوی p_i مثال از مجموعه‌ی P و n_i مثال از مجموعه‌ی N باشد آنگاه آنترپوی یا اطلاعات مورد انتظار که برای رده‌بندی اشیاء در تمامی زیرمجموعه‌های S_i نیاز داریم:

$$E(A) = \sum_{i=1}^n \frac{p_i + n_i}{p+n} I(p_i, n_i)$$

سپس بهره‌ی اطلاعاتی بدست آمده از طریق انشعاب خصیصه‌ی A به شکل زیر محاسبه می‌شود:

$$Gain(A) = I(p, n) - E(A)$$

این محاسبات برای هر خصیصه تکرار می‌شود و آن خصیصه‌ای که بالاترین بهره‌ی اطلاعاتی را نتیجه دهد به عنوان خصیصه‌ی تفکیک انتخاب می‌شود. ایده‌ی اصلی در پشت این شاخص‌های تفکیک تقریباً یکسان است ولی برخی از جزئیات مربوط به الگوریتم‌ها با یکدیگر تفاوت دارد. یک تعریف دقیق از الگوریتم ID3 و مکانیزم تفکیک آن در کوئینلان (۱۹۸۶) موجود است. مورد کاربردی ۴.۵ نشان می‌دهد در صورتیکه از تکنیک‌های داده‌کاوی درست برای یک مسئله‌ی کسب‌وکار استفاده شود، بهره اطلاعاتی به چه شکل خواهد بود.

¹ Ross Quinlan

² Entropy

شرکت اینفلوئنس هِلث^۱ تنها ارائه‌دهنده‌ی پلتفرم فعال‌سازی و مشارکت دیجیتالی یکپارچه‌ی مشتریان در صنعت مراقب سلامت است. این شرکت از طریق یک مداخله‌ی چندکاناله تعاملی و شخصی‌سازی شده به ارائه‌دهندگان، کارفرمایان و پرداخت‌کنندگان این امکان را می‌دهد تا تأثیر مثبتی بر تصمیم‌گیری مشتریان و رفتارهای سلامت آنان بگذارند. از سال ۱۹۹۶، این شرکت که در شهر بیرمنگهام ایالت آلاباما واقع شده است به بیش از ۱،۱۰۰ سازمان ارائه‌دهنده‌ی خدمات مراقبت سلامت کمک کرده است تا با تغییر نتایج کیفی و مالی بر مشتریان خود تأثیر بگذارند.

مراقبت سلامت یک کسب‌وکار شخصی است. نیازهای هر بیمار، متفاوت بوده و به یک پاسخ خاص نیاز دارد. از طرف دیگر، - با توجه به افزایش هزینه‌ی ارائه خدمات مراقبت سلامت - بیمارستان‌ها و سیستم‌های سلامت به طور فزاینده‌ای باید با دستیابی به جمعیت بزرگتری از بیماران، به صرفه‌جویی در مقیاس دست پیدا کنند. چالش آنها این است که بتوانند روش شخصی‌سازی شده‌ای را ارائه دهند و در عین حال در یک مقیاس بزرگ فعالیت کنند. اینفلوئنس هِلث در کمک به مشتریان بخش مراقبت سلامت خود برای حل این مشکل متخصص بوده و به آنها کمک می‌کند تا بیماران فعلی و بالقوه خود را بهتر شناخته و هر یک از آنها را در زمان مناسب با خدمات سلامت مناسب هدف قرار دهند. فناوری تحلیل پیشگویانه‌ی پیشرفته شرکت آی‌بی‌ام به اینفلوئنس هِلث اجازه می‌دهد تا مشتریان را در کشف عواملی که بیشترین تأثیر را بر تصمیمات مراقبت سلامت بیماران می‌گذارد کشف کنند. با ارزیابی صدها میلیون چشم انتظار که نیازمند دریافت خدمات مراقبت سلامت خاص هستند، اینفلوئنس هِلث می‌تواند بر درآمدها و نرخ پاسخگویی کمپین‌های مراقبت سلامت خود افزوده و نتایج آنها را برای مشتریان و بیماران آنها بهبود بخشد.

هدف قراردادن مشتریان زرنگ

صنعت امروزی مراقبت سلامت بیش از هر زمان دیگر در حال رقابتی شدن است. اگر استفاده از خدمات این سازمان‌ها دچار افت شود، سود آنها نیز کاهش می‌یابد. امروزه

Sources: H. M. Zolbanin, D. Delen, & A. H. Zadeh, "Predicting Overall Survivability in Comorbidity of Cancers: A Data Mining Approach," *Decision Support Systems*, 74, 2015, pp. 150–161; D. Delen, "Analysis of Cancer Data: A Data Mining Approach," *Expert Systems*, 26(1), 2009, pp. 100–112; J. Thongkam, G. Xu, Y. Zhang, & F. Huang, "Toward Breast Cancer Survivability Prediction Models Through Improving Training Space," *Expert Systems with Applications*, 36(10), 2009, pp. 12200–12209; D. Delen, G. Walker, & A. Kadam, "Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods," *Artificial Intelligence in Medicine*, 34(2), 2005, pp. 113–127.

سوالاتی برای مورد کاربردی ۴.۵

۱. از داده‌کاوی چگونه می‌توان برای درمان بیماری‌هایی مثل سرطان استفاده کرد؟
۲. به نظر شما وعده‌ها و چالش‌های اساسی داده‌کاوها در پژوهش‌های پزشکی و بیولوژیکی چیست؟

¹ Influence Health Inc.

تحلیل خوشه یک روش اصلی داده کاوی برای دسته‌بندی آیتم‌ها، رویدادها یا مفاهیم در قالب گروه‌بندی‌های مشترک به نام خوشه‌ها است. این روش کاربرد گسترده‌ای در زیست‌شناسی، پزشکی، ژنتیک، تحلیل شبکه‌های اجتماعی، انسان‌شناسی، باستان‌شناسی، نجوم، تشخیص شخصیت و حتی در توسعه سیستم‌های اطلاعات مدیریت دارد. با افزایش محبوبیت داده کاوی، از تکنیک‌های زیربنایی آن در کسب‌وکار و به ویژه در بازاریابی استفاده شده است. از تحلیل خوشه به طور گسترده‌ای برای تشخیص موارد مشکوک به کلاهبرداری (هم کارت‌های اعتباری و هم تجارت الکترونیک) و بخش‌بندی مشتریان بازار در سیستم‌های مدیریت ارتباط با مشتریان معاصر استفاده شده است. با توجه به قدرت بالای تحلیل خوشه، کاربردهای بیشتر این تحلیل در کسب‌وکار ادامه دارد. تحلیل خوشه یک ابزار توضیحی تحلیل داده برای حل مسائل دسته‌بندی و رده‌بندی است. هدف از این تحلیل، مرتب‌سازی و قراردادن موارد (مثلاً افراد، اشیاء، رویدادها) در قالب گروه‌ها یا خوشه‌ها است به نحوی که اعضای داخل هر خوشه دارای قویترین میزان وابستگی و اعضای بین خوشه‌ها دارای ضعیف‌ترین وابستگی باشند. یک مثال تک بُعدی واضح از تحلیل خوشه، تعیین دامنه‌های امتیازی است که نمرات کلاس در یک کالج بر اساس آن تعیین می‌شود. این مشابه مسئله تحلیل خوشه‌ای است که خزانه‌داری ایالات متحده در زمان استقرار براکت‌های مالیاتی جدید در دهه ۸۰ با آن مواجه شد. یک مثال فرضی از خوشه‌بندی نیز کتاب‌های هری پاتر نوشته‌ی جی کی رولینگز آورده شده است. کلاه گروه‌بندی تعیین می‌کند که دانش‌آموزان سال اولی به کدام خانه (خوابگاه) در مدرسه‌ی هوگ‌وارتز تخصیص پیدا کنند. یک مثال دیگر شامل تعیین جایگاه مهمانان در یک مراسم عروسی است. اهمیت تحلیل خوشه در این است که می‌تواند وابستگی‌ها و ساختارهای موجود در داده‌ها را که تا پیش از این قابل تشخیص نبوده‌اند ولی به محض آشکار شدن، حساسیت و مفید بودن آنها مشخص شده است را آشکار کند.

از نتایج تحلیل خوشه می‌توان برای موارد زیر استفاده کرد:

- شناسایی یک طرح یا الگوی دسته‌بندی و رده‌بندی (مثلاً نوع مشتریان)
- توصیه‌ی مدل‌های آماری برای توصیف جمعیت‌ها
- نشان دادن قواعد برای تخصیص موارد جدید به دسته یا گروه‌ها برای شناسایی، هدف قراردادن و تشخیص و عیب‌یابی
- ارائه سنجه‌هایی برای تعریف، اندازه و تغییر مفاهیم کلی و گسترده‌ی قبلی
- یافتن موارد معمول برای برجسب‌گذاری و ارائه‌ی دسته‌ها و گروه‌ها
- کاهش اندازه و پیچیدگی فضای مسئله برای سایر روش‌های داده کاوی
- شناسایی مقادیر پرت در یک دامنه‌ی شخاص (مثلاً تشخیص رویدادهایی که به ندرت اتفاق می‌افتند).

تعیین تعداد بهینه خوشه‌ها در الگوریتم‌های خوشه‌بندی معمولاً لازم است که تعداد خوشه‌های مورد نظر مشخص شود. اگر این عدد از قبل معلوم نباشد، باید به طریقی تعداد آنها را مشخص کرد. متأسفانه، هیچ روش بهینه‌ای برای محاسبه‌ی تعداد مورد انتظار خوشه‌ها وجود ندارد. از این رو، روش‌های هیوریستیک متعددی پیشنهاد شده که متداول‌ترین آنها در ادامه ذکر شده است:

- درصد واریانس توضیح داده شده را به عنوان تابعی از تعداد خوشه‌ها در نظر بگیرید یعنی عددی را برای تعداد خوشه‌ها انتخاب کنید که در صورت افزودن خوشه‌ی دیگری به آن، به مدل بهتری از داده‌ها دست پیدا نکنیم. به بیان خاص، اگر نمودار درصد واریانس توضیح داده شده توسط خوشه‌ها را ترسیم کنیم، نقطه‌ای وجود دارد که در آن بهره‌ی حاشیه‌ی افت می‌کند که نشان دهنده‌ی این است که آن نقطه، تعداد بهینه‌ی خوشه‌ها را نشان می‌دهد.

- تعداد خوشه‌ها را معادل $(n/2)^{1/2}$ در نظر بگیرید که n تعداد نقاط داده‌ها است.
- از معیار اطلاعاتی آکائیکه^۱ استفاده کنید. این معیار یک سنجی نکویی برازش (بر اساس مفهوم آنتروپی) است که برای تعیین تعداد خوشه‌ها بکار می‌رود.

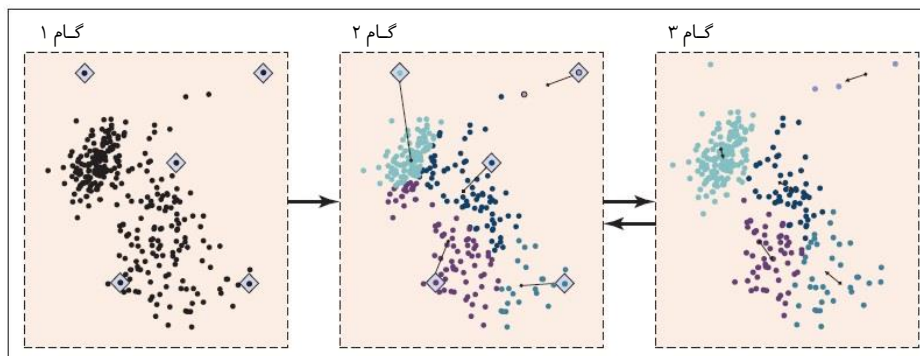
روش‌های تحلیل خوشه می‌تواند بر اساس یک یا چند روش عمومی زیر انجام شود:

- روش‌های آماری (از جمله روش‌های سلسله‌مراتبی و غیرسلسله‌مراتبی) مثل میانگین‌کی یا مد‌کی.
- شبکه‌های عصبی (با معماری معروف به نگاشت‌های خودسازمانده‌ی کننده).
- منطق فازی (مثلاً الگوریتم میانگین سی فازی).
- الگوریتم ژنتیک.

هرکدام از این روش‌ها با یک یا چند دسته یا گروه عمومی کار می‌کند:

- **تقسیم‌کننده.** در خصوص دسته‌های تقسیم‌کننده، تمامی آیت‌ها از داخل یک خوشه آغاز شده و تجزیه می‌شوند.
- **تجمع‌کننده.** در خصوص دسته‌های تجمع‌کننده، تمامی آیت‌ها از خوشه‌های انفرادی آغاز شده و خوشه‌ها با یکدیگر ترکیب می‌شوند.

اکثر روش‌های تحلیل خوشه شامل بکارگیری یک **سنجی فاصله^۲** برای محاسبه‌ی نزدیکی بین زوج آیت‌ها می‌باشد. متداول‌ترین سنجی‌های فاصله شامل فاصله‌ی اقلیدسی (فاصله‌ی بین دو نقطه که توسط یک خط‌کش اندازه‌گیری می‌شود) و فاصله‌ی منهن (یا **فاصله‌ی مستقیم/خط** یا **فاصله‌ی تاکسی**) بین دو نقطه است. اغلب، این سنجی‌ها مبتنی بر فواصل حقیقی هستند که اندازه‌گیری می‌شوند ولی لزومی ندارد که اینگونه باشد مثل مورد توسعه‌ی سیستم‌های اطلاعاتی. از میانگین‌های موزون می‌توان برای برقراری این فواصل استفاده کرد. به عنوان مثال، در یک پروژه‌ی توسعه‌ی سیستم اطلاعاتی، تک تک ماژول‌های سیستم را می‌توان بر اساس تشابه بین ورودی‌ها، خروجی‌ها، فرایندها و داده‌های خاص بکار رفته به یکدیگر مرتبط کرد. سپس این عوامل در قالب یک سنجی فاصله به صورت زوجی تجمع می‌شوند.



¹ Akaike Information Criterion

² Distance measure

شکل ۴.۱۳ نمایش گرافیکی گام‌های الگوریتم کی میانگین

الگوریتم خوشه‌بندی کی میانگین الگوریتم کی میانگین (کی معرف تعداد خوشه‌های از پیش تعیین شده است) را شاید بتوان پرکاربردترین الگوریتم خوشه‌بندی دانست. این الگوریتم ریشه در تحلیل‌های آماری سنتی دارد. همانطور که از نام آن پیداست، این الگوریتم هر نقطه‌ی داده (مشتری، رویداد، شیء و غیره) را به خوشه‌ای تخصیص می‌دهد که به مرکز نزدیک‌ترین است (مرکز خوشه^۱ یا مرکز ثقل نیز گفته می‌شود). مرکز به صورت میانگین تمامی نقاط داخل خوشه محاسبه می‌شود؛ بدین معنی که مختصات مرکز به صورت میانگین حسابی هر بُعد به صورت جداگانه برای تمامی نقاط داخل خوشه محاسبه می‌شود. گام‌های الگوریتم به صورت گرافیکی در شکل ۴.۱۳ نشان داده شده است:

گام آغازین: تعداد خوشه‌ها را انتخاب کنید (یعنی مقدار K).

گام ۱: نقاط تصادفی k را به عنوان مراکز آغازین خوشه به صورت تصادفی ایجاد کنید.

گام ۲: هر نقطه را به نزدیک‌ترین مرکز خوشه تخصیص دهید.

گام ۳: مراکز جدید خوشه را مجدداً محاسبه کنید.

گام تکرار: گام‌های ۲ و ۳ را تا زمانیکه به معیار همگرایی برسید تکرار کنید (معمولاً تخصیص نقاط به خوشه‌ها به یک ثابت و پایداری می‌رسد).

کاوش قانون وابستگی^۲ (تحلیل وابستگی^۳ یا تحلیل سبد خرید^۴) یک روش متداول داده‌کاوی است که عمدتاً برای توضیح اینکه داده‌کاوی چیست و چه کاری می‌تواند انجام دهد برای مخاطبان علاقمند که اطلاعات تکنولوژیکی کمتری دارند بکار می‌رود. اکثر شما ممکن است در مورد ارتباط معروفی که بین فروش ماء‌الشعیر و پوشک بچه در خواروبارفروشی‌ها کشف شد شنیده باشید. یک سوپرمارکت زنجیره‌ای بزرگ (شاید وال‌مارت شاید هم نه، در خصوص اینکه کدام سوپرمارکت بوده اتفاق نظری وجود ندارد) تحلیلی بر روی عادات خرید مشتریان خود انجام داد و یک همبستگی معنادار آماری را بین خرید ماء‌الشعیر و خرید پوشک بچه کشف کرد. این نظریه مطرح شد که دلیل این کار این بوده است که پدران (احتمالاً مردان جوان) برای خرید پوشک بچه به سوپرمارکت مراجعه می‌کردند (به ویژه پنجشنبه‌ها) و چون نمی‌توانستند مثل قبل به بارهای ورزشی بروند، ماء‌الشعیر را نیز از همانجا خریداری می‌کردند. در نتیجه‌ی این یافته، گفته شد که سوپرمارکت زنجیره‌ای، پوشک‌ها را در کنار ماء‌الشعیرها قرار داد که موجب افزایش فروش هر دو شد.

در واقع، هدف از کاوش قانون وابستگی این است که روابط (وابستگی‌های) جالب بین متغیرها (آیتم‌ها) را در پایگاه‌های داده بزرگ پیدا کنیم. به دلیل کاربرد موفق این روش در مسائل کسب‌وکار خرد، آنها را عمدتاً تحلیل سبد خرید می‌نامند. ایده‌ی اصلی در تحلیل سبد خرید، شناسایی ارتباطات قوی بین محصولات (یا خدمات) مختلف است که معمولاً همراه هم خریداری می‌شوند (در

¹ Centroid

² Association Rule Mining

³ Affinity Analysis

⁴ Market-basket analysis

داخل یک سبد کنار هم قرار می‌گیرند؛ خواه سبد فیزیکی در یک خواروبارفروشی یا یک سبد مجازی در یک وبسایت تجارت الکترونیک). به عنوان مثال، ۶۵ درصد از آنهایی که بیمه کامل خودرو را خریداری کرده‌اند، اقدام به خرید بیمه سلامت نیز کرده‌اند؛ ۸۰ درصد از آنهایی که به صورت آنلاین کتاب خریداری کرده‌اند، موسیقی را نیز به طور آنلاین خریداری می‌کنند؛ ۶۰ درصد از آنهایی که دارای فشار خون بالا هستند و اضافه وزن دارند دارای کلتسرول نیز می‌باشند؛ ۷۰ درصد از مشتریانی که رایانه لپ‌تاپ و نرم‌افزار محافظت در برابر ویروس خریداری می‌کنند، برنامه‌های خدماتی توسعه یافته نیز می‌خرند.

ورودی تحلیل سبد خرید در زمانیکه محصولات و/یا خدمات همراه هم خریداری شوند، داده‌های تراکنش ساده‌ی نقطه‌فروش است که در قالب یک نمونه تراکنش جدول‌بندی می‌شوند (مثل محتوای یک رسید خرید). نتیجه‌ی این تحلیل، اطلاعات ذی‌قیمتی است که می‌توان از آن برای درک بهتر رفتار خرید مشتری و دستیابی به سود حداکثری از معاملات کسب‌وکار استفاده کرد. یک کسب‌وکار به طرق زیر می‌تواند از چنین دانشی بهره‌بردارد: (۱) قراردادان اقلام در کنار یکدیگر برای راحتی بیشتر مشتری در زمان برداشتن همزمان آنها و افزایش حجم فروش، (۲) تبلیغ اقلام به صورت بسته‌ای و (۳) قرار دادن اقلام به دور از یکدیگر تا مشتری مجبور شود برای جستجوی آنها در راهروها قدم بزند و با اینکار ممکن است اقلام دیگری به چشمش خورده و آنها را خریداری کند.

کاربردهای تحلیل سبد خرید شامل بازاریابی متقاطع، فروش جانبی، طراحی فروشگاه، طراحی کاتالوگ، طراحی سایت تجارت الکترونیک، بهینه‌سازی تبلیغات آنلاین، قیمت‌گذاری محصول و پیکربندی تبلیغات/فروش می‌باشند. در واقع، تحلیل سبد خرید به کسب‌وکار کمک می‌کند تا از طریق الگوهای خرید مشتریان به نیازها و اولویت‌های آنها پی ببرند. خارج از دنیای کسب‌وکار، از قوانین وابستگی برای کشف روابط بین نشانه‌ها و بیماری‌ها، تشخیص، ویژگی‌های بیمار و درمان‌ها (که می‌توان از آن در یک سیستم پشتیبان تصمیم پزشکی استفاده کرد) و ژن‌ها و عملکردهای آنها (که در پروژه‌های ژنومیک می‌توان از آن استفاده کرد) استفاده می‌شود. چند حوزه و کاربرد کاوش قانون وابستگی به شرح زیر می‌باشند:

- **تراکنش‌های فروش.** ترکیب محصولات خرده‌فروشی که با هم خریداری شده‌اند را می‌توان برای بهبود جانمایی و جایگزاری محصول در طبقات فروش و قیمت‌گذاری ترفیعی محصولات (نباید برای هر دو محصولی که اغلب با هم خریداری می‌شوند تخفیف یا ترفیع یکسانی در نظر گرفت).
- **تراکنش‌های کارت اعتباری.** اقلام خریداری شده با کارت اعتباری می‌توانند بینش‌هایی را در خصوص سایر محصولات که مشتریان احتمالاً خریداری خواهند کرد یا استفاده مشکوک از شماره کارت اعتباری ارائه دهند.
- **خدمات بانکداری.** از الگوهای متوالی خدماتی که مشتریان استفاده می‌کنند (چک کردن حساب به دنبال حساب پس‌انداز) می‌توان برای شناسایی سایر خدماتی که ممکن است برایشان جذاب باشد استفاده کرد (حساب سرمایه‌گذاری).
- **محصولات و خدمات بیمه:** از مجموعه محصولات و خدمات بیمه‌ای که مشتریان خریداری می‌کنند (بیمه خودرو و بیمه خانه) می‌توان برای ارائه‌ی سایر محصولات و خدمات بیمه‌ای (بیمه عمر) استفاده کرد یا ترکیب‌های نامعمول ادعاهای بیمه می‌توانند نشانه‌ی کلاهبرداری باشند.
- **خدمات مخابراتی:** گروه‌هایی از خدمات و امکانات که عمدتاً توسط مشتریان خریداری می‌شوند (مثلاً انتظار تماس، شماره‌انداز تلفن، تماس سه‌طرفه) کمک می‌کنند تا ساختار محصولات و خدمات را بهتر انتخاب کرده و به حداکثر درآمد دست پیدا کنند. همین موضوع در خصوص ارائه دهندگان خدمات مخابراتی چندکاناله با تلفن، تلویزیون و خدمات اینترنتی نیز صادق است.

- **سوابق پزشکی:** ترکیب‌های خاصی از شرایط می‌توانند نشان‌دهنده‌ی افزایش ریسک عوارض و پیچیدگی‌های مختلف بوده یا روش‌های درمانی خاص در مراکز درمانی خاص را می‌توان به انواع خاصی از عفونت‌ها مرتبط دانست.

سوال خوبی که در رابطه با الگوها/روابطی که کاوش قانون وابستگی می‌تواند کشف کند می‌توانیم مطرح کنیم این است که « آیا تمامی قوانین وابستگی برای ما جالب و مفید هستند؟ » برای پاسخ به چنین سوالی، کاوش قانون وابستگی از دو سنج‌ی متداول استفاده می‌کند: **حمایت و اطمینان و لیفت**. پیش از تعریف این عبارات، اجازه بدهید نگاهی فنی به یک قانون وابستگی داشته باشیم:

$$X \Rightarrow Y [Supp(\%), Conf(\%)]$$

$$\{\text{Laptop Computer, Antivirus software}\} \Rightarrow \{\text{Extended Service Plan}\} [30\%, 70\%]$$

در اینجا X (محصول یا خدمت که سمت چپ، LHS یا مقدم و پیشین نامیده می‌شود) با Y وابستگی دارد (محصولات و/یا خدمات که سمت راست، RHS یا مؤخر یا بعدی نامیده می‌شود). S حمایت و C اطمینان برای این قانون خاص هستند. فرمول‌های ساده برای $Supp$ ، $Conf$ و $Lift$ در اینجا ارائه شده است.

$$Support = Supp(X \Rightarrow Y) = \frac{\text{Number of Baskets that contains both } X \text{ and } Y}{\text{Total Number of Baskets}}$$

$$Confidence = Conf(X \Rightarrow Y) = \frac{Supp(X \Rightarrow Y)}{Supp(X)}$$

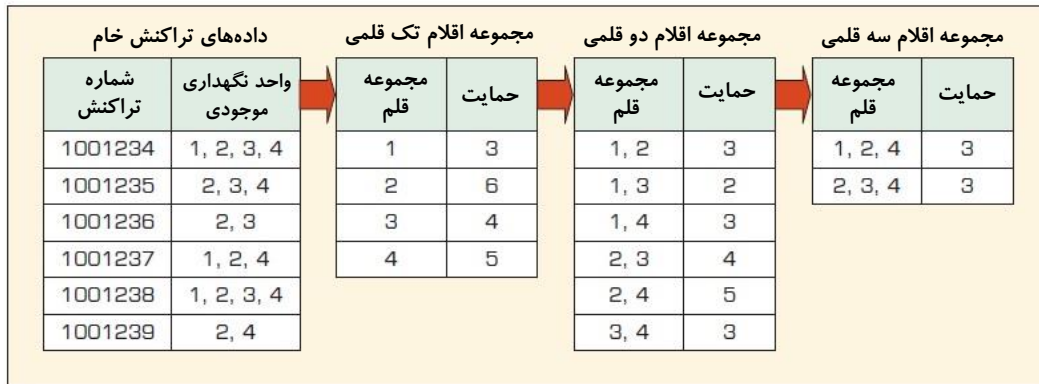
$$Lift(X \Rightarrow Y) = \frac{Conf(X \Rightarrow Y)}{Expeced Conf(X \Rightarrow Y)} = \frac{\frac{S(X \Rightarrow Y)}{S(X)}}{\frac{S(X) * S(Y)}{S(X)}} = \frac{S(X \Rightarrow Y)}{S(X) * S(Y)}$$

حمایت (S) مجموعه‌ای از محصولات است که می‌سجد که این محصولات و/یا خدمات (یعنی $LHS + RHS = \text{Laptop Computer, Antivirus Software, Extended Service Plan}$) هر چند وقت یکبار در یک معامله یا تراکنش در کنار یکدیگر ظاهر می‌شوند یعنی نسبت تراکنش‌ها در مجموعه داده‌ها که تمامی محصولات و/یا خدمات ذکر شده در یک قانون خاص را شامل می‌شوند. در این مثال، ۳۰ درصد از تمامی تراکنش‌ها در پایگاه داده‌ی یک فروشگاه فرضی هر سه این محصولات را در سبد خود داشته‌اند. اطمینان یک قانون به این معنی است که محصولات و/یا خدمات در سمت راست هر چند وقت یکبار با محصولات و/یا خدمات در سمت چپ می‌آیند یعنی نسبتی از تراکنش‌ها که شامل سمت راست و همچنین سمت چپ هستند. به عبارت دیگر، احتمال شرطی یافتن سمت راست قانون در تراکنش‌ها که سمت چپ قانون قبلاً وجود داشته باشد. مقدار لیفت یک قانون وابستگی

نسبت عبارتست از نسبت اطمینان قانون به اطمینان مورد انتظار از آن قانون. اطمینان مورد انتظار قانون به صورت حاصلضرب مقادیر حمایت سمت راست و سمت چپ تقسیم بر حمایت سمت چپ محاسبه می‌شود.

الگوریتم‌های متعددی برای کشف قوانین وابستگی وجود دارد. برخی از شناخته‌شده‌ترین الگوریتم‌ها شامل آپریوری، اکلان و FP-Growth هستند. این الگوریتم‌ها فقط نیمی از کار را انجام می‌دهند که همانا شناسایی مجموعه اقلام تکرار شونده در یک فرایند انطباق مستقیم است ولی این فرایند در پایگاه داده‌های بزرگ می‌تواند بسیار زمان‌بر باشد. اگرچه آیتم‌های بسیاری در هر بخش از قانون ممکن است وجود داشته باشد در عمل، بخش سمت راست قانون معمولاً حاوی یک آیتم تکی است. در بخش بعدی به توضیح یکی از معروف‌ترین الگوریتم‌های شناسایی مجموعه اقلام تکرار شونده خواهیم پرداخت.

الگوریتم آپریوری الگوریتم آپریوری^۱ پرکاربردترین الگوریتم برای کشف قوانین وابستگی است. برای یک مجموعه از اقلام (مثلاً مجموعه تراکنش‌های خرده‌فروشی که لیست تک تک اقلام خریداری شده در هر یک را نشان می‌دهد)، این الگوریتم تلاش می‌کند تا زیرمجموعه‌هایی را بیابد که حداقل تعداد اقلام مشترک را با یکدیگر دارند (یعنی با حمایت مینیمم انطباق دارد). آپریوری از یک روش پایین به بالا استفاده می‌کند که زیرمجموعه‌های تکرار شونده را به اندازه‌ی یک قلم در هر زمان گسترش می‌دهد (روشی به نام تولید کاندید^۲ که اندازه‌ی زیرمجموعه‌های تکرار شونده را از زیرمجموعه‌ی یک قلمی به دو قلمی و سپس سه قلمی و ... افزایش می‌دهد) و گروه کاندیدها در هر سطح برای اطمینان از حداقل حمایت با داده‌ها تست می‌شوند. الگوریتم زمانی خاتمه می‌یابد که هیچ گسترشی امکان‌پذیر نباشد.



شکل ۴.۱۴ شناسایی مجموعه اقلام تکرار شونده در الگوریتم آپریوری

به عنوان مثال، فرض کنید یک خواروبارفروشی تراکنش‌های فروش خود را با روش واحد نگهداری موجودی ردیابی می‌کند و بنابراین می‌داند که کدام اقلام معمولاً با یکدیگر خریداری می‌شوند. پایگاه داده تراکنش‌ها به همراه گام‌های زیر برای شناسایی مجموعه اقلام تکرار شونده در شکل ۴.۱۴ نشان داده شده است. هر واحد نگهداری موجودی در پایگاه داده تراکنش‌ها متناظر با یک محصول است مثلاً « ۱ = کره »، « ۲ = نان »، « ۳ = آب » و الی آخر. گام اول در الگوریتم آپریوری این است که تواترها (یعنی حمایت‌ها) برای هر قلم شمارش شود (مجموعه اقلام تک قلمی). برای این مثال بسیار ساده، حداقل مقدار حمایت را ۳ در نظر می‌گیریم (یا ۵۰ به این معنی که اگر یک مجموعه اقلام حداقل در ۳ تراکنش از ۶ تراکنش در پایگاه داده ظاهر شده باشد، آن

¹ Apriori Algorithm

² Candidate Generation

مجموعه قلم را تکرارشونده می‌دانیم). از آنجائیکه تمامی مجموعه اقلام تک قلمی دارای حداقل عدد حمایت ۳ هستند، تمامی آنها را مجموعه اقلام تکرارشونده در نظر می‌گیریم. با وجود این، در صورتیکه هر یک از مجموعه اقلام تک قلمی تکرار نشده باشند، به عنوان یک عضو احتمالی از زوج‌های دو قلمی در نظر گرفته خواهند شد. بدین ترتیب، الگوریتم آپریوری، درخت تمامی مجموعه اقلام را هرس می‌کند. همانطور که در شکل ۴.۱۴ مشاهده می‌کنید، با استفاده از مجموعه اقلام تک قلمی، تمامی مجموعه اقلام دو قلمی ممکن تولید شده و از پایگاه داده تراکنش‌ها برای محاسبه‌ی مقادیر حمایت آنها استفاده می‌شود. از آنجائیکه مقدار حمایت مجموعه‌ی دو قلمی {۱، ۳} کمتر از ۳ است جزو مجموع اقلام دو قلمی که برای تولید مجموعه اقلام سطح بعدی بکار می‌روند استفاده نخواهد شد (مجموعه اقلام سه قلمی). این الگوریتم به طرز فریبنده‌ای ساده به نظر می‌رسد ولی فقط برای مجموعه داده‌های کوچک این چنین است. در مجموعه داده‌های بسیار بزرگتر، به ویژه آنهایی که حجم انبوهی از اقلام را در تعداد کم و حجم کوچکی از اقلام را در تعداد انبوه دارند، جستجو و محاسبه فرایندی بسیار پرزحمت و زمان‌بر خواهد شد.

سوالاتی برای مرور بخش ۴.۵

۱. حداقل سه روش اصلی داده‌کاوی را شناسایی کنید؟
۲. مثال‌هایی از موقعیت‌هایی بزنید که طبقه‌بندی، روش مناسبی برای داده‌کاوی باشد. مثال‌هایی از موقعیت‌هایی بزنید که رگرسیون روش مناسبی برای داده‌کاوی باشد.
۳. حداقل دو تکنیک طبقه‌بندی را نام برده و به اختصار توضیح دهید.
۴. برخی از معیارهای مقایسه و انتخاب بهترین تکنیک طبقه‌بندی چه هستند؟
۵. الگوریتم عمومی بکار رفته در درخت تصمیم را به اختصار توضیح دهید.
۶. شاخص جینی را تعریف کنید. این شاخص چه چیزی را اندازه‌گیری می‌کند؟
۷. مثال‌هایی از موقعیت‌ها بزنید که در آنها تحلیل خوشه، روش مناسبی برای داده‌کاوی باشد.
۸. تفاوت اصلی بین تحلیل خوشه و طبقه‌بندی چیست؟
۹. برخی روش‌های تحلیل خوشه را نام ببرید.
۱۰. مثال‌هایی از موقعیت‌هایی بزنید که در آنها قانون وابستگی روش مناسبی برای داده‌کاوی باشد.

بسیاری از شرکت‌های نرم‌افزاری، ابزارهای داده‌کاوی قدرتمندی را عرضه می‌کنند. مثال‌هایی از این شرکت‌ها عبارتند از آی‌بی‌ام (آی‌بی‌ام‌اس‌پی‌اس‌مدلر^۱ که سابقاً با نام کِلِمِنَتاین^۲ شناخته می‌شد)، ساس (اینترپرایز ماینر^۳)، دل (استاتیسیتیکا^۴ که سابقاً با نام استات‌سافت استاتیسیتیکا دیتا ماینز^۵ شناخته می‌شد)، اس‌ای‌پی (اینفینیت اینسایت^۶ که سابقاً KXEN اینفینیت اینسایت

¹ IBM SPSS Modeler

² Clementine

³ Enterprise Miner

⁴ Statistica

⁵ StatSoft Statistica Data Miner

⁶ Infinite Insight

شناخته می‌شد)، سالفورد سیستمز^۱ (کارت^۲، مارس^۳، تری‌نت^۴، رندوم‌فارست^۵)، آگنوس^۶ (نالچ‌استودیو^۷، نالچ‌سیکر^۸) و مگ‌پی‌وتر^۹ (پلی‌آنالیست^{۱۰}) هستند. اکثر ابزارهای معروف داده‌کاوی توسط شرکت‌های با سابقه‌ی نرم‌افزارهای آماری (ساس، اس‌پی‌اس‌اس و استات‌سافت) توسعه یافته‌اند که عمدتاً به این دلیل است که آمار، پایه و اساس داده‌کاوی است و این شرکت‌ها نیز وسایل توسعه‌ی این سیستم‌های داده‌کاوی را با قیمت مناسب در اختیار دارند. اکثر عرضه‌کنندگان ابزارهای هوش کسب‌وکار (مثلاً آی‌بی‌ام کوگنوس^{۱۱}، اوراکل هایپریون^{۱۲}، اس‌ای‌پی بی‌زینس آجکتز^{۱۳}، تابلو^{۱۴}، تیبکو^{۱۵}، کیولیک^{۱۶}، مایکرواستراتژی^{۱۷}، ترادیتا و مایکروسافت) سطحی از قابلیت‌های داده‌کاوی را در قالب سایر مجموعه‌های نرم‌افزاری خود عرضه می‌کنند. این ابزارهای هوش کسب‌وکار همچنان تمرکزشان بر مدل‌سازی چندبُعدی و مصورسازی داده‌ها بوده و رقباتی جدی برای عرضه‌کنندگان ابزارهای داده‌کاوی محسوب نمی‌شوند.

علاوه بر این ابزارهای تجاری، چندین ابزار نرم‌افزاری داده‌کاوی منبع باز و/یا رایگان نیز به صورت آنلاین موجود است. محبوب‌ترین ابزار داده‌کاوی رایگان و منبع باز به ویژه در حوزه‌های آموزشی، وکا^{۱۸} نام دارد که توسط تعدادی از پژوهشگران دانشگاه وایکاتو در نیوزیلند توسعه یافت (این ابزار را می‌توانید از cs.waikato.ac.nz/ml/weka دانلود کنید). وکا شامل تعداد زیادی از الگوریتم‌ها برای کارهای مختلف داده‌کاوی بوده و یک رابط کاربری شهودی دارد. اخیراً، چندین ابزار داده‌کاوی رایگان منبع باز و با قابلیت‌های سطح بالا عرضه شده‌اند: سردسته‌ی آنها **کی‌نایم**^{۱۹} (knime.org) و **راپید‌ماینر**^{۲۰} (rapidminer.com) هستند. رابط‌های کاربری ارتقاء یافته‌ی گرافیکی آنها، بکارگیری تعداد زیادی از الگوریتم‌ها و استفاده از مجموعه ویژگی‌های متنوع مصورسازی داده‌ها آنها را از سایر ابزارهای رایگان متمایز کرده است. این دو ابزار نرم‌افزاری رایگان بر روی هر دو سیستم عامل ویندوز و مک اجرا می‌شوند. ریپید‌ماینر با یک تغییر جدید در محصولاتش، نسخه‌ای کوچکتر از ابزار تحلیلی خود را به رایگان عرضه کرده است و نسخه کامل آن به صورت تجاری به فروش می‌رسد. از این رو، ریپید‌ماینر که زمانی نامش در گروه ابزارهای رایگان/منبع باز فهرست می‌شد امروزه اغلب در گروه ابزارهای تجاری مطرح است. تفاوت عمده بین ابزارهای تجاری مثل ساس اینترپرایز ماینر^{۲۱}، آی‌بی‌ام

¹ Salford Systems

² CART

³ MARS

⁴ TreeNet

⁵ RandomForest

⁶ Agnoss

⁷ KnowledgeSTUDIO

⁸ KnowledgeSEEKER

⁹ Megputer

¹⁰ PolyAnalyst

¹¹ IBM Cognos

¹² Oracle Hyperion

¹³ SAP Business Objects

¹⁴ Tableau

¹⁵ Tibco

¹⁶ Qlik

¹⁷ MicroStrategy

¹⁸ Weka

¹⁹ KNIME

²⁰ RapidMiner

²¹ SAS Enterprise Miner

اس‌پی‌اس‌مدلر و استاتیسیتیکا و ابزارهای رایگان از قبیل وکا، رپیدماینر (نسخه‌ی کامیونیتی) و کی‌نایم در کارایی محاسبات آنها است. پردازش یک مجموعه داده‌ی نسبتاً بزرگ با ابزارهای نرم‌افزاری رایگان مدل زمان بیشتری طول می‌کشد و برخی وظایف داده‌کاوی ممکن است برای برخی الگوریتم‌ها به طور کامل انجام نشود (یعنی متوقف شدن نرم‌افزار به دلیل استفاده‌ی ناکارآمد در حافظه‌ی رایانه). جدول ۴.۲ چندین محصول اصلی به همراه وب‌سایت‌های آنها ارائه شده است. یک مجموعه‌ی کامل از قابلیت‌های تحلیلی و هوش کسب‌وکار که محبوبیت فزاینده‌ای در مطالعات داده‌کاوی پیدا کرده، **مایکروسافت اس‌کیوال سرور**^۱ می‌باشد (قابلیت‌های تحلیلی بیشتری از قبیل هوش کسب‌وکار و ماژول‌های مدلسازی پیشگویانه را از نسخه‌ی اس‌کیوال سرور ۲۰۱۲ گنجانده است) که داده‌ها و مدل‌ها در یک محیط پایگاه داده رابطه‌ای یکسان ذخیره شده و موجب آسان‌تر شدن مدیریت مدل می‌شوند. **مایکروسافت اینترپرایز کنسرسیوم**^۲ به عنوان یک منبع جهانی برای دسترسی به مجموعه مایکروسافت اس‌کیوال سرور برای مقاصد آکادمیک – آموزش و پژوهش عمل می‌کند. این کنسرسیوم به این منظور ایجاد شد است تا دانشگاه‌های سرتاسر جهان بتوانند بدون نیاز به تهیه‌ی سخت‌افزار و نرم‌افزار لازم در پردیس‌های دانشگاهی خود به این فناوری سازمانی دست پیدا کنند. این کنسرسیوم طیف گسترده‌ای از ابزارهای توسعه‌ی هوش کسب‌وکار (مثلاً داده‌کاوی، مکعب‌سازی، گزارش‌دهی کسب‌وکار) و همچنین چندین مجموعه داده‌های بزرگ و واقعی از سَمز کلاب، دیلاردز و تاپسون فودز را ارائه می‌کند. مایکروسافت اینترپرایز کنسرسیوم، رایگان بوده و فقط برای مقاصد دانشگاهی قابل استفاده است. کالج بیزینس سَم ام والتون^۳ از دانشگاه آرکانزاس میزبان این سیستم بوده به اعضای کنسرسیوم و دانشجویان آنها اجازه دسترسی به این منابع را با استفاده از یک اتصال رومیزی از راه دور می‌دهد. جزئیات عضویت در کنسرسیوم و راهنماهای ساده و مثال‌ها را می‌توانید در walton.uark.edu/enterprise مشاهده کنید.

جدول ۴.۲ گزینه نرم‌افزارهای داده‌کاوی

نام محصول	وب سایت (آدرس)
IBM SPSS Modeler	www-01.ibm.com/software/analytics/spss/products/modeler/
IBM Watson Analytics	ibm.com/analytics/watson-analytics/
SAS Enterprise Miner	sas.com/en_id/software/analytics/enterprise-miner.html
Dell Statistica	statsoft.com/products/statistica/product-index
PolyAnalyst	megaputer.com/site/polyanalyst.php
CART, RandomForest	salford-systems.com
Insightful Miner	solutionmetrics.com.au/products/iminer/default.html
XLMiner	solver.com/xlminer-data-mining
SAP InfiniteInsight (KXEN)	help.sap.com/ii
GhostMiner	GhostMiner

¹ Microsoft SQL Server

² Microsoft Enterprise Consortium

³ Sam M. Walton's Business College

msdn.microsoft.com/en-us/library/bb510516.aspx

SQL Server Data Mining

knowledgeminer.com

Knowledge Miner

teradata.com/products-and-services/teradata-warehouseminer/

Teradata Warehouse Miner

oracle.com/technetwork/database/options/odm/

Oracle Data Mining (ODM)

fico.com/en/analytics/decision-management-suite/

FICO Decision Management

orange.biolab.si/

Orange Data Mining tool

zementis.com

Zemetis Predictive Analytics

در می ۲۰۱۶، **KDnuggets.com** سیزدهمین نظرسنجی سالانه خود در خصوص نرم‌افزار را با سوال زیر انجام داد: « شما طی ۱۲ ماه گذشته از کدام نرم‌افزار برای پروژه‌های تحلیل، داده‌کاوی، علم داده‌ها و یادگیری ماشینی استفاده کرده‌اید؟ ». سطح مشارکت جامعه‌ی علم داده‌ها و تحلیل و عرضه‌کنندگان نرم‌افزارها در این نظرسنجی بسیار بالا بوده و ۲,۸۹۵ رأی‌دهنده شرکت داشته‌اند که از بین ۱۰۲ ابزار مختلف انتخاب کرده‌اند. برخی از یافته‌های جالب این نظرسنجی به شرح زیر می‌باشند:

- نرم‌افزار « آر » با ۴۹ درصد سهم همچنان رتبه اول را به خود اختصاص داده ولی استفاده از پایتون افزایش یافته و با ۴۵.۸٪ به « آر » نزدیک شده است (از ۳۰.۳٪).
- رپیدمایزر همچنان محبوب‌ترین پلتفرم عمومی برای داده‌کاوی/علم داده‌ها با ۳۳ درصد سهم است. ابزارهای قابل توجهی که بیشترین رشد را داشته‌اند شامل جی^۱، دیتاایکو^۲، ام‌ال‌لیب^۳، اچ‌تو^۴، یادگیری ماشینی آمازون^۵، سای‌کیت-لرن^۶ و آی‌بی‌ام واتسون هستند.
- ابزارها در طیف‌های گسترده‌ای مورد استفاده قرار گرفته‌اند و متوسط تعداد ابزارهای بکار رفته معادل ۶ بوده است (این عدد در سال ۲۰۱۵ معادل ۴.۸ بوده است).
- استفاده از ابزارهای هُدوپ/کلان‌داده‌ها از ۲۹ درصد در سال ۲۰۱۵ به ۳۹ درصد افزایش یافته است که آپاچی اسپارک^۷، ام‌ال‌لیس و اچ‌تو^۴ در صدر قرار دارند.
- بیشترین مشارکت از مناطق به ترتیب ایالات متحده/کانادا (۴۰٪)، اروپا (۳۹٪)، آسیا (۹.۴٪)، آمریکای لاتین (۵.۸٪)، آفریقا/خاورمیانه (۲.۹٪) و استرالیا/نیوزیلند (۲.۲٪) بوده است.
- در سال ۲۰۱۶، ۸۶ درصد از رأی‌دهندگان از نرم‌افزارهای تجاری و ۷۵ از نرم‌افزارهای رایگان استفاده کرده‌اند. در حدود ۲۵ درصد فقط از نرم‌افزارهای تجاری و ۱۳ درصد فقط از نرم‌افزارهای رایگان/منبع باز استفاده کرده‌اند. اکثریت ۶۱ درصدی هم از نرم‌افزارهای تجاری و هم رایگان استفاده کرده‌اند که این عدد در سال ۲۰۱۵ معادل ۶۴ درصد بوده است.

¹ g

² Dataiku

³ MLlib

⁴ H2O

⁵ Amazon Machine Learning

⁶ Scikit-learn

⁷ Apache Spark

- استفاده از ابزارهای هَدوپ/کلان داده‌ها از ۲۹ درصد در سال ۲۰۱۵ و ۱۷ درصد در سال ۲۰۱۴ به ۳۹ درصد افزایش یافت که عمدتاً به دلیل رشد زیاد نرم‌افزارهای آپاچی اسپارک^۱، ام‌ال‌لیس و اچ‌تو^۲ بوده که جزو ابزارهای کلان داده‌ها آنها را قرار می‌دهیم.
- برای دومین سال، رأی‌گیری **KDnuggets.com** شامل ابزارهای یادگیری عمیق نیز بوده است. امسال، ۱۸ درصد از رأی‌دهندگان از ابزارهای یادگیری عمیق استفاده کرده‌اند که آمار سال ۲۰۱۵ را به دو برابر رسانده است – گوگل تِنسورفلو به جایگاه اول رفته و جایگزین اکوسیستم Theano/Pylearn2 شده است.
- در گروه زبان‌های برنامه‌نویسی، بر محبوبیت پایتون، جاوا ابزارهای یونیکس و اسکالا^۳ افزوده شده در حالیکه C/C++، پرل^۴، جولیا^۵، F#، کلوزر^۶ و لیسپ^۶ نزول کرده‌اند.

برای کاهش انحراف ناشی از رأی‌دهی چندباره، **KDnuggets.com** از اعتبارسنجی از طریق پست الکترونیکی استفاده کرده و با اینکار قصد داشت تا به نتایجی دست یابد که بیشتر نماینده‌ی واقعیت موجود در دنیای تجزیه و تحلیل باشند. نتایج ۴۰ ابزار نرم‌افزاری برتر (به ازای آراء دریافت شده) در شکل ۴.۱۵ نشان داده شده‌اند. نمودار میله‌ای افقی نیز متمایزسازی بین ابزارهای رایگان/منبع باز، تجاری و کلان داده/هدوپ را با استفاده از الگوهای رنگ‌بندی نشان می‌دهد.

مورد کاربردی ۴.۶ در مورد یک مطالعه پژوهشی است که در آن تعدادی از ابزارهای نرم‌افزاری و تکنیک‌های داده‌کاوی برای ساخت مدل‌های داده‌کاوی جهت پیش‌بینی موفقیت مالی فیلم‌های هالیوود در گیشه زمانیکه فقط در حد ایده و فکر هستند بکار رفته‌اند.

¹ Apache Spark

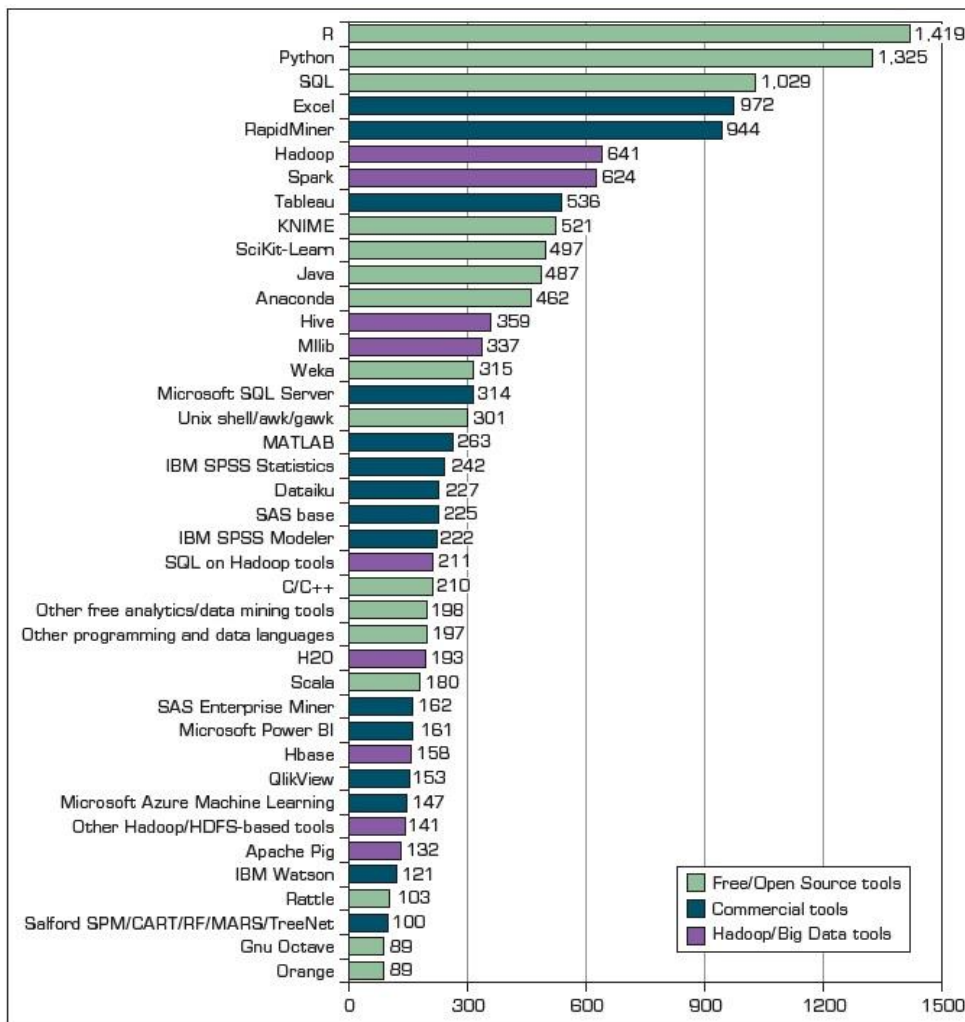
² Scala

³ Perl

⁴ Julia

⁵ Clojure

⁶ Lisp



شکل ۴.۱۵ ابزارهای نرم‌افزاری متداول داده‌کاوی (نتایج رأی‌گیری)

Source: Used with permission from KDnuggets.com

سوالاتی برای مرور بخش ۴.۶

۱. محبوب‌ترین ابزارهای تجاری داده‌کاوی کدام‌ها هستند؟
۲. به نظر شما چرا محبوب‌ترین ابزارها توسط شرکت‌های آماری تولید می‌شوند؟
۳. محبوب‌ترین ابزارهای رایگان داده‌کاوی کدام‌ها هستند؟ دلیل افزایش محبوبیت آنها چیست (به ویژه نرم‌افزار آر)؟
۴. تفاوت‌های اصلی بین ابزارهای نرم‌افزاری تجاری و رایگان داده‌کاوی چه هستند؟
۵. پنج معیار اصلی شما برای انتخاب یک ابزار داده‌کاوی چیست؟ توضیح دهید.

پیش‌بینی استقبال یک فیلم خاص در گیشه (یعنی موفقیت مالی)، مسئله‌ی چالشی و جذابی است. طبق نظرات برخی کارشناسان این حوزه، صنعت فیلم‌سازی به دلیل دشواری‌های مربوط به پیش‌بینی تقاضای محصول که موجب شده تا کسب‌وکار فیلم‌سازی به فعالیتی پرمخاطره در هالیوود بدل شود، «دنیای حدس و گمان‌های دشوار» نامیده می‌شود. جک ولانتی^۱ (رئیس و مدیرعامل قدیمی انجمن سینمایی آمریکا) یکبار اشاره کرد که «هیچکس نمی‌تواند بگوید یک فیلم در بازار چگونه عمل خواهد کرد ... تا زمانیکه فیلم در یک سالن نمایش تاریک اکران شده و جرقه‌ها بین پرده سینما و مخاطبان به پرواز درآیند.» روزنامه‌ها و مجلات صنعت سرگرمی مملو از مثال‌ها، بیابیه‌ها و تجاربی هستند که از چنین ادعایی حمایت می‌کنند.

همانند بسیاری از پژوهشگران دیگر که سعی کرده‌اند تا این مسئله‌ی چالش‌برانگیز دنیای واقعی را روشن کنند، رامش شادرا^۲ و دورسون دیلن^۳ به بررسی استفاده از داده‌کاوی برای پیش‌بینی عملکرد مالی صنعت فیلم‌سازی در گیشه و پیش از اینکه وارد مرحله تولید شود (زمانیکه فیلم در حد یک ایده و نظر مفهومی است) پرداخته‌اند. آنها در مدل‌های پیش‌بینی خود، مسئله‌ی پیش‌بینی یا رگرسیون را به یک مسئله‌ی رده‌بندی یا دسته‌بندی تبدیل کرده‌اند به این مفهوم که به جای پیش‌بینی تخمین نقطه‌ای استقبال فیلم در گیشه، آنها یک فیلم را بر اساس استقبال گیشه در یکی از نه دسته طبقه‌بندی می‌کنند که از «شکست کامل» تا «بلاک‌باستر» متغیر است. جدول ۴.۳ تعریف این نه دسته را بر حسب میزان پذیرش و استقبال فیلم در گیشه نشان می‌دهد.

داده‌ها

داده‌ها از طیف وسیعی از پایگاه‌های داده مربوط به فیلم‌ها گردآوری شده (مثلاً شوپیز^۴، آی‌ام‌دی‌بی^۵، آی‌ام‌اس‌دی‌بی^۶، آل مووی^۷، باکس‌آفیس‌موجو^۸) و در قالب یک مجموعه داده ادغام شده‌اند. مجموعه داده‌ها برای جدیدترین مدل‌ها حاوی ۲۶۳۲ فیلم بوده که بین سال‌های ۱۹۹۸ تا ۲۰۰۶ تولید شده‌اند. خلاصه‌ای از متغیرهای مستقل به همراه مشخصه‌های آنها در جدول ۴.۴ آورده شده است. برای جزئیات تفصیلی بیشتر و دلایل بکارگیری این متغیرهای مستقل خواننده را به شادرا و دیلن (۲۰۰۶) ارجاع می‌دهیم.

متدولوژی

با استفاده از روش‌های مختلف داده‌کاوی از جمله شبکه‌های عصبی، درخت‌های تصمیم، ماشین‌های برداری پشتیبان و سه نوع روش ترکیبی، شادرا و دیلن (۲۰۰۶) مدل‌های پیش‌بینی را توسعه دادند. داده‌های سال‌های ۱۹۹۸ تا ۲۰۰۵ به عنوان داده‌های آموزشی برای ساخت مدل‌های پیش‌بینی استفاده شدند و داده‌های سال ۲۰۰۶ به عنوان داده‌های تستی برای ارزیابی و مقایسه‌ی دقت پیش‌بینی مدل بکار رفتند. شکل ۴.۱۶ یک اسکرین‌شات از آی‌بی‌ام‌اس‌پی‌اس‌مدلر (با نام سابق ابزار داده‌کاوی کلمنتاین) را نشان می‌دهد که نقشه فرایند بکار رفته برای مسئله‌ی پیش‌بینی را به تصویر کشیده است. قسمت بالا سمت چپ نقشه، فرایند

¹ Jack Valenti

² Ramesh Shadra

³ Dursun Delen

⁴ ShowBiz

⁵ IMDb

⁶ IMSDb

⁷ AllMovie

⁸ BoxofficeMojo

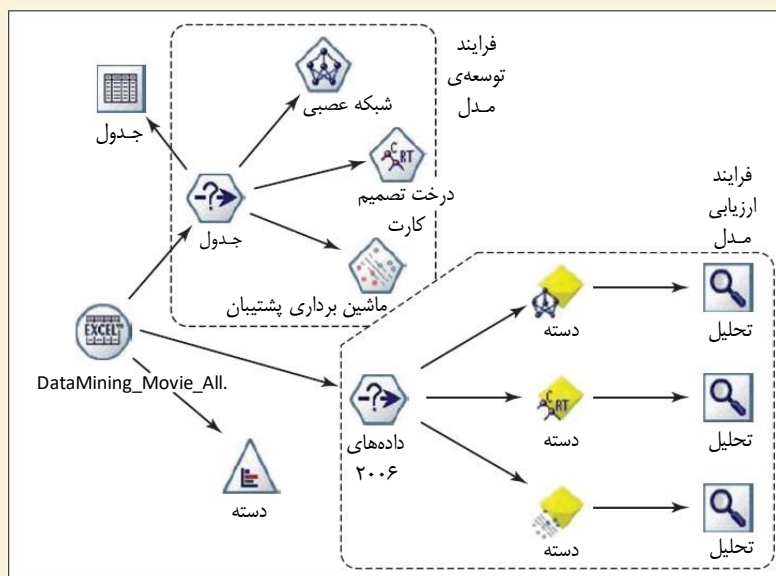
توسعه‌ی مدل را نشان می‌دهد و گوشه‌ی پایین سمت راست نقشه نیز فرایند ارزیابی مدل (یعنی تست یا امتیازدهی) را نشان می‌دهد (جزئیات بیشتر در خصوص ابزار آی‌بی‌ام‌اس‌پی‌اس‌مدلر را در وب سایت کتاب می‌توانید پیدا کنید).

جدول ۴.۳ دسته‌بندی فیلم‌ها بر اساس استقبال در گیشه

شماره دسته	۱	۲	۳	۴	۵	۶	۷	۸	۹
دامنه (میلیون دلار)	1 < شکست کامل	1 < 6 10>	10 < 6 20>	20 < 6 40>	40 < 6 65>	65 < 6 100>	100 < 6 150>	150 < 6 200>	200 < بلاک‌باستر

جدول ۴.۴ چکیده متغیرهای مستقل

متغیر مستقل	تعداد مقادیر	مقادیر ممکن
سیستم درجه‌بندی فیلم‌ها توسط انجمن تصاویر متحرک آمریکا ^۱	۵	G, PG, PG-13, R, NR
رقابت	۳	بالا، متوسط، پایین
ارزش ستاره	۳	بالا، متوسط، پایین
ژانر	۱۰	علمی-تخیلی، درام حماسی، درام مدرن، سیاسی، تریلر، وحشت، کمدی، کارتون، اکشن، مستند
جلوه‌های ویژه	۳	بالا، متوسط، پایین
دنباله	۲	دارد/ندارد
تعداد صفحات	۱	یک عدد صحیح مثبت بین ۱ و ۳,۸۷۶



شکل ۴.۱۶ اسکرین‌شات جریان فرایند سیستم پیش‌بینی استقبال در گیشه

Source: Reprint Courtesy of International Business Machines Corporation, © International Business Machines Corporation.

¹ MPAA Rating

نتایج

جدول ۴.۵ نتایج پیش‌بینی برای هر سه روش داده‌کاوی را به همراه نتایج سه روش جمعی نشان می‌دهد. اولین سنج‌هی عملکردی، درصد نرخ طبقه‌بندی یا دسته‌بندی صحیح است که آن را بینگو^۱ می‌نامند. در این جدول، نرخ طبقه‌بندی صحیح I-Away نیز گزارش شده است (یعنی در داخل یک دسته یا طبقه). نتایج نشان می‌دهند که ماشین برداری پشتیبان بهترین عملکرد را بین مدل‌های پیش‌بینی انفرادی داشته که پس از آن شبکه‌های عصبی مصنوعی قرار دارد. بدترین مدل بین این سه، الگوریتم درخت تصمیم کارت است. به طور کلی، مدل‌های جمعی نسبت به مدل‌های انفرادی عملکرد بهتری دارند که الگوریتم تلفیق در بین آنها بهترین عملکرد را داشته است. آنچه شاید برای تصمیم‌گیرندگان مهمتر است و در نتایج جدول خودش را نشان می‌دهد، پایین بودن معنادار انحراف استاندارد نتایج مدل‌های جمعی در مقایسه با مدل‌های انفرادی است.

نتیجه‌گیری

پژوهشگران معتقدند که این نتایج پیش‌بینی بهتر از هر نتیجه‌ی منتشر شده در ادبیات موضوع در این حوزه هستند. و رای دقت بسیار بالای نتایج پیش‌بینی این مدل‌ها در موفقیت فیلم‌ها در گیشه، این مدل‌ها را همچنین می‌توان برای تحلیل بیشتر (و بهینه‌سازی) متغیرهای تصمیم برای پیشینه کردن نرخ بازگشت مالی بکار برد. به طور خاص، پارامترهای بکار رفته برای مدل‌سازی را می‌توان با استفاده از مدل‌های پیش‌بینی که قبلاً آموزش دیده‌اند به نحوی تغییر داد تا به درک بهتری از تأثیر پارامترهای مختلف بر نتایج نهایی دست پیدا کنیم. در طی این فرایند، که عمدتاً از آن به تحلیل حساسیت یاد می‌شود، تصمیم‌گیر در یک شرکت سرگرمی با یک میزان دقت نسبتاً بالا می‌تواند پی ببرد که یک بازیگر خاص (یا یک تاریخ انتشار خاص یا افزودن جلوه‌های فنی بیشتر و غیره) چه اندازه موفقیت مالی می‌تواند برای فیلم به همراه داشته باشد و این سیستم بسیار ارزشمندی برای کمک به تصمیم‌گیری محسوب می‌شود.

جدول ۴.۵ نتایج پیش‌بینی جدول‌بندی شده برای مدل‌های انفرادی و جمعی

مدل‌های پیش‌بینی

مدل‌های جمعی			مدل‌های انفرادی			
تلفیق	درخت تقویت شده	جنگل تصادفی	کارت	شبکه عصبی مصنوعی	ماشین برداری پشتیبان	سنج‌هی عملکردی
۱۹۴	۱۸۷	۱۸۹	۱۴۰	۱۸۲	۱۹۲	شمارش (بینگو)
۱۲۰	۱۰۴	۱۲۱	۱۲۶	۱۲۰	۱۰۴	شمارش (I-Away)
۵۶.۰۷%	۵۴.۰۵%	۵۴.۶۲%	۴۰.۴۶%	۵۲.۶۰%	۵۵.۴۹%	دقت (% بینگو)
۹۰.۷۵%	۸۴.۱۰%	۸۹.۶۰%	۷۶.۸۸%	۸۷.۲۸%	۸۵.۵۵%	دقت (% I-Away)
۰.۶۳	۰.۸۴	۰.۷۶	۱.۰۵	۰.۸۷	۰.۹۳	انحراف استاندارد

سوالاتی برای مورد کاربردی ۴.۶

۱. چرا پیش‌بینی موفقیت مالی فیلم‌ها برای بسیاری از کارشناسان و متخصصان هالیوود مهم است؟
۲. چگونه می‌توان از داده‌کاوی برای پیش‌بینی موفقیت مالی فیلم‌ها پیش از آغاز فرایند تولید استفاده کرد؟
۳. به نظر شما هالیوود چگونه بدون کمک ابزارها و تکنیک‌های داده‌کاوی تا کنون عمل کرده و همچنان عمل می‌کند؟

¹ Bingo

داده‌هایی که در تکنیک داده‌کاوی گردآوری، ذخیره و تحلیل می‌شوند اغلب حاوی اطلاعاتی در مورد افراد حقیقی هستند. اینگونه اطلاعات ممکن است شامل داده‌های هویتی (نام، آدرس، شماره امنیت اجتماعی، شماره گواهینامه، شماره پرسنلی و ...)، داده‌های جمعیت‌شناختی (مثلاً سن، جنسیت، قومیت و نژاد، وضعیت تأهل، تعداد فرزندان)، داده‌های مالی (مثلاً حقوق، درآمد ناخالص خانوار، مانده‌ی حساب‌های قرض‌الحسنه یا جاری، مالکیت خانه، مشخصات حساب رهن یا وام، مانده‌ی کارت اعتباری، مشخصات حساب سرمایه‌گذاری)، تاریخچه‌ی خرید (یعنی چه چیزی از کجا و چه وقت خریداری شده - یا از سوابق تراکنش‌های فروشنده یا تراکنش‌های کارت‌های اعتباری مشتری) و سایر داده‌ها (سالگردها، بارداری، بیماری، از دست دادن اعضای خانواده، ورشکستگی) باشند. از طریق ارائه‌دهندگان شخص ثالث داده‌ها می‌توان به اکثر این داده‌ها دسترسی یافت. مسئله‌ی اصلی در اینجا حریم خصوصی شخصی است که داده‌ها به وی تعلق دارد. برای حفظ حریم خصوصی و حفاظت از حقوق فردی، متخصصان داده‌کاوی الزامات اخلاقی (و اغلب قانونی) را در نظر گرفته‌اند. یک راه برای اینکار، فرایند هویت‌زدایی از سوابق مشتری پیش از داده‌کاوی بر روی آنهاست که بدین ترتیب نتوان رد سوابق را به افراد پیدا کرد. بسیاری از منابع داده که به طور عمومی در دسترس هستند قبلاً هویت‌زدایی شده‌اند (مثلاً سی‌دی‌سی دیتا^۱، اس‌ای‌ای‌آر دیتا^۲، یونس دیتا^۳). پیش از دسترسی به این داده‌ها، از کاربران خواسته می‌شود تا تحت هیچ شرایطی سعی در یافتن هویت اشخاصی که داده‌ها مربوط به آنهاست نداشته باشند.

در گذشته‌ی نه چندان دور مثال‌هایی از شرکت‌هایی که داده‌های مشتریان را بدون اخذ رضایت آنها در اختیار دیگران قرار داده‌اند وجود داشته است. به عنوان مثال، اکثر شما به خاطر دارید که شرکت هواپیمایی جت‌بلو^۴ در سال ۲۰۰۳ اطلاعات و سوابق بیش از ۱ میلیون مسافر را در اختیار شرکت پیمانکار دولت ایالات متحده به نام تورچ کانسپتس^۵ قرار داد. شرکت تورچ نیز اطلاعات بیشتری از قبیل تعداد اعضای خانواده و شماره امنیت اجتماعی را که از یک دلال داده به نام اکسیوم^۶ خریداری کرده بود به این داده‌های مسافران افزود. این پایگاه داده ادغام شده قرار بود برای یک پروژه‌ی داده‌کاوی برای ایجاد پروفایل‌های تروریستی بالقوه بکار برود. تمامی این کارها بودن اطلاع یا رضایت مسافران انجام شده بود. زمانیکه اخبار این فعالیت‌ها به بیرون درز کرد، ده‌ها شکایت حقوقی نقض حریم خصوصی علیه شرکت‌های جت‌بلو، تورچ، اکسیوم و چندین سناتور ایالات متحده در دادگاه طرح شد (والد ۲۰۰۴). به همین شکل ولی کمی ضعیف‌تر، اخبار مرتبط با حریم خصوصی گزارش شده در مورد شرکت‌های معروف شبکه‌های اجتماعی بوده است که متهم به فروش داده‌های خاص مشتریان به سایر شرکت‌ها برای بازاریابی هدف‌مند شخصی‌سازی شده بوده است.

¹ CDC Data

² SEER Data

³ UNOS Data

⁴ JetBlue

⁵ Torch Concepts

⁶ Acxiom

یک داستان عجیب دیگر در مورد دغدغه‌های حریم خصوصی در سال ۲۰۱۲ اتفاق افتاد. در این اتفاق، شرکت تارگت حتی از هیچگونه داده‌های شخصی و/یا خصوصی استفاده نکرد. به لحاظ قانونی، هیچگونه تخطی از قوانین اتفاق نیوفتاد. خلاصه‌ی این داستان در مورد کاربردی ۴.۷ آورده شده است.

مورد کاربردی ۴.۷

پیش‌بینی الگوهای خرید مشتریان – داستان تارگت

در اوایل سال ۲۰۱۲، داستانی بدنام در مورد روش تحلیل‌های پیشگوییانه‌ی شرکت تارگت^۱ بر سر زبان‌ها افتاد. این داستان مربوط به دختر نوجوانی بود که فلایرها و کوپن‌های تبلیغاتی شرکت تارگت را برای انواع مایحتاجی که یک خانم باردار ممکن است از فروشگاه‌های مثل تارگت خریداری کند ارسال می‌کند. داستان به این شکل ادامه می‌یابد: مردی خشمگین وارد یک فروشگاه تارگت در خارج از مینیاپولیس شده و می‌خواهد که به مدیر صحبت کند: او گفت «این برای دختر من پست شده است! او هنوز دبیرستان را می‌گذراند و شما کوپن‌های لباس بچه برایش می‌فرستید؟ آیا می‌خواهید تشویقش کنید که باردار شود؟» مدیر اصلاً نمی‌دانست که این مرد در مورد چه چیزی صحبت می‌کند. به آدرس پستی نگاه کرد. هیچ شکی نبود که این بسته برای دختر این مرد پست شده بود و حاوی تبلیغات لباس بارداری، مبلمان مهدکودک و تصاویری از کودکان خندان بود. مدیر عذرخواهی کرده و چند روز بعد مجدداً برای عذرخواهی تماس گرفت. در پشت تلفن، پدر که به نظر خجالت‌زده می‌رسید گفت: «من با دخترم صحبت کردم. به نظر می‌رسد اتفاقاتی در منزل من افتاده است که من از آنها بی‌اطلاع بوده‌ام. او در ماه آگوست وضع حمل می‌کند. من یک عذرخواهی به شما بدهکارم.»

اینطور که مشخص شد، شرکت تارگت پیش از پدر متوجه شد که دختر نوجوان باردار است! شرکت اینطور متوجه شد. تارگت به هر مشتری یک شماره شناسایی مهمان می‌دهد (متصل به کارت اعتباری، نام یا آدرس پست الکترونیکی مشتری) که تاریخچه‌ی تمام چیزهایی که شخص خریداری کرده است را نگهداری می‌کند. تارگت اطلاعات دیگری از قبیل اطلاعات جمعیت‌شناختی را که از مشتری یا سایر منابع اطلاعاتی گردآوری کرده است را به این داده‌ها می‌افزاید. تارگت با استفاده از این اطلاعات نگاهی به داده‌های تاریخی خرید تمامی مشتریان زن که در گذشته برای خرید محصولات نوزادان ثبت‌نام کرده بودند انداخت. آنها داده‌ها را از هر جهت تحلیل کرده و به الگوهای مفیدی دست یافتند. به عنوان مثال، لوسیون و ویتامین‌های خاص محصولاتی بودند که الگوهای خرید جالبی داشتند. بسیاری از مردم لوسیون می‌خرند ولی آنها یک تحلیلگر متوجه شد این بود که زنان در ثبت نام کودک، مقادیر بسیار زیادی از لوسیون‌های بدون عطر را در اوایل سه ماهه‌ی دومشان خریداری می‌کردند. یک تحلیلگر دیگر متوجه شد که در حدود ۲۰ هفته‌ی اول، زنان باردار مواد مکمل مثل کلسیم، منیزیم و روی را بیشتر خریداری می‌کنند. بسیاری از خریداران، صابون و گلوله‌های پنبه‌ای می‌خرند ولی زمانیکه شخصی در کنار ضدعفونی‌کننده‌ها و پارچه‌های شستشو ناگهان شروع به خرید مقادیر زیادی صابون بدون عطر و کیسه‌های بسیار بزرگ گلوله‌های پنبه‌ای می‌کند بدین معناست که تاریخ وضع حمل آنها ممکن است نزدیک باشد. در پایان، تحلیلگرها توانستند حدوداً ۲۵ محصول را شناسایی کنند که زمانیکه با یکدیگر تحلیل شوند می‌توانند به هر خریدار یک امتیاز «پیش‌بینی بارداری» بدهند. مهمتر اینکه، آنها می‌توانند تاریخ وضع حمل شخص را تخمین بزنند و شرکت تارگت، کوپن‌هایی را برای مراحل خاص بارداری شخص ارسال کند. اگر از منظر قانونی به این داستان نگاه کنیم، ممکن است اینگونه نتیجه بگیریم که تارگت از هیچگونه اطلاعاتی که حریم خصوصی مشتری را نقض کند استفاده نکرده است و آنها از داده‌های تراکنشی که هر فروشگاه زنجیره‌ای دیگر آنها را گردآوری و ذخیره

¹ Target

(و شاید تحلیل) می‌کند استفاده می‌کنند. آنچه در این سناریو ناراحت کننده است شاید مفهوم مورد هدف باشد یعنی بارداری. حدود مفاهیم یا رویدادهای خاص مثل بیماری‌های مرحله آخر، طلاق و ورشکستگی باید با احتیاط تعیین شود.

سوالاتی برای مورد کاربردی ۴.۷

۱. در مورد داده‌های آن برای حریم خصوصی چه فکر می‌کنید؟ مرز بین کشف دانش و نقض حریم خصوصی کجاست؟
۲. آیا تارگت پا را فراتر از حد و حدود گذاشت؟ آیا کار غیرقانونی را انجام داد؟ به نظر شما تارگت باید چه کار می‌کرد؟ به نظر شما تارگت باید چه کار کند؟ (از این کارها دست بکشد)؟

Sources: K. Hill, "How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did," *Forbes*, February 16, 2012; R. Nolan, "Behind the Cover Story: How Much Does Target Know?," February 21, 2012. NYTimes.com.

داده‌های کاوی یک ابزار تحلیلی قدرتمند است که مدیران اجرایی سازمان‌ها را قادر می‌سازد تا از توضیح ماهیت آنچه که گذشته (نگاه رو به عقب) به سمت پیش‌بینی آینده (نگاه رو به جلو) برای مدیریت بهتر عملیات سازمان‌شان حرکت کنند (تصمیمات دقیق و به موقع بگیرند). داده‌های کاوی به بازاریاب‌ها در یافتن الگوهایی که رموز رفتار مشتریان را بازگشایی می‌کند کمک می‌کند. از نتایج داده‌های کاوی می‌توان برای افزایش درآمد و کاهش هزینه‌ها از طریق شناسایی فرصت‌های کسب‌وکار و موارد مشکوک به کلاهبرداری استفاده کرد که قلمروی کاملاً جدیدی از مزیت‌های رقابتی را در اختیار ما قرار می‌دهد. داده‌های کاوی به عنوان یک رشته‌ی در حال تکامل و بلوغ اغلب با افسانه‌هایی همراه است که در جدول ۴.۶ فهرست شده‌اند (دیلین ۲۰۱۴، زایما ۲۰۰۳). افراد مبتکر در حوزه داده‌های کاوی با درک این موضوع که این افسانه‌ها صرفاً افسانه هستند توانسته‌اند به مزیت رقابتی عظیمی دست پیدا کنند.

اگرچه گزاره ارزشی و لزوم وجود آن بر همگان واضح و روشن است ولی آنهایی که مسئول اجرای پروژه‌های داده‌های کاوی هستند (از نوآموز تا دانشمند با تجربه در حوزه‌ی داده‌ها) برخی اوقات مرتکب اشتباهاتی می‌شوند که منجر به نتایجی می‌شوند که چندان مطلوب نیستند. در ادامه به ذکر ۱۶ اشتباه داده‌های کاوی (اشتباهات بزرگ^۱، تله‌ها^۲ یا اشتباهات/احمقانه^۳) که اغلب در عمل رخ می‌دهد (نیسبت و همکاران^۴ ۲۰۰۹، شولتز^۵ ۲۰۰۴، اسکالاک^۶ ۲۰۰۱) و دانشمندان علوم داده باید از آنها مطلع باشند و هر آنچه در توان دارند را برای اجتناب از آنها انجام دهند می‌پردازیم.

۱. انتخاب مسئله‌ی اشتباه برای داده‌های کاوی. هر مسئله‌ی کسب‌وکار را نمی‌توان از طریق داده‌های کاوی حل کرد (یعنی سندروم گلوله‌ی جادویی). زمانیکه هیچ داده‌ی نماینده (بزرگ و غنی) وجود نداشته باشد، نمی‌توان یک پروژه‌ی داده‌های کاوی قابل اجرا را تعریف نمود.

¹ Blunders

² Pitfall

³ Blooper

⁴ Nisbet et al.

⁵ Shultz

⁶ Skalak

۲. نادیده گرفتن دیدگاه حامی مالی شما در مورد داده کاوی و اینکه قادر است چه کاری را انجام دهد و چه کاری را انجام ندهد. مدیریت انتظارات در اینجا کلید موفقیت پروژه‌های داده کاوی است.

جدول ۴.۵ افسانه‌های داده کاوی

افسانه	واقعیت
داده کاوی پیش‌بینی‌های فوری و شفافی به ما می‌دهد.	داده کاوی یک فرایند چندمرحله‌ای است که نیازمند طراحی و بکارگیری کنشگرانه و با تأمل است.
داده کاوی هنوز برای کاربردهای اصلی کسب‌وکار قابل اجرا نیست.	وضعیت جاری فناوری تقریباً برای هر نوع کسب‌وکار با هر اندازه‌ای آماده است.
داده کاوی نیازمند یک پایگاه داده‌ی مجزا و اختصاصی است.	به دلیل پیشرفت‌های به عمل آمده در فناوری پایگاه‌های داده، نیازی به یک پایگاه داده اختصاصی نیست.
فقط کسانی که درجات تحصیلی بالایی دارند می‌توانند داده کاوی انجام دهند.	با استفاده از ابزارهای جدیدتر مبتنی بر وب، مدیران با هر سطح تحصیلاتی می‌توانند داده کاوی انجام دهند.
داده کاوی فقط برای شرکت‌هایی است که حجم زیادی از داده‌های مشتری دارند.	اگر داده‌ها دقیقاً بیانگر یک سازمان و مشتریانش باشند آنگاه هر شرکتی می‌تواند از داده کاوی استفاده کند.

۳. اگرچه داده کاوی یک فرایند کشف دانش است ولی باید یک هدف/مقصود را برای موفقیت در ذهن داشت. چون ضرب‌المثلی هست که می‌گوید « اگر نمی‌دانید که به کجا می‌روید، هیچوقت به آنجا نخواهید رسید. »
۴. تعریف یک پروژه بر مبنای اساسی که داده‌های شما نتواند از آن پشتیبانی کند. داده کاوی تماماً در مورد داده‌ها است یعنی اینکه بزرگ‌ترین محدودی شما در یک پروژه‌ی داده کاوی با آن مواجه هستید، غنای داده‌ها است. آگاهی از محدودیت‌های داده‌هایتان به شما کمک خواهد کرد تا پروژه‌های امکان‌پذیری را تعریف کنید که انتظارات را برآورد کرده و نتیجه‌بخش باشند.
۵. زمان کافی برای آماده‌سازی داده‌ها تعریف نکنید. آماده‌سازی داده‌ها بیش از آنچه معمولاً فکر می‌کنید زمان نیاز دارد. طبق دانش فعلی ما، تا یک سوم کل زمان پروژه صرف اخذ، درک و آماده‌سازی داده‌ها می‌شود. برای موفقیت در این امر، تا پیش از آنکه داده‌ها را به خوبی پردازش (تجمیع، پاکسازی، تبدیل) نکرده باشید از مدل‌سازی پرهیز کنید.
۶. فقط به نتایج تجمیعی نگاه کنید و رکوردهای فردی را در نظر نگیرید. بهترین حالت داده کاوی وقتی است که داده‌ها در حالت نمایش دانه‌بندی شده باشند. سعی کنید تا از تجمیع غیرضرور خودداری کرده و داده‌ها برای کمک به الگوریتم‌های داده کاوی ساده‌سازی کنید.
۷. در پیگیری و دنبال کردن نتایج و رویه داده کاوی، نامرتب عمل کنید. چونکه داده کاوی یک فرایند کشفی است که شامل تکرارها و آزمایشات بسیاری می‌شود، کاربر به احتمال زیاد قادر به پیگیری یافته‌ها نخواهد بود. موفقیت در این امر نیازمند یک برنامه‌ریزی، اجرا و پیگیری/ثبت منظم و سیستماتیک تمامی وظایف داده کاوی است.
۸. استفاده از داده‌های آینده برای پیش‌بینی آینده. به دلیل فقدان شرح و درک داده‌ها، تحلیلگرها در اغلب مواقع متغیرهایی که ناشناخته هستند را در نظر می‌گیرند. با اینکار، مدل‌های پیش‌بینی آنها نتایج فوق‌العاده دقیقی را تولید می‌کند (پدیده‌ای

که اغلب آن را طلای احمق‌ها می‌نامند). اگر نتایج پیش‌بینی شما بیش از اندازه دقیق باشند، معمولاً در چنین وضعیتی اولین کاری که باید انجام دهید این است که به دنبال یک متغیر اشتباه از آینده باشید.

۹. نادیده گرفتن یافته‌های مشکوک و سریع رد شدن از آنها. یافته‌های غیرمنتظره اغلب نشانگر اکتشافات حقیقی در پروژه‌های داده‌کاوی هستند. تحقیق و بررسی مناسب اینگونه موارد غیرمنتظره می‌تواند منجر به اکتشافات خوشایندی گردد.

۱۰. کارتان را با یک پروژه‌ی پیچیده‌ی معروف آغاز کنید که شما را به یک سوپرستار تبدیل کند. پروژه‌های داده‌کاوی اغلب با شکست مواجه خواهند شد اگر در مورد ابتدا تا انتهای آنها به خوبی فکر نشده باشد. موفقیت اغلب با یک پیشرفت سیستماتیک و منظم از پروژه‌های کوچکتر/ساده‌تر تا بزرگتر/پیچیده‌تر به دست می‌آید. هدف باید این باشد که ارزش افزوده را به صورت تدریجی و پیوسته نشان دهیم به جای اینکه پروژه‌ی بزرگی را آغاز کنیم که بدون اینکه نتایج ارزشمندی را نتیجه دهد منابع بسیار زیادی را صرف کند.

۱۱. اجرای مکرر و کورکورانه‌ی الگوریتم‌های داده‌کاوی. اگرچه ابزارهای امروزی داده‌کاوی قادر به استفاده از داده‌ها و تنظیم پارامترهای الگوریتمی برای تولید نتایج هستند، شخص باید برای دستیابی به بهترین نتایج ممکن از نحوه‌ی تبدیل داده‌ها و تعیین مقادیر مناسب برای پارامترها آگاه باشد. هر الگوریتم از روش خاص خودش برای تولید داده‌ها استفاده می‌کند و آگاهی از این موضوع برای بهره‌برداری حداکثری از هر نوع مدل ضروری است.

۱۲. نادیده گرفتن نظرات افراد خبره در موضوع. درک دامنه‌ی مسئله و داده‌های مربوطه نیازمند یک همکاری بالا بین افراد خبره در داده‌کاوی و دامنه مسئله است. همکاری این دو گروه به خبرگان داده‌کاوی برای فراتر رفتن از بازنمایی نحوی و دستیابی به یک ماهیت معنایی داده‌ها بسیار کمک خواهد کرد (یعنی معنای حقیقی متغیرها).

۱۳. هرآنچه در مورد داده‌ها به شما می‌گویند را باور کنید. اگرچه برای درک بهتر داده‌ها و مسئله‌ی کسب‌وکار بهتر است با افراد خبره در آن حوزه صحبت کنید ولی یک دانشمند علوم داده نباید همه چیز را مسلم بداند. اعتبارسنجی و صحت‌گذاری از طریق یک تحلیل بحرانی، عامل مهمی برای درک و پردازش دقیق و عمیق داده‌ها است.

۱۴. اینطور فرض کنید که کسانی که داده‌ها را در اختیار دارند به طور کامل با شما همکاری خواهند کرد. بسیاری از پروژه‌های داده‌کاوی به دلیل عدم آگاهی/درک افراد خبره‌ی داده‌کاوی از سیاست‌های سازمانی با شکست مواجه می‌شوند. یکی از بزرگ‌ترین موانع در پروژه‌های داده‌کاوی می‌تواند افرادی باشد که داده‌ها را در اختیار داشته و کنترل می‌کنند. درک و مدیریت سیاست‌ها یک عامل کلیدی برای شناسایی، دسترسی و درک صحیح داده‌ها و تولید یک پروژه‌ی داده‌کاوی موفق است.

۱۵. نتایج را به روشی متفاوت از آنچه اسپانسر شما آنها را می‌سنجد اندازه‌گیری کنید. نتایج باید برای کاربر نهایی (مدیر/تصمیم‌گیر) که از آنها استفاده می‌کند جذابیت داشته باشد. از این رو، تولید نتایج در شکل و اندازه‌ای که برای کاربر نهایی جذاب باشد، احتمال درک حقیقی و بکارگیری مناسب از نتایج داده‌کاوی را به طور عظیمی افزایش می‌دهد.

۱۶. به نصیحتی که در ضرب‌المثل معروف «اگر شما آن را بسازید، آنها خواهند آمد»: نگران اینکه چطور آن را سرو کنید نباشید آمده عمل کنید. معمولاً، خبرگان داده‌کاوی فکر می‌کنند که به محض اینکه مدل‌هایی که پاسخگوی انتظارات/نیازها/خواسته‌های کاربر نهایی (یعنی مشتری) باشد را بسازند کار تمام است. بدون جاری‌سازی مناسب، ارزش افزوده‌ای که از داده‌کاوی حاصل خواهد شد محدود است. از این رو، جاری‌سازی، آخرین گام ضروری در یک فرایند داده‌کاوی است که در آن مدل‌ها برای تصمیم‌گیری سریع‌تر و بهتر با زیرساخت پشتیبانی تصمیم سازمانی ادغام می‌شوند.

- \mathbb{R}^n is a vector space over \mathbb{R} with the usual addition and scalar multiplication.
- \mathbb{C}^n is a vector space over \mathbb{C} with the usual addition and scalar multiplication.
- \mathbb{R}^n is also a vector space over \mathbb{C} with the usual addition and scalar multiplication.
- \mathbb{C}^n is also a vector space over \mathbb{R} with the usual addition and scalar multiplication.
- The dimension of \mathbb{R}^n over \mathbb{R} is n .
- The dimension of \mathbb{C}^n over \mathbb{C} is n .
- The dimension of \mathbb{R}^n over \mathbb{C} is $n/2$.
- The dimension of \mathbb{C}^n over \mathbb{R} is $2n$.
- A basis for \mathbb{R}^n over \mathbb{R} is $\{e_1, e_2, \dots, e_n\}$.
- A basis for \mathbb{C}^n over \mathbb{C} is $\{e_1, e_2, \dots, e_n\}$.
- A basis for \mathbb{R}^n over \mathbb{C} is $\{e_1, e_2, \dots, e_{n/2}\}$.
- A basis for \mathbb{C}^n over \mathbb{R} is $\{e_1, e_2, \dots, e_n, ie_1, ie_2, \dots, ie_n\}$.
- The dimension of \mathbb{R}^n over \mathbb{R} is n .
- The dimension of \mathbb{C}^n over \mathbb{C} is n .
- The dimension of \mathbb{R}^n over \mathbb{C} is $n/2$.
- The dimension of \mathbb{C}^n over \mathbb{R} is $2n$.
- A basis for \mathbb{R}^n over \mathbb{R} is $\{e_1, e_2, \dots, e_n\}$.
- A basis for \mathbb{C}^n over \mathbb{C} is $\{e_1, e_2, \dots, e_n\}$.
- A basis for \mathbb{R}^n over \mathbb{C} is $\{e_1, e_2, \dots, e_{n/2}\}$.
- A basis for \mathbb{C}^n over \mathbb{R} is $\{e_1, e_2, \dots, e_n, ie_1, ie_2, \dots, ie_n\}$.

teradatauniversitynetwork.com .1

teradatauniversitynetwork.com .2

CreditRisk.xlsx **Online File W4.1**

[ftp://ics.uc,i.edu/pub/machine-learning-](ftp://ics.uc,i.edu/pub/machine-learning-databases/statlog/german)
[CreditRisk](ftp://ics.uc,i.edu/pub/machine-learning-databases/statlog/german) [databases/statlog/german](ftp://ics.uc,i.edu/pub/machine-learning-databases/statlog/german)

MovieTrain.xlsx **Online File W4.2**

Online

File W4.3, MovieTest.xlsx

baskets1ntrans.xlsx

-
-

Wine, Canned Veg → Frozen Meal

•

.۹

Customer Churn Analysis

k

knime.org

•

•

CustomerChurnData.csv

•

•

•

•

•

•

Customer Churn Analysis

.۱)

.۲)

.۳)

ics.uci.edu/~mlearn/MLRepository.html

Credit Screening Databases, Housing

Databases

[\(data.gov\)](http://data.gov)

[\(www.cdc.gov/DataStatistics\)](http://www.cdc.gov/DataStatistics)

[\(http://seer.cancer.gov/data\)](http://seer.cancer.gov/data)

[\(www.nhtsa.gov/FARS\)](http://www.nhtsa.gov/FARS)

KDnuggets.com/datasets/index.html

	GMAT	GPA	GMAT
	11	111	11
	12	112	11
	11	113	11
	13	114	11
	14	115	11
	15	116	11
	16	117	11
	17	118	11
	18	119	11
	19	120	11
	20	121	11
	21	122	11
	22	123	11
	23	124	11
	24	125	11
	25	126	11
	26	127	11
	27	128	11
	28	129	11
	29	130	11
	30	131	11
	31	132	11
	32	133	11
	33	134	11
	34	135	11
	35	136	11
	36	137	11
	37	138	11
	38	139	11
	39	140	11
	40	141	11
	41	142	11
	42	143	11
	43	144	11
	44	145	11
	45	146	11
	46	147	11
	47	148	11
	48	149	11
	49	150	11
	50	151	11
	51	152	11
	52	153	11
	53	154	11
	54	155	11
	55	156	11
	56	157	11
	57	158	11
	58	159	11
	59	160	11
	60	161	11
	61	162	11
	62	163	11
	63	164	11
	64	165	11
	65	166	11
	66	167	11
	67	168	11
	68	169	11
	69	170	11
	70	171	11
	71	172	11
	72	173	11
	73	174	11
	74	175	11
	75	176	11
	76	177	11
	77	178	11
	78	179	11
	79	180	11
	80	181	11
	81	182	11
	82	183	11
	83	184	11
	84	185	11
	85	186	11
	86	187	11
	87	188	11
	88	189	11
	89	190	11
	90	191	11
	91	192	11
	92	193	11
	93	194	11
	94	195	11
	95	196	11
	96	197	11
	97	198	11
	98	199	11
	99	200	11

Decision Tree	cs.ualberta.ca/~aixplore	AI Exploratorium	.۱
	egain.com	fico.com	.۲
		dmreview.com	.۳
			.۴
		statsoft.com	.۵
		sas.com	.۶
		spss.com	.۷
		teradata.com	.۸
		fico.com	.۹
		salfordsystems.com	.۱۰
		rulequest.com	.۱۱
		KDnuggets.com	.۱۲

تکنیک‌های یادگیری ماشین برای تجزیه و تحلیل پیشگویانه

مدل‌سازی پیشگویانه را شاید بتوان پرکاربردترین شاخه در علم داده‌ها و تجزیه و تحلیل کسب‌وکار دانست که به تصمیم‌گیرندگان اجازه می‌دهد تا از طریق یادگیری از گذشته (یعنی داده‌های تاریخی) بتوانند آینده را تخمین بزنند. در این فصل، ما به مطالعه‌ی ساختارهای داخلی، قابلیت‌ها/محدودیت‌ها و کاربردهای محبوب‌ترین تکنیک‌های مدل‌سازی پیشگویانه از قبیل شبکه‌های عصبی مصنوعی، ماشین‌های برداری پشتیبان، کی نزدیک‌ترین همسایه، یادگیری بایزی و مدل‌های ترکیبی خواهیم پرداخت. اکثر این تکنیک‌ها می‌توانند برای مسائل پیش‌بینی از نوع دسته‌بندی و رگرسیون بکار روند. در اغلب مواقع، از این تکنیک‌ها برای مسائل پیش‌بینی پیچیده که سایر تکنیک‌های

- [۵.۱. پرونده‌ی آغازین: مدل‌سازی پیشگویانه به درک و مدیریت بهتر رویه‌های پزشکی پیچیده کمک می‌کند.](#)
- [۵.۲. مفاهیم پایه‌ی شبکه‌های عصبی](#)
- [۵.۳. معماری شبکه‌های عصبی](#)
- [۵.۴. ماشین‌های برداری پشتیبان](#)
- [۵.۵. رویکرد فرایندمحور در بکارگیری ماشین برداری پشتیبان](#)
- [۵.۶. روش نزدیک‌ترین همسایگی برای پیش‌بینی](#)
- [۵.۷. روش بایز ساده برای دسته‌بندی](#)
- [۵.۸. شبکه‌های بایزی](#)
- [۵.۹. مدل‌سازی ترکیبی](#)

قدیمی نمی‌توانند نتایج رضایت‌بخشی ارائه دهند استفاده می‌شود. علاوه بر تکنیک‌های مطرح شده در این فصل، دیگر تکنیک‌های مدل‌سازی پیش‌بینی مطرح شامل رگرسیون (خطی یا غیرخطی)، رگرسیون لجستیک (برای مسائل پیش‌بینی از نوع دسته‌بندی) و انواع مختلف درخت تصمیم می‌باشد.

۵.۱. پرونده‌ی آغازین: مدل‌سازی پیشگویانه به درک و مدیریت بهتر رویه‌های پزشکی پیچیده کمک می‌کند.

۵.۲. مفاهیم پایه‌ی شبکه‌های عصبی

۵.۳. معماری شبکه‌های عصبی

۵.۴. ماشین‌های برداری پشتیبان

۵.۵. رویکرد فرایندمحور در بکارگیری ماشین برداری پشتیبان

۵.۶. روش نزدیک‌ترین همسایگی برای پیش‌بینی

۵.۷. روش بایز ساده برای دسته‌بندی

۵.۸. شبکه‌های بایزی

۵.۹. مدل‌سازی ترکیبی

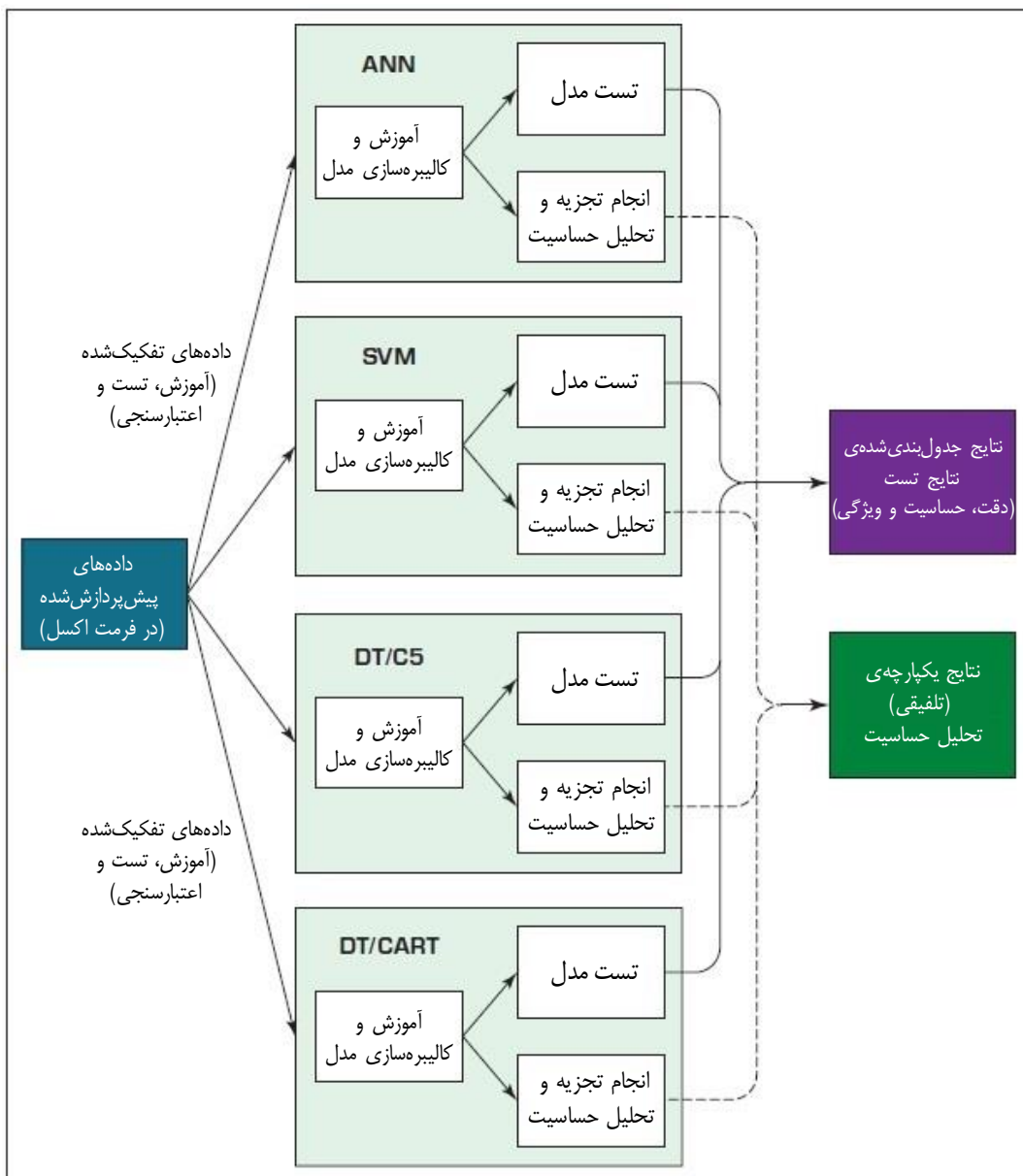
مراقبت سلامت تبدیل به یکی از مهمترین مسائلی شده است که تأثیر مستقیمی بر کیفیت زندگی در ایالات متحده و سرتاسر جهان دارد. اگرچه تقاضا برای دریافت خدمات مراقبت سلامت به دلیل بالا رفتن سن جمعیت در حال افزایش است، عرضه‌ی

اینگونه خدمات در حفظ سطح و کیفیت خدمات دچار مشکلاتی می‌باشد. برای از بین بردن این فاصله، سیستم‌های مراقبت سلامت باید اثربخشی و کارایی عملیاتی خود را به طور چشمگیری بهبود بخشند. اثربخشی (انجام دادن کار درست مثل تشخیص یا درمان دقیق) و کارایی (درست انجام دادن کار مثل استفاده از حداقل منابع یا زمان) دو رکن اساسی هستند که سیستم‌های مراقبت درمان بر آنها بنا نهاد شده‌اند. یک روش امیدبخش برای بهبود مراقبت سلامت، بهره‌گیری از تکنیک‌های مدلسازی پیشگویانه بر روی منابع داده بزرگ و غنی برای پشتیبانی از تصمیم‌گیری به موقع و دقیق است.

طبق آمار انجمن قلب آمریکا، بیماری قلبی-عروقی علت اصلی بیش از ۲۰ درصد مرگ و میرها در ایالات متحده است. از سال ۱۹۰۰، بیماری قلبی-عروقی به استثنای سال ۱۹۱۸ که سال شیوع آنفولانزای اسپانیایی بوده، قاتل شماره یک بشر شناخته شده است. تعداد قربانیان این بیماری به تنهایی بیش از چهار علت اصلی مرگ و میرها می‌باشد: سرطان، بیماری مزمن ریوی، تصادفات و انواع دیابت‌ها. از تمامی مرگ و میرهای ناشی از بیماری قلبی-عروقی، بیش از نیمی از آنها مربوط به بیماری کرونری هستند. بیماری‌های قلبی-عروقی نه تنها آسیب‌های زیادی را به سلامت شخصی و رفاه جامعه وارد می‌کنند بلکه حجم بسیار زیادی از منابع مراقبت سلامت در ایالات متحده و سایر بخش‌های جهان را نیز صرف می‌کند. هزینه‌های مستقیم و غیرمستقیم بیماری‌های قلبی-عروقی در یک سال چیزی بالغ بر ۵۰۰ میلیارد دلار برآورد می‌شود. یکی از رویه‌های جراحی متداول که برای درمان طیف گسترده‌ای از بیماری‌های قلبی-عروقی انجام می‌شود جراحی بای‌پس سرخرگ کرونری^۱ نام دارد. اگرچه هزینه‌ی انجام چنین عملی به عوامل مربوط به ارائه‌دهنده‌ی خدمات و بیمار بستگی دارد ولی به طور میانگین بین ۵،۰۰۰ تا ۱۰،۰۰۰ دلار در ایالات متحده متغیر است. در یک مثال جالب، دیلن و همکاران (۲۰۱۲) با استفاده از روش‌های مختلف مدلسازی پیشگویانه، یک مطالعه تحلیلی را به منظور پیش‌بینی نتیجه‌ی یک عمل جراحی بای‌پس سرخرگ کرونری انجام داده و برای درک بهتر اهمیت عوامل پیشگوی بیماری، از تحلیل حساسیت مبتنی بر تلفیق اطلاعات بر روی مدل‌های آموزش دیده استفاده کرده‌اند. هدف اصلی از این مطالعه این بوده است که نشان دهند تجزیه و تحلیل توصیفی و پیشگویانه‌ی مجموعه داده‌های بزرگ و غنی، اطلاعات ارزشمندی را برای تصمیم‌گیری کارا و اثربخش در مراقبت سلامت فراهم می‌کند.

شکل ۵.۱ یک فرایند توسعه و تست مدل را که دیلن و همکاران (۲۰۱۲) بکار برده‌اند نمایش می‌دهد. آنها از چهار نوع مدل پیش‌بینی متفاوت (شبکه‌های عصبی مصنوعی، ماشین‌های برداری پشتیبان و دو نوع درخت تصمیم C5 و CART) استفاده کرده و برای کالیبره کردن پارامترهای مدل‌سازی برای هر نوع مدل دفعات بسیاری اجرای آزمایشی داشته‌اند. پس از توسعه‌ی مدل‌ها، پژوهشگران مجموعه داده‌ها را تست کرده و نهایتاً، مدل‌های آموزش دیده تحت تحلیل حساسیت قرار گرفتند که به اندازه‌گیری سهم و نقش متغیرها پرداختند. جدول ۵.۱ نتایج تست را برای چهار نوع مدل پیش‌بینی مختلف نشان می‌دهد.

¹ Coronary Artery Bypass Grafting (CABG)



شکل ۵.۱ نقشه فرایندی آموزش و تست چهار مدل پیشگویانه

جدول ۵.۱ نتایج دقت پیش‌بینی هر چهار نوع مدل بر اساس مجموعه داده‌های تست

ماتریس‌های درهم‌ریختگی				نوع مدل
ویژگی	حساسیت	دقت	مثبت (۱) / منفی (۰)	

شبکه عصبی مصنوعی	مثبت (۱)	۷۴۹	۲۳۰	۷۴.۷۲٪	۷۶.۵۱٪	۷۲.۹۳٪
	منفی (۰)	۲۶۵	۷۱۴			
ماشین برداری پشتیبان	مثبت (۱)	۸۷۶	۱۰۳	۸۷.۷۴٪	۸۹.۴۸٪	۸۶.۰۱٪
	منفی (۰)	۱۳۷	۸۴۲			
C5	مثبت (۱)	۸۷۶	۱۰۳	۷۹.۶۲٪	۸۰.۲۹٪	۷۸.۹۶٪
	منفی (۰)	۱۳۷	۸۴۲			
CART	مثبت (۱)	۶۶۰	۳۱۹	۷۱.۱۵٪	۶۷.۴۲٪	۷۴.۸۷٪
	منفی (۰)	۲۴۶	۷۳۳			

در این مطالعه، دیلن و همکاران (۲۰۱۲) توان داده‌کاوی در پیش‌بینی نتایج و تحلیل عوامل پیشگویانه برای رویه‌های پزشکی پیچیده از قبیل جراحی بای‌پس سرخرگ کرونری را نشان دادند. پژوهشگران نشان دادند که استفاده از چندین روش پیش‌بینی (برخلاف یک روش) در یک محیط آزمایشی رقابتی، می‌تواند نتایج پیشگویانه و توضیحی بهتری را تولید کند. از بین چهار مدل بکار رفته، روش ماشین‌های برداری پشتیبان بهترین نتایج را با دقت ۸۸ درصد بر روی نمونه داده‌های تست ارائه کرده‌اند. از طریق تحلیل حساسیت تلفیقی اطلاعات، رتبه‌بندی اهمیت متغیرهای وابسته مشخص شد. برخی از متغیرهای شناسایی شده در این تحلیل با مهمترین متغیرهای شناسایی شده در مطالعات بالینی و بیولوژیکی که در مطالعات قبلی مؤید صحت و اثربخشی متدولوژی داده‌کاوی پیشنهادی بودند.

از نقطه نظر مدیریتی، سیستم‌های پشتیبان تصمیم‌های بالینی که از نتایج مطالعات داده‌کاوی استفاده می‌کنند جایگزین مدیران و متخصصان پزشکی نخواهند شد بلکه هدف از این سیستم‌ها این است که از آنها در تصمیم‌گیری دقیق و به موقع و تخصیص بهینه‌ی منابع به منظور افزایش کیفیت و کمیت خدمات پزشکی پشتیبانی کنند. هنوز راه زیادی مانده تا شاهد کاربرد گسترده‌ی این کمک‌ها در صنعت مراقبت سلامت باشیم. مهمترین علل مقاومت در برابر بکارگیری این سیستم‌ها، علل سیاسی، رفتاری و اخلاقی می‌باشند. شاید نیازها و مشوق‌های دولتی بتوانند موجب بکارگیری سریع‌تر این سیستم‌ها شوند.

سوالاتی برای مرور پرونده‌ی آغازین

۱۸. چرا مطالعه‌ی رویه‌های پزشکی اهمیت دارد؟ ارزش حاصل از پیش‌بینی نتایج آن چیست؟
۱۹. به نظر شما مهمترین عوامل برای درک و مدیریت بهتر مراقبت سلامت چه هستند؟ هر دو جنبه‌ی مدیریتی و بالینی مراقبت سلامت را در نظر بگیرید.
۲۰. تأثیر مدل‌سازی پیشگویانه بر مراقبت سلامت و پزشکی چیست؟ آیا مدل‌سازی پیشگویانه می‌تواند جایگزین پرسنل پزشکی یا مدیریتی شود؟
۲۱. نتایج این مطالعه چه بودند؟ چه کسانی می‌توانند این نتایج را ببینند؟ چگونه می‌توان نتایج را پیاده‌سازی کرد؟
۲۲. در اینترنت جستجو کرده و دو مطالعه موردی دیگر را که از مدل‌سازی پیشگویانه برای درک و مدیریت رویه‌های پزشکی پیچیده استفاده کرده‌اند شناسایی کنید.

همانطور که در این فصل خواهید دید، از تکنیک‌های مدل‌سازی پیشگویانه می‌توان برای طیف گسترده‌ای مسائل از کسب‌وکار معمول گرفته تا ارزیابی نیازهای مشتری برای درک و افزایش کارایی فرایندهای تولید و بهبود مراقبت سلامت و پزشکی استفاده کرد. این پرونده‌ی آغازین یک کاربرد نوآورانه از مدل‌سازی پیشگویانه را برای پیش‌بینی، درک و مدیریت بهتر رویه‌های جراحی بای‌پس سرخرگ کرونری نشان داده است. همانطور که نتایج نشان می‌دهند، این تکنیک‌های تجزیه و تحلیلی پیچیده قادر به پیش‌بینی و توضیح اینگونه پدیده‌های پیچیده هستند. پزشکی مبتنی بر شواهد عبارت نسبتاً جدیدی است که در حوزه‌ی مراقبت سلامت بکار رفته و هدف اصلی آن کاوش در تجارب گذشته به منظور کشف دانش جدید و مفید به منظور بهبود رویه‌های پزشکی و مدیریتی در صنعت مراقبت سلامت است. همانطور که می‌دانیم، مراقبت سلامت نیازمند همه‌گونه مساعدت است. در مقایسه با پژوهش‌های رایج که ذاتاً بالینی و بیولوژیکی هستند، مطالعات مبتنی بر داده‌ها یک دیدگاه کل‌نگر بر سیستم‌های پزشکی و مدیریتی دارند.

Sources: D. Delen, A. Oztekin, and L. Tomak, "An Analytic Approach to Better Understanding and Management of Coronary Surgeries," *Decision Support Systems*, Vol. 52, No. 3, 2012, pp. 698–705; and American Heart Association, "Heart Disease and Stroke Statistics," heart.org (accessed May 2018).

شبکه‌های عصبی بیانگر یک استعاره‌ی ذهنی برای پردازش اطلاعات هستند. این مدل‌ها به جای اینکه کپی دقیقی از نحوه عملکرد حقیقی مغز باشند بیشتر از منظر بیولوژیکی الهام گرفته‌اند. شبکه‌های عصبی به دلیل توانایی‌شان در «یادگیری» از روی داده‌ها، ماهیت غیرپارامتری (یعنی بدون مفروضات سفت و سخت) و توانایی تعمیم دادنشان در بسیاری از کاربردهای پیش‌بینی و دسته‌بندی کسب‌وکار بسیار موفق عمل کرده‌اند. **رایانش عصبی**^۱ به یک متدولوژی تشخیص الگو برای یادگیری ماشین اشاره دارد. مدل بدست آمده از رایانش عصبی را اغلب یک **شبکه عصبی مصنوعی**^۲ یا یک **شبکه عصبی**^۳ می‌نامند. شبکه‌های عصبی در بسیاری از کسب‌وکارها برای **تشخیص الگو**، پیش‌بینی، پیشگویی و دسته‌بندی بکار رفته‌اند. رایانش شبکه عصبی یکی از مؤلفه‌های اصلی هر جعبه ابزار تجزیه و تحلیل کسب‌وکار و علم داده‌ها است. کاربردهای شبکه‌های عصبی شامل امور مالی، بازاریابی، ساخت و تولید، عملیات، سیستم‌های اطلاعاتی و غیره است.

از آنجائیکه ما در فصل ۶ (که به موضوع یادگیری عمیق و رایانش شناختی اختصاص دارد) به طور خاص به بررسی شبکه‌های عصبی و به ویژه مدل‌سازی پیش‌بینی چندلایه ادراکی بیش‌خو خواهیم پرداخت در این بخش فقط به معرفی مختصر دنیای وسیع مدل‌ها، روش‌ها و کاربردهای شبکه‌های عصبی خواهیم پرداخت.

مغز انسان دارای قابلیت‌های خارق‌العاده‌ای برای پردازش اطلاعات و حل مسئله است که رایانه‌های پیشرفته و امروزی از بسیاری از جنبه‌ها قادر به رقابت با آن نیستند. فرض بر این است که یک مدل یا یک سیستم که به واسطه‌ی نتایج پژوهش‌های مغز انسان حمایت می‌شود و ساختاری مشابه شبکه‌های عصبی بیولوژیکی دارد می‌تواند کارکردهای هوشمندانه‌ی مشابهی را از خود نشان دهد. بر اساس این رویکرد پایین به بالا، شبکه‌های عصبی مصنوعی (که با نام‌های **مدل‌های ارتباط‌گرا**^۴، **مدل‌های پردازش**

¹ Neural Computing

² Artificial Neural Networks (ANN)

³ Neural Network

⁴ Connectionist models

توزیع یافته موازی^۱، سیستم‌های نورومورفیک^۲ یا به بیان ساده شبکه‌های عصبی نیز شناخته شده‌اند) به صورت مدل‌هایی که از ساختار بیولوژیکی الهام گرفته‌اند برای وظایف مختلف توسعه یافته‌اند.

شبکه‌های عصبی بیولوژیکی از تعداد بسیار زیادی نورون‌های متصل بهم تشکیل شده‌اند. هر نورون دارای زائده‌های انگشت‌مانندی به نام آسکون‌ها و دندریت‌ها است که به نورون امکان می‌دهد تا از طریق ارسال و دریافت سیگنال‌های الکتریکی و شیمیایی با نورون‌های مجاور خود ارتباط برقرار کند. ساختار شبکه‌های عصبی مصنوعی که کم و بیش مشابه هم‌تای بیولوژیکی خود است از اجزای بهم متصل و پرازنده‌ی ساده‌ای به نام نورون‌های مصنوعی^۳ تشکیل شده است. در زمان پردازش اطلاعات، اجزای پردازنده در شبکه‌های عصبی مصنوعی همانند نورون‌های بیولوژیکی به طور همزمان و باهم عمل می‌کنند. شبکه‌های عصبی مصنوعی دارای برخی ویژگی‌های مطلوب همانند شبکه‌های عصبی بیولوژیکی هستند مثل قابلیت‌های یادگیری، خودسازماندهی و تحمل خرابی.

پژوهشگران برای بیش از نیم قرن است که در مورد شبکه‌های عصبی مصنوعی به تحقیق و بررسی پرداخته‌اند. مطالعه‌ی رسمی شبکه‌های عصبی مصنوعی در سال ۱۹۴۳ و توسط مک‌کالاک و پیتز^۴ آغاز شد. مک‌کالاک و پیتز (۱۹۴۳) که از نتایج مشاهدات و آزمایشات بیولوژیکی الهام گرفته بودند، مدل ساده‌ای از یک نورون مصنوعی دودویی را معرفی کردند که برخی از کارکردهای نورون‌های بیولوژیکی را ضبط می‌نمود. مک‌کالاک و پیتز با استفاده از ماشین‌های اطلاعات/پردازنده برای مدل سازی مغط، مدل شبکه عصبی خود را با استفاده از تعداد زیادی از نورون‌های دودویی مصنوعی بهم متصل ساختند. پژوهش در حوزه شبکه‌های عصبی از اواخر دهه ۵۰ و اوایل ۶۰ بسیار متداول شد. توجه به شبکه‌های عصبی پس از تحلیل عمیق یک مدل شبکه عصبی اولیه (به نام پرسپترون که فاقد لایه‌های پنهان بود) و همچنین ارزیابی بدبینانه‌ی پژوهش‌ها توسط مینسکی و پاپرت در سال ۱۹۶۹ کاهش یافت.

طی دو دهه‌ی گذشته، به دلیل پیدایش توپولوژی‌های جدید شبکه‌ها، کارکردهای فعال‌سازی جدید و الگوریتم‌های یادگیری جدید و همچنین پیشرفته‌های به دست آمده در علوم شناختی و عصب‌شناختی، مطالعات شبکه‌های عصبی مصنوعی با یک تجدید حیات شگفت‌آوری روبرو بوده است. پیشرفت‌های به عمل آمده در حوزه‌های نظری و متدولوژی بر بسیاری از موانع که تا چند سال گذشته سد راه پژوهش‌های شبکه‌های عصبی بوده‌اند فائق آمده‌اند. با توجه به نتایج جذاب مطالعات متعدد در این زمینه، شبکه‌های عصبی از پذیرش و مقبولیت روزافزونی برخوردار شده‌اند. علاوه بر این، ویژگی‌های مطلوب پردازش اطلاعات صعبی بر جذابیت شبکه‌های عصبی برای حل مسائل پیچیده افزوده است. شبکه‌های عصبی مصنوعی در مسائل پیچیده متعدد و در حوزه‌های کاربردهای مختلف بکار رفته‌اند. کاربردهای موفق شبکه‌های عصبی موجب شده تا بکارگیری شبکه‌های عصبی در صنعت و کسب‌وکار مجدداً مورد توجه قرار گیرد. با ظهور شبکه‌های عصبی عمیق (به عنوان بخشی از پدیده‌ی نسبتاً نوین یادگیری عمیق)، معروفیت شبکه‌های عصبی (با یک باز نمود معماری «عمیق تر» و قابلیت‌های تجزیه و تحلیل ارتقاء یافته‌تر) به سطح پیش‌بینی نشده‌ای رسیده است و سطح انتظارات را از این نسل جدید شبکه‌های عصبی به شدت بالا برده است. شبکه‌های عصبی عمیق به تفصیل در فصل ۶ توضیح داده شده‌اند.

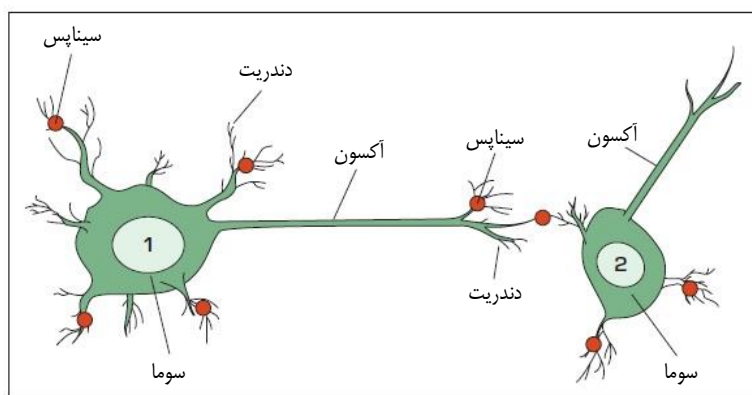
¹ Parallel distributed processing models

² Neuromorphic systems

³ Artificial Neurons

⁴ McCulloch & Pitts

مغز انسان از سلول‌های خاصی به نام نورون‌ها تشکیل شده است. این سلول‌ها زمانیکه شخص دچار آسیب می‌شود از بین نمی‌روند و خودشان را مجدداً احیا می‌کنند (بقیه‌ی سلول‌ها پس از تولید مثل و جایگزین کردن خودشان، از بین می‌روند). این پدیده ممکن است بتواند علت حفظ و نگهداشت اطلاعات برای یک دوره زمانی طولانی در مغز انسان و شروع به فراموشی آن با کهولت سن را توضیح دهد. ذخیره‌سازی اطلاعات، مجموعه بسیار زیادی از سلول‌های عصبی را شامل می‌شود. در هر بخشی از مغز بین ۵۰ تا ۱۵۰ میلیارد شبکه عصبی از بیش از ۱۰۰ نوع مختلف وجود دارد. سلول‌های عصبی به گروه‌هایی به نام شبکه تقسیم‌بندی شده‌اند. هر شبکه حاوی چندین هزار سلول عصبی بهم متصل است. بدین ترتیب، مغز را می‌توان به صورت مجموعه‌ای از شبکه‌های عصبی در نظر گرفت. توانایی یادگیری و واکنش نشان دادن به تغییرات محیطی نیازمند هوش است. مغز و سیستم عصبی مرکزی، تفکر و رفتار هوشمندانه را کنترل می‌کنند. افرادی که دچار آسیب مغزی شده‌اند در یادگیری و واکنش نشان دادن به تغییرات محیطی مشکل دارند. اگرچه، بخش‌های سالم و آسیب ندیده‌ی مغز می‌توانند از طریق یادگیری جدید این موضوع را جبران کنند.



شکل ۵.۲ بخشی از یک شبکه‌ی عصبی بیولوژیکی: دو سلول/عصب بهم متصل

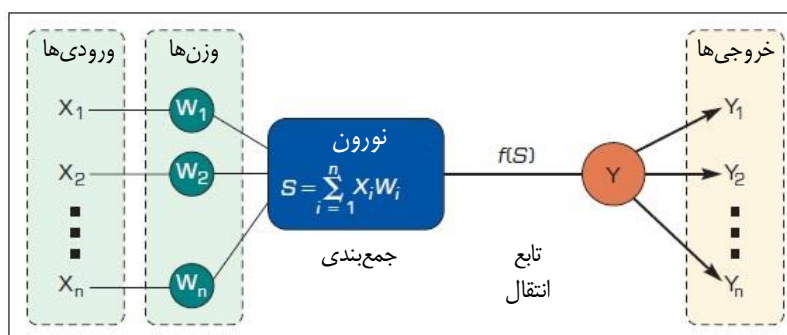
بخشی از این شبکه که از دو سلول تشکیل شده در شکل ۵.۲ نشان داده شده است. خود سلول از یک هسته (بخش پردازنده‌ی مرکزی عصب) تشکیل شده است. در سمت چپ سلول ۱، دندریت‌ها قرار دارند که سیگنال‌های ورودی را به سلول هدایت می‌کنند. در سمت راست، آکسون‌ها قرار دارند که سیگنال‌های خروجی را از طریق ترمینال‌های آکسون به سلول شماره ۲ منتقل می‌کنند. این ترمینال‌های آکسون‌ها با دندریت‌های سلول ۲ ادغام می‌شوند. سیناپس‌ها سیگنال‌ها را به صورت دست نخورده یا تغییر یافته می‌توانند انتقال دهند. سیناپس همچنین قادر است تا قدرت و استحکام بین عصب‌ها را افزایش یا کاهش داده و موجب تحریک یا مهار عصب بعدی شود. اطلاعات به این شکل در شبکه‌های عصبی ذخیره می‌شود.

یک شبکه عصبی مصنوعی از یک شبکه عصبی بیولوژیکی تقلید می‌کند. رایانش عصبی در واقع از مجموعه بسیار محدودی از مفاهیم سیستم‌های عصبی بیولوژیکی استفاده می‌کند (بینش فناوری ۵.۱). شبکه عصبی مصنوعی بیشتر شبیه مغز انسان است تا اینکه مدل دقیقی از آن باشد. مفاهیم عصبی معمولاً به صورت شبیه‌سازی‌های نرم‌افزاری انبوهی از فرایندهای موازی که مشغول پردازش عناصر بهم متصل (که نورون‌های مصنوعی یا نورود^۱ نیز نامیده می‌شوند) در یک معماری شبکه هستند پیاده‌سازی و

¹ Neurode

اجرا می‌شوند. نورون‌های مصنوعی، ورودی‌هایی مشابه تکانه‌های الکتروشیمیایی را که دندریت‌های نورون‌های بیولوژیکی از سایر نورون‌ها دریافت می‌کنند دریافت می‌کنند. خروجی نورون‌های مصنوعی متناظر با سیگنال‌های ارسالی از یک نورون بیولوژیکی از طریق آکسون است. وزن این سیگنال‌های مصنوعی را می‌توان به روشی مشابه تغییرات فیزیکی که در سیناپس‌ها رخ می‌دهد تغییر داد (شکل ۵.۳).

چندین پارادایم شبکه‌های عصبی مصنوعی برای کاربرد در حوزه‌های مختلف پیشنهاد شده است. شاید آسان‌ترین روش برای متمایز کردن مدل‌های مختلف عصبی، نحوه‌ی تقلید آنها از ساختار مغز انسان، پردازش اطلاعات و یادگیری برای انجام وظایف محوله باشد.



شکل ۵.۳ پردازش اطلاعات در یک نورون مصنوعی

بیش فناوری ۰.۱ ارتباط بین شبکه‌های عصبی مصنوعی و بیولوژیکی

در فهرست زیر برخی از ارتباطات بین شبکه‌های عصبی مصنوعی و بیولوژیکی ارائه شده است.

مصنوعی	بیولوژیکی
گره	سوما
ورودی	دندریت
خروجی	آکسون
وزن	سیناپس
سریع	آهسته
تعداد نورون‌های بسیار کم (ده‌ها تا صدها هزار)	تعداد نورون‌های بسیار زیاد (10^9)

اجزای پردازنده‌ی اصلی یک شبکه‌ی عصبی که همان نورون‌ها هستند مشابه نورون‌های مغز انسان می‌باشد و از آنها الهام گرفته است. این نورون‌های مصنوعی، اطلاعات را از سایر نورون‌ها یا محرک‌های ورودی بیرونی دریافت کرده، ورودی‌ها را تغییر داده و سپس آنها را به سایر نورون‌ها یا خروج‌های بیرونی انتقال می‌دهند. این بسیار مشابه کاری است که در حال حاضر مغز انسان انجام می‌دهد. انتقال اطلاعات از یک نورون به نورون دیگر را می‌توان روشی برای فعال‌سازی یا تحریک یا یک پاسخ از نورون‌های خاص بر اساس اطلاعات یا محرک‌های دریافتی در نظر گرفت.

اینکه اطلاعات چگونه توسط یک شبکه عصبی پردازش می‌شود ذاتاً به ساختار آن بستگی دارد. شبکه‌های عصبی می‌توانند یک یا چند لایه نورون داشته باشند. ممکن است ارتباطات بسیار زیادی بین این نورون‌ها وجود داشته باشد یا اینکه فقط لایه‌های خاصی به یکدیگر متصل باشند. ارتباطات بین نورون‌ها دارای وزن است. در واقع، « دانش » موجود در شبکه در این وزن‌های بهم متصل به صورت فشرده قرار گرفته است. هر نورون، مجموع وزنی مقادیر نورون ورودی را محاسبه کرده، این ورودی را تبدیل نموده و مقدار عددی آن را به عنوان ورودی به نورون‌های بعدی منتقل می‌کند. معمولاً، ولی نه همیشه این فرایند تبدیل ورودی به خروجی در سطح نورون‌های فردی به شکل غیرخطی انجام می‌شود.

شبکه‌های عصبی به نجات جان انسان‌ها در صنعت معدنکاری کمک می‌کنند

مورد کاربردی ۵.۱

در صنعت معدن کاری، اکثر آسیب‌ها و تلفات به دلیل سقوط سنگ می‌باشد (یعنی افتادن سقف/دیواره). روشی که سالیان سال برای تعیین انسجام دیواره‌ها در معادن استفاده می‌شود این است که با یک میله به دیواره کوبیده و به صدای منتشر شده از آن گوش می‌دهند. یک معدن کار با تجربه می‌تواند بر اساس صدای منتشر شده تشخیص دهد که آیا دیواره محکم/توپر است یا خیر. این روش ذهنی و فردی است. شورای پژوهش صنعتی و علمی آفریقای جنوبی^۱ دستگاهی را ابداع کرده است که به معدن کارها در این خصوص کمک می‌کند. در این دستگاه از یک مدل شبکه عصبی آموزش دیده استفاده شده است. این دستگاه زمانیکه به دیواره کوبیده می‌شود، صدای منتشر شده را ضبط می‌کند. سپس پیش از آنکه وارد مدل شبکه عصبی آموزش دیده شود پیش پردازش می‌شود و بر آن اساس تعیین می‌کند که آیا دیواره قابل اطمینان است یا خیر. تبوهو نیارلی^۲ دارای مدرک کارشناسی ارشد در مهندسی الکترونیک از دانشگاه کیپ‌تاون آفریقای جنوبی که به عنوان یک مهندس پژوهشگر در شورای پژوهش صنعتی و علمی آفریقای جنوبی فعالیت می‌کند از یک نرم‌افزار محبوب مدل‌سازی شبکه‌های عصبی مصنوعی به نام نیورو سولوشنز^۳ متعلق به شرکت نیورو دایمنشنز^۴ برای توسعه‌ی مدل‌های پیش‌بینی از نوع دسته‌بندی استفاده کرد. معماری چندلایه‌ی نوع پرسپترون شبکه عصبی مصنوعی که وی ساخت دارای دقت پیش‌بینی بیش از ۷۰ درصد بود. در سال ۲۰۱۸، سیستم نمونه پیش از آنکه به عنوان یک کمک برای تصمیم‌گیری مورد استفاده قرار بگیرد تحت مجموعه آزمون‌های نهایی قرار گرفت و وارد فاز تجاری‌سازی شد. شکل زیر یک اسنپ‌شات از فضای کاری مدل نیورو سولوشنز را به نام بردبورد نشان می‌دهد.

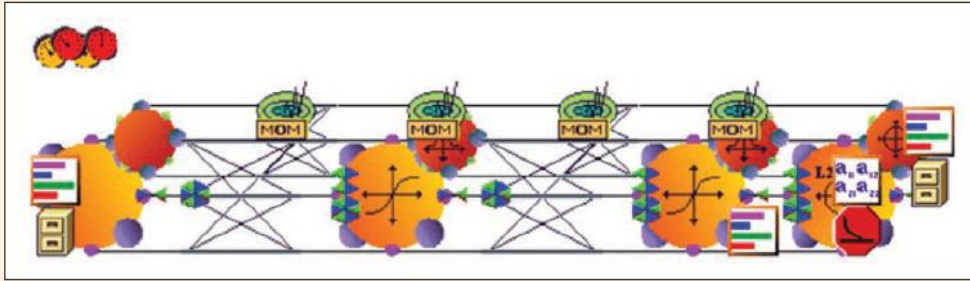
Source: Used with permission from NeuroSolutions, customer success story, neurosolutions.com/resources/nyareli.html (accessed May 2018).

¹ Council for Scientific and Industrial Research (CSIR)

² Teboho Nyareli

³ NeuroSolutions

⁴ NeuroDimensions



سوالاتی برای مورد کاربردی ۵.۱

۱۷. شبکه‌های عصبی چگونه به نجات جان انسان‌ها در صنعت معدن کاری کمک می‌کنند؟
۱۸. چالش‌ها، راه‌حل پیشنهادی و نتایج چه بودند؟

سوالاتی برای مرور بخش ۵.۲

۷. شبکه عصبی مصنوعی چیست؟
۸. وجوه اشتراک و افتراق بین شبکه‌های عصبی بیولوژیکی و مصنوعی چه هستند؟
۹. چه نوع مسائل کسب‌وکاری را می‌توان با استفاده از شبکه‌های عصبی مصنوعی حل کرد؟

چندین معماری شبکه عصبی برای حل انواع مختلف مسائل طراحی شده است (هیکین ۲۰۰۹). متداول‌ترین آنها شامل مدل پیش‌بازخورد (پرسپترون چندلایه با بازگشت به عقب)^۱، حافظه‌ی انجمنی (تداعی)^۲، شبکه‌های بازگشتی^۳، نقشه‌های خودسازمان‌دهنده کوهونن^۴ و شبکه‌های هاپفیلد^۵ هستند. معماری شبکه پیش‌بازخورد پرسپترون چندلایه، نورون‌ها را در یک جهت فعال می‌کند (و ارتباط بین متغیرهای ورودی و متغیر خروجی را یاد می‌گیرد) (از لایه ورودی به لایه خروجی و از بین یک لایه به چندین لایه‌ی پنهان می‌رود). در فصل ۶ به تفصیل به بررسی این معماری شبکه عصبی خواهیم پرداخت. بر خلاف معماری شبکه عصبی پیش‌بازخورد، شکل ۵.۴ یک نمایش تصویری از معماری شبکه عصبی بازگشتی را نشان می‌دهد که ارتباطات بین لایه‌ها تک‌سویه نیستند بلکه ارتباطات بسیاری در هر جهت بین لایه‌ها و نورون‌ها وجود دارد که یک ساختار ارتباطی پیچیده را تشکیل می‌دهند. بسیاری از خبرگان معتقدند که این ارتباطات چندسویه تقلید بهتری از ساختار نورون‌های بیولوژیکی در مغز انسان می‌باشد.

¹ Feedforward multilayer perceptron with backpropagation

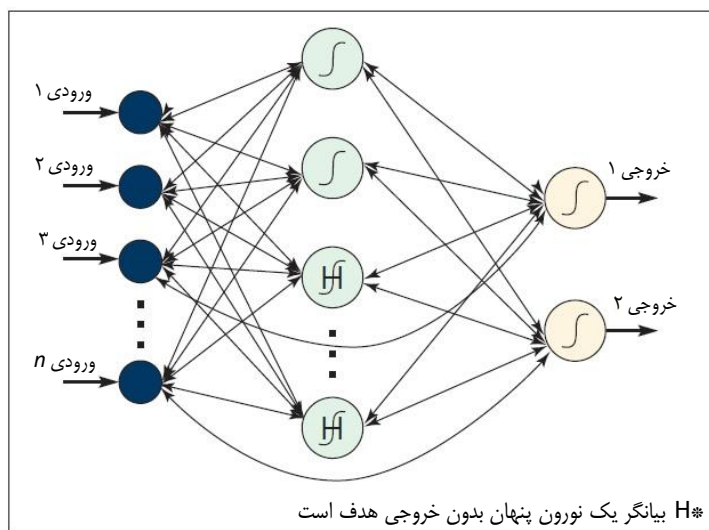
² Associative memory

³ Recurrent networks

⁴ Kohonen's self-organizing feature maps

⁵ Hopfield networks

نقشه‌های خودسازمان‌دهنده کوهونن (شبکه‌های کوهونن یا به اختصار اس‌اِم) که اولین بار توسط تئوو کوهونن^۱ استاد فنلاندی معرفی شدند روشی را برای نمایش داده‌های چندبُعدی در فضاهای معمولاً یک یا دو بُعدی ارائه می‌کند.



شکل ۵.۴ معماری شبکه عصبی بازگشتی

یکی از جالب‌ترین جنبه‌های اس‌اِم این است که می‌توانند یاد بگیرند تا داده‌ها را بدون نظارت دسته‌بندی کنند (یعنی هیچ بردار خروجی وجود ندارد). به خاطر داشته باشید که در تکنیک‌های یادگیری با نظارت مثل پس‌انتشار، داده‌های آموزش شامل زوج برداری - یک بردار ورودی و یک بردار هدف هستند. اس‌اِم به خاطر برخورداری از قابلیت خودسازماندهی عمدتاً برای خوشه‌بندی‌ها استفاده می‌شوند که یک گروه از اقلام به تعدادی تصادفی از گروه‌های طبیعی اختصاص می‌یابند. شکل ۵.۵a یک شبکه‌های کوهوننی بسیار کوچک از $4 * 4$ گره متصل به لایه ورودی را نشان می‌دهد (با سه ورودی) که بیانگر یک بردار دوبُعدی می‌باشد.

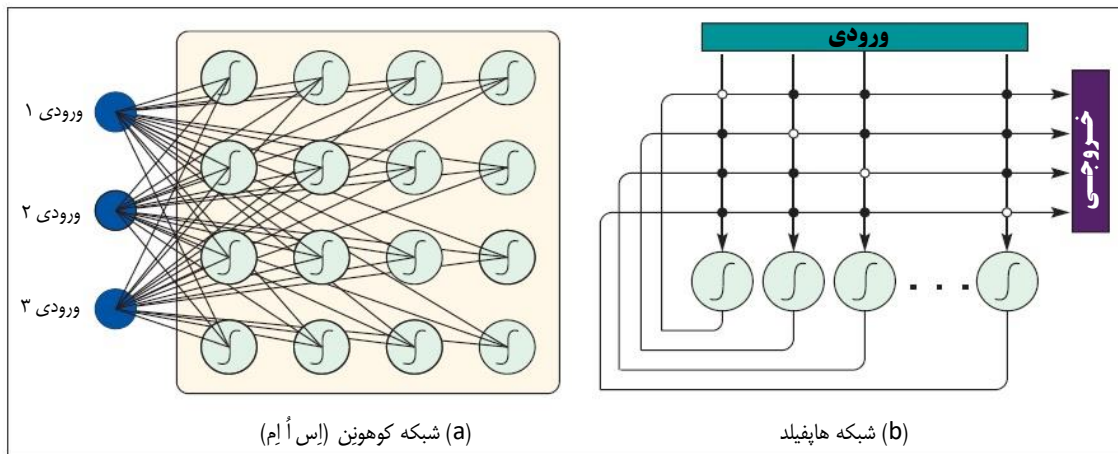
شبکه هاپفیلد یک معماری شبکه عصبی جالب دیگر است که اولین بار توسط جان هاپفیلد در سال ۱۹۸۲ معرفی شد. هاپفیلد طی مجموعه‌ای از مقالات پژوهشی در اوایل دهه ۸۰ نشان داد که چگونه شبکه‌های نورون‌های غیرخطی بهم متصل در حل مسائل محاسباتی پیچیده می‌توانند مؤثر باشند. این شبکه‌ها راه‌حل‌های نوین و سریعی را برای گروهی از مسائل دارای یک هدف مطلوب و چندین محدودیت هستند ارائه می‌کنند (یعنی مسائل بهینه‌سازی محدودیت). یکی از مزیت‌های اصلی شبکه‌های عصبی هاپفیلد این است که ساختار آنها را می‌توان بر روی یک برد مدار الکترونیک و حتی یک مدل یکپارچه‌سازی مقیاس بسیار بزرگ^۲ ایجاد کرد که به عنوان یک حل‌کننده آنلاین با یک پردازشگر موازی-توزیع یافته^۳ مورد استفاده قرار گیرد. از لحاظ

¹ Teuvo Kohonen

² Very Large-Scale Integration (VLSI)

³ Parallel-Distributed Process

معماری، یک شبکه‌ی هاپفیلد عام به صورت یک لایه‌ی بزرگ از نورون‌ها که تمامی آنها بهم متصل هستند نمایش داده می‌شود یعنی هر نورون به تمامی نورون‌های دیگری که در شبکه قرار دارند متصل است (شکل ۵.۵b).



شکل ۵.۵ نمایش گرافیکی ساختارهای شبکه‌های عصبی مصنوعی کوهون و هاپفیلد

نهایتاً، معماری یک مدل شبکه عصبی تحت تأثیر وظیفه و کارکردی است که قرار است انجام دهد. به عنوان مثال، مدل‌های شبکه عصبی به عنوان طبقه‌بندها، ابزارهای پیش‌بینی، مکانیزم‌های بخش‌بندی مشتریان و بهینه‌سازهای عمومی مورد استفاده قرار گرفته‌اند. همانطور که در ادامه این فصل خواهیم دید، طبقه‌بندهای شبکه‌های عصبی معمولاً مدل‌هایی چندلایه هستند که اطلاعات در آنها از یک لایه به لایه بعدی حرکت می‌کند و هدف نهایی از اینکار، نگاشت یک ورودی به شبکه در یک گروه یا دسته‌ی خاص است که توسط یک خروجی شبکه شناسانده می‌شود. یک مدل عصبی که به عنوان یک بهینه‌ساز بکار می‌رود می‌تواند یک لایه متشکل از نورون‌ها باشد که کاملاً به یکدیگر متصل هستند و می‌تواند مقادیر نورون را تا زمانیکه مدل به وضعیت پایدار برسد مکرراً محاسبه کند. این وضعیت پایدار بیانگر جواب بهینه برای مسئله است. مورد کاربردی ۵.۲ خلاصه‌ای از مدل‌سازی پیشگویانه (مثلاً شبکه‌های عصبی) را در رسیدگی به مسائل نوظهور در صنعت برق ارائه کرده است.

مدل‌سازی پیشگویانه در توانمندی‌های ژنراتورهای برق

مورد کاربردی ۵.۲

صنعت برق به تولید و توزیع انرژی الکتریکی (الکتریسیته یا برق) هم به شهروندان و هم کسب‌وکارها در هر کجا و در هر زمان می‌پردازد. الکتریسیته را می‌توان از تعدادی کثیری از منابع تولید کرد. اغلب مواقع، الکتریسیته در یک نیروگاه برق با استفاده از ژنراتورهای الکترومکانیکی که توسط موتورهای حرارتی با سوخت احتراق شیمیایی (سوزاندن زغال‌سنگ، نفت یا گاز طبیعی) یا هسته‌ای (راکتور هسته‌ای) کار می‌کنند تولید می‌شود. الکتریسیته را به روش‌های دیگری از قبیل انرژی جنبشی (از طریق سقوط/جریان آب یا باد که توربین‌ها را به حرکت در می‌آورد)، انرژی خورشیدی (از طریق انرژی متساطع شده از خورشید به

صورت نور یا گرما) یا انرژی حرارت مرکزی زمین (از طریق جریان آب داغ یا بخار که از لایه‌های عمیق زمین می‌آید) می‌توان تولید کرد. انرژی الکتریسیته پس از تولید از طریق زیرساخت‌های شبکه برق توزیع می‌شود.

اگرچه برخی روش‌های تولید انرژی به روش‌های دیگر ارجحیت دارند ولی تمامی شکل‌های تولید الکتریسیته دارای جنبه‌های مثبت و منفی می‌باشند. برخی از این روش‌ها از لحاظ زیست‌محیطی مطلوب هستند ولی از لحاظ اقتصادی توجیه‌پذیر نمی‌باشند. برخی دیگر از لحاظ اقتصادی برتری دارند ولی از لحاظ زیست‌محیطی ممنوعیت دارند. در یک اقتصاد بازار، گزینه‌هایی که هزینه‌های کل آنها کمتر باشد معمولاً انتخاب و پذیرفته می‌شوند. ولی مشخص نیست که کدام روش بدون اینکه آسیب‌های دائمی به محیط زیست وارد کند می‌تواند به بهترین نحو پاسخگوی نیازهای اساسی ما باشد. روندهای جاری نشان می‌دهند که افزایش سهم انرژی‌های تجدیدپذیر و تولید توزیع‌یافته از منابع مختلف نوید بخش کاهش/متعادل‌سازی مخاطرات زیست‌محیطی و اقتصادی خواهد بود.

صنعت برق نیازمند تلاش‌های بسیار پیچیده و مطابق با مقررات می‌باشد. شرکت‌ها برای مشارکت در این حوزه می‌توانند چهار نقش متمایز را بر عهده بگیرند: تولیدکنندگان برق، انتقال‌دهندگان برق، توزیع‌کنندگان برق و خرده‌فروشان. برای متصل کردن تمامی تولیدکنندگان به تمامی مشتریان از ساختاری پیچیده به نام شبکه برق استفاده می‌شود. اگرچه در تمامی جنبه‌های صنعت الکتریسیته شاهد افزایش شدید رقابت هستیم و بیشترین سهم بازار در اختیار تولیدکنندگان برق می‌باشد. تولیدکنندگان برق به منظور حفظ مزیت رقابتی باید با اتخاذ تصمیمات درست در زمان درست، حداکثر بهره‌برداری را از منابع کمیاب خود انجام دهند. استات‌سافت یکی از ارائه‌دهندگان راه‌حل‌های تجزیه و تحلیل سفارشی، ابزارهای پشتیبانی تصمیم یکپارچه‌ای را برای ژنراتورهای برق توسعه داده است. این ابزارهای نرم‌افزاری مبتنی بر داده‌کاوی با بهره‌گیری از داده‌های اخذ شده از فرایند تولید به تکنیسین‌ها و مدیران کمک می‌کنند تا پارامترهای فرایند بهینه را به سرعت برای رسیدن به حداکثر خروجی نیروی برق و در عین حال به حداقل رساندن ریسک تأثیرات منفی و مخرب تعیین کنند. در ادامه چند مثال از کارهایی که این ابزارهای تحلیلی پیشرفته که شامل شبکه‌های عصبی مصنوعی و ماشین‌های برداری پشتیبان هستند می‌توانند برای تولیدکنندگان برق انجام دهند آورده شده است.

- **مسئله‌ی بهینه‌سازی پارامترهای عملیات.** یک واحد چندسیکلونی ۳۰۰ مگاواتی که از سوخت زغال‌سنگ استفاده می‌کند نیازمند بهینه‌سازی دمای شعله‌ی بالا برای جلوگیری از تشکیل سرباره و سوختن اضافی است. **راه‌حل:** با استفاده از ابزارهای مدل‌سازی پیشگویانه شرکت استات‌سافت (همراه با داده‌های تاریخی سه دقیقه‌ای برای ۱۲ ماه)، تنظیمات بهینه برای پارامترهای کنترل برای نسبت‌های استوکیومتری (عنصرسنجی)، جریان زغال‌سنگ، هوای اصلی، سومین هوا و جریان دمپرهای اسپلیت هوای ثانویه شناسایی و پیاده‌سازی شد. **نتایج:** پس از بهینه‌سازی پارامترهای کنترلی، دمای شعله واکنش قوی نشان داد که منجر به احتراق پاک‌تر در درجه حرارت بالاتر و پایدارتر شعله گردید.
- **مسئله‌ی پیش‌بینی مشکلات پیش از رخ دادن آنها:** یک کوره‌ی DRB-4Z سوخت زغال‌سنگ ۴۰۰ مگاواتی برای اجتناب از خرابی‌های پرهزینه و نوسانات نیازمند بهینه‌سازی عملیات NOx بود. علل اصلی اسلیپ آمونیاک در یک فرایند کاهش انتخابی غیرکاتالیستی را برای کاهش NOx شناسایی کنید.

- **راه‌حل:** از متدولوژی‌های تحلیل پیشگویانه (به همراه داده‌های تاریخی فرایند) برای پیش‌بینی و کنترل تغییرات و نوسانات استفاده کرده و سپس فرایندها را برای دستیابی به عملکرد بهتر هدف‌گذاری کنید و بدین ترتیب NOx و تغییرپذیری را کاهش دهید.
 - **نتایج:** تنظیمات بهینه برای ترکیب پارامترهای کنترلی منجر به انتشار کمتر NOx و نوسان‌پذیری کمتر در طی عملیات با بار کم شد که شامل پیش‌بینی خرابی‌ها یا نگهداری و تعمیرات غیرمنتظره بوده است.
 - **مسئله‌ی کاهش انتشار گازهای مونوکسید کربن و اکسیدهای نیتروژن:** اگرچه انتشار گازهای NOx برای بارهای بالاتر در محدوده قابل قبول بود ولی کوره‌ی DRB-4Z سوخت زغال‌سنگ ۴۰۰ مگاواتی عملیات NOx پایین و تحت بار پایین (۱۷۵-۵۰ مگاوات) بهینه نشده بود.
 - **راه‌حل:** با استفاده از فناوری‌های مدل‌سازی پیشگویانه‌ی داده-محور با داده‌های تاریخی، تنظیمات بهینه برای پارامترها برای تغییرات جریان هوا شناسایی شد که منجر به مجموعه‌ای از دامنه‌های پارامتری ورودی مشخص و قابل حصول شد که به سادگی در سیستم‌های کنترل دیجیتالی قابل پیاده‌سازی بودند.
 - **نتایج:** پس از بهینه‌سازی، انتشار گازهای NOx در عملیات تحت بار پایین با انتشار گازهای NOx تحت بارهای سنگین قابل مقایسه بود.
- همانطور که این مثال‌ها نشان می‌دهند، فرصت‌های متعددی برای استفاده از تجزیه و تحلیل‌های پیشرفته در صنعت برق وجود دارد. استفاده از مدل‌های پیشگویانه و داده‌ها می‌تواند به تصمیم‌گیرندگان در دستیابی به بهترین کارایی سیستم تولیدشان و به حداقل رساندن تأثیر زیست‌محیطی کمک کند.

Source: Based on the StatSoft, Success Stories, statsoft.com/Portals/0/Downloads/EPRI.pdf (accessed June 2018) and the statsoft.fr/pdf/QualityDigest_Dec2008.pdf (accessed February 2018).

سوالاتی برای مورد کاربردی ۵.۲

۱. نگرانی‌های اصلی زیست‌محیطی در صنعت برق چه هستند؟
۲. حوزه‌های کاربردی اصلی مدل‌سازی پیشگویانه در صنعت برق کدام‌ها هستند؟
۳. چگونه از مدل‌سازی پیشگویانه برای رسیدگی و حل و فصل مسائل مختلف در صنعت برق استفاده می‌شود؟

سوالاتی برای مرور بخش ۵.۳

۱. معروف‌ترین معماری‌های شبکه‌های عصبی کدام‌ها هستند؟
۲. از معماری شبکه‌های عصبی مصنوعی کوهونن برای حل چه نوع مسائلی استفاده شده است؟
۳. نحوه کارکرد معماری شبکه عصبی مصنوعی هاپفیلد چگونه است؟ در چه نوع مسائلی می‌توان از آن استفاده کرد؟

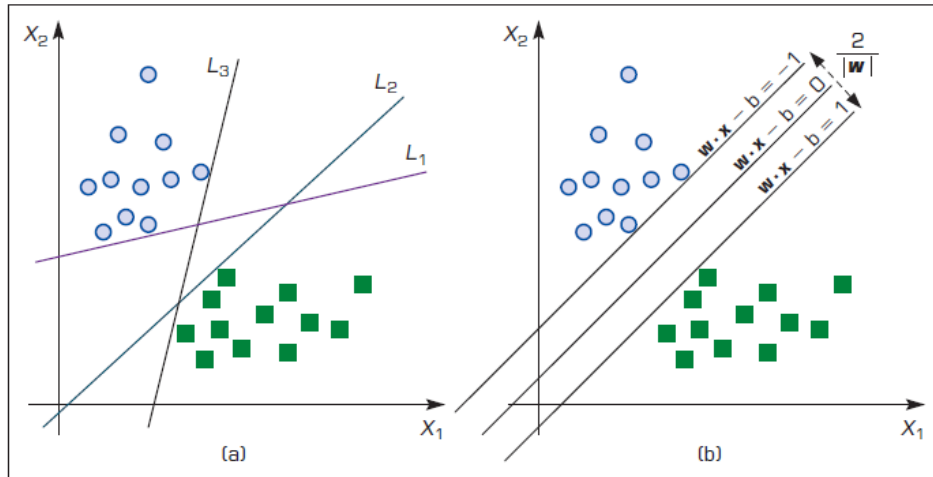
ماشین‌های برداری پشتیبان از محبوب‌ترین تکنیک‌های یادگیری ماشین هستند که این محبوبیت عمدتاً به دلیل توان پیش‌بینی برتر و بنیان نظری قوی آنها می‌باشد. ماشین‌های برداری پشتیبان جزو تکنیک‌های یادگیری با نظارت هستند که توابع ورودی-خروجی را از مجموعه‌ای از داده‌های یادگیری برچسب‌دار تولید می‌کنند. تابع بین بردارهای ورودی و خروجی می‌تواند از نوع دسته‌بندی (تخصیص موارد به دسته یا گروه‌های از پیش تعریف شده) یا رگرسیون (تخمین مقادیر عددی پیوسته برای یک خروجی مطلوب) باشند. برای دسته‌بندی اغلب از توابع هسته‌ی غیرخطی برای تبدیل داده‌های ورودی (معمولاً ارتباطات بسیار پیچیده‌ی غیرخطی) به یک فضای بُعد بالا^۱ که داده‌های ورودی را می‌توان در آن فضا به صورت خطی از یکدیگر تفکیک نمود استفاده می‌شود. سپس هایپرپلان‌های بیشترین حاشیه^۲ برای تفکیک بهینه‌ی دسته یا گروه‌های خروجی از یکدیگر در داده‌های آموزش ایجاد می‌شوند. با فرض یک مسئله‌ی پیش‌بینی از نوع طبقه‌بندی، تعداد زیادی از دسته‌بندی‌های خطی (هایپرپلان‌ها) می‌توانند داده‌ها را به چندین بخش فرعی تفکیک کنند که هر بخش فرعی بیانگر یکی از دسته‌ها یا گروه‌ها می‌باشد (شکل ۵.۶a) که دو گروه یا دسته با دایره و مربع نشان داده شده‌اند). با وجود این، فقط یک هایپرپلان می‌تواند بیشترین تفکیک بین دسته یا گروه‌ها را انجام دهد (شکل ۵.۶b) که یک هایپرپلان و دو هایپرپلان بیشترین حاشیه تفکیک، دو دسته یا گروه را از یکدیگر جدا کرده‌اند).

داده‌های بکار رفته در ماشین‌های برداری پشتیبان می‌توانند بیش از دو بُعد داشته باشند (یعنی دو دسته یا کلاس متمایز). در این صورت، ما داده‌ها را با استفاده از هایپرپلان $n-1$ بُعدی تفکیک و جدا می‌کنیم که n تعداد ابعاد (یعنی برچسب دسته یا کلاس) می‌باشد. این را می‌توان به شکل یک طبقه‌بند خطی دید که در آن ما می‌خواهیم $n-1$ هایپرپلان را به گونه‌ای بیابیم که فاصله‌ی هایپرپلان‌ها تا نزدیک‌ترین نقاط داده به بیشترین میزان خود برسد. فرض بر این است که هرچه حاشیه یا فاصله‌ی بین این هایپرپلان‌های موازی بزرگتر و بیشتر باشد، قدرت تعمیم‌دهی طبقه‌بند (یعنی توان پیش‌بینی مدل ماشین برداری پشتیبان) نیز بهتر خواهد بود. اگر چنین هایپرپلانی وجود داشته باشد، آنها را با استفاده از مدل‌سازی بهینه‌سازی کوادراتیک می‌توان به صورت ریاضی بیان کرد. این هایپرپلان‌ها تحت عنوان هایپرپلان بیشترین-حاشیه شناخته شده و چنین طبقه‌بند خطی را طبقه‌بند بیشترین-حاشیه^۳ می‌نامند.

¹ High-dimensional

² Maximum margin hyperplane

³ Maximum-margin classifier



شکل ۵.۶ تفکیک دو گروه یا دسته با استفاده از هایپرپلان‌ها

ماشین‌های برداری پشتیبان علاوه بر زیربنای ریاضی قوی در نظریه‌ی یادگیری آماری از عملکرد رقابتی بالایی در چندین مسئله‌ی پیش‌بینی واقعی از قبیل تشخیص‌های پزشکی، بیوانفورماتیک، تشخیص صدا/چهره، پیش‌بینی تقاضا، پردازش تصویر و متن‌کاوی برخوردارند که ماشین‌های برداری پشتیبان را جزو محبوب‌ترین ابزارهای تجزیه و تحلیل برای کشف دانش و داده‌کاوی قرار داده است. ماشین‌های برداری پشتیبان نیز همانند شبکه‌های عصبی مصنوعی توانایی بسیار خوبی برخوردارند تا به عنوان تخمین‌گرهای جهانی هرگونه تابع چندمتغیره با هر میزان دقت مطلوب استفاده شوند. از این رو، ماشین‌های برداری پشتیبان از اهمیت ویژه‌ای در مدل‌سازی مسائل، سیستم‌ها و فرایندهای بسیار پیچیده‌ی غیرخطی برخوردارند.

در خلاصه مطالعه‌ی پژوهشی که در مورد کاربردی ۵.۳ ارائه شده است، ماشین‌های برداری پشتیبان در پیش‌بینی و مشخص کردن عوامل ریسک شدت آسیب‌دیدگی در تصادفات خودروها بهتر از سایر روش‌های یادگیری ماشین عمل کرده است.

شناسایی عوامل ریسک شدت آسیب‌دیدگی در تصادفات خودروها با استفاده از تجزیه و تحلیل پیشگوبانه

مورد کاربردی ۵.۳

با پیشرفت فناوری‌ها، اقدامات ایمنی جدید و ارتقاء یافته‌ای برای پیشگیری از تصادفات و/یا کاهش تأثیر آسیب‌های وارده به سرنشینان در وسائل نقلیه و جاده‌ها تعبیه و بکار گرفته می‌شود. علیرغم گستردگی این تلاش‌ها، تصادفات وسائل نقلیه و آسیب‌های ناشی از آنها در سرتاسر جهان در حال افزایش است. به عنوان مثال، طبق آمار اداره ملی ایمنی ترافیک بزرگراه در ایالات متحده بیش از ۶ میلیون تصادف جاده‌ای در سال، جان بیش از ۳۰,۰۰۰ نفر را گرفته و موجب آسیب‌دیدگی بیش از ۲۰ میلیون نفر می‌شود (NHTSA 2014). طبق آخرین گزارش ارائه شده این اداره به کنگره‌ی ایالات متحده در آوریل ۲۰۱۴ بیان شده است که در سال ۲۰۱۲، تلفات جاده‌ای در ایالات متحده به ۳۳,۵۶۱ نفر رسید که نسبت به سال گذشته به میزان ۱,۰۸۲ نفر افزایش داشته است (فربدمن ۲۰۱۴). در همان سال، برآورد شده بود که تعداد آسیب‌ها در تصادفات مسائل نقلیه‌ی موتوری به ۲.۳۶ میلیون نفر برسد در حالیکه این رقم در سال ۲۰۱۲ معادل ۲.۲۲ میلیون نفر بوده است. در نتیجه، به طور میانگین در سال ۲۰۱۲

در هر ساعت ۴ نفر جان خود را در تصادفات جاده‌ای در ایالات متحده از دست داده و نزدیک به ۲۷۰ نفر نیز آسیب می‌دیدند. علاوه بر روند افزایشی تعداد تلفات و آسیب‌ها، این تصادفات جاده‌ای بیش از ۲۳۰ میلیارد دلار هزینه برای مالیت‌دهندگان در بر داشته است. از این رو، بحث ایمنی راه‌ها و جاده‌ها یک مشکل اساسی در ایالات متحده به حساب می‌آید.

علل ریشه‌ای تصادفات جاده‌ای و شدت آسیب‌های ناشی از آنها مورد توجه ویژه‌ی پژوهشگران (دانشگاهیان، دولت و صنعت) و عموم مردم بوده است زیرا اینگونه تحقیق و تفحص‌ها نه تنها به پیشگیری از تصادفات کمک می‌کنند بلکه از شدت آنها نیز کاسته و زندگی و پول بسیاری را نجات می‌دهند. علاوه بر روش‌های پژوهش مهندسی آزمایشگاهی، یک روش دیگر برای رسیدگی به این مسئله، شناسایی محتمل‌ترین عواملی تأثیرگذار بر شدت جراحات و آسیب‌ها از طریق کاوش در داده‌های تاریخی تصادفات و سائل نقلیه می‌باشد. درک عمیق شرایط پیچیده‌ای که رانندگان و/یا سرنشینان در یک تصادف ممکن است کشته یا به شدت زخمی شوند می‌تواند از مخاطرات مربوطه پیشگیری به عمل آورده و موجب نجات جان انسان‌ها شود. عوامل بسیاری وجود دارد که بر شدت آسیب و جراحات ناشی از تصادفات و سائل نقلیه تأثیر می‌گذارند. این عوامل شامل ویژگی‌های رفتاری یا جمعیت‌شناختی سرنشینان (مثلاً میزان مصرف دارو و/یا الکل، بسته بودن کمربند ایمنی یا استفاده از سایر سیستم‌های ایمنی، جنسیت و سن راننده)، مشخصه‌های موقعیتی (سطح/نوع/موقعیت جاده، تعداد خودروها و/یا اشیاء درگیر در تصادف، جهت و راستای تصادف)، عوامل زیست محیطی در زمان تصادف (شرایط آب و هوا، شرایط نور و/یا میدان دید، وقت از شبانه روز و غیره) و مشخصه‌های فنی وسیله‌ی نقلیه (سن، وزن، نوع بدنه و غیره).

هدف اصلی از این مطالعه تحلیلی، تعیین شایع‌ترین عوامل و اهمیت نسبی آنها در افزایش شدت آسیب‌ها و جراحات ناشی از تصادفات و سائل نقلیه بوده است. تصادفات بررسی شده در این مطالعه شامل مجموعه‌ای از نمونه‌ها بوده‌اند که در مناطق مختلف جغرافیایی اتفاق افتاده‌اند. برای اینکه نمونه‌ی منسجمی داشته باشیم، مجموعه داده‌ها فقط شامل انواع خاصی از تصادفات بوده است: برخورد مستقیم تک یا چند وسیله‌ی نقلیه، برخوردهای زاویه‌دار تک یا چند وسیله‌ی نقلیه و برخورد یک وسیله‌ی نقلیه با اشیاء ثابت. برای دستیابی به نتایج دقیق و معتبر در این تحقیق از متداول‌ترین تکنیک‌های یادگیری-ماشین برای تعیین اهمیت عوامل مربوط به تصادف و مقایسه‌ی تکنیک‌های مختلف یادگیری-ماشین استفاده شد.

روش پژوهش

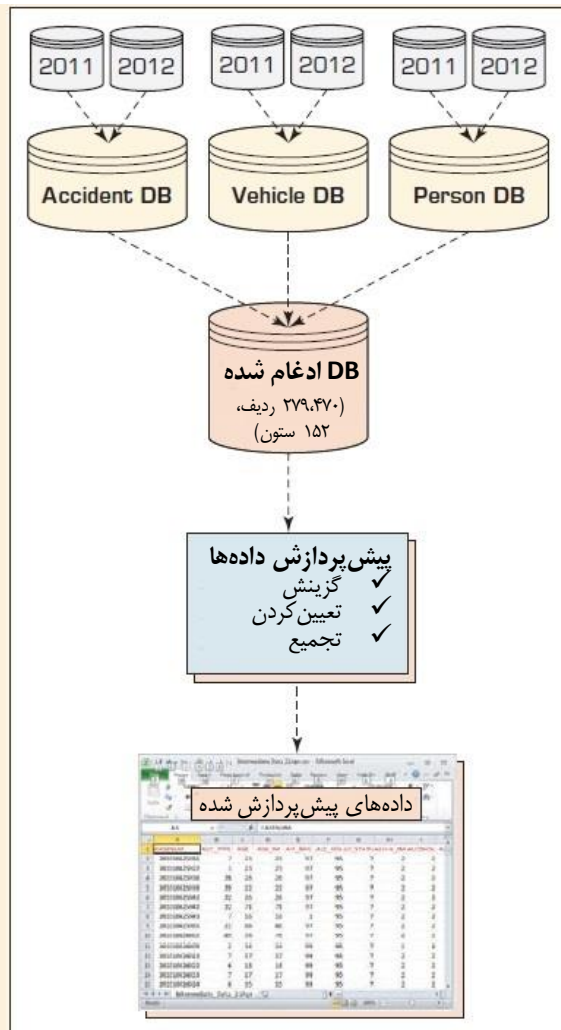
متدولوژی بکار رفته در این مطالعه از یک فرایند تجزیه و تحلیل استاندارد بسیار معروف به نام فرایند استاندارد صنعتی متقاطع برای داده‌کاوی یا (کریسپ-دی‌ام) پیروی می‌کند. همانند هر پروژه‌ی تجزیه و تحلیل، بخش قابل توجهی از زمان پروژه به اخذ، یکپارچه‌سازی و پیش‌پردازش داده‌ها اختصاص یافت. سپس داده‌های پیش‌پردازش شده که آماده‌ی تجزیه و تحلیل بودند برای ساخت مدل‌های متعدد و مختلف پیش‌بینی بکار گرفته شدند. پژوهشگران با استفاده از مجموعه‌ای از سنج‌های استاندارد به ارزیابی نتایج این مدل‌ها و مقایسه‌ی آنها پرداختند. در گام نهایی، از تجزیه و تحلیل حساسیت برای شناسایی شایع‌ترین عوامل ریسک که بر شدت آسیب‌ها و جراحات تأثیر دارد استفاده شد.

به منظور انجام مؤثر و کارای تک تک وظایف متدولوژی پیشنهادی، از چندین ابزار نرم‌افزاری آماری و داده‌کاوی استفاده شد. به طور خاص از نرم‌افزارهای جی‌ام‌پی (یک ابزار نرم‌افزارهای آماری و داده‌کاوی)، مایکروسافت اکسل و تابلو برای بازرسی، درک و پیش‌پردازش داده‌ها و از نرم‌افزارهای آی‌بی‌ام، اس‌پی‌اس‌مدلر و کی‌نایم برای ادغام داده‌ها، ساخت مدل‌های پیشگویانه و تحلیل حساسیت استفاده شد.

از مجموعه داده‌های سیستم تخمین‌های عمومی سیستم ملی نمونه‌برداری خودرو^۱ برای تصادفات در سال‌های ۲۰۱۱ و ۲۰۱۲ استفاده شد. مجموعه کامل داده‌ها به فرم سه فایل متنی/تخت مجزا - تصادف، وسیله نقلیه و شخص بدست آمد. فایل تصادفات حاوی مشخصه‌های خاص مربوط به شرایط جاده‌ها، شرایط زیست محیطی و شرایط مربوط به تصادف بوده است. فایل وسیله نقلیه شامل تعداد زیادی از متغیرها در مورد ویژگی‌های خاص وسیله نقلیه‌ی درگیر در تصادف بوده است. فایل اشخاص نیز جزئیات جمعیت‌شناختی در مورد سرنشینان (یعنی راننده و مسافران) در تصادف را شامل می‌شده است. به منظور ادغام داده‌ها در یک پایگاه داده، داده‌های این دو سال در داخل هر فایل ادغام شده و فایل‌های بدست آمده با استفاده از شناساگرهای منحصر بفرد تصادف، وسیله نقلیه و شخص با یکدیگر ترکیب شده و یک مجموعه داده تکی را بوجود آوردند. پس از ادغام/تجمیع داده‌ها، رکوردهای سطح شخصی که یک رکورد به ازای هر شخص در تصادف وسیله نقلیه گزارش شده بود بدست آمد. در این نقطه از فرایند، (پیش از پاک‌سازی، پیش‌پردازش و تقسیم داده‌ها)، مجموعه کامل داده‌ها شامل ۲۷۹،۴۷۰ رکورد منحصر بفرد (یعنی اشخاص/سرنشینان حاضر در تصادف) و بیش از ۱۵ متغیر (ترکیبی از مشخصه‌ها و ویژگی‌های تصادف، شخص و وسیله نقلیه) بود. شکل ۵.۷ نمایش گرافیکی از گام‌های پردازش داده‌ها را نشان می‌دهد.

از بین تمامی متغیرها - چه آنهایی که مستقیماً از پایگاه داده‌های جی‌ای‌اس بدست آمده‌اند چه آنهایی که از طریق محاسبات مجددی که بر روی متغیرهای موجود جی‌ای‌اس بدست آمده‌اند - ۲۹ متغیر به عنوان متغیر مرتبط که تأثیر بالقوه‌ای بر تعیین شدت آسیب و جراحات وارده در تصادفات و سائل نقلیه دارند برگزیده شدند. گستره‌ی متغیرهای انتخابی به نحوی بود که انتظار می‌رفت تا بتوان شرح کاملی از اشخاص و وسائل نقلیه‌ی درگیر در تصادفات ارائه کرد: مشخصه‌های شرایط زیست محیطی در زمان تصادف، شرایط پیرامون محل تصادف و زمان و مکان تصادف.

¹ National Automotive Sampling System General Estimates System (NASS GES)



شکل ۵.۷ فرایند اخذ/ادغام/آماده‌سازی داده‌ها

جدول ۵.۲ خلاصه‌ای از متغیرهای ایجاد شده و بکار رفته برای این مطالعه را ارائه کرده به طور مختصر توضیح می‌دهد. جدول ۵.۳ دقت پیش‌بینی هر چهار نوع مدل را نشان می‌دهد. این جدول نتایج ماتریس‌های درهم ریختگی، دقت کل، حساسیت، ویژگی و منحنی مشخصه عملکرد سیستم را که با استفاده از اعتبارسنجی متقاطع ۱۰ مرتبه‌ای برای هر چهار نوع مدل بدست آمده‌اند نشان می‌دهد. همانطور که نتایج نشان می‌دهند، ماشین‌های برداری پشتیبان، دقیق‌ترین تکنیک دسته‌بندی با دقت پیش‌بینی کل بالای ۹۰ درصد و حساسیت و ویژگی نسبتاً بالا و مقدار منحنی مشخصه عملکرد سیستم ۰.۹۲۸ است (ماکزیمم ۱.۰۰۰). بهترین مدل بعدی، الگوریتم‌های درخت تصمیم C5 بودند که دقت آنها کمی بهتر از شبکه‌های عصبی مصنوعی بوده است. آخرین رتبه از نظر دقت پیش‌بینی، رگرسیون لجستیک بوده است که دقت نسبتاً خوبی داشته است ولی به اندازه‌ی روش‌های یادگیری ماشین خوب نبوده است.

جدول ۰.۲ فهرست متغیرهای مورد مطالعه

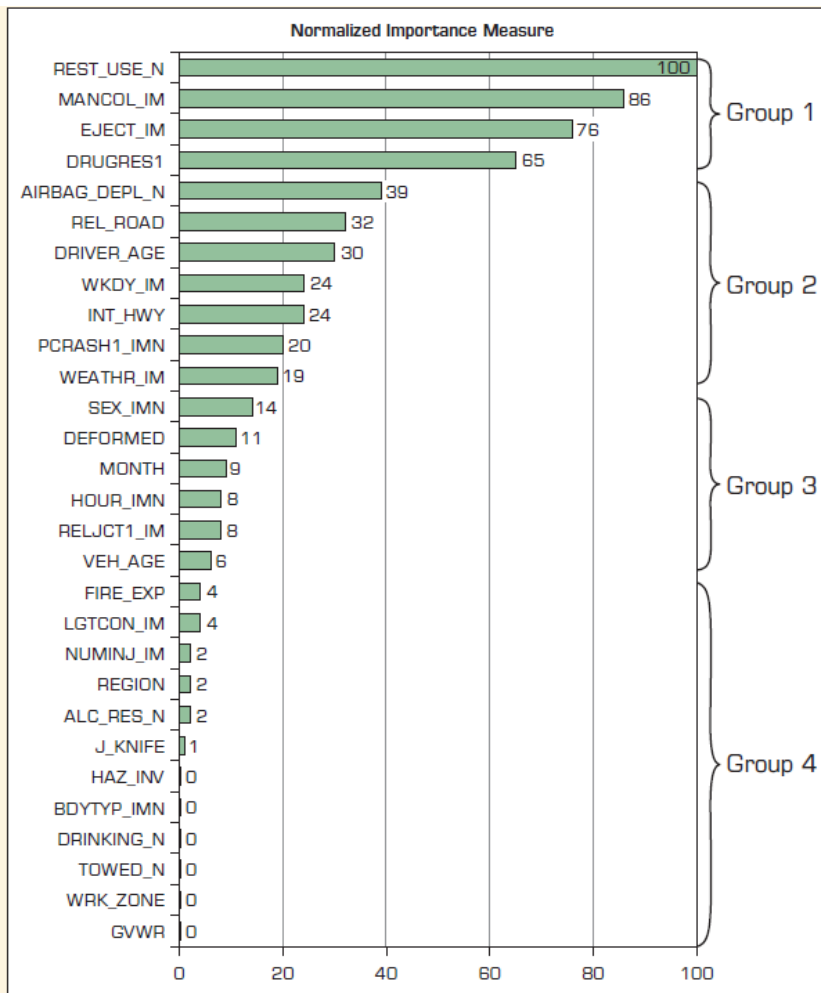
متغیر	شرح	نوع داده	آماره توصیفی	مفقود (%)
AIR_BAG	کیسه هوا باز شده است	دودویی	بله: ۵۲، خیر: ۲۶	۵.۲
ALC_RES	نتایج تست الکل	عددی	۱۲.۶۸ (۱۵.۰۵)	۰.۴
BDYTYP_IMN	نوع بدنه‌ی خودرو	اسمی	سدان: ۳۴، نیمه اس‌بیوی: ۱۳	۳.۲
DEFORMED	میزان آسیب	اسمی	جدی: ۴۳، جزئی: ۲۲	۳.۷
DRINKING	میزان الکل مصرف شده	دودویی	بله: ۴، خیر: ۶۷	۲۸.۸
AGE	سن شخص	عددی	۳۶.۴۵ (۱۸.۴۹)	۶.۹
DRUGRES1	نتایج تست مواد	دودویی	بله: ۲، خیر: ۷۲	۲۵.۵
EJECT_IM	پرتاب به بیرون	دودویی	بله: ۲، خیر: ۹۳	۴.۹
FIRE_EXP	وقوع آتشسوزی	دودویی	بله: ۳، خیر: ۹۷	۰.۰
GVWR	گروه وزنی وسیله نقلیه	اسمی	کوچک: ۹۲، بزرگ: ۵	۲.۹
HAZ_INV	وجود مواد خطرناک	دودویی	بله: ۱، خیر: ۹۹	۰.۰
HOUR_IMN	ساعت از شبانه روز	اسمی	بعد از ظهر: ۳۹، ظهر: ۲۲	۱.۲
INT_HWY	بزرگراه بین ایالتی	دودویی	بله: ۱۳، خیر: ۸۶	۰.۷
J_KNIFE	تا شدن خودرو	دودویی	بله: ۴، خیر: ۹۵	۰.۲
LGTCOIN_IM	شرایط روشنایی	اسمی	طول روز: ۷۰، تاریک: ۲۵	۰.۳
MANCOL_IM	نحوه‌ی تصادف	اسمی	از روبرو: ۳۴، از بغل: ۲۸	۰.۰
MONTH	ماه از سال	اسمی	اکتبر: ۱۰، دسامبر: ۹	۰.۰
NUMINI_IM	تعداد مجروحان	عددی	۱.۲۳ (۴.۱۳)	۰.۰
PCRASH1_IMN	حرکت پیش از تصادف	اسمی	با سرعت: ۵۲، متوقف: ۱۴	۱.۳
REGION	منطقه جغرافیایی	اسمی	جنوب: ۴۲، غرب میانه: ۲۴	۰.۰
REL_ROAD	نسبت به راه ترافیک	اسمی	بزرگراه: ۸۵، میانه: ۹	۰.۱
RELICT1_IM	در تقاطع	دودویی	بله: ۴، خیر: ۹۶	۰.۰
REST_USE_N	استفاده از سیستم نگهدارنده	دودویی	بله: ۷۶، خیر: ۴	۷.۴
SEX_IMN	جنسیت راننده	دودویی	مرد: ۵۴، زن: ۴۳	۳.۱
TOWED_N	خودرو حمل شده	دودویی	بله: ۴۹، خیر: ۵۱	۰.۰
VEH_AGE	سن خودرو	عددی	۸.۹۶ (۴.۱۸)	۰.۰
WEATHR_IM	شرایط آب و هوا	اسمی	صاف: ۷۳، ابری: ۱۴	۰.۰
WKDY_IM	روز هفته	اسمی	جمعه: ۱۷، پنجشنبه: ۱۵	۰.۰
WRK_ZONE	منطقه کاری	دودویی	بله: ۲، خیر: ۹۸	۰.۰
INJ_SEV	شدت جراحات	دودویی	کم: ۷۹، زیاد: ۲۱	۰.۰

جدول ۵.۳ جدول بندی نتایج تمامی پیش‌بینی‌ها بر اساس اعتبارسنجی متقاطع ۱۰ مرتبه‌ای

نوع مدل	ماتریس درهم ریختگی		دقت	حساسیت	ویژگی	AUC
	کم	زیاد				
شبکه‌های عصبی	کم	۱۲،۸۶۴	۱۰،۴۶۴	۸۸،۷۷	۸۹،۷۸	۰،۸۵۸
	زیاد	۲،۴۰۹	۱۰،۴۷۷			
ماشین‌های	کم	۱۳،۱۹۲	۱۰،۱۳۶	۹۰،۴۱	۹۲،۰۷	۰،۹۲۸
	زیاد	۱،۴۷۵	۱۱،۴۱۱			
...	کم	۱۲،۶۷۵	۱۰،۶۵۳	۸۵،۵۱	۸۱،۵۵	۰،۸۷۵
	زیاد	۱،۹۹۱	۱۰،۸۹۵			
رگرسیون	کم	۸،۹۶۱	۲،۷۴۲	۷۶،۹۷	۷۶،۸۷	۰،۸۲۷
	زیاد	۳،۵۲۵	۱۱،۹۸۶			

اگرچه سنج‌های دقت بدست آمده از هر چهار نوع مدل به اندازه‌ای بالا بودند که صحت و اعتبار متدولوژی پیشنهادی را تأیید کنند ولی هدف اصلی این مطالعه، شناسایی و اولویت‌بندی عوامل ریسک اصلی تأثیرگذار بر شدت آسیب‌دیدگی رانندگان در تصادف وسیله‌ی نقلیه بوده است. برای دستیابی به این هدف، یک تحلیل حساسیت بر روی تمامی مدل‌های پیش‌بینی انجام شد. سنج‌های اهمیت متغیر با تمرکز بر نوع تک تک مدل‌ها با استفاده از روش اعتبارسنجی یک طرفه محاسبه شده و نتایج بدست آمده برای هر نوع مدل جمع‌بندی شد. به منظور تلفیق مناسب نتایج تحلیل حساسیت برای هر چهار نوع مدل، سهم مدل در مقادیر اهمیت متغیر تلفیقی/ترکیبی بر اساس دقت اعتبارسنجی متقاطع آنها تعیین شد. یعنی اینکه بهترین مدل، مدلی است که بزرگترین وزن/سهم را داشته باشد و ضعیف‌ترین مدل، کوچکترین وزن/سهم را دارد. مقادیر اهمیت متغیر تلفیقی جدول‌بندی، نرمال‌سازی و به صورت گرافیکی در شکل ۵.۸ نشان داده شده‌اند.

بررسی نتایج تحلیل حساسیت حاکی از وجود چهار گروه متمایز از ریسک‌ها دارد که هر یک از چهار تا هشت متغیر تشکیل شده‌اند. گروه برتر به ترتیب اهمیت از بالا به پایین شامل REST_USE_N (آیا از کمربندی ایمنی یا هر سیستم نگهدارنده‌ی دیگری استفاده شده است)، MANCOL_IM (نحوه‌ی برخورد)، EJECT_IM (آیا راننده از خودرو به بیرون پرتاب شده است) و DRUGRES1 (نتایج تست مواد مخدر یا دارو) هستند. بر اساس نتایج تحلیل حساسیت تلفیقی هر چهار نوع مدل، این چهار عامل ریسک نسبت به سایر عوامل از اهمیت بسیار بالاتری برخوردارند.



شکل ۵.۸ مقادیر اهمیت متغیر

Sources: D. Delen, L. Tomak, K. Topuz, & E. Eryarsoy, "Investigating Injury Severity Risk Factors in Automobile Crashes with Predictive Analytics and Sensitivity Analysis Methods," *Journal of Transport & Health*, 4, 2017, pp. 118–131; D. Friedman, "Oral Testimony Before the House Committee on Energy and Commerce, by the Subcommittee on Oversight and Investigations," April 1, 2014, www.nhtsa.gov/Testimony (accessed October 2017); National Highway Traffic Safety Administration (NHTSA's) (2018) General Estimate System (GES), www.nhtsa.gov (accessed January 20, 2018).

سوالاتی برای مورد کاربردی ۵.۳

۱. مهمترین انگیزه‌ها در تحقیق و بررسی تحلیلی تصادفات خودروها چه هستند؟
۲. داده‌های بکار رفته در این مورد کاربردی چگونه گردآوری، ادغام و پیش‌پردازش شده‌اند؟
۳. نتایج این مطالعه چه بوده است؟ چگونه می‌توان از این یافته‌ها برای کاربردهای عملی استفاده کرد؟

$$\{(x_1, c_1), (x_2, c_2), \dots, (x_n, c_n)\}$$

$$x \in [-1, 1] \quad c \in [0, 1] \quad m$$

$$w \cdot x - b = 0$$

$$b \quad w$$

$$w \cdot x - b = 1$$

$$w \cdot x - b = -1$$

$$b \quad |w| \quad 2/|w|$$

$$i$$

$$w \cdot x_i - b \geq 1$$

$$w \cdot x_i - b \leq -1$$

$$c_i(w \cdot x_i - b) \geq 1 \quad 1 \leq i \leq n$$

$$1 \leq i \leq n \quad c_i(w \cdot x_i - b) \geq 1 \quad |w|$$

$$\begin{aligned} \text{Minimize} \quad & (1/2) \|w\|^2 \\ \text{Subject to} \quad & c_i(w \cdot x_i - b) \geq 1, \quad 1 \leq i \leq n \end{aligned}$$

1/2 به منظور راحتی در محاسبات ریاضی استفاده می‌شود.

نوشتن قاعده‌ی دسته‌بندی به فرم ثانویه‌اش به ما نشان می‌دهد که دسته‌بندی فقط تابعی از بردارهای پشتیبان است یعنی اینکه داده‌های آموزش بر روی حاشیه قرار می‌گیرند. ثانویه‌ی ماشین‌های برداری پشتیبان را به شکل زیر می‌توان نوشت:

$$\max \sum_{i=1}^n \alpha_i - \sum_{ij} \alpha_i \alpha_j c_i c_j x_i^T x_j$$

α بیانگر بازنمود ثانویه‌ی برداری وزنی بر حسب مجموعه داده‌های آموزش است.

$$w = \sum_i \alpha_i c_i x_i$$

$$c_i(w \cdot x_i - b) \geq 1 - \xi_i, \quad 1 \leq i \leq n$$

$$\xi_i$$

$$\min \|w\|^2 + C \sum_i \xi_i \quad \text{such that} \quad c_i(w \cdot x_i - b) \geq 1 - \xi_i, \quad 1 \leq i \leq n$$

$|w|$

C

$$k(x, \hat{x}) = (x \cdot \hat{x})$$

$$k(x, \hat{x}) = (x \cdot \hat{x} + 1)$$

$$k(x, \hat{x}) = (x \cdot \hat{x})$$

$$k(x, \hat{x}) = \exp(-\gamma \|x - \hat{x}\|^2) \quad \text{for } \gamma > 0$$

$$k(x, \hat{x}) = \exp\left(-\frac{\|x - \hat{x}\|^2}{2\sigma^2}\right)$$

$$k(x, \hat{x}) = \tan h(kx \cdot \hat{x} + c) \quad \text{for } k > 0, c < 0$$

¹ Vladimir Vapnik

$K(x, y)$

X

$$\sum_{i,j} K(x_i, x_j) c_i c_j \geq 0$$

$\{x_1, x_2, \dots, x_n\} \subset X$ $\{x_1, x_2, \dots, x_n\}$ $\varphi(x)$

$$K(x, y) = \varphi(x) \cdot \varphi(y)$$

φ

φ

-
-
-
-
-

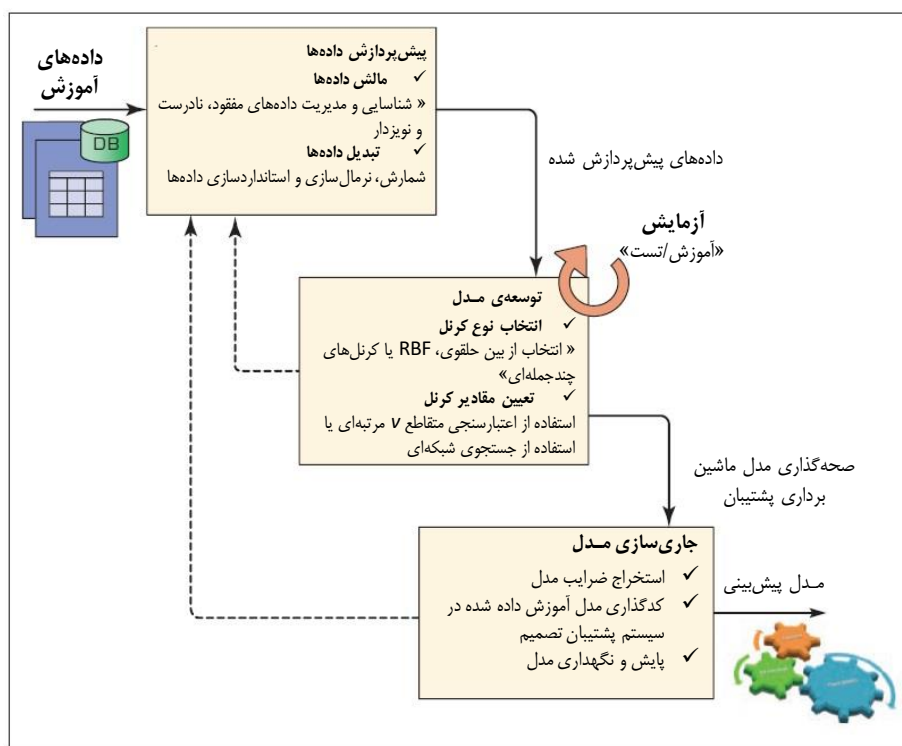
سوالاتی برای مرور بخش ۵.۴

¹ Mercer

² Kernel trick

۱. ماشین‌های برداری پشتیبان چگونه کار می‌کنند؟
۲. مزایا و معایب ماشین‌های برداری پشتیبان چه هستند؟
۳. مفهوم "هایپرپلان‌های حاشیه-ماکزیمم" چیست؟ چرا در ماشین‌های برداری پشتیبان اهمیت دارند؟
۴. "ترفند کرنل" چیست؟ چگونه از آن در ماشین‌های برداری پشتیبان استفاده می‌شود؟

ماشین‌های برداری پشتیبان اخیراً عمدتاً به دلیل نتایج بهتری که در دسته‌بندی‌ها ارائه کرده‌اند به تکنیک محبوبی برای مسائل دسته‌بندی تبدیل شده‌اند. اگرچه به نظر می‌رسد که استفاده از آنها راحت‌تر از شبکه‌های عصبی مصنوعی باشد ولی کاربرانی که با پیچیدگی‌های ماشین‌های برداری پشتیبان آشنایی ندارند اغلب نتایج نامطلوبی را بدست می‌آورند. در این بخش ما یک رویکرد فرایند-محور را برای استفاده در ماشین‌های برداری پشتیبان ارائه می‌کنیم که به احتمال زیاد نتایج بهتری را بدست خواهد داد. یک نمایش گرافیکی از فرایند سه مرحله‌ای در شکل ۵.۹ ارائه شده است.



شکل ۵.۹ شرح فرایند ساده‌ی توسعه‌ی مدل‌های ماشین برداری پشتیبان

رقومی‌سازی داده‌ها در ماشین‌های برداری پشتیبان، هر نمونه داده باید به صورت یک بردار از اعداد حقیقی نمایش داده شود. از این رو، اگر صفات رسته‌ای داشته باشیم ابتدا باید آنها را به داده‌های عددی تبدیل کنیم. یک توصیه متداول برای اینکار، استفاده از m متغیر شبه‌دودویی برای نمایش صفت دسته‌ی m است ($3 \cup m$). در عمل، فقط یکی از m متغیر مقدار ۱ را به

خود می‌گیرند و مابقی بر اساس دسته یا گروه حقیقی مقدار ۰ به خود می‌گیرند. به عنوان مثال، یک صفت سه دسته‌ای مثل {قرمز، سبز، آبی} را می‌توان به شکل $(0, 1, 0)$, $(0, 0, 1)$ و $(1, 0, 0)$ نمایش داد.

نرمال‌سازی داده‌ها همانند شبکه‌های عصبی مصنوعی، ماشین‌های برداری پشتیبان نیز نیازمند نرمال‌سازی و/یا مقیاس‌دهی مقادیر عددی هستند. مزیت اصلی نرمال‌سازی این است که صفاتی که در طیف‌های عددی بزرگتر قرار می‌گیرند نمی‌توانند بر طیف‌های عددی کوچکتر تسلط پیدا کنند. مزیت دیگر آن این است که به انجام محاسبات عددی در طی فرایند تکرارپذیر ساخت مدل کمک می‌کند. از آنجائیکه مقادیر کرنل به حاصلضرب داخلی بردارها بستگی دارد (مثلاً کرنل خطی و کرنل چندجمله‌ای)، هرچه مقادیر صفات بزرگتر باشند، فرایند آموزش کندتر خواهد بود. از توصیه‌های نرمال‌سازی هر صفت به دامنه‌ی $[-1, +1]$ یا $[0, 1]$ استفاده کنید. البته، ما باید از روش نرمال‌سازی یکسانی برای مقیاس‌دهی داده‌های آموزش، پیش از تست استفاده کنیم.

انتخاب نوع کرنل و پارامترهای آن اگرچه فقط چهار کرنل متداول در بخش قبل معرفی شدند ولی باید تصمیم بگیریم که از کدامیک از آنها استفاده کنیم (یا اینکه از تمامی آنها در یک زمان با استفاده از یک روش طراحی آزمایش ساده استفاده کنیم).

به محض انتخاب نوع کرنل، آنگاه باید مقدار پارامتر جریمه‌ی C و پارامترهای کرنل را انتخاب کنیم. به بیان کلی‌تر، تابع مبنای شعاعی اولین انتخاب منطقی برای نوع کرنل است. هدف کرنل تابع مبنای شعاعی این است که داده‌های غیرخطی را بر روی فضای بُعد بالا ترسیم کند. با اینکار (برخلاف کرنل خطی)، مواردی که رابطه‌ی بین بردارهای ورودی و خروجی در آنها شدیداً غیرخطی است را می‌توان مدیریت کرد. علاوه بر این، باید در نظر داشت که کرنل خطی فقط یک حالت خاص از کرنل تابع مبنای شعاعی است. در کرنل‌های تابع مبنای شعاعی باید دو پارامتر را تعیین کرد: C و g . چون از قبل مشخص نیست که بهترین مقادیر برای یک مسئله‌ی پیش‌بینی چه هستند به همین دلیل باید از یک روش جستجوی پارامتر استفاده کرد. هدف از جستجو، شناسایی مقادیر بهینه برای پارامترهای C و g به نحوی است که طبقه‌بند بتواند داده‌های نامعین (یعنی داده‌های تست) را به دقت پیش‌بینی کند. دو روش متداول جستجوی پارامتر عبارتند از اعتبارسنجی متقاطع و جستجوی شبکه.

جاری‌سازی مدل به محض توسعه‌ی مدل پیش‌بینی « بهینه‌ی » ماشین برداری پشتیبان، گام بعدی این است که آن را با سیستم پشتیبان تصمیم ادغام کنیم. برای این کار، دو گزینه وجود دارد: (۱) تبدیل مدل به یک شیء محاسباتی (مثلاً سرویس وب، جاوا بین یا مدل شیء مؤلفه) که مقادیر پارامتر ورودی را گرفته و خروجی را پیش‌بینی می‌کند و (۲) استخراج ضرایب مدل و یکپارچه‌سازی مستقیم آن با سیستم پشتیبان تصمیم. مدل‌های ماشین‌های برداری پشتیبان فقط در صورتی مفید هستند که رفتار دامنه‌ی زیربنایی آنها ثابت باقی بماند. اگر به دلایلی این رفتار تغییر کند، دقت مدل نیز تغییر خواهد کرد. از این رو، عملکرد مدل باید به طور مداوم ارزیابی شده و در صورتیکه دقت خود را از دست داده‌اند تصمیم‌گیری شود.

اگرچه برخی افراد، ماشین‌های برداری پشتیبان را نوع خاصی از شبکه‌های عصبی مصنوعی در نظر می‌گیرند ولی اکثراً اینها را دو تکنیک یادگیری ماشین رقیب با کیفیت‌های متفاوت می‌دانند. در اینجا به ذکر برخی نکات که تفاوت‌های ماشین‌های برداری

پشتیبان و شبکه‌های عصبی مصنوعی را نشان می‌دهد می‌پردازیم. از لحاظ تاریخی، توسعه‌ی شبکه‌های عصبی مصنوعی پس از یک مسیر هیورستیک و آزمایش گسترده انجام شد. برعکس، توسعه‌ی ماشین‌های برداری پشتیبان شامل یک نظریه‌ی یادگیری آماری درست و سپس پیاده‌سازی و آزمایشات آن بود. یک مزیت مهم ماشین‌های برداری پشتیبان این است که راه‌حل‌های آنها منحصر بفرد و عمومی هستند در حالیکه شبکه‌های عصبی مصنوعی چندین راه‌حل مینیمم موضعی دارند. دو مزیت دیگر ماشین‌های برداری پشتیبان این است که تفسیر هندسی ساده‌ای دارند و جواب **پراکنده‌ای** به مسئله می‌دهند. دلیل اینکه اغلب ماشین‌های برداری پشتیبان عملکرد بهتری نسبت به شبکه‌های عصبی مصنوعی دارند این است که در عمل بهتر با مسئله‌ی « بیش‌برازش » کنار می‌آیند که مسئله‌ی بزرگی در شبکه‌های عصبی مصنوعی محسوب می‌شود.

اگرچه ماشین‌های برداری پشتیبان این مزیت‌ها را دارند (از نقطه نظر عملی) ولی محدودیت‌هایی نیز دارند. یک مسئله‌ی مهم که به طور کامل حل نشده، انتخاب نوع کرنل و پارامترهای تابع کرنل است. محدودیت دوم و شاید مهم‌تر ماشین‌های برداری پشتیبان، سرعت و اندازه آنها در سیکل‌های آموزش و تست است. ساخت مدل در ماشین‌های برداری پشتیبان نیازمند محاسبات پیچیده و زمان‌بر است. از نقطه نظر عملی، شاید جدی‌تر مسئله‌ی آنها، پیچیدگی الگوریتمی بالا و نیاز به حافظه‌ی زیاد در برنامه‌نویسی کوادراتیک در کارهای بزرگ است. علیرغم این محدودیت‌ها، از آنجائیکه ماشین‌های برداری پشتیبان مبتنی بر یک مبنای نظری درست و محکم هستند و جواب‌هایی که تولید می‌کنند نیز منحصر بفرد و عمومی هستند، امروزه بدون شک می‌توان آنها را جزو محبوب‌ترین و متداول‌ترین تکنیک‌های مدل‌سازی پیش‌بینی در حوزه‌ی داده‌کاوی دانست. کاربرد و محبوبیت آنها فقط زمانی افزایش خواهد یافت که در ابزارهای داده‌کاوی تجاری به کار گرفته شوند.

سوالاتی برای مرور بخش ۵.۵

۱. گام‌ها و نقاط تصمیم اصلی در توسعه‌ی یک مدل ماشین برداری پشتیبان چه هستند؟
۲. چگونه می‌توان بهترین نوع کرنل و پارامترهای کرنل را تعیین نمود؟
۳. در مقایسه با شبکه‌های عصبی مصنوعی، مزیت‌های ماشین‌های برداری پشتیبان چه هستند؟
۴. کاربردهای متداول ماشین‌های برداری پشتیبان چه هستند؟ برای شناسایی حوزه‌های کاربردی متداول و ابزارهای نرم‌افزاری خاص ماشین‌های برداری پشتیبان که در این حوزه‌ها بکار گرفته شده‌اند در اینترنت به جستجو بپردازید.

الگوریتم‌های داده‌کاوی به لحاظ حجم محاسباتی و ریاضیاتی بسیار پیچیده هستند. دو الگوریتم متداول که در بخش قبل به بررسی آنها پرداختیم (شبکه‌های عصبی مصنوعی و ماشین‌های برداری پشتیبان) دارای الگوریتم‌های مشتق شده‌ی بسیار زمان‌بر و با

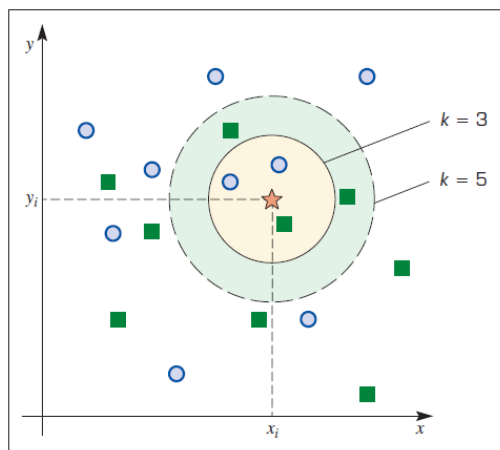
محاسبات ریاضیاتی بسیار پیچیده و سنگین هستند. برعکس، الگوریتم **کی نزدیک‌ترین همسایه**¹ (یا به اختصار k NN) به عنوان یک روش پیش‌بینی رقیب، بیش از اندازه ساده به نظر می‌رسد. درک نحوه کارکرد و هدف این الگوریتم (و توضیح دادن آن به دیگران) بسیار آسان است. **کی نزدیک‌ترین همسایه** یک روش پیش‌بینی برای دسته‌بندی و همچنین برای مسائل پیش‌بینی رگرسیونی است. **کی نزدیک‌ترین همسایه** یک نوع یادگیری مبتنی بر نمونه (یا یادگیری تنبل) است زیرا تابع آن فقط به صورت محلی تقریب زده می‌شود و تمامی محاسبات تا زمان پیش‌بینی واقعی به تعویق می‌افتند.

الگوریتم **کی نزدیک‌ترین همسایه** جزو ساده‌ترین الگوریتم‌های یادگیری ماشین است: به عنوان مثال، در پیش‌بینی از نوع دسته‌بندی، یک مورد بر اساس رأی اکثریت همسایه‌های دسته‌بندی می‌شود که یک شیء به مشترک‌ترین دسته یا گروه در بین **کی** نزدیک‌ترین همسایگانش تخصیص می‌یابد (که k یک عدد صحیح مثبت است). اگر $k = 1$ باشد آنگاه مورد به دسته یا گروه نزدیک‌ترین همسایه‌اش تخصیص می‌یابد. برای اینکه این مفهوم را با یک مثال به تصویر بکشیم، نگاهی به شکل ۵.۱۰ داشته باشیم که یک فضای ساده‌ی دو بُعدی بیانگر مقادیر دو متغیر (x, y) است؛ مورد جدید (یا شیء) با ستاره و موارد شناخته شده (یا مثال‌ها) با دایره و مربع نمایش داده می‌شوند. هدف این است که مورد جدید را بر اساس مجاورت (تشابه) به یکی از این دو به دایره یا مربع‌ها تخصیص دهیم. اگر مقدار $k = 1$ باشد آنگاه مورد جدید به مربع‌ها تخصیص می‌یابد زیرا نزدیک‌ترین مثال به ستاره، مربع است. اگر $k = 3$ باشد آنگاه مورد جدید را باید به دایره تخصیص دهیم زیرا دو دایره و یک مربع وجود دارد و از این رو، بر اساس قاعده‌ی ساده‌ی رأی اکثریت، مورد جدید به دایره تخصیص می‌یابد. به همین منوال، اگر مقدار $k = 5$ در نظر بگیریم، آنگاه مورد جدید باید به گروه یا دسته‌ی مربع تخصیص داده شود. این مثال بسیار ساده، اهمیت مقداری که به k تخصیص می‌دهیم را نشان می‌دهد.

همین روش را می‌توان برای پیش‌بینی‌های رگرسیونی بکار برد که فقط کافی است میانگین مقادیر **کی** نزدیک‌ترین همسایه‌ها را محاسبه کرده و نتیجه‌ی آن را به موردی که قرار است پیش‌بینی شود تخصیص دهیم. وزن‌دهی به همسایه‌ها می‌تواند مفید باشد به نحویکه همسایه‌های نزدیک‌تر، نسبت به همسایه‌های دورتر سهم بیشتری را در میانگین به خود اختصاص دهند. یک روش متداول وزن‌دهی این است که به هر همسایه وزن $1/d$ را اختصاص بدهیم که d فاصله تا همسایه است. این الگو یک حالت تعمیم یافته از درون‌یابی خطی است.

همسایه‌ها از مجموعه‌ای از موارد انتخاب می‌شوند که دسته‌بندی صحیح (یا در خصوص رگرسیون، مقدار عددی مقدار خروجی) مشخص است. این را می‌توان به صورت مجموعه داده‌های آموزش برای الگوریتم در نظر گرفت اگرچه صراحتاً هیچ گام آموزشی مورد نیاز نیست. الگوریتم **کی** نزدیک‌ترین همسایه نسبت به ساختار محلی داده‌ها حساس است.

¹ k -nearest neighbor



شکل ۵.۱۰ اهمیت مقدار k در الگوریتم کی نزدیک‌ترین همسایه

یکی از دو تصمیم بحرانی که یک تحلیلگر در بکارگیری الگوریتم کی نزدیک‌ترین همسایه باید اتخاذ کند، تعیین معیار تشابه است (تصمیم دیگر، تعیین مقدار k است که بعداً توضیح خواهیم داد). در الگوریتم کی نزدیک‌ترین همسایه، معیار تشابه، یک سنجی فاصله است که به صورت ریاضی قابل محاسبه می‌باشد. با در نظر گرفتن یک مورد جدید، این الگوریتم پیش‌بینی‌هایی را بر اساس نتیجه‌ی k همسایه که نزدیک‌ترین فاصله را با آن نقطه دارند انجام می‌دهد. از این رو، برای انجام پیش‌بینی با الگوریتم کی نزدیک‌ترین همسایه، باید سنجی‌ای را برای اندازه‌گیری فاصله‌ی بین مورد جدید و موارد مثال‌های موجود تعریف کنیم. یکی از متداول‌ترین گزینه‌ها برای اندازه‌گیری این فاصله، معادله ۲ اقلیدسی است که فاصله‌ی خطی ساده بین دو نقطه را در یک فضای تک بُعدی اندازه‌گیری می‌کند. یک گزینه‌ی دیگر، روش مستقیم‌الخط (فاصله منهتن) معادله ۳ است. هر دو سنجی فاصله، موارد ویژه‌ای برای فاصله‌ی مینکوفسکی (معادله ۱) محسوب می‌شوند.

$$d(i, j) = \sqrt{\left(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q \right)} \quad (\text{معادله ۱})$$

که $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ و $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ دو شیء داده‌ای p بُعدی (مثلاً یک مورد جدید و یک مثال در مجموعه داده‌ها) و q یک عدد صحیح مثبت است.

اگر $q = 1$ باشد آنگاه d را فاصله‌ی منهتن می‌نامند.

$$d(i, j) = \sqrt{\left(|x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}| \right)} \quad (\text{معادله ۲})$$

اگر $q = 2$ باشد آنگاه d را فاصله‌ی اقلیدسی می‌نامند.

$$d(i, j) = \sqrt{\left(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2 \right)} \quad (\text{معادله ۳})$$

مشخصاً، این معیارها فقط بر روی داده‌های عددی قابل محاسبه هستند. اما در مورد داده‌های اسمی چطور؟ دو روش برای اندازه‌گیری فاصله‌ی بین داده‌های غیر عددی وجود دارد. در ساده‌ترین حالت، برای یک متغیر اسمی چندمقداره، اگر مقدار آن متغیر برای مورد جدید و برای مورد مثال، یکسان باشند، فاصله برابر ۰ است و در غیراینصورت برابر ۱ خواهد بود. در مواردی مثل دسته‌بندی یا دسته‌بندی متن، سنج‌های پیچیده‌تری وجود دارد مثل سنج‌ی همپوشانی (یا فاصله‌ی هامینگ^۱). اغلب، اگر سنج‌ی فاصله از طریق یک طراحی آزمایش تعیین شود که در آن سنج‌های مختلفی برای شناسایی بهترین مورد برای مسئله‌ی مفروض مورد تست و آزمایش قرار گرفته‌اند آنگاه دقت دسته‌بندی الگوریتم کی نزدیک‌ترین همسایه را به میزان زیادی می‌تواند بهبود بخشد.

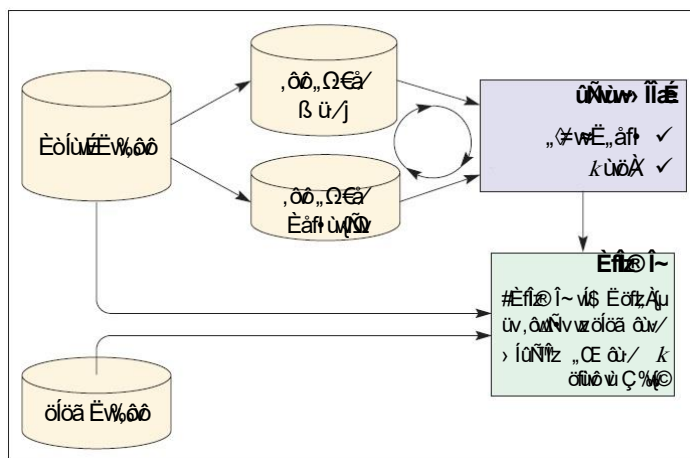
بهترین انتخاب مقدار k به داده‌ها بستگی دارد. به طور کلی، هرچه مقدار k بزرگتر باشد، تأثیر نویز بر روی دسته‌بندی (یا رگرسیون) کمتر خواهد بود ولی موجب می‌شود تا مرزهای بین دسته‌ها یا گروه‌ها نیز کمرنگ‌تر شوند. مقدار «بهینه‌ی» k را می‌توان به کمک چند تکنیک هیوریستیک مثل اعتبارسنجی متقاطع بدست آورد. حالت خاصی که در آن، دسته یا گروه پیش‌بینی شده، نزدیک‌ترین نمونه آموزشی است (یعنی $k = 1$) را الگوریتم نزدیک‌ترین همسایه می‌نامند.

اعتبارسنجی متقاطع اعتبارسنجی متقاطع یک تکنیک طراحی آزمایش معتبر است که از آن می‌توان برای تعیین مقادیر بهینه برای مجموعه‌ای از پارامترهای نامعلوم مدل استفاده کرد. از این تکنیک برای اکثر تکنیک‌های یادگیری ماشین که در آنها چندین پارامتر مدل باید تعیین شود استفاده می‌شود. ایده‌ی اصلی این روش طراحی آزمایش این است که نمونه داده‌ها را به طور تصادفی به چندین نمونه‌ی جزء و گسسته از هم تقسیم می‌کند (یعنی v قسمت). برای هر مقدار بالقوه‌ی k ، از مدل کی نزدیک‌ترین میانگین برای پیش‌بینی قسمت v ام استفاده می‌شود در حالیکه از $v-1$ قسمت برای مثال‌ها و ارزیابی خطا استفاده می‌شود. گزینه معمول برای ارزیابی این خطا، خطای جذر میانگین مربعات برای پیش‌بینی‌های رگرسیونی و درصد نمونه‌هایی که به درستی دسته‌بندی نشده‌اند (یعنی نرخ ضربه) برای پیش‌بینی‌های از نوع دسته‌بندی است. این فرایند آزمون در هر مرتبه در مقابل مثال‌های باقیمانده، v مرتبه تکرار می‌شود. در پایان v چرخه‌ی تکرار، خطاهای محاسبه شده جمع شده و سنج‌ی برازش مدل را نتیجه می‌دهد (یعنی پیش‌بینی این مدل با مقدار فعلی k تا چه اندازه خوب بوده است). در انتها، مقدار k که کوچکترین خطای کل را نتیجه داده باشد به عنوان مقدار بهینه‌ی آن مسئله انتخاب می‌شود. شکل ۵.۱۱ یک فرایند ساده را نشان می‌دهد که از داده‌های آموزش برای تعیین مقادیر بهینه‌ی k و سنج‌های فاصله استفاده می‌کند که سپس برای پیش‌بینی موارد جدید استفاده می‌شود. همانطور که در مثال ساده‌ی قبل دیدیم، دقت الگوریتم کی نزدیک‌ترین همسایه بسته به مقادیر مختلف k می‌تواند متفاوت باشد. علاوه بر این، توان پیش‌بینی الگوریتم کی نزدیک‌ترین همسایه با حضور داده‌های نادقیق یا نامربوط و نویزدار کاهش می‌یابد. تلاش‌های پژوهشی بسیاری در خصوص انتخاب، نرمال‌سازی/مقیاس‌دهی جهت‌یابی به نتایج پیش‌بینی معتبر انجام پذیرفته است. یک رویکرد بسیار ویژه، استفاده از الگوریتم‌های تکاملی (مثل الگوریتم ژنتیک) برای بهینه‌سازی مجموعه‌ی ویژگی‌های تعریف شده در سیستم پیش‌بینی کی نزدیک‌ترین همسایه است. در مسائل طبقه‌بندی دودویی (دو دسته‌ای)، بهتر است که k عددی فرد در نظر گرفته شود تا از آراء مساوی اجتناب شود.

یکی از عیوب دسته‌بندی بر اساس رأی اکثریت در الگوریتم کی نزدیک‌ترین همسایه این است که گروه‌ها و دسته‌هایی که فراوانی مثال‌ها در آنها بیشتر است بر پیش‌بینی بردار جدید تسلط می‌یابند زیرا زمانیکه همسایه‌ها به دلیل تعداد زیادشان محاسبه می‌شوند

¹ Hamming distance

جزو کی نزدیک‌ترین همسایه‌ها قرار می‌گیرند. یک روش برای غلبه بر این مشکل این است که با در نظر گرفتن فاصله‌ی نقطه تست با هر یک از کی نزدیک‌ترین همسایه‌ها به روش دسته‌بندی وزن بدهیم. یک روش دیگر این است که نمایش داده‌ها را یک سطح خلاصه‌تر کنیم. نسخه‌ی ساده‌ی این الگوریتم را به آسانی می‌توان از طریق محاسبه‌ی فواصل نمونه‌ی تست با تمامی بردارهای ذخیره شده اجرا کرد ولی از لحاظ محاسباتی بسیار سنگین خواهد بود به ویژه زمانی که حجم داده‌های آموزش افزایش یابد. طی سالیان گذشته، تعداد بسیار زیادی از الگوریتم‌های جستجوی نزدیک‌ترین همسایه ارائه شده است که عموماً به دنبال کاهش تعداد ارزیابی فواصل هستند. استفاده از یک الگوریتم مناسب جستجوی نزدیک‌ترین همسایه موجب شده تا کی نزدیک‌ترین میانگین حتی برای مجموعه داده‌های بزرگ از لحاظ محاسباتی قابل ردیابی باشد. در مورد قابلیت‌های برتر کی نزدیک‌ترین میانگین در بازشناسی و گروه‌بندی تصویر به مورد کاربردی ۵.۴ مراجعه کنید.



شکل ۵.۱۱ فرایند تعیین مقادیر بهینه برای سنجه‌ی فاصله و k

بازشناسی و گروه‌بندی مؤثر تصویر با استفاده از کی نزدیک‌ترین همسایه

مورد کاربردی ۵.۴

بازشناسی تصویر یکی از کاربردهای نوظهور داده‌کاوی است که شامل پردازش، تحلیل و گروه‌بندی اشیاء دیداری مثل عکس و تصویر است. در فرایند بازشناسی (یا گروه‌بندی)، تصاویر ابتدا به یک فضای ویژگی چندبُعدی تبدیل شده و سپس با استفاده از تکنیک‌های یادگیری ماشین، به تعداد محدودی از دسته‌ها گروه‌بندی می‌شوند. حوزه‌های کاربردی بازشناسی و گروه‌بندی تصویر از کشاورزی تا امنیت ملی، بازاریابی شخصی‌سازی شده یا حفاظت محیط‌زیست را شامل می‌شوند. بازشناسی تصویر، جزو لاینفک یکی از رشته‌های هوش مصنوعی به نام *بینایی رایانه‌ای* است. بینایی رایانه‌ای به عنوان یک رشته‌ی تکنولوژیکی به دنبال توسعه‌ی سیستم‌های رایانه‌ای است که قادر به «دیدن» و واکنش نسبت به محیط اطراف خود باشند. مثال‌هایی از کاربرد بینایی رایانه‌ای شامل سیستم‌های خودکارسازی فرایند (ربات‌های صنعتی)، ناوبری (وسائل نقلیه مستقل)، پایش/تشخیص (نظارت دیداری)، جستجو و مرتب‌سازی (اندیس‌گذاری پایگاه‌های داده تصاویر و یک توالی از تصاویر)، درگیرسازی (تعامل انسان-رایانه) و بازرسی (فرایندهای ساخت و تولید) می‌باشند.

علیرغم رشد و پیشرفت سریع رشته‌ی تشخیص دبداری و بازشناسی دسته یا گروه، کارهای بسیار زیادی باید برای رسیدن به عملکرد سطح انسانی انجام شود. رویکردهای جاری قادر هستند تا فقط با تعداد محدودی از دسته‌ها یا گروه‌ها (حدوداً ۱۰۰ مورد) کار کنند و از لحاظ محاسباتی نیز بسیار پرهزینه هستند. بسیاری از تکنیک‌های یادگیری ماشین (از جمله شبکه‌های عصبی مصنوعی، ماشین‌های برداری پشتیبان و کی نزدیک‌ترین الگوریتم) برای توسعه‌ی سیستم‌های رایانه‌ای برای بازشناسی و گروه‌بندی دیداری بکار می‌روند. اگرچه نتایج قابل توجهی از اینکار بدست آمده است اما به بیان کلی، هیچیک از این ابزارها در شکل فعلی خود قادر به توسعه‌ی سیستم‌هایی که بتوانند با انسان‌ها رقابت کنند نمی‌باشند.

چندین پژوهشگر از بخش علوم رایانه‌ای دانشکده‌ی علوم رایانه و مهندسی برق در دانشگاه کالیفرنیا-برکلی از یک روش ترکیبی نوآورانه برای گروه‌بندی تصویر استفاده کردند (ژانگ و همکاران ۲۰۰۶). آنها بازشناسی دسته‌ی دیداری را در چارچوب اندازه‌گیری تشابه‌ها یا فواصل ادارکی در نظر گرفتند تا مثال‌هایی از دسته‌ها یا گروه‌ها را ایجاد کنند. رویکردی که این پژوهشگران برای بازشناسی و گروه‌بندی اتخاذ کردند بسیار انعطاف‌پذیر بود که امکان بازشناسی بر اساس رنگ، بافت و به ویژه شکل را فراهم می‌کرد. اگرچه دسته‌بندی‌های نزدیک‌ترین همسایه در این شرایط، طبیعی هستند ولی در خصوص نمونه‌برداری محدود، دارای واریانس بالایی بودند. به شکل دیگر می‌توان از ماشین‌های برداری پشتیبان استفاده کرد ولی محاسبات و بهینه‌سازی به این روش نیز بسیار زمان‌بر است. پژوهشگران یک روش ترکیبی از این دو روش را پیشنهاد کرده‌اند که می‌توان از آن برای حالت‌های چنددسته‌ای استفاده کرد و پیچیدگی محاسباتی چه در زمان آموزش و چه در زمان تست را کاهش داده و در عمل به نتایج عالی دست یافته است. ایده‌ی اصلی این بود که نزدیک‌ترین همسایه‌ها به یک نمونه‌ی کوثری را یافته و یک ماشین برداری پشتیبان محلی را آموزش دهند تا تابع فاصله را بر روی مجموعه همسایه‌ها حفظ کند.

روش پژوهشگران را می‌توان بر روی مجموعه داده‌های بزرگ و چنددسته‌ای اعمال کرد زمانیکه عملکرد بهترین نسبت به نزدیک‌ترین همسایه و ماشین‌های برداری پشتیبان داشته و زمانیکه مسئله غیرقابل ردیابی می‌شود همچنان کارایی خود را حفظ می‌کند. از طیف گسترده‌ای از توابع فاصله استفاده شده و آزمایشات آنها نشان داد که عملکرد بسیار خوبی بر روی تعدادی از مجموعه داده‌های الگو برای گروه‌بندی شکل و بافت (MNIST, USPS, CURET) و تشخیص شیء (Caltech-101) برخوردار بوده‌اند.

گروه دیگری از پژوهشگران (بویمان و ایرانی^۱ ۲۰۰۸) بیان کرده‌اند که دو روش پرکاربرد برای گروه‌بندی و دسته‌بندی تصاویر (یعنی رویکردهای مدل‌محور از نوع شبکه‌های عصبی مصنوعی و ماشین‌های برداری پشتیبان و رویکردهای غیرپارامتری از نوع کی نزدیک‌ترین همسایه) نتایجی پایین‌تر از سطح انتظار را به همراه داشته‌اند. این پژوهشگران همچنین مدعی شده‌اند که یک روش ترکیبی می‌تواند موجب بهبود گروه‌بندی و بازشناسی تصویر شود. آنها یک دسته‌بند بایز ساده مبتنی بر کی نزدیک‌ترین همسایه را پیشنهاد کردند که از فواصل کی نزدیک‌ترین همسایه در فضای توصیف‌گرهای محلی تصویر (نه در فضای تصاویر) استفاده می‌کند. پژوهشگران مدعی شده‌اند که اگرچه روش کی نزدیک‌ترین همسایه بیش از اندازه ساده و کارا بوده و به هیچ‌فاز آموزش/یادگیری نیاز ندارد ولی از لحاظ عملکردی جزو بالاترین دسته‌بندهای پارامتری مبتنی بر یادگیری برای گروه‌بندی تصاویر است. مقایسه‌های تجربی روش آنها بر روی چندین پایگاه داده‌ی گروه‌بندی تصویر نشان داده شده است (Caltech-101, Caltech-101, Graz-01, 256).

¹ Boiman & Irani

علاوه بر گروه‌بندی و بازشناسی تصویر، کی نزدیک‌ترین همسایه با موفقیت برای مسائل دسته‌بندی پیچیده از قبیل بازیابی محتوا^۱ (تشخیص دست‌خط، تحلیل محتوای ویدئویی، زبان بدن و زبان اشاره، بیان ژن^۲ (حوزه‌ی دیگری که در آن، کی نزدیک‌ترین همسایه از عملکرد بهتری نسبت به سایر تکنیک‌های پیشرفته برخوردار است و در واقع ترکیبی از کی نزدیک‌ترین همسایه و ماشین‌های برداری پشتیبان است)، تعامل پروتئین-پروتئین^۳ و پیشگویی ساختار سه بعدی^۴ (یک الگوریتم کی نزدیک‌ترین همسایه مبتنی بر گراف‌ها که اغلب برای پیشگویی ساختار تعامل پروتئین-پروتئین بکار می‌رود) بکار گرفته شده است.

Sources: H. Zhang, A. C. Berg, M. Maire, & J. Malik, "SVMKNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition," *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2, 2006, pp. 2126–2136; O. Boiman, E. Shechtman, & M. Irani, "In Defense of Nearest-Neighbor Based Image Classification," *IEEE Conference on Computer Vision and Pattern Recognition, 2008 (CVPR)*, 2008, pp. 1–8.

سوالاتی برای مورد کاربردی ۵.۴

۱. چرا بازشناسی/دسته‌بندی تصویر یک مسئله‌ی پراهمیت ولی دشوار است؟
۲. از کی نزدیک‌ترین همسایه چگونه می‌توان برای کاربردهای بازشناسی/دسته‌بندی تصویر استفاده کرد؟

سوالاتی برای مرور بخش ۵.۶

۱. ویژگی مهم الگوریتم کی نزدیک‌ترین همسایه چیست؟
۲. مزایا و معایب کی نزدیک‌ترین همسایه در مقایسه با شبکه‌های عصبی مصنوعی و ماشین‌های برداری پشتیبان چه هستند؟
۳. عوامل کلیدی موفقیت برای پیاده‌سازی و اجرای کی نزدیک‌ترین همسایه چه هستند؟
۴. منظور از معیار تشابه (یا فاصله) چیست؟ چگونه می‌توان آن را برای متغیرهای با مقادیر عددی و اسمی بکار برد؟
۵. کاربردهای متداول کی نزدیک‌ترین همسایه چه هستند؟

روش بیز ساده^۵ یک روش دسته‌بندی ساده‌ی مبتنی بر احتمالات است (یک تکنیک یادگیری ماشین که برای مسائل پیش‌بینی از نوع دسته‌بندی بکار می‌رود) که از نظریه‌ی معروف بیز مشتق شده است. در این تکنیک، متغیر خروجی باید دارای مقادیر اسمی باشد. اگرچه متغیرهای ورودی می‌توانند آمیزه‌ای از انواع اسمی و عددی باشند ولی متغیر خروجی را باید پیش از آنکه در یک دسته‌بند بیز بکار گرفته شود از طریق یک نوع روش داده‌بندی به صورت عددی درآورد. واژه‌ی « ساده » از فرض قوی ولی غیرواقعی مستقل بودن متغیرهای ورودی گرفته شده است. به بیان ساده‌تر، یک دسته‌بند بیزی ساده فرض می‌کند که متغیرهای ورودی به یکدیگر

¹ Content retrieval

² Gene expression

³ Protein-to-protein interaction

⁴ 3D structure prediction

⁵ Naïve Bayes

وابستگی ندارند و حضور (یا غیاب) یک متغیر خاص در آمیزه‌ی متغیرهای پیشگو هیچگونه ارتباطی با بود یا نبود سایر متغیرها ندارد.

مدل‌های دسته‌بندی بیزی ساده را به شکلی کاملاً کارا (بسیار سریع و حداقل تلاش محاسباتی) و اثربخش (کاملاً دقیق) در یک محیط یادگیری ماشین نظارت شده توسعه داد. با استفاده از مجموعه‌ای از داده‌های آموزش (نه لزوماً داده‌های حجیم)، پارامترهای مدل‌های دسته‌بندی بیزی ساده را می‌توان با استفاده از روش ماکزیمم احتمال بدست آورد. به عبارت دیگر، به دلیل فرض مستقل بودن، می‌توانیم مدل‌های بیزی ساده را بدون نیاز به اعمال قواعد و الزامات نظریه‌ی بیزی توسعه دهیم. ابتدا اجازه دهید تا نظریه‌ی بیزی را مرور کنیم.

برای اینکه به مزایای روش دسته‌بندی بیز ساده پی ببریم ابتدا باید تعریف پایه‌ی نظریه‌ی بیز و دسته‌بندی‌کننده‌ی بیزی دقیق (بدون فرض ساده‌ی مستقل بودن) را درک کنیم. نظریه‌ی بیز (که آن را *قاعده‌ی بیز* نیز می‌نامند) به احترام ریاضی‌دان انگلیسی توماس بیز (۱۷۶۱-۱۷۰۱) نام‌گذاری شده است که یک فرمول ریاضی برای تعیین احتمالات شرطی می‌باشد. در این فرمول، Y بیانگر فرضیه و X بیانگر داده/شواهد است. این نظریه/قاعده‌ی بسیار محبوب، روشی را برای بازنگری/بهبود احتمالات پیش‌بینی با استفاده از شواهد اضافی ارائه می‌کند.

فرمول زیر ارتباط بین احتمالات دو رویداد Y و X را نشان می‌دهد. $P(Y)$ احتمال پیشین^۱ وقوع Y است. از این لحاظ به آن «پیشین» گفته می‌شود که هیچگونه اطلاعاتی را در مورد X در نظر نمی‌گیرد. $P(X>Y)$ احتمال شرطی وقوع Y با فرض X است که آن را *احتمال پسین*^۲ می‌نامند زیرا از مقدار مشخص شده‌ی X مشتق شده (یا به آن بستگی دارد) است. آن را همچنین *درست‌نمایی*^۳ می‌نامند. $P(X)$ احتمال پیشین وقوع X است که آن را *شواهد*^۴ می‌نامند که به عنوان یک مقدار ثابت نرمال‌کننده عمل می‌کند.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad \rightarrow \quad \text{Posterior} = \frac{\text{Likelihood} * \text{Prior}}{\text{Evidence}}$$

$P(Y|X)$: احتمال پسین Y به شرط X

$P(X|Y)$: احتمال شرطی X به شرط Y (درست‌نمایی)

$P(Y)$: احتمال پیشین Y

$P(X)$: احتمال پیشین X (شواهد یا احتمال غیرشرطی X)

به منظور نمایش عددی این فرمول‌ها، به یک مثال ساده نگاه کنیم. بر اساس گزارش هواشناسی، می‌دانیم که احتمال وقوع بارندگی در روز شنبه ۴۰ درصد می‌باشد. از روی داده‌های تاریخی می‌دانیم که در صورت وقوع بارندگی، احتمال بارش در روز یکشنبه ۱۰ درصد می‌باشد و در صورتیکه روز شنبه بارندگی نداشته باشیم، احتمال بارندگی در روز یکشنبه معادل ۸۰ درصد است. فرض کنیم

¹ Prior probability

² Posterior probability

³ Likelihood

⁴ Evidence

"بارش روز یکشنبه" را رویداد Y و "بارش روز دوشنبه" را رویداد X در نظر بگیریم. بر اساس این توضیح فرمول‌ها را به شکل زیر می‌توانیم بنویسیم:

$$P(Y) = \text{احتمال بارش باران در روز یکشنبه} = 0.40$$

$$P(X|Y) = \text{احتمال بارش باران در روز یکشنبه در صورتیکه روز شنبه باران باریده باشد} = 0.10$$

$$P(X) = \text{احتمال بارش باران در روز دوشنبه} = \text{جمع احتمالات "بارش باران در روزهای شنبه و یکشنبه" و "باریدن باران در روز شنبه و بارش باران در روز یکشنبه"} = 0.40 * 0.10 + 0.60 * 0.80 = 0.52$$

اکنون اگر قرار باشد که احتمال اینکه "آیا روز شنبه باران باریده است؟" را با این شرط که "روز یکشنبه باران باریده است" محاسبه کنیم باید از نظریه‌ی بیز استفاده کنیم. این نظریه به ما اجازه می‌دهد تا احتمال وقوع یک رویداد پیشین را با توجه به نتیجه‌ی یک رویداد پسین محاسبه کنیم.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} = \frac{0.10 * 0.40}{0.52} = 0.0769$$

از این رو، در این مثال، اگر روز یکشنبه باران باریده باشد، ۷.۶۹ درصد احتمال دارد که روز شنبه نیز باران باریده باشد.

دسته‌بندی‌کننده‌ی بیز از نظریه‌ی بیز و بدون فرض استقلال قوی استفاده می‌کند. در یک مسئله‌ی پیش‌بینی از نوع دسته‌بندی، دسته‌بندی‌کننده‌ی بیز به شکل زیر عمل می‌کند: با در نظر گرفتن یک نمونه‌ی جدید، تمامی نمونه‌های دیگری را که دقیقاً شبیه آن هستند را پیدا می‌کند (یعنی تمامی متغیرهای پیشگویی که دارای مقادیر یکسان با نمونه‌ای که دسته‌بندی می‌شود هستند)؛ برچسب گروه یا دسته‌هایی که تمامی آنها به آن تعلق دارند را تعیین می‌کند و نمونه‌ی جدید را در مناسب‌ترین گروه دسته‌بندی می‌کند. اگر مقدار هیچیک از نمونه‌ها با گروه یا دسته‌ی جدید دقیقاً تطابق نداشته باشد آنگاه دسته‌بندی‌کننده موفق نمی‌شود تا نمونه‌ی جدید را به یک برچسب گروه یا دسته تخصیص دهد (چونکه دسته‌بندی‌کننده نمی‌تواند شواهد قوی برای انجام اینکار بیابد). در اینجا به ذکر مثالی بسیار ساده می‌پردازیم. ما با استفاده از دسته‌بندی‌کننده‌ی بیز تصمیم می‌گیریم که در موقعیت زیر گلف بازی کنیم یا خیر (آفتابی، درجه حرارت بالا، رطوبت بالا و هوای ساکن). جدول ۵.۴ نمونه‌های تاریخی را که برای ترسیم مشخصه‌های فرایند دسته‌بندی نیاز داریم را ارائه کرده است. بر اساس داده‌های تاریخی، به نظر می‌رسد که سه نمونه با موقعیت ما انطباق داشته باشند (نمونه‌های ۱، ۶ و ۷ که در جدول ۵.۴ مشخص شده‌اند). از این سه نمونه، دو مورد دارای برچسب «خیر» و یک مورد دارای برچسب «بله» هستند. چونکه اکثریت نمونه‌های منطبق با برچسب «خیر» مشخص شده‌اند، موقعیت/نمونه‌ی جدید نیز با برچسب «خیر» مشخص خواهد شد. اکنون موقعیتی را در نظر بگیرید که هوا آفتابی، بسیار گرم، رطوبت بالا و وزش باد است. چونکه هیچ نمونه‌ای وجود ندارد که با این مقدار منطبق باشد، دسته‌بندی‌کننده‌ی بیز هیچ نتیجه‌ای را به ما نمی‌دهد. برای یافتن تطبیق‌های دقیق، به یک مجموعه کلان‌داده نیاز داریم. حتی برای مجموعه داده‌های بزرگ، با افزایش تعداد متغیرهای پیشگو، احتمال اینکه نتوانیم تطبیق دقیقی را بیابیم نیز به میزان زیادی افزایش می‌یابد. زمانیکه مجموعه داده و تعداد متغیرهای پیشگو بیشتر می‌شوند، زمان جستجوی موارد تطبیق دقیق نیز بالا می‌رود. تمامی اینها دلایلی هستند که چرا از دسته‌بندی‌کننده‌ی بیز ساده که مشتق از دسته‌بندی‌کننده‌ی بیز است اغلب در تحلیل‌های پیشگوانه و روش‌های داده‌کاوی استفاده می‌شود. در

دسته‌بندی‌کننده‌ی بیز ساده، شرط پیدا کردن یک انطباق دقیق مطرح نیست. دسته‌بندی‌کننده‌ی بیز ساده هر متغیر پیشگو را به عنوان یک عامل مستقل در پیش‌بینی متغیر خروجی در نظر گرفته و از این رو، سودمندی و کاربردی بودن آن را به عنوان یک ابزار پیش‌بینی از نوع دسته‌بندی به میزان قابل توجهی افزایش می‌دهد.

جدول ۵.۴ مجموعه داده‌های نمونه برای روش‌های پیش‌بینی از نوع دسته‌بندی

متغیر خروجی (Y)	متغیرهای ورودی (X)				
بازی گلف	باد	رطوبت	دما	چشم‌انداز	شماره نمونه
خیر	خیر	بالا	گرم	آفتابی	۱
بله	خیر	بالا	گرم	ابری	۲
بله	خیر	نرمال	خنک	بارانی	۳
خیر	بله	نرمال	خنک	بارانی	۴
خیر	بله	نرمال	خنک	ابری	۵
خیر	خیر	بالا	گرم	آفتابی	۶
بله	خیر	بالا	گرم	آفتابی	۷
بله	خیر	نرمال	ملایم	بارانی	۸
بله	بله	نرمال	ملایم	آفتابی	۹

همانند سایر روش‌های یادگیری ماشین، بیز ساده نیز از یک فرایند جاری‌سازی/امتیازدهی و توسعه مدل دومرحله‌ای استفاده می‌کند: (۱) آموزش که در آن پارامترها/مدل تخمین زده می‌شود و (۲) تست که در آن دسته‌بندی/پیش‌بینی بر روی مورد جدید انجام می‌شود. فرایند به شرح زیر انجام می‌پذیرد:

فاز آموزش

گام ۱. داده‌ها را گرفته، پاکسازی کرده و آنها را در قالب یک فایل تخت سازماندهی کنید (یعنی ستون‌ها به عنوان متغیرها و ردیف‌ها به عنوان موردها).

گام ۲. اطمینان حاصل کنید که متغیرها به صورت اسمی باشند و در غیراینصورت (یعنی اگر هر یک از متغیرها از نوع عددی/پیوسته بود)، باید متغیرهای عددی را با استفاده از تکنیک گسسته‌سازی به متغیرهای اسمی تبدیل کرد.

گام ۳. احتمال پیشین تمامی برچسب دسته‌ها را برای متغیر وابسته محاسبه کنید.

گام ۴. درست‌نمایی تمامی متغیرهای پیشگو و مقادیر ممکن آنها با توجه به متغیر وابسته را محاسبه کنید. در خصوص داده‌هایی که مقادیر مختلط دارند (رسته‌ای یا پیوسته)، درست‌نمایی هر متغیر (احتمال شرطی) با روش مناسبی که برای آن نوع متغیر خاص بکار می‌رود محاسبه می‌شود. درست‌نمایی متغیرهای پیشگوی اسمی و عددی به شکل زیر محاسبه می‌شود:

- برای متغیرهای رسته‌ای، درست‌نمایی (احتمال شرطی) به صورت کسری ساده از نمونه‌های آموزش برای مقدار متغیر با توجه به متغیر وابسته تخمین زده می‌شود.
- برای متغیرهای عددی، درست‌نمایی با (۱) محاسبه‌ی میانگین و واریانس هر متغیر پیشگو برای هر مقدار متغیر وابسته (یعنی دسته یا گروه) و سپس (۲) محاسبه‌ی درست‌نمایی با استفاده از فرمول زیر محاسبه می‌شود:

$$P(x = v|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$

اغلب مواقع، متغیرهای ورودی/مستقل پیوسته/عددی با استفاده از روش‌های مناسب به صورت گسسته درآورده شده و سپس از روش تخمین متغیر رسته‌ای برای محاسبه‌ی احتمالات شرطی (پارامترهای درست‌نمایی) استفاده می‌شود. در صورتیکه این روش به درستی انجام شود، مدل‌های بیز ساده‌ی بهتری را نتیجه می‌دهد.

با استفاده از دو مجموعه پارامترهای تولید شده در گام‌های ۳ و ۴ در فاز آموزش، هر نمونه‌ی جدید را می‌توان با استفاده از فرمول زیر به یک برچسب دسته یا گروه تخصیص داد:

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}$$

$$P(C|F_1, \dots, F_n) = \frac{P(C)P(F_1, \dots, F_n|C)}{P(F_1, \dots, F_n)}$$

از آنجائیکه مخرج این کسر (برای تمامی برچسب دسته‌ها) ثابت است، می‌توانیم آن را از فرمول حذف کنیم و به فرمول ساده‌تری برسیم که فرمول احتمال مشترک است.

$$\text{classify}(f_1, \dots, f_n) = \underset{c}{\operatorname{argmax}} p(C = c) \prod_{i=1}^n p(F_i = f_i | C = c)$$

این مثال ساده‌ای است که این محاسبات را نمایش می‌دهد. در این مثال، ما از همان داده‌های جدول ۵.۴ استفاده کرده‌ایم. هدف این است که مورد زیر را دسته‌بندی کنیم: هوای آفتابی، درجه حرارت بالا، رطوبت بالا و باد نمی‌وزد. این متغیر وابسته در چه دسته یا گروهی قرار می‌گیرد؟ (بازی = بله یا خیر).

از داده‌ها متوجه می‌شویم که $\text{Prior}(Yes) = 5/9$ و $\text{Prior}(No) = 4/9$. برای متغیر چشم‌انداز، درست‌نمایی‌ها به صورت $\text{Likelihood}(No/Sunny) = 2/3$ ، $\text{Likelihood}(No/Overcast) = 1/2$ ، $\text{Likelihood}(No/Rainy) = 1/3$ محاسبه می‌شود.

مقادیر درست‌نمایی سایر متغیرها (دما، رطوبت و باد) را می‌توان به همین شکل محاسبه/تعیین کرد. نتایج در جدول ۵.۵ نشان داده شده‌اند.

جدول ۵.۵ محاسبات دسته‌بندی بیز ساده				درستی نتایج
کسر (%)		نسبت		
بازی = خیر	بازی = بله	بازی: خیر	بازی: بله	
۰.۶۷	۰.۳۳	۲/۳	۱/۳	چشم‌انداز = آفتابی
۰.۵۰	۰.۵۰	۲/۴	۲/۴	دما = گرم
۰.۵۰	۰.۵۰	۲/۴	۲/۴	رطوبت = بالا
۰.۳۳	۰.۶۷	۲/۶	۴/۶	باد = خیر
۰.۴۴	۰.۵۶	۴/۹	۵/۹	پیشین
۰.۰۲۵	۰.۰۳۱	حاصلضرب		
۰.۰۵۶	۰.۰۷۰	تقسیم بر شواهد		

بر اساس نتایج جدول ۵.۵، جواب $\text{Play Golf} = \text{Yes}$ خواهد بود زیرا مقدار عددی آن بزرگتر است (۰.۰۳۱). اگر قرار باشد از فرمول پسین کامل برای دو برچسب دسته یا گروه استفاده کنیم که نیازمند این بود که مخرج کسر را در نظر بگیریم، می‌بینیم که مقدار ۰.۰۷ برای «بله» و ۰.۰۵۶ برای «خیر» بدست می‌آید. از آنجائیکه مخرج کسر برای تمامی برچسب گروه یا دسته‌ها مشترک است، خروجی عددی را تغییر می‌دهد ولی تخصیص دسته یا گروه تغییری نمی‌کند.

اگرچه روش بیز ساده امروزه چندان در پروژه‌های تحلیل پیشگویانه کاربرد ندارد (به دلیل عملکرد نسبتاً ضعیف آن در پیش‌بینی در طیف گسترده‌ای حوزه‌های کاربردی)، یکی از انشعابات آن به نام شبکه بیزی به طرز شگف‌آوری بین دانشمندان علم داده‌ها محبوبیت یافته است.

مورد کاربردی ۵.۵ مثال جالبی از بکارگیری تعداد زیادی تکنیک تحلیل پیشگویانه برای تعیین شرایط متغیر بیماران مبتلا به کرون را به منظور مدیریت بهتر این بیماری مژمن به تصویر کشیده است. در کنار روش بیز ساده چندین روش آماری و یادگیری ماشین توسعه یافته، تست و مقایسه شده‌اند. بهترین مدل از لحاظ عملکرد برای توضیح اهمیت رتبه‌بندی شده‌ی تمامی متغیرهای مستقل در پیش‌گویی میزان پیشرفت بیماری بکار گرفته شده است.

مقدمه و انگیزش

بر اساس آمار بنیاد کولیت و کرون^۱ (crohnscolitisfoundation.org) ۱.۶ میلیون آمریکایی به بیماری التهابی روده که شامل بیماری کرون و کولیت اولسراتیو می‌باشد، مبتلا هستند. بیماری کرون موجب التهاب مزمن و آسیب دستگاه گوارش می‌شود. همچنین می‌تواند بر هر بخشی از سیستم گوارشی تأثیر بگذارد. علل بروز این بیماری کاملاً مشخص نیست ولی پژوهش‌ها بیانگر این است که می‌تواند به دلیل ترکیبی از عوامل از جمله ژنتیک، سیستم ایمنی و شرایط محیطی بروز کند. سیستم‌هایی که قادر به تشخیص میزان پیشرفت بیماری یا تشخیص زودهنگام بیماری باشند می‌توانند در بهره‌برداری بهینه از منابع نهادهای مراقبت سلامت کمک کرده و منجر به نتایج بهتری برای بیماران شوند. هدف از این مطالعه‌ی موردی، استفاده از سوابق پزشکی الکترونیکی^۲ برای پیش‌بینی و توضیح التهاب بیماران مبتلا به کرون بوده است.

متدولوژی

داده‌های بکار رفته در این مطالعه از یکی از بزرگ‌ترین پایگاه داده‌های EMR کشور یعنی مرکز آمارهای سلامت سررند بدست آمده‌اند. این مرکز میزبان اطلاعات غنی و متنوع مرتبط با بیماران، محیط‌های مراقبت سلامت، هزینه‌ها، انواع بازپرداخت‌ها و داده‌های مربوط به نسخه‌های تجویز شده از چندین ارائه دهنده‌ی خدمات مراقبت سلامت و بیمارستان در ایالات متحده می‌باشد. داده‌های ذخیره شده در این پایگاه داده شامل داده‌های سطح بیمار است که در زمان مراجعه‌ی بیمار به بیمارستان، مراکز اورژانس، کلینیک‌های تخصصی، کلینیک‌های عمومی و آسایشگاه‌ها ثبت شده است. این پایگاه داده حاوی داده‌های طولی بیماران بدون امکان شناسایی شخص و با استامپ زمانی بوده‌اند. پایگاه داده مطابق با جدول ۵.۶ در قالب جداول داده‌ها سازماندهی شده است. یک جریان فرایندی سطح بالا از متدولوژی پژوهش در شکل ۵.۱۲ نشان داده شده است. اگرچه در این دیاگرام جریان فرایند، جزئیات هر گام نمایش داده نشده است ولی یک دیدگاه سطح بالا از توالی گام‌هایی که در مطالعه‌ی مدلسازی پیشگويانه‌ی جاری با استفاده از داده‌های EMR انجام شده را ارائه کرده است. سه نوع مدل نشان داده شده در دیاگرام بر اساس عملکرد نسبتاً بهترشان در مقایسه با سایر روش‌های یادگیری ماشین مثل بیز ساده، نزدیک‌ترین همسایه و شبکه‌های عصبی انتخاب شدند. جزئیات گام‌های متوازن‌سازی داده‌ها و استانداردسازی داده‌ها در مقاله‌ی ردی، دیلن و آگراوال (۲۰۱۸) توضیح داده شده است.

نتایج

نتایج پیش‌بینی با استفاده از مجموعه داده‌های تست و اجرای ۱۰ مرتبه‌ای روش اعتبارسنجی متقاطع ۱۰ مرتبه‌ای تولید شده‌اند. عملکرد هر مدل با استفاده از سنجه‌ی AUC مورد ارزیابی قرار گرفت که سنجه‌ی عملکردی برتر نسبت به سنجه‌ی دقت پیش‌بینی بوده است زیرا منحنی مشخصه‌ی عملکرد سیستم که برای تولید AUC بکار رفته بود به مقایسه‌ی عملکرد دسته‌بندی‌کننده در بین کل توزیع دسته‌ها و هزینه خطاها پرداخته و از این رو به طور گسترده‌ای به عنوان سنجه‌ی عملکردی در کاربردهای یادگیری

¹ Crohn's and Colitis Foundation

² Electronic Medical Records (EMR)

ماشین پذیرفته شده است. میانگین مقدار AUC برای ۱۰ مرتبه اجرای روش اعتبارسنجی متقاطع ۱۰ مرتبه‌ای برای سه نوع مدل نهایی - رگرسیون لجستیک، رگرسیون عادی و ماشین‌های گرادیان تقویتی^۱ محاسبه شد (جدول ۵.۷).

پس از تولید AUC برای ۱۰۰ مدل، پژوهشگران یک آزمون تحلیل واریانس تعقیبی^۲ و یک آزمون تفاضل معنادار واقعی توکی^۳ را برای مقایسه‌ی چندین آزمون به منظور تعیین طبقه‌بندی‌کننده‌ای که عملکرد آن بر اساس AUC متفاوت از سایر روش‌ها بوده است انجام دادند. نتایج آزمون نشان داد که میانگین AUC برای رگرسیون عادی و رگرسیون لجستیک تفاوت معناداری ندارد. با وجود این، AUC حاصل از رگرسیون‌های عادی و لجستیک دارای تفاوت معناداری با مدل ماشین‌های گرادیان تقویتی بوده‌اند (جدول ۵.۸).

اهمیت نسبی متغیرهای مستقل با افزودن کل مقدار کاهش شاخص جینی برای یک متغیر پیشگوی خاص محاسبه شده و میانگین آن به ازای تمامی درخت‌های مشخص شده در پارامتر تنظیم‌کننده‌ی ماشین‌های گرادیان تقویتی (که در این پژوهش ۱۰۰۰ درخت وجود داشته است) محاسبه شده است. این کاهش میانگین شاخص جینی به مقیاس ۱۰۰-۰ نرمال شده است که هرچه عدد بالاتر باشد به معنی قوی‌تر بودن متغیر پیشگو می‌باشد. نتایج اهمیت متغیر در شکل ۵.۱۳ نشان داده شده‌اند که هیچ متغیر پیشگویی به تنهایی وجود نداشته بلکه ترکیبی از متغیرها بوده که پیشگویی را انجام داده‌اند. محل بیماری کرون در زمان تشخیص مثل روده‌ی کوچک و روده‌ی بزرگ؛ پارامترهای آزمایشگاهی در خط مبنا مثل شمارش گلبول‌های سفید؛ مقدار متوسط هموگلوبین؛ حجم متوسط گلبولی؛ سدیم، شمارش گلبول‌های قرمز، حجم متوسط پلاکت؛ کراتینین؛ هماتوکریت و هموگلوبین جزو قوی‌ترین متغیرهای پیشگو بودند.

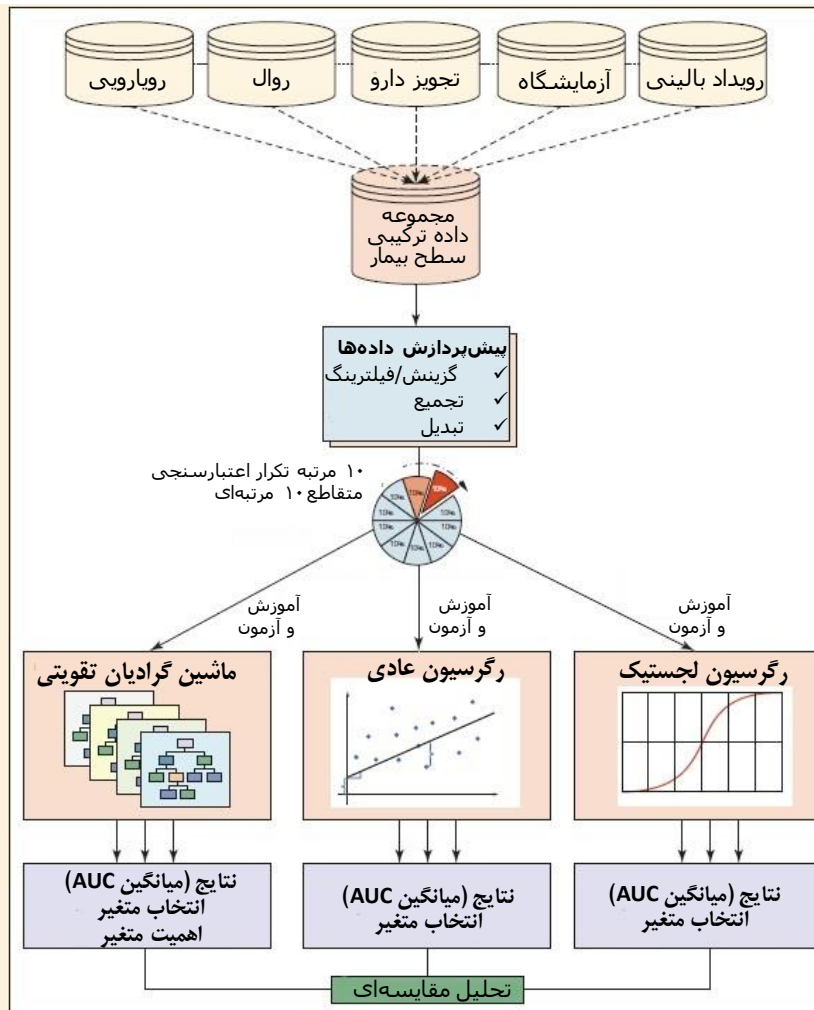
جدول ۵.۶ فراداده‌های جداول استخراج شده پایگاه داده‌های EMR

مجموعه داده	شرح
روبارویی	روبارویی شامل جمع‌شناختی، صدور صورتحساب، شرایط مراقبت سلامت، نوع پرداخت‌کننده و غیره
تجویز دارو	درخواست‌های دارو که توسط ارائه دهنده‌ی خدمات مراقبت سلامت ارسال شده‌اند
آزمایشگاه	داده‌های آزمایشگاهی شامل هماتولوژی، ترکیبات خون و تحلیل ادرار
رویداد بالینی	داده‌های رویدادهای بالینی حاوی اطلاعات مربوط به انواع مختلف سنج‌ها از قبیل شاخص توده بدنی، وضعیت استعمال دخانیات، نمره‌ی درد و غیره
روال کار	روال‌های بالینی انجام شده بر روی بیمار

¹ Gradient Boosting Machines

² Post hoc analysis of variance (ANOVA)

³ Tukey's Honest Significant Difference (HSD)



شکل ۵.۱۲ نمودار جریان فرایند گام‌های سطح بالای پژوهش داده‌کاوی

یکی از قوی‌ترین متغیرهای پیشگوی جمعیت‌شناختی که شدت التهاب را دو برابر می‌کند، سن بیمار است. سایر متغیرهای مرتبط با محیط و شرایط مراقبت سلامت و رویارویی از جمله حجم رختخواب‌های بیمارستان، اولویت تشخیص و منطقه نیز از توانایی پیشگویانه برخوردار بودند. اکثریت پژوهشگران در حوزه‌ی بیمار کرون، محل وقوع بیماری در بدن، سن بیمار در زمان تشخیص، وضعیت استعمال دخانیات، نشانگرهای بیولوژیکی و فاکتور نکرور دهنده‌ی تومور را عوامل پیش‌بینی پاسخ در برابر درمان تشخیص دادند که برخی از آنها پیش‌بینی‌کننده‌ی شدت التهاب نیز بودند.

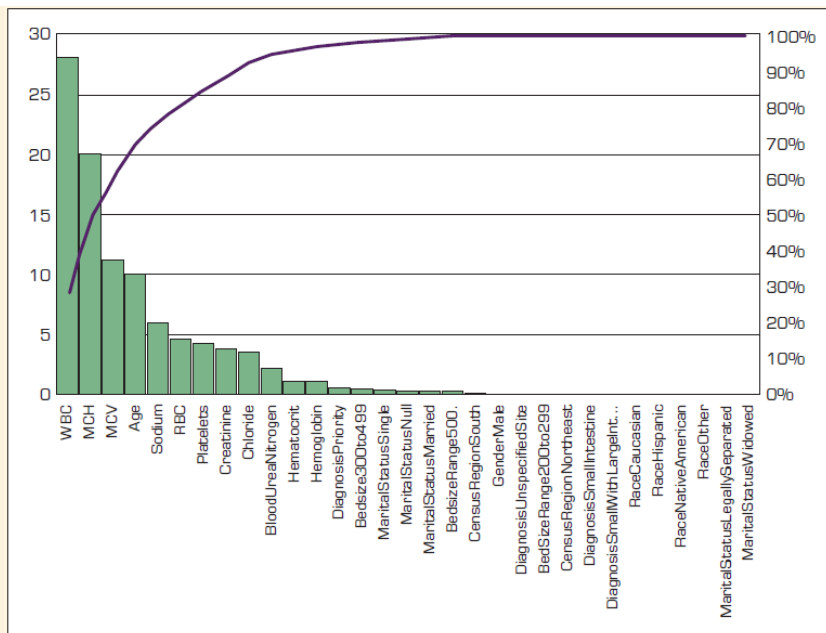
این مطالعه توانست نشان دهد که بیماری را می‌توان با استفاده از ابزارهای پشتیبان تصمیم که برای پیش‌بینی وضعیت التهاب در آینده به تحلیل‌های پیشرفته متکی هستند را می‌توان به موقع مدیریت کرد که پس از آن امکان مداخلات پزشکی متناسب با آن پیش‌بینی را فراهم می‌کند. به کمک این اطلاعات، ارائه‌دهندگان خدمات مراقبت سلامت می‌توانند با مداخله‌ی زودهنگام و اتخاذ اقدامات درمانی خاص برای آن بیمار، وضعیت وی را بهبود بخشند.

جدول ۵.۷ مقدار AUC برای هر بار تکرار آزمایش برای سه مدل

ماشین‌های گرادیان تقویتی	رگرسیون عادی	رگرسیون لجستیک	مرتبه‌ی تکرار
۰.۹۳۹۳	۰.۸۲۶۷	۰.۷۹۲۹	۱
۰.۹۲۶۲	۰.۸۰۷۸	۰.۷۸۷۸	۲
۰.۹۳۶۹	۰.۸۱۴۵	۰.۸۰۸۰	۳
۰.۹۱۲۴	۰.۸۴۸۷	۰.۸۴۶۱	۴
۰.۹۴۱۴	۰.۸۲۸۱	۰.۸۲۴۳	۵
۰.۸۸۷۸	۰.۸۵۴۳	۰.۷۶۸۱	۶
۰.۹۳۵۶	۰.۸۱۵۴	۰.۸۱۶۷	۷
۰.۹۳۳۰	۰.۸۱۷۶	۰.۸۱۷۴	۸
۰.۹۴۶۷	۰.۸۲۸۱	۰.۸۴۵۲	۹
۰.۹۲۳۰	۰.۸۲۹۴	۰.۸۰۵۰	۱۰
۰.۹۲۸۲	۰.۸۲۷۱	۰.۸۱۳۱	میانگین AUC
۰.۹۳۴۳	۰.۸۲۷۴	۰.۸۱۶۷	میانگین AUC

جدول ۵.۸ آنالیز واریانس با چندین مقایسه با استفاده از آزمون توکی

نوع مدل	تعداد مشاهدات	میانگین AUC	دسته‌بندی توکی
ماشین گرادیان تقویتی	۱۰۰	۰.۹۲۸	A
رگرسیون عادی	۱۰۰	۰.۸۲۷	B
رگرسیون لجستیک	۱۰۰	۰.۸۱۲	B



شکل ۵.۱۳ اهمیت متغیر نسبی مدل ماشین گرادیان تقویتی

Source: B. K. Reddy, D. Delen, & R. K. Agrawal, "Predicting and Explaining Inflammation in Crohn's Disease Patients Using Predictive Analytics Methods and Electronic Medical Record Data," *Health Informatics Journal*, 2018.

سوالاتی برای مورد کاربردی ۵.۵

۱. بیماری کرون چیست و چرا اهمیت دارد؟
۲. بر اساس یافته‌های این مورد کاربردی، درباره‌ی استفاده از تحلیل‌ها در مدیریت بیماری کرون چه می‌توانید بگویید؟
۳. برای پیش‌بینی بهتر نتایج این بیماری مزم از چه روش‌ها و مجموعه داده‌هایی می‌توان استفاده کرد؟

سوالاتی برای مرور بخش ۵.۷

۱. ویژگی مهم الگوریتم بیز ساده چیست؟ منظور از « ساده » در این الگوریتم چیست؟
۲. مزایا و معایب الگوریتم بیز ساده در مقایسه با سایر روش‌های یادگیری ماشین چه هستند؟
۳. در الگوریتم بیز ساده از چه نوع داده‌هایی می‌توان استفاده کرد؟ چه نوع پیش‌بینی‌هایی را می‌توان از آن بدست آورد؟
۴. فرایند توسعه و تست یک دسته‌بندی‌کننده‌ی بیز ساده چیست؟

شبکه‌های بیزی اولین بار در مقاله‌ای توسط جودا پرل به عنوان «پشتیبان انتشار خودکار و چندجهته‌ی شواهدی که به سرعت به سمت یک تعادل جهان شمول همگرایی می‌یابند» تعریف شدند (پرل ۱۹۸۵). بعدها، پرل با تداوم کارهایش در این حوزه موفق به کسب جایزه‌ی پرافتخار ای‌سی‌ام تورینگ به خاطر سهمی که در رشته‌ی هوش مصنوعی و توسعه‌ی شبکه‌های بیزی داشته گردید. با این موفقیت، شبکه‌های بیزی بیش از پیش مورد توجه عموم قرار گرفت و به یک پارادایم نوین در هوش مصنوعی، تحلیل‌های پیشگویانه و علم داده‌ها تبدیل شد.

شبکه‌های بیزی، ابزاری قدرتمند برای نمایش ساختار وابستگی در یک قالب گرافیکی، صریح و شهودی می‌باشد. این شبکه‌ها بیانگر حالات مختلف یک مدل چندمتغیره و ارتباطات احتمالی آنها هستند. از لحاظ نظری، یک سیستم را می‌توان به کمک شبکه‌های بیزی مدل‌سازی کرد. در یک مدل فرضی، برخی حالات ممکن است با تواتر بیشتری رخ بدهند در حالیکه حالات دیگری نیز حضور دارند. به عنوان مثال، اگر یک دانشجوی سال اول برای فصل بعدی دانشگاه ثبت نام نکرده باشد، احتمال اینکه بتواند مساعده‌ی مالی بگیرد پایین است که این نشان دهنده‌ی وجود رابطه بین این دو متغیر است. در چنین شرایطی است که احتمالات شرطی (نظریه‌ی زیربنایی شبکه‌های بیزی) در تحلیل و تعیین موقعیت وارد صحنه می‌شود.

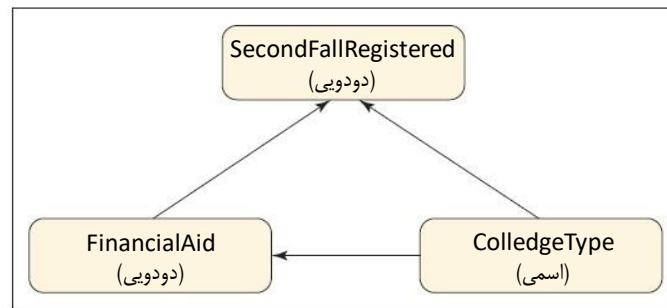
شبکه‌های بیزی به خاطر توانایی در ثبت و استنتاج موقعیت‌ها و تعاملات پیچیده، غیرخطی و نسبتاً نامطمئن به مدل‌های گرافیکی احتمالی محبوبی تبدیل شده‌اند (کولر و فریدمن ۲۰۰۹). اگرچه مشخصه‌های نظری مبتنی بر احتمالات شبکه‌های بیزی آنها را سریعاً به موضوعی جذاب برای پژوهش دانشگاهی و به ویژه مطالعه‌ی روابط علت و معلولی تبدیل کرد ولی کاربرد آنها در علم داده‌ها و تحلیل‌های کسب‌وکار همچنان جدید و نوپا است. به عنوان مثال، پژوهشگران اخیراً مدل‌های شبکه‌های بیزی مبتنی بر تحلیل داده‌ها را در حوزه‌هایی توسعه داده‌اند که شامل پیش‌بینی و درک مشکلاتی از قبیل بقاء پوست در پیوند کلیه (توپوز و همکاران ۲۰۱۸)، پیش‌بینی خرابی در صنعت ریلی که ناشی از شرایط آب و هوایی باشد (وانگ و همکاران ۲۰۱۷)، پیش‌بینی کلاهبرداری در مواد غذایی (بوزمبارک و همکاران ۲۰۱۶) و تشخیص بیماری‌ها (می‌فروید و همکاران ۲۰۰۹) می‌باشد.

اساساً مدل شبکه‌های بیزی یک گراف غیرچرخه‌ای جهت‌دار است که گره‌های آن بیانگر متغیرها و کمان‌ها نشان دهنده‌ی وابستگی‌های شرطی بین متغیرها و مقادیر احتمالی آنها می‌باشند (پرل ۲۰۰۹). یک مثال ساده از شبکه‌های بیزی در مورد کاربردی ۳.۲ به آن پرداخته شد. برای جزئیات بیشتر در این خصوص لطفاً این مورد کاربردی را مجدداً مطالعه نمایید. فرض کنید که هدف ما این بوده است که از برخی داده‌ها/اطلاعات مربوط به دانشجو مثل (۱) نوع کالج اعلامی (تعداد حالات/گزینه‌های موجود برای کالج‌های بالقوه) و (۲) آیا دانشجو در ترم اول دانشگاه مساعده‌ی مالی دریافت کرده یا خیر برای پیش‌بینی این موضوع بوده که آیا دانشجوی سال اول در دانشگاه باقی خواهد ماند یا اخراج خواهد شد استفاده کنیم. هر دو مورد فوق را می‌توان با استفاده از داده‌های تاریخی به شکل احتمالات بیان نمود. ممکن است تصور کنید که پیوندهای علت و معلولی بین این سه متغیر وجود دارد یعنی نوع کالج و کمک مالی که مربوط به این می‌شود که آیا دانشجو برای ترم دوم به دانشگاه مراجعه خواهد کرد یا خیر و منطقی است که فکر کنیم که برخی کالج‌ها سابقه‌ی طولانی‌تری در ارائه مساعده‌های مالی نسبت به سایر کالج‌ها دارند (شکل ۵.۱۴).

جهت پیوندها در گراف‌های شبکه‌های بیزی متناظر با وابستگی‌های احتمالی یا شرطی بین هر دو متغیر می‌باشد. محاسبه‌ی احتمالات شرطی واقعی با استفاده از داده‌های تاریخی به پیش‌بینی و درک حفظ و نگهداشت دانشجویان با استفاده از دو متغیر «مساعده‌ی مالی» و «نوع کالج» کمک می‌کند. سپس از چنین شبکه‌ای می‌توان برای پاسخگویی به سوالات زیر استفاده کرد:

- آیا نوع کالج مورد نظر، کالج « مهندسی » است؟
- احتمال اینکه دانشجو در ترم پاییز بعدی ثبت نام کند چقدر است؟
- مساعده‌ی مالی چگونه بر نتیجه تأثیر خواهد گذاشت؟

ساخت مدل‌های احتمالی از جمله شبکه‌های بیزی پیچیده برای مسائل/موقعیت‌های پیچیده‌ی دنیای واقعی با استفاده از داده‌های تاریخی می‌توانند به پیش‌بینی آنچه احتمال دارد اتفاق بیوفتد در حالیکه رویداد دیگری رخ داده است کمک می‌کنند. اساساً، شبکه‌های بیزی معمولاً سعی دارند تا ارتباطات بینابین متغیرها (هم متغیرهای ورودی و هم خروجی) را با استفاده از یک ساختار احتمالی که اغلب توزیع مشترک^۱ نامیده می‌شود نمایش دهند. توزیع مشترک را می‌توان به صورت یک جدول که شامل تمامی ترکیب‌های ممکن است حالات (مقادیر متغیر) در یک مدل مفروض است نشان داد. برای مدل‌های پیچیده، چنین جدولی می‌تواند بسیار بزرگ شود زیرا یک مقدار احتمال را برای تمامی ترکیب‌های حالات ذخیره می‌کند. برای پیشگیری از چنین مشکلی، شبکه‌های بیزی تمامی گره‌های موجود در مدل را به یکدیگر متصل نمی‌کنند بلکه فقط گره‌هایی که به یکدیگر وصل می‌کنند که دارای وابستگی شرطی و/یا منطقی با یک احتمال خاص هستند که موجب کاهش چشمگیر محاسبات می‌شود.



شکل ۵.۱۴ ترسیمی ساده از ارتباطات علت و معلولی در نگهداشت دانشجو

توزیع‌های احتمال پیچیده و طبیعی را می‌توان با استفاده از فرمول وابستگی شرطی شبکه‌های بیزی به شکلی نسبتاً فشرده نمایش داد. در فرمول زیر، هر x_i بیانگر یک متغیر و Pa_{x_i} بیانگر والدیان آن متغیر است. با استفاده از اینها، قاعده‌ی زنجیره‌ی شبکه بیزی را می‌توان به شکل زیر بیان کرد (کولر آند فریدمن ۲۰۰۹):

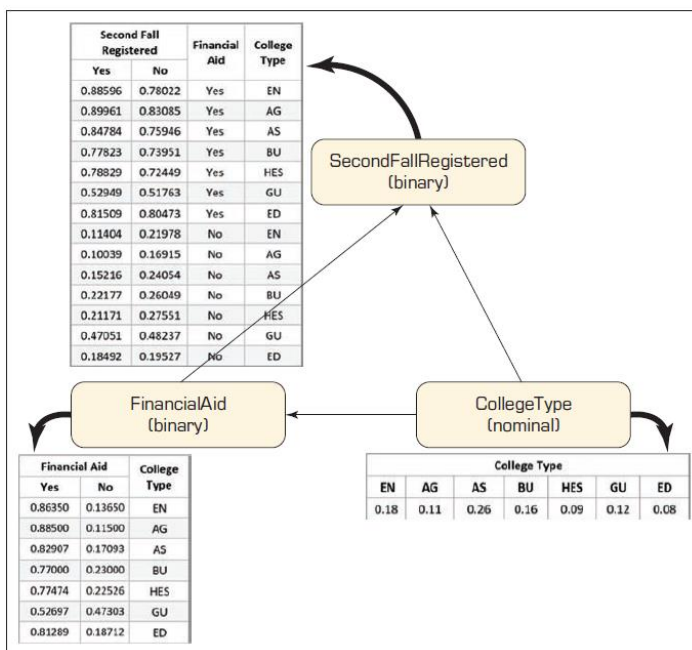
$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | Pa_{x_i})$$

برای مثال، شبکه‌ای ساده را برای مسئله‌ی پیش‌بینی حفظ و نگهداشت دانشجویان ایجاد می‌کنیم. به خاطر داشته باشید که مسئله‌ی ما استفاده از داده‌ها/اطلاعات از سوابق دانشجو (یعنی نوع کالج اعلام شده و اینکه آیا دانشجو در ترم اول مساعده‌ی مالی دریافت کرده یا خیر) برای پیش‌بینی این موضوع است که آیا دانشجو برای ترم دوم نیز در دانشگاه باقی خواهد ماند یا اینکه از دانشگاه اخراج خواهد شد. مدل گرافیکی شبکه‌ی بیزی ساخته شده در شکل ۵.۱۵ از ارتباطات و احتمالات شرطی بین هر سه گره را نشان می‌دهد.

¹ Joint distribution

دو روش متداول برای ساخت شبکه‌های بیزی وجود دارد: (۱) به روش دستی و به کمک یک فرد خبره در موضوع و (۲) به روش تحلیلی با یادگیری ساختار شبکه از روی داده‌های تاریخی با استفاده از روش‌های ریاضی پیشرفته. ساخت یک شبکه به صورت دستی حتی برای یک شبکه‌ی نه چندان بزرگ نیازمند صرف ساعت‌های وقت یک فرد خبره و یک مهندس ماهر و مطلع است. با افزایش حجم و اندازه‌ی شبکه، زمان صرف شده توسط مهندس و فرد خبره نیز به صورت نمایی افزایش می‌یابد. در برخی موارد، یافتن فردی خبره و مطلع برای یک حوزه‌ی خاص کار بسیار دشواری است. حتی در صورتیکه چنین شخصی وجود داشته باشد، ممکن است نتواند زمان کافی را برای ساخت مدل اختصاص داده و/یا اینکه دانش و اطلاعات وی به صورت ضمنی و پنهان باشد و به عنوان یک منبع دانشی نتوان استفاده چندانی از وی کرد. از این رو، در اکثر مطالعات قبلی، تکنیک‌های مختلفی توسعه یافته و ارائه شده‌اند که می‌توان از آنها برای یادگیری خودکار ساختار شبکه از روی داده‌ها استفاده کرد.

یکی از روش‌های اولیه که برای یادگیری ساختار خودکار شبکه‌ها از روی داده‌ها بکار گرفته شد، روش بیز ساده بود. روش دسته‌بندی بیز ساده، یک مدل احتمالی ساده است که وابستگی شرطی بین تمامی متغیرهای پیشگو و یک متغیر هدف/دسته‌ی مفروض را برای یادگیری ساختار در نظر می‌گیرد. الگوریتم دسته‌بندی مبتنی بر قاعده‌ی بیز است که بر اساس آن، احتمال مقدار هدف/دسته برای هر متغیر ویژگی مفروض محاسبه شده و سپس بالاترین مقدار پیش‌بینی شده برای آن ساختار انتخاب می‌شود.



شکل ۵.۱۵ جداول احتمال شرطی برای دو متغیر پیشگو و یک متغیر هدف

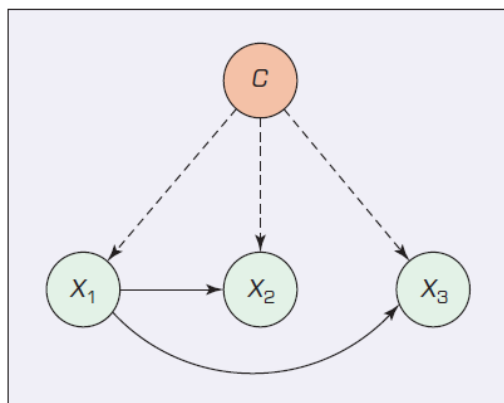
یک روش جدیدتر و محبوب‌تر برای یادگیری ساختار شبکه، بیزین ساده‌ی درخت‌افزودنی^۱ نام دارد. این روش، نسخه‌ی به‌روزآوری شده‌ی دسته‌بندی‌کننده‌ی بیز ساده است که از ساختار درختی برای تقریب تعاملات بین متغیرهای پیشگو و متغیر هدف استفاده می‌کند (فریدمن، گایجر و گلدزمیت ۱۹۹۷). در ساختار مدل درخت‌افزودنی، متغیر دسته‌ی فاقد والد بوده و هر یک از متغیرهای پیشگو دارای

¹ Tree Augmented Naïve Bayes (TAN)

یک متغیر دسته به عنوان والد به همراه حداکثر یک متغیر پیشگوی دیگر (یعنی صفت) هستند (شکل ۵.۱۶). بنابراین، کمان بین این دو متغیر نشان دهنده‌ی یک ارتباط جهت‌دار و علی و معلولی بین آنها است. نمایش رسمی والد‌ها برای متغیر x_i به شکل زیر می‌باشد:

$$Pa_{x_i} = \{C, x_{\delta(i)}\}$$

که درخت در اینجا تابعی در بازه‌ی $\delta(i) > 0$ و Pa_{x_i} مجموعه والد‌ها برای هر x_i می‌باشد. یک متغیر دسته یا کلاس (C) که فاقد والد است به صورت $Pa_C = \emptyset$ تعریف می‌شود. بیزین ساده‌ی درخت افزودنی از لحاظ نظری و تجربی عملکرد بهتری نسبت به الگوریتم بیزین ساده داشته و با توجه به عدم نیاز به فرایند جستجو، محاسبات آن ساده‌تر است (فریدمن و همکاران ۱۹۹۷). در روال ساخت یک درخت افزودنی از مفهوم بیزین درختی چاو و لیو^۱ استفاده شده است. در اینجا با یک مسئله‌ی بهینه‌سازی روبرو هستیم که به دنبال یافتن یک درخت پوشای موزون ماکزیمم^۲ در یک گراف با هدف بیشینه کردن درست‌نمایی $\delta(i)$ می‌باشد. آنگاه گام‌های ساخت درخت افزودنی را به شکل زیر می‌توان شرح داد (فریدمن و همکاران ۱۹۹۷):



شکل ۵.۱۶ ساختار شبکه‌ی بیز ساده‌ی درخت افزودنی

گام ۱. تابع اطلاعات متقابل شرطی را برای هر زوج (i, j) به شکل زیر محاسبه کنید:

$$I_P(x_i: x_j | C) = \sum_{x_i, x_j, C} P(x_i, x_j, C) \log \frac{P(x_i, x_j | C)}{P(x_i | C)P(x_j | C)}, \quad i \neq j$$

این تابع نشان می‌دهد که زمانیکه متغیر دسته یا کلاس مشخص باشد، چه میزان اطلاعات ارائه می‌شود.

گام ۲. یک گراف غیرجهت‌دار کامل را ساخته و از یک تابع اطلاعات متقابل شرطی برای تخصیص وزن به هر یالی که x_i را به x_j متصل می‌کند استفاده کنید.

¹ Chow & Liu tree Bayesian

² Maximally weighted spanning tree

گام ۳. یک درخت پوشای موزون بیشینه (ماکزیمم) ایجاد کنید.

گام ۴. گراف غیرجهت دار را با انتخاب یک متغیر ریشه و تعیین جهت تمامی یال‌ها به سمت خارج از آن به یک گراف جهت دار تبدیل کنید.

گام ۵. با افزودن یک رأس به نام C و ترسیم یک کمان از C به هر x_i یک مدل بیزین ساده‌ی درخت افزودنی ایجاد کنید.

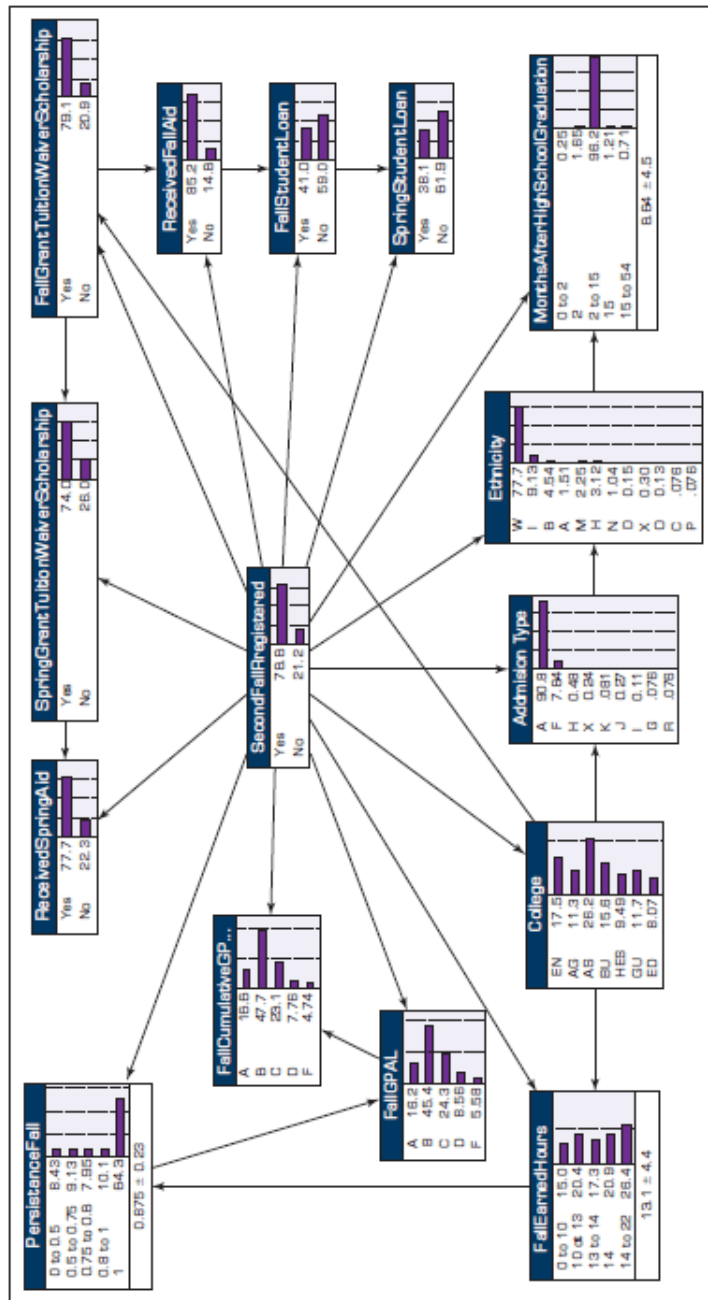
یکی از ویژگی‌های برتر شبکه‌های بیزی، سهولت وفق‌پذیری آنهاست. برای ساخت یک شبکه بیزی، می‌توانیم کار را با یک شبکه کوچک و دانش محدود شروع کرده و سپس با دستیابی به اطلاعات جدید، آن را گسترش دهیم. در چنین وضعیتی، نبود برخی مقادیر در مجموعه داده‌ها نمی‌تواند چندان مشکل ساز باشد زیرا می‌توان از بخشی از داده‌ها/مقادیر/دانش که موجود است برای ساخت احتمالات استفاده کرد. شکل ۵.۱۷ یک مثال کامل از شبکه بیزی را برای پروژه‌ی حفظ و نگهداشت دانشجویان نشان می‌دهد.

از منظر کاربردپذیری، چنین شبکه‌ی بیزی کاملی می‌تواند برای متخصصان و کارشناسان (یعنی مدیران مؤسسات آموزشی) بسیار سودمند باشد زیرا دیدگاهی کل‌نگر از تمامی ارتباطات را ارائه کرده و روش‌هایی را برای جستجو در اطلاعات ریز و دقیق با استفاده از انواع آنالیزهای «شرطی» فراهم می‌کند. در واقع، با این مدل شبکه می‌توان با انتخاب و تغییر سیستماتیک مقدار متغیر پیشگو در دامنه‌ی مقادیر (یعنی ارزیابی میزان تغییر ریسک اخراج دانشجو با تغییر مقدار یک متغیر پیشگو مثل Fall-PGA) احتمال ریسک ترک یک دانشجوی خاص را که درست‌نمایی یا احتمال پسین دانشجویی است که اخراج می‌شود محاسبه کرد.

در تفسیر مدل شبکه بیزی شکل ۵.۱۷، باید کمان‌ها، جهت فلش‌های روی کمان‌ها، تعاملات مستقیم و ارتباطات غیرمستقیم را در نظر گرفت. به عنوان مثال، دسته‌ی `fall grant/tuition waiver/scholarship` (یعنی `FallGrantTuitionWaiverScholarship`) و تمامی گره‌های متصل به آن مربوط به ریزش یا ترک دانشجو هستند (یعنی `SecondFallRegistered`). علاوه بر این، اگرچه `FallGrantTuitionWaiverScholarship` با `College` و `spring grant/tuition waiver/scholarship`

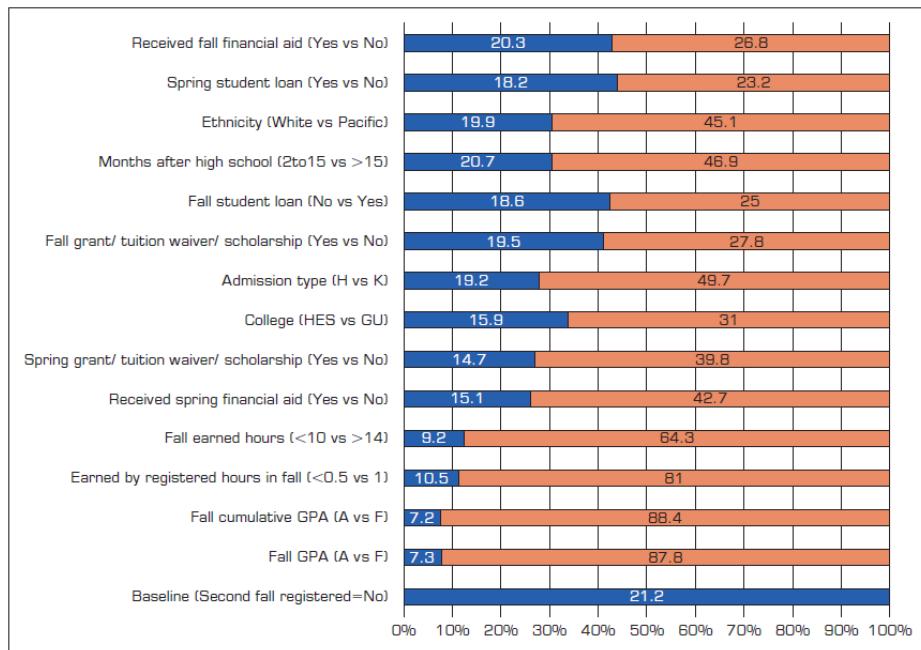
(`SpringGrantTuitionWaiverScholarship`) ارتباط مستقیم دارد، ولی از طریق `College` با نوع پذیرش (`AdmissionType`) نیز رابطه‌ی غیرمستقیم دارد. طبق مدل شبکه بیزی، یکی از مهم‌ترین متغیرهای پیشگوی تعاملی، ساعت اعتبار کسب شده‌ی دانشجو بر اساس نرخ ثبت‌نام است (`PersistenceFall`) که موجب `FallGPA` دانشجو و ترک وی می‌شود. در نتیجه، اگر `PersistenceFall` دانشجو کمتر از ۰.۸ باشد آنگاه نوع `College` بر ترک دانشجو تأثیر خواهد گذاشت. با وجود این، اگر این مقدار برابر ۱ باشد آنگاه نوع `College` تأثیر معناداری بر ریزش یا ترک دانشجو نخواهد داشت.

شکل ۵.۱۸ به عنوان یک دیدگاه جمعی بر سناریوهای شرطی، مثبت‌ترین و منفی‌ترین مقادیر هر متغیر پیشگو را به همراه احتمالات پسین آنها خلاصه کرده است. به عنوان مثال، دستیابی به A برای `Fall GPA` موجب کاهش احتمال پسین ترک دانشجو به ۷.۳ درصد شده یا برعکس، گرفتن F موجب افزایش احتمال ترک دانشجو به میزان ۸۷.۸٪ می‌شود در حالیکه خط‌مبنا معادل ۲۱.۲٪ است.



شکل ۵.۱۷ شبکه‌ی پیز برای پیش‌بینی ترک کالج توسط دانشجو

برخی افراد در مورد بکارگیری شبکه‌های بیزی دچار تردید هستند زیرا تصور می‌کنند که شبکه‌های بیزی در صورتیکه احتمالاتی که این مدل بر اساس ساخته می‌شود دقیق نباشد به خوبی نتیجه نخواهد داد. با وجود این، در اکثر موارد مشخص شده است که احتمالات تقریبی بدست آمده از داده‌ها و حتی احتمالات ذهنی که توسط کارشناسان خبره انجام شده، نتایج نسبتاً خوبی را ارائه داده است. شبکه‌های بیزی در برابر دانش و اطلاعات ناقص، نادقیق و معیوب، استوار است. در اغلب مواقع، ترکیب چندین مسیر دانشی ناقص به شبکه‌های بیزی اجازه می‌دهد که پیش‌بینی‌ها و نتیجه‌گیری‌های فوق‌العاده‌ای را به ما بدهند. مطالعات نشان داده است که افراد در برآورد احتمالات «رو به جلو» بهتر عمل می‌کنند. به عنوان مثال، مدیران در برآورد احتمال اینکه «اگر دانشجو در کالج اخراج شود، احتمال اینکه در کالج علوم و هنر بوده باشد چقدر است؟» بهتر عمل می‌کنند تا اینکه برعکس آن را بتوانند تخمین بزنند «اگر دانشجویی در کالج علوم و هنر مشغول به تحصیل باشد احتمال اینکه ترم بعدی را ثبت نام نکند چقدر است؟»



شکل ۵.۱۸ احتمال ترک دانشجو به تفکیک عوامل ریسک، عوامل تحلیل شرطی فردی

سوالاتی برای مرور بخش ۵.۸

۱. شبکه‌های بیزی چه هستند؟ ویژگی آنها چیست؟
۲. ارتباط بین شبکه‌های بیز ساده و شبکه‌های بایزین چیست؟
۳. فرایند توسعه‌ی یک مدل شبکه‌های بایزین چیست؟
۴. مزایا و معایب شبکه‌های بایزین در مقایسه با سایر روش‌های یادگیری ماشین چیست؟
۵. بایز ساده‌ی درخت افزودنی چیست و چه ارتباطی با شبکه‌های بایزین دارد؟

همادی‌ها (یا به بیان بهتر، مدل‌سازی همادی^۱) ترکیبی از نتایج بدست آمده از دو یا چند مدل تحلیلی است که در قالب یک خروجی مرکب ارائه می‌شوند. همادی‌ها عمدتاً برای مدل‌سازی پیش‌گویی در زمانیکه امتیازهای دو یا چند مدل برای دستیابی به پیش‌بینی بهتری با یکدیگر ترکیب می‌شوند بکار می‌روند. پیش‌بینی می‌تواند از نوع دسته‌بندی (پیش‌بینی یک برچسب دسته یا کلاس) یا رگرسیون/تخمین باشد (تخمین یک متغیر خروجی عددی). اگرچه در مدل‌سازی از نوع پیش‌بینی عمدتاً از همادی‌ها استفاده می‌شود ولی می‌توان از آن برای تحلیل‌هایی از قبیل خوشه‌بندی و کاهش قانون وابستگی نیز استفاده کرد. این بدان معناست که از مدل‌های همادی می‌توان برای وظایف یادگیری ماشین نظارت شده و نظارت نشده استفاده کرد. این رویه‌های یادگیری ماشین برای شناسایی و ساخت بهترین مدل ممکن (اغلب دقیق‌تری متغیر پیشگو از مجموعه داده‌ها) از بین تعداد زیادی از مدل‌های جایگزین استفاده می‌شده است. برای اینکار، تحلیلگرها و دانشمندان از یک فرایند تجربی مبسوط که عمدتاً مبتنی بر آزمون و خطا بوده برای بهبود عملکرد هر یک از مدل‌ها استفاده می‌کردند (که بر اساس سنجه‌های از پیش تعیین شده مثل دقت پیش‌بینی تعریف شده‌اند) تا بتوان از بهترین مدل‌ها برای انجام کارهای در دست استفاده کرد. رویکرد همادی، برعکس این تفکر است و به جای ساخت مدل‌ها و انتخاب یک مدل بهینه برای استفاده/استقرار، تعداد زیادی مدل را ایجاد کرده و از آنها برای وظایفی (مثلاً پیش‌بینی) که برایشان در نظر گرفته شده استفاده می‌کند.

پژوهشگران و متخصصان معمولاً به دو دلیل اصلی از همادی‌ها استفاده می‌کنند: برای دقت بهتر و نتایج پایدارتر/استوارتر/منسجم‌تر/معتبرتر. نشریات و مطالعات پژوهشی متعددی طی دو دهه‌ی گذشته انجام شده که نشان می‌دهند همادی‌ها تقریباً همیشه موجب افزایش دقت پیش‌بینی مسئله شده و به ندرت نتایج آنها ضعیف‌تر از مدل‌های انفرادی می‌باشد (ابوت ۲۰۱۴). همادی‌ها در دهه ۹۰ در ادبیات داده‌کاوی/تجزیه و تحلیل پدیدار شدند که موفقیت‌های محدودی که از کارهای ابتدایی انجام شده برای ترکیب پیش‌بینی‌ها بدست آمده بودند عامل اصلی آن بودند. در اوایل تا اواسط سال‌های ۲۰۰۰، همادی‌ها به عوامل اصلی برنده شدن در رقابت‌های مدل‌سازی پیشگویانه و داده‌کاوی تبدیل شدند. یکی از شناخته‌شده‌ترین جوایز رقابت‌های همادی‌ها، جایزه‌ی نتفلیکس است که یک رقابت آزاد برای ترغیب پژوهشگران و متخصصان برای پیش‌بینی رتبه‌بندی فیلم‌ها توسط کاربران و بر اساس رتبه‌بندی‌های قبلی بوده است. جایزه‌ی تیمی معادل ۱ میلیون دلار بود که می‌توانست RMSE الگوریتم‌های پیش‌بینی داخلی نتفلیکس را تا بالاترین میزان حاشیه سود ولی تا حداکثر ۱۰ امتیاز کاهش دهد. برنده این جایزه، نفر دوم و تمامی تیم‌هایی که در صدر جدول قرار داشتند از مدل‌های همادی استفاده کرده بودند. در نتیجه؛ پروژه‌ی برنده، ماحصل یک همادی متشکل از صدها مدل پیشگویانه بود.

زمانیکه به موضوع توجیه کاربرد همادی‌ها می‌رسیم، ورهایز^۲ (۲۰۱۶) بهترین تعریف را ارائه می‌دهد – اگر بخواهید در یک رقابت تحلیل پیشگویانه برنده شوید یا حداقل به جایگاه قابل قبولی در جدول امتیازات دست پیدا کنید، باید هوشمندانه و با آغوش باز از مدل‌های همادی استفاده کنید. کاگل^۳ به پلتفرم اصلی دانشمندان علوم داده برای نمایش استعدادهایشان بدل شده است. طبق ورهایز، رقابت‌های کاگل همانند مسابقات فرمول یک در حوزه علم داده‌ها است. برندگان در این رقابت‌ها با چهار رقم اعشار

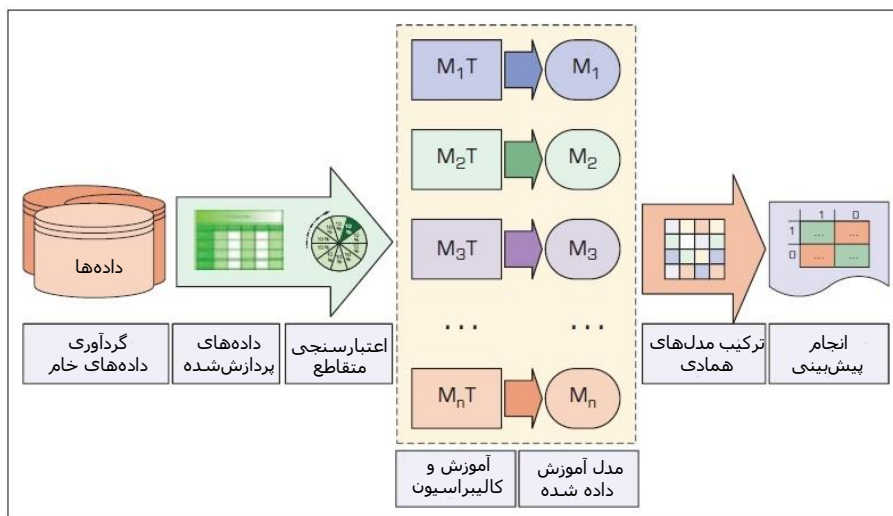
¹ Model ensembles, ensemble modeling

² Vorhies

³ Kaggle

اختلاف مشخص می‌شوند و همانند خودروهای مسابقات فرمول یک ما آنها را با رانندگانی که در خیابان‌ها می‌بینیم اشتباه نمی‌گیریم. میزان زمان اختصاص داده شده و تکنیک‌های بکار رفته همیشه برای یک پروژه‌ی معمول تولید علم داده‌ها مناسب نیستند ولی همانند پدل شیفترها و سیستم‌های تعلیق اگزاتیک، بسیاری از این بهبود به زندگی روزمره‌ی و روش کاری متخصصان تجزیه و تحلیل راه پیدا کرده‌اند. علاوه بر رقابت‌های کاگل، سازمان‌های معتبر از قبیل گروه موضوعی ویژه‌ی انجمن ماشین‌های حسابگر (ای‌سی‌ام)^۱ در خصوص داده‌کاوی و کشف دانش^۲ و کنفرانس آسیا-اقیانوسیه در خصوص کشف دانش و داده‌ها^۳ به طور منظم به سازماندهی رقابت‌ها و برخی اوقات جوایز مالی برای جامعه‌ی دانشمندان علوم داده می‌پردازند تا بتوانند شایستگی‌های خود را به نمایش بگذارند. برخی شرکت‌های معروف در حوزه تجزیه و تحلیل مثل شرکت اس‌ای‌اس و ترادیتا نیز رقابت‌های مشابهی را هم برای دانشجویان و هم فارغ التحصیلان دانشگاه‌ها در سرتاسر جهان و معمولاً همراستا با کنفرانس‌های منظم خودشان سازماندهی می‌کنند.

علت محبوبیت و ضرورت مدل‌های همادی تنها دقت آنها نیست. بارها و بارها مشاهده شده است که همادی‌ها می‌توانند دقت مدل را افزایش دهند ولی همچنین می‌تواند استواری، پایداری و قابلیت اطمینان مدل‌ها را نیز بهبود بخشند. در موقعیت‌هایی که قابلیت اطمینان پیش‌بینی از اهمیت اساسی برخوردار باشد، استواری، پایداری و قابلیت اطمینان همادی‌ها به اندازه‌ی دقت آنها اهمیت می‌یابد. در مدل‌های همادی، با ترکیب و ادغام چندین مدل (یک نوع میانگین‌گیری) در قالب یک نتیجه‌ی پیش‌بینی، هیچ مدلی به تنهایی نمی‌تواند بیانگر مقدار پیش‌بینی شده‌ی نهایی مدل‌ها باشد که به نوبه‌ی خود، درست‌نمایی پیش‌بینی‌های غیرعادی و بی‌ارزش را کاهش می‌دهد. شکل ۵.۱۹ یک تصویر گرافیکی از مدل‌های همادی را برای مسائل پیش‌بینی از نوع دسته‌بندی نشان می‌دهد. اگرچه شکل‌های مختلفی از آنها وجود دارند ولی اکثر روش‌های مدل‌سازی همادی از این فرایند پیروی می‌کنند. از چپ به راست، شکل ۵.۱۹ وظایف عمومی ثبت و آماده‌سازی داده‌ها را پس از اعتبارسنجی متقاطع و ساخت و تست مدل و نهایتاً تجمیع/ترکیب مدل‌های انفرادی و ارزیابی پیش‌بینی‌های آنها به تصویر کشیده است.



شکل ۵.۱۹ نمایش گرافیکی مدل‌های همادی برای مدل‌سازی پیش‌بینی

¹ Association for Computing Machinery (ACM)

² Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD)

³ Pacific-Asia Conference in Knowledge Discovery and Data Mining (PAKDD)

یک راه دیگر برای بررسی همادی‌ها این است که از منظر «خرد جمعی» یا «جمع‌سپاری» به آن نگاه کنیم. در کتاب معروف «خرد جمع» (سورویکی ۲۰۰۵)، مؤلف پیشنهاد می‌کند که اگر به جای آنکه به نظر یک فرد خبره اتکا کنیم، نظرات تعداد زیادی از افراد خبره را بتوانیم (طی فرایندی به نام جمع‌سپاری) به شکل یک تصمیم جمعی کنیم که بر عقیده و نظر یک فرد خبره ارجحیت و برتری دارد آنگاه می‌توانیم به تصمیمات بهتری دست پیدا کنیم. وی چهار ویژگی ضروری را برای اینکه نظر گروهی به خوبی پاسخگو بوده و تبدیل به اثر مخالف تصمیمات ضعیف که ناشی از «جنون جمع» هستند نشوند نام برده است: تنوع عقاید، استقلال، تمرکززدایی و تجمع.

سه ویژگی اول مربوط به نحوه‌ی اتخاذ تصمیمات فردی هستند - آنها باید اطلاعاتی داشته باشند که متفاوت از اطلاعات سایر افراد گروه است و تحت تأثیر سایر افراد گروه قرار نمی‌گیرد. ویژگی آخر صرفاً بیانگر تصمیماتی است که باید ترکیب شوند. این چهار مشخصه یا ویژگی به نظر می‌رسد که سنگ زیربنای ساخت مدل‌های همادی بهتر نیز باشند. هر مدل پیشگویانه، در تصمیم‌نهایی حرفی برای گفتن دارد. تنوع عقاید را می‌توان از طریق همبستگی خود مقادیر پیشگویانه سنجید - اگر تمامی پیش‌بینی‌ها شدیداً با یکدیگر همبستگی داشته باشند یا به بیان دیگر اگر مدل‌ها همگی تقریباً با یکدیگر همخوانی داشته باشند، آنگاه تجمع آنها هیچ مزیت قابل توجهی نخواهد داشت. برای ویژگی تمرکززدایی می‌توان داده‌ها را مجدداً نمونه‌برداری کرده یا به موردها وزن داد؛ هر مدل یا از رکوردهای متفاوتی از یک مجموعه داده مشترک استفاده کرده یا حداقل از رکوردهایی استفاده می‌کند که اوزان آنها با سایر مدل‌ها متفاوت است (ابوت ۲۰۱۴).

یکی از شایع‌ترین مفاهیم در آمار و مدل‌سازی پیشگویانه که بسیار با مدل‌های همادی نیز مرتبط است، معامله بایاس-واریانس است. از این رو، پیش از آنکه وارد بررسی انواع مختلف مدل‌های همادی شویم لازم است که اصل معامله بایاس-واریانس را مرور و درک کنیم (چون در رشته‌ی آمار و یادگیری ماشین کاربرد دارد). در تحلیل پیشگویانه، بایاس به خطا و واریانس به ثبات (یا فقدان آن) در دقت پیشگویی مدل‌های بکار رفته بر روی سایر مجموعه داده‌ها اشاره دارد. بهترین مدل‌ها آنهایی هستند که بایاس پایین (خطای پایین، دقت بالا) و واریانس پایین (ثبات دقت داده‌ها از یک مجموعه داده به مجموعه داده دیگر) داشته باشند. متأسفانه، همواره یک معامله بین این دو سنجه در ساخت مدل‌های پیشگویانه وجود دارد که بهبود یکی منجر به بدتر شدن دیگری می‌شود. شما می‌توانید به بایاس پایینی در خصوص داده‌های آموزش دست پیدا کنید ولی مدل در خصوص داده‌های اعتبارسنجی ممکن است واریانس بالایی پیدا کند زیرا مدل‌ها ممکن است بیش‌آموزش شده باشند. به عنوان مثال، الگوریتم کی میانگین همسایه با $k = 1$ مثالی از یک مدل با بایاس پایین است (بهترین حالت برای مجموعه داده‌های آموزش) ولی بر روی مجموعه داده‌های تست/اعتبارسنجی دارای واریانس بالایی می‌تواند باشد. استفاده از اعتبارسنجی متقاطع به همراه مدل‌های همادی مناسب بهترین روش جاری برای حل مشکل معامله بایاس-واریانس در مدل‌سازی پیشگویانه است.

همادی‌ها یا گروهی از مدل‌های پیشگویانه که با یکدیگر کار می‌کنند، راهبرد اصولی برای توسعه‌ی مدل‌های تحلیلی دقیق و استوار بوده‌اند. اگرچه همادی‌ها مدت زمان طولانی است که وجود دارند ولی محبوبیت و اثربخشی آنها تنها در چند دهه‌ی گذشته و به توجه به بهبود مستمر آنها به موازات رشد سریع قابلیت‌های نرم‌افزاری و سخت‌افزاری مشخص شده است. زمانیکه ما به مدل‌های همادی رجوع می‌کنیم، بسیاری از ما بلافاصله به همادی‌های درخت تصمیم مثل جنگل تصادفی و درخت‌های تقویتی فکر می‌کنیم در حالیکه به بیان کلی، مدل‌های همادی را می‌توان به چهار گروه در دو بُعد دسته‌بندی کرد که در شکل ۵.۲۰ نشان داده شده است. بُعد اول، نوع روش (محور افقی x در شکل ۵.۲۰) می‌باشد که در آن همادی‌ها را می‌توان در انواع تقویت

یا دسته‌بندی دسته‌بندی کرد. بعد دوم، نوع مدل است (محور عمودی لا در شکل ۵.۲۰) که در آن همادی‌ها در قالب انواع همگون و ناهمگون دسته‌بندی می‌شوند (ابوت ۲۰۱۴).

نوع مدل	ناهمگون	وزن‌دهی ساده/پیچیده مدل	[حوزه پژوهشی فعال-نادر]
		<ul style="list-style-type: none"> ✓ پشته‌سازی (فرا مدل‌سازی) ✓ تلفیق اطلاعات 	<ul style="list-style-type: none"> وزن‌دهی سیستماتیک نمونه داده‌ها برای مدل‌سازی بهتر
نوع روش	همگون	<ul style="list-style-type: none"> ✓ درخت‌های ترکیبی ✓ جنگل تصادفی ✓ سایر انواع بگینگ تک مدلی (مثلاً شبکه عصبی مصنوعی) 	<ul style="list-style-type: none"> ✓ AdaBoost ✓ XGBoost سایر انواع بوستینگ تک مدلی
		دسته‌بندی	تقویت

شکل ۵.۲۰ دسته‌بندی ساده‌ی مدل‌های همادی

همانطور که از نامشان پیداست، همادی‌های همگون^۱، نتایج دو یا چند مدل از یک نوع را ترکیب می‌کنند مثل درخت تصمیم. در واقع، اکثریت قریب به اتفاق مدل‌های همادی همگون با استفاده از ترکیبی از ساختارهای درخت‌های تصمیم توسعه می‌یابند. دو گروه متداول از همادی‌های همگون که از درخت‌های تصمیم استفاده می‌کنند عبارتند از دسته‌بندی و بوستینگ. مدل‌های همادی ناهمگون^۲، نتایج حاصل از دو یا چند نوع مدل متفاوت مثل درخت تصمیم، شبکه‌های عصبی مصنوعی، رگرسیون لجستیک، ماشین‌های بردار پشتیبان و غیره را ترکیب می‌کنند. همانطور که در بحث «خرد جمعی» ذکر شد، یکی از عوامل کلیدی موفقیت در مدل‌سازی همادی، استفاده از مدل‌هایی کاملاً متفاوت با یکدیگر است که از منظرهای متفاوتی به داده‌ها نگاه می‌کنند. همادی‌های ناهمگون به خاطر روش ترکیب و ادغام مدل‌های مختلف، مدل‌های تلفیق اطلاعات^۳ (دیلین و شاردو ۲۰۱۰) یا پشته‌سازی^۴ اطلاعات نیز نامیده می‌شوند.

¹ Homogenous-type ensemble

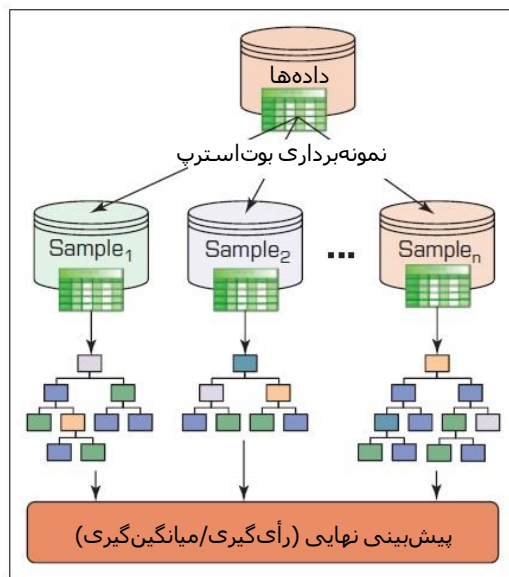
² Heterogenous-type ensemble

³ Information fusion models

⁴ Stacking

دسته‌بندی^۱ ساده‌ترین و متداول‌ترین روش مدل‌سازی همادی است. لئو برایمن^۲ یکی از صاحب‌نظران بسیار مورد احترام در دنیای آمار و تحلیل اولین کسی است که در سال ۱۹۹۶ به توضیح الگوریتم دسته‌بندی (یعنی تجمیع بوت‌استرپ^۳) در دانشگاه برکلی کالیفرنیا پرداخت. ایده‌ی پشت دسته‌بندی بسیار ساده ولی قوی است: ساخت چندین درخت تصمیم از روی داده‌هایی که مجدداً نمونه‌برداری شده‌اند و ترکیب مقادیر پیش‌بینی شده از طریق میانگین‌گیری یا رأی‌گیری. روش میانگین‌گیری که برای بکار برد، نمونه‌برداری بوت‌استرپ (نمونه‌برداری با جایگزینی) بوده است که مقادیر تکراری را برای برخی از رکوردهای موجود در داده‌های آموزش تولید می‌کند. با این روش انتخاب، به طور میانگین حدود ۳۷ درصد از رکوردها به هیچ وجه در مجموعه داده‌های آموزش در نظر گرفته نخواهند شد (ابوت ۲۰۱۴).

اگرچه دسته‌بندی اولین بار برای درخت تصمیم توسعه یافت ولی این ایده را می‌توان برای هر الگوریتم مدل‌سازی پیشگویانه که نتایجی تولید می‌کنند که مقادیر پیش‌بینی شده آنها دارای نوسان کافی است بکار برد. سایر الگوریتم‌های مدل‌سازی پیشگویانه که نامزدهای بالقوه‌ای برای مدل‌های همادی از نوع دسته‌بندی هستند شامل شبکه‌های عصبی، بیز ساده، کی میانگین همسایه (برای k با مقادیر پایین) و به میزان کمتر، حتی رگرسیون لجستیک می‌باشند. در صورتیکه مقدار k بزرگ باشد، کی نزدیک‌ترین همسایه، گزینه‌ی خوبی برای دسته‌بندی نیست. در این روش پیش‌بینی بر اساس رأی‌گیری یا میانگین‌گیری انجام می‌شود و به همین دلیل پیش‌بینی‌ها بسیار پایدار و با واریانس پایین هستند.



شکل ۵.۲۱ مدل‌های همادی درخت تصمیم از نوع دسته‌بندی

از دسته‌بندی می‌توان هم برای مسائل دسته‌بندی و هم رگرسیون/تخمین استفاده کرد. در مسائل پیش‌بینی از نوع دسته‌بندی، تمامی نتایج مدل‌های شرکت کننده با استفاده از یک مکانیزم رأی‌گیری اکثریت ساده یا پیچیده/موزون با یکدیگر ترکیب می‌شوند.

¹ Bagging

² Leo Breiman

³ Bootstrap Aggregation

برچسب دسته یا کلاسی که بالاترین/بیشترین رأی را بگیرد به عنوان پیش‌بینی تجمیعی/همادی برای آن نمونه یا رکورد در نظر گرفته می‌شود. در مسائل پیش‌بینی از نوع رگرسیون/تخمین، زمانیکه متغیر خروجی/هدف یک عدد باشد آنگاه تمامی نتایج مدل‌های شرکت‌کننده (تخمین‌های عددی) یا با استفاده از یک مکانیزم میانگین‌گیری موزون ساده یا پیچیده ترکیب می‌شوند. شکل ۵.۲۱ نمایش گرافیکی یک الگوریتم دسته‌بندی درخت تصمیم را به تصویر می‌کشد.

یکی از سوالات اساسی در بحث دسته‌بندی این است که «چه تعداد نمونه‌های بوت‌استرپ یا نسخه‌ی کیبی^۱ باید ایجاد شود؟» برایمان می‌گوید، «حس من این است که وقتیکه γ [متغیر وابسته] از نوع عددی باشد به تعداد کمی از این نمونه‌ها نیاز است و هرچه تعداد دسته‌ها افزایش یابد [برای مسائل پیش‌بینی از نوع دسته‌بندی] به تعداد بیشتری نمونه نیاز خواهیم داشت.» وی معمولاً از ۲۵-۱۰ نمونه‌ی بوت‌استرپ با بهبودهای چشمگیر که با حداکثر ۱۰ نمونه کیبی رخ می‌داد استفاده می‌کرد. بیش‌برازش مدل‌ها یک الزام اساسی برای ایجاد همادی‌های دسته‌بندی شده‌ی خوب است. با بیش‌برازش هر مدل، میزان بایاس کاهش می‌یابد ولی درخت‌های تصمیم معمولاً دقت ضعیف‌تری بر روی داده‌های آزمایش دارند. اما دسته‌بندی یک تکنیک کاهش واریانس است و میانگین‌گیری پیش‌بینی‌ها موجب پایداری بیشتر پیش‌بینی‌های انجام شده بر روی داده‌های جدید می‌شود.

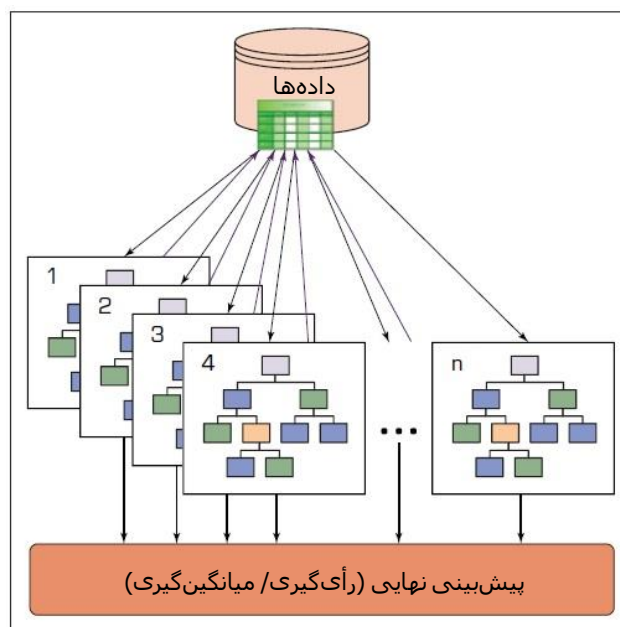
همانطور که قبلاً گفته شد، تنوع پیش‌بینی‌ها یک عامل کلیدی برای ساخت همادی‌های مؤثر است. یک روش برای سنجش تنوع پیش‌بینی‌ها این است که همبستگی بین مقادیر پیش‌بینی شده را بررسی کنیم. اگر همبستگی بین پیش‌بینی‌ها همواره بالا باشد (بیش از ۰.۹۵) آنگاه هر مدل می‌تواند اطلاعات پیشگویانه تکمیلی جزئی را وارد همادی‌ها کرده و بدین ترتیب دقت پیش‌بینی نیز به میزان کمی بهبود می‌یابد. به طور کلی، بهترین حالت این است که همبستگی بین پیش‌بینی‌ها کمتر از ۰.۹۰ باشد. همبستگی‌ها را باید به جای ۰.۱۶، ۰.۵۰ مقدار دسته‌بندی از روی احتمالات پیش‌بینی شده یا گرایش‌ها مدل محاسبه کنیم. نمونه‌برداری بوت‌استرپ در تکنیک دسته‌بندی، کلید اصلی دستیابی به تنوع مدل‌ها می‌باشد. متدولوژی نمونه‌برداری بوت‌استرپ را می‌توانیم به صورت وزن‌دهی برای هر رکورد در نظر بگیریم که برخی از رکوردها چندین مرتبه در داده‌های آموزش تکرار شده‌اند (وزن آنها ۱، ۲، ۳ یا بیشتر است) و برخی دیگر از رکوردها در مجموعه داده‌های آموزش لحاظ نشده‌اند (معادل ۰ می‌باشند). (بوت ۲۰۱۴).

تقویت^۲ را شاید بتوان دومین روش متداول همادی پس از دسته‌بندی دانست. یوآو فرویند و رابرت ایبی شاپیر برای اولین بار الگوریتم تقویت را در مقالاتی جداگانه در اوایل دهه ۹۰ و سپس در مقاله‌ای مشترک در سال ۱۹۹۶ معرفی کردند (فرویند و شاپیر ۱۹۹۶). آنها الگوریتم معروف تقویت را به نام AdaBoost معرفی کردند. ایده‌ی پشت مفهوم تقویت، همانند دسته‌بندی بسیار ساده است. ابتدا، یک مدل طبقه‌بندی نسبتاً ساده‌ای را ایجاد کنید؛ فقط کافی است که این مدل کمی بهتر از یک مدل احتمال تصادفی باشد. بنابراین، برای مسئله‌ی طبقه‌بندی دودویی فقط کافی است که طبقه‌بندی، کمی بهتر از ۰.۵۰ باشد. در این گام اول، همانند کاری که در ساخت یک مدل پیشگویانه انجام می‌دهیم از هر رکورد با وزن‌های برابر در الگوریتم استفاده می‌شود. خطای مقادیر پیش‌بینی شده برای هر مورد یادداشت می‌شود. وزن رکوردها/موارد/نمونه‌هایی که به درستی طبقه‌بندی شده‌اند را ثابت نگهداشته یا حتی ممکن است کاهش دهیم و وزن رکوردهایی که به اشتباه طبقه‌بندی شده‌اند را افزایش می‌دهیم و سپس بر اساس این اوزان جدید (یعنی مجموعه داده‌های آموزش تبدیل شده/موزون) یک مدل نمونه ثانویه می‌سازیم. به عبارت دیگر، برای مدل ثانویه، رکوردهایی که به اشتباه طبقه‌بندی شده‌اند از طریق اوزان "تقویت می‌شوند" تا تأثیر قوی‌تر و جدی‌تری

¹ Replicates

² Boosting

در ساخت مدل پیش‌بینی جدید داشته باشند. در هر بار تکرار، همچنان به افزایش وزن رکوردهایی که به اشتباه پیش‌بینی شده‌اند (آنهایی که طبقه‌بندی آنها کار دشواری است) می‌پردازیم تا جاییکه به درستی طبقه‌بندی شوند. این فرایند اغلب ده‌ها یا حتی صدها مرتبه تکرار می‌شود. پس از ده‌ها یا صدها تکرار، پیش‌بینی‌های نهایی بر اساس میانگین موزون پیش‌بینی‌های بدست‌آمده از تمامی مدل‌ها محاسبه می‌شود. شکل ۵.۲۲ فرایند ساده‌ی تقویت را در ساخت مدل‌های همادی درخت تصمیم به تصویر می‌کشد. همانطور که نشان داده شده است، هر درخت، جدیدترین مجموعه داده را (با اندازه‌های یکسان ولی جدیدترین وزن‌هایی که تقویت شده‌اند) برای ساخت درخت دیگری در نظر می‌گیرد. از بازخورد مواردی که به غلط پیش‌بینی شده‌اند به عنوان یک شاخص برای تعیین اینکه کدام موارد و تا چه اندازه‌ای برای نمونه‌ها/موارد آموزش تقویت داده شوند استفاده می‌شود.



شکل ۵.۲۲ مدل‌های همادی تقویتی برای درخت تصمیم

اگرچه ساختار و کاربرد تکنیک‌های دسته‌بندی و تقویت مشابه است ولی از استراتژی‌های متفاوتی برای بهره‌برداری از مجموعه داده‌های آموزش و ساخت بهترین مدل‌های همادی پیش‌بینی ممکن استفاده می‌کنند. دو تفاوت اساسی بین دسته‌بندی و تقویت به شرح زیر است: دسته‌بندی از یک نمونه بوت‌استرپ برای ایجاد درخت‌های تصمیم استفاده می‌کند در حالیکه تقویت از یک مجموعه داده آموزش کامل استفاده می‌کند. اگرچه روش دسته‌بندی، درخت‌های ساده و مستقلی را ایجاد می‌کند و تکنیک تقویت، درخت‌های وابسته‌ای را می‌سازد که جمعاً به یک مدل همادی نهایی دست پیدا می‌کنند. روش‌های تقویت به گونه‌ای طراحی شده‌اند تا با یادگیرنده‌های ضعیف یعنی مدل‌های ساده کار کنند؛ مدل‌های مؤلفه‌ای در یک همادی تقویت شده، مدل‌های ساده‌ای هستند که علیرغم واریانس پایین، بایاس بالایی دارند. زمانیکه از الگوریتم‌هایی که پیشگوهای ناپایداری هستند استفاده شود، در روش تقویت بهتر از روش دسته‌بندی می‌توان به بهبود دست یافت. پرکاربردترین تکنیک‌ها در مدل‌های تقویت شده، درخت‌های تصمیم هستند. بیز ساده نیز اگرچه بکار می‌رود ولی برای یک مدل ساده بهبودهای

چندانی را نتیجه نمی‌دهد. به لحاظ تجربی، تقویت معمولاً در مقایسه با درخت‌های تصمیم یا حتی همادی‌های از نوع دسته‌بندی دقت بالاتری دارد.

دسته‌بندی و تقویت جزو اولین روش‌های همادی بودند که در نرم‌افزارهای تحلیل پیشگویانه که عمدتاً از الگوریتم‌های درخت تصمیم استفاده می‌کردند بکار رفتند. از زمان ظهورشان، روش‌های بسیار زیاد دیگری برای ایجاد همادی‌ها به ویژه در نرم‌افزارهای منبع باز (هر دو آنها بخشی از پلتفرم‌های تحلیلی باز مثل کی‌نایم یا آر‌نچ و مثل کتابخانه‌های کلاس در نرم‌افزارهای آر و پایتون هستند) توسعه یافته و در دسترس است. محبوب‌ترین و موفق‌ترین واریانت‌های دسته‌بندی و تقویت عبارتند از جنگل تصادفی و تقویت گرادیان تصادفی.

جنگل تصادفی مدل جنگل تصادفی (آر‌اف)^۱ اولین بار توسط برایمن (۲۰۰۱) به عنوان یک شکل اصلاح‌یافته از الگوریتم دسته‌بندی ساده معرفی شد. الگوریتم آر‌اف همانند دسته‌بندی با یک مجموعه داده نمونه‌برداری شده بوت‌استرپ شروع شده و برای هر نمونه بوت استرپ یک درخت تصمیم می‌سازد. در مقایسه با دسته‌بندی ساده ولی یک تفاوت مهم در الگوریتم آر‌اف وجود دارد: در هر تفکیک در درخت و با شروع از اولین تفکیک، به جای آنکه تمامی متغیرهای ورودی را به عنوان کاندید در نظر بگیریم فقط یک زیرمجموعه تصادفی از متغیرها در نظر گرفته می‌شود. از این رو در آر‌اف، از تکنیک نمونه‌برداری بوت‌استرپ هم برای انتخاب تصادفی موارد و هم انتخاب تصادفی ویژگی‌ها (یعنی متغیرهای ورودی) استفاده می‌شود.

تعداد موارد و تعداد متغیرهایی که باید به همراه تعداد درخت‌هایی که ایجاد شود همگی پارامترهایی هستند که برای تصمیم‌گیری در ساخت مدل‌های آر‌اف بکار می‌روند. روش معمول این است که تعداد پیش فرض متغیرهایی که به عنوان کاندید در هر نقطه تفکیک در نظر گرفته می‌شود، ریشه مربع کل تعداد ورودی‌های کاندید باشد. به عنوان مثال، اگر ۱۰۰ ورودی کاندید برای مدل داشته باشیم، ۱۰ ورودی تصادفی برای هر تفکیک در نظر گرفته می‌شود. این بدان معناست که احتمال اینکه ورودی‌های یکسانی برای تفکیک‌ها در هر گره‌ی والد و فرزند در یک درخت وجود داشته باشند بسیار ضعیف است و بدین ترتیب درخت را ملزم به یافتن روش‌های دیگری برای دستیابی به حداکثر دقت در تفکیک‌های بعدی می‌کند. از این رو، یک مکانیزم گوناگونی دوگانه در بطن فرایند ساخت درخت ایجاد می‌شود - انتخاب تصادفی موارد و متغیرها. مدل‌های آر‌اف پیش‌بینی‌هایی را ارائه می‌کنند که معمولاً دقیق‌تر از نتایج دسته‌بندی ساده و اغلب دقیق‌تر از نتایج تکنیک تقویت هستند (یعنی AdaBoost).

تقویت گرادیان تصادفی الگوریتم تقویت ساده AdaBosst فقط یکی از چندین الگوریتم تقویت است که در حال حاضر در ادبیات موضوع به صورت مستند به آن پرداخته شده است. در حوزه نرم‌افزارهای تجاری، AdaBoost همچنان پرکاربردترین تکنیک تقویت است ولی ده‌ها واریانت دیگر از این تکنیک را می‌توان در بسته‌های نرم‌افزاری منبع باز پیدا کرد. یک الگوریتم تقویت جالب که اخیراً به دلیل عملکرد بالای خود محبوبیت یافته است، الگوریتم تقویت گرادیان تصادفی (اس جی بی)^۲ است که توسط جری فریدمن^۳ در دانشگاه استانفورد ایجاد شد. سپس فریدمن، نسخه‌ی پیشرفته‌ای از این الگوریتم را به نام درخت

¹ Random Forest (RF)

² Stochastic Gradient Boosting (SGB)

³ Jerry Friedman

رگرسیون افزودنی چندگانه (مارت)^۱ توسعه داد که بعدها به نام تری‌نت^۲ در ابزار نرم‌افزاری سیستم‌های سالفورد نامیده شد. الگوریتم مارت نیز همانند سایر الگوریتم‌های تقویت، درخت‌های ساده و متوالی را ایجاد کرده و آنها را به صورت جمعی ادغام می‌کند. معمولاً، درخت‌های ساده چیزی بیش از گنده درخت بود و می‌توانند حداکثر تا ۶ گرهی انتهایی داشته باشند. طبق روال الگوریتم، پس از ساخت اولین درخت، خطاها (که باقی‌مانده نیز نامیده می‌شوند) محاسبه می‌شود. درخت دوم و تمامی درخت‌های بعدی از این باقی‌مانده‌ها به عنوان متغیر هدف استفاده می‌کنند. درخت‌های بعدی به شناسایی الگوریتم‌هایی می‌پردازند که ورودی‌ها را به خطاهای کوچک و بزرگ مرتبط می‌کنند. پیش‌بینی ضعیف خطاها منجر به خطاهای بزرگ در پیش‌بینی درخت بعدی شده و پیش‌بینی قوی خطاها منجر به خطاهای کوچک‌تر برای پیش‌بینی درخت بعدی می‌شود. معمولاً، صدها درخت ساخته شده و از طریق ادغام افزودنی پیش‌بینی‌ها، پیش‌بینی نهایی بدست می‌آید که مدل‌های ثابت قطعه به قطعه هستند زیرا هر درخت نیز به خودی خود یک مدل ثابت قطعه به قطعه است. با وجود این، به ندرت می‌توان این پیچیدگی‌های مربوط به درخت‌ها را متوجه شد زیرا تقریباً صدها درخت در مدل همادی وجود دارد (ابوت ۲۰۱۴). الگوریتم تری‌نت که یک نمونه از تقویت‌گرایان تصادفی است از زمان پیدایش خود در چندین مسابقه‌ی مدل‌سازی داده‌کاوی برنده شده و جایگاه خود را به عنوان یک پیشگوی دقیق که نیازی به پاکسازی داده‌ها پیش از مدل‌سازی ندارد اثبات کرده است.

پشته‌سازی^۳ (یا تممیم پشته‌ای یا آبرفراگیر)، بخشی از روش‌های همادی نامتجانس است. از نظر برخی متخصصان تجزیه و تحلیل، این روش بهترین تکنیک همادی می‌تواند باشد ولی به دلیل پیچیدگی‌هایی که دارد کمتر از سایر روش‌ها درک شده است. به دلیل روال آموزش دو مرحله‌ای این تکنیک، برخی آن را یک روش مدل‌سازی همادی بیش از اندازه پیچیده می‌دانند. به بیان ساده، در پشته‌سازی یک مدل همادی از گروهی متنوع از فراگیرهای قوی ساخته می‌شود. در این فرایند، یک گام فراداده به نام آبرفراگیر^۴ وجود دارد. این فرا طبقه‌بندهای واسطه میزان دقت طبقه‌بندهای اصلی را پیش‌بینی کرده و به عنوان مبنایی برای تعدیل و اصلاحات بکار می‌روند (ورهایز ۲۰۱۶). فرایند پشته‌سازی در شکل ۵.۲۳ به تصویر کشیده شده است. همانگونه که در شکل ۵.۲۳ نشان داده شده است، در ساخت یک مدل همادی از نوع پشته‌ای، ابتدا چندین طبقه‌بند قوی و گوناگون با استفاده از نمونه‌های بوت‌استرپ از داده‌های آموزش، آموزش داده شده و طبقه‌بندهای لایه ۱ را ایجاد می‌کنند (هر یک از آنها برای دستیابی به بهترین نتایج پیش‌بینی ممکن به حد اعلی بهینه‌سازی شده‌اند). سپس خروجی‌های طبقه‌بندهای لایه ۱ برای آموزش یک طبقه‌بند لایه ۲ بکار می‌رود (یعنی یک آبرطبقه‌بند) (وولپر ۱۹۹۲). ایده‌ی اصلی این کار این است که ببینیم آیا داده‌های آموزش به شکلی مناسب فرا گرفته شده‌اند یا خیر. به عنوان مثال، اگر یک طبقه‌بند خاص، منطقه‌ی خاصی از یک فضای ویژگی را به اشتباه فرا گرفته باشد و بر این اساس، نمونه‌هایی مربوط به آن منطقه را به اشتباه طبقه‌بندی نماید، آنگاه طبقه‌بند لایه‌ی ۲ قادر با یادگیری این رفتار و سایر رفتارهای طبقه‌بندهای دیگر بوده و می‌تواند چنین آموزش نادرستی را اصلاح نماید. برای آموزش طبقه‌بندهای لایه ۱ معمولاً از انتخاب اعتبارسنجی متقابل استفاده می‌شود که کل داده‌های آموزش به k زیرمجموعه‌ی ناسازگار متقابل تقسیم می‌شوند و هر طبقه‌بند لایه ۱ ابتدا بر روی $k-I$ زیرمجموعه (مجموعه متفاوتی از) داده‌های آموزش، آموزش داده می‌شود. سپس هر طبقه‌بند بر روی مجموعه‌ی k ام که در حین آموزش دیده نشده است ارزیابی

¹ Multiple additive regression trees (MART)

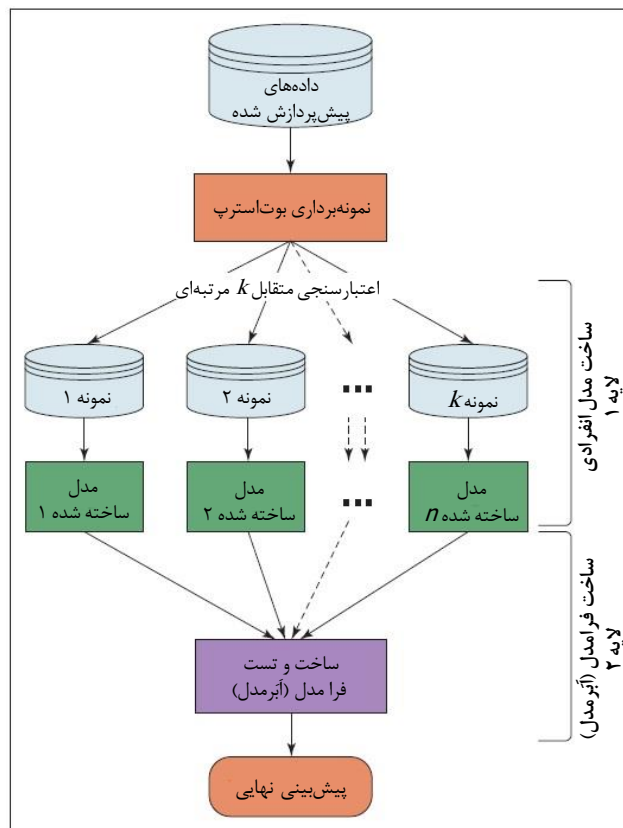
² TreeNet

³ Stacking

⁴ Meta learner, super learner

می‌شود. خروجی این طبقه‌بندها بر روی بلوک‌های شبه‌آموزشی آنها به همراه برچسب‌های صحیح برای این بلوک‌ها، مجموعه داده‌های آموزش را برای طبقه‌بند لایه ۲ فراهم می‌کنند.

تلفیق اطلاعات^۱ به عنوان بخشی از مدل‌های عمادی ناهمگون به ادغام (تلفیق) خروجی (یعنی پیش‌بینی) انواع مختلف داده‌ها از قبیل درخت‌های تصمیم، شبکه‌های عصبی مصنوعی، رگرسیون لجستیک، ماشین‌های بردار پشتیبان، بیز ساده و کی نزدیک‌ترین همسایه. واریانت‌های آنها می‌پردازد. تفاوت بین پشته‌سازی و تلفیق اطلاعات در این است که تلفیق اطلاعات فاقد هرگونه «فرا مدلسازی» یا «آبرفاگیر» می‌باشد و صرفاً نتایج حاصل از طبقه‌بندهای قوی و ناهمگون را با استفاده از رأی‌گیری ساده یا موزون (برای دسته‌بندی) یا میانگین‌گیری ساده یا موزون (برای رگرسیون) ادغام می‌کند. از این رو، نسبت به روش پشته‌سازی، ساده‌تر بوده و به کار محاسباتی کمتری نیاز دارد. در فرایند ادغام نتایج چندین مدل می‌توان از روش رأی‌گیری ساده (که هر مدل دارای یک رأی است) یا ترکیب موزونی از رأی‌گیری (که هر مدل بر اساس دقت پیش‌بینی خود سهم دارد یعنی مدل‌های دقیق‌تر وزن بیشتری دارند) استفاده کرد. صرفنظر از روش ترکیب بکار رفته، این نوع همادی‌های ناهمگون از ارزش بسیار بالایی در پروژه‌های داده‌کاوی و مدل‌سازی پیشگویانه برخوردارند. شکل ۵.۲۴ فرایند ساخت مدل‌های همادی از نوع تلفیق اطلاعات را به تصویر کشیده است.

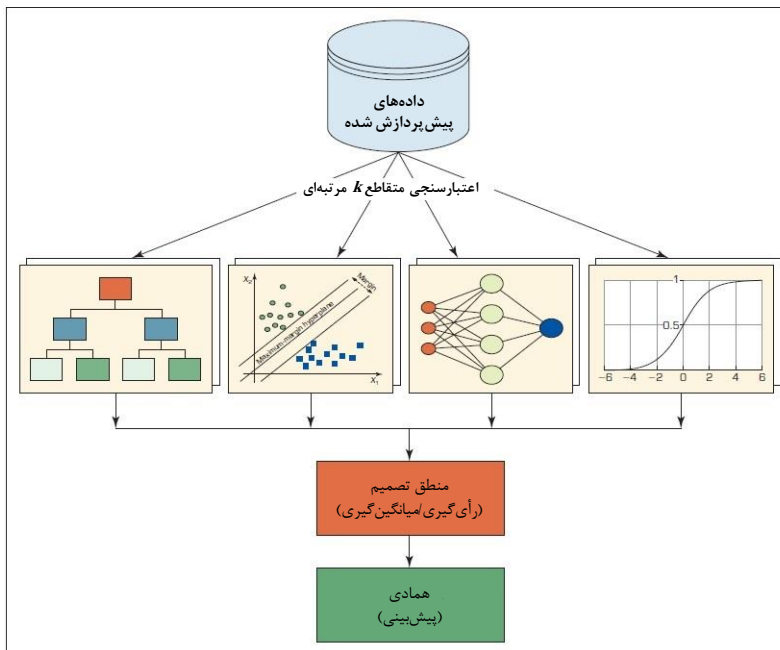


شکل ۵.۲۳ مدل‌های همادی از نوع پشته‌سازی

¹ Information fusion

به عنوان یک دانشمند علم داده‌ها، اگر از شما خواسته شود که یک مدل پیش‌بینی (یا هر نوع مدل تجزیه و تحلیل دیگری) را بسازید، از شما انتظار می‌رود تا برخی از مدل‌های همادی متداول را به همراه مدل‌های انفرادی استاندارد توسعه دهید. اگر این کار را به درستی انجام دهید متوجه خواهید شد که همادی‌ها اغلب دقت بیشتری داشته و تقریباً همیشه پایداری و قابلیت اطمینان بیشتری نسبت به مدل‌های انفرادی دارند. اگرچه ممکن است این مدل‌های همادی، شاه کلید تمامی مشکلات به نظر برسند ولی آنها نیز نواقصی دارند که در ادامه به دو مورد اصلی آنها می‌پردازیم.

پیچیدگی مدل‌های همادی پیچیده‌تر از مدل‌های انفرادی هستند. تیغ آکام، یک اصل بنیادی است که بسیاری از دانشمندان علم داده‌ها از آن استفاده می‌کنند؛ ایده این است که مدل‌های ساده‌تر، قدرت تعمیم‌دهی بهتری دارند و به همین خاطر بهتر است که پیچیدگی مدل‌ها را کاهش داد یا به عبارت دیگر، مدل‌ها را به شکلی ساده کرد که گنجاندن هر شرط، ضریب یا تفکیک در یک مدل به واسطه‌ی توان کاهش خطای مدل به میزان کافی توجیه‌پذیر باشد. یک روش برای کمی ساده ارتباط بین دقت و پیچیدگی از نظریه‌ی اطلاعات به شکل معیارهای نظری اطلاعات مثل معیار اطلاعات آکائیکه، معیار اطلاعات بیزی و حداقل طول توضیح گرفته شده است. آماردانان از قدیم و دانشمندان علم داده‌ها اخیراً از این معیارها برای گزینش متغیرها در مدل‌سازی پیشگویانه استفاده می‌کنند. بر اساس معیارهای نظری اطلاعات، خطای مدل را باید برای توجیه پیچیدگی‌های اضافی مدل باید کاهش داد. بنابراین سوال این است که «آیا مدل‌های همادی از اصل تیغ آکام تخطی می‌کنند؟» همادی‌ها روی هم رفته بسیار پیچیده‌تر از مدل‌های انفرادی هستند. طبق ابوت (۲۰۱۴)، اگر دقت مدل همادی بر روی داده‌های آموزش بهتر از مدل‌های انفرادی باشد تا زمانیکه پیچیدگی یک مدل را از منظرهای متفاوتی در نظر بگیریم (نه فقط پیچیدگی محاسباتی بلکه پیچیدگی رفتاری) آنگاه پاسخ ما «خیر» است. از این رو، نگران این نیستیم که افزایش پیچیدگی محاسباتی لزوماً موجب افزایش پیچیدگی مدل‌ها خواهد شد زیرا برخی اوقات، همادی‌ها پیچیدگی محاسباتی را به میزان زیادی کاهش می‌دهند.



شکل ۵.۲۴ فرایند ساخت مدل‌های همادی از نوع تلفیق اطلاعات

شفافیت تفسیر مدل‌های همادی می‌تواند بسیار دشوار باشد. اگر شما یک مدل همادی آر اف با ۲۰۰ درخت را بسازید چگونه می‌توانید توضیح دهید که چرا یک پیش‌بینی دارای مقدار مشخصی است؟ می‌توانید هر کدام از درخت‌ها را به صورت انفرادی بررسی کنید اگرچه چنین کاری مشخصاً ممکن نیست. برای این منظور، همادی اکثراً به صورت مدل‌های جعبه سیاه در نظر گرفته می‌شوند بدین معنی که نحوه کارکرد آنها برای مدل‌ساز یا فرد خبره در آن حوزه شفاف نیست. اگرچه برای قضاوت در مورد سطح مشارکت (یک شبه-متغیر برای سنجش میزان اهمیت) می‌توانید به آمارهای تفکیک نگاه کنید (کدام متغیرها اغلب برای تفکیک زود هنگام و اولیه در آن ۲۰۰ درخت انتخاب می‌شوند) ولی هر متغیر که در مدل‌های همادی آموزش دیده سهم دارد. تحقیق و بررسی ۲۰۰ درخت در مقایسه با یک درخت تصمیم می‌تواند بسیار دشوار بوده و روشی شهودی برای تفسیر نتایج و پیش‌بینی مدل نمی‌باشد. روش دیگر برای تعیین اینکه مهمترین ورودی‌های مدل کدامها هستند این است که تحلیل حساسیت انجام دهیم.

علاوه بر پیچیدگی و شفافیت، محاسبات مدل‌های همادی نیز به مراتب دشوارتر و پرهزینه‌تر بوده و ساخت و جاری‌سازی آنها نیز بسیار دشوار است. جدول ۵.۹ مزایا و معایب مدل‌های همادی را در مقایسه با مدل‌های انفرادی نمایش می‌دهد. به طور خلاصه، مدل‌های همادی برای مدل‌سازی که علاقمند به افزایش دقت مدل‌ها از طریق کاهش خطاهای مدل‌ها هستند خط مقدم جدیدتری محسوب می‌شود. شواهد این سلطه‌ی مدل‌های همادی را می‌توان در مسابقات داده‌کاوی و تجزیه و تحلیل مشاهده کرد: همادی‌ها همواره برنده هستند.

خبر خوب برای مدل‌سازان این است که بسیاری از تکنیک‌های ساخت مدل‌های همادی در قالب نرم‌افزارها ارائه شده‌اند. متداول‌ترین الگوریتم‌های همادی (دسته‌بندی، تقویت، پشته‌سازی و واریانت‌های آنها) تقریباً در هر نرم‌افزار تجاری یا منبع باز در دسترس هستند. در بسیاری از محصولات نرم‌افزاری از ساخت مدل‌های همادی سفارشی چه بر اساس یک الگوریتم انفرادی یا مجموعه‌ای ناهمگون از الگوریتم‌ها پشتیبانی می‌شود.

مدل‌های همادی برای هر راه‌حلی مناسب نیستند - کاربردپذیری آنها بر اساس اهداف مدل‌سازی که در طی تعریف و درک مسئله تعریف می‌شود تعیین می‌گردد - ولی آنها باید بخشی از جعبه‌ابزار مدل‌سازی هر دانشمند علم داده‌ها و مدل‌ساز باشند.

جدول ۵.۹ فهرست خلاصه‌ای از مزایا و معایب مدل‌های همادی در مقایسه با مدل‌های انفرادی

مزایا	شرح
دقت	مدل‌های همادی معمولاً در مقایسه با مدل‌های انفرادی مدل‌های دقیق‌تری را نتیجه می‌دهند
استواری	مدل‌های همادی در مقابل مقادیر پرت و نویز داده‌ها در مقایسه با مدل‌های انفرادی استواری بیشتری دارند.
پایداری (قابلیت اطمینان)	مدل‌های همادی به خاطر کاهش واریانس در مقایسه با مدل‌های انفرادی نتایج پایدارتر، مطمئن‌تر و قابل قبول‌تری را ارائه می‌دهند.
پوشش	مدل‌های همادی در مقایسه با مدل‌های انفرادی قدرت پوشش‌دهی بهتری از الگوهای پیچیده‌ی پنهان دارند.
معایب	شرح
پیچیدگی	مدل‌های همادی بسیار پیچیده‌تر از مدل‌های انفرادی هستند.
هزینه‌بر بودن محاسبات	در مقایسه با مدل‌های انفرادی، به زمان و قدرت بیشتری نیاز دارند.

مورد کاربردی ۵.۶

یک سیستم پشتیبان تصمیم تحلیلی پیشگویانه برای دادگاه‌های جرائم مواد مخدر: حکم زندانی شدن یا نشدن؟

مقدمه و انگیزش

بسیاری از کسب‌وکارها، سازمان و ارگان‌های دولتی برای یادگیری از تجارب گذشته و استفاده کارا تر و اثربخش‌تر از منابع محدودشان به منظور تحقق اهداف و مقاصد خود از تجزیه و تحلیل استفاده می‌کنند. علیرغم تمامی وعده‌هایی که تجزیه و تحلیل داده است ولی ماهیت چندبعدی و چندرشته‌ای آن برخی اوقات می‌تواند کاربرد مناسب آن را خدشه‌دار می‌کند. این موضوع به ویژه در خصوص استفاده از تحلیل‌های پیشگویانه در چندین رشته‌ی علوم اجتماعی صادق است زیرا این حوزه‌ها تحت سلطه‌ی تحلیل‌های توصیفی (مدل‌سازی آماری علی-توضیحی) بوده‌اند و ممکن است براحتی نتوان به مجموعه مهارت‌های مورد نیاز برای ساخت مدل‌های تحلیلی پیشگویانه دسترسی پیدا کرد. یک مرور بر ادبیات موضوع نشان می‌دهد که دادگاه‌های جرائم مواد مخدر یکی از این حوزه‌های کاربردی است. اگرچه پژوهشگران بسیاری به مطالعه‌ی این پدیده‌ی اجتماعی، مشخصه‌ها، الزامات و پیامدهای آن از دیدگاه تحلیلی توصیفی پرداخته‌اند اما در حال حاضر مدل‌های تحلیلی پیشگویانه که بتوانند به طور دقیق و مناسب پیش‌بینی کنند که چه کسی برنامه‌های درمانی اعتیاد را با موفقیت پشت سر خواهد گذاشت بسیار کمیاب هستند. برای پر کردن این شکاف و کمک به مقامات به منظور مدیریت بهتر منابع و بهبود نتایج، در این مطالعه به توسعه و مقایسه‌ی چندین مدل تحلیلی پیشگویانه (با استفاده از مدل‌های انفرادی و همادی) به منظور شناسایی افرادی که این برنامه‌های درمان را با موفقیت پشت سر خواهند گذاشت استفاده شده است.

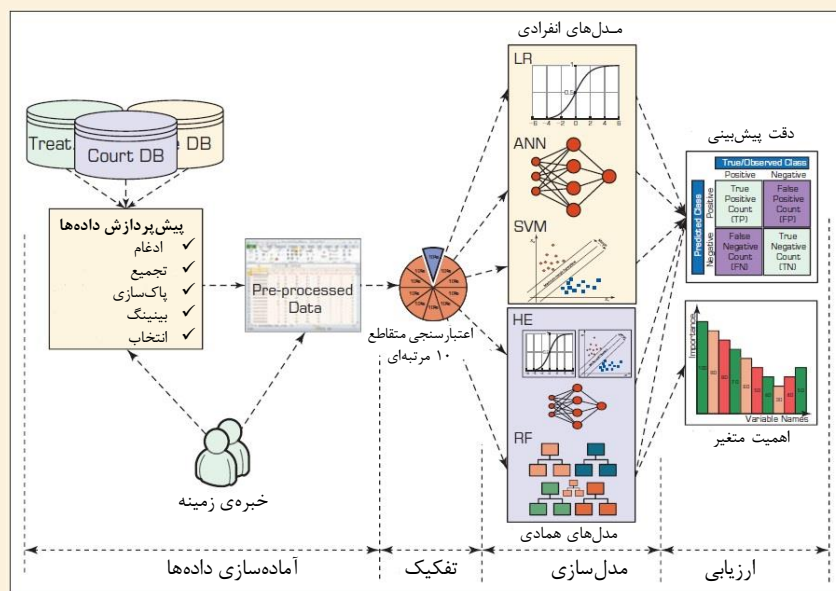
ده سال پس از آنکه رئیس‌جمهور ریچارد نیکسون برای اولین بار "جنگ علیه مواد مخدر" را اعلام کرد، رئیس‌جمهور رونالد ریگان یک دستور اجرایی که منجر به اجرای جدی‌تر جنگ علیه مواد مخدر می‌شد را امضا کرد و اعلام کرد «ما پرچم صلحی را که برای بسیاری از تلاش‌های مقابله با مواد مخدر به اهتزاز درآورده بودیم را پایین می‌آوریم. ما پرچم جنگ را بالا برده‌ایم.» شدت گرفتن جنگ علیه مواد مخدر منجر به افزایش بی‌سابقه ۱۰ برابری تعداد شهروندانی شد که به دلیل جرایم مواد مخدر طی دو دهه بعد زندانی شدند. افزایش بی‌سابقه‌ی تعداد پرونده‌های مواد مخدر در دادگاه‌ها موجب قفل شدن سیستم قضایی کیفری بیش و زندان‌ها شد. فراوانی پرونده‌های مربوط به مواد مخدر که منجر به افزایش زمان رسیدگی به پرونده‌ها نسبت به سایر جرائم شده بود هزینه‌های هنگفتی را به ادارات دادگستری ایالتی و فدرال تحمیل کرد. با توجه به افزایش تقاضا، سیستم‌های قضایی شروع به یافتن روش‌های نوآورانه برای تسریع رسیدگی به پرونده‌های مواد مخدر کردند. شاید سیستم‌های پشتیبان تصمیم مبتنی بر تحلیل، راه‌حل این مشکل باشند. برای پشتیبانی از این ادعا، هدف مطالعه‌ی جاری، ایجاد و مقایسه‌ی چندین مدل پیشگویانه بود که از نمونه بزرگی از داده‌های بدست آمده از دادگاه‌های مواد مخدر در مکان‌های مختلف برای پیش‌بینی افرادی که برنامه‌های درمانی را با موفقیت به پایان رسانده‌اند استفاده کرده است. پژوهشگران معتقدند که این تلاش می‌تواند منجر به کاهش هزینه‌های سیستم قضایی جنایی گردد.

متدولوژی

متدولوژی بکار رفته در این پژوهش شامل یک فرایند چندمرحله‌ای است که از روش‌های تحلیلی پیشگویانه در زمینه‌ی علوم اجتماعی بهره برده است. گام اول این فرایند که متمرکز بر درک دامنه‌ی مسئله و نیاز به انجام این مطالعه بوده است در بخش قبلی ارائه شد. در گام‌های فرایند، پژوهشگران از یک رویکرد ساختارمند و سیستماتیک برای توسعه و ارزیابی مجموعه مدل‌های پیشگویانه با استفاده از مجموعه بزرگ و غنی از داده‌های حقیقی استفاده کردند. گام‌ها شامل درک داده‌ها، پیش‌پردازش داده‌ها، ساخت مدل و ارزیابی مدل بوده است. در این بخش به مرور و بررسی این گام‌ها می‌پردازیم. این رویکرد همچنین شامل تکرارهای متعدد آزمایشات و اصلاحات متعدد به منظور بهبود وظایف و بهینه‌سازی پارامترهای مدل‌سازی برای دستیابی به بهترین نتیجه‌ی ممکن بوده است. یک نمایش تصویری از متدولوژی در شکل ۵.۲۵ نشان داده شده است.

نتایج

خلاصه‌ای از عملکرد مدل‌ها بر اساس دقت، حساسیت، ویژگی و AUC در جدول ۵.۱۰ ارائه شده است. همانطور که نتایج نشان می‌دهند، مدل آر اف دارای بهترین دقت طبقه‌بندی و بیشترین مقدار AUC بین مدل‌ها می‌باشد. مدل همادی ناهمگون به فاصله‌ی کمی بعد از مدل آر اف قرار دارد و ماشین‌های بردار پشتیبان، شبکه‌های عصبی مصنوعی و رگرسیون لجستیک از لحاظ عملکرد در رتبه سوم به بعد قرار دارند. آر اف همچنین بالاترین ویژگی و دومین حساسیت بالا را داشته است. حساسیت در این مطالعه، شاخصی از توانایی مدل برای پیش‌بینی درست نتیجه‌ی شرکت‌کنندگانی است که با موفقیت برنامه‌های درمان را به پایان برده‌اند. ویژگی از طرف دیگر تعیین می‌کند که یک مدل چگونه نتایج نهایی آنهایی که با موفقیت برنامه‌های درمان را طی نکرده‌اند را پیش‌بینی می‌کند. در نتیجه، می‌توان نتیجه گرفت که آر اف در خصوص مجموعه داده‌های دادگاه‌های مواد مخدر در این مطالعه عملکرد بهتری نسبت به سایر مدل‌ها دارد.



شکل ۵.۲۵ متدولوژی پژوهش به شکل جریان کار

جدول ۵.۱۰ عملکرد مدل‌های پیش‌بینی با استفاده از اعتبارسنجی متقابل ۱۰ مرتبه‌ای بر روی مجموعه داده‌های متوازن

نوع مدل	ماتریس درهم ریختگی		دقت	حساسیت	ویژگی	AUC
	T	G				
شبکه‌های عصبی	۱،۰۷۲	۶،۸۳۱	۸۵.۵۳	۸۵.۷۴	۸۵.۴۵	۰.۵۰۵
	۶،۸۶۱	۱،۰۴۲				
ماشین‌های برداری	۹۹۲	۶،۹۱۱	۸۸.۶۷	۸۹.۶۳	۸۷.۷۸	۰.۵۹۱۷
	۷،۱۰۴	۷۹۹				
...	۱،۵۸۲	۶،۳۲۱	۸۸.۸۳	۸۶.۸۶	۸۱.۸۸	۰.۵۸۸۹
	۷،۱۳۵	۷۶۸				
...	۹۰۵	۶،۹۹۸	۹۱.۸۶	۹۳.۴۴	۸۹.۸۲	۰.۵۹۲۷
	۷،۴۱۲	۴۹۱				
همادی‌های	۱،۰۱۸	۶،۸۸۵	۹۰.۶۱	۹۳.۶۶	۸۷.۹۶	۰.۵۹۱۶
	۷،۴۳۷	۴۶۶				

اگرچه مدل آر اف به طور کلی عملکرد بهتری در مقایسه با سایر مدل‌ها دارد، اما از لحاظ تعداد پیش‌بینی‌های منفی غلط پس از مدل همادی‌های ناهمگون قرار می‌گیرد. به همین منوال، مدل همادی‌های ناهمگون نیز عملکردی کمی بهتر در تعداد پیش‌بینی‌های منفی صحیح دارد. پیش‌بینی‌های مثبت غلط بیانگر شرکت‌کنندگانی هستند که موفق به تکمیل برنامه درمان نشده‌اند ولی مدل‌ها آنها را به اشتباه جزو کسانی طبقه‌بندی کرده‌اند که برنامه را با موفقیت به پایان برده‌اند. پیش‌بینی‌های منفی غلط بیانگر افرادی است که برنامه را به پایان برده‌اند ولی به اشتباه جزو افراد ناموفق طبقه‌بندی شده‌اند. پیش‌بینی‌های مثبت غلط مترادف با افزایش هزینه و زیان‌های فرصت هستند در حالیکه پیش‌بینی‌های منفی غلط دارای تبعات اجتماعی هستند. صرف منابع بر روی مجرمانی که در طی دوران درمان خود در برهه‌ای مجدداً مرتکب جرم شده و از ادامه‌ی برنامه باز مانده‌اند موجب شده است تا تعداد دیگری از مجرمان از مشارکت در برنامه درمانی باز بمانند. محروم کردن مجرمان از ادامه درمان برخلاف اهداف اولیه‌ی دادگاه‌های مواد مخدر مبنی بر بازگرداندن مجرمان غیرخوشونت‌گر به جامعه است.

به طور خلاصه، مدل‌سازی آماری علی-توضیحی رایج از استنتاج آماری و سطوح اهمیت برای تست و ارزیابی قدرت توضیحی مدل‌های فرضی زیربنایی یا تحقیق و بررسی در خصوص ارتباط بین متغیرها استفاده می‌کنند. اگرچه از یک رویکرد مشروع برای درک ارتباطات بین داده‌ها برای ساخت مدل استفاده شده است، تحلیل‌های توصیفی قادر به پیش‌بینی نتایج مشاهدات نمی‌باشند. به بیان دیگر، قدرت توضیحی نسبی (جزئی) بیانگر قدرت پیشگویانه مدل نیست و تحلیل‌های پیشگویانه یک الزام برای ساخت مدل‌های تجربی هستند که بتوانند به خوبی نتایج را پیش‌بینی کنند. از این رو، با اتکا بر یافته‌های این مطالعه می‌توانیم بگوییم که بکارگیری تحلیل‌های پیشگویانه (به جای استفاده‌ی صرف از تحلیل‌های توصیفی) برای پیش‌بینی نتایج دادگاه‌های مواد مخدر کاملاً قابل اثبات و معتبر می‌باشد.

سوالاتی برای مورد کاربردی ۵.۶

۱. دادگاه‌های مواد مخدر چه هستند و چه خدماتی برای جامعه انجام می‌دهند؟
۲. موارد تشابه و افتراق بین روش‌های پایه‌ی سنتی (نظری) و پیشرفته (یادگیری ماشین) در مطالعه‌ی دادگاه‌های مواد مخدر چه هستند؟
۳. آیا سیستم‌ها و موقعیت‌های اجتماعی دیگری برای بکارگیری تحلیل‌های پیشگویانه به ذهن شما می‌رسد؟

Source: Zolbanin, H., and Delen, D. (2018). To Imprison or Not to Imprison : An Analytics-Based Decision Support System for Drug Courts. *The Journal of Business Analytics* (forthcoming).

سوالاتی برای مرور بخش ۵.۹

۱. یک مدل همادی چیست و کجا می‌توان از آن برای کارهای تحلیلی استفاده کرد؟
۲. انواع مختلف مدل‌های همادی چه هستند؟
۳. چرا مدل‌های همادی نسبت به سایر روندهای یادگیری ماشین محبوبیت بیشتری پیدا می‌کنند؟
۴. تفاوت بین مدل‌های همادی از نوع دسته‌بندی و تقویت چیست؟
۵. مزایا و معایب مدل‌های همادی چه هستند؟

Neural computing involves a set of methods that emulates the way the human brain works.

The basic processing unit is a neuron. Multiple neurons are grouped into layers and linked together.

- There are differences between biological and artificial neural networks.
- In an artificial neural network, knowledge is stored in the weight associated with each connection between two neurons.
- Neural network applications abound in almost all business disciplines as well as in virtually all other functional areas.
- Business applications of neural networks include finance, bankruptcy prediction, time-series forecasting, and so on.
- There are various neural network architectures for different types of problems.
- Neural network architectures can be applied not only to prediction (classification or estimation)

-
-
-
-
-
-
- کاربردهای بسیار زیادی برای ماشین‌های بردار پشتیبان وجود دارد و کاربردهای جدیدی نیز در حال ظهور در حوزه‌های مختلف از جمله مراقبت سلامت، امور مالی، امنیت و انرژی هستند.
- الگوریتم نزدیک‌ترین همسایه یک نوع الگوریتم یادگیری مبتنی بر مثال است که در آن کلیه محاسبات تا زمان پیش‌بینی واقعی به تعویق انداخته می‌شوند.
- الگوریتم نزدیک‌ترین همسایه (یا کی نزدیک‌ترین همسایه) یک تکنیک ساده‌ی یادگیری ماشین است که هم برای مسائل پیش‌بینی از نوع طبقه‌بندی و هم رگرسیون/تخمین استفاده می‌شود.
- پارامتر k بیانگر تعداد همسایه‌هایی است که در یک مسئله پیش‌بینی از آنها استفاده می‌شود.
- تعیین مقدار « بهینه‌ی k » نیازمند یک آزمایش از نوع اعتبارسنجی متقاطع است.
- الگوریتم نزدیک‌ترین همسایه از یک سنج‌ی فاصله برای شناسایی همسایه‌های نزدیک/مناسب استفاده می‌کند.
- متغیرهای ورودی الگوریتم نزدیک‌ترین همسایه باید در قالب عددی باشند؛ تمامی متغیرهای غیر عددی/اسمی باید به متغیرهای عددی شبه-دودویی تبدیل شوند.
- طبقه‌بند‌های بیزین پایه و زیربنای نظریه‌ی بیز (یعنی احتمالات شرطی) هستند.
- بیز ساده یک روش طبقه‌بندی ساده‌ی مبتنی بر احتمالات است که برای مسائل پیش‌بینی از نوع طبقه‌بندی بکار گرفته می‌شود.
- در روش بیز ساده از متغیرهای ورودی و خروجی با مقادیر اسمی استفاده می‌شود و مقادیر عددی می‌بایست به صورت اسمی درآیند.
- عبارت کلیدی ساده به یک فرض غیرواقع‌بینانه ولی در عین حال کاربردی یعنی وابستگی متغیرهای پیشگو و ورودی اشاره دارد.
- شبکه‌ی بیزی یک تکنیک یادگیری ماشین نسبتاً جدید است که محبوبیت آن در بین دانشمندان علم داده‌ها، دانشگاهیان و نظریه‌پردازان در حال گسترش است.
- شبکه بیزی، ابزاری قدرتمند برای نمایش ساختار وابستگی به یک شکل گرافیکی و صریح است.
- از شبکه بیزی می‌توان برای پیش‌بینی و توضیح (ارتباطات بینابین متغیرها) استفاده کرد.
- شبکه‌های بیزی را می‌توان به صورت دستی (بر اساس دانش فرد خبره در آن حوزه) یا به صورت خودکار با استفاده از داده‌های تاریخی ایجاد کرد.
- در ساخت خودکار یک شبکه بیزی می‌توان از روش بیز ساده یا بیز ساده‌ی درخت افزوده استفاده کرد.

- شبکه‌های پیازی، یک مدل عالی برای انجام تحلیل‌های منطقی برای انواع مختلف سناریوهای فرضی ارائه می‌کنند.
- همادی‌ها (یا مدل‌های همادی) ترکیب‌هایی از نتایج هستند که از دو یا چند مدل تحلیلی بدست آمده و در قالب یک خروجی ترکیبی ارائه می‌شوند.
- اگرچه همادی‌ها عمدتاً برای مدل‌سازی پیش‌بینی در زمانیکه امتیازهای دو یا چند مدل برای دستیابی به یک پیش‌بینی بهتر با هم ترکیب می‌شوند بکار می‌روند ولی می‌توان از آنها برای خوشه‌بندی و وابستگی نیز استفاده کرد.
- همادی‌ها را می‌توان برای مسائل پیش‌بینی از نوع طبقه‌بندی (از طریق رأی‌گیری) و تخمین/رگرسیون (از طریق میانگین‌گیری) بکار برد.
- همادی‌ها عمدتاً به دو دلیل بکار می‌روند: دستیابی به دقت بهتر و دستیابی به نتایج پایدارتر/معتبرتر.
- تاریخچه‌ی اخیر در حوزه علم داده‌ها نشان داده است همادی‌ها، برنده‌ی اصلی رقابت‌ها هستند.
- همادی‌ها به دو نوع وجود دارند: همگون و ناهمگون. اگر مدل‌های ترکیب شده از یک نوع باشند (مثلاً درخت تصمیم) آنگاه آن را همادی همگون و در غیراینصورت ناهمگون می‌نامند.
- سه روش برای مدل‌سازی همادی‌ها وجود دارد: دسته‌بندی، تقویت و پشته‌سازی.
- جنگل تصادفی یک روش همادی مبتنی بر درخت تصمیم، همگن و از نوع دسته‌بندی است.
- تقویت گرادیان تصادفی یک روش همادی مبتنی بر درخت تصمیم همگن از نوع تقویت است.
- تلفیق اطلاعات و پشته‌سازی، مدل‌های همادی ناهمگون هستند که در آنها انواع مختلف مدل‌ها با یکدیگر ترکیب می‌شوند.
- معایب همادی‌ها شامل پیچیدگی و فقدان شفافیت است.

			AdaBoost
		k	

Training Sample						
Firm	WC/TA	RE/TA	EBIT/TA	MVE/TD	S/TA	BR/NB
1	0.1650	0.1192	0.2035	0.8130	1.6702	1
2	0.1415	0.3868	0.0681	0.5755	1.0579	1
3	0.5804	0.3331	0.0810	1.1964	1.3572	1
4	0.2304	0.2960	0.1225	0.4102	3.0809	1
5	0.3684	0.3913	0.0524	0.1658	1.1533	1
6	0.1527	0.3344	0.0783	0.7736	1.5046	1
7	0.1126	0.3071	0.0839	1.3429	1.5736	1
8	0.0141	0.2366	0.0905	0.5863	1.4651	1
9	0.2220	0.1797	0.1526	0.3459	1.7237	1
10	0.2776	0.2567	0.1642	0.2968	1.8904	1
11	0.2689	0.1729	0.0287	0.1224	0.9277	0
12	0.2039	-0.0476	0.1263	0.8965	1.0457	0
13	0.5056	-0.1951	0.2026	0.5380	1.9514	0
14	0.1759	0.1343	0.0946	0.1955	1.9218	0
15	0.3579	0.1515	0.0812	0.1991	1.4582	0
16	0.2845	0.2038	0.0171	0.3357	1.3258	0
17	0.1209	0.2823	-0.0113	0.3157	2.3219	0
18	0.1254	0.1956	0.0079	0.2073	1.4890	0
19	0.1777	0.0891	0.0695	0.1924	1.6871	0
20	0.2409	0.1660	0.0746	0.2516	1.8524	0

Test Data						
Firm	WC/TA	RE/TA	EBIT/TA	MVE/TD	S/TA	BR/NB
A	0.1759	0.1343	0.0946	0.1955	1.9218	?
B	0.3732	0.3483	-0.0013	0.3483	1.8223	?
C	0.1725	0.3238	0.1040	0.8847	0.5576	?
D	0.1630	0.3555	0.0110	0.3730	2.8307	?
E	0.1904	0.2011	0.1329	0.5580	1.6623	?
F	0.1123	0.2288	0.0100	0.1884	2.7186	?
G	0.0732	0.3526	0.0587	0.2349	1.7432	?
H	0.2653	0.2683	0.0235	0.5118	1.8350	?
I	0.1070	0.0787	0.0433	0.1083	1.2051	?
J	0.2921	0.2390	0.0673	0.3402	0.9277	?

Vertical_Distance_to_Hydrology	1
Vertical_Distance_to_Roadways	1
Hillshade_9am	1
Hillshade_Noon	1
Hillshade_9pm	1
Horizonatl_Distance_To_Fire_Poitns	1
Wilderness_Area	1
Soil_Type	1

$$Y = a + bX$$

$$R^2$$

$$Y = a + b_1X + b_2X^2$$

$$R^2$$

$$R^2$$

$$Y \quad X$$

$$R^2$$

neurodimension.com .Y

archive.ics.uci.edu/ml/index.php

GPA

SAT

... census.gov

...

... .8

...

...

...

... .9

...

Year	GMAT	GPA	GMAT
2007	11	1.1	11
2008	12	1.2	12
2009	11	1.1	11
2010	12	1.2	12
2011	13	1.3	13
2012	14	1.4	14
2013	15	1.5	15
2014	16	1.6	16
2015	17	1.7	17
2016	18	1.8	18
2017	19	1.9	19

...

California Scientific Software (calsci.com) .10

Neural Ware (neuralware.com) .11

Ward Systems Group (wardsystems.com) .12

StatSoft.com .13

Credit Screening Database, Housing Database .14

NeuroSolutions (neurosolutions.com) .15

California Scientific Software (calsci.com) .16

Neural Ware (neuralware.com) .17

Ward Systems Group (wardsystems.com) .18

StatSoft.com .19

Credit Screening Database, Housing Database .20

NeuroSolutions (neurosolutions.com) .21

California Scientific Software (calsci.com) .22

Neural Ware (neuralware.com) .23

Ward Systems Group (wardsystems.com) .24

¹ California Scientific Software

² Neural Ware

³ Ward Systems Group

	neoxi.com	.۱۸
Gee Whiz	neuroshell.com	.۱۹
	easynn.com	.۲۰
	Houseprices.tvq	
	tibco.com	.۲۱
	neuralware.com	.۲۲
	ibm.com	.۲۳
	sas.com	.۲۴
	Teradata.com	.۲۵

۶

یادگیری عمیق و رایانش شناختی

هوش مصنوعی مجدداً راه خود را این بار بسیار قویتر و نویدبخش تر از گذشته به دنیای رایانش و زندگی های ما پیدا کرده است. این ظهور مجدد و پیش‌بینی نشده و سطح انتظارات جدید را به میزان زیادی می‌توان مربوط به یادگیری عمیق و رایانش شناختی دانست. این دو واژه بسیار جدید بیانگر نوین‌ترین پیشرفت‌ها در هوش مصنوعی و یادگیری ماشین امروزی هستند. یادگیری عمیق که شکل تکامل یافته‌ی شبکه‌های عصبی مصنوعی است، بنیان و اساس کارکرد یادگیری ماشین را تغییر می‌دهد. به لطف مجموعه‌های بزرگ داده‌ها و ارتقاء منابع محاسباتی، یادگیری عمیق توانسته است تأثیر عمیقی بر نحوه‌ی کشف الگوهای پیچیده با استفاده از ویژگی‌های استخراج شده از داده‌ها بگذارد (بر خلاف دانشمندان علوم داده که بردار ویژگی را به الگوریتم یادگیری

می‌دهند). رایانش شناختی که اولین بار توسط آی‌بی‌ام واتسون و به واسطه‌ی موفقیتش در برابر بهترین بازیکنان در نمایش جئوپاردی! شهرت یافت این امکان را به وجود آورده است تا بتوان با گروه جدیدتری از مسائل سر و کله زد. انواعی از مسائل که تصور میشد فقط توسط هوش و خلاقیت انسانی قابل حل هستند و با عدم قطعیت و ابهام همراه هستند. در این فصل به بررسی مفاهیم، روش‌ها و کاربردهای این دو روند فناوری پیشرفته‌ی هوش مصنوعی می‌پردازیم.

۶.۱. پرونده‌ی آغازین: مبارزه با کلاهبرداری به کمک یادگیری عمیق و هوش مصنوعی.

۶.۲. مقدمه‌ای بر یادگیری عمیق

۶.۳. اصول و مبانی شبکه‌های عصبی " کم عمق "

۶.۴. فرایند توسعه‌ی سیستم‌های مبتنی بر شبکه‌های عصبی

۶.۵. روشن کردن جعبه‌ی سیاه شبکه‌های عصبی مصنوعی

۶.۶. شبکه‌های عصبی عمیق

۶.۷. شبکه‌های عصبی پیچشی (همگشتی)

۶.۸. شبکه‌های بازگشتی و شبکه‌های حافظه طولانی کوتاه مدت

۶.۹. چارچوب‌های رایانه‌ای برای پیاده‌سازی یادگیری عمیق

۶.۱۰. رایانش شناختی

دانسکه بانک^۱ یک بانک جهانی وابسته به شمال اروپا می‌باشد که ریشه‌های محلی و ارتباطات قوی با سایر بخش‌های جهان دارد. این بانک که در اکتبر ۱۸۷۱ بنیان نهاده شد به مدت ۱۴۵ سال است که به افراد و مشاغل در مناطق شمال اروپا به تحقق آمال و آرزوهایشان کمک کرده است. مقر اصلی آن در دانمارک و بازارهای اصلی آن در دانمارک، فنلاند، نروژ و سوئد می‌باشد. پیشگیری از کلاهبرداری و تقلب، اولویت اصلی بانک‌ها می‌باشد. طبق آمار انجمن مورد تأیید ارزیابان کلاهبرداری، کسب‌وکارها هر ساله بیش از ۳.۵ تریلیون دلار در کلاهبرداری‌ها از دست می‌دهند. این مسئله در صنعت مالی بسیار شایع بوده و هر ماهه بر پیچیدگی و گستردگی آن افزوده می‌شود. با گسترش فعالیت‌های بانک‌داری آنلاین مشتریان از طریق کانال‌ها و دستگاه‌های مختلف، فرصت‌های بیشتری برای وقوع کلاهبرداری ایجاد شده است. علاوه بر این، کلاهبردارها نیز خلاق‌تر و از لحاظ دانش تکنولوژیکی مطلع‌تر شده‌اند و آنها نیز از فناوری‌های پیشرفته‌ای مثل یادگیری ماشین استفاده می‌کنند.

روش‌های قدیمی برای شناسایی کلاهبرداری‌ها مثل استفاده از موتورهای قواعد نوشته شده توسط انسان فقط درصدی کوچکی از پرونده‌های کلاهبرداری را تشخیص داده و تعداد موارد مثبت غلط بسیار بیشتری را تولید می‌کنند. اگرچه موارد منفی غلط برای بانک‌ها هزینه به همراه دارند ولی پیگیری تعداد زیادی از موارد مثبت غلط نه تنها پول و زمان را هزینه می‌کند بلکه بر رضایت و اعتماد مشتری نیز خدشه وارد می‌کند. به منظور بهبود پیش‌بینی‌های احتمالی و شناسایی درصد بیشتری از موارد واقعی کلاهبرداری و کاهش هشدارهای اشتباه، بانک‌ها به شکل جدیدی از تجزیه و تحلیل‌ها نیاز دارند که شامل استفاده از هوش مصنوعی می‌باشد.

دانسکه بانک همانند سایر بانک‌های جهانی شاهد تغییر زلزله‌وار در تعاملات مشتریان است. در گذشته، اکثر مشتریان برای انجام تراکنش‌ها و معاملات خود به شعب بانکی مراجعه می‌کردند. امروزه، تقریباً تمامی معاملات و تراکنش‌ها به صورت دیجیتالی از طریق یک تلفن همراه، تبلت، دستگاه خودپرداز یا مرکز تماس انجام می‌پذیرد. بدین ترتیب "فضای سطحی" بیشتری برای وقوع کلاهبرداری فراهم می‌شود. بانک‌ها نیازمند این بودند که مکانیزم‌های تدافعی تشخیص کلاهبرداری را مدرن‌سازی کنند. نرخ تشخیص کلاهبرداری بانک در حد پایین ۴۰ درصد بوده و روزانه حداکثر ۱،۲۰۰ مورد مثبت غلط را مدیریت می‌کرده و ۹۹.۵ درصد از تمامی مواردی که بانک به آنها رسیدگی می‌کرد هیچ ارتباطی به کلاهبرداری نداشتند. چنین تعداد زیادی از هشدارهای غلط نیازمند سرمایه‌گذاری پولی، زمانی و نیروی انسانی برای تحقق و تفحص بود که نهایتاً به بن بست می‌رسیدند. دانسکه بانک به همکاری تینک بیگ آنالیتیکز^۲ که یکی از شرکت‌های زیرمجموعه ترادیتا است یک تصمیم استراتژیک برای بکارگیری تکنیک‌های تحلیلی نوآورانه از جمله هوش مصنوعی برای تشخیص بهتر موارد کلاهبرداری و کاهش موارد مثبت غلط اتخاذ کرد.

دانسکه بانک از یادگیری عمیق در دستگاه‌های واحد پردازنده گرافیکی استفاده کرد که برای یادگیری عمیق نیز بهینه‌سازی شده بودند. این سیستم نرم‌افزاری جدید به تیم تحلیل در شناسایی موارد بالقوه‌ی کلاهبرداری و در عین حال، اجتناب از موارد

¹ Danske Bank

² Think Big Analytics

مثبت غلط کمک کرد. تصمیمات عملیاتی از سمت کاربران به سمت سیستم‌های هوش مصنوعی شیفت پیدا کرده‌اند. با وجود این، هنوز در برخی موارد به مداخله‌ی انسانی نیاز است. به عنوان مثال، این مدل می‌تواند ناهنجاری‌هایی از قبیل خریدهای کارت نقدی که در سرتاسر جهان رخ می‌دهند را شناسایی کنید ولی برای تعیین اینکه آیا این مورد یک کلاهبرداری است یا اینکه مشتری بانک صرفاً یک خرید آنلاین انجام داده و پرداختی را برای کشور چین ارسال کرده و روز بعد کالایی را از یک فروشگاه در لندن خریداری کرده است به یک تحلیلگر انسانی نیاز است.

رویکرد تحلیلی دانسکه بانک از متدولوژی "قهرمان/چالشگر" بهره می‌برد. با این روش، سیستم‌های یادگیری عمیق می‌توانند برای تعیین مؤثرترین مدل، آنها را در لحظه با یکدیگر مقایسه کنند. هر چالشگر، داده‌ها را به صورت بلادرنگ پردازش می‌کند و همچنان که در مسیر تشخیص رفتارهایی که احتمال کلاهبرداری دارند پیش می‌رود به یادگیری می‌پردازد. اگر یک فرایند به زیر آستانه مشخص شده برسد، آنگاه داده‌های بیشتری به آن تزریق می‌شود مثل موقعیت‌یابی جغرافیایی مشتری یا تراکنش‌های اخیر دستگاه‌های خودپرداز. زمانیکه عملکرد یک چالشگر بهتر از دیگری باشد، تبدیل به قهرمان می‌شود که نقشه راهی را برای موفقیت سایر مدل‌ها در تشخیص کلاهبرداری‌ها در اختیارشان قرار می‌دهد.

نتیجه‌گیری

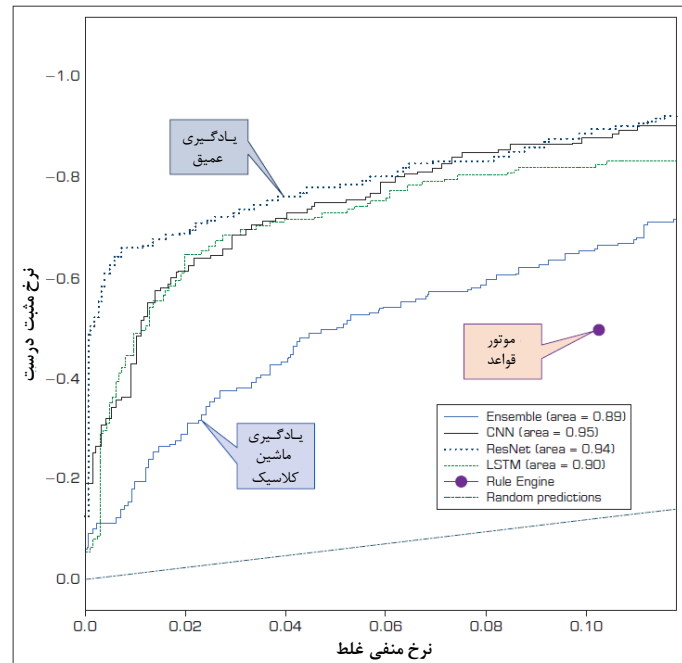
دانسکه بانک یک راه‌حل تحلیلی سازمانی پیشرفته را با بهره‌گیری از هوش مصنوعی و یادگیری عمیق پیاده‌سازی کرد. بانک توانست:

- تعداد موارد مثبت غلط را تا ۶۰ درصد کاهش دهد و انتظار می‌رود تا این میزان را به ۸۰ درصد برساند.
- موارد مثبت درست را به میزان ۵۰ درصد افزایش دهد.
- منابع خود را بر روی موارد حقیقی کلاهبرداری متمرکز کند.

گراف زیر (شکل ۶.۱) نشان می‌دهد که چگونه نرخ موارد مثبت غلط و درست با بهره‌گیری از تحلیل‌های پیشرفته (از جمله یادگیری عمیق) بهبود یافته‌اند. نقطه قرمز بیانگر موتور قواعد قدیمی است که قادر بودند تا فقط ۴۰ درصد از کل کلاهبرداری‌ها را تشخیص دهند. یادگیری عمیق به میزان چشمگیری بر اساس یادگیری ماشین بهبود یافته و به دانسکه بانک اجازه داده است تا کلاهبرداری‌ها را بهتر و با مثبت‌های غلط کمتری تشخیص دهند.

تحلیل‌های سازمانی به سرعت در حال رشد بوده و وارد سیستم‌های یادگیری جدید مبتنی بر هوش مصنوعی شده‌اند. در همین حین، سخت‌افزار و پردازنده‌ها نیز قدرتمندتر و تخصصی‌تر شده‌اند و الگوریتم‌های بیشتری در دسترس هستند از جمله الگوریتم‌های منبع باز. این قابلیت‌ها به بانک‌ها راه‌حل‌هایی قدرتمندی را برای شناسایی و پیشگیری از کلاهبرداری‌ها ارائه می‌کند. همانطور که دانسکه بانک آموخت، ارزشی که از ایجاد و جاری‌سازی یک راه‌حل تحلیلی سازمانی که پاسخگوی نیازهای خاص آن بوده و از منابع داده آن بهره بگیرد ایجاد می‌شود بیش از ابزارهای سنتی و قدیمی است. دانسکه بانک به کمک هوش مصنوعی و یادگیری عمیق اکنون توانایی این را دارد تا کلاهبرداری‌ها را بهتر و بدون اینکه بار موارد مثبت غلط زیادی را بر دوش بکشد شناسایی کند. این راه‌حل همچنین به مهندسان بانک، دانشمندان علم داده‌ها و افسران رسیدگی و تحقیق از پلیس بین‌الملل، پلیس محلی و سایر ارگان‌ها اجازه می‌دهد تا برای کشف کلاهبرداری‌ها از جمله حلقه‌های کلاهبرداری پیچیده پرده بردارند. این راه‌حل تحلیلی سازمانی با برخورداری از قابلیت‌های پیشرفته در حال حاضر در سایر حوزه‌های کسب‌وکار بانک نیز مورد استفاده قرار گرفته و خلق ارزش می‌کند.

چونکه این فناوری‌ها هنوز در حال رشد و تکامل هستند، پیاده‌سازی و اجرای راه‌حل‌های یادگیری عمیق و هوش مصنوعی برای شرکت‌ها به تنهایی می‌تواند دشوار باشد. آنها از طریق شراکت با شرکتی که قابلیت‌های اثبات شده برای اجرای راه‌حل‌های تکنولوژیکی دارد می‌توانند به نتایج ارزشمندی دست پیدا کنند. همانطور که در این مورد نشان داده شده است، تینک بیگ آنالیتیکز از تخصص لازم برای پیکربندی چارچوب‌های سخت‌افزاری و نرم‌افزاری تخصصی برای عملیاتی‌سازی فرایندهای جدید برخوردار است. این پروژه شامل یکپارچه‌سازی راه‌حل‌های منبع-باز، جاری‌سازی مدل‌های تولید و سپس بکارگیری تحلیل‌های یادگیری عمیق برای گسترش و بهبود مدل‌ها بوده است. یک چارچوب برای مدیریت و ردیابی مدل‌ها در سیستم تولید ایجاد شد. این مدل‌ها به سیستم زیربنایی این امکان را می‌دادند تا تصمیمات مستقل را به صورت بلادرنگ اتخاذ کنند که همسو با اصول راهنمای امنیتی بانک بوده‌اند. این راه‌حل، سطوح جدیدی از جزئیات را از قبیل سری‌های زمانی و توالی رویدادها برای کمک بهتر به بانک در تحقیق و بررسی‌های کلاهبرداری‌ها فراهم کرد. کل راه‌حل بسیار سریع پیاده‌سازی و اجرا شد - از زمان شروع تا اجرایی شدن فقط پنج ماه به طول انجامید. شکل ۶.۲ یک چارچوب تعمیم یافته از راه‌حل‌های تحلیلی سازمانی مبتنی بر هوش مصنوعی و یادگیری عمیق را نشان می‌دهد.

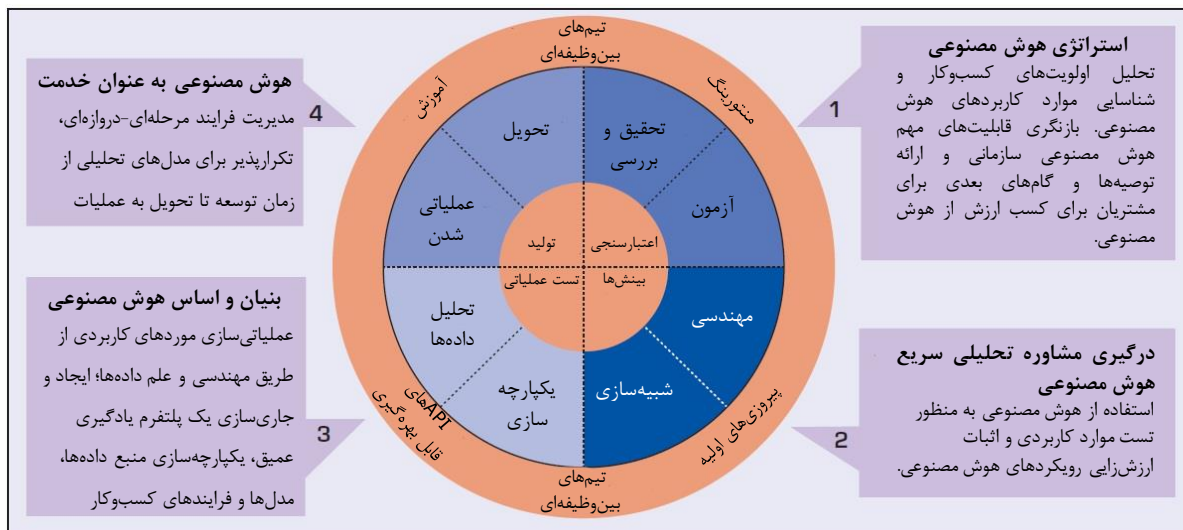


شکل ۶.۱ یادگیری عمیق موجب بهبود مثبت‌های درست و منفی‌های درست می‌شود.

به طور خلاصه، دانسکه بانک یک پروژه‌ی چندمرحله‌ای را به منظور بکارگیری تکنیک‌های یادگیری ماشین و توسعه‌ی مدل‌های یادگیری عمیق برای تست آن تکنیک‌ها اجرا کرد. مدل‌های یکپارچه به شناسایی مشکل فزاینده‌ی کلاهبرداری کمک کردند. برای یک چکیده‌ی دیداری به ویدئوی زیر نگاه کنید: (<https://www.teradata.com/Resources/Videos/Danske-Bank>)

(Innovating-in-Artificial-Intelligence) و یا وبلاگ زیر را مطالعه کنید

(<http://blogs.teradata.com/customers/danske-bank-innovating-artificial-intelligence-deep-learningdetect-sophisticated-fraud/>).



شکل ۶.۲ یک چارچوب تعمیم یافته از راه‌حل‌های تحلیلی مبتنی بر هوش مصنوعی و یادگیری عمیق

سوالاتی برای مرور پرونده‌ی آغازین

۲۳. کلاهبرداری در بانک‌داری به چه معناست؟
۲۴. موسسات بانکی امروزه با چه نوع کلاهبرداری‌هایی مواجه هستند؟
۲۵. به نظر شما تأثیرات کلاهبرداری بر بانک‌ها و مشتریان آنها چه می‌باشند؟
۲۶. روش‌های قدیمی و جدید شناسایی و پیشگیری از کلاهبرداری را مقایسه کنید؟
۲۷. به نظر شما چرا روش‌های یادگیری عمیق از دقت پیش‌بینی بهتری برخوردارند؟
۲۸. در مورد توازن بین موارد مثبت غلط و منفی غلط (خطاهای نوع ۱ و ۲) در مقوله‌ی پیش‌بینی فعالیت‌های مشکوک به کلاهبرداری بحث کنید.

همانطور که در این فصل خواهید دید، هوش مصنوعی به طور کلی و روش‌های یادگیری ماشین به طور خاص به سرعت در حال رشد و تکامل هستند. استفاده از منابع داده رقوم‌سازی شده بزرگ درون و برون سازمانی، ساختنیافته و ساختنیافته به همراه سیستم‌های رایانشی پیشرفته (ترکیب نرم‌افزار و سخت‌افزار)، راه را به سوی حل مشکلاتی که تا چندین سال پیش لاینحل به نظر می‌رسیدند باز کرده است. یادگیری عمیق و رایانش شناختی از طریق بکارگیری منابع سریع‌اً در حال گسترش کلان‌داده‌ها به سازمان‌ها کمک می‌کنند تا تصمیمات دقیق و به‌موقعی را اتخاذ کنند. همانگونه که در این پرونده دیدیم، این

نسل جدید از سیستم‌های هوش مصنوعی در مقایسه با همتایان قدیمی‌تر خود بهتر قادر به حل مسائل هستند. در حوزه‌ی تشخیص کلاهبرداری، روش‌های قدیمی همواره تا حدودی مفید بوده‌اند و نرخ مثبت‌های غلط آنها بسیار بالا بوده و منجر به تحقیق و تفحص‌های نادر و نارضایتی مشتریان‌شان شده است. در خصوص مسائل دشوار مثل تشخیص کلاهبرداری، فناوری‌های نوین هوش مصنوعی از قبیل یادگیری عمیق توانسته‌اند آنها را با دقت و کاربردی‌تری بالایی حل کنند.

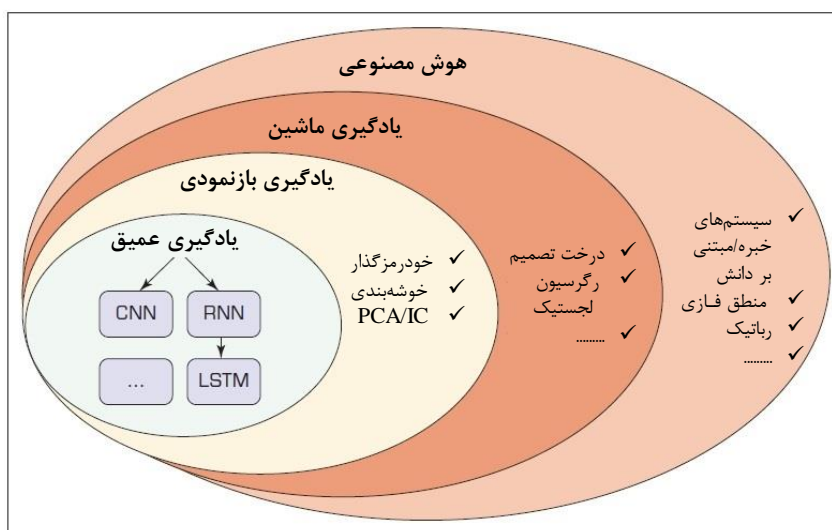
Source: Teradata Case Study. "Danske Bank Fights Fraud with Deep Learning and AI." <https://www.teradata.com/Resources/Case-Studies/Danske-Bank-Fight-Fraud-With-Deep-Learning-and-AI> (accessed August 2018). Used with permission

حدود یک دهه قبل، محاوره و گفتگو با یک دستگاه الکترونیکی (به صورت هوشمند و به زبان انسان) غیرقابل تصور و چیزی بود که تنها می‌توانستیم در فیلم‌های علمی-تخیلی مشاهده کنیم. امروزه، به لطف پیشرفت‌های به عمل آمده در فناوری‌ها و روش‌های هوش مصنوعی، تقریباً همه ما این پدیده‌ی غیرقابل تصور را تجربه کرده‌ایم. ممکن است قبلاً چندین بار از سیری یا دستیار گوگل خواسته باشید که تا شماره‌ای را از دفترچه تلفن گوشی شما گرفته یا آدرسی را پیدا کرده و مسیرهای مشخصی را در حین رانندگی به شما نشان دهند. برخی اوقات زمانیکه خسته هستید ممکن است از گوگل هوم یا آمازون الکسا خواسته باشید تا موسیقی مورد علاقه‌تان را بر روی دستگاه یا تلویزیون شما پخش کنند. ممکن است زمانیکه عکس گروهی از دوستان‌تان را بر روی فیسبوک بارگذاری کرده‌اید و مشاهده کرده‌ایم که برچسب‌های نام‌گذاری پیشنهادی دقیقاً با چهره دوستان شما در تصویر تطابق دارد شگفت‌زده شده باشید. برای ترجمه‌ی یک متن از یک زبان خارجی نیازی نیست تا ساعت‌ها درگیر استفاده از فرهنگ واژگان باشیم فقط کافی است تصویر آن متن را در مترجم گوگل وارد کرده و چند ثانیه صبر کنید. اینها تعداد معدودی از کاربردهای بسیار زیاد و روزافزون یادگیری عمیق هستند که امیدبخش زندگی آسان‌تر برای انسان‌ها هستند.

یادگیری عمیق به عنوان جدیدترین و شاید در حال حاضر، معروف‌ترین عضو خانواده‌ی هوش مصنوعی و یادگیری ماشین، هدفی مشابه سایر روش‌های یادگیری ماشین که پیش از آن وجود داشته‌اند دارد: تقلید فرایند فکری و ذهنی انسان‌ها با استفاده از الگوریتم‌های ریاضی برای یادگیری از داده‌ها به عنوان شکل و روشی که انسان‌ها قادر به یادگیری هستند. پس یادگیری عمیق در واقع چه تفاوتی با آنها دارد؟ در اینجا به برخی از مشخصه‌های متمایزکننده‌ی یادگیری عمیق نسبت به یادگیری ماشین سنتی اشاره می‌کنیم. عملکرد الگوریتم‌های یادگیری ماشین سنتی مثل درخت‌های تصمیم، ماشین‌های بردار پشتیبانی، رگرسیون لجستیک شبکه‌های عصبی شدیداً متکی بر نمایش و بازنمود داده‌ها است. این بدان معناست که اگر ما (متخصصان تجزیه و تحلیل یا دانشمندان علم داده‌ها) اطلاعات مرتبط و کافی را به فرمت مناسبی در اختیار الگوریتم‌های یادگیری ماشین سنتی قرار دهیم آنها قادر هستند تا الگوها را یاد گرفته و وظایف پیش‌بینی (طبقه‌بندی یا تخمین)، خوشه‌بندی یا وابستگی را دقت قابل قبولی انجام دهند. به عبارت دیگر، این الگوریتم‌ها نیازمند این هستند که انسان‌ها به صورت دستی، ویژگی‌هایی که از لحاظ نظری و/یا منطقی با اهداف مسئله‌ی موجود مرتبط هستند را شناسایی کرده و آنها را به یک شکل مناسب وارد الگوریتم کنند. به عنوان مثال، برای استفاده از یک درخت تصمیم به منظور پیش‌بینی احتمال بازگشت یک مشتری، مدیر بازاریابی باید اطلاعاتی از قبیل مشخصه‌های اجتماعی-اقتصادی مشتری از قبیل درآمد، شغل، سطح تحصیلات و غیره به همراه تعاملات تاریخی و جمعیت‌شناختی آن مشتری با شرکت را برای الگوریتم فراهم کند. اما الگوریتم به خودی خود قادر به تعریف و تعیین این مشخصه‌های اجتماعی-اقتصادی و استخراج اینگونه ویژگی‌ها به عنوان مثال از فرم‌های نظرسنجی تکمیل شده توسط مشتری یا بدست آمده از رسانه‌های اجتماعی نمی‌باشد.

اگرچه چنین رویکرد یادگیری ماشین انسانی ساختیافته‌ای برای وظایف رسمی و نسبتاً انتزاعی به خوبی کار کرده است ولی استفاده از آن برای وظایف غیررسمی ولی به ظاهر ساده (برای انسان‌ها) مثل تشخیص چهره یا بازشناخت صدا می‌تواند بسیار چالش‌برانگیز باشد زیرا این وظایف به دانش بسیار زیادی از دنیای اطراف‌مان نیاز دارند (گودفلا و همکاران ۲۰۱۶). به عنوان مثال، یک الگوریتم یادگیری ماشین را نمی‌توان به آسانی آموزش داد تا بتواند صرفاً با استفاده از چند ویژگی دستوری و مفهومی، مفهوم واقعی جمله‌ای که توسط شخصی بیان شده است را درک کند. انجام اینگونه کارها نیازمند دانشی «عمیق» در مورد جهانی است که مدون‌سازی و نمایش صریح آن کار آسانی نیست. آنچه یادگیری عمیق به روش‌های یادگیری ماشین کلاسیک افزوده این واقعیت توانایی اکتساب خودکار دانش مورد نیاز برای انجام وظایف غیررسمی و در نتیجه، استخراج برخی ویژگی‌های پیشرفته است که منجر به عملکرد سطح بالای سیستم می‌شود.

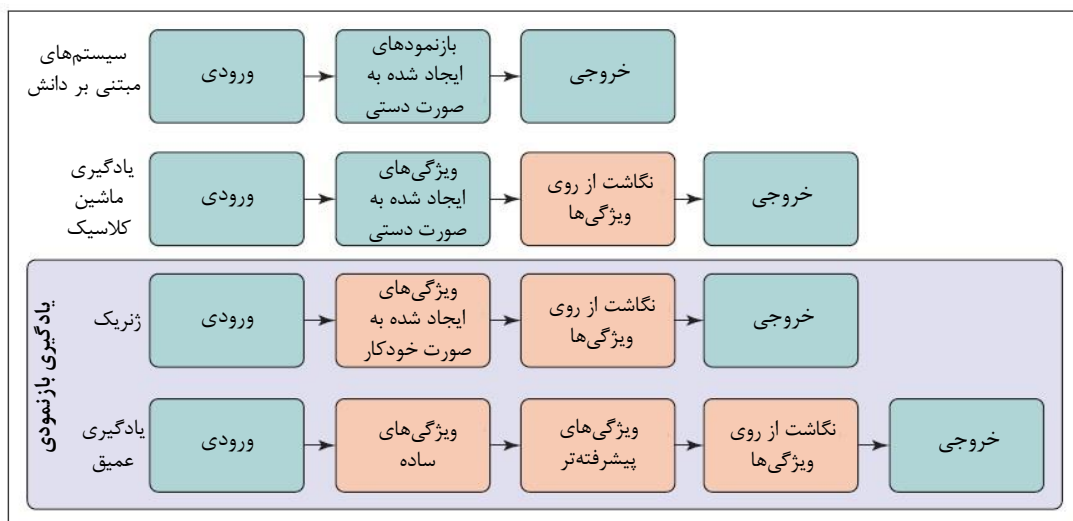
برای دستیابی به درکی نزدیک از یادگیری عمیق باید ببینیم که در تصویر بزرگ تمامی روش‌های دیگر خانواده‌ی هوش مصنوعی چه جایگاهی دارد. یک دیاگرام ارتباطات سلسله‌مراتبی ساده یا یک روش نمایش طبقه‌بندی می‌تواند چنین درک کل‌نگری را به ما بدهد. گودفلا و همکارانش (۲۰۱۶) در تلاش برای اینکار، یادگیری عمیق را به عنوان بخشی از روش‌های خانواده‌ی یادگیری بازنمودی دسته‌بندی کردند. تکنیک‌های یادگیری بازنمودی شامل یک نوع یادگیری ماشین است که در آن تأکید بر ویژگی‌های یادگیری و کشف توسط سیستم علاوه بر کشف نگاشت‌ها از روی آن ویژگی‌ها بر روی هدف/خروجی است. شکل ۶.۳ از یک نمودار ون برای ترسیم جایگیری یادگیری عمیق در خانواده‌ی روش‌های یادگیری مبتنی بر هوش مصنوعی استفاده می‌کند.



شکل ۶.۳ یک نمودار ون که جایگیری یادگیری عمیق را در خانواده‌ی روش‌های یادگیری مبتنی بر هوش مصنوعی نشان می‌دهد

شکل ۶.۴ به بررسی تفاوت‌های بین گام‌ها/وظایفی که باید در زمان ساخت یک مدل یادگیری عمیق معمول انجام داد در مقابل گام‌ها/وظایفی که در زمان ساخت مدل‌ها در الگوریتم‌های یادگیری ماشین کلاسیک باید انجام شوند می‌پردازد. همانطور که در دو جریان کار بالا نشان داده شده است، در سیستم‌های مبتنی بر دانش و روش‌های یادگیری ماشین کلاسیک، دانشمندان برای دستیابی به نتیجه‌ی مطلوب باید ویژگی‌ها (یعنی بازنمود) را به صورت دستی ایجاد کنند. پایینی‌ترین جریان کارها نشان می‌دهند که یادگیری عمیق، رایانه‌ها را قادر می‌سازد تا برخی ویژگی‌های پیچیده را از مفاهیم ساده‌ای که کشف آنها به صورت دستی برای انسان کار بسیار دشوار و پردردسری است (و شاید در برخی موارد ناممکن) بدست بیاورند و سپس آن ویژگی‌های پیشرفته را بر روی خروجی مورد نظرشان نگاشت کنند.

از یک نقطه نظر متدولوژیکی، اگرچه یادگیری عمیق معمولاً حوزه‌ی جدیدی در یادگیری ماشین محسوب می‌شود ولی ایده اصلی و اولیه‌ی آن به دهه ۸۰ و تنها چند دهه پس از پیدایش شبکه‌های عصبی مصنوعی که لکان و همکارانش (۱۹۸۹) مقاله‌ای را در مورد بکارگیری شبکه‌های پس‌انتشار برای تشخیص کدپستی‌های دست‌نویس منتشر کردند باز می‌گردد. در واقع، یادگیری عمیق به نظر می‌رسد که می‌تواند با بکارگیری لایه‌های متعددی از نورون‌های بهم متصل و مجموعه داده‌های بسیار بزرگتر، متغیرها را به صورت خودکار تعریف و مسائل را حل نماید ولی اینکار فقط به قیمت تلاش محاسباتی بسیار زیاد میسر خواهد شد. این نیاز محاسباتی بسیار زیاد و نیاز به مجموعه داده‌های بسیار بزرگ دو دلیل اصلی بودند که ایده‌ی اولیه‌ی اینکار برای بیش از دو دهه و تا زمانیکه زیرساخت‌های محاسباتی و تکنولوژیکی پیشرفته‌ای برای تحقق عملی یادگیری عمیق پدیدار شوند به تعویق افتاد. اگرچه مقیاس شبکه‌های عصبی در دهه گذشته و به واسطه‌ی پیشرفت فناوری‌های مربوطه به شدت افزایش یافته است ولی همچنان تخمین زده می‌شود که دستیابی به شبکه‌های عصبی عمیق مصنوعی که از تعداد نورون‌ها و سطح پیچیدگی همچون مغز انسان برخوردار باشند چندین دهه دیگر به طول خواهد انجامید.



شکل ۶.۴ تفاوت‌های اساسی بین روش‌های یادگیری ماشین کلاسیک و یادگیری عمیق/یادگیری بازنمودی

علاوه بر زیرساخت‌های رایانه‌ای همانطور که گفته شد، در دسترس بودن مجموعه داده‌های رقومی‌سازی بزرگ و غنی، دلیل اصلی دیگری برای توسعه‌ی برنامه‌های کاربردی یادگیری عمیق در سال‌های اخیر بوده است. سابق بر این، دستیابی به عملکرد خوب از الگوریتم یادگیری عمیق کار بسیار دشواری بود که نیازمند مهارت‌های گسترده و تجربه/درک عمیق برای طراحی شبکه‌های خاص بود که به همین دلیل، تعداد بسیار معدودی قادر به توسعه‌ی یادگیری عمیق برای مقاصد کاربردی و/یا پژوهشی بودند. مجموعه داده‌های آموزشی بزرگ ولی تا حدودی زیادی این فقدان دانش را جبران کرده و میزان مهارت مورد نیاز برای پیاده‌سازی و اجرای شبکه‌های عصبی عمیق را کاهش داده‌اند. با وجود این، اگرچه حجم مجموعه داده‌های موجود در سال‌های اخیر رشد نمایی داشته است، یک چالش بزرگ به ویژه برای شبکه‌های نظارت شده یادگیری عمیق در حال حاضر، برچسب‌گذاری موارد در

این مجموعه داده‌های عظیم است. در نتیجه، پژوهش‌های بسیاری در حال اجرا است که تمرکز آنها بر چگونگی بهره‌گیری از حجم انبوهی از داده‌های برچسب‌گذاری نشده برای یادگیری نظارت نشده یا نیمه نظارت شده یا نحوه‌ی توسعه‌ی روش‌هایی برای برچسب‌گذاری مثال‌ها به طور دسته‌جمعی در یک زمان قابل قبول است.

در بخش بعدی این فصل به معرفی کلی شبکه‌های عصبی که خاستگاه یادگیری عمیق است خواهیم پرداخت. پس از مروری بر این شبکه‌های عصبی "کم عمق"، به معرفی انواع مختلف معماری‌های یادگیری عمیق و چگونگی عملکرد آنها، برخی کاربردهای متداول این معماری‌های یادگیری عمیق و برخی چارچوب‌های رایانه‌های معروف برای پیاده‌سازی و اجرای یادگیری عمیق در عمل می‌پردازیم. از آنجائیکه اصول و مبنای یادگیری عمیق و شبکه‌های عصبی مصنوعی یکسان هستند، در بخش بعد به بررسی اجمالی معماری شبکه‌های عصبی (شبکه‌های عصبی پرسپترون چندلایه) برای تمرکز بر اصول ریاضی آنها و توضیح نحوه‌ی استخراج انواع مختلف رویکردها/معماری‌های یادگیری عمیق از این اصول و مبنای می‌پردازیم. مورد کاربردی ۶.۱ یک مثال جالب از آنچه تکنیک‌های تحلیلی پیشرفته و یادگیری عمیق می‌توانند در رشته‌ی فوتبال بدست آورند ارائه کرده است.

کشف ستاره‌ی بعدی فوتبال به کمک هوش مصنوعی

مورد کاربردی ۶.۱



فوتبال. این بازی زیبا. هر طور که آن را بنامید، یک استارت‌آپ هلندی با بکارگیری هوش مصنوعی در این حوزه در حال تغییر شکل دادن محبوب‌ترین ورزش دنیا است. شرکت سای‌اسپورتز^۱ که در سال ۲۰۱۲ توسط دو معتاد به فوتبال بنیانگذاری شد در حال نوآوری است. این شرکت تحلیل ورزش از تکنیک‌های یادگیری ماشین، یادگیری عمیق و هوش مصنوعی برای ضبط و تحلیل داده‌ها استفاده کرده و راه را برای نوآوری در هر چیزی از استخدام بازیکن گرفته تا واقعیت مجازی طرفداران باز می‌کند.

فناوری سطح بالا در انتخاب بازیکن

در عصر قراردادهای هشت رقمی، استخدام بازیکن یک بازی بسیار پرمخاطره‌ای است. بهترین تیم‌ها لزوماً آنهايي نیستند که بهترین بازیکنان را دارند بلکه بهترین ترکیب بازیکنان است که اهمیت دارد. مربیان برای دهه‌های متمادی از مشاهده، شهود و داده‌های ابتدایی استفاده کرده‌اند ولی باشگاه‌های آگاه و زیرک در حال حاضر از تحلیل‌های پیشرفته برای شناسایی ستاره‌های

¹ SciSports

نوظهور و بازیکنانی که قدر و ارزش آنها دست کم گرفته شده است بهره می‌برند. به گفته‌ی جیلز برویر، مدیرعامل و بنیان‌گذار سای‌اسپورتز، "شاخص SciSkill عملکرد هر بازیکن حرفه‌ای فوتبال در جهان را در قالب یک شاخص عمومی و جهانی ارزیابی می‌کند." این شرکت از الگوریتم‌های یادگیری ماشین برای محاسبه‌ی کیفیت، استعداد و ارزش بیش از ۲۰۰,۰۰۰ بازیکن استفاده می‌کند. این به باشگاه‌ها در یافتن استعدادها، جستجوی بازیکنانی که دقیقاً با نیازهایشان منطبق باشند و تحلیل رقبایشان کمک می‌کند.

هر هفته، بیش از ۱,۵۰۰ مورد منطبق در ۲۱۰ لیگ به کمک فناوری SciSkill تجزیه و تحلیل می‌شوند. باشگاه‌های تراز اول فوتبال در اروپا و سایر قاره‌ها از طریق شراکت با سای‌اسپورتز که مجهز به این بینش هستند قادر به یافتن و جذب بازیکنان مناسب هستند. این منجر به جذب ناخواسته و غیرمنتظره بازیکنان شده است. به عنوان مثال، یک بازیکن لیگ دسته دوم فوتبال تمایلی به تمدید قرارداد خود نداشت و به همین دلیل به عنوان یک عامل آزاد از قرارداد خارج شد. یک باشگاه جدید شاخص SciSkill را بررسی کرده و متوجه این داده‌های کنجکاوبرانگیز شد. باشگاه در ابتدا چندان مطمئن نبود زیرا به نظر می‌رسید که در امر مربیگری ناشی و بی‌تجربه باشد ولی داده‌ها حقیقت ماجرا را بیان کردند. باشگاه او را به عنوان مهاجم سوم جذب کرد و او به سرعت تبدیل به امتیازآورترین بازیکن شد. حق امتیاز وی در عرض دو سال به رقم بسیار بالایی فروخته شدند و او در حال حاضر پرگل‌ترین بازیکن حرفه‌ای فوتبال هلند است.

تحلیل بازی سه بُعدی بلادرنگ

شرکت‌های سنتی داده‌های فوتبال، فقط داده‌های مربوط به بازیکنانی که توپ را در اختیار دارند تولید می‌کنند و سایر موارد را از بررسی کیفیت بازیکن خارج می‌کنند. سای‌اسپورتز با دیدن این فرصت برای ضبط حجم انبوهی از داده‌ها به غیر از آنچه برای توپ اتفاق می‌افتد یک سیستم دوربین به نام بال‌جیمز^۱ را توسعه داد.

بال‌جیمز یک فناوری ردیابی بلادرنگ است که به طور خودکار به تولید داده‌های سه‌بعدی از ویدئوها می‌پردازد. چهارده دوربین که در اطراف استادیوم تعبیه شده‌اند هر تحرک در زمین بازی را ضبط می‌کنند. سپس بال‌جیمز داده‌هایی از قبیل دقت، جهت و سرعت پاس‌ها، قدرت دویدن و قدرت پرش را تولید می‌کند. برویر می‌گوید: "بدین ترتیب تصویر کاملی از بازی بدست می‌آید. از این داده‌ها می‌توان به شکل‌های بسیار جالبی استفاده کرد مثلاً به طرفداران اجازه می‌دهد تا با استفاده از واقعیت مجازی بازی را از هر زاویه‌ای که بخواهند تماشا کنند." وی می‌افزاید که داده‌ها حتی می‌توانند به مربیان که بر روی نیمکت نشسته‌اند کمک کنند. "زمانیکه می‌خواهند بفهمند که آیا بازیکنی خسته شده است می‌توانند بر اساس تحلیل‌ها بازیکن را تعویض کنند."

یادگیری ماشین و یادگیری عمیق

سای‌اسپورتز به مدل‌سازی تحرکات در زمین بازی با استفاده از الگوریتم‌های یادگیری ماشین می‌پردازد که با کسب تجربه‌ی بیشتر توسط بازیکن، موجب بهبود عملکرد او می‌شود. بال‌جیمز به طور خودکار به هر اقدام در زمین بازی یک مقدار تخصیص می‌دهد مثلاً ضربه‌ی کورنر. در طول زمان، این مقادیر بر اساس نرخ موفقیت‌شان تغییر می‌کنند. به عنوان مثال، زدن گل امتیاز بالایی دارد ولی نقش داشتن در گل - که ممکن است قبلاً امتیاز یا مقدار پایینی داشته - می‌تواند با پیشرفت بیشتر در بازی امتیاز بالاتری کسب کند. ووتر روزنبرگ^۲، افسر ارشد فناوری سای‌اسپورتز می‌گوید هوش مصنوعی و یادگیری ماشین نقش بسیار مهمی را در آینده‌ی سای‌اسپورتز و تحلیل فوتبال ایفا خواهند کرد. "مدل‌های ریاضی فعلی به مدل‌سازی دانش و

¹ BallJames

² Wouter Rosenberg

بینش‌های موجود در فوتبال می‌پردازند در حالیکه هوش مصنوعی و یادگیری ماشین امکان کشف ارتباطات جدیدی را فراهم می‌کنند که خود افراد قادر به برقراری آنها نیستند."

بال‌جیمز برای کامپایل دقیق تصاویر سه بُعدی باید بازیکنان، داورها و توپ را تشخیص دهد. سیستم پردازش جریان رویداد اس‌ای‌اس امکان بازساخت آنی تصویر را با استفاده از مدل‌های یادگیری عمیق فراهم می‌کند. به گفته‌ی روزنبرگ، "با ترکیب مدل‌های یادگیری عمیق خودمان با 'SAS'Viya، می‌توانیم مدل‌هایمان را در فضای ابری، بر روی دوربین‌ها یا هرکجا که منابع ما قرار دارد آموزش دهیم." توانایی استقرار مدل‌های یادگیری عمیق بر روی دوربین‌ها و سپس استنتاج لحظه‌ای یک علم بسیار پیشرفته است. روزنبرگ می‌گوید: "برخورداری از یک پلتفرم متحدالشکل برای مدیریت کل زنجیره‌ی تولید سه بُعدی فوق‌العاده چیز ارزشمندی است و بدون SAS Viya این پروژه امکان‌پذیر نبود."

افزودن Oomph به منبع باز

سای اسپورترز تا پیش از این انحصاراً از منبع باز برای ساخت مدل‌ها استفاده می‌کرد و از یک پلتفرم سر-به-سر بهره می‌برد که به تیم‌های تحلیل اجازه می‌داد تا به زبان مورد نظرشان کار کرده و یک موجودی دارایی تحلیلی را در سرتاسر سازمان به اشتراک بگذارند. طبق گفته‌ی برویر، شرکت بدین ترتیب قادر بود تا کارکنانی با مهارت‌های منبع-باز مختلف را جذب کرده و زنجیره تولید خود را با استفاده از یک پلتفرم مدیریت کند. به گفته‌ی برویر، "افسر ارشد فناوری من به من می‌گوید که عاشق این است که دانشمندان علوم داده‌ی ما می‌توانند کلیه پژوهش‌هایشان را در منبع باز انجام دهند و نیازی نیست که نگران تولید مدل‌ها باشیم. ما آنچه در نرم‌افزار پایتون باید ۱۰۰ خط کد برایش نوشته شود در نرم‌افزار اس‌ای‌اس تنها پنج خط کد لازم دارد. بدین ترتیب سرعت ارائه محصول به بازار که در تحلیل ورزشی بسیار مهم و حیاتی است افزایش می‌یابد."



شرکت سای اسپورترز از زمان شکل‌گیری خود به سرعت به یکی از پرنسپالترین شرکت‌های تحلیل ورزشی جهان تبدیل شده است. برویر می‌گوید که همه‌کاره بودن پلتفرم اس‌ای‌اس یک عامل مهم در این روند بوده است. "به کمک اس‌ای‌اس، ما توانایی این را داریم که قدرت پردازش را بسته به نیاز کم و زیاد کنیم، مدل‌ها را در لحظه تولید کنیم، هر چیزی که می‌خواهیم را در قالب یک مدل توسعه داده و آنها را با منبع باز ادغام کنیم. هدف غایی ما این است که تحلیل بلادرنگ داده‌ها را در اختیار میلیاردها طرفدار بازی فوتبال در سرتاسر جهان بگذاریم. از طریق شراکت با اس‌ای‌اس است که این امر محقق می‌شود."

Sources: SAS Customer Stories. "Finding the Next Football Star with Artificial Intelligence." www.sas.com/en_us/customers/scisports.html (accessed August 2018). Copyright (c) 2018 SAS Institute Inc., Cary, NC, USA. All Rights Reserved. Used with permission.

۱۹. سای اسپورتز چه کار می کند؟ برای اطلاعات بیشتر به وب سایت آن مراجعه کنید.
۲۰. تحلیل های پیشرفته چگونه می توانند به تیم های فوتبال کمک کنند؟
۲۱. نقش یادگیری عمیق در راه حل های ارائه شده توسط سای اسپورتز چیست؟

سوالاتی برای مرور بخش ۶.۲

۱۰. یادگیری عمیق چیست؟ چه کاری می تواند انجام دهد؟
۱۱. تفاوت اساسی یادگیری عمیق با یادگیری ماشین سنتی چیست؟
۱۲. روش های مختلف یادگیری در هوش مصنوعی را نام برده و به اختصار توضیح دهید.
۱۳. یادگیری بازنمودی چیست و چه ارتباطی با یادگیری عمیق دارد؟

" "

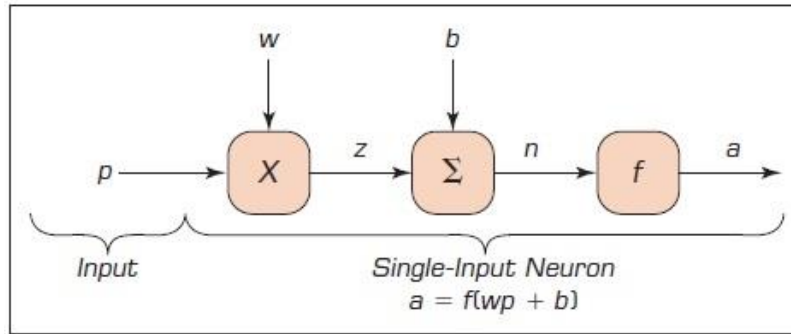
در این تصویر، p بیانگر یک ورودی عددی است. هر ورودی با یک وزن تنظیم پذیر w و یک مقدار ثابت بایاس b وارد نورون می شود. مقدار وزن از طریق یک ضرب تابع وزنی در ورودی قرار می گیرد و یک تابع ورودی خالص^۲ که توسط Σ نمایش داده می شود، مقدار بایاس را به ورودی موزون z می افزاید. خروجی تابع ورودی خالص (که به عنوان ورودی خالص نیز شناخته می شود) وارد تابع دیگری تحت عنوان تابع انتقال^۳ (یا فعال سازی) می شود که برای تبدیل و تولید خروجی واقعی a با f نمایش داده می شود. به عبارت دیگر:

$$a = f(wp + b)$$

¹ Processing Elements (PEs)

² Net input function

³ Transfer



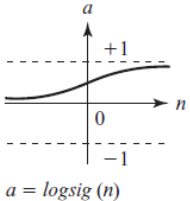
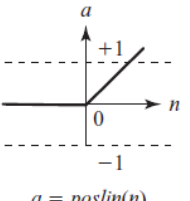
شکل ۶.۵ نمایش یک نورون مصنوعی تک ورودی

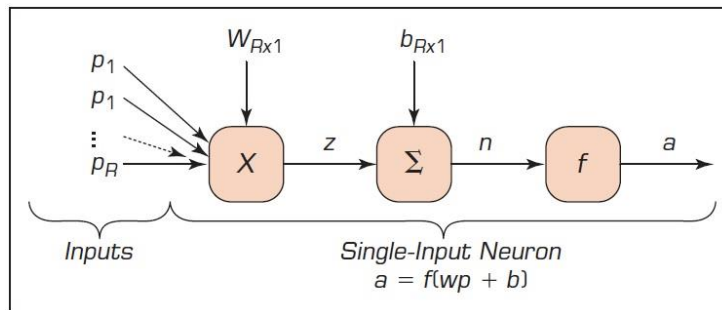
یک مثال عددی: اگر $w = 2$, $p = 3$ و $b = -1$ آنگاه $a = f(2 * 3 - 1) = f(5)$

انواع مختلف توابع انتقال در طراحی شبکه‌های عصبی بکار می‌روند. جدول ۶.۱ برخی از متداول‌ترین توابع انتقال و عملیات مرتبط آنها را نشان می‌دهد. توجه داشته باشید که در عمل، انتخاب توابع انتقال مناسب برای یک شبکه نیازمند دانش عمیق و گسترده‌ای از شبکه‌های عصبی است - مشخصه‌های داده‌ها و همچنین هدف خاصی که این شبکه برای آن ایجاد شده است.

اگر در مثال قبل ما یک تابع انتقال هارد لیمیت داشته باشیم، خروجی حقیقی a به صورت $a = \text{hardlim}(5) = 1$ محاسبه می‌شود. دستورالعمل‌هایی برای انتخاب تابع انتقال مناسب برای هر مجموعه از نورون‌ها در یک شبکه وجود دارد. این اصول راهنما به ویژه برای نورون‌هایی که در لایه‌ی خروجی شبکه واقع شده‌اند کاربرد دارد. به عنوان مثال، اگر ماهیت خروجی یک مدل از نوع دودویی باشد، توصیه می‌شود که از توابع انتقال سیگموئید در لایه خروجی استفاده شود که یک مقدار خروجی بین ۰ و ۱ را نتیجه می‌دهد که بیانگر احتمال شرطی $y = 1$ به شرط x یا $P(y = 1|x)$ در کتابچه‌های زیادی از شبکه‌های عصبی، اصول راهنمای مبسوطی در لایه‌های مختلف در یک شبکه عصبی ارائه شده است که بیان می‌کنند که بهترین تجارب باید از تجربه حاصل شوند.

جدول ۶.۱ توابع انتقال (فعال‌سازی) متداول در شبکه‌های عصبی		
عملیات	فرم	تابع انتقال
$a = +1 \text{ if } n > 0$ $a = 0 \text{ if } n < 0$	<p style="text-align: center;">$a = \text{hardlim}(n)$</p>	هارد لیمیت
$a = n$	<p style="text-align: center;">$a = \text{purelin}(n)$</p>	خطی

$a = \frac{1}{1 + e^{-n}}$	 <p style="text-align: center;">$a = \text{logsig}(n)$</p>	لاگ-سیگموئید
$a = +1 \text{ if } n > 0$ $a = 0 \text{ if } n < 0$	 <p style="text-align: center;">$a = \text{poslin}(n)$</p>	خطی مثبت



شکل ۶.۶ یک نورون چند ورودی با ورودی‌های انفرادی R

معمولاً، یک نورون بیش از یک ورودی دارد. در این صورت، هر ورودی p_i را می‌توان به صورت یک عنصر بردار ورودی \mathbf{p} نمایش داد. هر کدام از مقادیر ورودی‌ها وزن تنظیم‌پذیر w_i بردار وزنی \mathbf{W} را دارند. شکل ۶.۶ یک نورون چند ورودی با ورودی‌های انفرادی R را نمایش می‌دهد. برای این نورون، ورودی خالص n را می‌توان به شکل زیر نمایش داد:

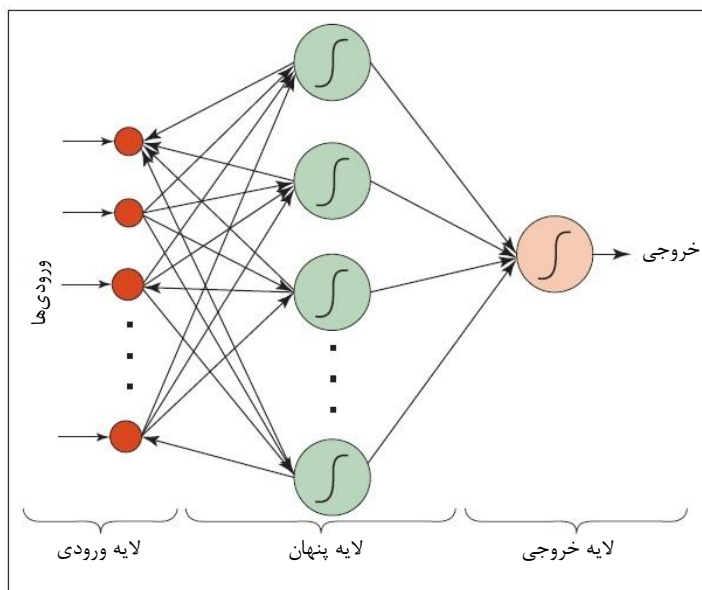
$$n = w_{1,1}p_1 + w_{1,2}p_2 + w_{1,3}p_3 + \dots + w_{1,R}p_R + b$$

با در نظر گرفتن بردار ورودی \mathbf{p} به صورت یک بردار $R * I$ و بردار وزنی \mathbf{W} به صورت یک بردار $I * R$ آنگاه می‌توان n را به صورت ماتریسی زیر مجدداً نوشت:

$$n = \mathbf{Wp} + b$$

که \mathbf{Wp} یک مقیاس‌گر است (یعنی بردار $1 * 1$).

علاوه بر این، هر شبکه عصبی معمولاً از چندین نورون متصل بهم تشکیل شده و به صورت لایه‌های متوالی قرار گرفته‌اند که بدین ترتیب خروجی‌های یک لایه به صورت ورودی‌های لایه بعدی عمل می‌کنند. شکل ۶.۷ یک شبکه عصبی معمول با چهار نورون در لایه ورودی (یعنی اول)، چهار نورون در لایه پنهان (میانی) و یک نورون در لایه خروجی (آخر) را نشان می‌دهد.



شکل ۶.۷ یک شبکه عصبی با سه لایه و هشت نورون

هر کدام از نورون‌ها، وزن، تابع وزن‌دهی، بایاس و تابع انتقال خاص خودشان را داشته و ورودی(های) خودشان را پردازش می‌کنند. اگرچه ورودی‌ها، توابع وزن‌دهی و توابع انتقال در یک شبکه ثابت هستند ولی مقادیر وزن‌ها و بایاس‌ها قابل تغییر می‌باشند. فرایند تعدیل و تغییر وزن‌ها و بایاس‌ها در یک شبکه عصبی را عمدتاً آموزش می‌نامند. در واقع، در عمل یک شبکه عصبی را نمی‌توان به طور مؤثری برای مسائل پیش‌بینی بکار برد مگر اینکه با تعداد مثال‌های کافی و خروجی‌های واقعی مشخص به خوبی آموزش دیده باشند. هدف از فرایند آموزش این است که اوزان و بایاس‌های شبکه را به گونه‌ای تنظیم کنیم که خروجی شبکه برای هر مجموعه از ورودی‌ها (یعنی هر نمونه) به میزان کافی نزدیک به مقدار هدف متناظر با آن باشد.

مورد کاربردی ۶.۲ یک مثال از کاربرد تحلیل‌های پیشرفته در شرکت‌های ساخت و تولید بازی برای درک بهتر مشتریان و مشارکت دادن آنها ارائه می‌کند.

شرکت‌های بازی‌سازی از تحلیل‌های پیشرفته برای امتیازدهی بازیکنان استفاده می‌کنند.

مورد کاربردی ۶.۲

علاقمندان به بازی‌های ویدئویی از نژاد خاصی هستند. مطمئناً آنها زمان بسیار زیادی را صرف بازی کردن می‌کنند ولی به ایجاد شبکه‌های اجتماعی نیز می‌پردازند. علاقمندان به بازی‌های ویدئویی همانند ورزشکاران حرفه‌ای به رقابت با یکدیگر می‌پردازند.

آنها به صورت آنلاین با سایر بازیکنها رقابت می‌کنند. آنهایی که به جایگاه اول یا حتی دوم یا سوم دست پیدا می‌کنند حق چانه‌زنی دارند. همانند ورزشکارانی که زمان زیادی را صرف آموزش می‌کنند، علاقمندان به بازی‌های ویدئویی نیز چندین ساعت در روز را به تمرین اختصاص می‌دهند. علاوه بر این، با افزایش پیچیدگی بازی‌ها، این افراد نیز مهارت‌های منحصری فردی را برای غلبه بر رقبایشان کسب می‌کنند.

سطح جدیدی از بازی‌ها

تاریخچه‌ی بازی‌های ویدئویی به روزهای PAC-MAN باز می‌گردد. دسترسی گسترده به اینترنت موجب شد تا بازی‌های ویدئویی از طریق دستگاه‌های الکترونیکی متنوع مثل رایانه‌های شخصی و دستگاه‌های موبایل به راحتی در دسترس مردم قرار بگیرد و بدین ترتیب بر محبوبیت بازی‌های ویدئویی افزوده شود. دنیای بازی‌های رایانه‌ای در حال حاضر یک تجارت قوی و سودآور محسوب می‌شود.

بر اساس گزارش بازار جهانی بازی‌های نیوزو از آوریل ۲۰۱۷، وضعیت بازار بازی‌های جهانی به این شکل بوده است:

- درآمد ۱۰۹ میلیارد دلاری
- افزایش ۷.۸ درصدی نسبت به سال گذشته
- ۲.۲ میلیارد نفر بازیکن در سرتاسر جهان
- بازار ۴۲ درصدی بازی‌های موبایل

شرکت‌های بازی‌های ویدئویی می‌توانند از این محیط بهره‌برداری کرده و اطلاعات ارزشمندی در مورد مشتریان و به‌ویژه رفتارها و انگیزه‌های آنها بدست آورند. این داده‌های مشتریان، شرکت‌ها را قادر می‌سازد تا به تجربه بهتری از بازی‌ها دست پیدا کرده و بازیکن‌ها را بیش از پیش مشارکت دهند.

صنعت بازی از دیرباز با ارائه‌ی مصورسازی‌های گرافیکی خیره‌کننده برای مشتریان جذبیت داشته است. با پیشرفت فناوری‌ها، بر جلوه‌های خاص گرافیک‌ها نیز افزوده شده است. شرکت‌ها همچنان از روش‌های بسیار خلاقانه برای تولید بازی‌ها استفاده می‌کنند تا بتوانند مشتریان را جذب خود کنند تا زمان بیشتری را صرف بازی و وابستگی به خود نمایند. آنچه شرکت‌های بازی‌های ویدئویی هنوز انجام نداده‌اند این است که از فناوری به طور کامل برای درک عوامل بافتاری پیشران در درگیری پایدار با برندها بهره‌برداری کنند.

بازیکنان را بشناسید

در دنیای امروز بازی‌های ویدئویی، خلق یک محصول هیجان‌انگیز دیگر به تنهایی کافی نیست. بازی‌ها باید از لحاظ دیداری و شنیداری در عصری که مردم به دنبال جلوه‌های صوتی و گرافیکی خارق‌العاده و چشمگیر هستند برایشان جذابیت داشته باشند. بازی‌ها همچنین باید برای دسترسی به گروه بازیکنان هدف به طور مناسبی بازاریابی شده باشند. همچنین فرصت‌هایی برای پول درآوردن از کاراکترهای بازی‌ها به شکل کالاهای تجاری یا حقوق ساخت فیلم‌ها وجود دارد. تولید یک بازی موفق نیازمند همکاری و به اشتراک‌گذاری اطلاعات توسط برنامه‌نویسان، طراحان، سناریونویسان، موسیقی‌دانان و بازاریاب‌ها می‌باشد. نقش بازیکن‌ها و داده‌های آنها اینجا مطرح می‌شود.

به عنوان مثال، حجم شبکه‌ی یک بازیکن - تعداد و انواع افرادی که بازی می‌کنند - معمولاً با صرف زمان و پول بیشتر برای بازی همبستگی دارد. هرچه بازیکن‌ها ارتباطات بیشتری داشته باشند، احتمال اینکه بازی‌های بیشتری را با افراد بیشتری انجام دهند بیشتر خواهد بود که به دلیل لذت بردن از این تجربه است.

این داده‌ها همچنین به شرکت‌ها کمک می‌کنند تا انواع بازی‌هایی که هر فرد تمایل دارد بازی کند را بهتر درک کنند. این بینش‌ها شرکت‌ها را قادر می‌سازد تا بازی‌های بیشتری را در کنار ژانرهایی که ممکن است تأثیر مثبتی بر رضایت و درگیرسازی مشتری داشته باشند توصیه کند. شرکت‌ها همچنین می‌توانند از این داده‌ها در کمپین‌های بازاریابی و برای هدف قراردادن بازیکن‌های جدید یا ترغیب بازیکن‌های فعلی برای ارتقاء حق عضویت‌هایشان استفاده کنند.

پول در آوردن از رفتارهای بازیکنان

فیلترینگ مشارکتی یک کارکرد تحلیلی پیشرفته است که به صورت خودکار و از طریق گردآوری اولویت‌ها یا سلاقی جمع‌کشی از بازیکن‌ها به پیش‌بینی علایق یک بازیکن می‌پردازد. در این کارکرد فرض بر این است که اگر کاربر A در خصوص مسئله‌ای با کاربر B هم عقیده باشد آنگاه کاربر A در مقایسه با یک کاربر تصادفی در خصوص مسئله‌ای متفاوت دیگری به احتمال زیاد با کاربر B هم عقیده خواهد بود. این نشان می‌دهد که پیش‌بینی‌ها بر اساس داده‌های بدست آمده از تعداد زیادی بازیکن دیگر به یک بازیکن مشخص بستگی دارد.

خرده‌فروشی‌ها اغلب از سیستم‌های فیلترینگ برای توصیه‌ی محصولات خود استفاده می‌کنند. تجزیه و تحلیل‌ها می‌توانند بر اساس آنچه سایر مشتریانی که خریده‌های مشابهی را انجام داده‌اند تعیین کنند که مشتری کدام محصول را ترجیح خواهد داد. مثال‌های بسیاری در صنایع مختلف از جمله مراقبت سلامت، مالی، تولید و مخابرات وجود دارد.

کارکرد تحلیلی فیلترینگ مشارکتی مزایای متعددی برای شرکت‌های بازی‌های ویدئویی آنلاین ارائه می‌کند:

- **بازاریاب‌ها می‌توانند کمپین‌های مؤثرتری را راه‌اندازی کنند.** ارتباطات بین بازیکن‌ها طبیعتاً منجر به شکل‌گیری خوشه‌ها می‌شود. بازاریاب‌ها می‌توانند مشخصه‌های مشترک بازیکن‌ها را تعیین کرده و از آن بینش‌ها در طرح‌های تبلیغاتی خود استفاده کنند. برعکس، آنها می‌توانند بازیکنانی که به یک خوشه تعلق ندارند را ایزوله کرده و ویژگی منحصربفردی که موجب شده است تا چنین رفتارهایی داشته باشد را مشخص کنند.
- **شرکت‌ها می‌توانند نگهداشت بازیکن را افزایش دهند.** یک عضویت قوی در یک جامعه از بازیکن‌ها می‌تواند احتمال ریزش یا ترک آن عضو را کاهش دهد. هرچه مشوق‌های قویتری برای تعلق بازیکن‌ها به یک گروه در نظر گرفته شود، آنها تمایل بیشتری برای مشارکت در رقابت‌ها و مسابقات خواهند داشت. بدین ترتیب نام‌نویسی مشتریان برای بازی‌ها بیشتر خواهد شد.
- **بینش داده‌ها منجر به افزایش رضایت مشتری می‌شود.** خوشه‌ها بیانگر تمایل نسبت به انواع خاصی از بازی‌ها است که رفتارها و سلاقی متمایز بازیکن‌ها را نشان می‌دهد. شرکت‌ها می‌توانند تجارب منحصربفرد و خاص هر بازیکن را خلق کنند. ترغیب افراد بیشتر برای بازی کردن و افزایش زمان بازی موجب افزایش رضایت بازیکن خواهد شد.

شرکت‌ها پس از درک اینکه مشتریان چه بازی‌هایی را دوست دارند و پس از کشف ارتباطات آنها با سایر بازیکن‌ها می‌توانند مشوق‌های درست و مناسبی را برای حفظ بازیکنان ایجاد کنند. بدین ترتیب حفظ یک پایگاه مشتری و جریان‌های درآمدی پایدار در اینجا تضمین می‌شود.

تقویت وفاداری و درآمد

صرفنظر از ژانر، هر بازی ویدئویی بازیکنانی مشتاقی را دارد که به دنبال رقابت با یکدیگر هستند. هیجان فتح کردن است که موجب مشارکت و درگیر شدن آنها می‌شود. در طول زمان، شبکه‌های متمایزی از بازیکنان شکل می‌گیرند که هر شرکت‌کننده ارتباطات اجتماعی خاص خود را تشکیل می‌دهد و بدین ترتیب تعاملات بازی‌ها بسیار بیشتر و شدیدتر می‌شود. صنعت بازی‌سازی در حال حاضر از مصورسازی و تجزیه و تحلیل داده‌ها برای درک بهتر رفتارهای مشتریان و کشف انگیزه‌های آنان بهره می‌برد. فقط نگاه به بخش‌بندی مشتریان، دیگر کافی نیست. شرکت‌ها اکنون برای درک اولویت‌های مشتریان از قبیل بازی‌های محبوب، سطح دشواری مورد انتظار یا ژانر بازی‌ها به ریزبخش‌هایی از مشتریان نگاه می‌کنند که فراتر از آمارهای جمعیت‌شناختی از قبیل سن یا موقعیت جغرافیایی نگاه می‌کنند. شرکت‌ها با دستیابی به بینش‌های تحلیلی نسبت به استراتژی‌ها و رفتارهای بازیکنان می‌توانند تجارب متناسب با آن رفتارها را در بازی‌سازی خلق کنند. شرکت‌های تولید بازی‌های ویدئویی با درگیر کردن بازیکنان با بازی‌ها و ویژگی‌های مطلوبشان می‌توانند به افزایش سود و توسعه جریان‌های درآمدی جدید از طریق اقدامات بازاریابی دست پیدا کنند. برای اینکه ببینید شرکت‌ها چگونه از تجزیه و تحلیل برای رمزگشایی ارتباطات بازیکنان که پیش‌ران و محرک اصلی رفتارهای کاربر است استفاده می‌کنند ویدئوی کوتاه [https://www.teradata.com/Resources/Videos/Art-of-Analytics-](https://www.teradata.com/Resources/Videos/Art-of-Analytics) (**The-Sword**) را تماشا کنید.

Source: Teradata Case Study. <https://www.teradata.com/Resources/Case-Studies/Gaming-Companies-Use-Data-Analytics> (accessed August 2018).

سوالاتی برای مورد کاربردی ۶.۲

۴. چالش‌های اصلی شرکت‌های بازی‌سازی چه هستند؟
۵. تجزیه و تحلیل چگونه می‌تواند به حفظ مزیت رقابتی شرکت‌های بازی‌سازی کمک کند؟
۶. شرکت‌های بازی‌سازی از چه نوع داده‌ها می‌توانند برای تجزیه و تحلیل‌ها استفاده کنند؟

بینش فناوری ۶.۱ به اختصار مؤلفه‌ها (یا عناصر) مشترک یک شبکه عصبی مصنوعی را به همراه ارتباطات کارکردی آنها نشان می‌دهد.

بینش فناوری ۶.۱ عناصر یک شبکه‌های عصبی مصنوعی

یک شبکه عصبی از عناصر پردازشی تشکیل شده است که به روش‌های مختلفی ساختار شبکه را شکل می‌دهند. واحد پردازش پایه در یک شبکه عصبی را نورون می‌نامند. سپس چندین نورون در کنار هم قرار گرفته و شبکه‌ای از نورون‌ها را شکل می‌دهند. نورون‌ها

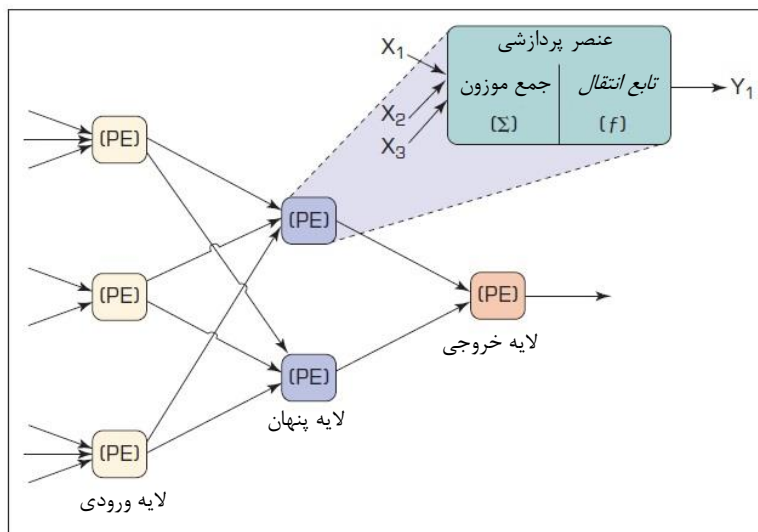
را می‌توان به روش‌های مختلفی سازماندهی کرد؛ این الگوهای شبکه‌ای مختلف را توپولوژی یا معماری شبکه می‌نامند. یکی از متداول‌ترین روش‌ها به نام پرسپترون چندلایه‌ی پیش‌بازخور^۱ به تمامی نورون‌ها اجازه می‌دهد تا ورودی در یک لایه را به ورودی لایه بعدی متصل کنند ولی امکان هیچگونه پیوند بازخور در آن وجود ندارد (هایکین ۲۰۰۹).

عنصر پردازشی در یک شبکه عصبی مصنوعی، یک نورون مصنوعی است. هر نورون، ورودی‌ها را دریافت کرده، آنها را پردازش کرده و یک خروجی را تحویل می‌دهد (شکل ۶.۵). ورودی می‌تواند یک داده‌های خام یا خروجی سایر عناصر پردازشی باشد. خروجی می‌تواند یک نتیجه‌ی نهایی (مثلاً ۱ به معنی بله و ۰ به معنی خیر) بوده یا ورودی برای سایر نورون‌ها باشد.

ساختار شبکه

هر شبکه عصبی مصنوعی از مجموعه‌ای از نورون‌ها که به صورت لایه‌ای گروه‌بندی شده‌اند تشکیل شده است. یک ساختار معمول در شکل ۶.۸ نشان داده شده است. توجه داشته باشید که سه لایه‌ی ورودی، واسطه (یا لایه‌ی پنهان) و خروجی وجود دارد. لایه‌ی پنهان، لایه‌ای از نورون‌ها است که ورودی را از لایه قبلی گرفته و آن ورودی‌ها را به خروجی‌های لایه‌های بعدی تبدیل می‌کند. بین لایه‌های ورودی و خروجی می‌توان چندین لایه پنهان قرار داد اگرچه معمولاً از یک لایه پنهان استفاده می‌شود. در این حالت، لایه‌ی پنهان صرفاً ورودی‌ها را به یک ترکیب غیرخطی تبدیل کرده و ورودی‌های تبدیل شده را به لایه خروجی انتقال می‌دهد. متداول‌ترین تفسیر از لایه پنهان این است که آن را یک مکانیزم استخراج-ویژگی^۲ می‌دانند بدین معنی که لایه‌ی پنهان، ورودی‌های اصلی در مسئله را به ترکیب سطح بالاتری از این ورودی‌ها تبدیل می‌کند.

در شبکه عصبی مصنوعی، زمانیکه اطلاعات پردازش می‌شوند بسیاری از عناصر پردازشی، محاسبات خود را به طور همزمان انجام می‌دهند. این پردازش موازی مشابه نحوه‌ی کارکرد مغز انسان است.



¹ Feedforward-multilayered perceptron

² Feature-extraction mechanism

شکل ۶.۸ شبکه عصبی با یک لایه پنهان

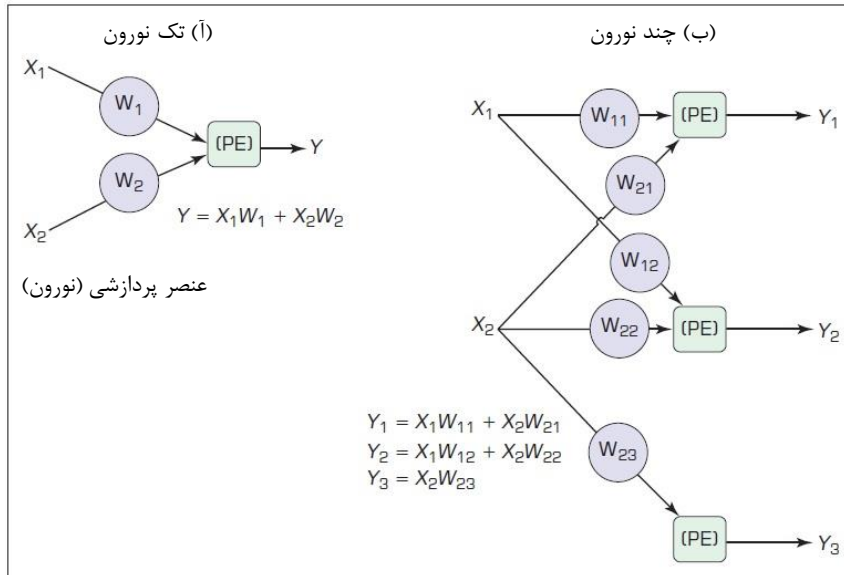
هر ورودی متناظر با یک صفت یا ویژگی است. به عنوان مثال، اگر مسئله این باشد که در مورد تأیید یا عدم تأیید یک وام تصمیم‌گیری کنیم، صفاتی که می‌توانیم تعریف کنیم شامل میزان درآمد، سن و وضعیت مالکیت منزل متقاضی است. مقدار عددی یا نمایش عددی مقادیر غیر عددی یک صفت، به عنوان ورودی شبکه در نظر گرفته می‌شود. چندین نوع داده از قبیل متن، تصویر و صدا را می‌توان به عنوان ورودی در نظر گرفت. ممکن است برای تبدیل داده‌های سمبولیک/غیر عددی یا عددی/مقیاس به ورودی‌های معنادار نیاز به پیش‌پردازش داده‌ها باشد.

خروجی یک شبکه حاوی راه‌حل مسئله است. به عنوان مثال، در خصوص متقاضی یک وام، خروجی می‌تواند "بله" یا "خیر" باشد. شبکه عصبی مصنوعی مقادیر عددی را به خروجی تخصیص می‌دهد که ممکن است نیاز باشد که با استفاده از یک مقدار آستانه آنها را به خروجی‌های رسته‌ای تبدیل کرد. مثلاً ۱ برای "بله" و ۰ برای "خیر".

اوزان ارتباطات، عناصر کلیدی یک شبکه عصبی مصنوعی هستند. آنها بیانگر قوت نسبی (یا مقدار ریاضی) داده‌های ورودی یا ارتباطات بسیار زیادی هستند که داده‌ها را از لایه‌ای به لایه‌ی دیگر انتقال می‌دهند. به عبارت دیگر، اوزان بیانگر اهمیت نسبی هر ورودی نسبت به یک عنصر پردازشی و نهایتاً نسبت به خروجی است. اوزان از این لحاظ حیاتی هستند که الگوهای اطلاعاتی آموخته شده را در خود ذخیره می‌کنند. یک شبکه از طریق تغییرات و تنظیمات مکرر اوزان است یاد می‌گیرد.

تابع مجموع^۱، جمع اوزان تمامی عناصر ورودی که وارد هر عنصر پردازشی می‌شوند را محاسبه می‌کند. یک تابع مجموع، مقدار هر ورودی را ضرب در وزن آن کرده و جمع کل مقادیر موزون را محاسبه می‌کند. فرمول برای n ورودی (که با X نمایش داده شده است) در یک عنصر پردازشی در شکل 6.9a و برای چندین عنصر پردازشی در شکل 6.9b نشان داده شده است.

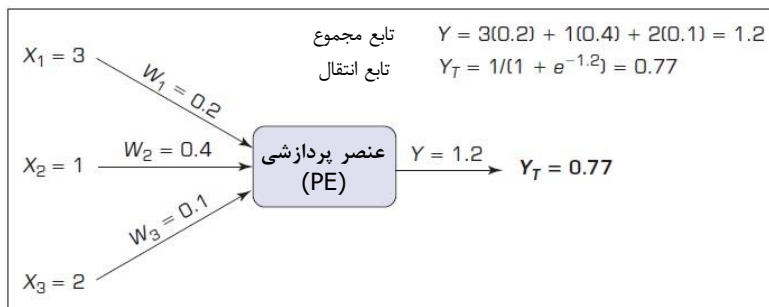
¹ Summation Function



شکل ۶.۹ تابع مجموع برای (آ) تک نورونی/عنصر پردازشی و (ب) چند نورونی/عنصر پردازشی

تابع انتقال

تابع مجموع، شبیه‌سازی داخلی یا سطح فعال‌سازی نورون را محاسبه می‌کند. بر اساس این سطح، نورون ممکن است یک خروجی تولید کرده یا نکند. ارتباط بین سطح فعال‌سازی داخلی و خروجی ممکن است به صورت خطی یا غیرخطی باشد. این ارتباط توسط یکی از چندین نوع تابع تبدیل (انتقال) بیان می‌شود. اینکه کدام تابع فعال‌سازی را انتخاب کنیم بر عملکرد شبکه تأثیر می‌گذارد. شکل ۶.۱۰ محاسبه‌ی یک تابع فعال‌سازی سیگموئیدی ساده را نشان می‌دهد. سطح خروجی به نحوی اصلاح می‌شود که در یک دامنه‌ی قابل قبول از مقادیر (معمولاً بین ۰ و ۱) قرار بگیرد. این تبدیل پیش از آنکه خروجی به سطح بعدی برسد انجام می‌شود. بدون چنین تغییری، مقدار خروجی ممکن است بسیار بزرگ شود مخصوصاً زمانی که چندین لایه نورون وجود داشته باشد. برخی اوقات به جای یک تابع تبدیل از یک مقدار آستانه استفاده می‌شود. یک مقدار آستانه در واقع یک مقدار بازدارنده برای خروجی یک نورون است تا نورون‌ها سطح بعدی را تحریک کند. به عنوان مثال، هر مقدار ۰.۵ یا کمتر از آن معادل ۰ و هر مقدار ۰.۵ و بالاتر از آن معادل ۱ خواهد بود. تبدیل می‌تواند در خروجی هر عنصر پردازشی رخ داده یا اینکه فقط در گره‌های خروجی پایانی انجام شود.



شکل ۶.۱۰ مثالی از تابع انتقال شبکه عصبی مصنوعی

مقامات وایلدترک^۱ می‌گویند "کسانی هستند که می‌خواهند حیواناتی مثل شیرها و یوزپلنگ‌ها را بکشند. می‌خواهیم به آنها بگوییم که تعداد زیادی از این حیوانات باقی نمانده است." هرچه بیشتر بتوانیم رفتار آنها را مطالعه کنیم، بیشتر می‌توانیم به آنها کمک کنیم تا از این حیوانات محافظت کنند و تنوع زیستی زمین که برای همه ما مهم است را حفظ کنند. ردپای این حیوانات می‌تواند اطلاعات ارزشمندی را در اختیار ما قرار دهند. به کجا می‌روند؟ چه تعداد از آنها باقیمانده است؟ با رصد کردن ردپای گونه‌های در معرض انقراض مثل یوزپلنگ‌ها می‌توانیم چیزهای بسیاری بیاموزیم.

وایلدترک که یک سازمان غیرانتفاعی است در سال ۲۰۰۴ توسط زوئی جوئل^۲ یک دامپزشک و اسکای علی‌باهی^۳ بیولوژیست حیات وحش که برای سالیان متمادی در آفریقا به رصد کرگدن‌های سیاه و سفید می‌پرداخته است بنیان نهاده شد. آنها زمانیکه در اوایل دهه ۹۰ در زیمبابوه بودند به گردآوری و ارائه‌ی داده‌هایی پرداختند که نشان می‌دهد تکنیک‌هایی که برای رصد کرگدن‌های سیاه بکار می‌رفت تأثیر منفی بر باروری کرگدن‌های ماده داشته است و شروع به توسعه‌ی یک تکنیک شناسایی ردپا کردند. پژوهشگران از سرتاسر جهان که نیازمند روشی کم هزینه و غیرتهاجمی برای رصد حیات وحش بودند از این رویکرد استقبال کردند.

هوش مصنوعی می‌تواند به ما کمک کند تا برخی از مهارت‌هایی که ردیاب‌های بومی استفاده می‌کردند را مجدداً ابداع کنیم. پژوهشگران وایلدترک در حال اکتشاف ارزش هوش مصنوعی در این موضوع هستند. آنها فکر می‌کنند که راه‌حل‌های هوش مصنوعی برای ارتقاء تلاش‌های انسانی و نه جایگزین شدن با آنها طراحی شده‌اند. یک رایانه به کمک یادگیری عمیق و با استفاده از داده‌های کافی را به نحوی می‌توان آموزش داد تا وظایف انسانی همانند شناسایی تصاویر ردپاها و تشخیص الگوها را به شکلی که ردیاب‌های بومی انجام می‌دادند انجام دهند- ولی این توانایی را می‌توان در مقیاس بسیار بزرگتر و سریع‌تر انجام داد. تجزیه و تحلیل در واقع از کل موضوع حمایت کرده و بینش‌هایی را در مورد جمعیت گونه‌ها در اختیار ما قرار می‌دهد که تا پیش از آن نداشته‌ایم.

تکنیک تشخیص ردپای وایلدترک، ابزاری برای رصد غیرتهاجمی گونه‌های در معرض انقراض از طریق تصاویر دیجیتالی ردپاها است.

اندازه‌های بدست آمده از این تصاویر توسط مدل‌های ریاضی سفارشی تجزیه و تحلیل می‌شوند که به شناسایی گونه‌ها، جنسیت و گروه سنی آنها کمک می‌کند. هوش مصنوعی می‌تواند توانایی وفق و سازگاری را از طریق الگوریتم‌های یادگیری پیش‌رونده افزوده و ماجرای کامل‌تری را تعریف کند.

گام بعدی برای تعریف کل داستان در آینده، دستیابی به داده‌های جمع‌سپاری است. مردم عادی طبیعتاً نمی‌توانند یک کرگدن را با دارت بزنند ولی می‌توانند از یک ردپا عکس بگیرند. داده‌های وایلدترک از همه جا می‌آید. در اینجاست که هوش مصنوعی وارد عمل می‌شود. هوش مصنوعی از طریق داده‌ها، انجام وظایف رایانه‌ای مکرر و با حجم بالا و بدون خستگی می‌تواند فرایند یادگیری تکرارپذیر را به صورت خودکار درآورد.

علی‌باهی می‌گوید: "چالش ما این است که چگونه از هوش مصنوعی برای خلق محیطی استفاده کنیم که هم برای ما و هم کلیه‌ی گونه‌های موجودات در این جهان جا داشته باشد."

¹ WildTrack

² Zoe Jewell

³ Sky Alibhai

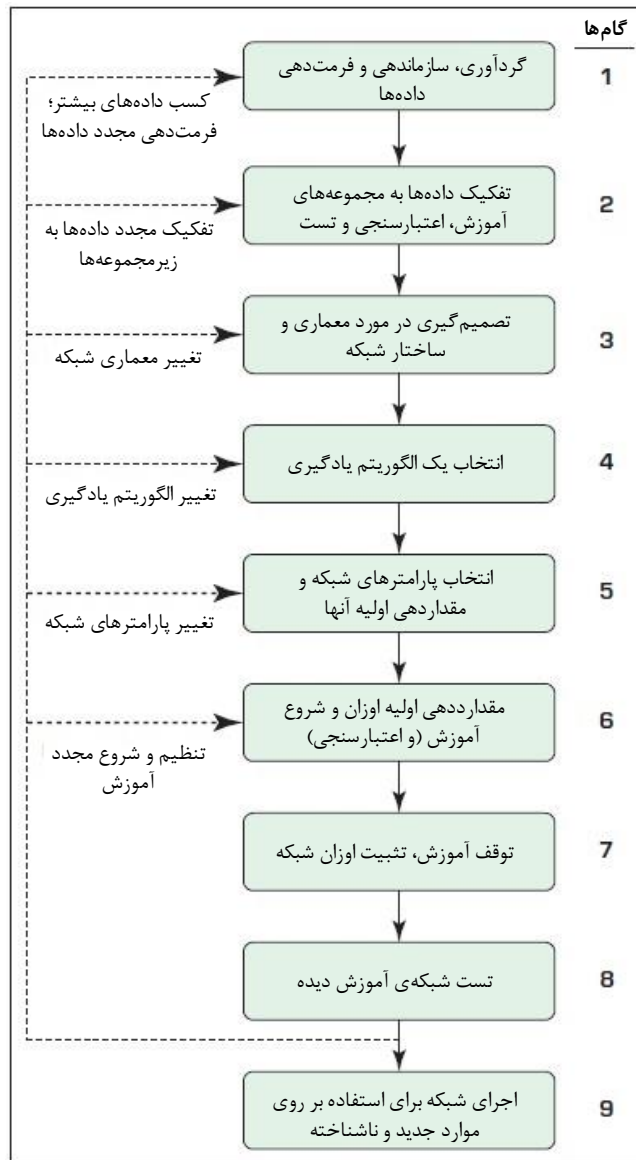
سوالاتی برای مورد کاربردی ۶.۳

۱. وایلدترک چیست و چه کاری انجام می‌دهد؟
۲. تحلیل‌های پیشرفته چگونه می‌توانند به وایلدترک کمک کنند؟
۳. یادگیری عمیق چه نقشی در این مورد کاربردی ایفا می‌کند؟

سوالاتی برای مرور بخش ۶.۳

۵. یک نورون مصنوعی (یعنی عنصر پردازشی) چگونه عمل می‌کند؟
۶. متداول‌ترین توابع فعال‌سازی شبکه‌های عصبی مصنوعی را نام برده و به اختصار توضیح دهید.
۷. منظور از ام‌ال‌پی چیست و چگونه کار می‌کند؟
۸. کارکرد اوزان در شبکه عصبی مصنوعی را توضیح دهید.
۹. توابع مجموع و فعال‌سازی در معماری شبکه عصبی مصنوعی از نوع ام‌ال‌پی را توضیح دهید.

اگرچه فرایند توسعه‌ی شبکه عصبی مصنوعی مشابه متدولوژی‌های طراحی ساختارمند سیستم‌های اطلاعات رایانه‌ای سنتی است ولی برخی فازهای آنها منحصربفرد بوده یا جنبه‌های منحصربفردی دارند. در فرایند توضیح داده شده در بالا، فرض می‌کنیم که گام‌های اولیه‌ی توسعه‌ی سیستم از قبیل تعیین نیازمندی‌های اطلاعاتی، انجام تحلیل امکان‌سنجی و جلب حمایت مدیریت ارشد برای پروژه با موفقیت به انجام رسیده باشد. این گام‌ها برای هر سیستم اطلاعاتی عمومیت دارند. همانطور که در شکل ۶.۱۱ نشان داده شده است، فرایند توسعه‌ی یک کاربرد شبکه عصبی مصنوعی شامل نه گام است. در گام ۱، داده‌های مورد نیاز برای آموزش و تست شبکه گردآوری می‌شوند. ملاحظات مهم در اینجا این هستند که این مسئله‌ی خاص یک راه‌حل شبکه عصبی دارد و داده‌های کافی برای آن وجود داشته و قابل دسترس است. در گام ۲، داده‌های آموزش باید شناسایی شده و یک طرح برای تست عملکرد شبکه تهیه شود.



شکل ۶.۱۱ توسعه‌ی فرایند یک مدل شبکه عصبی مصنوعی

در گام‌های ۳ و ۴ یک معماری شبکه و یک روش یادگیری انتخاب می‌شود. نوع شبکه‌ای که می‌خواهیم ایجاد کنیم به در دسترس بودن ابزارهای توسعه خاص یا قابلیت‌های پرسنل بستگی دارد. همچنین، برخی از انواع مسائل از نرخ موفقیت بالایی با پیکربندی‌های خاص برخوردار بوده‌اند (مثلاً شبکه‌های عصبی چندلایه پیش‌بازخورد برای پیش‌بینی ورشکستگی (آلتمن^۱، ۱۹۶۸)،

¹ Altman

ویلسون و شاردا^۱ (۱۹۹۴) و السون، دیلن و منگ^۲ (۲۰۱۲)). ملاحظات مهم در اینجا تعداد دقیق نورون‌ها و لایه‌ها است. برخی از بسته‌های نرم‌افزاری از الگوریتم‌های ژنریک (عمومی) برای انتخاب طراحی شبکه استفاده می‌کنند.

پارامترهای متعددی برای تنظیم شبکه برای دستیابی به سطح عملکرد یادگیری مطلوب وجود دارد. بخشی از فرایند در گام ۵ مقداردهی اولیه به اوزان و پارامترهای شبکه و اصلاح پارامترها پس از دریافت بازخورد عملکرد آموزش است. اغلب، مقادیر اولیه در تعیین کارایی و طول آموزش بسیار اهمیت دارند. برخی روش‌ها برای ارتقاء سطح عملکرد، مقدار پارامترها را در حین آموزش تغییر می‌دهند. گام ۶ داده‌های کاربردی را به نوع و فرمت مورد نیاز شبکه عصبی تبدیل می‌کند. این ممکن است نیازمند نوشتن نرم‌افزاری برای پیش‌پردازش داده‌ها یا انجام مستقیم این عملیات در یک بسته نرم‌افزاری شبکه عصبی مصنوعی باشد. تکنیک‌ها و فرایندهای انبارش و دستکاری داده‌ها باید به شکلی طراحی شوند که در صورت نیاز، آموزش مجدد شبکه عصبی به شکلی راحت و مؤثر امکان‌پذیر باشد. نمایش و مرتب‌سازی داده‌های کاربردی اغلب بر کارایی و دقت نتایج تأثیر می‌گذارد. در گام‌های ۷ و ۸، آموزش و تست به ترتیب با ارائه‌ی داده‌های ورودی و خروجی معلوم به شبکه انجام می‌شود. شبکه به محاسبه‌ی خروجی‌ها پرداخته و اوزان را تا جایی تعدیل می‌کند که خروجی‌های محاسبه شده در محدوده‌ی تولرانس قابل قبول قرار بگیرند. خروجی‌های مطلوب و ارتباطات آنها با داده‌های ورودی از داده‌های تاریخی بدست می‌آیند (یعنی بخشی از داده‌های گردآوری شده در گام ۱).

در گام ۹، مجموعه‌ی پایداری از اوزان به دست می‌آید. آنگاه شبکه می‌تواند خروجی‌های مطلوب را مجدداً تولید کند. شبکه اکنون به عنوان یک سیستم مستقل یا بخشی از یک سیستم نرم‌افزاری دیگر آماده‌ی استفاده است که داده‌های ورودی جدید به آن داده شده و خروجی آن به صورت تصمیم توصیه شده نمایش داده می‌شود.

در یادگیری نظارت شده، فرایند یادگیری به صورت القایی است بدین معنی که اوزان ارتباطات از موارد موجود گرفته می‌شوند. فرایند معمول یادگیری شامل سه وظیفه است (شکل ۶.۱۲):

۱. محاسبه‌ی خروجی‌های موقت.

۲. مقایسه‌ی خروجی‌ها با اهداف مطلوب.

۳. تنظیم اوزان و تکرار فرایند.

همانند هر تکنیک یادگیری ماشین نظارت شده‌ی دیگر، آموزش شبکه عصبی معمولاً از طریق تعریف یک تابع عملکرد (F) (تابع هزینه^۳ یا تابع زیان^۴) و بهینه‌سازی (کمینه‌سازی) آن تابع از طریق تغییر پارامترهای مدل انجام می‌شود. معمولاً، تابع عملکرد چیزی به جز یک سنجه‌ی خطا (یعنی اختلاف بین ورودی و هدف) بین تمامی ورودی‌های یک شبکه نیست. چندین نوع سنجه‌ی خطا وجود دارد (مثلاً جمع مجذور خطاها، میانگین مجذور خطاها، آنتروپی متقاطع یا حتی سنجه‌های سفارشی) که تمامی آنها برای ضبط اختلاف بین ورودی‌ها و خروجی‌های شبکه طراحی شده‌اند.

¹ Wilson & Sharda

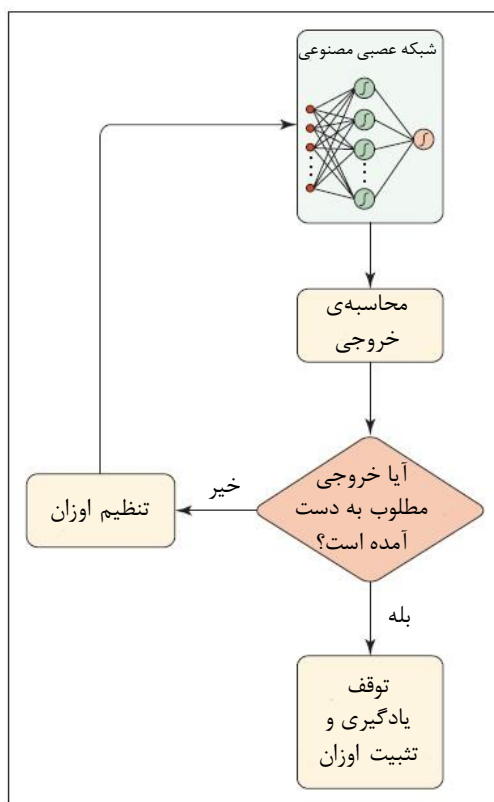
² Olson, Delen, Meng

³ Cost Function

⁴ Loss Function

فرایند آموزش با محاسبه‌ی خروجی‌ها برای مجموعه‌ی مشخصی از ورودی‌ها با استفاده از اوزان و بایاس‌های تصادفی آغاز می‌شود. پس از محاسبه‌ی خروجی‌های شبکه، تابع عملکرد را می‌توان محاسبه کرد. اختلاف بین خروجی واقعی (Y یا Y_T) و خروجی مطلوب (Z) برای مجموعه معینی از ورودی‌ها را خطای دلتا می‌نامند.

هدف این است که دلتا به کمترین مقدار ممکن برسانیم (یعنی در صورت امکان به ۰ برسانیم) که با تعدیل اوزان شبکه امکان‌پذیر است. کلید اینکار، تغییر اوزان در راستای مناسب و ایجاد تغییراتی است که دلتا (یعنی خطا) را کاهش دهد. شبکه‌های عصبی مصنوعی مختلف، دلتا را به روش‌های متفاوتی محاسبه می‌کنند که به الگوریتم یادگیری مورد استفاده بستگی دارد. صدها الگوریتم یادگیری برای موقعیت‌ها و شرایط مختلف شبکه‌های عصبی مصنوعی وجود دارد.



شکل ۶.۱۲ فرایند یادگیری نظارت شده یک شبکه عصبی مصنوعی

بهینه‌سازی عملکرد (یعنی به حداقل رساندن خطا یا دلتا) در شبکه عصبی معمولاً از طریق الگوریتمی به نام **گرادیان نزولی تصادفی**^۱ انجام می‌شود که یک بهینه‌ساز تکرارپذیر مبتنی بر گرادیان است که برای یافتن مینیمم (یعنی پایین‌ترین نقطه) در توابع عملکردی مثل شبکه‌های عصبی بکار می‌رود. ایده‌ی پشت این الگوریتم این است که مشتق تابع عملکرد با توجه به وزن یا

^۱ Stochastic Gradient Descent (SGD)

بایاس جاری نشان دهنده‌ی مقدار تغییرات در سنجه‌های خطا به ازای هر واحد تغییر در آن وزن یا بایاس است. این مشتقات را گرادین‌های شبکه می‌نامند. محاسبه‌ی گرادین‌های شبکه در شبکه‌های عصبی نیازمند بکارگیری الگوریتمی به نام **پس‌انتشار^۱** است که متداول‌ترین الگوریتم یادگیری شبکه‌های عصبی است که از **قاعده‌ی زنجیره‌ی محاسبه^۲** برای محاسبه‌ی مشتقات توابعی که از سایر توابعی که مشتقات آنها مشخص است تشکیل شده‌اند بکار می‌رود (جزئیات ریاضی این الگوریتم در رام‌هارت، هینتون و ویلیامز ۱۹۸۶ در دسترس است).

پس‌انتشار (مخفف پس انتشار خطا) پرکاربردترین الگوریتم یادگیری نظارت شده در رایانش عصبی است (پرینسیپ، یولیانو و لوفبر ۲۰۰۰). پیاده‌سازی و اجرای الگوریتم‌های پس‌انتشار با استفاده از SGD نسبتاً آسان‌تر می‌شود. یک شبکه عصبی با یادگیری پس‌انتشار شامل یک یا چند لایه‌ی پنهان است. این نوع شبکه را پیش‌بازخور می‌نامند زیرا هیچ ارتباطی بین خروجی یک عنصر پردازشی و ورودی یک گره در همان لایه یا لایه قبلی وجود ندارد. الگوهای صحیح در طی فرایند آموزش با خروجی شبکه عصبی مقایسه شده و از بازخورد برای تنظیم اوزان استفاده می‌شود تا زمانیکه تمامی الگوهای آموزشی به شکل درست و صحیح دسته‌بندی شوند (تولرانس خطا از قبل تعیین شده است).

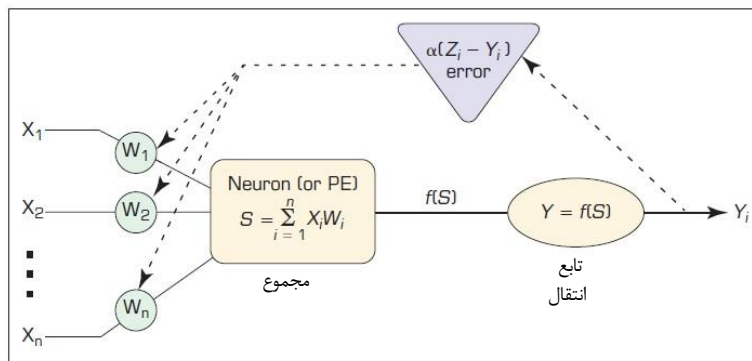
با شروع از لایه خروجی، خطاهای بین خروجی تولید شده توسط شبکه و خروجی‌های مطلوب برای تصحیح/تنظیم اوزان برای ارتباطات بین نورون‌ها بکار می‌روند (شکل ۶.۱۳). برای هر نورون خروجی j ، مقدار خطا $\Delta = (Z_j - Y_j)(df/dx)$ که Z و Y خروجی‌های مطلوب و واقعی هستند. با استفاده از تابع سیگموئید، $f = [1 + \exp(-x)]^{-1}$ که x متناسب با جمع ورودی‌های موزون نورون، مشتق تابع سیگموئید $df/dx = f(1 - f)$ و مقدار خطا نیز تابع ساده‌ای از خروجی‌های مطلوب و واقعی است. فاکتور $f(1 - f)$ یک تابع لجستیک است که تصحیح خطا را در محدوده مناسب حفظ می‌کند. وزن هر ورودی به نورون j متناسب با این خطای محاسبه شده تغییر می‌کند. یک عبارت پیچیده‌تر می‌تواند به کار گرفته شود که به روشی مشابه از نورون‌های خروجی از طریق لایه‌های پنهان کار می‌کند تا اصلاحات مربوط به وزن نورون‌های داخلی را محاسبه کند. این روش پیچیده یک رویکرد تکرارپذیر برای حل یک مسئله‌ی بهینه‌سازی غیرخطی است از لحاظ مفهومی بسیار شبیه رگرسیون خطی چندگانه است.

در پس‌انتشار، الگوریتم یادگیری شامل رویه‌های زیر است:

۱. مقداردهی اولیه‌ی اوزان با مقادیر تصادفی و تعیین سایر پارامترها
۲. وارد کردن بردار ورودی و خروجی مطلوب
۳. محاسبه‌ی خروجی واقعی از طریق محاسبات و حرکت به رو جلو در لایه‌ها
۴. محاسبه‌ی خطا
۵. تغییر اوزان از طریق روش رو به عقب از لایه‌ی خروجی بین لایه‌های پنهان

¹ Backpropagation

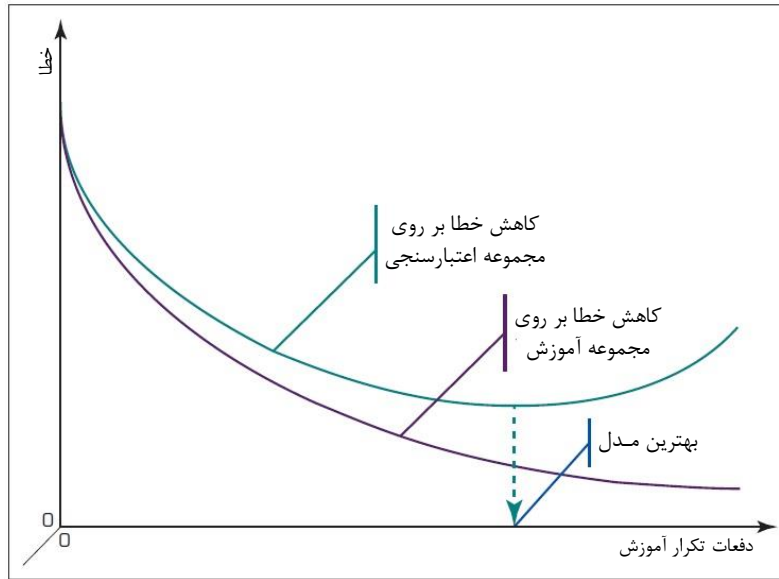
² Chain rule of calculus



شکل ۶.۱۳ پس انتشار خطا برای یک نورون

این رویه برای کل مجموعه بردارهای ورودی تکرار می‌شود تا زمانی که خروجی مطلوب و خروجی واقعی در محدوده تolerانس از پیش تعیین شده قرار بگیرند. با توجه به نیازمندی‌های محاسباتی برای یک تکرار، آموزش یک شبکه‌ی بزرگ می‌تواند زمان زیادی به طول بینجامد به همین خاطر، مجموعه‌ای از موارد به صورت رو به جلو اجرا شده و یک خطای تجمعی برای سرعت‌دهی به فرایند یادگیری به صورت رو به عقب تغذیه می‌شود. برخی اوقات، بسته به اوزان تصادفی اولیه و پارامترهای شبکه، شبکه قادر نیست تا سطح عملکردی رضایت‌بخشی را پوشش دهد. در چنین مواردی، اوزان تصادفی جدید باید تولید شوند و پارامترهای شبکه یا حتی ساختار آن باید پیش از تکرار بعدی، اصلاح شوند. هدف پژوهش جاری، توسعه‌ی الگوریتم‌هایی است که بتوان از آنها برای انتخاب پارامترهای شبکه برای بیشینه‌کردن عملکرد خروجی مطلوب استفاده کرد. در واقع، اکثر ابزارهای نرم‌افزاری تجاری شبکه‌های عصبی مصنوعی امروزه از الگوریتم‌های ژنتیک برای کمک به کاربران برای "بهینه‌سازی" پارامترهای شبکه به نحو نیمه‌خودکار استفاده می‌کنند.

یک دغدغه‌ی اصلی در آموزش هرگونه مدل یادگیری ماشین، بیش‌برازش است. این زمانی اتفاق می‌افتد که مدل آموزش‌دیده کاملاً برای مجموعه داده‌های آموزشی برازش داشته باشد ولی در مقایسه با مجموعه داده‌های بیرونی بسیار ضعیف عمل کند. بیش‌برازش موجب بروز مشکلات جدید در خصوص تعمیم‌پذیری مدل می‌شود. گروه بزرگی از استراتژی‌ها به نام استراتژی‌های منظم‌سازی برای پیشگیری از بیش‌برازش مدل‌ها از طریق ایجاد تغییرات یا تعریف محدودیت‌ها برای پارامترهای مدل یا تابع عملکردی تعریف شده‌اند. در مدل‌های کوچک شبکه‌های عصبی مصنوعی کلاسیک، یک استراتژی منظم‌سازی متداول برای اجتناب از بیش‌برازش، این است که تابع عملکرد برای یک مجموعه داده‌ی اعتبارسنجی و همچنین مجموعه داده‌های آموزش مجزا پس از هر بار تکرار مورد ارزیابی قرار بگیرد. هر زمان که بهبودی در عملکرد داده‌های اعتبارسنجی بوجود نیاید، فرایند آموزش را باید متوقف کرد. شکل ۶.۱۴ یک گراف معمول از سنجی خطا را با تعداد دفعات تکرار آموزش نشان می‌دهد. همانطور که مشاهده می‌کنید، در آغاز، مقدار خطا هم برای داده‌های آموزش و هم اعتبارسنجی با افزایش دفعات تکرار، کاهش می‌یابد ولی از یک نقطه‌ی خاص (با خط‌چین نشان داده شده)، خطای مجموعه داده‌های اعتبارسنجی شروع به افزایش می‌کند در حالیکه خطای داده‌های آموزش شروع به کاهش می‌کند. این بدان معناست که پس از این تعداد تکرار، مدل برای مجموعه داده‌هایی که برایشان آموزش دیده دچار بیش‌برازش می‌شود و در صورتیکه از داده‌های بیرونی استفاده شود عملکرد مناسبی نخواهد داشت. این نقطه در واقع بیانگر تعداد دفعات تکرار توصیه شده برای آموزش یک شبکه عصبی مفروض است.



شکل ۶.۱۴ بیش‌برازش یک مدل شبکه عصبی مصنوعی - با افزایش تعداد دفعات تکرار، نرخ خطای مجموعه داده‌های آموزش و اعتبارسنجی نیز به تدریج تغییر می‌کند.

بیش فناوری ۶.۲ نرم‌افزار شبکه عصبی مصنوعی

ابزارهای بسیاری برای توسعه‌ی شبکه‌های عصبی وجود دارند. برخی از این ابزارها همانند پوسته‌های نرم‌افزاری عمل می‌کنند. آنها مجموعه‌ای از معماری‌ها، الگوریتم‌های یادگیری و پارامترهای استاندارد را به همراه قابلیت دستکاری داده‌ها ارائه می‌کنند. برخی از ابزارهای توسعه می‌توانند از چندین پارادایم و الگوریتم یادگیری شبکه پشتیبانی کنند.

پیاده‌سازی شبکه‌های عصبی نیز در اکثر ابزارهای داده‌کاوی و تحلیل‌های پیشگویانه جامع از قبیل اس‌ای‌اس اینترپرایس ماینز، آی‌بی‌ام اس‌پی‌اس‌مدلر (نام سابق کلمنتاین) و استاتیسیتیکا دیتا ماینر در دسترس هستند. وکا، رپیدماینر، آرنج و کی‌نایم ابزارهای نرم‌افزاری داده‌کاوی رایگان منبع باز هستند که شامل قابلیت‌های شبکه‌های عصبی می‌باشند. این ابزارهای رایگان را می‌توان از وبسایت‌هایشان بارگذاری کرد: با جستجوی ساده‌ی نام این ابزارها برآحتی می‌توانیم به صفحات بارگذاری آنها بروید. همچنین، اکثر ابزارهای نرم‌افزاری تجاری برای مقاصد ارزیابی قابل بارگذاری و استفاده هستند (معمولاً از لحاظ زمان/قابلیت‌ها محدودیت‌هایی دارند).

بسیاری از ابزارهای تخصصی شبکه‌های عصبی، ساخت و جاری‌سازی یک مدل شبکه عصبی را در عمل آسان‌تر می‌کنند. هرگونه فهرستی از این ابزارها ناقص خواهد بود. منابع آنلاین از قبیل ویکی‌پدیا، دایرکتوری نرم‌افزارهای گوگل یا یاهو! و فروشندگان در **pcai.com** مکان‌های خوبی برای یافتن آخرین اطلاعات در خصوص عرضه‌کنندگان نرم‌افزارهای شبکه‌های عصبی هستند. برخی از عرضه‌کنندگان که سابقه طولانی دارند و کاربردهای صنعتی نرم‌افزارهای شبکه عصبی خودشان را گزارش داده‌اند شامل کالیفرنیا ساینتیفیک (برین‌میکر)، نیورال‌ور، نیورودایمنژن، وارد سیستمز گروپ (نیوروشل) و مگاپیوتر هستند. بازهم این لیست کامل نیست.

برخی ابزارهای توسعه‌ی شبکه عصبی مصنوعی، افزودنی‌های صفحات گسترده هستند. اکثر افراد قادر به خواندن صفحات گسترده، پایگاه‌های داده و فایل‌های متنی هستند. برخی از آنها رایگان‌افزار و برخی دیگر نرم‌افزار مشروط هستند. برخی سیستم‌های شبکه عصبی مصنوعی در محیط جاوا برای اجرای مستقیم بر روی وب توسعه یافته‌اند و از طریق یک رابط مرورگر وب قابل دسترس هستند. سایر محصولات شبکه عصبی مصنوعی برای برقراری ارتباط با سیستم‌های خبره در قالب محصولات توسعه ترکیبی طراحی شده‌اند.

برنامه‌نویسان ممکن است ترجیح دهند از زبان‌های برنامه‌نویسی عمومی‌تر مثل C، C#، C++، جاوا و غیره یا کتابخانه‌های آماده‌ی و پایتون یا صفحات گسترده برای برنامه‌نویسی مدل، انجام محاسبات و اجرای نتایج استفاده کنند. یک روش متداول در این حوزه استفاده از یک کتابخانه از روتین‌های شبکه عصبی مصنوعی است. بسیاری از عرضه‌کنندگان نرم‌افزارهای شبکه عصبی مصنوعی و پلتفرم‌های منبع-باز، اینگونه کتابخانه‌های قابل برنامه‌نویسی را عرضه می‌کنند. به عنوان مثال، hav.Software یک کتابخانه از کلاس‌های C++ را برای پیاده‌سازی شبکه‌های عصبی مستقل، جاری ساده، پیش‌بازخورد و جاری ترتیب تصادفی ارائه می‌کند. نرم‌افزارهای محاسباتی از قبیل MATLAB نیز شامل کتابخانه‌های خاص شبکه‌های عصبی هستند.

سوالاتی برای مرور بخش ۶.۴

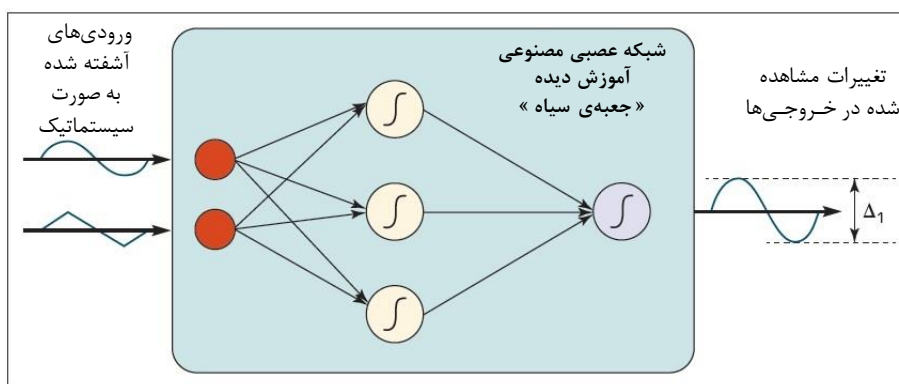
۱. نه گام در اجرای یک پروژه‌ی شبکه عصبی را نام ببرید؟
۲. برخی از پارامترهای طراحی برای توسعه‌ی یک شبکه عصبی چه هستند؟
۳. فرایند سه مرحله‌ای یادگیری در شبکه عصبی مصنوعی را ترسیم کرده و به اختصار توضیح دهید.
۴. یادگیری پس‌انتشار چگونه کار می‌کند؟
۵. بیش‌برازش در یادگیری شبکه عصبی مصنوعی چیست؟ چگونه اتفاق می‌افتد و چگونه می‌توان از آن پیشگیری کرد؟
۶. انواع مختلف نرم‌افزارهای شبکه عصبی چه هستند؟

از شبکه‌های عصبی به عنوان ابزاری مؤثر برای حل مسائل بسیار پیچیده‌ی دنیای واقعی در حوزه‌های کاربردی متنوعی استفاده شده است. اگرچه شبکه‌های عصبی مصنوعی در بسیاری از سناریوها به عنوان پیشگوها و/یا شناساگرهای خوشه‌ای سطح بالاتری اثبات شده‌اند، در برخی کاربردها همچنان نیاز است بدانیم که این مدل چگونه کار می‌کند. شبکه‌های عصبی مصنوعی را معمولاً جعبه‌ی سیاه می‌نامند که اگرچه قادر به حل مسائل پیچیده هستند ولی توضیحی برای این قابلیت‌ها و توانایی‌هایشان نداریم. این فقدان شفافیت را عمدتاً سندروم "جعبه‌ی سیاه" می‌نامند.

بسیار مهم است که بتوانیم "ماهیت درونی" یک مدل را توضیح دهیم و چنین توضیحی تضمین می‌کند که این شبکه به خوبی آموزش دیده و در صورت اجرا در یک محیط تحلیلی کسب‌وکار به شکلی مطلوب و مورد انتظار رفتار خواهد کرد. چنین نیازی به "پی بردن به باطن مدل‌ها" ممکن است برای یک مجموعه آموزش نسبتاً کوچک (به خاطر هزینه‌ی بالایی ضبط داده‌ها) یا الزامات قانونی بسیار شدید، قابل استناد باشد. یک مثال از چنین کاربردی، بکارگیری کیسه‌های هوا در وسایل نقلیه است. در اینجا هم هزینه‌ی ضبط داده‌ها (وسایل نقلیه‌ای که تصادف می‌کنند) و هم ملاحظات الزامات قانونی (خطر از دست رفتن جان انسان‌ها) دارای اهمیت هستند. یک مثال بارز دیگر برای اهمیت توضیح نحوه عملکرد مدل‌ها، رسیدگی به درخواست‌های وام

است. اگر درخواست وام یک متقاضی رد شود، او حق دارد که علت آن را بداند. داشتن یک سیستم پیش‌بینی که قادر به تشخیص درخواست‌های خوب و بد از هم باشد اگر نتواند توجیهی برای پیش‌بینی‌هایش ارائه کند به تنهایی کافی نیست.

تکنیک‌های متنوعی برای تحلیل و ارزشیابی شبکه‌های عصبی آموزش دیده پیشنهاد شده است. این تکنیک‌ها تفسیر روشنی از نحوه کارکرد یک شبکه عصبی ارائه می‌کنند به بیان دقیق‌تر اینکه تک تک ورودی‌ها چگونه منجر به تولید خروجی یک شبکه خاص می‌شوند. در اینجا از تحلیل حساسیت به عنوان تکنیک اصلی برای روشن‌نگری این جعبه‌ی سیاه نام برده شده است. تحلیل حساسیت، روشی برای استخراج ارتباطات علت و معلولی در بین ورودی‌ها و خروجی‌ی یک مدل شبکه عصبی آموزش دیده است. در فرایند تحلیل حساسیت، توانایی یادگیری شبکه عصبی آموزش دیده غیرفعال می‌شود تا اوزان شبکه تحت تأثیر قرار نگیرند. رویه اصلی در تحلیل حساسیت این است که ورودی‌های شبکه به صورت سیستماتیک را در دامنه‌ی مقادیر مجاز، آشفته می‌کنند و تغییر متناظر با آن در خروجی برای هر متغیر ورودی ثبت می‌شود (پرینسیپ و همکاران ۲۰۰۰). شکل ۶.۱۵ یک تصویر گرافیکی از این فرایند را نشان می‌دهد. اولین ورودی بین مقدار میانگین به علاوه و منهای تعداد تعریف شده‌ای از انحراف استاندارد‌ها تغییر داده می‌شود (یا برای متغیرهای رسته‌ای، از تمامی مقادیر ممکن آن استفاده می‌شود) در حالیکه تمامی سایر متغیرهای ورودی نسبت به مقادیر میانگین‌شان (یا مُد) ثابت نگهداشته می‌شوند. خروجی شبکه برای تعداد مشخص شده‌ای از گام‌ها در بالا و پایین میانگین محاسبه می‌شود. این فرایند برای هر ورودی تکرار می‌شود. در نتیجه، گزارشی برای خلاصه‌سازی نوسانات هر خروجی متناظر با میزان تغییر و نوسان در هر ورودی تهیه می‌شود. گزارش تولید شده اغلب حاوی یک نمودار ستونی با مقادیر عددی که بر روی محور افقی x نمایش داده شده‌اند می‌باشد که مقادیر حساسیت نسبی هر متغیر ورودی را گزارش می‌دهد. یک مثال بارز از تحلیل حساسیت بر روی مدل شبکه عصبی مصنوعی در مورد کاربردی ۶.۴ آورده شده است.



شکل ۶.۱۵ تحلیل حساسیت بر روی یک مدل شبکه عصبی مصنوعی

طبق آمار سازمان ایمنی ترافیک بزرگراه‌های ملی آمریکا (ان‌اچ‌تی‌اس‌ای)^۱، بیش از ۶ میلیون تصادف جاده‌ای هر ساله در ایالت متحده جان بیش از ۴۱،۰۰۰ نفر را می‌گیرد. علل تصادفات و شدت آسیب‌ها و جراحات مربوطه برای پژوهشگران ایمنی ترافیک و جاده‌ها از اهمیت بسیار بالایی برخوردار است. هدف از اینگونه پژوهش‌ها، نه تنها کاهش تعداد تصادفات، بلکه کاهش شدت آسیب‌ها و جراحات نیز می‌باشد. یک راه برای کاهش شدت آسیب‌ها، شناسایی عمیق‌ترین عواملی است که بر شدت آسیب‌ها تأثیر می‌گذارد. درک شرایطی که رانندگان و مسافران تحت آنها بیشترین احتمال آسیب‌دیدگی یا کشته شدن در یک تصادف وسیله نقلیه را دارند می‌تواند به بهبود وضعیت ایمنی رانندگی‌ها کمک کند. عواملی که به طور بالقوه موجب بالا رفتن ریسک شدن آسیب‌دیدگی سرنشینان خودروها در صورت بروز تصادفات می‌شوند شامل مشخصه‌های جمعیت‌شناختی و/یا رفتاری شخص (مثلاً سن، جنسیت، بستن کمربند ایمنی، استفاده از مواد مخدر یا الکل در حین رانندگی)، عوامل محیطی و/یا شرایط راه‌ها و جاده‌ها در زمان بروز تصادف (مثلاً شرایط سطح جاده، شرایط نور و روشنایی یا آب و هوا) و همچنین مشخصه‌های فنی خود وسیله نقلیه (مثلاً سن خودرو، نوع بدنه) می‌باشد.

دیلن و همکاران (۲۰۰۶) در یک مطالعه‌ی داده‌کاوی اکتشافی از نمونه بزرگی از داده‌ها - ۳۰،۳۵۸ پرونده تصادف گزارش شده توسط پلیس که از سیستم تخمین عمومی ان‌اچ‌تی‌اس‌ای بدست آمده بود - برای شناسایی عواملی که اهمیت اساسی در تشدید احتمال مصدومیت شدید در حین تصادفات دارند استفاده کردند. تصادفات مورد بررسی در این مطالعه شامل یک نمونه جغرافیایی در تصادفات چنیدن خودرو با یکدیگر، تصادفات یک خودرو با مانع ثابت و تصادفات غیربرخوردی یک خودرو (واژگون شدن) بوده‌اند.

برخلاف بسیاری از مطالعات پیشین که در این حوزه انجام شده‌اند و عمدتاً از مدل‌های خطی تعمیم یافته‌ی رگرسیونی استفاده کرده‌اند که ارتباطات وظیفه‌ای و کارکردی بین شدت آسیب و عوامل مربوط به تصادف به صورت خطی فرض می‌شوند (یک حالت بسیار ساده شده از موقعیت‌های دنیای واقعی). دیلن و همکارانش (۲۰۰۶) تصمیم گرفتند تا به شکل متفاوتی عمل کنند. از آنجائیکه شبکه‌های عصبی مصنوعی در ضبط و ثبت ارتباطات پیچیده و غیرخطی بین متغیرهای پیشگو (عوامل تصادف) و متغیر هدف (میزان شدت آسیب‌ها) عملکرد بهتری دارند، آنها تصمیم گرفتند تا از مجموعه‌ای از مدل‌های شبکه‌های عصبی مصنوعی برای تخمین اهمیت عوامل تصادف بر روی سطح شدت آسیب وارده به راننده استفاده کنند.

دیلن و همکارانش (۲۰۰۶) از یک نقطه نظر متدولوژیکی، از یک فرایند دو مرحله‌ای پیروی کردند. در گام اول، آنها مجموعه‌ای از مدل‌های پیش‌بینی را ایجاد کردند (برای هر سطح شدت آسیب‌دیدگی) تا بتوانند ارتباطات عمیق بین عوامل مربوط به تصادف و سطح مشخصی از شدت آسیب‌دیدگی ثبت کنند. در گام دوم، آنها یک تحلیل حساسیت بر روی مدل‌های شبکه عصبی آموزش دیده انجام دادند تا اهمیت اولویت‌بندی شده‌ی عوامل مربوط به تصادفات را در رابطه با شدت آسیب‌دیدگی تعیین کنند. در فرمولاسیون این مطالعه، مسئله‌ی پیش‌بینی پنج دسته‌ای به منظور دستیابی به سطح دانه‌بندی مورد نیاز اطلاعات برای شناسایی ارتباطات علت و معلولی "درست و حقیقی" بین عوامل مربوط به تصادف و سطوح مختلف شدت آسیب‌دیدگی به چندین مدل طبقه‌بندی دودویی تجزیه شد. شکل ۶.۱۶ هشت مدل شبکه عصبی مختلف که در این تحلیل حساسیت برای شناسایی عوامل تعیین کننده‌ی اصلی سطوح شدت آسیب توسعه یافته و بکار رفته‌اند را نشان می‌دهد.

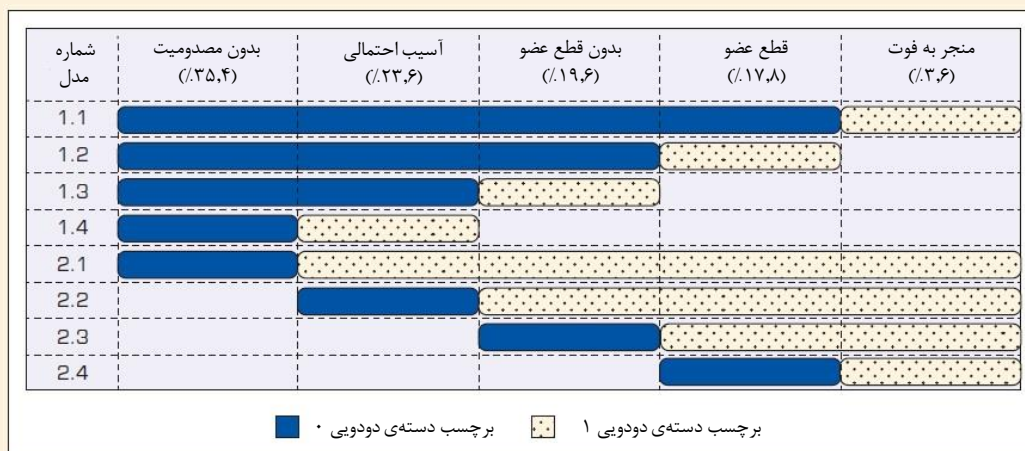
¹ National Highway Traffic Safety Administration (NHTSA)

نتایج، تفاوت‌های قابل توجهی را بین مدل‌های ساخته شده برای سطوح مختلف شدت آسیب‌دیدگی نشان می‌دهند. این بدان معناست که تأثیرگذارترین عوامل در مدل‌های پیش‌بینی، شدیداً به سطح شدت آسیب‌دیدگی بستگی دارند. به عنوان مثال، این مطالعه نشان می‌دهد که استفاده از متغیر کمربندی ایمنی، مهمترین عامل تعیین کننده برای پیش‌بینی سطوح بالاتر شدت مصدومیت و آسیب‌دیدگی است ولی برای سطوح پایین تر شدت آسیب‌دیدگی جزو کم اهمیت‌ترین پیشگوها بوده است. یک یافته‌ی جالب دیگر مربوط به جنسیت است: جنسیت راننده جزو مهمترین پیشگوها برای سطوح آسیب‌دیدگی پایین است ولی برای سطوح شدت آسیب‌دیدگی بالا جزو عوامل مهم نبوده است که نشان می‌دهد که آسیب‌های جدی تر به جنسیت راننده ارتباطی ندارند. یک یافته‌ی جالب دیگر در این مطالعه نشان می‌دهد که با افزایش شدت آسیب‌دیدگی، عامل سن اهمیت بیشتری می‌یابد بدین معنا که افراد سالمندتر در مقایسه با افراد جوان تر در بروز تصادفات به احتمال بیشتری دچار مصدومیت‌های شدید (و مرگ و میر) می‌شوند.

Sources: Delen, D., R. Sharda, & M. Bessonov. (2006). "Identifying Significant Predictors of Injury Severity in Traffic Accidents Using a Series of Artificial Neural Networks." *Accident Analysis and Prevention*, 38(3), pp. 434-444; Delen, D., L. Tomak, K. Topuz, & E. Eryarsoy (2017). "Investigating Injury Severity Risk Factors in Automobile Crashes with Predictive Analytics and Sensitivity Analysis Methods." *Journal of Transport & Health*, 4, pp. 118-131.

سوالاتی برای مورد کاربردی ۶.۴

۱. تحلیل حساسیت چگونه می‌تواند در مورد این جعبه‌ی سیاه (یعنی شبکه‌های عصبی) روشن‌گری کند؟
۲. چرا باید از یک ابزار جعبه‌ی سیاه مثل شبکه‌های عصبی به جای ابزارهای آماری شفاف‌تر مثل رگرسیون لجستیک استفاده کرد؟
۳. در این مورد، شبکه‌های عصبی و تحلیل حساسیت چگونه توانستند به شناسایی عوامل شدت-آسیب در تصادفات جاده‌ای کمک کنند؟



شکل ۶.۱۶ نمایش گرافیکی هشت مدل شبکه عصبی مصنوعی دودویی

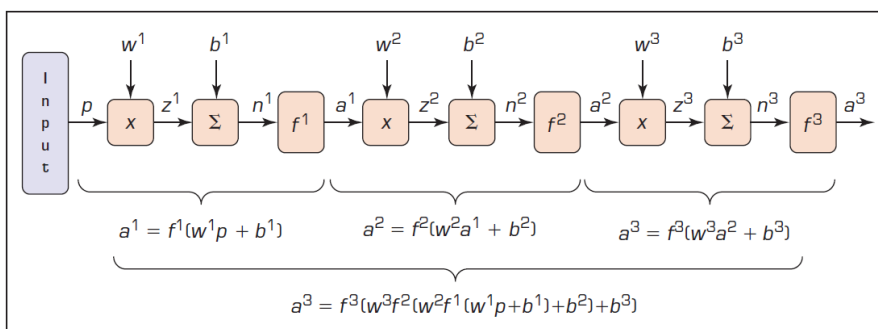
سوالاتی برای مرور بخش ۶.۵

۵. سندروم معروف به جعبه سیاه چیست؟
۶. چرا قابلیت توضیح دادن ساختار یک مدل شبکه عصبی مصنوعی مهم است؟
۷. تحلیل حساسیت در شبکه عصبی مصنوعی چگونه کار می‌کند؟
۸. برای یافتن سایر روش‌های توضیح دادن شبکه عصبی مصنوعی در اینترنت جستجو کنید. نتایج را گزارش دهید.

تا همین اواخر (تا پیش از پیدایش پدیده‌ی یادگیری عمیق)، اکثر کاربردهای شبکه عصبی شامل معماری‌های شبکه با تعداد کمی لایه‌ی پنهان و تعداد محدودی نورون در هر لایه بودند. حتی در کاربردهای نسبتاً پیچیده شبکه‌های عصبی نیز تعداد نورون‌ها در شبکه‌ها به زحمت از چند هزار تجاوز می‌کرد. در واقع، قابلیت پردازش رایانه‌ها در آن زمان یک عامل محدود کننده به شمار می‌رفت که واحدهای پردازش مرکزی به سختی می‌توانستند شبکه‌هایی با بیش از چند لایه را در زمان قابل قبولی اجرا کنند. در سال‌های اخیر، توسعه‌ی واحدهای پردازش گرافیکی (جی پی یو) به همراه زبان‌های برنامه‌نویسی مربوطه (مثلاً CUDA از شرکت NVIDIA) انسان را قادر ساخته تا از آنها برای مقاصد تحلیل داده‌ها استفاده کند منجر به کاربردهای پیشرفته‌تر شبکه‌های عصبی شده است. فناوری جی پی یو ما را قادر ساخته است تا شبکه‌های عصبی با بیش از یک میلیون نورون را اجرا کنیم. این شبکه‌های بزرگتر می‌توانند عمیق‌تر به درون ویژگی‌های داده‌ها نفوذ کرده و الگوهای پیچیده‌تری که به شکل دیگری امکان تشخیص آنها وجود ندارد را استخراج کنند. اگرچه شبکه‌های عمیق می‌توانند تعداد بسیار زیادی از متغیرهای ورودی را مدیریت کنند ولی برای اینکه به خوبی آموزش ببینند نیازمند مجموعه داده‌های نسبتاً بزرگتری هستند و استفاده از مجموعه داده‌های کوچک برای آموزش شبکه‌های عمیق معمولاً منجر به بیش‌برازش مدل برای آموزش داده‌ها و نتایج ضعیف و نامعتبر در صورت بکارگیری مدل برای داده‌های بیرونی می‌شود. به لطف فناوری‌ها و ابزارهای ضبط داده‌ی مبتنی بر اینترنت و اشیاء، در حال حاضر مجموعه داده‌های بزرگتر در بسیاری از حوزه‌های کاربردی برای آموزش شبکه‌های عصبی عمیق‌تر وجود دارد. ورودی یک مدل شبکه عصبی مصنوعی معمولاً یک آرایه به اندازه‌ی $R \times 1$ است که R تعداد متغیر ورودی می‌باشد. در شبکه‌های عمیق، ولی می‌توانیم از تَنسورها (آرایه‌های n بُعدی) به عنوان ورودی استفاده کنیم. به عنوان مثال، در شبکه‌های تشخیص تصویر، هر ورودی (یعنی تصویر) را می‌توان با یک ماتریس و کدهای رنگی در پیکسل‌های تصویر نشان داد یا برای مقاصد پردازش ویدئویی، هر ویدئو را می‌توان با چندین ماتریس نمایش داد (یعنی یک تَنسور سه بُعدی) که هر کدام از آنها نمایانگر یک تصویر در ویدئو هستند. به عبارت دیگر، تَنسورها به ما این امکان را می‌دهند تا ابعاد بیشتری را در تحلیل مجموعه داده‌ها در نظر بگیریم (مثلاً زمان، مکان).

به استثنای این تفاوت‌های کلی، انواع مختلف شبکه‌های عمیق شامل تغییرات مختلف در معماری شبکه‌های عصبی استاندارد هستند که آنها را به قابلیت‌های متمایزی برای مدیریت انواع خاص داده‌ها برای مقاصد پیشرفته مجهز می‌کنند. در بخش بعدی به بحث در خصوص برخی از این انواع خاص شبکه‌ها و مشخصه‌های آنها می‌پردازیم.

شبکه‌های عمیق پرسپترون چندلایه پیشخور که با نام شبکه‌های پیش‌بازخور عمیق^۱ نیز شناخته می‌شوند از عمومی‌ترین انواع شبکه‌های عمیق هستند. این شبکه‌ها صرفاً شبکه‌های عصبی بزرگی هستند که می‌توانند حاوی لایه‌های بسیاری از نورون‌ها بوده و از تانسورها به عنوان ورودی خود استفاده کنند. انواع و مشخصه‌های عناصر شبکه (یعنی توابع اوزان، توابع انتقال) بسیار شبیه مدل‌های شبکه عصبی مصنوعی استاندارد هستند. این مدل‌ها را به این دلیل پیشخور می‌نامند که جریان اطلاعاتی که از آنها عبور می‌کنند همواره رو به جلو است و هیچگونه ارتباطات رو به عقب (یعنی ارتباطاتی که در آنها خروجی یک مدل به خودش برمی‌گردد) مجاز نمی‌باشد. شبکه‌های عصبی که در آنها ارتباطات بازخوردی مجاز باشد را شبکه‌های عصبی جاری می‌نامند. معماری شبکه‌های عصبی جاری معمول و همچنین یک نوع خاص از آن به نام شبکه‌های حافظه کوتاه مدت طولانی در بخش‌های بعدی این فصل مورد بحث و بررسی قرار خواهند گرفت.



شکل ۶.۱۷ نمایش برداری سه لایه‌ی اول در یک شبکه‌ی پرسپترون چندلایه‌ی پیشخور

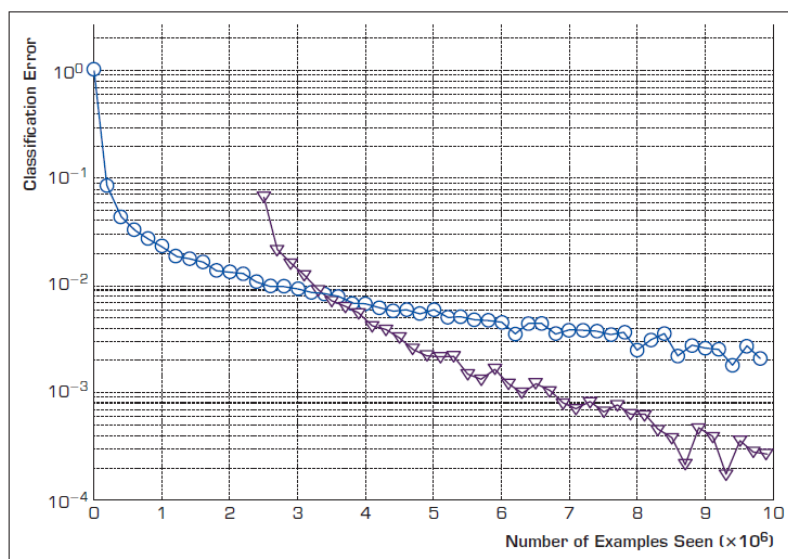
به بیان کلی، یک ترتیب متوالی از لایه‌ها باید بین لایه‌های ورودی و خروجی در معماری شبکه پرسپترون چندلایه پیشخور وجود داشته باشد. این بدان معناست که بردار ورودی باید به طور متوالی از بین تمامی لایه‌ها عبور کند و نمی‌تواند هیچکدام از آنها را جا بیندازد. علاوه بر این، مستقیماً نمی‌تواند به هر لایه‌ای به جز اولین لایه متصل شود و خروجی هر لایه، ورودی لایه بعدی خواهد بود. شکل ۶.۱۷ یک نمایش برداری از سه لایه‌ی اول یک شبکه‌ی پرسپترون چندلایه پیشخور را نشان می‌دهد. همانطور که مشاهده می‌کنید، فقط یک بردار وارد هر لایه می‌شود که یا بردار ورودی اصلی (p برای لایه اول) یا بردار خروجی از لایه پنهان قبلی در معماری شبکه است (a^{i-1} برای لایه i ام). با وجود این، حالات خاصی از معماری شبکه‌های پرسپترون چندلایه پیشخور وجود دارد که برای اهداف خاصی طراحی شده‌اند که این اصول ممکن است در آنها برقرار نباشد.

بهینه‌سازی تابع عملکرد در بسیاری از کاربردهای حقیقی شبکه‌های ام‌ال پی عمیق یک مسئله‌ی چالش‌برانگیز است. مسئله این است که بکارگیری الگوریتم‌های آموزش گرادینتی با مقاردهی اولیه‌ی تصادفی اوزان و بایاس‌ها که برای یافتن مجموعه پارامترهای بهینه در شبکه‌های عصبی کم‌عمق بسیار مؤثر است به جای اینکه بهترین مقادیر مطلق را برای پارامترها نتیجه دهد ممکن است ما را درگیر راه‌حل‌های بهینه‌ی موضعی کند. با افزایش عمق شبکه، احتمال رسیدن به یک جواب بهینه مطلق با استفاده از

¹ Deep Feedforward Networks

مقداردهی اولیه تصادفی با استفاده از الگوریتم‌های گرادیانی کاهش می‌یابد. در چنین مواردی، معمولاً وابسته بودن به پارامترهای شبکه با استفاده از برخی روش‌های یادگیری عمیق بدون نظارت مثل شبکه‌های عمیق می‌تواند مفید باشد (هینتون، آسیندرو و ته ۲۰۰۶). شبکه‌های عمیق یک نوع از دسته‌ی بزرگ شبکه‌های عصبی عمیق به نام مدل‌های مولد هستند. ظهور شبکه‌های عمیق در سال ۲۰۰۶، سرآغاز رنسانس یادگیری عمیق جاری محسوب می‌شود (گودفلو و همکاران ۲۰۱۶) زیرا تا پیش از آن، بهینه‌سازی مدل‌های عمیق کار بسیار دشواری محسوب می‌شود. در واقع، امروزه کاربرد اصلی شبکه‌های عمیق، بهبود مدل‌های طبقه‌بندی از طریق حفظ پارامترهای آنها می‌باشد.

با استفاده از این روش‌های یادگیری نظارت نشده می‌توانیم با شروع از لایه اولیه و استفاده از خروجی هر لایه به عنوان ورودی لایه بعدی و مقداردهی اولیه پارامترها در سرتاسر شبکه لایه‌های ام‌ال پی را آموزش دهیم. سپس از این پارامترها می‌توان به جای پارامترهای مقداردهی شده به صورت تصادفی به عنوان مقادیر اولیه در یادگیری نظارت شده استفاده کرد. شکل ۶.۱۸ خطاهای طبقه‌بندی که ناشی از آموزش یک شبکه‌ی ام‌ال پی عمیق هستند را با (دایره‌های آبی) و بدون (مثلث‌های سیاه) پارامترهای مربوطه نشان می‌دهد (بنجیائو ۲۰۰۹). در این مثال، خط آبی بیانگر نرخ خطای مشاهده شده‌ی تست مدل طبقه‌بندی (بر روی ۱۰۰۰ مثال) آموزش دیده با استفاده از یک روش نظارت شده‌ی محض با ۱۰ میلیون مثال است در حالیکه خط سیاه نشانگر نرخ خطای همان مجموعه داده‌های تست با ۲.۵ میلیون مثال است که در ابتدا برای آموزش نظارت نشده‌ی پارامترهای شبکه و سپس ۷.۵ میلیون مثال دیگر همراه با پارامترهای مقداردهی اولیه شده برای آموزش یک مدل طبقه‌بندی نظارت شده بکار رفته‌اند. دیاگرام‌ها به وضوح بهبود نرخ خطای طبقه‌بندی در مدل را نشان می‌دهند.



شکل ۶.۱۸ تأثیر پارامترهای شبکه بر بهبود نتایج یک شبکه عصبی عمیق از نوع طبقه‌بندی

یک سوال مهم در خصوص مدل‌های پرسپترون چندلایه پیش‌بازخورد عمیق این است که "آیا معنی دارد که چنین شبکه‌هایی را فقط با چند لایه ایجاد کنیم ولی در هر لایه نورون‌های زیادی قرار دهیم؟" به عبارت دیگر، سوال این است که چرا به شبکه‌های پرسپترون چندلایه پیش‌بازخورد عمیق با لایه‌های زیاد نیاز داریم وقتی می‌توانیم همان تعداد نورون را فقط در چند لایه قرار دهیم

(یعمی شبکه‌های گسترده‌تر به جای شبکه‌های عمیق‌تر). بر اساس نظریه‌ی تقرب جهانی (سایینکو ۱۹۸۹، هورنیک ۱۹۹۱)، یک شبکه‌ی ام‌ال پی تک لایه بسیار بزرگ قادر است تا هر تابعی را تقریب بزند. اگرچه این موضوع در حد نظریه است ولی چنین لایه‌ای با تعداد نورون‌های زیاد ممکن است بیش از اندازه بزرگ بوده و قادر به یادگیری صحیح الگوهای زیربنایی نباشد. یک شبکه‌ی عمیق‌تر می‌تواند تعداد نورون‌های مورد نیاز در هر لایه و خطای تعمیم را کاهش دهد. اگرچه این موضوع از لحاظ نظری همچنان تحت پژوهش است ولی از لحاظ عملی، استفاده از تعداد لایه‌های بیشتر در یک شبکه به نظر مؤثرتر و از لحاظ محاسباتی مقرون به صرفه‌تر از بکارگیری تعداد نورون‌های بسیار زیاد در لایه‌های معدود است.

شبکه‌های پرسپترون چندلایه همانند شبکه‌های عصبی مصنوعی را همچنین می‌توان برای مقاصد پیش‌بینی، طبقه‌بندی و خوشه‌بندی بکار برد. به ویژه زمانی که تعداد متغیرهای ورودی بسیار زیادی درگیر بوده یا در مواردیکه ماهیت ورودی باید به صورت یک آرایه‌ی n بعدی باشد باید از یک طراحی شبکه چندلایه عمیق بهره جست.

مورد کاربردی ۶.۵ یک مورد عالی از بکارگیری تحلیل‌های پیشرفته برای مدیریت بهتر جریان ترافیک در شهرهای پر ازدحام ارائه می‌کند.

تحلیل‌های محدودیت سرعت به وزارت ترابری جورجیا در مدیریت ازدحام رفت و آمدها کمک می‌کند

مورد کاربردی ۶.۵

تاریخچه

زمانیکه وزارت ترابری جورجیا تصمیم به بهینه‌سازی استفاده از کلان‌داده‌ها و تحلیل‌های پیشرفته برای دستیابی به بینش و آگاهی در زمینه حمل و نقل شد، برای توسعه‌ی پروژه‌ی پایلوت محدودیت سرعت متغیر این وزارت خانه^۱ با ترادپتا همکاری نمود. مفهوم محدودیت سرعت متغیر (وی‌اس‌ال)^۲ در بسیاری از بخش‌های جهان بکار گرفته شده ولی در ایالات متحده همچنان مفهوم جدیدی محسوب می‌شود. طبق توضیحات وزارت ترابری جورجیا،

وی‌اس‌ال، محدودیت‌های سرعتی هستند که بر اساس شرایط جاده، ترافیک و آب و هوا تغییر می‌کنند. علائم الکترونیکی که در صورت بد شدن شرایط آب و هوا یا ازدحام رفت و آمد، ترافیک را کند می‌کنند تا جریان به نرمی انجام شود و تصادفات کاهش یابد. این فناوری پیشرفته و کم هزینه به صورت لحظه‌ای به رانندگان هشدار می‌دهد تا با توجه به شرایط جاده، سرعت خود را کاهش دهند. علائم افزایش سرعت در شرایط تغییر مسیر یا دنده عقب از بروز تصادفات ناشی از توقف‌های ناگهانی پیشگیری می‌کند.

کم‌سازی مزایای خدمات مشتری، ایمنی و اثربخشی وی‌اس‌ال برای وزارت ترابری بی‌اندازه مهم است و جزو نیازهای گسترده‌تر برای درک تأثیرات سرمایه‌گذاری‌های انجام شده در سیستم‌های حمل و نقل هوشمند و همچنین سایر سیستم‌ها و زیرساخت‌های حمل و نقل قرار می‌گیرد.

¹ Georgia Department of Transportation Variable Speed Limit Pilot Project (GDOT)

² VSL

پروژه‌ی پایلوت وی‌اس‌ال بر روی بزرگراه بین ایالتی I-285 در آتلانتا

وزارت حمل و نقل جورجیا یک پروژه‌ی پایلوت وی‌اس‌ال را بر روی نیمه شمالی بزرگراه I-285 که به دور آتلانتا کشیده شده اجرا کرد. این مسیر ۳۶ میلی از بزرگراه مجهز به ۸۸ علامت محدودیت سرعت الکترونیکی بوده که محدودیت سرعت را با بازه‌های ۱۰ مایل در ساعت از ۶۵ مایل در ساعت تا حداقل ۳۵ مایل در ساعت تنظیم می‌کند. دو هدف از این پروژه دنبال می‌شود:

۱. تجزیه و تحلیل سرعت در بزرگراه قبل و بعد از اجرای وی‌اس‌ال.

۲. اندازه‌گیری تأثیر وی‌اس‌ال بر شرایط رانندگی.

برای دستیابی به یک دید اولیه از ترافیک، راه‌حل علم داده‌های ترادیتا به شناسایی محل‌ها و مدت زمان "کند شدن‌های دائمی" پرداخت. اگر سرعت در بزرگراه‌ها بیش از "سرعت مرجع" باشد، آنگاه ترافیک به صورت آزاد و روان در حال جریان است. اگر سرعت در هر نقطه از بزرگراه کمتر از سرعت مرجع باشد آنگاه با ازدحام ترافیک مواجه هستیم. زمانیکه ازدحام ترافیک و کند شدن جریان حرکت وسایل نقلیه چندین دقیقه‌ی متوالی ادامه داشته باشد یک کند شدن دائمی تعریف می‌شود.

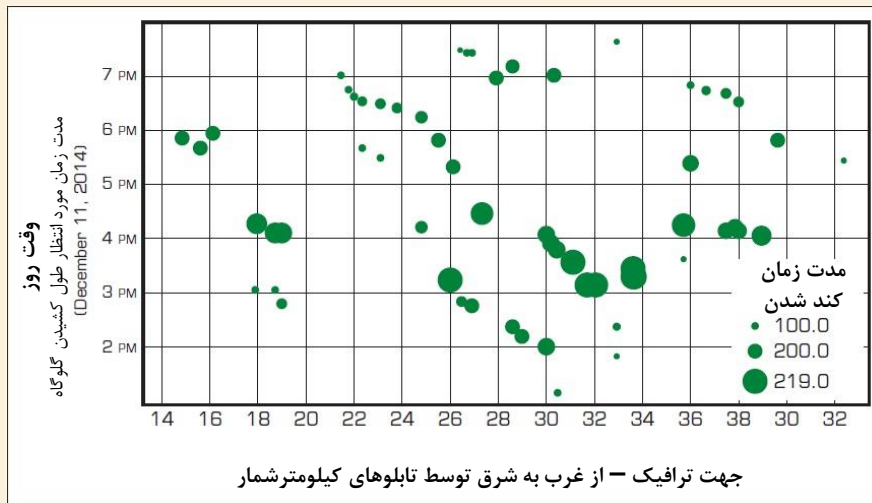
با ارائه‌ی یک تعریف تحلیلی از کند شدن‌ها، می‌توان داده‌های حجیم و بسیار متغیر سرعت را به الگوهایی برای تحقیق و بررسی دقیق‌تر تبدیل کرد. تحلیل‌های اولیه داده‌ها نشان داد که تواتر و مدت زمان کند شدن‌ها در یک بزرگراه در جهت‌های عقربه ساعت و خلاف عقربه ساعت می‌تواند بسیار متفاوت باشد.

برای اینکه نحوه تأثیرگذاری کند شدن‌ها بر ترافیک بزرگراه‌ها را بهتر درک کنیم، بهتر است تعریف جدید را در نظر گرفته و بر روی یک موقعیت خاص تمرکز کنیم. شکل ۶.۱۹ یک بعد از ظهر خاص ولی معمول آتلانتا را در بزرگراه I-285 نشان می‌دهد و اینجا بخشی از بزرگراه است که ترافیک در جهت عقربه‌های ساعت از غرب به شرق بین تابلوهای کیلومترشمار MM10 در غرب تا انتهای شرقی در MM46 جریان دارد.

اولین کند شدن مهم در ساعت ۳ بعد از ظهر نزدیک تابلوی کیلومتر MM32 رخ داد. اندازه دوایر بیانگر مدت زمان (بر حسب دقیقه) هستند. کند شدن در MM32 تقریباً چهار ساعت به طول انجامید. با تداوم این کند شدن، سرعت رفت و آمد پشت این ترافیک نیز کاهش یافت. کند شدن در MM32 تبدیل به یک گلوگاه شد که موجب شد تا ترافیک پشت آن نیز کند شود. "دنباله‌ی" ترافیک در قسمت بالا سمت چپ شکل ۶.۲۰ اطلاعات متوالی کند شدن‌ها را در MM32 و سپس در غرب نشان می‌دهد که هر کدام در بعد از ظهر شروع شده و خیلی طولانی نبوده‌اند.

اندازه‌گیری تغییرات سرعت بزرگراه‌ها

الگوی کند شدن‌ها در بزرگراه‌ها و همچنین زمان و مکان‌های متفاوت آنها موجب شد تا تأثیر آنها را بر رانندگان بررسی کنیم. اگر وی‌اس‌ال بتواند به رانندگان کمک کند تا ماهیت توقف و حرکت در کند شدن‌ها را بهتر پیش‌بینی کنند آنگاه کمی‌سازی تأثیر آنها می‌تواند برای وزارت ترابری جورجیا بسیار مفید باشد. وزارت ترابری جورجیا بسیار علاقمند بود بداند که زمانیکه رانندگان برای اولین بار با کند شدن مواجه می‌شوند چه می‌کنند. "اگرچه نمی‌دانیم که علت کند شدن‌ها چیست ولی می‌دانیم که رانندگان سرعت خود را تنظیم کرده‌اند. اگر کند شدن به دلیل بروز تصادف بوده باشد، آنگاه کاهش سرعت به طور ناگهانی اتفاق می‌افتد ولی اگر کند شدن به دلیل افزایش حجم ترافیک بوده باشد آنگاه کاهش سرعت ممکن است تدریجی اتفاق افتاده باشد."

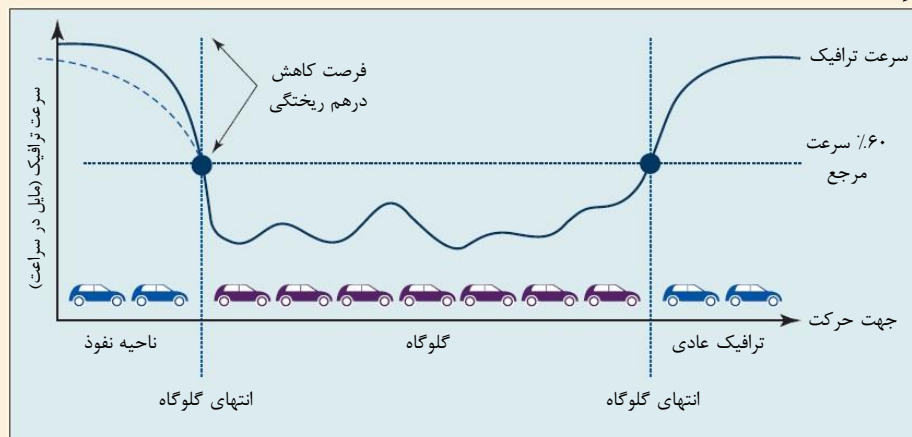


شکل ۶.۱۹ جریان ترافیک در جهت عقربه‌های ساعت در بعد از ظهر

شناسایی گلوگاه‌ها و بهم ریختگی ترافیک

یک گلوگاه به صورت یک کند شدن در یک مکان خاص شروع می‌شود. چیزی شبیه یک "نقطه‌ی پینچ" در بزرگراه رخ می‌دهد. سپس، در طول یک دوره‌ی زمانی، سرعت ترافیک در پشت این نقطه‌ی پینچ کاهش می‌یابد. یک گلوگاه، بخشی از بزرگراه است که سرعت ترافیک به کمتر از ۶۰ درصد سرعت مرجع رسیده و به طول چندین مایل ممکن است به همین وضع باقی بماند. شکل ۶.۲۰ یک نمایش مفهومی از یک گلوگاه را نشان می‌دهد.

اگرچه گلوگاه‌ها توسط نقطه‌ی پینچ یا کند شدن شروع می‌شوند که نوکِ صف را تشکیل می‌دهند ولی این انتهای صف است که برای ما جالب توجه است. بخشی که در پشت صف قرار دارد جایی است که ترافیک از یک جریان آزاد و روان به یک جریان کند و آهسته تغییر شکل می‌دهد که این تغییر در انتهای صف بسیار سریع می‌شود. رانندگانی که با سرعت مجاز بزرگراه در حال حرکت هستند ممکن است به طور غیرمنتظره‌ای با ترافیک آهسته مواجه شوند. این شرایط مستعد بروز تصادفات بوده و جایی است که وی‌اِس‌اِل می‌تواند ارزش واقعی خود را نشان دهد.



شکل ۶.۲۰ نمایش گرافیکی یک گلوگاه در بزرگراه

بینش جدید و قدرتمند در مورد ازدحام بزرگراه

در دسترس بودن منابع جدید کلان داده‌ها که بتوانند شرایط ترافیکی را در بزرگراه‌ها توضیح دهند، فرصت‌های جدید و غنی را برای توسعه و تحلیل سنجه‌های عملکردی بزرگراه‌ها ارائه می‌کنند. ما با استفاده از فقط یک منبع داده در مورد سرعت بزرگراه‌ها توانستیم دو سنجه‌ی جدید و متمایز را با استفاده از قابلیت‌های پیشرفته‌ی علم داده‌های ترادیتا معرفی کنیم.

اول، با تعریف و اندازه‌گیری کند شدن‌های دائمی ما به مهندسان ترافیک کمک کردیم تا تواتر و مدت زمان کند شدن‌ها در مکان‌های مختلف بزرگراه را درک کنند. تمایز بین اندازه‌گیری یک کند شدن دائمی و یک کند شدن موقت بسیار چالش‌برانگیز بوده و به علم داده‌ها نیاز دارد. همچنین توانایی مقایسه‌ی تعداد، مدت زمان و مکان وقوع کند شدن‌ها را به شکلی آموزنده‌تر و قانع‌کننده‌تر از سنجه‌های ساده‌ی میانگین، واریانس یا مقادیر پرت در مورد سرعت بزرگراه‌ها فراهم می‌کند.

سنجه‌ی دوم، توانایی اندازه‌گیری تلاطم و بهم‌ریختگی ناشی از گلوگاه‌ها بوده است. با شناسایی محل وقوع گلوگاه‌ها و محدود کردن آنها در نواحی بحرانی تأثیرگذارشان، می‌توانیم سرعت‌ها و آشفتگی کاهش شتاب ترافیک را در آن نواحی اندازه‌گیری کنیم. علم داده‌ها و قابلیت‌های تحلیلی نشان می‌دهند که زمانیکه وی‌اس‌اِل در یک ناحیه‌ی بحرانی از گلوگاه فعال باشد، آشفتگی و بهم‌ریختگی کاهش می‌یابد.

در این مقوله موضوعات بسیار زیادی را می‌تواند بررسی نمود. به عنوان مثال، طبیعی است که فرض کنیم چون بیشترین ترافیک در راه‌ها و جاده‌ها در ساعات شلوغی است، وی‌اس‌اِل بیشترین مزیت را در طی این دوره‌های زمانی پرازدحام ارائه می‌کند. با وجود این، عکس این مطلب ممکن است درست باشد که مزیت بسیار مهمی برای برنامه‌ی وی‌اس‌اِل می‌باشد.

اگرچه این پروژه‌ی کوچکی بود و فقط برای اثبات یک مفهوم انجام شد ولی ترکیبی از پروژه‌های مشابه که فراتر از بحث حمل و نقل هستند تحت عنوان "شهرهای هوشمند" در ایالات متحده و سایر کشورهای جهان در حال اجرا می‌باشند. هدف این است که از انواع مختلف داده‌هایی که از حسگرها یا چندرسانه‌ای، گزارشات رویدادهای نادر تا تصاویر ماهواره‌ای و تحلیل‌های پیشرفته‌ای که شامل یادگیری عمیق و رایانش شناختی هستند بتوان برای تبدیل ماهیت پویای شهرها به سمت بهترین شهرها برای تمامی ذینفعان استفاده کنیم.

Source: Teradata Case Study. "Georgia DOT Variable Speed Limit Analytics Help Solve Traffic Congestion." [https:// www.teradata.com/Resources/Case-Studies/Georgia-DOT-Variable-Speed-Limit-Analytics](https://www.teradata.com/Resources/Case-Studies/Georgia-DOT-Variable-Speed-Limit-Analytics) (accessed July 2018); "Georgia DOT Variable Speed Limits." [www.dot.ga.gov/ DriveSmart/SafetyOperation/Pages/VSL.aspx](http://www.dot.ga.gov/DriveSmart/SafetyOperation/Pages/VSL.aspx) (accessed August 2018). Used with permission from Teradata.

سوالاتی برای مورد کاربردی ۶.۵

۱. ماهیت مسائلی که وزارت ترابری جورجیا به دنبال حل آنها با علم داده‌ها بود چه هستند؟
۲. به نظر شما از چه نوع داده‌هایی برای تحلیل‌ها استفاده شد؟
۳. چه سنجه‌های علم داده‌هایی در این پروژه توسعه یافتند؟ این سنجه‌های دیگری به ذهن شما می‌رسند؟

در بخش بعدی، به بحث و بررسی در خصوص شکل بسیار محبوب و معروفی از معماری ام‌ال‌پی عمیق به نام شبکه عصبی همگشتی یا پیچشی که به طور خاص برای کاربردهای بینایی رایانه‌ای (یعنی بازشناخت تصویر، پردازش متون دست‌نویس) طراحی شده است می‌پردازیم.

سوالاتی برای مرور بخش ۶.۶

۱. منظور از "عمیق" در شبکه‌های عصبی عمیق چیست؟ شبکه‌های عصبی عمیق را با شبکه‌های عصبی کم عمق مقایسه کنید.
۲. منظور از واحد پردازنده گرافیکی چیست؟ چه ارتباطی با شبکه‌های عصبی عمیق دارد؟
۳. یک شبکه‌ی عمیق پرسپترون چندلایه پیشخور چگونه کار می‌کند؟
۴. در خصوص تأثیر اوزان تصادفی در توسعه ام‌ال‌پی عمیق توضیح دهید.
۵. کدام استراتژی بهتر است: تعداد لایه‌های بیشتر یا تعداد نورون‌های بیشتر؟

شبکه‌های عصبی همگشتی (سی‌ان‌ان) محبوب‌ترین روش‌های یادگیری عمیق هستند. سی‌ان‌ان‌ها در واقع شکل خاصی از معماری ام‌ال‌پی عمیق هستند که در ابتدا برای کاربردهای بینایی رایانه‌ای (یعنی پردازش تصویر، پردازش ویدئو، بازشناخت متن) طراحی شدند ولی بر روی مجموعه داده‌های غیرتصویری نیز کاربرد دارند. مشخصه‌ی اصلی شبکه‌های همگشتی این است که به جای تکثیر ماتریس عمومی، حداقل یک لایه دارند که یک تابع وزنی همگشتی را شامل می‌شود. شکل ۶.۲۱ یک واحد همگشتی معمول را نشان می‌دهد.

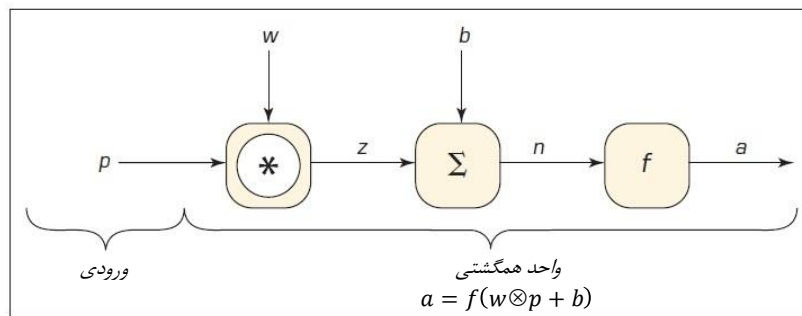
همگشت یا پیچش که معمولاً با نشانه‌ی * نمایش داده می‌شود یک عملیات خطی است که هدف از آن استخراج الگوهای ساده از الگوهای داده‌ای پیچیده است. به عنوان مثال، در پردازش یک تصویر که حاوی چندین شیء و رنگ است، توابع همگشت می‌توانند الگوهای ساده مثل وجود خطوط عمودی یا افقی یا لبه‌ها را در بخش‌های مختلف تشویر استخراج کنند. ما در بخش بعد به طور مفصل به بحث و بررسی در خصوص توابع همگشت خواهیم پرداخت.

در یک سی‌ان‌ان، یک لایه که حاوی تابع همگشت باشد را لایه‌ی همگشت (پیچش) می‌نامند. بعد از این لایه اغلب یک لایه‌ی ادغام^۱ یا (زیرنمونه‌برداری^۲) وجود دارد. لایه‌های ادغام مسئول ترکیب و ادغام تانسورهای بزرگ در قالب یک تانسور کوچکتر و کاهش تعداد پارامترهای مدل در عین حفظ ویژگی‌های مهم آنها هستند. انواع مختلف لایه‌های ادغام نیز در بخش‌های بعد مورد بررسی و بحث قرار خواهند گرفت.

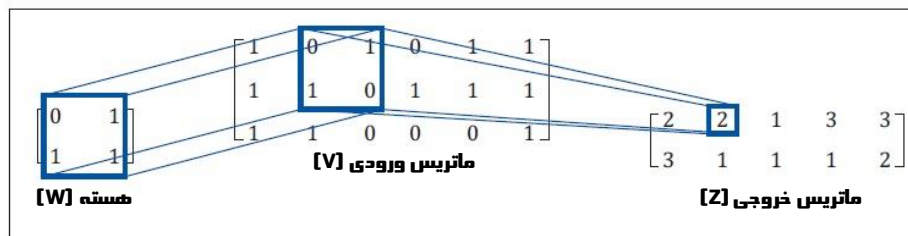
¹ Pooling

² Subsampling

در شرح شبکه‌های ام‌ال‌پی، گفته شد که تابع وزن معمولاً یک تابع دستکاری ماتریسی است که بردار وزن را در بردار ورودی ضرب کرده و در هر لایه یک بردار خروجی تولید می‌کند. با برخورداری از یک تانسور/بردار ورودی بسیار بزرگ که در اکثر کاربردهای یادگیری عمیق به همین شکل است ما به تعداد پارامترهای وزنی زیادی نیاز داریم تا هر ورودی به هر نورون را بتوان به یک پارامتر وزنی تخصیص داد. به عنوان مثال، هر ماتریس ورودی حاوی ۲۲،۵۰۰ عدد صحیح خواهد بود که هر کدام از آنها را باید به پارامتر وزنی خاص خودش در هر نورونی که در سرتاسر شبکه حرکت می‌کند تخصیص داد. از این رو، با داشتن حتی فقط یک لایه نیاز است تا ما هزاران پارامتر را تعریف و آموزش کنیم. همانگونه که ممکن است حدس زده باشید، این واقعیت، زمان و توان پردازشی مورد نیاز برای آموزش شبکه را به شدت افزایش خواهد داد زیرا در هر مرتبه تکرار آموزش، تمامی پارامترهای وزنی باید توسط الگوریتم اس‌جی‌دی به‌روزرسانی شوند. راه‌حل این مسئله، استفاده از تابع همگشت است.



شکل ۶.۲۱ یک واحد شبکه‌ی همگشتی



شکل ۶.۲۲ همگشت یک هسته‌ی ۲ * ۲ با یک ماتریس ورودی ۶ * ۲

$W_{r \times c}$

V

$$Z_{i,j} = \sum_{k=1}^r \sum_{l=1}^c w_{k,l} v_{i+k-1, j+l-1}$$

به عنوان مثال، فرض کنید ماتریس ورودی به یک لایه و هسته‌ی همگشت به شکل زیر باشد:

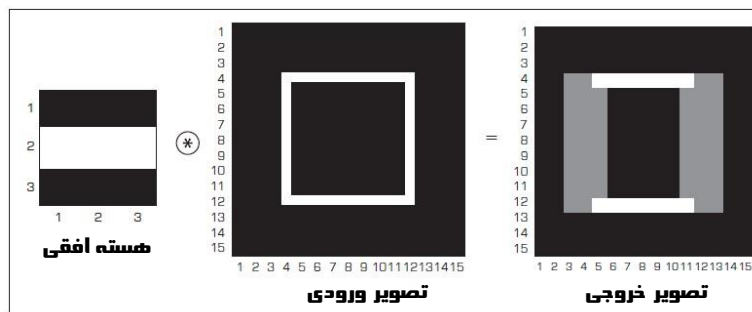
$$V = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad W = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

شکل ۶.۲۲ نشان می‌دهد که چگونه می‌توان خروجی همگشت را محاسبه کرد. همانطور که نشان داده شده است، هر عنصر ماتریس خروجی از جمع حاصلضرب یک به یک عناصر هسته در یک زیرمجموعه 2×2 از عناصر ماتریس ورودی بدست می‌آید (در این مثال، ماتریس 2×2 زیرا هسته نیز 2×2 است. بنابراین، در مثال نشان داده شده، عنصری که در ستون دوم ردیف اول ماتریس خروجی قرار دارد در واقع $0(0) + 1(1) + 1(1) + 1(0) = 2$

مشاهده می‌شود که بزرگی هر عنصر در ماتریس خروجی مستقیماً به هسته‌ی منطبق (با ماتریس 2×2) و ماتریس ورودی بستگی دارد. به عنوان مثال، عنصری که در ستون چهارم ردیف اول ماتریس خروجی قرار دارد، حاصل از همگشت هسته به واسطه‌ی بخشی از ماتریس ورودی است که دقیقاً با هسته یکسان است. این بیانگر این موضوع است که با بکارگیری عملیات همگشت، ما در واقع ماتریس ورودی را به یک ماتریس خروجی تبدیل می‌کنیم که در آن بخش‌هایی که دارای ویژگی خاصی هستند در جعبه‌ی مربع شکل قرار داده شده‌اند.

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

شکل ۶.۲۳ خروجی عملیات همگشت زمانی ماکزیمم می‌شود که هسته دقیقاً با بخشی از ماتریس ورودی که به واسطه‌ی آن همگشت پیدا می‌کند منطبق شود.



شکل ۶.۲۴ یک مثال از بکارگیری همگشت برای استخراج ویژگی‌ها از تصویر (خطوط افقی در این مثال)

این مشخصه‌ی توابع همگشت به ویژه در کاربردهای پردازش تصاویر مفید هستند. به عنوان مثال، اگر ماتریس ورودی بیانگر پیکسل‌های یک تصویر باشد، یک هسته‌ی ویژه که بیانگر یک شکل خاص است (مثلاً یک خط مورب) را می‌توان برای استخراج بخش‌هایی از تصویر که به شکل خاصی است همگشت نمود. شکل ۶.۲۴ به عنوان مثال نتیجه‌ی بکارگیری یک هسته‌ی خط افقی 3×3 بر روی یک تصویر مربع 15×15 را نشان می‌دهد.

مشخصاً، هسته‌ی افقی یک خروجی را تولید می‌کند که در آن محل خطوط افقی در تصویر ورودی اصلی شناسایی شده است.

همگشت با استفاده از یک هسته $2 \times C$ تعداد ردیف‌ها و ستون‌های ورودی را به اندازه‌ی $r-I$ و $C-I$ کاهش خواهد داد. در مورد اخیر، به عنوان مثال، ماتریس خروجی با استفاده از یک هسته‌ی 2×2 برای همگشت، یک ردیف و یک ستون کمتر از ماتریس ورودی دارد. برای پیشگیری از این تغییر اندازه، می‌توانیم بخش بیرونی ماتریس ورودی را پیش از همگشت با عدد صفر پر کنیم یعنی $r-I$ ردیف و $C-I$ ستون از عدد صفر را به ماتریس ورودی اضافه کنیم. از طرف دیگر، اگر بخواهیم ماتریس خروجی از این هم کوچکتر کنیم، می‌توانیم جابجایی‌های هسته را بزرگتر کنیم. طبیعتاً، هسته در زمان انجام همگشت، در هر واحد زمان یک گام جابجا می‌شود. با افزایش این مقدار به دو گام، اندازه‌ی ماتریس خروجی نیز به میزان کسری از ۲ کاهش می‌یابد.

اگرچه مزیت اصلی بکارگیری همگشت در شبکه‌های عمیق، به اشتراک‌گذاری پارامتر است که زمان و توان پردازشی مورد نیاز برای آموزش را با کاهش تعداد پارامترهای وزنی کاهش می‌دهد، مزیت‌های دیگری را نیز به همراه دارد. یک لایه‌ی همگشت در یک شبکه دارای یک ویژگی به نام هم‌وردایی برای کارهای ترجمه است (گودفلو و همکاران ۲۰۱۶). این بدان معناست که هر تغییری در ورودی منجر به تغییر در خروجی به همان شکل خواهد شد. به عنوان مثال، جابجایی یک شیء در تصویر ورودی به اندازه ۱۰ پیکسل در یک جهت مشخص منجر به جابجایی بازنمود آن در تصویر خروجی به اندازه ۱۰ پیکسل در همان جهت خواهد شد. گذشته از کاربردهای پردازش تصویر، این ویژگی بالاخص برای تحلیل داده‌های سری زمانی با استفاده از شبکه‌های همگشتی مفید است که همگشت در آنجا می‌تواند نوعی خط زمانی را تولید کند که نشان می‌دهد هر ویژگی چه زمانی در ورودی ظاهر می‌شود.

شایان ذکر است که تقریباً در تمامی کاربردهای شبکه‌های همگشتی، تعداد زیادی از عملیات همگشت به موازات برای استخراج انواع مختلف ویژگی‌ها از داده‌ها مورد استفاده قرار می‌گیرند زیرا یک ویژگی به تنهایی نمی‌تواند برای توضیح کامل ورودی‌ها برای طبقه‌بندی یا بازشناخت کافی باشد. همچنین همانطور که گفته شد، در اکثر کاربردهای دنیای واقعی، ما باید ورودی‌ها را به صورت تانسورهای چندبُعدی نمایش دهیم. به عنوان مثال، در پردازش تصاویر رنگی برخلاف عکس‌های سیاه و سفید، به جای تانسورهای ۲ بُعدی (یعنی ماتریس‌ها) که بیانگر رنگ پیکسل‌ها هستند (یعنی سیاه و سفید)، باید از تانسورهای ۳ بُعدی استفاده کنیم چون هر پیکسل باید با استفاده از شدت رنگ‌های قرمز، آبی و سبز تعریف شود.

اکثر مواقع، پس از یک لایه‌ی همگشت، لایه‌ی دیگری به نام لایه ادغام (یا زیرنمونه‌برداری) وجود دارد. هدف لایه‌ی ادغام این است که عناصر موجود در ماتریس ورودی را برای تولید یک ماتریس خروجی کوچکتر ادغام کرده و در عین حال ویژگی‌های مهم آن ماتریس را حفظ کند. طبیعتاً، یک تابع ادغام شامل یک پنجره‌ی یکی‌سازی $2 \times C$ (مشابه هسته در تابع همگشت) است که حول ماتریس ورودی جابجا شده و در هر جابجایی برخی از آمارهای عناصر موجود در پنجره‌ی یکی‌سازی را به گونه‌ای محاسبه می‌کند که بتوان آن را در تصویر خروجی قرار داد. به عنوان مثال، نوع خاصی از تابع ادغام به نام ادغام میانگین، میانگین عناصر ماتریس ورودی در پنجره‌ی یکی‌سازی را گرفته و آن مقدار میانگین را به عنوان عنصر ماتریس خروجی در مکان مربوطه قرار می‌دهد. به

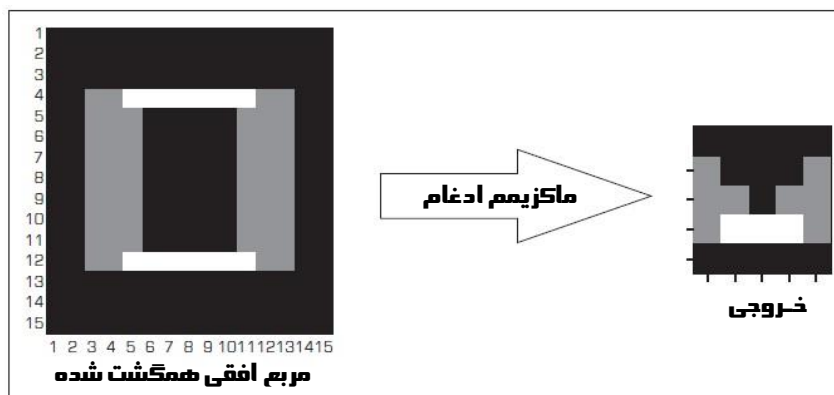
همین منوال، تابع ادغام ماکزیمم (ژو و همکاران) ماکزیمم مقادیر در پنجره را به عنوان عنصر خروجی در نظر می‌گیرد. برخلاف همگشت، برای تابع ادغام، با توجه به اندازه‌ی پنجره‌ی یکی‌سازی (یعنی r و c)، اندازه‌ی گام را باید به دقت انتخاب نمود تا هیچ‌گونه همپوشانی در یکی‌سازی‌ها بوجود نیاید. عملیات ادغام با استفاده از یک پنجره‌ی یکی‌سازی $r \times c$ موجب کاهش تعداد ردیف‌ها و ستون‌های ماتریس ورودی به صورت کسری از r و c می‌شود. به عنوان مثال، با استفاده از پنجره‌ی یکی‌سازی 3×3 ، یک ماتریس 15×15 ادغام شده و به یک ماتریس 5×5 تبدیل خواهد شد.

ادغام، علاوه بر کاهش تعداد پارامترها به ویژه برای کاربردهای پردازش تصویر یادگیری عمیق مفید هستند که وظیفه‌ی بحرانی در آنها تعیین این است که آیا یک ویژگی (مثلاً یک حیوان خاص) در تصویر وجود دارد یا خیر اگرچه موقعیت مکانی دقیق آن در تصویر اهمیتی ندارد. با وجود این، اگر مکان ویژگی‌ها در یک مقوله‌ی خاص اهمیت داشته باشد، استفاده از یک تابع ادغام می‌تواند به طور بالقوه گمراه کننده باشد.

ادغام را می‌توان به عنوان عملیاتی در نظر گرفت که ورودی‌های بزرگی که ویژگی‌های آنها قبلاً توسط لایه‌ی همگشت استخراج شده‌اند را خلاصه‌سازی می‌کند و فقط بخش‌های مهم در هر همسایگی کوچک در فضای ورودی را به ما نمایش می‌دهد. به عنوان مثال، در مورد مثال پردازش تصویر در شکل ۶.۲۴ اگر یک لایه‌ی ادغام ماکزیمم را با استفاده از یک پنجره‌ی یکی‌سازی پس از لایه‌ی همگشت قرار دهیم، خروجی چیزی شبیه آنچه در شکل ۶.۲۵ مشاهده می‌کنید خواهد بود. یک تصویر همگشت شده‌ی 15×15 به یک تصویر 5×5 تبدیل شده در حالیکه ویژگی‌های اصلی (خطوط افقی) ثابت باقی مانده‌اند.

برخی اوقات، از ادغام برای اصلاح اندازه ماتریس‌هایی که از لایه‌ی قبلی آمده و تبدیل آنها به اندازه خاص مورد نیاز لایه‌ی بعدی در شبکه استفاده می‌شود.

انواع مختلف عملیات ادغام از قبیل ادغام ماکزیمم، ادغام میانگین، نرم L_2 یک همسایگی مستطیلی^۱ و ادغام میانگین موزون وجود دارد. انتخاب عملیات ادغام مناسب و همچنین تصمیم در خصوص بکارگیری یک لایه‌ی ادغام در شبکه تماماً به ویژگی‌های مسئله‌ای که می‌خواهیم حل کنیم بستگی دارد. یک مجموعه اصول راهنما در ادبیات موضوع برای کمک به طراحان شبکه در چنین تصمیم‌گیری‌های وجود دارد (بورو و همکاران ۲۰۱۱، بورو، پونس و لوکان ۲۰۱۰، شرر، مولر و بنکه ۲۰۱۰).



شکل ۶.۲۵ یک مثال از بکارگیری ماکزیمم ادغام بر روی تصویر خروجی برای کاهش اندازه آن

¹ L2 Norm of a Rectangular Neighbourhood

کاربردهای حقیقی یادگیری عمیق به طور کلی و سی‌ان‌ان‌ها به طور خاص به میزان زیادی به در دسترس بودن مجموعه داده‌های بزرگ بستگی دارد. از لحاظ نظری، سی‌ان‌ان‌ها در بسیاری از مسائل کاربردی استفاده می‌شوند و امروزه پایگاه داده‌های بسیار بزرگ و غنی برای اینگونه کاربردها در دسترس می‌باشد. با وجود این، بزرگ‌ترین چالش این است که در کاربردهای یادگیری نظارت شده، برای آموزش مدل پیش از اینکه بتوانیم از آن برای پیش‌بینی/شناسایی سایر موارد نامعلوم استفاده کنیم به مجموعه داده‌هایی نیاز داریم که قبلاً برچسب‌گذاری شده باشند. اگرچه استخراج ویژگی‌های مجموعه داده‌ها با استفاده از لایه‌های سی‌ان‌ان یک فعالیت نظارت نشده است، ویژگی‌های استخراج شده بدون داشتن برچسب، چندان برای توسعه‌ی یک شبکه‌ی طبقه‌بندی به نحوی نظارت‌شده فایده نخواهد داشت. به همین دلیل است که شبکه‌های طبقه‌بندی و دسته‌بندی تصویر شامل دو مسیر اصلی هستند: استخراج ویژگی دیداری و دسته‌بندی تصویر.

ایمیج‌نت^۱ (<http://www.image-net.org>) یک پروژه‌ی پژوهشی مستمر است که پایگاه‌های داده بزرگی از تصاویر را در اختیار پژوهشگران قرار می‌دهد که هر تصویر به مجموعه‌ای از واژگان مترادف از وردنت^۲ (یک پایگاه داده سلسله‌مراتب واژگان) متصل است. هر مجموعه واژگان مترادف بیانگر مفهومی خاص در وردنت است. در حال حاضر، وردنت شامل بیش از ۱۰۰,۰۰۰ مجموعه واژگان مترادف است که هر یک به طور میانگین با ۱,۰۰۰ تصویر در ایمیج‌نت نمایش داده می‌شود. ایمیج‌نت یک پایگاه داده‌ی عظیم برای توسعه‌ی شبکه‌های عمیق برای پردازش تصویر است که حاوی بیش از ۱۵۰ میلیون تصویر برچسب‌گذاری شده در ۲۲,۰۰۰ دسته است. ایمیج‌نت به دلیل اندازه و دسته‌بندی مناسب، تاکنون پرکاربردترین مجموعه داده‌ی الگوبرداری برای ارزیابی کارایی و دقت شبکه‌های عمیق که توسط پژوهشگران یادگیری عمیق طراحی شده‌اند می‌باشد.

یکی از اولین شبکه‌های همگشتی طراحی شده برای دسته‌بندی تصاویر با استفاده از مجموعه داده‌های ایمیج‌نت، آکسنت^۳ نام داشت (کریژنسکی، سوتس‌کُور و هینتون ۲۰۱۲). این شبکه از پنج لایه‌ی همگشت و سه لایه‌های کاملاً متصل به آن تشکیل شده بود (نمای شماتیک از آکسنت در شکل ۶.۲۶). یکی از نقش‌های این معماری نسبتاً ساده که آموزش آن را بسیار سریع‌تر و مقرون به صرفه‌تر نمود، استفاده از توابع انتقال واحد یکسوشده خطی^۴ در لایه‌های همگشت به جای توابع سیگموئید سنتی بود. با اینکار طراحان به مسئله‌ی معروف به مشکل محوشدگی گرادیان^۵ که ناشی از مشتقات بسیار کوچک توابع سیگموئید در برخی نواحی تصویر بود رسیدگی کردند. یک نقش مهم دیگر این شبکه، معرفی مفهوم لایه‌های دراپ‌آوت در سی‌ان‌ان‌ها به عنوان یک تکنیک منظم‌سازی برای کاهش بیش‌برازش بود. یک لایه‌ی دراپ‌آوت معمولاً پس از لایه‌های کاملاً متصل آمده و برای غیرفعال کردن برهی از نورون‌ها یک احتمال تصادفی را بر روی آنها اعمال می‌کند.

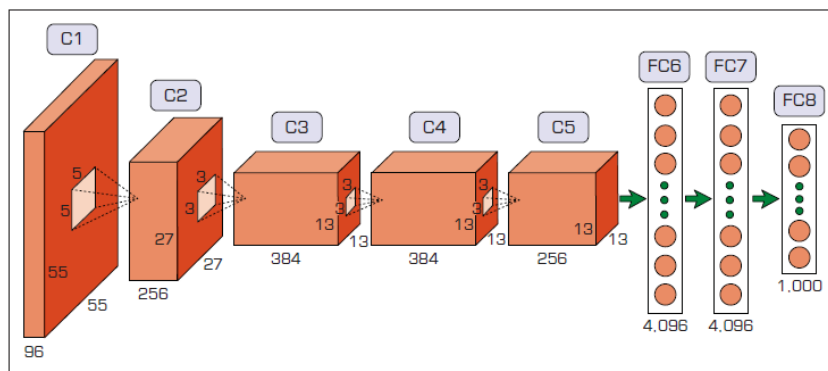
¹ ImageNet

² WordNet

³ AlexNet

⁴ Rectified linear Unit (ReLU)

⁵ Vanishing gradient problem



شکل ۶.۲۶ معماری یک آکسنت یک شبکه همگشتی برای دسته‌بندی تصویر

در سال‌های اخیر، علاوه بر تعداد کثیری از دانشمندان که به نمایش قابلیت‌های یادگیری عمیق پرداخته‌اند برخی از شرکت‌های بسیار معروف تراز اول صنعت از قبیل مایکروسافت، گوگل و فیسبوک نیز در چالش سالانه‌ی بازساخت دیداری ایمِج‌نت در مقیاس بزرگ^۱ مشارکت کرده‌اند. هدف در اینجا، طراحی و آموزش شبکه‌هایی است که قادر به دسته‌بندی ۱.۲ تصویر ورودی در ۱,۰۰۰ دسته باشند. به عنوان مثال، گوگل‌نت^۲ یک معماری شبکه همگشتی عمیق است که توسط پژوهشگران گوگل طراحی شده و در سال ۲۰۱۴ به عنوان معماری برتر با یک شبکه‌ی ۲۲ لایه و نرخ خطای دسته‌بندی فقط ۶.۶۶ درصد که فقط ۱۵.۱٪ بدتر از خطای دسته‌بندی انسانی است برگزیده شد (روساکوفسکی و همکاران ۲۰۱۵). نقش اصلی معماری گوگل‌نت، معرفی ماژولی به نام *اینسپشن*^۳ بود. ایده‌ی اینسپشن این است که چون هیچ نظری در مورد اندازه‌ی هسته‌ی همگشت که بهترین نتیجه را بر روی یک مجموعه داده بدهد نداریم بهتر است که از چندین همگشت استفاده کرده و اجازه دهیم تا شبکه در مورد آن تصمیم بگیرد. از این رو، همانطور که در شکل ۶.۲۷ نشان داده شده است، در هر لایه‌ی همگشت، داده‌هایی که از لایه‌ی قبلی می‌آیند از بین چندین نوع همگشت عبور کرده و پیش از ورود به لایه‌ی بعدی، بهم الحاق می‌شوند. این معماری به مدل اجازه می‌دهد تا هم ویژگی‌های موضعی را از طریق همگشت‌های کوچکتر و ویژگی‌های انتزاعی را از طریق همگشت‌های بزرگتر مد نظر قرار دهند.

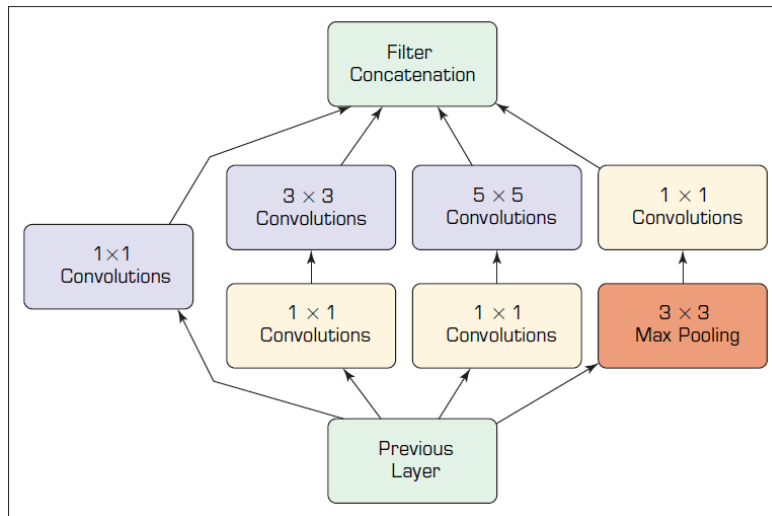
گوگل اخیراً خدمت جدیدی را به نام گوگل لنز^۴ راه‌اندازی نمود که از الگوریتم‌های شبکه عصبی مصنوعی یادگیری عمیق (به همراه سایر تکنیک‌های هوش مصنوعی) برای ارائه‌ی اطلاعات در مورد تصاویری که کاربران از اشیاء مجاور خود گرفته‌اند استفاده می‌کند. این شامل شناسایی اشیاء، محصولات، گیاهان، جانوران و مکان‌ها و ارائه‌ی اطلاعات در مورد آنها بر روی اینترنت است. برخی از سایر ویژگی‌های این سرویس عبارتند از قابلیت ذخیره‌سازی اطلاعات تماس از روی تصویر یک کارت ویزیت بر روی تلفن، شناسایی انواع گیاهان و گونه‌های جانوران، شناسایی کتاب‌ها و فیلم‌ها از روی عکس روی جلد آنها و ارائه‌ی اطلاعات (مثلاً فروشگاه‌ها، تئاترها، خرید، رزرواسیون) در مورد آنها. شکل ۶.۲۸ دو مثال از بکارگیری اپلیکیشن گوگل لنز را بر روی دستگاه موبایل اندروید نشان می‌دهد.

¹ ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

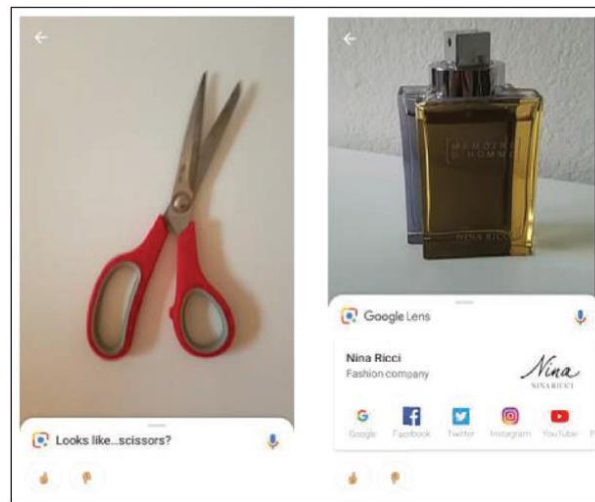
² GoogLeNet

³ Inception

⁴ Google Lens



شکل ۶.۲۷ نمایش مفهومی ویژگی اینسپشن در GoLeNet



شکل ۶.۲۸ دو مثال از بکارگیری گوگل لنز، یک سرویس مبتنی بر شبکه‌های عمیق همگشتی برای بازشناخت تصویر

اگرچه بعدها شبکه‌های دقیق‌تری از لحاظ کارایی و نیازمندی‌های پردازش (یعنی تعداد لایه‌ها و پارامترهای کمتر) توسعه یافتند، گوگل‌نت همچنان یکی از بهترین معماری‌ها تا به امروز به شمار می‌رود. گذشته از آکس‌نت و گوگل‌نت، چندین معماری شبکه‌ی همگشتی دیگر از قبیل VGGNet، Residual Networks (ResNet) و Xception نیز برای پردازش تصاویر توسعه یافته و بکار گرفته شده‌اند که تمامی آنها متکی بر پایگاه داده‌ی ایمیج‌نت هستند.

در ماه می ۲۰۱۸ فیسبوک در تلاش برای فعالیت بسیار سنگین برچسب‌گذاری تصاویر در مقیاس بزرگ یک پروژه‌ی یادگیری عمیق بازشناخت تصویر با آموزش نیمه نظارت شده را راه‌اندازی نمود (ماهاجان و همکاران ۲۰۱۸). در این پروژه از هشتگ‌هایی که کاربران بر روی تصاویر ارسالی بر روی اینستاگرام ایجاد کرده بودند به عنوان برچسب استفاده شده و یک مدل بازشناخت تصویر

یادگیری عمیق را بر آن اساس آموزش دادند. این مدل با استفاده از ۳.۵ میلیارد تصویر اینستاگرامی برچسب‌گذاری شده با حدود ۱۷,۰۰۰ هشتگ با استفاده از ۳۳۶ واحد پردازنده گرافیکی که به موازات هم کار می‌کردند آموزش داده شد. فرایند آموزش چندین هفته به طول انجامید. یک نسخه‌ی اولیه از مدل (فقط با ۱ میلیارد تصویر و ۱,۵۰۰ هشتگ آموزش داده شد) بر روی مجموعه داده‌های ایمِج‌نت تست شده و طبق گزارشات تهیه شده، عملکرد آن از لحاظ دقت به میزان بیش از ۲ درصد بهتر از سایر مدل‌های پیشرفته و جدید بود. این دستاورد بزرگ فیسبوک مطمئناً درهایی را به دنیای جدید پردازش تصاویر با استفاده از یادگیری عمیق باز کرد زیرا می‌تواند اندازه مجموعه داده‌های تصاویر موجود را که برای مقاصد آموزشی برچسب‌گذاری شده‌اند به شدت افزایش دهد.

استفاده از روش‌های یادگیری عمیق و تحلیل پیشرفته برای دسته‌بندی تصاویر تا سطح تشخیص چهره‌ی انسان تکامل یافته و به یک کاربرد بسیار متداول برای مقاصد مختلف تبدیل شده است که در مورد کاربردی ۶.۶ بحث شده است.

از تشخیص تصویر تا تشخیص چهره

مورد کاربردی ۶.۶

تشخیص یا بازشناخت چهره اگرچه مشابه بازشناخت تصویر به نظر می‌رسد ولی بسیار پیچیده‌تر از آن است. هدف از تشخیص چهره، شناسایی افراد با توجه به دسته یا گروهی که به آن تعلق دارند (انسان) است و این شناسایی می‌بایست بر روی یک محیط سه بُعدی غیرایستا (یعنی شخص در حال حرکت) انجام شود. تشخیص چهره برای چندین دهه یک رشته‌ی پژوهشی فعال در هوش مصنوعی بوده که موفقیت‌چندانی تا این اواخر به دست نیاورده است. به لطف نسل جدید الگوریتم‌ها (یعنی یادگیری عمیق) به همراه مجموعه داده‌های بزرگ و توان محاسباتی، فناوری تشخیص چهره تأثیر مهمی بر کاربردهای دنیای واقعی گذاشته است. از امنیت تا بازاریابی، تشخیص چهره و کاربردهای متنوع این فناوری با سرعتی اعجاب‌آور در حال افزایش است.

برخی از مثال‌های بارز تشخیص چهره (هم در دستاوردهای فناوری و هم در استفاده‌ی خلاقانه از منظر فناوری) از چین آمده‌اند. امروزه در چین، تشخیص چهره یک موضوع بسیار داغ هم برای توسعه‌ی کسب‌وکار و هم منظرهای توسعه برنامه‌های کاربردی است. تشخیص چهره به یک اکوسیستم پربار با صدها شرکت نوپا در چین بدل شده است. در محیط‌های شخصی و/یا تجاری، مردم چین به طور گسترده‌ای از دستگاه‌هایی استفاده می‌کنند که امنیت آنها مبتنی بر تشخیص خودکار چهره‌هایشان است.

دولت چین اخیراً پروژه‌ای را تحت عنوان "چشمان تیزبین"^۱ که شاید بتوان آن را بزرگ‌ترین مورد کاربردی یادگیری عمیق و تشخیص چهره در جهان دانست با هدف استقرار یک سیستم نظارت مبتنی بر تشخیص چهره در سطح کشور آغاز کرده است. هدف این پروژه، یکپارچه‌سازی دوربین‌های امنیتی که قبلاً در مکان‌های عمومی نصب شده‌اند با دوربین‌های شخصی که در ساختمان‌ها نصب شده‌اند و بهره‌برداری از یادگیری عمیق برای تحلیل ویدئوهای بدست‌آمده از آن دوربین‌ها می‌باشد. با وجود میلیون‌ها دوربین و میلیارد‌ها خط کد برنامه‌نویسی، چین در حال ساخت یک آینده‌ی مقتدرانه با فناوری بالاست. با این سیستم، دوربین‌های موجود در برخی شهرها می‌توانند به رصد ایستگاه‌های اتوبوس و قطار و همچنین فرودگاه‌ها بپردازند تا بتوانند خطرناک‌ترین مجرمان چین را شناسایی و دستگیر کنند. نمایشگرهای بزرگ می‌توانند تصویر چهره‌ی عابرین پیاده‌ای که از

¹ Sharp Eyes

محل‌های غیرمجاز عبور می‌کنند را نمایش داده و نام و تصویر افرادی که دیون خود را پرداخت نمی‌کنند را منتشر کنند. از اسکرین‌های تشخیص چهره در ورودی‌های مجتمع‌های مسکونی نیز استفاده می‌شود.

یک مثال جالب از این سیستم نظارتی، "بازی شرم" نام دارد (موزور ۲۰۱۸). یک تقاطع در جنوب پل چانگ‌هونگ در شهر ژیانگ‌یانگ تا پیش از این یک کابوس به شمار می‌رفت. خودروها با سرعت حرکت می‌کردند و عابرین پیاده از مکان‌های غیرمجاز وارد خیابان می‌شدند. سپس در تابستان سال ۲۰۱۷، پلیس دوربین‌های متصل به فناوری تشخیص چهره و یک صفحه نمایش بزرگ در آنجا نصب نمود. تصاویر قانون‌شکنان به همراه نام و شماره شناسایی دولتی آنها نمایش داده می‌شد. مردم در ابتدا از دیدن چهره‌شان بر روی صفحه نمایش هیجان زده می‌شدند تا وقتی که نشریات دولتی به آنها گفتند که این نوعی مجازات و تنبیه محسوب می‌شود. با استفاده از این، شهروندان نه تنها سوژه‌ی این بازی شرم‌آور شدند بلکه امتیاز شهروندی منفی نیز برایشان لحاظ می‌شد. برعکس، از منظر مثبت قضیه، اگر تصویر مردم در حالیکه رفتار مناسبی از خود نشان می‌دادند بر روی صفحه نمایش پخش می‌شد، مثل برداشتن یک تکه زباله از خیابان و انداختن آن در سطح زباله یا کمک به شخصی سالمند برای عبور از خیابان، آنها امتیاز شهروندی مثبتی دریافت می‌کردند که برای جوایز کوچک قابل استفاده بود.

چین در حال حاضر ۲۰۰ میلیون دوربین نظارتی نصب شده دارد که چهار برابر تعداد دوربین‌های ایالات متحده است. این سیستم عمدتاً برای ردیابی مظنونین، شناسایی رفتارهای مشکوک و پیش‌بینی جرایم بکار می‌رود. به عنوان مثال، برای یافتن یک مجرم، تصویر یک مظنون را می‌توان بر روی سیستم بارگذاری کرد، آن را با میلیون تصویر چهره که از ویدئوهای میلیون‌ها دوربین امنیتی در سرتاسر کشور بدست آمده است تطبیق داد. بدین ترتیب می‌توان افرادی با میزان تشابه بسیار بالا را شناسایی کرد. این سیستم همچنین با یک پایگاه داده‌ی عظیم از اطلاعات پرونده‌ها و سوابق پزشکی، رزروهای مسافرتی، خریدهای آنلاین و حتی فعالیت‌های رسانه‌های اجتماعی هر شهروند یکپارچه بوده و می‌تواند هر فردی را در کشور پایش کرده (۱.۴ میلیارد نفر)، و ردیابی کنند که شخص کجا قرار دارد و در حال انجام چه کاری است (دنیپر ۲۰۱۸). دولت انتظار دارد تا با کمک پروژه‌ی چشمان تیزبین نهایتاً بتواند یک "امتیاز اعتبار اجتماعی" را به هر فردی در کشور اختصاص دهد تا میزان قابل اعتماد بودن شخص مشخص شود.

اگرچه چنین کاربرد بدون محدودیتی از یادگیری عمیق (یعنی جاسوسی شهروندان) برخلاف هنجارها و قوانین اخلاقی و حریم خصوصی بسیاری از کشورهای غربی از جمله ایالات متحده است، اما به روشی متداول در کشورهایی مثل چین که قوانین و ملاحظات حریم خصوصی در آنها ضعیف‌تر است تبدیل شده است. حتی کشورهای غربی شروع به برنامه‌ریزی برای بکارگیری فناوری‌های مشابه در مقیاس‌های کوچک فقط برای مقاصد امنیتی و پیشگیری از جرائم کرده‌اند. سیستم شناسایی نسل بعدی اِف‌بی‌آی به عنوان مثال یک کاربرد قانونمند از فناوری‌های تشخیص چهره و یادگیری عمیق است که برای شناسایی مظنونین بالقوه، تصاویر بدست آمده از صحنه‌های جرم را با پایگاه‌های داده ملی تصاویر مجرمان تطبیق می‌دهد.

Sources: Mozur, P. (2018, June 8). "Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras." *The New York Times*. <https://www.nytimes.com/2018/07/08/business/chinasurveillance-technology.html>; Denyer, S. (2018, January). "Beijing Bets on Facial Recognition in a Big Drive for Total Surveillance." *The Washington Post*. https://www.washingtonpost.com/news/world/wp/2018/01/07/feature/inchina-facial-recognition-is-sharp-end-of-a-drive-for-totalsurveillance/?noredirect=on&utm_term=.e73091681b31

سوالاتی برای مورد کاربردی ۶.۶

۱. چالش‌های فنی در تشخیص چهره چه هستند؟
۲. به غیر از بحث‌های امنیت و نظارت از فناوری تشخیص چهره برای چه مقاصد دیگری می‌توان استفاده کرد؟

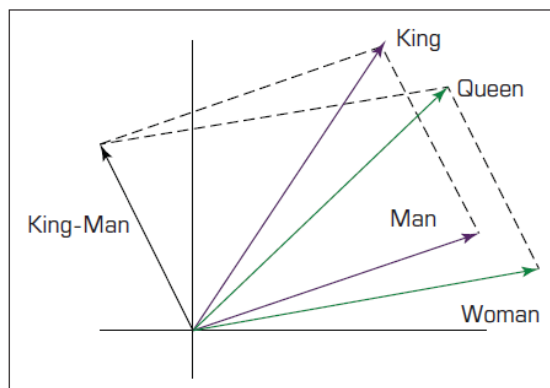
علاوه بر پردازش تصویر که در واقع دلیل اصلی محبوبیت و توسعه‌ی شبکه‌های همگشتی بوده است، در برخی وظایف متن‌کاوی گسترده نیز مفید بوده‌اند. به ویژه از سال ۲۰۱۳ و زمانیکه گوگل پروژه‌ی **word2vec** خود را منتشر کرد (میکولوف و همکاران ۲۰۱۳؛ میکولوف، سوتسکور، چن، کورادو و دین ۲۰۱۳)، کاربردهای یادگیری عمیق برای متن‌کاوی به طور چشمگیری افزایش یافته است. Word2vec یک شبکه عصبی دولایه است که متن‌های بزرگ را به عنوان ورودی دریافت کرده و هر واژه در متن را به یک بردار عددی با هر اندازه‌ی مورد نظر (معمولاً بین ۱۰۰ تا ۱۰۰۰) و با ویژگی‌های بسیار جذاب تبدیل می‌کند. اگرچه word2vec به خودی خود یک الگوریتم یادگیری عمیق نیست، خروجی آن (بردارهای واژگان) به طور گسترده‌ای در بسیاری از پروژه‌های تجاری و پژوهشی یادگیری عمیق بکار گرفته شده‌اند. یکی از جالب‌ترین ویژگی‌های بردارهای واژگان که توسط الگوریتم word2vec ایجاد شده، حفظ و نگهداری وابستگی‌های نسبی واژگان است. به عنوان مثال، عملیات بردار

$$\text{vector}('King') - \text{vector}('Man') + \text{vector}('woman')$$

و

$$\text{vector}('London') - \text{vector}('England') + \text{vector}('France')$$

منجر به یک بردار بسیار نزدیک به $\text{vector}('Queen')$ و $\text{vector}('Paris')$ می‌شود. شکل ۶.۲۹ یک بردار ساده از مثال اول را در یک فضای برداری دو بُعدی نشان می‌دهد. علاوه بر این، بردارها به گونه‌ای مشخص می‌شوند که آنهایی که دارای بافتار مشابه هستند در یک فضای برداری n بُعدی بسیار نزدیک به یکدیگر قرار داده می‌شوند. به عنوان مثال، در مدل word2vec با استفاده از یک متن با حدود ۱۰۰ میلیارد واژه، نزدیک‌ترین بردارها به $\text{vector}('Sweden')$ بر حسب فاصله‌ی کسینوسی در جدول ۶.۲ نشان داده شده است که به شناسایی نام کشورهای اروپایی نزدیک به منطقه‌ی اسکاندیناوی یعنی همان منطقه‌ای که سوئد در آنجا واقع شده است می‌پردازد. علاوه بر این، از آنجائیکه word2vec بافتارهایی که یک واژه در آن بکار رفته و تواتر استفاده از آن در هر بافتار را برای حدس معنای آن واژه در نظر می‌گیرد، ما را قادر می‌سازد تا به جای فقط عبارت سمبولیک/نحوی آن، بافتار معنایی آن را نمایش دهیم. در نتیجه، word2vec به چندین مسئله‌ی تغییر واژگان می‌پردازد که تا در فعالیت‌های متن‌کاوی سنتی بسیار مشکل‌ساز بودند. به عبارت دیگر، word2vec قادر است تا واژگان از جمله حروف چاپی، اختصارات و مکالمات غیررسمی را به درستی شناسایی و بیان کند. به عنوان مثال، واژگان *Frnce*, *France* و *Frans* همانند واژه‌ی درست و اصلی *France* بردار واژگان یکسانی را می‌گیرند. بردارهای واژگان همچنین می‌توانند انواع دیگری از وابستگی‌ها از قبیل تمایز موجودیت‌ها (مثلاً $\text{vector}['human'] - \text{vector}['ethics'] \sim \text{vector}['animal'] - \text{vector}['Iraq'] - \text{vector}['violence']$) یا وابستگی‌های ژئوپولتیکی (مثلاً $\text{vector}['Jordan']$) را تعیین کنند.



شکل ۶.۲۹ نمایش برداری معمول از تعبیه کلمات در یک فضای دو بُعدی

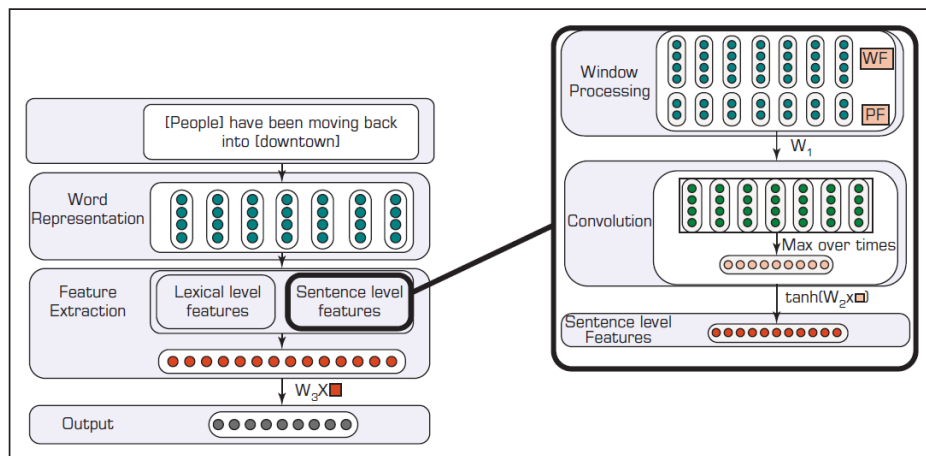
با ارائه‌ی چنین نمایش معناداری از داده‌های متنی، در سال‌های اخیر word2vec توانسته بسیاری از پروژه‌های متن‌کاوی مبتنی بر یادگیری عمیق را در حوزه‌های مختلف به پیش برد (مثلاً پزشکی، علوم رایانه، رسانه‌های اجتماعی، بازاریابی) و انواع مختلف شبکه‌های عمیق برای تعبیه‌های واژگان که با این الگوریتم خلق شده‌اند برای اهداف مختلف بکار رفته است. به طور خاص، گروه بزرگی از مطالعات، از شبکه‌های همگشتی بر روی تعبیه‌های واژگان و با هدف استخراج نسبت^۱ از مجموعه داده‌های متنی استفاده کرده‌اند. استخراج نسبت، یکی از زیروظایف پردازش زبان طبیعی است که بر تعیین اینکه آیا دو یا چند موجودیت نام‌گذاری شده که در یک متن شناسایی شده‌اند دارای ارتباطات خاصی هستند یا خیر تمرکز دارد (مثلاً "A منجر به B می‌شود"، "B در اثر A بوجود می‌آید"). برای مثال، زنگ و همکاران (۲۰۱۴) یک شبکه همگشتی عمیق را برای دسته‌بندی روابط بین موجودیت‌های مشخص شده در جملات ایجاد کردند.

جدول ۶.۲ یک مثال از پروژه‌ی word2vec که نزدیک‌ترین

بردارهای واژگان به واژه‌ی "سوئد" را نشان می‌دهد

فاصله‌ی کسینوسی	واژه
۰.۷۶۰۱۲۴	نروژ
۰.۷۱۵۴۶۰	دانمارک
۰.۶۲۰۰۲۲	فنلاند
۰.۵۸۸۱۳۲	سوئیس
۰.۵۸۵۶۳۵	بلژیک
۰.۵۷۴۶۳۱	هلند
۰.۵۶۲۳۶۸	ایسلند
۰.۵۴۷۶۲۱	استونی
۰.۵۳۱۴۰۸	اسلوونی

¹ Relation extraction



شکل ۶.۳۰ معماری سی‌ان‌ان برای استخراج نسبت در متن کاوی

تا به امروز، پژوهشگران از یک فرمت ماتریسی برای نمایش هر جمله استفاده کرده‌اند. هر ستون برای ماتریس‌های ورودی در واقع یک تعبیه واژه (بردار) مرتبط با یکی از واژگان جمله است. زنگ و همکاران از یک شبکه‌ی همگشتی که در کادر سمت راست شکل ۶.۳۰ نشان داده شده است برای یادگیری خودکار ویژگی‌های سطح جمله استفاده کرده و ویژگی‌هایی که برخی ویژگی‌های لغوی پایه دارند (مثلاً ترتیب دو واژه مورد نظر در یک جمله و توکن‌های چپ و راست برای هر یک از آنها) را الحاق کرده‌اند. بردار ویژگی الحاق شده سپس توسط یک تابع انتقال *softmax* وارد لایه‌ی طبقه‌بندی می‌شود که نوع ارتباطات بین دو واژه‌ی مورد نظر را بین انواع از پیش تعیین شده مشخص می‌کند. تابع انتقال *softmax* پرکاربردترین تابع برای لایه‌های طبقه‌بندی است به ویژه زمانیکه تعداد دسته‌ها بیش از دو عدد باشد. برای مسائل طبقه‌بندی با دو دسته یا گروه نتیجه، از توابع انتقال لاگ-سیگموئید نیز استفاده می‌شود. روش پیشنهادی زنگ و همکاران به درستی روابط بین عبارات مشخص شده در جمله‌ی مجموعه داده نمونه را با ۸۲.۷ درصد دقت دسته‌بندی کرده است.

در مطالعه‌ای مشابه، نگوین و گریشمان (۲۰۱۵) از یک شبکه همگشتی چهارلایه با اندازه هسته‌های مختلف در هر لایه همگشتی برای دسته‌بندی نوع ارتباطات بین دو واژه‌ی مشخص شده در هر جمله استفاده کردند. در ماتریس ورودی، هر ردیف یک تعبیه‌ی واژه‌ی مرتبط با یک واژه در توالی یکسانی از جمله بود. علاوه بر این، این پژوهشگران دو ستون دیگر را به عنوان ماتریس‌های ورودی برای نمایش موقعیت نسبی هر واژه با توجه هر کدام از عبارات مشخص شده اضافه کرده‌اند. ویژگی‌های استخراج شده به صورت خودکار از یک لایه‌ی طبقه‌بندی با تابع *softmax* عبور داده می‌شود تا نوع ارتباطات تعیین شود. نگوین و گریشمان مدل خود را با استفاده از ۸,۰۰۰ مثال آموزش داده (۱۹ دسته یا گروه از ارتباطات از پیش تعریف شده) و مدل آموزش داده شده را بر روی مجموعه‌ای از ۲,۷۱۷ مجموعه داده‌های اعتبارسنجی تست کرده و به یک دقت طبقه‌بندی ۶۱.۳۲ درصدی دست یافتند (یعنی ۱۱ برابر بهتر از حدس زدن).

اینگونه رویکردهای متن کاوی با استفاده از شبکه‌های عمیق همگشتی را می‌توان به حوزه‌های کاربردی مختلفی گسترش داد. یک چالش بزرگ در اینجا همانند پردازش تصویر، فقدان مجموعه داده‌های بزرگ و کافی برای آموزش نظارت شده شبکه‌های عمیق است. یک روش نظارت از راه دور برای آموزش توسط (مینتز و همکاران ۲۰۰۹) برای رفع این چالش پیشنهاد شده است. در این روش، حجم زیادی از داده‌های آموزش را می‌توان با همراستا کردن اطلاعات مبنای دانش با متن‌ها تولید کرد. در واقع، این روش مبتنی بر این فرض است که رابطه‌ی خاصی بین یک زوج موجودیت در پایگاه دانشی وجود داشته باشد آنگاه هر سند متنی که این

زوج موجودیت در آن اشاره شده باشد بیانگر این رابطه خواهد بود. با وجود این، با توجه به اینکه این فرض واقع‌بینانه نیست، ریدل، یائو و مک‌کالوم (۲۰۱۰) با مدل سازی مسئله به صورت یک مسئله‌ی یادگیری چندمنونه‌ای آن را واقع‌بینانه کردند. آنها توصیه کردند که برچسب‌هایی به جای یک نمونه به یک بسته از نمونه‌ها تخصیص داده شود تا نویز روش نظارت از راه دور را کاهش داده و مجموعه داده‌های آموزش برچسب‌گذاری شده و واقع‌بینانه‌تری را ایجاد کنند (کومار ۲۰۱۷).

سوالاتی برای مرور بخش ۶.۷

۶. سی‌ان‌ان چیست؟
۷. از سی‌ان‌ان برای چه نوع کاربردهایی می‌توان استفاده کرد؟
۸. تابع همگشت در سی‌ان‌ان چیست و چگونه کار می‌کند؟
۹. ادغام در سی‌ان‌ان چیست و چگونه کار می‌کند؟
۱۰. ایمپجنت چیست و چه ارتباطی با یادگیری عمیق دارد؟
۱۱. اهمیت آکسنت چیست؟ معماری آن را ترسیم کرده و توضیح دهید.
۱۲. گوگل‌نت چیست؟ چگونه کار می‌کند؟
۱۳. تعبیه واژگان چه هستند و چگونه کار می‌کنند؟

تفکر و درک انسان به میزان زیادی به بافتار^۱ متکی است. به عنوان مثال برای ما خیلی مهم است که بدانیم یک سخنور خاص از یک زبان بسیار طعنه‌آمیز (بر اساس سخنرانی‌های قبلی وی) استفاده می‌کند تا بتوانیم معنی تمامی لطیفه‌های او را متوجه شویم. یا درک معنای حقیقی واژه‌ی *fall* (یعنی فصل یا فرو ریختن) در جمله‌ی "*It is a nice day of fall*" بدون دانش و آگاهی در مورد سایر واژگان در جملات مرتبط فقط یک حدس و گمان خواهد بود نه لزوماً درک معنای آن کلمه. دانش در خصوص بافتار معمولاً بر اساس مشاهده‌ی رویدادهایی که در گذشته رخ داده‌اند شکل می‌گیرد. در واقع، افکار انسانی، مستمر و دائم هستند و ما از هرگونه اطلاعاتی که قبلاً در مورد یک رویداد از فرایند تحلیل بدست آورده‌ایم استفاده می‌کنیم به جای اینکه دانش گذشته را به دور بیندازیم و بخواهیم همه چیز را هر زمان که با رویدادها یا موقعیت‌های مشابهی مواجه می‌شویم از ابتدا بسازیم. از این رو، در نحوه‌ی پردازش اطلاعات توسط انسان‌ها، تکرار وجود دارد.

اگرچه شبکه‌های ام‌ال‌پی عمیق و همگشتی به طور تخصصی برای پردازش یک شبکه‌ی ایستا^۲ از مقادیری مثل یک تصویر یا یک ماتریس از تعبیه واژگان استفاده می‌شود برخی اوقات یک توالی از مقادیر ورودی نیز برای عملیاتی‌سازی شبکه و انجام یک وظیفه‌ی مشخص بسیار مهم بوده و باید مد نظر قرار گیرند. یک نوع متداول دیگر از شبکه‌های عصبی، شبکه‌های عصبی بازگشتی هستند (روملهارت و همکاران ۱۹۸۶) که مشخصاً برای پردازش ورودی‌های متوالی طراحی شده‌اند. یک شبکه عصبی بازگشتی اصولاً به مدل‌سازی یک سیستم پویا می‌پردازد که (حداقل یکی از نورون‌های پنهان آن) وضعیت سیستم (یعنی خروجی یک نورون پنهان) در هر نقطه زمانی t به ورودی‌های سیستم در آن زمان و وضعیت آن در نقطه زمانی قبلی $t - 1$ بستگی دارد. به عبارت دیگر، شبکه‌های عصبی بازگشتی نوعی از شبکه‌های عصبی هستند که دارای حافظه بوده و از آن حافظه برای تعیین خروجی‌های آینده

¹ Context

² Static

آن استفاده می‌کنند. به عنوان مثال، در طراحی یک شبکه عصبی برای بازی شطرنج، بسیار مهم است که چندین حرکت قبلی را در آموزش شبکه در نظر بگیریم زیرا یک حرکت اشتباه توسط بازیکن می‌تواند منجر به باخت نهایی در بازی در ۱۰ الی ۱۵ حرکت بعدی شود. همچنین، برای درک معنای واقعی یک جمله در یک نوشته، برخی اوقات لازم است تا به اطلاعات موجود در جملات یا پاراگراف‌های قبلی اتکا کنیم. این بدان معناست که برای دستیابی به یک درک حقیقی و درست باید یک بافتار را به طور متوالی و جمعی در طول زمان ایجاد کنیم. از این رو، در نظر گرفتن عنصر حافظه برای شبکه‌های عصبی که تأثیر حرکت‌های قبلی (مثلاً در بازی شطرنج) و جملات و پاراگراف‌های پیشین (در یک مقاله) را برای تعیین بهترین خروجی در نظر می‌گیرند بسیار حیاتی است. این حافظه، بافتار مورد نیاز برای یادگیری و درک را به تصویر کشیده و خلق می‌کند.

در شبکه‌های ایستا همانند سی‌ان‌ان‌های از نوع ام‌ال‌پی، ما به دنبال یافتن توابعی هستیم (یعنی اوزان و بایاس‌های شبکه) که بتوانند ورودی‌ها را بر روی برخی از خروجی‌هایی که تا حد امکان به هدف واقعی نزدیک هستند، نگاشت کنند. از طرف دیگر، در شبکه‌های پویا مثل شبکه‌های عصبی بازگشتی، ورودی‌ها و خروجی‌ها به صورت متوالی هستند. از این رو، یک شبکه‌ی پویا، بیشتر یک سیستم پویا است تا یک تابع زیرا خروجی آن نه تنها به ورودی، بلکه به خروجی‌های قبلی نیز بستگی دارد. اکثر شبکه‌های عصبی بازگشتی از معادله‌ی کلی زیر برای تعریف مقادیر واحدهای پنهان خود استفاده می‌کنند (گودفلا و همکاران ۲۰۱۶):

$$a^{(t)} = f(a^{(t-1)}, p^{(t)}, \theta)$$

در این معادله، $a^{(t)}$ بیانگر وضعیت سیستم در زمان t و $p^{(t)}$ و θ بیانگر ورودی به واحد در زمان t و پارامترها هستند. با بکارگیری همین معادله‌ی عمومی برای محاسبه‌ی وضعیت سیستم در زمان $t - 1$ خواهیم داشت:

$$a^{(t-1)} = f(a^{(t-2)}, p^{(t-1)}, \theta)$$

به عبارت دیگر،

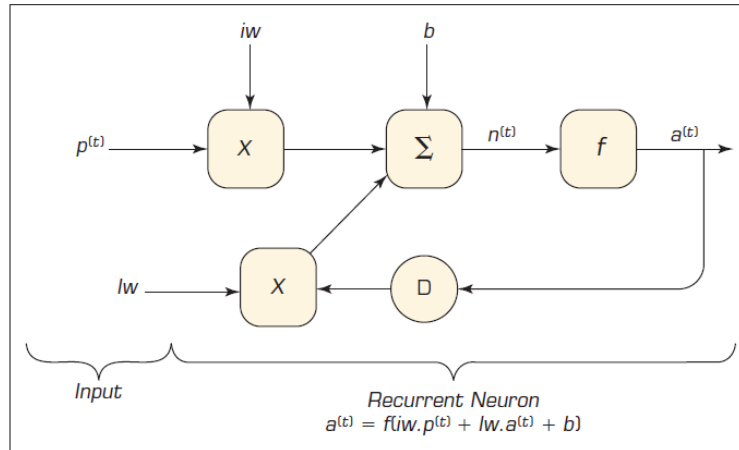
$$a^{(t)} = f(f(a^{(t-2)}, p^{(t-1)}, \theta), p^{(t)}, \theta)$$

و این معادله را می‌توان چندین مرتبه برای هر توالی با طول مشخص توسعه داد. به لحاظ گرافیکی، یک واحد بازگشتی در یک شبکه را می‌توان در قالب یک دیاگرام مدار همانند شکل ۶.۳۱ نمایش داد. در این تصویر، D بیانگر خط تأخیر با اتصال عرضی یا به بیان ساده، عنصر تأخیر آن شبکه است که در هر نقطه‌ی زمانی t حاوی $a^{(t)}$ یعنی مقدار خروجی قبلی واحد است. برخی اوقات به جای فقط یک مقدار، ما چندین مقدار خروجی قبلی را برای در نظر گرفتن تأثیر تمامی آنها در D ذخیره می‌کنیم. همچنین، iw و iw بیانگر بردارهای وزنی اعمال شده بر روی ورودی و تأخیر هستند.

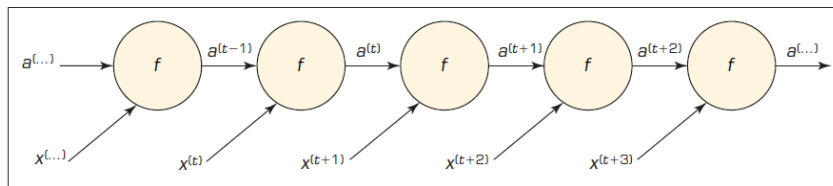
به لحاظ فنی، هر شبکه‌ای با بازخور را می‌توان در واقع یک شبکه عمیق نامید زیرا حتی با داشتن یک لایه، حلقه‌ی ایجاد شده توسط بازخور را می‌توان یک شبکه ام‌ال‌پی ایستا با لایه‌های زیاد در نظر گرفت. با وجود این، در عمل، هر شبکه عصبی بازگشتی شامل ده‌ها لایه است که هر کدام بازخور خاص خودشان یا حتی با لایه‌های قبلی را دارند که موجب عمیق‌تر و پیچیده‌تر شدن شبکه‌های عصبی بازگشتی می‌شود.

به دلیل وجود بازخورها، محاسبه‌ی گرادیان‌ها در شبکه‌های عصبی بازگشتی می‌تواند تا حدودی متفاوت از الگوریتم پس‌انتشار عمومی بکار رفته در شبکه‌های ام‌ال‌پی ایستا باشد. دو روش جایگزین برای محاسبه‌ی گرادیان‌ها در شبکه‌های عصبی بازگشتی

وجود دارد: یادگیری جاری بلادرنگ (آرتی آرال) و پس انتشار زمانی که توضیح آنها خارج از دامنه‌ی این فصل است. با وجود این، هدف اصلی در اینجا ثابت یکسان: به محض محاسبه‌ی گرادیان‌ها، روال‌های یکسانی برای بهینه‌سازی یادگیری پارامترهای شبکه بکار گرفته می‌شوند.



شکل ۶.۳۱. یک واحد برگشتی معمول

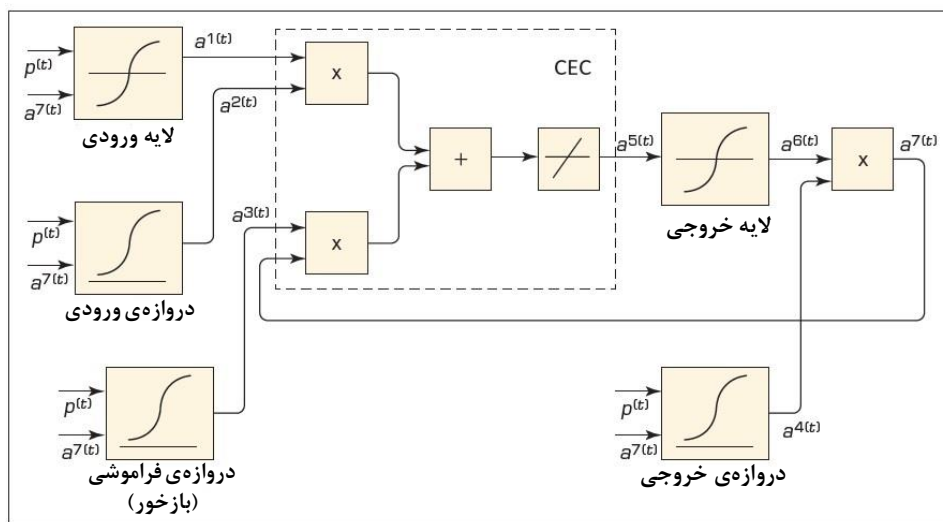


شکل ۶.۳۲. نمای باز شده از یک واحد برگشتی معمول

شبکه‌های ال‌ای‌تی‌ام (هوشرایتر آند اشمیت‌هوبر ۱۹۹۷) شکل‌های دیگری از شبکه‌های عصبی برگشتی هستند که امروزه به عنوان مؤثرترین تکنیک مدل‌سازی توالی شناخته شده و مبنای بسیاری از کاربردهای عملی هستند. در یک شبکه‌ی پویا، اوزان را حافظه‌ی بلند مدت می‌نامند در حالیکه بازخورها نقش حافظه‌ی کوتاه مدت را دارند. در واقع، فقط حافظه‌ی کوتاه مدت (یعنی بازخوردها، رویدادهای قبلی)، بافتار را برای شبکه فراهم می‌کنند. در یک شبکه‌ی عصبی برگشتی معمول، اطلاعات موجود در حافظه‌ی کوتاه مدت به طور مستمر با اطلاعات جدیدی که در طول زمان وارد شبکه می‌شود جایگزین می‌شود. به همین دلیل است که شبکه‌های عصبی برگشتی، زمانیکه بین اطلاعات مرتبط شکاف وجود داشته باشد و فضای مورد نیاز نیز کم باشد به خوبی عمل می‌کنند. به عنوان مثال، برای پیش‌بینی واژه‌ی مفقوده در جمله‌ی "The referee blew his whistle"، ما فقط کافی است چند واژه‌ی قبل را بدانیم (یعنی referee) تا بتوانیم به درستی پیش‌بینی کنیم. از آنجائیکه شکاف بین اطلاعات مرتبط (یعنی referee) و جائیکه مورد نیاز است (یعنی پیش‌بینی whistle) کوچک است، یک شبکه‌ی عصبی برگشتی به آسانی می‌تواند این وظیفه‌ی پیش‌بینی و یادگیری را انجام دهد.

با وجود این، برخی اوقات اطلاعات مرتبط مورد نیاز برای انجام یک وظیفه بسیار دورتر از جایی است که نیاز است تا در آنجا انجام شود. از این رو، این احتمال وجود دارد که این اطلاعات در زمانی که برای ایجاد بافتار مناسب مورد نیاز بوده است با اطلاعات

دیگری در حافظه‌ی کوتاه مدت جایگزین شده باشد. به عنوان مثال، برای پیش‌بینی واژه‌ی آخر در جمله‌ی "I went to a carwash yesterday. It cost 5\$ to wash my car" ، بین اطلاعات مربوطه (یعنی carwash) و جایی که مورد نیاز است شکاف بزرگتری وجود دارد. برخی اوقات ممکن است برای دستیابی به اطلاعات مربوطه برای پیش‌بینی معنای درست یک واژه نیاز باشد با پاراگراف‌های قبلی رجوع کنیم. در چنین مواردی، شبکه‌های عصبی برگشتی معمولاً عملکرد خوبی ندارند زیرا نمی‌توانند اطلاعات را برای مدت زمان طولانی در حافظه‌ی کوتاه مدت خود حفظ کنند. خوشبختانه، شبکه‌های ال‌اس‌تی‌ام چنین نقضی ندارند. عبارت شبکه حافظه‌ی کوتاه مدت طولانی به شبکه‌ای اشاره دارد که در آن ما سعی داریم تا آنچه که در گذشته نسبتاً طولانی مدت رخ داده است را به خاطر بیاوریم (یعنی بازخورها، خروجی‌های قبلی لایه‌ها) تا بتوان از آن در انجام وظایف مورد نیاز بهره گرفت. از یک نقطه نظر معماری، مفهوم حافظه (یعنی به یاد آوردن آنچه در گذشته رخ داده است) با بکارگیری چهار لایه‌ی اضافی در معماری شبکه برگشتی معمول در شبکه‌های ال‌اس‌تی‌ام میسر می‌شود: سه لایه‌ی دروازه‌ای به نام دروازه‌ی ورودی، دروازه‌ی فراموشی (بازخور) و دروازه‌ی خروجی و یک لایه‌ی اضافی به نام چرخه‌ی خطای ثابت یا وارد وضعیت که وظیفه‌ی یکپارچه‌سازی سه لایه‌ی دیگر و تعامل آنها با سایر لایه‌ها را بر عهده دارد. هر دروازه چیزی به جز یک لایه با دو ورودی نیست که یکی از ورودی شبکه و دیگری بازخور از خروجی نهایی کل شبکه است. این دروازه‌ها شامل توابع انتقال لاگ-سیگموئید هستند. بنابراین، خروجی آنها عددی بین ۰ و ۱ بوده و نحوه‌ی عبور هر مؤلفه از داخل شبکه را توضیح می‌دهند. همچنین، چرخه‌ی خطای ثابت یک لایه است که بین لایه‌های ورودی و خروجی در یک معماری شبکه برگشتی قرار گرفته و خروجی‌های دروازه‌ها را برای طولانی‌تر شدن حافظه‌ی کوتاه مدت بکار می‌گیرد.



شکل ۶.۳۳ معماری یک شبکه‌ی حافظه کوتاه مدت طولانی

داشتن یک حافظه‌ی کوتاه مدت طولانی به این معناست که ما می‌خواهیم تأثیر خروجی‌های قبلی را برای مدت زمان طولانی‌تر حفظ کنیم. با وجود این، ما معمولاً نمی‌خواهیم تا هرآنچه که در گذشته اتفاق افتاده است را به یاد بیاوریم. به همین خاطر، استفاده از دروازه‌ها امکان یادآوری گزینشی خروجی‌های قبلی را به ما می‌دهد. دروازه‌ی ورودی اجازه می‌دهد تا ورودی‌های به CEC را به صورت گزینشی انتخاب کنیم؛ دروازه‌ی فراموشی نیز بازخوردهای ناخواسته‌ی پیشین را از CEC پاک می‌کند و دروازه‌ی خروجی امکان گزینش خروجی‌ها از CEC را فراهم می‌کند. شکل ۶.۳۳ یک تصویر ساده از یک معماری ال‌اس‌تی‌ام معمول را نشان می‌دهد.

در مجموع، دروازه‌ها در ال‌اس‌تی‌ام مسئول کنترل جریان اطلاعات در سرتاسر شبکه و تغییر مقیاس زمانی یکپارچه‌سازی بر اساس توالی ورودی و به صورت پویا هستند. در نتیجه، شبکه‌های ال‌اس‌تی‌ام قادر به یادگیری آسان‌تر وابستگی‌های طولانی مدت مابین یک توالی از ورودی‌ها در مقایسه با شبکه‌های عصبی برگشتی معمولی هستند.

مورد کاربردی ۶.۷ استفاده از پردازش متن در بافتار درک نظرات و احساسات مشتری نسبت به طراحی و توسعه‌ی محصولات و خدمات جدید و بهبود یافته را به تصویر کشیده است.

ایجاد نوآوری از طریق درک احساسات مشتری

مورد کاربردی ۶.۷

تحلیل رفتار مشتری و محصول می‌تواند بینش و ادراک ارزشمندی را در خصوص اینکه مشتریان چه می‌خواهند، چگونه با محصولات ارتباط برقرار می‌کنند و کجا با قابلیت استفاده‌ی محصولات با مشکل مواجه می‌شوند فراهم کند. این بینش‌ها می‌توانند منجر به طراحی و توسعه‌ی ویژگی‌های جدید یا حتی محصولات جدید شوند.

درک احساس مشتریان و اطلاع از آنچه واقعاً در مورد محصولات یا یک برند فکر می‌کنند، نقاط درد رایج و قدیمی هستند. تحلیل سفر مشتری می‌تواند بینش‌هایی را در خصوص این حوزه‌ها به ما ارائه دهند اگرچه این راه‌حل‌ها تماماً به گونه‌ای طراحی نشده‌اند که بتوانند منابع حیاتی داده‌های ساختارنیافته از قبیل مرکز پاسخگویی، بازخوردهای رسانه‌های اجتماعی یا یادداشت‌ها را یکپارچه‌سازی کنند. در دنیای امروز، یادداشت‌های ساختارنیافته بخشی از ارتباطات اساسی در هر صنعتی محسوب می‌شوند به عنوان مثال:

- متخصصان پزشکی که مشاهدات بیماران را ثبت می‌کنند.
- تکنیسین‌های خودرو که اطلاعات ایمنی را یادداشت می‌کنند.
- خرده‌فروشان که رسانه‌های اجتماعی را برای نظرات مشتریان رصد می‌کنند.
- مراکز تماس که بازخوردهای مشتریان را پایش کرده و یادداشت برداری می‌کنند.

گردهم آوردن این یادداشت‌ها که معمولاً در قالب متون بدون شکل و فرم موجود هستند با سایر داده‌ها به منظور تجزیه و تحلیل می‌تواند کار دشواری باشد. این بدان خاطر است که هر صنعت، عبارات و اصطلاحات، اختصارات و زبان محاوره‌ای خاص خودش را دارد. یافتن معنا و بینش در کسب‌وکار در وهله اول نیازمند این است که متن را به شکل و فرم ساختیافته تبدیل کنیم. این فرایند دستی می‌تواند پرهزینه، زمان‌بر و مستعد بروز خطا باشد به ویژه زمانیکه حجم داده‌ها بسیار بالا است. یک روش که شرکت‌ها می‌توانند از یادداشت‌ها و بدون نیاز به کدینگ متن استفاده کنند این است که از خوشه‌بندی متن استفاده کنند. این تکنیک تحلیلی سریعاً به شناسایی واژگان یا جملات مشترک برای دستیابی به بینش سریع می‌پردازد.

متن و یادداشت‌ها می‌توانند منجر به محصولات جدید یا بهبود یافته شوند

بهره‌گیری از بینش و احساسات مشتری کشف شده در طی تحلیل متن و تحلیل احساسات می‌تواند منجر به نوآوری شوند. شرکت‌هایی از قبیل خودروسازان می‌توانند از این اطلاعات برای بهبود خدمات مشتریان و ارائه‌ی یک تجربه سطح بالا به مشتری استفاده کنند. با آگاهی از آنچه مشتری در مورد محصولات جاری دوست دارد یا ندارد، شرکت‌ها می‌توانند طراحی خود را از طریق افزودن ویژگی‌های جدید به یک وسیله‌ی نقلیه و ارتقاء تجربه‌ی رانندگی بهبود بخشند. تشکیل خوشه واژگان نیز به شرکت‌ها

اجازه می‌دهد تا مسائل ایمنی را شناسایی کنند. اگر یک خودروساز متوجه شوند که مشتریان متعددی، احساسات منفی را مبنی بر خروج دود سیاه رنگ از وسیله نقلیه‌شان ابراز می‌کنند، شرکت قادر به پاسخگویی به آنها خواهد بود. به همین منوال، خودروسازان می‌توانند به مسائل ایمنی که دغدغه و نگرانی مشتریان است رسیدگی کنند. با گروهبندی اظهارنظرات، شرکت‌ها توانایی این را دارند تا بر روی مشتریان خاصی که مشکل مشابهی داشته‌اند تمرکز کنند. این به شرکت‌ها اجازه می‌دهد تا امتیازها یا پاداش‌های ویژه‌ای را به آنها بدهد که مشکل خروج دود سیاه از خودروی خود داشته‌اند عرضه کنند.

درک احساسات می‌تواند به اطلاع‌رسانی بهتر سیاست‌ها و خط‌مشی‌های یک شرکت خودروساز کمک کند. به عنوان مثال، مشتریان دارای ارزش طول عمرهای متفاوتی هستند. یک مشتری که فقط یکبار شکایت می‌کند ولی ارزش طول عمر طولانی دارد اولویت بالاتری نسبت به یک مشتری که دارای ارزش طول عمر کوتاه‌تر بوده ولی مسائل و مشکلات متعددی است دارد. یک نفر ممکن است ۵,۰۰۰ دلار صرف خرید یک خودروی دست دوم کرده باشد. شخص دیگری ممکن است تاریخچه‌ی خرید خودروهای جدید از کارخانه را داشته و ۳۰,۰۰۰ دلار صرف خرید خودرو از نمایشگاه کرده باشد.

تحلیل یادداشت‌ها امکان دستیابی به نتایج ارزش بالا را در کسب‌وکار فراهم می‌کند

مدیریت چرخه عمر محصولات و خدمات همچنان برای بسیاری از شرکت‌ها دغدغه‌ساز است. حجم انبوه داده‌هایی که در حال حاضر وجود دارند موجب پیچیدگی چرخه‌ی عمر، ایجاد چالش‌های جدید بر سر راه نوآوری شده است. در همین حین، افزایش سریع بازخوردهای مشتریان از طریق رسانه‌های اجتماعی موجب شده تا سازمان‌ها از یک استراتژی برای هضم، اندازه‌گیری یا بکارگیری اطلاعات در چرخه نوآوری محصول خود برخوردار نباشند بدین معنی که حجم اطلاعات حیاتی که بیانگر افکار، احساسات یا عواطف مشتریان است موجود نمی‌باشد. تحلیل متن و احساسات، یک راه‌حل برای این مسئله است. تجزیه کردن موضوعات از انبوه متن‌ها به شرکت‌ها اجازه می‌دهد تا به مسائل، شکایات یا احساسات مثبت یا منفی مشتریان در خصوص محصولات پی ببرند. این بینش‌ها می‌توانند منجر به نتایج ارزش سطح بالا از قبیل بهبود محصولات یا خلق محصولات جدیدی شود که تجربه بهتری را به مشتری داده، به موقع به مسائل ایمنی آنها پاسخ داده و محبوب‌ترین خطوط محصول در نزد مشتریان را شناسایی می‌کنند.

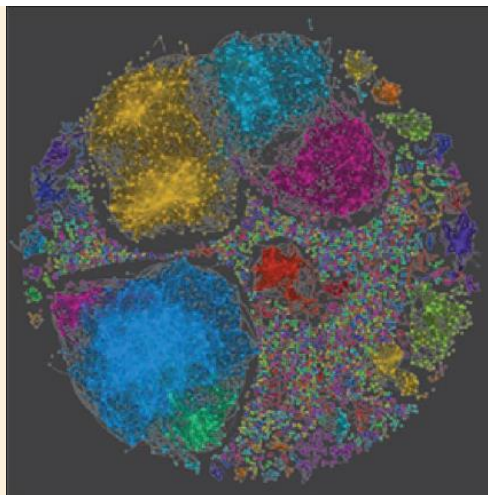
مثال: مصورسازی مسائل خودرویی به کمک "The Safety Cloud"

ترادیتا آرت آو آنالیتیکز^۱ از علم داده‌ها، ترادیتا آستر آنالیتیکز^۲ و تکنیک‌های مصورسازی برای تبدیل داده‌ها به یک کار هنری بی‌مانند استفاده می‌کند. برای نمایش بینش‌های منحصربفردی که از خوشه‌بندی متن بدست می‌آید، دانشمندان علم داده‌ها از آرت آو آنالیتیکز برای خلق "The Safety Cloud" استفاده کرده‌اند.

دانشمندان از الگوریتم‌های تحلیلی پیشرفته بر روی یادداشت‌های بازرس ایمنی و مرکز تماس بدست آمده از شرکت خودروساز استفاده کرده‌اند. تحلیل‌ها به شناسایی و استخراج سیستماتیک واژگان و جملات مشترک تعبیه شده در حجم انبوه داده‌ها پرداخته‌اند. خوشه‌آبی بیانگر از کار افتادن فرمان هیدرولیک خودرو است. خوشه‌ی صورتی نیز از کار افتادگی موتور خودرو و خوشه زرد، دود سیاه از اگزوز می‌باشد.

¹ Teradata Art of Analytics

² Teradata Aster Analytics



خوشه نارنجی خرابی یا از کار افتادن ترمز است. خودروساز می‌تواند از این اطلاعات برای مشکل نمودن بزرگی مسئله و اینکه آیا مرتبط با موضوع ایمنی است یا خیر استفاده کند و اقدامات لازم را برای رفع آن اتخاذ نماید.

برای یک خلاصه‌ی دیداری می‌توانید ویدئو را در <http://www.teradata.com/Resources/Videos.Art-of-Analytics-Safety-Cloud> تماشا کنید.

Source: Teradata Case Study. "Deliver Innovation by Understanding Customer Sentiments."
<http://assets.teradata.com/resourceCenter/downloads/CaseStudies/EB9859.pdf>
 (accessed August 2018). Used with permission.

سوالاتی برای مورد کاربردی ۶.۷

۱. به نظر شما چرا بر محبوبیت تحلیل احساسات افزوده می‌شود؟
۲. تحلیل احساسات چگونه کار می‌کند؟ چه چیزی تولید می‌کند؟
۳. علاوه بر مثال‌های مشخص در این مورد کاربردی، آیا می‌توانید به کسب‌وکارها و صنایع دیگری که می‌توانند از تحلیل احساسات بهره ببرند فکر کنید؟

شبکه‌های ال‌اس‌تی‌ام از زمان پیدایش خود در اواخر دهه ۹۰ (هوخرایتراند اسمیت‌هوبر^۱ ۱۹۹۷) به طور گسترده‌ای در بسیاری از کاربردهای مدلسازی متوالی از جمله کپشن‌گذاری تصاویر (یعنی توضیح خودکار محتوای تصاویر) (وینیالز، توشو، بنجیو و ارهان^۲ ۲۰۱۷ و ۲۰۱۵؛ ژو و همکاران^۳ ۲۰۱۵)، تولید و بازشناخت دست‌خط (گریوز^۴ ۲۰۱۳؛ گریوز آند اسمیت‌هوبر^۵ ۲۰۰۹؛ کیسرز و

¹ Hochreiter & Schmidhuber

² Vinyals, Toshev, Bengio, Erhan

³ Xu et al.

⁴ Graves

⁵ Graves & Schmidhuber

همکاران^۱ (۲۰۱۷)، تجزیه^۲ (لیانگ و همکاران ۲۰۱۶؛ وینیالز، کیزر و همکاران ۲۰۱۵)، شناخت گفتار (گریوز آند جیلتی^۳، ۲۰۱۴؛ گریوز، جیلتی آند محمد ۲۰۱۳؛ گریوز، محمد آند هینتون^۴ ۲۰۱۳) و ترجمه ماشینی (بادانو، چو آند بنجیو^۵، ۲۰۱۴؛ سوتسکور، وینیالز آند لی^۶ ۲۰۱۴).

در حال حاضر، ما با راه‌حل‌های متعدد یادگیری عمیق که بر مبنای شناخت گفتار کار می‌کنند احاطه شده‌ایم مثل سیری اپل، گوگل ناو^۷، کورتانا مایکروسافت و آمازون الکسا که برخی از آنها به صورت روزانه مورد استفاده قرار می‌گیرند (مثلاً چک کردن آب و هوا، درخواست جستجو در وب، تماس با یک دوست و درخواست مسیر بر روی نقشه). از آنجائیکه می‌توانیم یک سخنرانی یا گفتار را ضبط کرده، ضبط دیجیتال را بر روی پلتفرم‌های گفتار به متن مبتنی بر فناوری ابر بارگذاری کنیم و متن نوشتاری را در عرض چند ثانیه دانلود کنیم، یادداشت‌برداری کار چندان دشوار و خسته‌کننده‌ای نیست. سرویس گفتار به متن مبتنی بر ابر گوگل به عنوان مثال از ۱۲۰ زبان پشتیبانی کرده و توانایی تبدیل گفتار به متن را به صورت آنی یا با استفاده از فایل‌های صوتی ضبط شده دارد. سرویس گوگل به طور خودکار نویز داخل فایل صوتی را از بین برده و متن را به شکل دقیق با ویرگول، علامت سوال و فاصله نشانه‌گذاری می‌کند و کاربر قادر است تا با بکارگیری عبارات و جملاتی که به احتمال زیاد در یک گفتار استفاده خواهند شد، متن را به صورت سفارشی‌سازی شده درآورد.

ترجمه ماشینی به زیررشته‌ای از هوش مصنوعی اشاره دارد که از برنامه‌های رایانه‌ای برای ترجمه گفتار یا متن از یک زبان به زبان دیگر بهره می‌برد. یکی از جامع‌ترین سیستم‌های ترجمه ماشینی، پلتفرم ترجمه ماشینی عصبی گوگل (جی‌ان‌ام‌تی^۸) است. جی‌ان‌ام‌تی در واقع یک شبکه‌ی ال‌اس‌تی‌ام با هشت لایه‌ی رمزگذار و هشت لایه‌ی رمزگشا است که توسط گروهی از پژوهشگران گوگل در سال ۲۰۱۶ طراحی شد (وو و همکاران ۲۰۱۶). جی‌ان‌ام‌تی برای ترجمه تخصصی جملات کامل در یک زمان استفاده می‌شود که نقطه‌ی مقابل نسخه‌ی قبلی پلتفرم مترجم گوگل است که به صورت جمله به جمله ترجمه می‌کرد. این شبکه قادر است تا به طور طبیعی با تفکیک واژگان به مجموعه‌ای از واحدهای زیرواژه‌ی مشترک، واژگان ناشناس را ترجمه کند. جی‌ان‌ام‌تی در حال حاضر از ترجمه خودکار جملات بین بیش از ۱۰۰ زبان پشتیبانی می‌کند. شکل ۶.۳۴ نشان می‌دهد که چگونه یک جمله‌ی نمونه با جی‌ان‌ام‌تی و یک مترجم انسانی از زبان فرانسه به انگلیسی ترجمه شده است. این شکل رتبه‌بندی انجام شده بین ترجمه‌هایی که جی‌ان‌ام‌تی بین زبان‌های مختلف انجام داده و سخنوران انسانی انجام داده‌اند را با ترجمه‌های انسانی نشان می‌دهد.

¹ Keyzers et al.

² Parsing

³ Graves and Jaitly

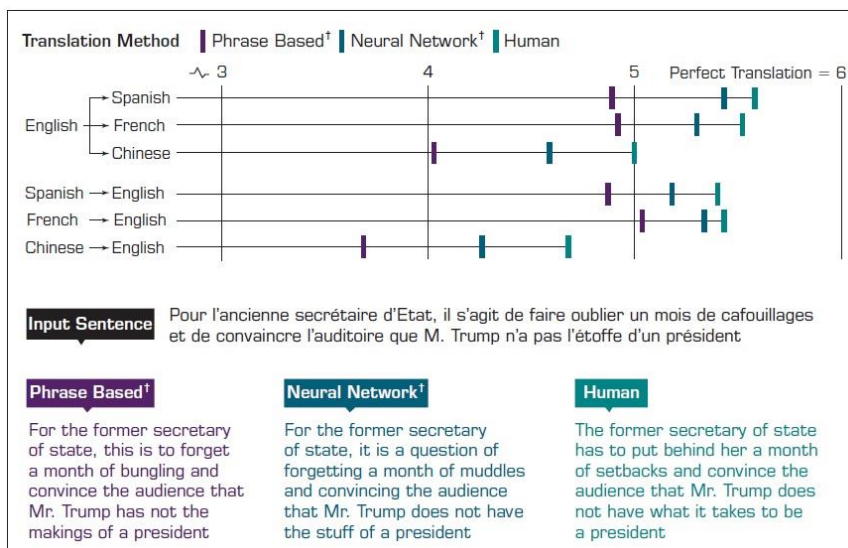
⁴ Mohamed and Hinton

⁵ Bahdanau, Cho, Bengio

⁶ Sutskever, Vinyals, Le

⁷ Google Now

⁸ Google Neural Machine Translation (GNMT)



شکل ۶.۳۴ مثال نشان‌دهنده‌ی عملکرد نزدیک به ترجمه انسانی توسط مترجم ماشینی عصبی گوگل

اگرچه ترجمه ماشینی به لطف ال‌اس‌تی‌ام‌ها دچار تحول و انقلابی شگرف شده است ولی همچنان با چالش‌هایی مواجه است که موجب شده تا کیفیت ترجمه‌های کاملاً خودکار را نداشته باشند. همانند کاربردهای پردازش تصاویر، داده‌های آموزشی کافی برای بسیاری از زوج زبان‌ها وجود ندارد که بتواند شبکه را بر اساس آنها آموزش داد. در نتیجه؛ ترجمه‌های بین زبان‌های نادر معمولاً از طریق یک زبان واسطه (معمولاً انگلیسی) انجام می‌شود که احتمال بروز خطاها را بسیار افزایش می‌دهد.

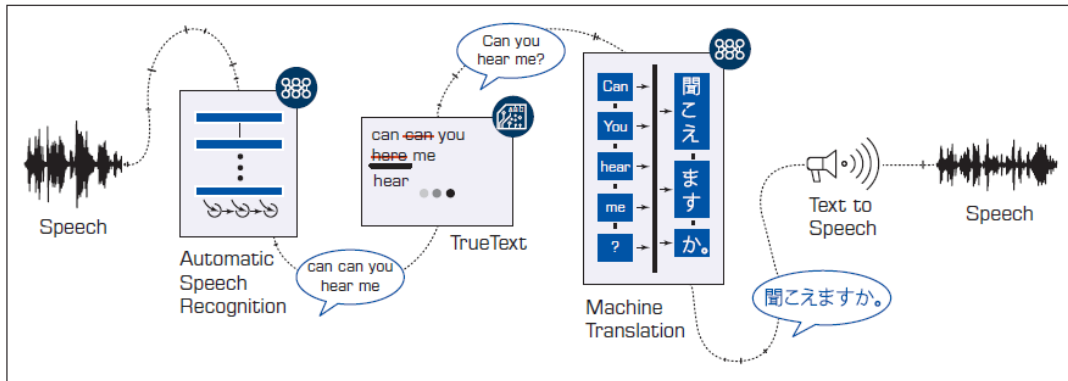
در ۲۰۱۴، مایکروسافت، سرویس مترجم اسکایپ^۱ خود را که یک سرویس رایگان ترجمه صوتی بود راه‌اندازی کرد که شامل بازساخت گفتار و ترجمه‌ی ماشینی با قابلیت ترجمه‌ی همزمان مکالمات به ۱۰ زبان است. افرادی که به زبان‌های متفاوتی صحبت می‌کنند می‌توانند با استفاده از این سرویس از طریق یک تماس ویدئویی یا صوتی اسکایپ به زبان خودشان صحبت کنند و سیستم قادر به تشخیص صدای آنها و ترجمه‌ی جمله به جمله از طریق یک بات مترجم به شکل آنی برای طرف مقابل است. برای ترجمه‌های دقیق‌تر، شبکه‌های عمیق که در قسمت پشتیبان این سیستم بکار گرفته شده‌اند به جای استفاده از زبان‌های رسمی متداول که در متون و نوشتار بکار می‌روند با استفاده از زبان محاوره‌ای (یعنی مواردی از قبیل صفحات وب ترجمه شده، زیرنویس فیلم‌ها و جملات مورد استفاده از محاورات روزمره در وب سایت‌های شبکه‌های اجتماعی) آموزش دیده‌اند. خروجی ماژول تشخیص گفتار این سیستم وارد TrueText که یک فناوری مایکروسافت برای نرمال‌سازی متن است می‌شود که قادر به شناسایی اشتباهات و خطاهایی که افراد معمولاً در گفتگوها و مکالمات خود مرتکب می‌شوند بوده و آنها را برای ترجمه‌های بهتر اصلاح می‌کنند. شکل ۶.۳۵ یک فرایند چهار مرحله‌ای مورد استفاده از مترجم اسکایپ مایکروسافت را نشان می‌دهد که هرکدام از این مراحل متکی بر شبکه‌های عصبی عمیق از نوع ال‌اس‌تی‌ام هستند.

سوالاتی برای مرور بخش ۶.۸

۵. آر‌ان‌ان چیست؟ چه تفاوتی با سی‌ان‌ان دارد؟

¹ Skype Translator

۶. اهمیت و معنای "بافتار"، "توالی" و "حافظه" در آر ان ان چیست؟
۷. کارکرد یک واحد شبکه عصبی برگشتی معمولی را ترسیم کرده و توضیح دهید.
۸. شبکه ال اس تی ام چیست و چه تفاوتی با آر ان ان دارند؟
۹. سه نوع کاربرد مختلف ال اس تی ام را نام برده و به اختصار توضیح دهید.
۱۰. ترجمه ی ماشینی عصبی گوگل و مترجم اسکایپ مایکروسافت چگونه کار می کنند؟



شکل ۶.۳۵ فرایند چهار مرحله‌ای ترجمه‌ی گفتار با استفاده از شبکه‌های عمیق در مترجم اسکایپ مایکروسافت

GPU

GPU

GPU

GPU

¹ Theano
² Torch
³ Caffe
⁴ PyLearn2
⁵ Tensorflow
⁶ MXNet

تورچ^۱ (کولوبرت و همکاران ۲۰۱۱) یک چارچوب رایانشی علمی منبع باز برای پیاده‌سازی الگوریتم‌های یادگیری ماشین با استفاده از GPU ها می‌باشد. چارچوب تورچ یک کتابخانه مبتنی بر LuaJIT است که یک نسخه ی کامپایل شده از زبان برنامه نویسی محبوب Lua می باشد. در واقع، تورچ تعدادی ویژگی ارزشمند را به Lua می‌افزاید که تحلیل‌های یادگیری عمیق را امکان‌پذیر می‌سازند. این روش از آرایه‌های n بُعدی پشتیبانی می‌کند (یعنی تنسورها) در حالیکه جداول (یعنی آرایه‌های دو بُعدی) به طور طبیعی تنها روش ساختاردهی داده‌ها است که توسط Lua بکار گرفته می‌شود. علاوه بر این، تورچ شامل کتابخانه‌های روتین برای دستکاری (یعنی اندیس‌گذاری، تفکیک، ترانهادن) تنسورها، جبر خطی، کارکردهای شبکه‌های عصبی و بهینه‌سازی است. مهمتر اینکه، اگرچه Lua به طور پیش فرض از واحد پردازنده مرکزی برای اجرای برنامه‌ها استفاده می‌کند، تورچ امکان استفاده از واحدهای پردازنده گرافیکی را برای اجرای برنامه‌هایی که به زبان Lua نوشته شده‌اند فراهم می‌کند. ویژگی‌های ساده و بسیار سریع نوشتاری LuaJIT به همراه انعطاف‌پذیری آن موجب شده تا تورچ به یک چارچوب بسیار محبوب برای کاربردهای یادگیری عمیق تبدیل شود به گونه‌ای که آخرین نسخه‌ی آن یعنی تورچ ۷ به طور گسترده‌ای توسط چندین شرکت بزرگ در حوزه‌های یادگیری عمیق از قبیل فیسبوک، گوگل و آی‌بی‌ام در آزمایشگاه‌های پژوهشی و همچنین کاربردهای تجاری استفاده شده است.

<http://caffe.berkeleyvision.org>

C++

MATLAB

.prototxt

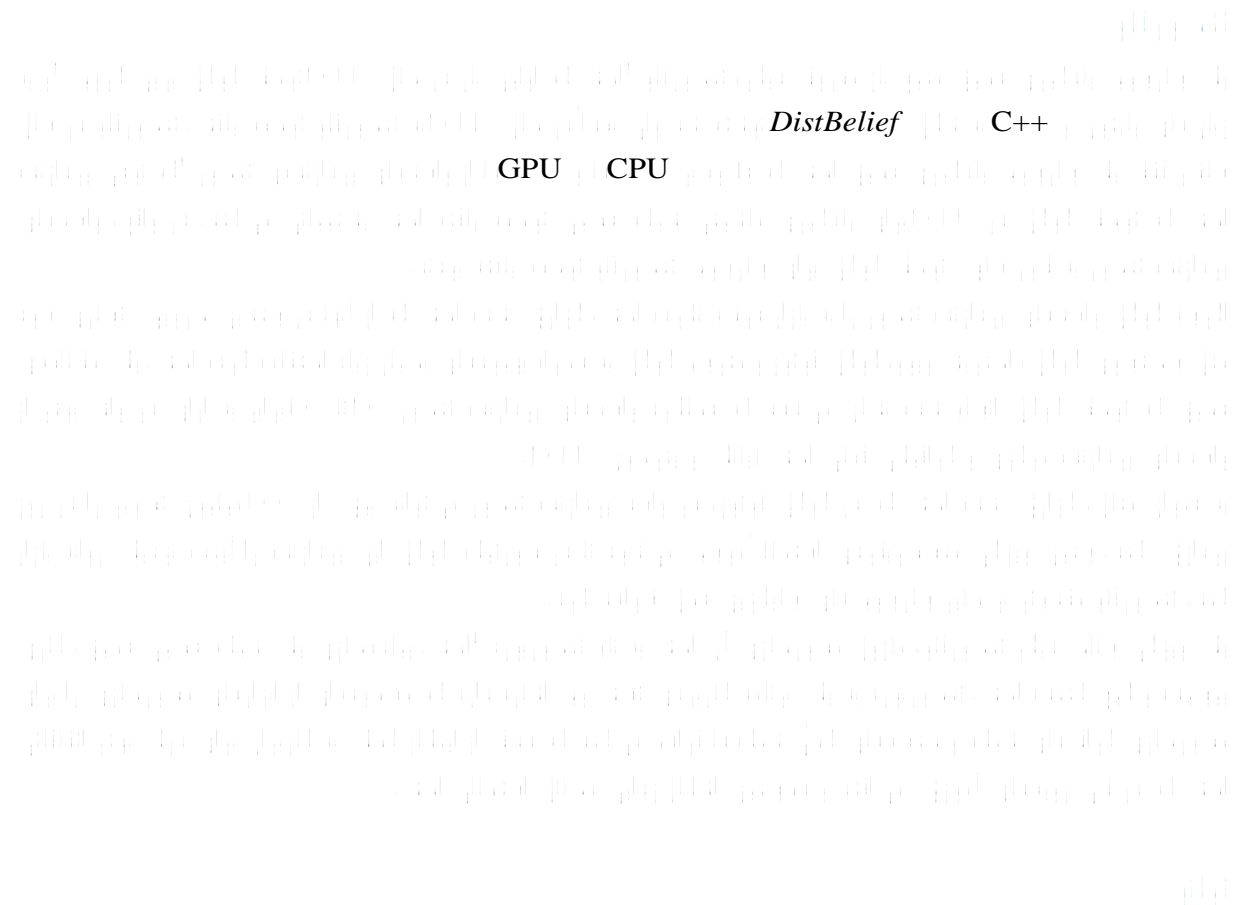
¹ Torch

² Caffe

³ Yangqing Jia

⁴ JavaScript Object Notation (JSON)

NVIDIA K40



¹ Solver

² TensorFlow

³ Google Brain Group

⁴ Tensor Processing Unit (TPU)

⁵ TensorBoard

<http://deeplearning.net/software/theano>

GPU CPU

session

GPU

CPU

C

C

GPU

[\(https://keras.io/\)](https://keras.io/)

سوالاتی برای مرور بخش ۶.۹

۶. علیرغم علاقه کوتاه به پیاده‌سازی یادگیری عمیق، به نظر شما چرا چندین چارچوب رایانشی متفاوت برای آن وجود دارد؟

۷. *CUDA*، *NVIDIA*، *CPU* و یادگیری عمیق را تعریف کرده و در خصوص روابط بین آنها اظهار نظر کنید.

¹ Theano

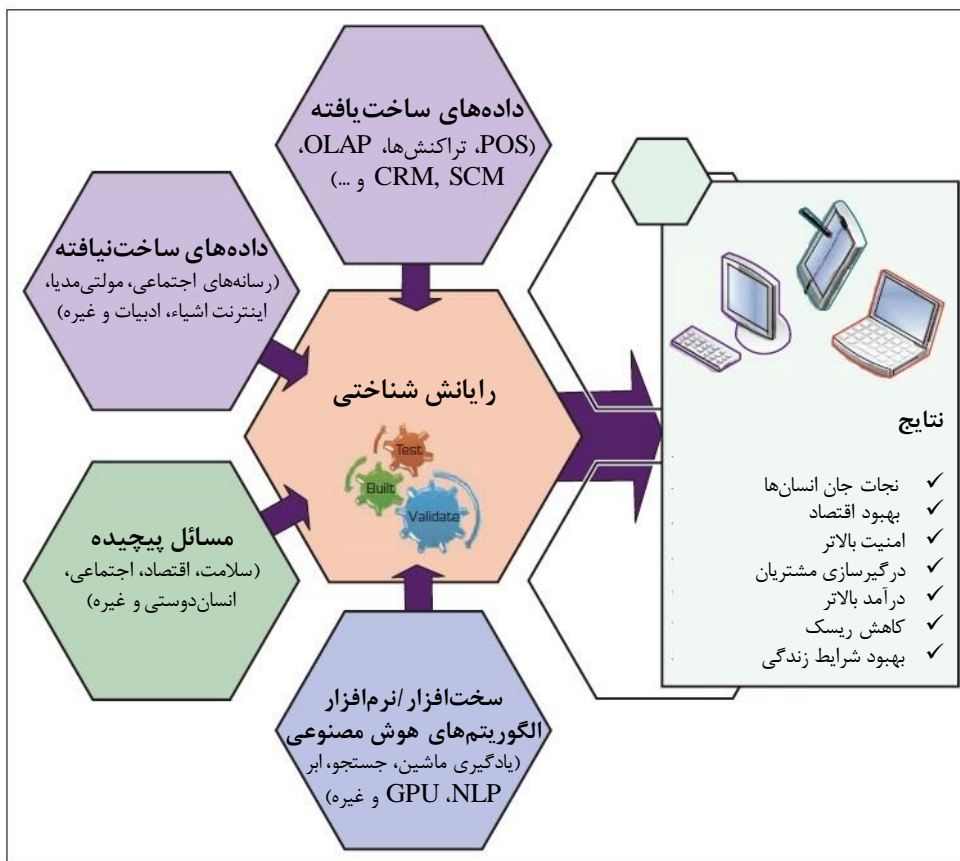
² Keras

۸. ویژگی‌های چارچوب‌های مختلف یادگیری عمیق را نام برده و به اختصار توضیح دهید.

۹. کراس چیسیت و چه تفاوتی با سایر چارچوب‌ها دارد؟

ما شاهد افزایش چشمگیر در نحوه رشد و تکامل فناوری هستیم. اموری که انجام آنها قبلاً چندین دهه به طول می‌انجامید اکنون در عرض چند ماه انجام می‌شود و چیزهایی که ما فقط در فیلم‌های علمی و تخیلی می‌دیدیم یکی بعد از دیگری به واقعیت می‌پیوندد. از این رو، می‌توانیم بگوییم که در یک یا دو دهه‌ی اخیر دستاوردهای تکنولوژیکی، نحوه‌ی زندگی، یادگیری و کار مردم را شدیداً دستخوش تغییر خواهند کرد. تعاملات بین انسان‌ها و فناوری، شهودی‌تر و شفاف‌تر خواهد شد. رایانش شناختی نقش بسزایی را در این تغییر و دگرگونی ایفا خواهد کرد. به بیان کلی‌تر، رایانش شناختی به سیستم‌های رایانشی اشاره دارد که از مدل‌های ریاضی برای تقلید (یا شبیه‌سازی جزئی) فرایند شناختی انسانی برای یافتن راه‌حل مسائل و موقعیت‌های پیچیده که پاسخ‌های بالقوه ممکن است دقیق نباشند استفاده می‌کند. اگرچه عبارت رایانش شناختی اغلب به عنوان جایگزین هوش مصنوعی و موتورهای جستجوی هوشمند بکار می‌رود و شاید ارتباط بسیار نزدیکی با سیستم رایانه‌ای شناختی آی‌بی‌ام به نام واتسون و موفقیت آن در نمایش تلویزیونی *جئوپاردی!* داشته باشد. جزئیات موفقیت واتسون در *جئوپاردی!* را می‌توانید در مورد کاربردی ۶.۸ مطالعه کنید.

بر اساس کنسرسیوم رایانش شناختی (۲۰۱۸)، رایانش شناختی، گروه جدیدی از مسائل را قابل محاسبه می‌کند و به موقعیت‌های بسیار پیچیده‌ای می‌پردازد که دارای ابهام و عدم قطعیت هستند. به بیان دیگر، می‌تواند به انواع مسائلی که تصور می‌شود فقط خلاقیت و نبوغ انسانی قادر به حل آنها است بپردازد. در موقعیت‌های پویا، ناپایدار و مملو از اطلاعات امروزی، داده‌ها به طور مکرر تغییر می‌کنند و اغلب با یکدیگر تعارض دارند. با یادگیری بیشتر و بازتعریف مقاصد، اهداف کاربران نیز تغییر می‌کند. برای پاسخگویی به این ماهیت سیال درک کاربران از مسائل، سیستم رایانش شناختی یک تلفیق از نه تنها منابع اطلاعاتی بلکه تأثیرات، بافتارها و بینش‌ها را نیز پیشنهاد می‌دهد. برای دستیابی به چنین سطح بالایی از عملکرد، سیستم‌های رایانشی اغلب باید به وزن‌دهی شواهد متعارض پرداخته و «بهترین» پاسخ را به جای پاسخ «درست» پیشنهاد کنند. شکل ۶.۳۶ یک چارچوب کلی برای رایانش شناختی را به تصویر کشیده که از فناوری‌های داده و هوش مصنوعی برای حل مسائل پیچیده و واقعی استفاده می‌شود.



شکل ۶.۳۶ چارچوب مفهومی رایانش شناختی و نتایج مورد انتظار از آن

همانطور که از نامش می‌توان حدس زد، رایانش شناختی بسیار مشابه فرایند ذهن انسان، مکانیزم استنتاج و سیستم شناختی عمل می‌کند. این سیستم‌های محاسباتی بسیار پیچیده می‌توانند داده‌ها را از منابع اطلاعاتی مختلف یافته و تلفیق کرده و برای دستیابی به بهترین پاسخ‌های ممکن برای یک مسئله یا سوال به شواهد و مدارک ذاتی درون داده‌ها وزن بدهند. برای تحقق این امر، سیستم‌های شناختی شامل فناوری‌های خودیادگیرنده هستند که از داده‌کافی، تشخیص الگو، یادگیری عمیق و پردازش زبان طبیعی برای تقلید روش کارکرد مغز انسان استفاده می‌کنند.

استفاده از سیستم‌های رایانه‌ای برای حل انواع مسائلی که انسان‌ها با آنها سر و کار دارند معمولاً نیازمند حجم انبوهی از داده‌های ساخت یافته و ساخت نیافته است که باید در الگوریتم‌های یادگیری ماشینی وارد شود. در طول زمان، سیستم‌های شناختی قادر به بازتعریف روش یادگیری و تشخیص الگوها و روش پردازش داده‌ها برای پیش‌بینی مسأله جدید و مدل‌سازی و پیشنهاد جواب‌های احتمالی هستند.

برای دستیابی به آن قابلیت‌ها، سیستم‌های رایانش شناختی باید از ویژگی‌های کلیدی زیر که توسط کنسرسیوم رایانش شناختی تعریف شده است برخوردار باشند (۲۰۱۸):

- **وفق پذیر:** سیستم‌های شناختی باید به اندازه‌ی کافی انعطاف‌پذیر باشند تا بتوانند همراه با تغییر اطلاعات و تکامل اهداف، به یادگیری بپردازند. این سیستم‌ها باید قادر به هضم داده‌های پویا در لحظه و ایجاد تغییرات همراه با تغییر داده‌ها و محیط باشند.
- **تعاملی:** تعامل انسان-ماشین یک مؤلفه‌ی حیاتی در سیستم‌های شناختی است. کاربران باید قادر به تعامل با ماشین‌های شناختی بوده و با تغییر نیازها، نیازهایشان را تعریف کنند. این فناوری‌ها باید قادر به تعامل با سایر پردازنده‌ها، دستگاه‌ها و پلتفرم‌های ابری باشند.
- **تکرار شونده و وضعیت‌پذیر:** فناوری‌های رایانش شناختی همچنین با پرسیدن سوالات یا اخذ اطلاعات بیشتر در صورتیکه مسئله‌ی بیان شده ناقص یا مبهم باشد قادر به شناسایی مسائل هستند. سیستم‌ها اینکار را با حذف اطلاعات در مورد موقعیت‌های مشابهی که قبلاً رخ داده‌اند انجام می‌دهند.
- **بافتاری (زمینه‌ای):** درک بافتار در فرایندهای ذهنی و فکری اهمیت بسیار کلیدی دارد و به همین دلیل سیستم‌های شناختی باید داده‌های بافتاری از قبیل سینتکس، زمان، مکان، دامنه، الزامات و پروفایل یک کاربر خاص، وظایف یا اهدافش را درک، شناسایی و کاوش کنند. سیستم‌های شناختی می‌توانند داده‌ها را از منابع اطلاعاتی متعددی از جمله داده‌های ساخت‌یافته و ساخت‌نیافته و داده‌های دیداری، شنیداری و حسگرها اخذ کنند.

رایانش شناختی اغلب به شکل جایگزین با هوش مصنوعی بکار می‌رود که یک عبارت چتری است که برای فناوری‌هایی که برای تصمیم‌گیری مبتکی بر داده‌ها و روش‌ها/محاسبات علمی هستند اطلاق می‌شود. اما تفاوت‌هایی بین این دو عبارت وجود دارد که به میزان زیادی به اهداف و کاربردهای آنها بستگی دارد. فناوری‌های هوش مصنوعی شامل یادگیری ماشینی، رایانش عصبی، NLP و جدیدترین آنها یادگیری عمیق هستند ولی محدود به اینها نیستند. در سیستم‌های هوش مصنوعی به ویژه در سیستم‌های یادگیری ماشینی، داده‌ها برای پردازش وارد الگوریتم‌ها می‌شوند (طی یک فرایند تکرار شونده و زمان‌بر که اغلب آموزش نامیده می‌شود)، بدین ترتیب که سیستم‌ها قادر به یادگیری متغیرها و ارتباطات بین آنها هستند که می‌توانند پیش‌بینی‌ها را در مورد یک مسئله یا موقعیت خاص انجام دهند.

کاربردهای مبتنی بر هوش مصنوعی و رایانش شناختی شامل دستیارهای هوشمند از قبیل آمازون الکسا، گوگل هوم و سیری شرکت اپل است. یک مقایسه‌ی ساده بین رایانش شناختی و هوش مصنوعی در جدول ۶.۳ آورده شده است (رینولدز و فلدمن^۱ ۲۰۱۴؛ CCC 2018). همانطور که در جدول ۶.۳ مشاهده می‌کنید، تفاوت‌های بین هوش مصنوعی و رایانش شناختی بسیار جزئی است. این موضوع مورد انتظار است چون رایانش شناختی اغلب به عنوان یک زیرمجموعه از هوش مصنوعی یا یک کاربرد از فناوری‌های هوش مصنوعی شناخته می‌شود که برای هدف خاص طراحی شده‌اند. هوش مصنوعی و رایانش شناختی هر دو از فناوری‌های مشابهی استفاده کرده و در بخش‌های مشابهی از صنعت بکار گرفته می‌شوند. تفاوت اصلی بین این دو در هدف آنها است: هدف رایانش شناختی، کمک به انسان‌ها برای حل مسائل پیچیده است در حالیکه هدف هوش مصنوعی، خودکارسازی فرایندهایی است که توسط انسان انجام می‌شود و در حد نهایی خود، به دنبال این است که ماشین‌ها را برای انجام وظایفی که نیازمند هوش هستند جایگزین انسان‌ها کند.

¹ Reynolds and Feldman

در سال‌های اخیر، رایانش شناختی معمولاً برای توضیح سیستم‌های هوش مصنوعی که هدف آنها شبیه‌سازی فرایند تفکر انسانی است استفاده شده است. شناخت انسانی شامل تحلیل بلادرنگ محیط، بافتار و مقصود است توانایی یک شخص را برای حل مسائل مشخص می‌کند. برای اینکه یک سیستم رایانه‌ای بتواند مدل‌های شناختی را ایجاد کند که فرایندهای تفکر انسانی را تقلید کنند به فناوری‌های هوش مصنوعی از قبیل یادگیری ماشینی، یادگیری عمیق، شبکه‌های عصبی، NLP، متن‌کاوی و تحلیل احساس نیاز دارند.

جدول ۶.۳ رایانش شناختی در برابر هوش مصنوعی

مشخصه	رایانش شناختی	هوش مصنوعی
فناوری‌های بکار رفته	<ul style="list-style-type: none"> یادگیری ماشین پردازش زبان طبیعی شبکه‌های عصبی یادگیری عمیق متن‌کاوی تحلیل احساسات 	<ul style="list-style-type: none"> یادگیری ماشین پردازش زبان طبیعی شبکه‌های عصبی یادگیری عمیق
قابلیت‌های ارائه شده	شبیه‌سازی فرایندهای ذهن انسانی برای کمک به انسان‌ها در یافتن راه‌حل برای مسائل پیچیده	یافتن الگوهای پنهان در انواع منابع داده‌ها برای شناسایی مسائل و ارائه راه‌حل‌های ممکن
هدف	ارتقاء قابلیت و توانایی انسانی	خودکارسازی فرایندهای پیچیده با عمل کردن همانند انسان در موقعیت‌های خاص
صنایع	خدمات مشتری، بازاریابی، مراقبت سلامت، سرگرمی، بخش خدمات	ساخت و تولید، مالی، مراقبت سلامت، بانک‌داری، خرده‌فروشی، دولت

به طور کلی، از رایانش شناختی برای کمک به انسان‌ها در فرایند تصمیم‌گیری استفاده می‌شود. برخی مثال‌ها از کاربردهای رایانش شناختی شامل پشتیبانی از پزشکان در درمان بیماری می‌باشد. آی‌بی‌ام و اتسون برای تومورشناسی در مرکز سرطان مموریال سلون کترینگ است که به پزشکان تومورشناسی در درمان بیماران سرطانی بر اساس شواهد و مدارک کمک می‌کند. زمانی که پزشکان سوالات خود را مطرح می‌کنند، واتسون فهرستی از فرضیه‌ها را تولید کرده و گزینه‌های درمان را برای پزشکان ارائه می‌کند. اگرچه هوش مصنوعی برای حل یک مسئله خاص یا شناسایی الگوهای پنهان در داده‌ها متکی بر الگوریتم‌ها است، سیستم‌های رایانش شناختی به دنبال خلق الگوهایی است که فرایند استنتاج ذهنی انسان را برای کمک به انسان‌ها در حل مجموعه‌ای مسائل تلقید می‌کند.

در خصوص رسیدگی به موقعیت‌های پیچیده، بافتار از اهمیت بالایی برخوردار است و سیستم‌های رایانش شناختی امکان انجام محاسبات بر اساس بافتار را فراهم می‌کنند. آنها ویژگی‌های بافتاری از قبیل زمان، مکان، وظیفه، تاریخچه یا پروفایل را برای ارائه مجموعه خاصی از اطلاعات که برای یک فرد مناسب است در زمان و مکان خاصی شناسایی و استخراج می‌کنند. طبق کنسرسیوم رایانش شناختی آنها با حرکت در بین مجموعه‌های انبوهی از اطلاعات متنوع برای یافتن الگوها و بکارگیری آنها الگوها برای پاسخگویی به نیازهای کاربر در یک لحظه‌ی خاص می‌توانند موهبت‌های غیرمنتظره‌ای را به ارمغان بیاورند. به مفهومی دیگر، سیستم‌های رایانشی شناختی به دنبال بازتعریف ارتباطات بین انسان‌ها و محیط شدیداً در حال دیجیتالی شدن پیرامون آنها هستند. آنها ممکن است نقش یک دستیار یا مربی را برای کاربر ایفا کنند و ممکن است در بسیاری از موقعیت‌های حل مسئله به طور کاملاً مستقل عمل کنند. مرزهای فرایندها و حوزه‌هایی که این سیستم‌ها می‌توانند بر آنها تأثیر بگذارند همچنان در حال گسترش و پیدایش است. خروجی آنها ممکن است تجویزی، توصیه‌ای، دستوری یا صرفاً سرگرم کننده باشد. رایانش شناختی در مدت زمان کوتاهی از حضورش می‌گذرد در موقعیت‌های پیچیده و حوزه‌های بسیاری مفید بوده و در حال گسترش نیز می‌باشد. کاربردهای معمول رایانش شناختی شامل موارد زیر است:

- توسعه‌ی موتورهای جستجوی هوشمند و وفق‌پذیر
- استفاده‌ی مؤثر از پردازش زبان طبیعی
- تشخیص گفتار
- ترجمه‌ی زبان
- تحلیل احساسات بر اساس بافتار
- تشخیص چهره و تشخیص حالات چهره
- ارزیابی و پیشگیری از ریسک
- تشخیص و پیشگیری از کلاهبرداری
- ارزیابی و توصیه‌های رفتاری

تحلیل شناختی¹ عبارتی است که به پلتفرم‌های فناوری‌های رایانش شناختی از قبیل آی‌بی‌ام و اتسون اشاره دارد که به طور تخصصی در پردازش و تحلیل مجموعه داده‌های بزرگ و ساخت‌نیافته بکار می‌روند. معمولاً، مستندات پردازش واژه، نامه‌های الکترونیکی، ویدئوها، تصاویر، فایل‌های صوتی، پرزنت‌ها، صفحات وب، رسانه‌های اجتماعی و دیگر فرمت‌های داده‌ها را پیش از آنکه بتوان آنها را برای تحلیل‌های رایانشی و تولید بینش وارد موتورهای تحلیل و ابزارهای کلان‌داده‌ها کرد باید به صورت دستی با فراداده‌ها تگ‌گذاری شوند. مزیت اصلی بکارگیری تحلیل‌های شناختی نسبت به ابزارهای تحلیل کلان‌داده‌ها این است که برای تحلیل‌های شناختی از قبیل مجموعه داده‌ها نیازی به تگ‌گذاری از قبل نیست. سیستم‌های تحلیل شناختی می‌توانند از یادگیری ماشینی برای سازگاری بافتارهای مختلف با حداقل نظارت انسانی استفاده کنند. این سیستم‌ها می‌توانند مجهز به یک چَت‌بات یا دستیار جستجو باشند که پرس و جوها را درک کرده، بینش داده‌ها را توضیح داده و با انسان‌ها به زبان انسان تعامل برقرار می‌کند.

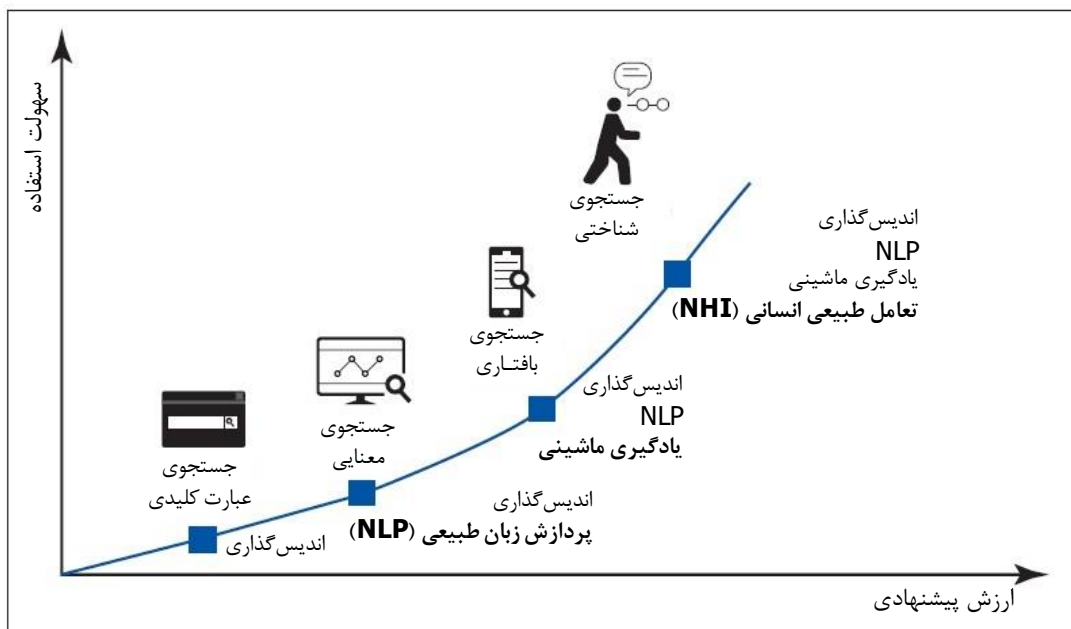
¹ Cognitive Analysis

جستجوی شناختی، یک روش جستجوی نسل جدید است که از هوش مصنوعی (اندیس‌گذاری پیشرفته، NLP و یادگیری ماشین) برای بازگرداندن نتایج مرتبط‌تر به کاربر استفاده می‌کند. فارستر، جستجوی شناختی و راه‌حل‌های اکتشاف دانش را به صورت «نسل جدید راه‌حل‌های جستجوی سازمانی تعریف می‌کند که از فناوری‌های هوش مصنوعی از قبیل پردازش زبان طبیعی و یادگیری ماشین برای هضم، درک، سازماندهی و استعلام محتوای دیجیتال از چندین منبع داده استفاده می‌کنند.» (گوئال تیری ۲۰۱۷). جستجوی شناختی با بکارگیری الگوریتم‌های رایانش شناختی از بین محتوای غیرقابل جستجو، اطلاعات قابل جستجو را ایجاد کرده و یک پلتفرم اندیس‌گذاری را ایجاد می‌کند.

جستجوی اطلاعات کار بسیار دشوار و پرهزمتی است. اگرچه موتورهای جستجوی جاری در یافتن اطلاعات مربوط در زمان مناسب بسیار خوب عمل می‌کنند، منابع آنها معمولاً محدود به داده‌های موجود در اینترنت است. جستجوی شناختی یک نسل جدید از جستجوهای سفارشی را برای استفاده در سازمان‌ها پیشنهاد می‌دهد. جستجوی شناختی به دلایل زیر متفاوت از جستجوی سنتی و قدیمی است (گوئال تیری ۲۰۱۷):

- **قادر به مدیریت انواع مختلف داده‌ها هستند.** جستجو دیگر محدود به متون ساخت‌نیافته در مستندات و صفحات وب نیست. راه‌حل‌های جستجوی شناختی می‌توانند داده‌های ساخت‌یافته در پایگاه‌های داده و حتی داده‌های سازمانی غیرمتداول از قبیل تصاویر، ویدئوها، صوت و سوابق تولید شده توسط ماشین‌ها/حسگرها از دستگاه‌های اینترنت اشیاء را نیز مدیریت کنند.
- **قادر به بافتاری‌سازی فضای جستجو هستند.** در بازیابی اطلاعات، بافتار بسیار اهمیت دارد. بافتار قادر است تا جستجوی سنتی مبتنی بر سینتکس/سمبول را به سطح جدیدی ارتقاء دهد که بر اساس معنا و مفهوم‌شناسی تعریف می‌شود.
- **از فناوری‌های هوش مصنوعی پیشرفته استفاده می‌کنند.** ویژگی متمایزکننده‌ی راه‌حل‌های جستجوی شناختی این است که از یادگیری ماشین و NLP برای درک و سازمان‌دهی داده‌ها، پیش‌بینی ماهیت استعلام جستجو، بهبود وابستگی نتایج و تنظیم خودکار وابستگی نتایج در طول زمان استفاده می‌کنند.
- **برنامه‌نویسان را قادر به تولید برنامه‌های جستجوی خاص هر سازمان می‌کند.** جستجو فقط یک کادر متنی بر روی یک پرتال سازمانی نیست. سازمان‌ها، برنامه‌های جستجو را طراحی و ایجاد می‌کنند که امکان جستجو را در برنامه‌های مشتری ۳۶۰، ابزارهای جستجوی دارو و بسیاری از برنامه‌های فرایند کسب‌وکار فراهم می‌کنند. دستیارهای دیجیتالی مجازی از قبیل آمازون الکسا، گوگل نو و سیری بدون جستجوی قدرتمندی که در پس صحنه وجود دارد بی‌فایده خواهند بود. سازمان‌هایی که می‌خواهند برنامه‌های مشابهی را برای مشتریان خود تولید کنند نیز می‌توانند از راه‌حل‌های جستجوی شناختی بهره ببرند. راه‌حل‌های جستجوی شناختی ابزارهای توسعه نرم‌افزار، APIها و ابزارهای طراحی دیداری را فراهم می‌کنند که به برنامه‌نویسان اجازه می‌دهند تا قدرت موتورهای جستجو را در برنامه‌های کاربردی تعبیه کنند.

شکل ۶.۳۷ تکامل تدریجی روش‌های جستجو را از جستجو قدیمی عبارات کلیدی تا جستجوی شناختی امروزی بر روی دو بُعد نشان می‌دهد – سهولت استفاده و ارزش پیشنهادی.



شکل ۶.۳۷ تکامل تدریجی روش‌های جستجو

آی‌بی‌ام و اتسون را شاید بتوان باهوش‌ترین سیستم رایانه‌ای دانست که تا کنون ساخته شده است. از زمان پیدایش رایانه‌ها و در نتیجه، هوش مصنوعی در اواخر دهه ۴۰، دانشمندان عملکرد این ماشین‌های «هوشمند» را با ذهن انسان مقایسه کرده‌اند. بر همین اساس، در اواسط تا اواخر دهه ۹۰، پژوهشگران آی‌بی‌ام یک ماشین هوشمند را ساخته و برای تست توانایی آن در برابر بهترین ذهن‌های انسانی از بازی شطرنج استفاده کردند. در ۱۱ می ۱۹۹۷ یک رایانه آی‌بی‌ام به نام دیپ بلو، استاد بزرگ شطرنج جهان را پس از شش مسابقه شکست داد: دو پیروزی برای دیپ بلو، یک پیروزی برای قهرمان جهان و سه مساوی. این مسابقه چندین روز به طول انجامید و تعداد بسیاری از رسانه‌ها در سرتاسر جهان این مسابقه را پوشش دادند. این یک نمایش کلاسیک از مبارزه‌ی ماشین در برابر انسان بود. جدای یک مسابقه‌ی شطرنج، هدف از توسعه‌ی این نوع هوش رایانه‌ای این بود که رایانه‌ها بتوانند انواع مختلف محاسبات پیچیده مورد نیاز برای کشف داروهای جدید و انجام مدل‌سازی‌های مالی گسترده مورد نیاز برای شناسایی روندها و تحلیل ریسک، مدیریت جستجو در پایگاه‌های داده بزرگ و محاسبات پیچیده مورد نیاز در رشته‌های مختلف علمی را انجام دهند.

پس از گذشت چندین دهه، پژوهشگران آی‌بی‌ام به ایده‌ی دیگری دست یافتند که شاید چالش‌برانگیزتر بود: یک ماشین که نه تنها می‌توانست در نمایش مسابقه‌ی تلوویزیونی آمریکا به نام جنوپاردی! بازی کند بلکه توانست بهترین‌ها را شکست دهد. جنوپاردی! در مقایسه با شطرنج بسیار چالش‌برانگیزتر است. اگرچه شطرنج دارای ساختار مشخصی است و قواعد بسیار ساده‌ای

دارد و به همین دلیل انتخاب بسیار خوبی برای پردازش رایانه‌ای است ولی جئوپاردی! نه ساده و نه ساختارمند است. جئوپاردی! یک بازی است که برای آزمودن هوش و خلاقیت انسان طراحی شده است. از این رو، رایانه‌ای که برای شرکت در این بازی طراحی می‌شود باید یک سیستم رایانش شناختی باشد که بتواند همانند انسان فکر کرده و عمل کند. معنا دادن به عدم دقت ذاتی در زبان انسان کلید موفقیت در این کار بود.

در سال ۲۰۱۰، یک تیم پژوهشی آی‌بی‌ام یک سیستم رایانه‌های شگفت‌انگیز به نام واتسون را توسعه دادند که ترکیب نوینی از سخت‌افزار و نرم‌افزارهای پیشرفته بود که برای پاسخگویی به سوالات طراح شده به زبان طبیعی انسان طراحی شده بود. این تیم، واتسون را به عنوان بخشی از پروژه‌ی دیپ‌کیو‌ای^۱ توسعه داده و نام اولین مدیرعامل شرکت آی‌بی‌ام، توماس جی واتسون را بر آن نهادند. تیمی که واتسون را ایجاد کردند به دنبال یک چالش پژوهشی مهم بودند: سیستمی که بتواند دیپ بلو را به رقابت طلبیده و وابستگی واضح و شفافی با منابع کسب‌وکار آی‌بی‌ام داشته باشد. هدف در اینجا، پیشبرد علوم محاسباتی و رایانشی از طریق کشف روش‌های جدید برای فناوری‌های رایانه‌ای به منظور تأثیرگذاری گسترده بر علوم، کسب‌وکار و جامعه بوده است. بر همین اساس، تیم پژوهشی آی‌بی‌ام چالش ساخت واتسون به عنوان یک سیستم رایانه‌ای را که بتواند در سطح قهرمانی در برنامه جئوپاردی! با انسان رقابت کند را بر عهده گرفت. این تیم به دنبال ساخت یک مسابقه‌دهنده‌ی خودکار و بلادرنگ در نمایش تلویزیونی بود که قادر به گوش‌دادن، درک و پاسخگویی باشد. مورد کاربردی ۶.۸ جزئیات شرکت واتسون آی‌بی‌ام را در این نمایش تلویزیونی ارائه کرده است.

آی‌بی‌ام واتسون با بهترین‌ها در جئوپاردی! مسابقه می‌دهد

مورد کاربردی ۶.۸

در سال ۲۰۱۱، به منظور تست قابلیت‌های شناختی، واتسون در اولین رقابت انسان در برابر ماشین در نمایش مسابقاتی جئوپاردی! به رقابت پرداخت. در یک مسابقه‌ی دو مرحله‌ای با امتیاز ترکیبی، واتسون موفق به شکست برد روتر، بزرگترین برنده‌ی تمامی ادوار و کین چینیگز، رکورددار طولانی‌ترین قهرمانی (۷۵ روز) شد. در این اپیزودها، واتسون به خوبی توانست رقبای انسانی خود را بر روی دستگاه سیگنال‌دهی بازی شکست دهد ولی در پاسخگویی به چندین سوال از جمله آنهایی که سرنخ‌های کوتاهی که فقط از چند کلمه تشکیل شده بودند دچار مشکل شد. واتسون به ۲۰۰ میلیون صفحه محتوای ساخت‌یافته و ساخت‌نیافته دسترسی داشته و فضایی در حدود ۴ ترابایت را اشغال کرده بود. در طی بازی، واتسون به اینترنت متصل نبود. موفقیت در چالش جئوپاردی! نیازمند انواع مختلف فناوری‌های متن‌کاوی و پردازش زبان طبیعی از جمله دسته‌بندی سوالات، تجزیه‌ی سوالات، اکتساب و ارزیابی خودکار منبع، تشخیص موجودیت و روابط، تولید منطقی فرم و بازنمود و استنتاج دانش بوده است. پیروزی در جئوپاردی! نیازمند اطمینان محاسباتی دقیق در پاسخگویی بوده است. سوالات و محتوا، مبهم و دارای نویز بوده و هیچکدام از الگوریتم‌ها به تنهایی بی‌عیب و نقص نیستند. از این رو، هر کامپوننت باید یک سطح اطمینان در خروجی خود تولید کرده و این اطمینان‌ها با یکدیگر ترکیب شده و یک سطح اطمینان کلی را برای پاسخ نهایی محاسبه کند. از اطمینان‌هایی برای تعیین اینکه آیا سیستم رایانه‌ای باید انتخاب کرده و ریسک کند که به سوال پاسخ بدهد یا خیر. در جئوپاردی! از این اطمینان برای تعیین اینکه آیا رایانه برای یک سوال، "ring in" یا "buzz in" خواهد کرد استفاده می‌شود. اطمینان باید در طی زمانی که سوال خوانده می‌شود و پیش از آنکه فرصت buzz in وجود داشته باشد محاسبه شود. این زمان چیزی حدود یک تا شش ثانیه و به طور متوسط حدود ۳ ثانیه است.

¹ DeepQA

واتسون یک مثال عالی از توسعه‌ی سریع فناوری رایانشی و توانایی‌های آن بوده است. اگرچه خلاقیت آنها به اندازه‌ی انسان نیست و سیستم‌های رایانه‌های همانند واتسون در حال تکامل و تغییر دنیایی که در آن زندگی می‌کنیم هستند.

سوالاتی برای مورد کاربردی ۶.۸

۱. به نظر شما منحصربفردترین ویژگی‌های واتسون چه هستند؟
۲. در چه بازی‌های چالشی دیگری دوست دارید که رقابت واتسون را با انسان ببینید؟
۳. تشابه و تفاوت‌های بین هوش واتسون و هوش انسان چه هستند؟

Sources: Ferrucci, D., E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, D. Kalyanpur, A. Lally, J. Murdock, E. Nyberg, J. Prager, N. Schlaefel, and C. Welty. (2010). "Building Watson: An Overview of the DeepQA Project." *AI Magazine*, 31(3), pp. 59–79; IBM Corporation. (2011). "The DeepQA Project." https://researcher.watson.ibm.com/researcher/view_group.php?id=2099 (accessed May 2018).

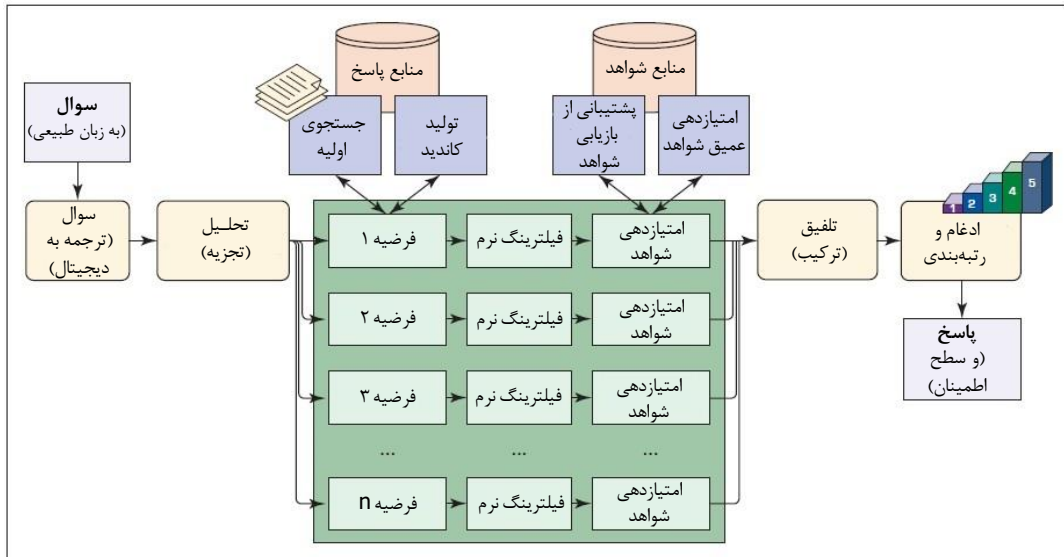
زیر نقاب واتسون چه چیزی قرار دارد؟ چگونه اینکار را می‌کند؟ سیستم پشت واتسون که دیپ‌کیو ای نام دارد یک معماری محاسباتی موازی، متمرکز بر متن کاوی و مبتنی بر شواهد و مدارک احتمالاتی است. برای چالش جنوپاردی، واتسون از بیش از ۱۰۰ تکنیک متفاوت برای تحلیل زبان طبیعی، شناسایی منابع، یافتن و تولید فرضیه‌ها، یافتن و رتبه‌بندی شواهد و مدارک و ادغام و رتبه‌بندی فرضیه‌ها استفاده کرده است. آنچه بیش از هر تکنیک ویژه‌ی دیگری که تیم آی‌بی‌ام بکار برده است این است که چگونه آنها را در دیپ‌کیو ای ترکیب کرده که روش‌هایی که با یکدیگر همپوشانی دارند بتوانند نقاط قوت خود را کنار یکدیگر قرار داده و موجب بهبود و افزایش دقت، اطمینان و سرعت شوند.

دیپ‌کیو ای یک معماری با متدولوژی همراه است که خاص چالش جنوپاردی! نبوده است. اصول کلی دیپ‌کیو ای به شرح زیر می‌باشند:

- **موازی‌کاری انبوه.** واتسون می‌خواست از موازی‌کاری انبوه در تفسیر و فرضیه‌های متعدد استفاده کند.
- **تعداد زیادی از افراد خبره.** واتسون می‌خواست قادر به یکپارچه‌سازی، بکارگیری و ارزیابی بافتاری طیف گسترده‌ای سوالات احتمالی و تحلیل محتوا باشد.
- **برآورد فراگیر اطمینان.** هیچ کامپوننتی از واتسون متعهد به پاسخگویی نبود بلکه تمامی کامپوننت‌ها، ویژگی‌ها و اطمینان‌های مربوطه، امتیازدهی تفسیرهای محتوا و سوالات مختلف را بر عهده داشتند. یک زیرلایه‌ی پردازش اطمینان یاد گرفت که چگونه امتیازها را جمع و ترکیب کند.
- **یکپارچه‌سازی دانش عمیق و سطحی.** واتسون می‌بایست در بکارگیری معناسناسی سطحی (کم عمق) و معناسناسی عمیق تعادل برقرار کرده و از تعداد زیادی از هستی‌شناسی‌های شکل گرفته شده بهره می‌برد.

شکل ۶.۳۸ معماری دیپ‌کیو ای را در یک سطح بسیار بالا به تصویر کشیده است. برای جزئیات بیشتر در مورد کامپوننت‌های معماری مختلف و نقش و قابلیت‌های خاص آنها می‌توانید به فروچی و همکاران (۲۰۱۰) مراجعه کنید.

چالش جئوپاردی! به آی بی ام کمک کرد تا به الزامات و نیازمندی‌هایی رسیدگی کند که منجر به طراحی معماری دیپ کیو ای و پیاده‌سازی واتسون گردید. پس از سه سال پژوهش سنگین و توسعه توسط تیم اصلی متشکل از حدود ۲۰ پژوهشگر و همچنین بودجه‌ی هنگفت تحقیق و توسعه، واتسون توانست از لحاظ دقت، اطمینان و سرعت در نمایش تلویزیونی جئوپاردی! در سطوح خبره‌ی انسانی ظاهر شود.



شکل ۶.۳۸ ترسیم سطح بالای معماری دیپ کیو ای

پس از نمایش، سوال بزرگ این بود که «خب حالا چه؟» آیا واتسون فقط برای یک نمایش تلویزیونی طراحی و ساخته شد؟ قطعاً خیر. نشان دادن کارهایی که واتسون قادر به انجام آنها بود به کل جهان، الهام‌بخش نسل بعدی سیستم‌های اطلاعاتی هوشمند شد. برای آی بی ام، طراحی و توسعه‌ی واتسون یک نمایش از آنچه بود که علوم محاسباتی و تحلیل‌های پیشرفته قادر به انجام آن بودند. پیام واضح بود: اگر یک ماشین هوشمند بتواند بهترین‌ها را در کاری که به بهترین نحو انجام می‌دهند شکست دهد، فکر کنید برای مشکلات و مسائل سازمان شما چه کارهایی می‌توانند انجام دهند.

فناوری‌های نوآورانه و آینده‌نگر که واتسون را تبدیل به تحسین شده‌ترین دستاور تکنولوژیکی این دهه کردند به عنوان زیربنای محاسباتی و رایانشی ابزارهای متعدد برای تحلیل داده‌های ساخت‌نیافته در مسائل پیش‌بینی مورد استفاده قرار می‌گیرند. این ابزارهای آزمایشی شامل تون آنالایزر^۱ و پرسونالیتی اینسایت^۲ هستند. این ابزارها با استفاده از محتوای بافتاری از توانایی پیش‌بینی نتایج برای رویدادهای اجتماعی پیچیده و رقابت‌های معروف جهانی برخوردارند.

After the show, the big question was “So what now?” Was developing Watson all

¹ Tone Analyzer

² Personality Insight

یکی از ابزارهایی که بر اساس واتسون آی بی ام توسعه یافت، واتسون تون آنالایزر نام دارد که از زبان‌شناسی محاسباتی برای تشخیص لحن در یک متن نوشتاری استفاده می‌کند. هدف بزرگتر آن این است که مدیران کسب‌وکار را مجاب کند تا از تون آنالایزر برای درک و فهم مکالمات و ارتباطات بین جوامع مشتریان هدف و پاسخگویی به موقع به نیازها و خواسته‌های آنها استفاده کنند. به عنوان مثال، یک نفر می‌تواند از این ابزارها برای پایش رسانه‌های اجتماعی و سایر محتوای مبتنی بر وب از جمله پست‌های، توییتهای، اظهارنظرها در خصوص محصولات و بوردهای بحث و گفتگو و همچنین مستندات طولانی‌تر مثل مقالات و پست‌های وبلاگ‌ها استفاده کند. یا شخصی می‌تواند از آن برای پایش تعاملات خدمات مشتریان و گفتگوها و مکالمات پشتیبانی مشتریان استفاده کند. اگرچه اینگونه به نظر می‌رسد که هر سیستم تشخیص متن دیگری را می‌توان بر اساس تحلیل احساسات ایجاد کرد ولی تون آنالایزر از این جهت با این سیستم‌ها تفاوت دارد که قادر به تحلیل و مشخص‌سازی محتوای بافتاری می‌باشد. واتسون تون آنالایزر به اندازه‌گیری عقاید و نظرات اجتماعی پرداخته و با استفاده از نسخه‌ای از Big-5 یعنی پنج گروه ویژگی‌های شخصیتی (باز و پذیرا بودن، موافق بودن، وظیفه‌شناس بودن، برون‌گرا بودن و روان‌رنجور بودن) به همراه سایر گروه‌بندی‌های عاطفی و احساسی به تشخیص لحن در یک محتوای بافتاری پرداخته است. به عنوان مثال، اسلویی (2017b) از واتسون تون آنالایزر آی بی ام برای پیش‌بینی برنده‌ی مسابقات آواز یوروویژن در سال ۲۰۱۷ پرداخت. اسلویی فقط با استفاده از اشعار مسابقات سال‌های گذشته الگویی را کشف کرد که می‌توانست اکثر برندگان را با سطوح بالای موافق بودن و وظیفه‌شناس بودن تشخیص دهد. نتایج نشان داد که کشور پرتغال برنده مسابقه خواهد بود و دقیقاً همین اتفاق افتاد. خودتان امتحان کنید:

- به آدرس واتسون تون آنالایزر بروید (<https://tone-analyzer-demo.ng.bluemix.net>).
- متن خودمان را در فیلد ورود متن وارد کنید.
- بر روی "تحلیل کن" کلیک کنید.
- خلاصه نتایج و همچنین جملات خاصی که لحن‌های خاص در آنها بسیار قوی است را مشاهده کنید.

یکی دیگر از ابزارهای آی بی ام واتسون که بر اساس زبان‌شناسی ساخته شده است واتسون پرسونالیتی اینسایت نام دارد که بسیار مشابه واتسون تون آنالایزر کار می‌کند. در یک مورد کاربردی سرگرمی دیگر، اسلویی (2017a) از واتسون پرسونالیتی اینسایت برای پیش‌بینی برنده‌ی بهترین فیلم در جوایز اسکار ۲۰۱۷ استفاده کرد. با استفاده از متن (فیلمنامه) فیلم‌های سال‌های گذشته، اسلویی یک شرح‌حال تعمیم یافته از برندگان را ایجاد کرده و سپس آنها را با شرح حال فیلم‌هایی که جدیداً نامزد دریافت جایزه شده بودند مقایسه کرد تا بتواند برنده‌ی آتی را شناسایی کند. اگرچه در این مورد خاص، اسلویی فیلم *Hidden Figures* را به اشتباه به عنوان برنده پیش‌بینی کرد ولی از یک متدولوژی منحصر‌بفرد و نوآورانه برای اینکار استفاده کرد که شایسته‌ی تقدیر بود. برای اینکه واتسون پرسونالیتی اینسایت را خودتان امتحان کنید فقط به صفحه‌ی

<https://personality-insights-demo.ng.bluemix.net/> رفته و محتوای بافتاری خودتان را بر بخش "Body of Text" وارد کرده و نتیجه را مشاهده کنید.

یکی از ارزشمندترین تلاش‌های واتسون (یک سیستم‌های رایانشی شناختی بزرگ در مقیاس واتسون) کمک به پزشکان و دیگر متخصصان پزشکی برای تشخیص بیماری و شناسایی بهترین گزینه‌های درمان برای هر یک از بیماران است. اگرچه واتسون در این حوزه جدید است اما در دنیای رایانش این موضوع چیز جدیدی نیست. در اوایل دهه ۷۰، چندین پژوهشگر در دانشگاه استانفورد یک سیستم رایانه‌ای به نام MYCN را برای شناسایی باکتری‌هایی که موجب بروز عفونت‌های شدید مثل مننژیت و

باکتری می‌شدند و برای توصیه‌ی آنتی‌بیوتیک‌ها با مقدار مناسب برای هر بیمار توسعه دادند (بوشانان آند شورت‌لایف ۱۹۸۴). این تلاش شش ساله متکی بر یک سیستم خبره‌ی قاعده‌محور که یک نوع سیستم هوش مصنوعی است بود که دانش تشخیص و درمان از طریق تعداد کثیری از خبرگان (یعنی پزشکان با تجربه‌ی بالا در حوزه‌ی پزشکی تخصصی) بدست می‌آمد. سیستم بدست آمده سپس بر روی بیماران جدید آزمایش می‌شد و عملکرد آن با نتایج سایر پزشکان مجرب که به عنوان منابع دانشی/خبرگان مقایسه می‌شد. نتایجی که برای MYCN مطلوب بودند، نشانه‌ی واضحی از این بود که سیستم‌های رایانه‌ی مبتنی بر هوش مصنوعی که به خوبی طراحی و پیاده‌سازی شده باشند می‌توانند کارایی و اثربخشی مطلوب و حتی بالاتر از بهترین خبرگان پزشکی را از خود نشان می‌دهند. پس از بیش از ۴ دهه، واتسون اکنون سعی دارد تا مأموریت MYCN برای استفاده از سیستم‌های رایانه‌ای هوشمند برای بهبود سلامت و رفاه انسان‌ها را از جایی که رها کرده است ادامه دهد و برای اینکار به دنبال کمک به پزشکان از طریق ارائه اطلاعات بافتاری برای تشخیص و درمان بهتر سریع‌تر و بهتر بیماران است. اولین صنعتی که از واتسون در آن استفاده شده است، مراقبت سلامت و پس از آن بخش امنیت، مالی، خرده‌فروشی، آموزش، خدمات عمومی و پژوهش است. در بخش بعدی، توضیحات کوتاهی از آنچه واتسون می‌تواند در این صنایع انجام دهد آورده شده است.

چالش‌هایی که بخش مراقبت سلامت امروزه با آنها مواجه است بسیار بزرگ و دارای وجوه متعددی هستند. با افزایش سن جمعیت ایالات متحده که ممکن است تا حدودی مربوط به شرایط بهتر زندگی و اکتشافات پزشکی به کمک نوآوری‌های تکنولوژیکی باشد، تقاضا برای خدمات مراقبت سلامت بیش از تأمین منابع در حال افزایش است. همانگونه که می‌دانیم، زمانیکه عدم توازن بین عرضه و تقاضا وجود داشته باشد، قیمت‌ها افزایش یافته و کیفیت کاهش می‌یابد. از این رو، ما برای کمک به تصمیم‌گیرندگان در بهینه‌سازی استفاده از منابع‌شان در محیط‌های بالینی و مدیریت به سیستم‌های شناختی مثل واتسون نیاز داریم.

بر اساس آمار خبرگان مراقبت سلامت، فقط ۲۰ درصد از دانشی که پزشکان برای تشخیص و درمان بیماران استفاده می‌کنند مبتنی بر شواهد است. با در نظر گرفتن اینکه حجم اطلاعات پزشکی در دسترس هر پنج سال دو برابر می‌شود و اکثر این داده‌ها ساخت‌نیافته هستند، پزشکان زمان کافی برای خواندن هر مجله‌ای که بتواند آخرین دستاوردها و پیشرفت‌ها را در اختیارشان قرار دهد ندارند. با توجه به تقاضای روزافزون برای خدمات و پیچیدگی تصمیمات پزشکی، ارائه دهندگان خدمات مراقبت سلامت چگونه می‌توانند به این مسائل رسیدگی کنند؟ پاسخ می‌تواند استفاده از واتسون یا سیستم‌های رایانشی مشابه باشد که توانایی کمک به پزشکان در تشخیص و درمان بیماران را -هم از طریق تحلیل حجم انبوه داده‌های ساخت‌یافته که از پایگاه داده‌های سوابق پزشکی الکترونیکی بدست می‌آیند و متون ساخت‌نیافته که از یادداشت‌های پزشکان و ادبیات منتشر شده می‌آیند - برای ارائه‌ی شواهد و مدارک برای تصمیم‌گیری بهتر و سریع‌تر دارند. اول، پزشک و بیمار می‌توانند نشانه‌ها و سایر عوامل مربوط به سیستم را به زبان طبیعی توضیح دهند. واتسون می‌تواند سپس بخش‌های اطلاعاتی مهم را شناسایی کرده و داده‌های بیمار را برای یافتن مرتبط‌ترین واقعیت‌ها در مورد تاریخچه‌ی خانوادگی، داروهای جاری و سایر شرایط موجود مورد کاوش قرار دهد. سپس می‌تواند اطلاعات را با یافته‌های جاری از آزمون‌ها ترکیب کرده و سپس فرضیه‌ها را از طریق ارزیابی منابع داده‌ای مختلف - دستورالعمل‌ها و راهنماهای درمان، داده‌های سوابق پزشکی الکترونیکی، یادداشت‌های پزشکان و پرستارها و مطالعات بالینی و پژوهش‌های داوری همکار- تشکیل داده و تست کند. سپس، واتسون می‌تواند گزینه‌های تشخیص و درمان بالقوه را با یک رتبه‌بندی اطمینان برای هر گزینه توصیه کند.

واتسون همچنین از طریق تلفیق یافته‌های پژوهشی پراکنده که منتشر شده‌اند از توان بالوقه برای تغییر و تحول مراقبت سلامت برخوردار است. واتسون می‌تواند روش یادگیری دانشجویان پزشکی را به طور چشمگیری تغییر دهد. به مدیران مراقبت سلامت می‌تواند کمک کند تا رویکرد کنشگرانه‌ای را در خصوص الگوهای تقاضای پیش‌رو اتخاذ کرده، منابع را به شکل بهینه تخصیص داده و رسیدگی به پرداخت‌ها را بهبود بخشند. مثال‌های اولیه از ارائه‌دهندگان پیشگام در مراقبت سلامت که از سیستم‌های شناختی همچون واتسون استفاده می‌کنند شامل ام‌دی اندرسون^۱، کلیولند کلینیک^۲ و مموریال اسلون کترینگ هستند.

¹ MD Anderson
² Cleveland Clinic

رایانه‌ها و شبکه‌ها، به‌ویژه در زمینه‌های امنیتی و حفاظتی، به‌کار می‌روند. این سیستم‌ها می‌توانند به تشخیص و ردیابی تهدیدات امنیتی کمک کنند. همچنین، در زمینه‌های آموزشی و پژوهشی، به‌کار می‌روند. این سیستم‌ها می‌توانند به بهبود کیفیت آموزش و پژوهش کمک کنند. در نهایت، در زمینه‌های اقتصادی و اجتماعی، به‌کار می‌روند. این سیستم‌ها می‌توانند به بهبود کیفیت زندگی و رفاه اجتماعی کمک کنند.

در ادامه، به بررسی کاربردهای دیگر سیستم‌های هوش مصنوعی در زمینه‌های مختلف می‌پردازیم. در زمینه پزشکی، سیستم‌های هوش مصنوعی می‌توانند به تشخیص و درمان بیماری‌ها کمک کنند. در زمینه کشاورزی، سیستم‌های هوش مصنوعی می‌توانند به بهبود عملکرد و بهره‌وری کشاورزی کمک کنند. در زمینه حمل و نقل، سیستم‌های هوش مصنوعی می‌توانند به بهبود ترافیک و کاهش آلودگی کمک کنند. در نهایت، در زمینه‌های اجتماعی و فرهنگی، سیستم‌های هوش مصنوعی می‌توانند به بهبود کیفیت زندگی و رفاه اجتماعی کمک کنند.

در ادامه، به بررسی کاربردهای دیگر سیستم‌های هوش مصنوعی در زمینه‌های مختلف می‌پردازیم. در زمینه پزشکی، سیستم‌های هوش مصنوعی می‌توانند به تشخیص و درمان بیماری‌ها کمک کنند. در زمینه کشاورزی، سیستم‌های هوش مصنوعی می‌توانند به بهبود عملکرد و بهره‌وری کشاورزی کمک کنند. در زمینه حمل و نقل، سیستم‌های هوش مصنوعی می‌توانند به بهبود ترافیک و کاهش آلودگی کمک کنند. در نهایت، در زمینه‌های اجتماعی و فرهنگی، سیستم‌های هوش مصنوعی می‌توانند به بهبود کیفیت زندگی و رفاه اجتماعی کمک کنند.

سوالاتی برای مرور بخش ۶.۱۰

۱. رایانش شناختی چیست و چه تفاوتی با پارادایم‌های رایانشی دیگر دارد؟

- نرم‌افزار شبکه عصبی امکان آزمایش ساده‌ی مدل‌های متعدد را فراهم می‌کند. اگرچه ماژول‌های شبکه‌های عصبی در تمامی ابزارهای نرم‌افزاری داده‌کاوی مهم گنجانده شده‌اند، بسته‌های نرم‌افزاری شبکه عصبی خاص نیز در بازار وجود دارد.
- شبکه‌های عصبی تقریباً در تمامی رشته‌های کسب‌وکار و حوزه‌های عملکردی و وظیفه‌ای کاربرد دارند.
- بیش‌برازش زمانی رخ می‌دهد که شبکه‌های عصبی برای دفعات تکرار بسیار زیاد با مجموعه داده‌های نسبتاً کوچک آموزش داده می‌شوند. برای پیشگیری از برازش، فرایند آموزش توسط یک فرایند ارزیابی با استفاده از یک مجموعه داده‌ی اعتبارسنجی مجزا کنترل می‌شود.
- شبکه‌های عصبی را با نام مدل‌های جعبه سیاه می‌شناسند. اغلب از تحلیل حساسیت برای روشنگری این جعبه‌ی سیاه برای ارزیابی اهمیت نسبی ویژگی‌های ورودی استفاده می‌شود.
- شبکه‌های عصبی عمیق، مفهوم و قاعده‌ی پذیرفته شده‌ی «دو لایه‌ی پنهان برای فرموله کردن مسائل پیش‌بینی پیچیده کافی است» را شکستند. آنها استفاده از لایه‌های پنهان بیشتر را برای بازنمود بهتر پیچیدگی‌های مجموعه داده‌ها ترویج می‌کنند.
- شبکه‌های عمیق MLP که به نام شبکه‌های پیشخور عمیق نیز شناخته می‌شوند عمومی‌ترین نوع شبکه‌های عمیق هستند.
- تأثیر اوزان تصادفی در فرایند یادگیری MLP عمیق مسئله بسیار مهمی است. به نظر می‌رسد تخصیص غیرتصادفی اوزان اولیه موجب بهبود چشمگیر فرایند یادگیری در MLP عمیق شود.
- اگرچه هیچ مبنای نظری پذیرفته شده‌ای برای این موضوع وجود ندارد، باورها بر این است و از لحاظ تجربی نیز نشان داده شده است که در شبکه‌های MLP عمیق، تعداد لایه‌های زیاد، بهتر و سریع‌تر از تعداد لایه‌های اندک با تعداد نورون‌های زیاد عمل می‌کنند.
- CNNها (شبکه‌های عصبی پیچشی) بدون شک محبوب‌ترین و موفق‌ترین روش‌های یادگیری عمیق هستند.
- CNNها در ابتدا برای کاربردهای بینایی رایانه‌ای (مثلاً پردازش تصویر، پردازش ویدئو، تشخیص متن) طراحی شدند ولی در مجموعه داده‌های غیرتصویری و غیرمتنی نیز کاربردهایی داشته‌اند.
- مشخصه‌ی اصلی شبکه‌های پیچشی در این است که حداقل یک لایه دارند که در آن از تابع وزنی پیچش به جای ضرب ماتریسی معمول استفاده شده است.
- تابع پیچش روشی برای حل مشکل تعداد پارامترهای وزنی بسیار زیاد در شبکه است که از مفهوم به اشتراک‌گذاری پارامتر استفاده می‌کند.
- در CNNها پس از یک لایه‌ی پیچش، لایه‌ی دیگری به نام ادغام وجود دارد. هدف لایه‌ی ادغام این است که عناصر ماتریس ورودی را برای تولید یک ماتریس خروجی کوچکتر، تجمیع و یکپارچه کرده و در عین حال، ویژگی‌های مهم را حفظ نماید.
- ایمیج‌نت یک پروژه‌ی پژوهشی در حال انجام است که یک پایگاه داده‌ی بزرگ از تصاویر را در اختیار پژوهشگران قرار می‌دهد که هر کدام از آن تصاویر به مجموعه‌ای از واژگان مترادف (به نام *synset*) در ووردنت (یک پایگاه داده‌ی سلسله‌مراتب واژگان) متصل است.

GPU	.61	
GPU	.62	
MLP	.63	
MLP	.64	
CNN	.65	
CNN	.66	
CNN	.67	
CNN	.68	
CNN	.69	
CNN	.70	
CNN	.71	
CNN	.72	
word2vec	.73	
CNN	RNN	.74
RNN	.75	
RNN	.76	
RNN	LSTM	.77
LSTM	.78	
LSTM	.79	
LSTM	.80	
CPU, NVIDIA, CUDA	.81	
CPU, NVIDIA, CUDA	.82	
CPU, NVIDIA, CUDA	.83	
CPU, NVIDIA, CUDA	.84	
CPU, NVIDIA, CUDA	.85	
CPU, NVIDIA, CUDA	.86	
CPU, NVIDIA, CUDA	.87	
CPU, NVIDIA, CUDA	.88	
CPU, NVIDIA, CUDA	.89	
CPU, NVIDIA, CUDA	.90	
CPU, NVIDIA, CUDA	.91	
CPU, NVIDIA, CUDA	.92	

۵. teradatauniversitynetwork.com

۶. [knime.org](http://www.knime.org)

۷. [teradata.com](http://www.teradata.com)

۸. [knime.org](http://www.knime.org)

۹. [knime.org](http://www.knime.org)

۱۰. [knime.org](http://www.knime.org)

۱۱. [Teradata.com](http://www.teradata.com)

۱۲. [knime.org](http://www.knime.org)

۱۲	SAS.com
۱۳	IBM.com
۱۴	TIBCO.com

در این فصل مرور جامعی بر تحلیل/کاوش متن و تحلیل/کاوش وب به همراه حوزه‌های کاربردی متداول آنها از قبیل موتورهای جستجو، تحلیل احساسات و تحلیل‌های رسانه‌ها/شبکه‌های اجتماعی خواهیم پرداخت. همانطور که در سال‌های اخیر شاهد بوده‌ایم، داده‌های ساخت‌نیافته‌ای که از طریق اینترنت اشیاء (وب، شبکه‌های حسگرها، سیستم‌های زنجیره تأمین مجهز به شناسایی فرکانس

۷.۱. پرونده‌ی آغازین: آمادوری گروپ احساسات مشتریان را به فرصت فروش قریب‌الوقوع تبدیل می‌کند.	۷۰۱
۷.۲. مروری بر تحلیل متن و متن کاوی	۷۰۲
۷.۳. پردازش زبان طبیعی (ان‌ال‌پی)	۷۰۳
۷.۴. کاربردهای متن کاوی	۷۰۴
۷.۵. فرایند متن کاوی	۷۰۵
۷.۶. تحلیل (آنالیز) احساسات	۷۰۶
۷.۷. مروری بر وب کاوی	۷۰۷
۷.۸. موتورهای جستجو	۷۰۸
۷.۹. کاربرد کاوی وب	۷۰۹
۷.۱۰. تحلیل اجتماعی	۷۱۰

رادیویی، شبکه‌های نظارتی و غیره) تولید شده‌اند با سرعت چشمگیری در حال افزایش هستند و از سرعت آن کاسته نخواهد شد. این ماهیت متغیر داده‌ها سازمان‌ها را مجبور کرده است تا از تحلیل متن و وب به عنوان جزء لاینفک زیرساخت هوش/تحلیل کسب‌وکارشان استفاده کنند.

- ۷.۱. پرونده‌ی آغازین: آمادوری گروپ احساسات مشتریان را به فرصت فروش قریب‌الوقوع تبدیل می‌کند.
- ۷.۲. مروری بر تحلیل متن و متن کاوی
- ۷.۳. پردازش زبان طبیعی (ان‌ال‌پی)
- ۷.۴. کاربردهای متن کاوی
- ۷.۵. فرایند متن کاوی
- ۷.۶. تحلیل (آنالیز) احساسات
- ۷.۷. مروری بر وب کاوی
- ۷.۸. موتورهای جستجو
- ۷.۹. کاربرد کاوی وب
- ۷.۱۰. تحلیل اجتماعی

آمدوری گروه^۱ یا به ایتالیایی گروه پوپو آمدوری یک شرکت پیشگام ساخت و تولید در ایتالیا است که به تولید و بازاریابی محصولات و فرآورده‌های غذایی می‌پردازد. مقرر اصلی این شرکت در سان ویتوره دی چسنا^۲ ایتالیا است و بیش از ۷,۰۰۰ کارمند و ۱۶ کارخانه تولیدی دارد. آمدوری به دنبال گسترش بازارهای خود و همسو شدن با تغییرات سبک زندگی و نیازهای رژیمی افراد جوان بین ۲۵ تا ۳۵ سال بود. این شرکت به فکر خلق روش‌های بامزه برای مشارکت دادن مشتریان هدف از طریق بهره‌برداری از توان بالقوه‌ی بازاریابی آنلاین و رسانه‌های اجتماعی افتاد. شرکت به دنبال ارتقاء شفافیت برند، وفاداری مشتری و سنجش واکنش مشتریان به کمپین‌های بازاریابی و محصولات بود.

آمدوری به همراه تکلا^۳ (یک شرکت کسب‌وکار دیجیتال) از پورتال آی‌بی‌ام وب اسفیر^۴ و نرم‌افزار مدیریت محتوای وب آی‌بی‌ام^۵ برای خلق و مدیریت محتوای تعاملی برای چهار مینی وب‌سایت استفاده کرد که به تبلیغ محصولات آماده‌ی طبخ که متناسب با ترجیحات و سبک زندگی جوانان است می‌پردازد. به عنوان مثال، این شرکت برای بازاریابی محصول سوسیس جدیدش به نام اویوا^۶ یک مینی سایت به نام « Evviva Il Würstel Italiano » را ایجاد کرد و به مشتریان اجازه داد تا تصاویر و ویدئوهای خودشان را از حضور در رویدادهایی که توسط شرکت آمدوری ترتیب داده شده بود بارگذاری کنند. به منظور ترویج مشارکت، جایگاهی را در کمپین تبلیغاتی بعدی خود برای برنده در نظر می‌گرفت.

پرسنل حوزه بازاریابی آمدوری به کمک این طرح و سایر کمپین‌های تبلیغاتی از بازدیدکنندگان مینی‌سایت‌ها خواستند تا داده‌ها را به اشتراک گذاشته، برنامه‌های کاربردی را دانلود کرده، خبرنامه‌های دوره‌ای را دریافت کرده و برای رویدادها نام‌نویسی کنند و از این طریق یک پایگاه داده از شرح حال مشتریان تهیه کنند. علاوه بر این، شرکت از فناوری فیسبوک اینسایت^۷ برای دستیابی به سنج‌ها بر روی صفحه فیسبوک خود استفاده کرد که شامل تعداد طرفدارهای جدید و محتوای مورد علاقه آنها بود.

¹ Amadori Group

² Vittore di Cesena

³ Tecla

⁴ IBM WebSphere Portal

⁵ IBM Web Content Manager

⁶ Evviva

⁷ Facebook Insight

این شرکت برای کمک به ارزیابی نظرات مردم در خصوص محصولات و نتیجه‌گیری در مورد هرگونه نوسان در محبوبیت برند آمادوری در بین مشتریان خود بر روی نرم‌افزار آی‌بی‌ام‌اس‌پی‌اس‌اس دیتا کالکشن^۱ سرمایه‌گذاری کرده است. به عنوان مثال، به محض معرفی محصولات جدید اویوا و سنجش واکنش مشتریان نسبت به طرح‌های تبلیغاتی محصولات و بازاریابی، تبلیغات تلویزیونی و تور گردش ساحلی، سیلی از گفتگوهای مشتریان به سمت آمادوری سرازیر شد.



پلتفرم دیجیتالی تعالمی از گردآوری سریع و دقیق داده‌ها از شرکای کسب‌وکار و مشتریان پشتیبانی می‌کند.	ابزارمند	
پلتفرم دیجیتالی همچنین یک دید یکپارچه از فرایندهای سرتاسری سازمان از برنامه تولید گرفته تا بازاریابی و فروش ارائه می‌کند.	متصل	
مدیریت محتوا، گردآوری داده‌ها و برنامه‌های تحلیل پیشگویانه به پایش و تحلیل رسانه‌های اجتماعی مرتبط با برند آمادوری پرداخته و به شرکت در پیش‌بینی مسائل و همراستایی بهتر محصولات و تبلیغات بازاریابی با نیازها و خواسته‌های مشتریان کمک می‌کند.	هوشمند	

شکل ۷.۱ تجارت هوشمندتر – بهبود مشارکت مشتری از طریق تحلیل‌ها

آمدوری اظهارنظرهای مشتریان در مورد محصول خود را از وب‌سایت و شبکه‌های رسانه‌های اجتماعی با استفاده از قابلیت تحلیل احساسات نرم‌افزار آی‌بی‌ام‌اس‌پی‌اس‌اس اخذ کرده و آن را با موفقیت با فعالیت‌های بازاریابی خود به نحو بلادرنگ تنظیم نمود. این نرم‌افزار فقط وابسته به جستجوی واژگان کلیدی نبوده بلکه قادر است تا نحو و ترکیب زبان‌ها، دلالت‌ها و حتی زبان محاوره را برای کشف الگوهای گفتاری پنهان مورد تجزیه و تحلیل قرار دهد تا از این طریق بتواند مشخص کند که آیا اظهارنظرها در مورد شرکت یا محصولات آمادوری بیانگر عقیده و نظر مثبت، منفی یا خنثی هستند. شکل ۷.۱ سه جنبه‌ی تحلیل‌های تجاری آمادوری را برای بهبود مشارکت مشتریان نشان می‌دهد.

آمدوری بر اساس موفقیتی که در بازاریابی مینی‌سایت‌ها بدست آورد یک وب‌سایت سازمانی جدید را بر روی همان پورتال آی‌بی‌ام‌ و فناوری مدیریت محتوا ایجاد کرد. شرکت در حال حاضر تمرکز خود را بر جذب بازدیدکنندگان به وب‌سایت جدید سازمانی معطوف کرده است. آمادوری به جای مینی‌سایت‌های مجزا، بخش‌هایی را با الگوها و گرافیک متفاوت در درون وب‌سایت

¹ IBM SPSS Data Collection

تعبیه کرده و یک رابط کاربر خاص را برای هر طرح تبلیغاتی یا بازاریابی خاص در نظر گرفته است. به گفته‌ی فابری، « به عنوان مثال، ما محصول جدیدی را عرضه می‌کنیم که از جوجه‌های ارگانیک و نگهداری شده در فضای آزاد تهیه شده‌اند، به عنوان بخشی از طرح بازاریابی، ما امکان مشاهده از طریق وب کم را در بخش جدیدی از سایت سازمانی مان قرار داده‌ایم تا بازدیدکنندگان بتوانند نحوه‌ی زندگی و رشد ماکیان را ببینند. ما یک گرافیک جدید ایجاد کردیم ولی آدرس وب‌سایت ما و پانویس و سرنویس سایت تغییری نکرده است تا بازدیدکنندگان متوجه شوند که همیشه در سایت آمادوری هستند.»

بازدیدکنندگان می‌توانند از بخشی به بخش دیگر بروند؛ زمان طولانی‌تری را در سایت بمانند و در مورد محصولات ارائه شده مطلع شوند. با افزودن هفتگی محتوای جدید، سایت آمادوری بزرگ‌تر شده و شهرت و اعتبار بیشتری در گوگل و سایر موتورهای جستجو کسب کرده است. فابری می‌گوید: « یکسال بعد از پیاده‌سازی وب‌سایت، ترافیک سایت به ۲۴۰,۰۰۰ بازدیدکننده با ۳۰٪ کاربر وفادار افزایش یافت.»

با افزایش محتوا و حجم ترافیک وب بسیار مهم است که بازدیدکنندگان بتوانند به آسانی آنچه را که می‌خواهند صرف‌نظر از نحوه‌ی دسترسی به سایت آمادوری پیدا کنند. تیم پروژه‌ی آمادوری با داشتن این هدف در ذهن، یک دسته‌بندی محتوا را بر اساس نقش و حوزه‌ی مورد نظر ایجاد کردند. به عنوان مثال، زمانیکه مردم از وب‌سایت آمادوری بازدید می‌کنند، آنها یک بنر را می‌بینند که از آنها دعوت می‌کند تا « محتوا را مجدداً سازماندهی کنند». آنها می‌توانند خودشان را در جایگاه مصرف‌کننده، خریداران یا روزنامه‌نگار/وبلاگ‌نویس قرار داده و سطح علاقه خود به اطلاعات سازمانی، آشپزی و/یا سرگرمی را تعیین کنند. محتوایی که بر روی سایت نمایان می‌شود بر اساس این انتخاب‌ها در هر لحظه تغییر می‌کند. فابری می‌گوید: « اگر بازدیدکننده خودش را در جایگاه یک خریدار حرفه‌ای علاقمند به اطلاعات سازمانی قرار دهد، آیکن‌هایی که در قسمت بالای صفحه نمایش می‌بیند او را دعوت می‌کند تا یک کاتالوگ دیجیتالی محصول را به صورت آنلاین مشاهده کرده یا اینکه فایل پی‌دی‌اف را دانلود کند. در همان بخش صفحه یک مصرف‌کننده‌ی علاقمند به آشپزی، آیکنی را می‌بیند که در بین صفحات دستور پخت یا تهیه‌ی غذاهای مختلف که از محصولات آمادوری استفاده می‌کنند جابجا می‌شود.» پروژه‌های تحلیل‌های پیشرفته‌ی آمادوری، مزایای چشمگیری را برای کسب‌وکار فراهم کرده‌اند و شرکت را قادر ساخته‌اند تا استفاده‌های نوآورانه‌تری از داده‌های اجتماعی داشته باشند.

- افزایش توانایی شرکت برای پایش و یادگیری سلامت برند با استفاده از تحلیل احساسات به میزان ۱۰۰ درصد.
- ارتقاء حضور شرکت در رسانه‌های اجتماعی به میزان ۱۰۰ درصد با استفاده از بینش‌های بازاریابی بلادرنگ و دستیابی به ۴۵,۰۰۰ طرفدار در فیسبوک در کمتر از یکسال.
- برقراری ارتباط مستقیم با بخش هدف از طریق یکپارچه‌سازی وب با رسانه‌های اجتماعی.
- افزایش فروش از طریق تسهیل طرح‌های تبلیغاتی به موقع از قبیل eCoupons.

همانطور که در این مورد مشاهده می‌کنید، در این عصر اینترنت و رسانه‌های اجتماعی، شرکت‌های مشتری‌محور در رقابت برای برقراری بهتر ارتباط با مشتریان خود برای دستیابی به درک عمیق نیازها، خواسته‌ها، دوست داشتن‌ها و نداشتن‌های آنها هستند. تحلیل‌های اجتماعی که مبتنی بر رسانه‌های اجتماعی هستند – هم داده‌های مرتبط با محتوا و هم شبکه‌های اجتماعی را ارائه می‌کنند – شرکت‌ها را قادر می‌سازد تا به بینش عمیق‌تری نسبت به گذشته دست پیدا کنند.

سوالاتی برای مرور پرونده‌ی آغازین

۲۹. بر اساس این پرونده‌ی آغازین و به عقیده‌ی خودتان، صنعت غذایی امروزه با چه چالش‌هایی روبرو است؟
۳۰. تحلیل‌ها چگونه می‌توانند به بقاء و تلاش کسب‌وکارها در بازار رقابتی صنعت غذایی کمک کنند؟
۳۱. اهداف اصلی آمادوری برای بکارگیری تحلیل‌ها چه بوده و هستند؟
۳۲. آیا کسب‌وکارهای دیگری را در صنایع غذایی می‌شناسید که از تحلیل‌ها برای رقابتی‌تر شدن و تمرکز بیشتر بر مشتری استفاده کنند؟ در صورت منفی بودن پاسخ، در اینترنت می‌توانید به جستجوی اطلاعات مرتبط و مفید برای این سوال بپردازید.

می‌توانیم اینگونه بگوییم که فناوری رایانه‌ای هم از منظر سخت‌افزار و هم نرم‌افزار سریع‌تر از هر چیز دیگری در ۵۰ سال اخیر پیشرفت داشته است. چیزهایی که بسیار بزرگ، بسیار پیچیده و حمل آنها ناممکن بود امروزه با راحتی در دامنه‌ی دسترسی فناوری اطلاعات قرار دارند. یکی از فناوری‌های توانمندساز را شاید بتوان تحلیل متن/متن‌کاوی و مشتقات آن به نام *تحلیل احساسات*^۱ دانست. سابق بر این، ما پایگاه‌های داده را برای ساختاردهی به داده‌ها به منظور آماده‌سازی برای پردازش توسط رایانه‌ها ایجاد می‌کردیم. از طرف دیگر، پردازش محتوای بافتاری همواره در حوزه توانمندی انسان بوده است. آیا ماشین‌ها قادر به انجام کارهایی که فقط با هوش و خلاقیت انسان‌ها انجام می‌شوند هستند؟ مشخصاً بله! این مورد امکان‌پذیری و گزاره‌ی ارزشی حاصل از گردآوری و پردازش عقاید و نظرات مشتریان برای توسعه و بهبود محصولات و خدمات، مدیریت برند تجاری شرکت و درگیرسازی و انرژی دادن به مشتریان برای برقراری ارتباطات نزدیک‌تر و سودمند متقابل را به تصویر کشیده است. آمادوری تحت نام «بازاریابی دیجیتالی» استفاده از متن‌کاوری، تحلیل احساسات و تحلیل رسانه‌های اجتماعی را برای پیشبرد سود و زیان شرکت از طریق افزایش رضایت مشتریان، افزایش فروش و ارتقاء وفاداری به برند به تصویر کشیده است.

Sources: IBM Customer Case Study. "Amadori Group Converts Consumer Sentiments into Near-Real-Time Sales."

Used with permission of IBM.

یکی از ویژگی‌های مهم عصر اطلاعات که ما در آن زندگی می‌کنیم، رشد سریع حجم داده‌ها و اطلاعات گردآوری شده، ذخیره شده و موجود در قالب الکترونیک است. اکثریت داده‌های کسب‌وکارها در قالب مستندات متنی که تقریباً ساخت‌نیافته هستند ذخیره می‌شوند. بر اساس مطالعه‌ای که توسط مریل لینیچ و گارتنر انجام شده، ۸۵ درصد از کل داده‌های سازمان‌ها به شکلی ساخت‌نیافته اخذ و ذخیره می‌شوند (مک‌نایت ۲۰۰۵). همین مطالعه همچنین نشان داده است که حجم این داده‌های ساخت‌نیافته هر ۱۸ ماه دو برابر می‌شود. از آنجائیه دانش در دنیای کسب‌وکار و دانش امروزی که قدرت محسوب می‌شود از داده‌ها و اطلاعات استخراج می‌شود، کسب‌وکارهایی که بتوانند به نحوی اثربخش و کارا از منابع داده‌های متنی خود استفاده کنند باید از دانش لازم برای اتخاذ تصمیمات بهتر برخوردار باشند تا منجر به یک مزیت رقابتی نسبت به سایر کسب‌وکارها شود. اینجاست که نیاز به تحلیل متن و متن‌کاوی در تصویر بزرگ کسب‌وکارهای امروزی قرار می‌گیرد.

¹ Sentiment Analysis

اگرچه هدف *تحلیل متن*^۱ و *متن‌کاوی*^۲ این است که داده‌های بافتاری ساخت‌نیافته را از طریق بکارگیری پردازش زبان طبیعی و تحلیل‌ها به اطلاعات اقدام‌پذیر تبدیل کنند، اگرچه تعاریف این عبارات به عقیده‌ی برخی افراد خبره و کارشناس این حوزه تا حدودی متفاوت است. از نظر آنها، «*تحلیل متن*» مفهوم وسیع‌تری است که شامل بازیابی اطلاعات (مثلاً جستجو و شناسایی اسناد مرتبط برای یک مجموعه عبارات مهم) و همچنین استخراج اطلاعات، داده‌کاوی و وب‌کاوی است در حالیکه «*متن‌کاوی*» عمدتاً بر کشف دانش و اطلاعات مفید از منابع داده‌های بافتاری متمرکز است. شکل ۷.۲ ارتباطات بین تحلیل متن و متن‌کاوی را با سایر حوزه‌های کاربردی مرتبط نشان می‌دهد. در قسمت پایین شکل ۷.۲ اصول اصلی (زیربنای خانه) که نقش مهمی را در توسعه‌ی این حوزه‌های کاربردی معروف‌تر و محبوب‌تر ایفا می‌کنند نشان داده شده است. بر اساس این تعریف از تحلیل متن و متن‌کاوی، به آسانی می‌توان این دو را از یکدیگر متمایز کرد:

تحلیل متن = بازیابی اطلاعات + استخراج اطلاعات + داده‌کاوی + وب‌کاوی

یا به بیان ساده

تحلیل متن = بازاریابی اطلاعات + متن‌کاوی

در مقایسه با *متن‌کاوی*، *تحلیل متن* عبارت نسبتاً جدیدی است. با تأکید اخیر که بر تحلیل داشتیم و در بسیاری از سایر حوزه‌های کاربردی فنی نیز چنین بوده است (مثلاً تحلیل مشتری، تحلیل رقابتی، تحلیل دیداری، تحلیل اجتماعی)، رشته‌ی متن نیز راه خود را به حوزه تحلیل‌ها باز کرده است. اگرچه عبارت *تحلیل متن* در حوزه‌ی کاربردهای کسب‌وکار متداول‌تر بوده است ولی از *متن‌کاوی* بیشتر در پژوهش‌های دانشگاهی استفاده می‌شود. اگرچه این دو عبارت در زمان‌هایی تعاریف متفاوتی داشته‌اند، *تحلیل متن* و *متن‌کاوی* معمولاً مترادف یکدیگر بکار می‌روند.

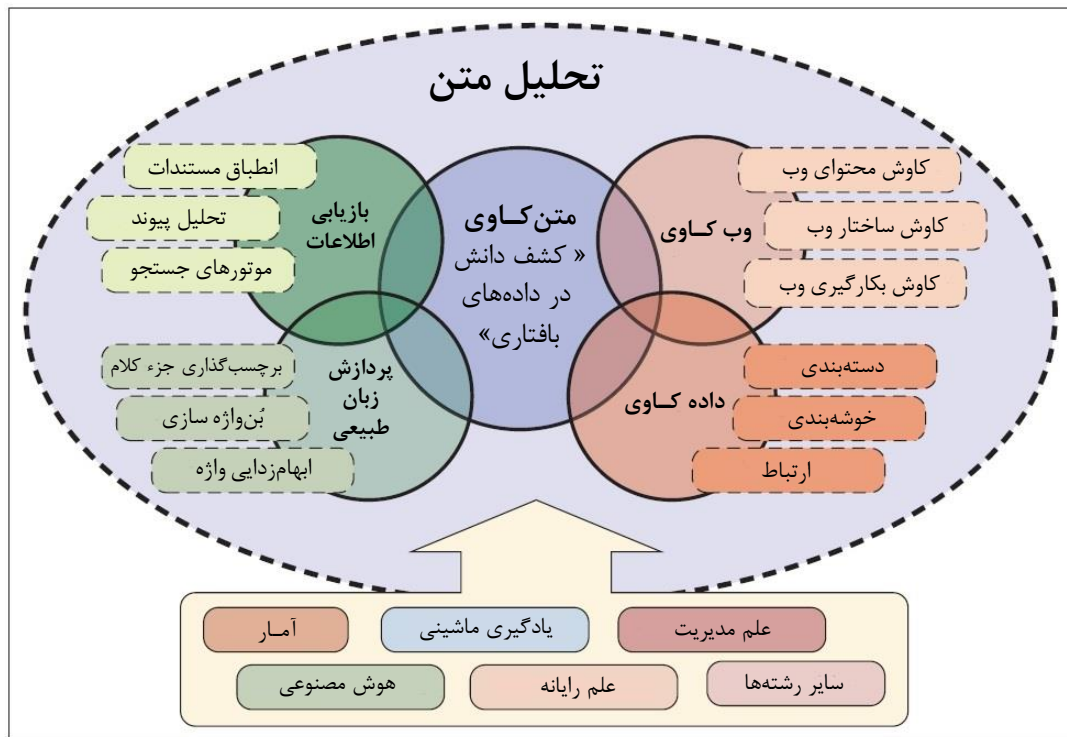
متن‌کاوی (که با عناوین *کاوش داده‌های متنی*^۳ یا *اکتشاف دانش در پایگاه داده‌های بافتاری*^۴ نیز شناخته می‌شود) یک فرایند نیمه‌خودکار استخراج الگوها (اطلاعات و دانش مفید) از حجم انبوه منابع داده‌های ساخت‌نیافته است. به خاطر داشته باشید که داده‌کاوی فرایند شناسایی الگوهای معتبر، جدید، مفید و نهایتاً قابل درک در داده‌های ذخیره شده در پایگاه داده‌های ساخت‌یافته است که داده‌ها در قالب رکوردهای ساخت‌یافته بر اساس متغیرهای رسته‌ای، ترتیبی یا پیوسته سازماندهی می‌شدند. *متن‌کاوی* همانند داده‌کاوی است از این منظر که هدف یکسانی دارند و از فرایندهای یکسانی استفاده می‌کنند ولی در *متن‌کاوی*، ورودی فرایند، مجموعه‌ای از فایل داده‌های ساخت‌نیافته (یا نیمه ساخت‌یافته) از قبیل مستندات ورد، فایل‌های پی‌دی‌اف، گزیده‌های متنی و فایل‌های اکس‌ام‌ال هستند. در واقع، *متن‌کاوی* را می‌توان یک فرایند (با دو گام اصلی) که با تحمیل ساختار بر روی منابع داده متنی شروع شده و پس از آن دانش و اطلاعات مرتبط با استفاده از ابزارها و تکنیک‌های داده‌کاوی از این داده‌های متنی ساخت‌یافته استخراج می‌شوند.

¹ Text Analysis

² Text Mining

³ Text Data Mining

⁴ Knowledge Discovery in textual databases



شکل ۷.۲ تحلیل متن، حوزه‌های کاربردی مربوطه و رشته‌های توانمندساز

مزایای متن‌کاوی در حوزه‌هایی که حجم انبوهی از داده‌های بافتاری تولید می‌شوند مثل قانون (حکم دادگاه‌ها)، پژوهش‌های دانشگاهی (مقالات پژوهشی)، امور مالی (گزارشات فصلی)، پزشکی (خلاصه وضعیت ترخیص)، زیست‌شناسی (تعاملات مولکولی)، فناوری (فایل‌های حقوق ثبت اختراع) و بازاریابی (نظرات مشتریان) واضح و روشن است. به عنوان مثال، تعاملات متنی بدون قالب با مشتریان به شکل شکایات (یا تعریف و تمجیدها) و ادعاهای گارانتی را می‌توان برای شناسایی مشخصه‌های محصولات یا خدماتی که عیوب و نواقصی دارند یا به عنوان ورودی برای بهبود تخصیص خدمات و توسعه‌ی محصول بکار برد. به همین منوال، برنامه‌های توسعه بازار و گروه‌های تمرکز نیز حجم انبوهی از داده‌ها را تولید می‌کنند. در صورتیکه بازخوردهای محصولات و خدمات را به یک فرمت و قالب کدگذاری شده محدود نکنیم، مشتریان می‌توانند به زبان خودشان آنچه را که در مورد محصولات و خدمات شرکت در نظر دارند بیان کنند. یکی دیگر از حوزه‌هایی که پردازش خودکار متون ساخت‌نیافته تأثیر بسزایی بر آن داشته است، ارتباطات الکترونیک و پست الکترونیک است. از داده‌کاوی نه تنها می‌توان برای دسته‌بندی و فیلتر کردن هرنامه‌ها استفاده کرد بلکه می‌توان برای اولویت‌بندی خودکار نامه‌های الکترونیکی بر اساس میزان اهمیت و همچنین پاسخ‌دهی خودکار استفاده کرد (ونگ آند لیو ۲۰۰۴). در ادامه مهمترین و متداول‌ترین حوزه‌های کاربردی متن‌کاوی آورده شده است:

- استخراج اطلاعات. شناسایی جملات و ارتباطات کلیدی در داخل متن از طریق جستجوی اشیاء و توالی‌های از پیش تعریف شده در متن به روش انطباق الگو.
- ردیابی موضوع. پیش‌بینی سایر اسناد مورد نظر کاربر بر اساس شرح حال کاربر و مستندات که مشاهده می‌کند.

- **چکیده‌سازی.** چکیده‌سازی یک سند برای صرفه‌جویی در زمان خواننده.
- **دسته‌بندی.** شناسایی مضامین اصلی یک سند و قراردادن آن سند در مجموعه‌ای از پیش‌تعریف شده‌ای از دسته‌ها و گروه‌ها بر اساس آن مضامین.
- **خوشه‌بندی.** گروه‌بندی اسناد مشابه بدون استفاده از یک مجموعه‌ای از پیش‌تعریف شده از دسته‌ها.
- **ایجاد ارتباط بین مفهوم.** متصل کردن اسناد مرتبط با هم از طریق شناسایی مفاهیم مشترک بین آنها و کمک به کاربران در یافتن آنچه شاید با روش‌های جستجوی قدیمی قادر به یافتن آنها نبوده‌اند.
- **پاسخ به سوالات.** یافتن بهترین پاسخ برای یک سوال از طریق انطباق دانش محور الگوها.

برای توضیح برخی عبارات و مفاهیم بکار رفته در متن کاوی به بینش فناوری ۷.۱ مراجعه کنید. مورد کاربردی ۷.۱ استفاده از متن کاوی را در صنعت بیمه توضیح می‌دهد. مورد کاربردی ۷.۱ نشان می‌دهد که نتفلیکس چگونه از متن کاوی و انواع مختلف منابع داده تولید شده توسط کاربران برای حفظ نوآوری در روش‌های کسب‌وکار خود، تولید بینش عمیق‌تر نسبت به مشتری و تولید محتوای موفق‌تر برای بینندگان خود استفاده کرده است.

بینش فناوری ۷.۱ اصطلاح‌شناسی متن کاوی

در ادامه فهرستی از برخی عبارات متداول و پرکاربرد در متن کاوی توضیح داده شده است:

- **داده‌های ساخت‌نیافته (در برابر داده‌های ساخت‌یافته).** داده‌های ساخت‌یافته دارای قالب از پیش تعیین شده هستند و معمولاً در داخل رکوردهایی با مقادیر داده‌ای ساده سازماندهی شده (متغیرهای رسته‌ای، ترتیبی و پیوسته) و در پایگاه داده ذخیره می‌شوند. برعکس، داده‌های ساخت‌نیافته دارای قالب از پیش تعیین شده نیستند و در به شکل اسناد متنی ذخیره می‌شوند. در واقع، داده‌های ساخت‌یافته برای پردازش رایانه‌ای در نظر گرفته شده‌اند در حالیکه داده‌های ساخت‌نیافته توسط انسان‌ها درک و پردازش می‌شوند.
- **پیکره‌ی زبانی.** در زبان‌شناسی، یک پیکره مجموعه بزرگ و ساخت‌یافته‌ای از متون (که معمولاً به صورت الکترونیکی ذخیره و پردازش می‌شوند) است که برای اکتشاف دانش آماده شده است.
- **عبارات.** یک عبارت یک واژه تکی یا یک جمله‌ی چندواژه‌ای است که مستقیماً از پیکره‌ی یک دامنه‌ی خاص و به روش‌های پردازش زبان طبیعی استخراج شده است.
- **مفاهیم.** مفاهیم ویژگی‌هایی هستند که به وسیله‌ی متدولوژی دست‌ی، آماری، مبتنی بر قواعد یا دسته‌بندی ترکیبی از مجموعه‌ای از اسناد تولید شده‌اند. مفاهیم در مقایسه با عبارات، نتیجه‌ی یک چکیدگی سطح بالاتر هستند.
- **ریشه‌یابی واژگان.** ریشه‌یابی فرایند تقلیل واژگان تا سطح ریشه است. به عنوان مثال، ریشه‌ی تمامی کلمات *stemmer*, *stemming*, *stemmed* عبارت *stem* است.
- **واژه‌ی پالایشی.** واژگان پالایشی^۱ (یا *ایستاده شده* یا *تصفیه‌شونده*) واژگانی هستند که پیش یا پس از پردازش داده‌های زبان طبیعی (یعنی متن) فیلتر و حذف شده باشند. اگرچه در خصوص فهرست این واژگان اتفاق آراء وجود ندارد ولی اکثر ابزارهای پردازش زبان طبیعی از فهرستی که شامل حروف تعریف (*a, an, the*)، حروف اضافه (*of, on, for*)، افعال کمکی (*is, are, was, were*) و واژگان خاص هر بافتار که چندان ارزش متمایزکننده‌ای ندارند تشکیل شده است استفاده می‌کنند.

¹ Stop Words

- **مترادف‌ها و چندمعناها.** مترادف‌ها واژگانی هستند که از لحاظ قواعد نحوی متفاوت (یعنی به شکل متفاوتی هجی یا نوشته می‌شوند) ولی دارای معانی یکسان یا مشابهی هستند (مثلاً *movie, film, motion picture*). برعکس، چندمعناها که متشابه نیز خوانده می‌شوند واژگانی هستند که از لحاظ قواعد نحوی یکسان (یعنی دقیقاً به یک شکل هجی یا نوشته می‌شوند) ولی معانی متفاوتی دارند (مثلاً *bow* به معنی «خم شدن به جلو»، «قسمت جلوی عرشه کشتی»، «سلاحی برای پرتاب تیر» یا «یک نوع روبان گره‌خورده» است).
- **قطعه‌بندی متون.** یک توکن یک قطعه متنی دسته‌بندی شده در یک جمله است. یک قطعه متنی متناظر با یک توکن بر اساس کارکردی که در جمله دارد دسته‌بندی می‌شود. این تخصیص معنا به قطعات متن‌ها را قطعه‌بندی متون^۱ می‌گویند. یک قطعه متنی می‌تواند شبیه هر چیزی باشد فقط باید بخش مفیدی از یک متن ساخت‌یافته باشد.
- **فرهنگ اصطلاحات.** مجموعه‌ای از عبارات و اصطلاحات خاص یک رشته است که برای محدود کردن عبارات استخراج شده از یک پیکره زبانی مورد استفاده قرار می‌گیرد.
- **بسامد واژگانی.** تعداد دفعاتی که یک واژه در یک سند خاص یافت می‌شود.
- **برچسب‌گذاری جزء کلام.** این فرایند برچسب‌گذاری واژگان در یک متن متناظر با جزء خاصی از کلام (اسامی، افعال، صفات، قید و غیره) بر اساس تعریف واژه و بافتاری است که در آن بکار رفته است.
- **ریخت‌شناسی.** شاخه‌ای از رشته‌ی زبانشناسی و بخشی از پردازش زبان طبیعی است که به مطالعه‌ی ساختار درونی واژگان (الگوهای شکل‌گیری واژگان در یک زبان یا بین زبان‌ها) می‌پردازد.
- **ماتریس سند-واژه^۲.** این عبارت به طرح بازنمایی متداول ارتباطات مبتنی بر بسامد بین عبارات و اسناد در یک قالب جدولی اشاره دارد که این عبارات در داخل ستون‌ها و اسناد در داخل ردیف‌ها آورده شده و بسامد بین عبارات و اسناد نیز در داخل سلول‌ها به صورت مقادیر صحیح وارد می‌شوند.
- **تجزیه‌ی مقادیر منفرد^۳ (اندیس‌گذاری پنهان مفهومی)^۴.** این روش تقلیل بُعد برای تبدیل ماتریس سند-واژه به یک حجم و اندازه‌ی قابل مدیریت از طریق یک روش بازنمود واسطه‌ای بسامدها با استفاده از روش دستکاری ماتریس^۵ مشابه تحلیل مؤلفه‌های اصلی^۶ بکار می‌رود.

نتفلیکس: استفاده از کلان‌داده‌ها برای مدیریت درگیرسازی گسترده:
رهاسازی قدرت تحلیل‌ها برای مدیریت محتوا و بینش مشتری

مورد کاربردی ۷.۱

مسئله

در دنیای کاملاً متصل امروزی، کسب‌وکارها برای ایجاد و برقراری روابط با مشتریانی که بیش از پیش مراجعه می‌کنند تحت فشار زیادی هستند.

¹ Tokenizing

² Term-by-document matrix

³ Singular value decomposition

⁴ Latent semantic indexing

⁵ Matrix manipulation

⁶ Principal component analysis

از لحاظ نظری، با ظهور منابع جدید داده‌ها، مدیریت روابط نزدیک‌تر با مشتریان نیز آسان‌تر شده، داده‌ها همچنان به رشد پیش‌بینی نشده‌ی خود ادامه داده و بر پیچیدگی فناوری‌ها افزوده می‌شود. این دستاوردها باید کسب‌وکارها را قادر کند تا شخصی‌سازی کمپین‌های بازاریابی را بهتر از قبل انجام داده و توصیه‌های دقیق‌تری را در خصوص محتوای مورد نیاز برای درگیری‌سازی و مشارکت، بکارگیری و خلق ارزش برای مشتریان ارائه دهند.

ولی دستیابی به درکی عمیق از مخاطبان، یک فرایند مستمر آزمون و یادگیری است که نیازمند توانایی گردآوری سریع و تحلیل موثق هزاران، میلیون‌ها و حتی میلیاردها رویدادها است که هر روزه در منابع داده‌ها، قیمت‌ها و مکان‌های مختلف تحت عنوان کلان‌داده‌ها یافت می‌شوند. پلتفرم‌های فناوری که برای گردآوری این داده‌ها و انجام تجزیه و تحلیل‌ها ساخته شده‌اند باید از قدرت کافی برای ارائه‌ی به موقع بینش‌ها و از انعطاف‌پذیری کافی برای تغییر و رشد چشم‌اندازهای فناوری و کسب‌وکار برخوردار باشند.

نتفلیکس یک پیشگام و نوآور بی‌بدیل در فضای محتوای OTT است که بافتاری که در آن فعالیت می‌کند را بهتر از هر کس دیگری درک کرده است. نتفلیکس، کسب‌وکار و برند خود را مبتنی بر ارائه‌ی تجربیات بسیار هدفمند و شخصی‌سازی شده برای هر مشترک قرار داده و حتی شروع به استفاده از بینش‌های بسیار دقیق و مفصل خود برای تغییر نحوه‌ی خرید، اعطای مجوز و توسعه‌ی محتوای نموده که باعث شده تا بسیاری در سرتاسر صنایع رسانه و سرگرمی متوجه این موضوع شوند.

نتفلیکس برای پشتیبانی از این تلاش‌ها از ترادیتا به عنوان یک مؤلفه‌ی حیاتی برای پلتفرم داده‌ها و تحلیل‌هایش استفاده می‌کند. اخیراً، دو شرکت برای انتقال نتفلیکس به پلتفرم «ابر ترادیتا» به شراکت پرداختند که توان و انعطاف‌پذیری مورد نیاز نتفلیکس را برایش فراهم کرد و توانست تا تمرکز خود را بر راهکارها و ابتکار عمل‌هایی که در بطن کسب‌وکارش قرار دارند معطوف کند.

مدلی برای کسب‌وکارهای متمرکز بر مشتری و داده‌محور

داستان نتفلیکس، مدلی برای شرکت‌های مبتنی بر مشتری و داده‌محور و در واقع هر کسب‌وکاری است که برای تلاش و بقاء در یک دنیای سریع‌اً در حال تغییر نیازمند درگیری‌سازی و مشارکت مخاطباتش است.

نتفلیکس که کار خود را با ارسال دی‌وی‌دی از طریق پست آغاز کرد تبدیل به بزرگ‌ترین ارائه‌دهنده‌ی محتوای OTT شده و دنیای رسانه را دگرگون کرد. به تصمیمات اخیر سایر شرکت‌های بزرگ رسانه‌ای برای شروع ارائه‌ی محتوای OTT توجه کنید.

یکی از مؤلفه‌های اصلی موفقیت نتفلیکس این است که چگونه موتورهای توصیه‌ی خود را بی‌وقفه برای سازگاری با سبک مورد ترجیح هر مشتری تغییر می‌دهد. اکثر فعالیت‌های پخش محتوای شرکت از توصیه‌ها نشأت می‌گیرد که موجب شده تا سطح مشارکت و وفاداری مشتریان به میزان زیادی افزایش یابد. هر تعاملی که یک کاربر یا مشترک نتفلیکس با سرویس برقرار می‌کند مبتنی بر تجزیه و تحلیل‌های بسیار دقیق است و هیچ دو تجربه‌ای مثل هم نیستند.

علاوه بر این، همانطور که گفته شد، نتفلیکس برای انجام خریدهای استراتژیک، اعطای مجوز و تصمیمات توسعه‌ی محتوا به درک مشتریان بالقوه و بالفعل - فردی و گروهی - پرداخته است. این شرکت دو مجموعه‌ی سریالی بسیار موفق را تولید کرده است - خانه پوشالی و نارنجی مد روز است - که تا حدودی به واسطه‌ی درک خارق‌العاده‌ی شرکت از مشتریان خود بوده است.

اگرچه آن تلاش‌ها و مغزهای کسب‌وکار هستند که قلب کسب‌وکار شرکت را شکل می‌دهند ولی فناوری‌هایی که از این راهکارها پشتیبانی می‌کنند نیز باید قدرتمندتر و قابل اعتمادتر از سایر رقبا باشند. پلتفرم تحلیل‌ها و داده‌ها باید قادر باشند تا:

- به سرعت و به شکلی قابل اعتماد، بار کاری سرسام‌آور را مدیریت کنند و از تحلیل دقیق میلیاردها رویداد تراکنشی روزانه در هر قالب و شکلی که داده‌ها ثبت می‌شوند پشتیبانی کند.
- با انواع روش‌های تحلیلی از جمله شبکه‌های عصبی، پایتون، پیگ^۱ و همچنین ابزارهای متنوع هوش کسب‌وکار مثل میکرواستراتژی^۲ کار کند.
- در مورد استثنا از قابلیت ارتجاعی برخوردار بوده و به آسانی خودش را با مقیاس‌ها وفق دهد.
- یک مخزن امن و افزونه برای تمامی داده‌های شرکت فراهم کند.
- متناسب با ساختار هزینه و حاشیه سود مورد نظر شرکت باشد.

انتقال تحلیل ترادیتا به فضای ابر

نتفلیکس و ترادیتا با در نظر گرفتن این ملاحظات یک شراکت موفق را با هدف انتقال انبار داده‌های ترادیتا متعلق به شرکت نتفلیکس به فضای ابر راه‌اندازی کردند.

قدرت و بلوغ: شهرت و اعتبار ترادیتا به خاطر عملکرد استثنایی‌اش به ویژه برای شرکتی مثل نتفلیکس که پلتفرم تحلیلیش با صدها پرس و جوی همزمان مواجه می‌شود اهمیت دارد. نتفلیکس همچنین به ابزارهای تحلیل و انبارش داده‌هایی نیاز داشت که به مدیریت بار کاری پیچیده شرکت کمک کند – ایجاد صف‌های مختلف برای کاربران مختلف و امکان فیلترینگ دائمی و قابل اطمینان نیازهای هر کاربرد.

اکوسیستم‌های تحلیلی هیبریدی و یک معماری داده‌ی یکپارچه

اتکای نتفلیکس به یک اکوسیستم تحلیلی هیبریدی که در شرایط مناسب از هُدوپ بهره می‌برد ولی سرعت و چابکی را به مخاطره نمی‌اندازد بهترین گزینه برای ترادیتا بوده است. محیط ابر نتفلیکس متکی بر یک کانکتور ترادیتا-هَدوپ است که نتفلیکس را قادر می‌سازد تا داده‌های ابری را بدون انقطاع از یک ارائه‌کننده‌ی دیگر به پلتفرم ابر ترادیتا انتقال دهد. نتیجه این‌که نتفلیکس می‌تواند بسیاری از تحلیل‌های خودش را در خارج از انبار داده‌ی کلاس جهانی در پلتفرم ابر ترادیتا انجام دهد که افزونگی، توانایی انقباض و انبساط در واکنش به شرایط محیطی متغیر و کاهش چشمگیر نیاز به جابجایی داده‌ها، آرامش ذهن را برای شرکت به ارمغان آورده است. همچنین رویکرد بدون کنترل و محدودیت نتفلیکس برای اینکه تحلیلهای هر ابزارهای تحلیلی که فکر می‌کنند به بهترین شکل پاسخگوی نیازهای آنها است استفاده کنند نیازمند یک پلتفرم تحلیلی منحصربفرد بود. داشتن یک شریک کسب‌وکار که به خوبی بتواند با مجموعه‌ی کاملی از برنامه‌های کاربردی تحلیلی کار کند از اهمیت حیاتی برخوردار بود.

معماری داده‌ی یکپارچه‌ی ترادیتا با سازماندهی مجدد اکثر شرکت‌هایی که نیازمند مجموعه‌ای امن و اقتصادی از خدمات، پلتفرم‌ها، برنامه‌های کاربردی و ابزارها برای مدیریت، پردازش و تحلیل هوشمندتر داده‌ها بودند توانست به این آرزو جامه‌ی عمل بپوشاند. در عوض، سازمان‌ها می‌توانند بیشترین بهره را از تمامی داده‌هایشان بدست بیاورند.

¹ Pig

² MicroStrategy

معماری داده‌ی یکپارچه‌ی ترادیتا شامل موارد زیر است:

- یک انبار داده‌ی یکپارچه که سازمان‌ها را قادر می‌کند تا به محیطی جامع و مشترک از داده‌ها دسترسی یافته و بینش‌های کسب شده را به سرعت و به شکلی مطمئن در سرتاسر سازمان عملیاتی سازند.
- یک پلتفرم اکتشافی قدرتمند که تحلیل‌های کشف را در اختیار شرکت‌ها قرار می‌دهد تا به سرعت بتوانند بینش‌های موجود در تمامی داده‌ها را از طریق تکنیک‌های مختلف که در دسترس تحلیلگران کسب‌وکارها قرار دارد استخراج کنند.

The Proof Is in the Eyeballs

Netflix scrupulously adheres to a few simple and powerful metrics when evaluating the success of its personalization capabilities: eyeballs. Are subscribers watching? Are they watching more? Are they watching more of what interests them?

With engagement always top of mind, it's no surprise that Netflix is among the world's leaders in personalizing content to successfully attract and retain profitable consumers. It has achieved this standing by drawing on its understanding that in a rapidly changing business and technology landscape, one key to success is constantly testing new ways of gathering and analyzing data to deliver the most effective and targeted recommendations. Working with technology partners that make such testing possible frees Netflix to focus on its core business.

Moving ahead, Netflix believes that making increased use of cloud-based technology will further empower its customer engagement initiatives.

By relying on technology partners that understand how to tailor solutions and provide peace of mind about the redundancy of Netflix's data, the company expects to continue its organic growth and expand its capacity to respond nimbly to technological change

and the inevitable ebbs and flows of business.

سوالاتی برای مورد کاربردی ۷.۱

۲۲. نتفلیکس چه کاری می‌کند؟ آنها چگونه به مدل کسب‌وکار فعلی خود دست پیدا کردند؟

۲۳. در مورد نتفلیکس، منظور از مشتری‌گرا و داده‌محور چیست؟

۲۴. نتفلیکس چگونه از فناوری‌های ترادیتا در تحلیل‌های خود استفاده کرد؟

Source: Teradata Case Study "Netflix: Using Big Data to Drive Big Engagement" <https://www.teradata.com/Resources/Case-Studies/Netflix-Using-Big-Data-to-Drive-Big-Engageme> (accessed July 2018).

سوالاتی برای مرور بخش ۷.۲

۱۴. تحلیل متن چیست؟ چه تفاوتی با متن‌کاوی دارد؟

۱۵. متن‌کاوی چیست؟ چه تفاوتی با داده‌کاوی دارد؟

۱۶. محبوبیت و معروفیت متن کاوی به عنوان یک ابزار تحلیل به چه دلیل در حال افزایش است؟
۱۷. برخی از متداول ترین حوزه‌های کاربردی متن کاوی چه هستند؟

۱۶. محبوبیت و معروفیت متن کاوی به عنوان یک ابزار تحلیل به چه دلیل در حال افزایش است؟

محبوبیت و معروفیت متن کاوی به عنوان یک ابزار تحلیل به چه دلیل در حال افزایش است؟

متن کاوی (Text Mining) یک فرآیند استخراج اطلاعات مفید از متن‌های غیر构造 شده است. این فرآیند شامل شناسایی الگوها، کلمات کلیدی و مفاهیم پنهان در داده‌های متنی است. در سال‌های اخیر، متن کاوی به دلیل افزایش حجم داده‌های متنی و نیاز به تحلیل این داده‌ها به سرعت محبوب شده است. یکی از دلایل اصلی این محبوبیت، پیشرفت‌های فزاینده در الگوریتم‌های یادگیری ماشین و پردازش زبان طبیعی است. این ابزارها قادرند حجم عظیمی از داده‌ها را به سرعت و دقت بالا تحلیل کنند و نتایج قابل فهمی ارائه دهند. همچنین، متن کاوی در صنایع مختلفی مانند بازاریابی، پزشکی، حقوق و مدیریت منابع انسانی کاربرد گسترده‌ای دارد. این ابزارها به متخصصان کمک می‌کند تا روندهای جدید را شناسایی کنند، نیازهای مشتریان را درک کنند و تصمیمات بهتری بگیرند. علاوه بر این، متن کاوی به دلیل قابلیت خودکارسازی و یکپارچه شدن با سایر ابزارهای تحلیلی، به یک ابزار ضروری در دنیای دیجیتال تبدیل شده است.

۱۷. برخی از متداول ترین حوزه‌های کاربردی متن کاوی چه هستند؟

برخی از متداول ترین حوزه‌های کاربردی متن کاوی چه هستند؟

متن کاوی در حوزه‌های مختلفی کاربرد دارد. یکی از رایج‌ترین کاربردها، بازاریابی و فروش است. در این حوزه، متن کاوی برای تحلیل نظرات مشتریان، شناسایی نیازها و سبک‌های زندگی، و پیش‌بینی روندهای بازار استفاده می‌شود. در صنعت پزشکی، متن کاوی برای استخراج اطلاعات از پرونده‌های پزشکی، مقالات علمی و گزارش‌های آزمایشگاهی به کار می‌رود. این ابزارها به پزشکان کمک می‌کند تا تشخیص‌های دقیق‌تری بزنند و درمان‌های شخصی‌تری ارائه دهند. در حوزه حقوق، متن کاوی برای تحلیل اسناد حقوقی، قراردادها و پرونده‌های قضایی استفاده می‌شود. این ابزارها به وکلای کمک می‌کند تا موارد مشابه را شناسایی کنند و استراتژی‌های دفاعی و prosecutorial را بهبود بخشند. در مدیریت منابع انسانی، متن کاوی برای تحلیل نظرات کارکنان، شناسایی علل ترک خدمت و بهبود فرآیندهای استخدام استفاده می‌شود. همچنین، در صنایع مختلف، متن کاوی برای نظارت بر شبکه‌های اجتماعی، شناسایی تهدیدات امنیتی و تحلیل روندهای بازار استفاده می‌شود. این ابزارها به مدیران کمک می‌کند تا تصمیمات بهتری بگیرند و سازمان‌های خود را در رقابت‌های جهانی تقویت کنند.

¹ Bag-of-words

² Medical Literature Analysis and Retrieval System Online (MEDLINE)

پردازش زبان طبیعی (ان‌ال‌پی) یک مؤلفه‌ی مهم در متن‌کاوی و زیر زیررشته از هوش مصنوعی و زبانشناسی رایانشی است که به مطالعه‌ی مسئله‌ی «درک» زبان طبیعی انسان و تبدیل نمایش‌های زبان انسانی (مثلاً اسناد متنی) به بازنمودهای رسمی‌تر (به فرم داده‌های عددی و سمبولیک) که کار با آنها برای برنامه‌های رایانه‌ای آسان‌تر است می‌پردازد. هدف ان‌ال‌پی این است که فراتر از دستکاری مبتنی بر علم نحو متون رفته و به درک حقیقی و پردازش زبان طبیعی بپردازد که محدودیت‌های دستور زبانی و معنایی و در عین حال بافتار را مد نظر قرار می‌دهد.

تعریف و دامنه‌ی درک کلمات یکی از بحث‌های اساسی در ان‌ال‌پی است. با توجه به مبهم بودن زبان طبیعی انسان و اینکه درک حقیقی معنا نیازمند دانش گسترده در خصوص آن موضوع است (فراتر از دانش واژگان؛ جملات و پاراگراف‌ها)، آیا رایانه‌ها هرگز قادر به درک زبان طبیعی به همان روش و دقت انسان‌ها خواهند بود؟ شاید خیر! ان‌ال‌پی راه درازی را از زمان شمارش ساده‌ی کلمات طی کرده است ولی راه طولانی‌تری را برای درک واقعی زبان طبیعی انسان پیش رو دارد. در ادامه به برخی از چالش‌های متداول که در پیاده‌سازی ان‌ال‌پی با آنها مواجه خواهیم شد می‌پردازیم:

- **برچسب‌گذاری جزء کلمه.** به سختی می‌توان عباراتی در یک متن را که متناظر با یک بخش یا جزء خاص از گفتار هستند (مثلاً اسامی، افعال، صفات یا قیود) برچسب‌گذاری کرد زیرا جزء کلمه یا گفتار نه تنها به تعریف آن عبارت بلکه به بافتاری که در آن بکار رفته است نیز بستگی دارد.
- **بخش‌بندی متن.** برخی زبان‌های نوشتاری مثل چینی، ژاپنی و تایلندی دارای مرزهای تک واژه‌ای نیستند. در این مثال‌ها، تجزیه‌ی متن نیازمند شناسایی مرزهای کلمات است که اغلب کار دشواری است. زمانیکه به تجزیه و تحلیل زبان‌های گفتاری می‌پردازیم نیز با چالش‌های مشابهی در بخش‌بندی گفتار مواجه می‌شویم چونکه اصواتی که حروف و واژگان متوالی را بیان می‌کنند در هم می‌آمیزند.
- **ابهام‌زدایی معنایی واژه.** بسیاری از واژگان بیش از یک معنا دارند. انتخاب بهترین معنا برای یک کلمه ممکن است فقط با در نظر گرفتن بافتاری که این واژه در آن بکار رفته است میسر شود.
- **ابهام نحوی.** دستور زبان‌های طبیعی دارای ابهام است یعنی اینکه اغلب چندین ساختار برای ساخت یک جمله امکان‌پذیر است. انتخاب مناسب‌ترین ساختار معمولاً نیازمند تلفیق اطلاعات معنایی و بافتاری است.
- **ورودی ناقص یا نامنظم.** لهجه و گویش‌های خارجی یا منطقه‌ای و مشکلات صوتی در گفتار و خطاهای دستوری یا نوشتاری در متون می‌توانند پردازش زبان را دشوارتر سازند.
- **کنش گفتاری¹.** یک جمله را اغلب می‌توان یک کنش از جانب گوینده در نظر گرفت. ساختار جمله به تنهایی ممکن است حاوی اطلاعات کافی برای تعریف این کنش نباشد. به عنوان مثال، «Can you pass the class?» به یک پاسخ ساده‌ی بله/خیر نیاز دارد در حالیکه «Can you pass the salt?» نیازمند یک کنش یا اقدام فیزیکی است که باید انجام شود.

یک رویای قدیمی جامعه‌ی هوش مصنوعی این است که الگوریتم‌هایی را داشته باشند که قادر به خواندن و کسب خودکار دانش از متن باشد. پژوهشگران دانشگاه استنفورد با بکارگیری یک الگوریتم یادگیری بر روی متن تجزیه شده، روش‌هایی را توسعه داده‌اند که قادر به شناسایی خودکار مفاهیم و ارتباطات بین آن مفاهیم در متن هستند. با بکارگیری یک رویه‌ی منحصربفرد برای

¹ Speech act

حجم انبوهی از متون، الگوریتم‌های آزمایشگاه قادر به اکتساب خودکار صدها هزار قلم دانشی و بکارگیری آنها برای تولید مخازن بسیار پیشرفته برای وُردنت هستند. وُردنت یک پایگاه داده از واژگان زبان انگلیسی، تعاریف آنها، مجموعه مترادف‌ها و روابط معنایی مختلف بین مجموعه مترادف‌ها است که به صورت دستی در آزمایش کدگذاری شده‌اند. اگرچه وُردنت یک منبع اصلی برای کاربردهای پردازش زبان طبیعی محسوب می‌شود ولی ساخت و حفظ آن به صورت دستی بسیار پرهزینه است. با القای خودکار دانش به وُردنت این پتانسیل وجود دارد که به منبعی بزرگتر و جامع‌تر برای پردازش زبان طبیعی با هزینه‌ای بسیار کمتر تبدیل شود. یکی از حوزه‌های اصلی که این‌ال‌پی و وُردنت در آن سرمایه‌گذاری کرده‌اند، مدیریت ارتباط با مشتریان است. به بیان کلی، هدف مدیریت ارتباط با مشتریان این است که با درک بهتر و پاسخگویی مؤثرتر به نیازهای واقعی و ادراک شده‌ی مشتریان، ارزش مشتری را به حداکثر برسانیم. یک حوزه‌ی مهم مدیریت ارتباط با مشتریان که این‌ال‌پی در آن سهم بسزایی دارد، تحلیل احساسات است. تحلیل احساسات تکنیکی است که برای تشخیص عقاید و نظرات مثبت و منفی نسبت به محصولات یا خدمات خاص با استفاده از تعداد زیادی منبع داده‌ی متنی بکار می‌رود (بازخورد مشتری در قالب پست‌های وبی). برای پوشش کامل تحلیل احساسات و وُردنت به بخش ۷.۶ مراجعه کنید.

تحلیل‌ها به طور کلی و تحلیل متن و متن کاوی به طور خاص را می‌توان در صنعت پخش برنامه‌های رادیویی و تلویزیونی بکار برد. مورد کاربردی ۷.۲ یک مثال از کاربرد طیف گسترده‌ای از توانایی‌های تحلیلی را برای اخذ نظرات بینندگان، پیش‌بینی رتبه‌بندی‌ها و خلق ارزش برای شرکت‌های پخش برنامه‌های رادیویی و تلویزیونی ارائه می‌کند.

شبکه‌های ای‌ام‌سی از تحلیل‌ها برای جذب بینندگان جدید، پیش‌بینی رتبه‌بندی‌ها و خلق ارزش افزوده برای تبلیغ‌کنندگان در یک دنیای چندکاناله استفاده می‌کند.

مورد کاربردی ۷.۲

طی ۱۰ سال گذشته، بخش تلویزیون کابلی در ایالات متحده با رشدی همراه بوده است که خلاقیت‌های پیش‌بینی نشده موجب شده تا محتوای با کیفیت بسیار بالا خلق شود. شبکه‌های ای‌ام‌سی در خط مقدم این عصر طلایی تلویزیون بوده و مجموعه‌ای از نمایش‌های بسیار موفق و معروف از قبیل *Breaking Bad, Mad Men* و *Walking Dead* را تولید کرده‌اند. شبکه‌های ای‌ام‌سی که برای بیش از ۳۰ سال خود را وقف تولید فیلم‌ها و برنامه‌های با کیفیت کرده‌اند مالک چندین برند معروف و برنده‌ی جایزه در تلویزیون کابلی، تولید و ارائه‌ی محتوای متمایز و متناسب با فرهنگ جامعه بوده است که مخاطبان بسیاری را از سرتاسر پلتفرم‌های متعدد درگیر کرده است.

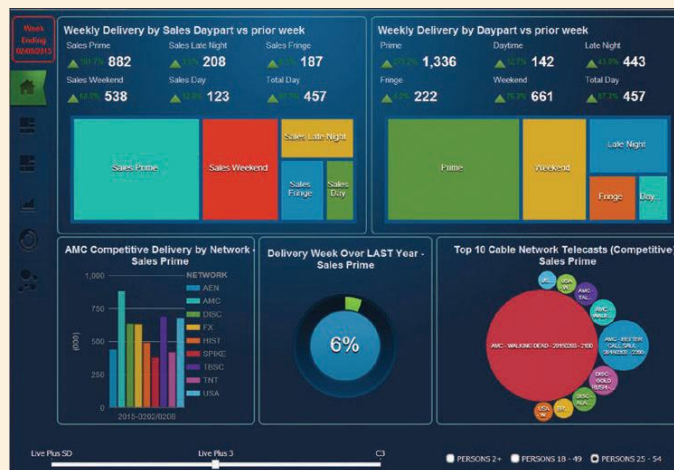
جلو افتادن در بازی

شبکه‌های ای‌ام‌سی علیرغم موفقیتش هیچ برنامه‌ای ندارد که در این بازی جلوتر از بقیه باقی بماند. طبق گفته‌های ویتالی تسیوین: ما هیچ علاقه‌ای به پیشرفت نداریم. اگرچه درصد قابل توجهی از کسب‌وکار ما را همچنان تلویزیون کابلی خطی تشکیل می‌دهد ولی باید برای نسل جدید که محتوا را به شکل‌ها و روش‌های بسیار متفاوتی استفاده می‌کنند جذابیت داشته باشیم. تلویزیون تبدیل به یک کسب‌وکار چندکاناله و مالتی‌استریم شده است و شبکه‌های کابلی برای بازاریابی و اتصال با مخاطبان خود در تمامی این جریان‌ها باید هوشمندتر عمل کنند. اتکا به به داده‌های رتبه‌بندی سنتی و ارائه‌دهندگان تحلیل‌های شخص ثالث یک استراتژی شکست خواهد بود: شما باید مالکیت داده‌هایتان را در اختیار بگیرید، آن را برای دستیابی به تصویری غنی‌تر از بینندگان‌تان، آنچه که می‌خواهند و چگونگی حفظ توجه آنها در یک بازار سرگرمی پرازدحام بکار بگیرید.

منطقه‌بندی بینندگان

چالشی که در اینجا وجود دارد حجم بسیار بالای اطلاعات موجود است - صدها میلیارد ردیف داده از ارائه‌دهندگان داده‌ها در صنایع از قبیل نیلسن و کام‌اسکور، از کانال‌هایی از قبیل سرویس پخش وب زنده و ویدئوی هنگام درخواست تلویزیون‌ای‌ام‌سی، از شرکای خرده‌فروش مثل آی‌تیونز و آمازون و از خدمات ویدئویی آنلاین شخص ثالث مثل نتفلیکس و هولو. به گفته‌ی تسیوین، «ما نمی‌توانیم بر چکیده‌های سطح بالا اتکا کنیم؛ ما باید بتوانیم داده‌های ساخت‌یافته و ساخت‌نیافته را دقیقاً به دقیقه و برای هر بیننده تحلیل کنیم. ما باید بدانیم چه کسی و چرا برنامه‌های ما را تماشا می‌کند و باید سریعاً متوجه شویم تا بتوانیم تصمیم بگیریم که آیا باید یک آگهی تبلیغاتی خاص را فردا شب در زمان پخش *Mad Men* پخش کنیم یا خیر.»

ای‌ام‌سی تصمیم گرفت که یک قابلیت تحلیل پیشرو در صنعت را در داخل سازمان توسعه دهد و این قابلیت را در کوتاه‌ترین زمان ممکن عرضه کند. به جای اجرای یک فرایند طولانی و پرهزینه ارزیابی و گزینش محصول و فروشنده، ای‌ام‌سی تصمیم گرفت تا از روابط موجود خود با آی‌بی‌ام به عنوان شریک استراتژیک مورد اعتمادش در حوزه فناوری بهره بگیرد. در عوض، زمان و پولی را صرف تدارکات می‌شد برای محقق‌سازی راه‌حل مسئله می‌شود و پیشرفت ای‌ام‌سی در توسعه‌ی ابزارهای تحلیلی را به میزان حداقل شش ماه جلو می‌اندازد.



یک داشبورد مبتنی بر وب که شبکه‌های ای‌ام‌سی از آن استفاده می‌کند.

تفویض اختیار به دپارتمان پژوهش

در گذشته، تیم پژوهشی ای‌ام‌سی بخش زیادی از زمان خود را صرف پردازش داده‌ها می‌کرد. امروزه به لطف ابزارهای تحلیلی جدیدشان می‌توانند اکثر انرژی خود را بر دستیابی به بینش‌های اقدام‌پذیر متمرکز کنند. تسیوین می‌گوید: «تحلیل‌هایی که ممکن است روزها و هفته‌ها به طول بینجامند اکنون در چند دقیقه یا ثانیه امکان‌پذیر شده‌اند.» وی می‌افزاید: «توسعه‌ی ابزارهای تحلیلی در داخل سازمان، صرفه‌جویی مستمر در هزینه‌ها را به همراه دارند. به جای پرداخت صدها هزار دلار به تأمین‌کنندگان و فروشندگان برون سازمانی ما می‌توانیم خودمان اینکار را سریع‌تر و دقیق‌تر و با هزینه بسیار کمتری انجام بدهیم. ما انتظار داریم تا شاهد بازگشت سریع سرمایه باشیم. هرچه دسترسی به منابع بینش‌های بالقوه بیشتر می‌شود و تحلیل‌ها اهمیت استراتژیک‌تری برای کسب‌وکارها پیدا می‌کنند، استفاده از یک روش درون سازمانی تنها راه ممکن برای هر شبکه‌ای است که حقیقتاً به دنبال کسب مزیت رقابتی از داده‌هایش می‌باشد.»

مدیریت تصمیم به کمک داده‌ها

بسیاری از نتایجی که این قابلیت تحلیلی جدید ارائه کرده است نشان می‌دهند که تغییر و دگرگونی حقیقی در نحوه‌ی عملیات ای‌ام‌سی بوجود آمده است. به عنوان مثال، دپارتمان هوش کسب‌وکار شرکت توانسته است تا مدل‌های آماری پیچیده‌ای را ایجاد کند که به شرکت در اطلاع‌رسانی استراتژی‌های بازاریابی و اتخاذ تصمیمات هوشمندتر در خصوص تبلیغات برای هر نمایش کمک می‌کنند.

با دستیابی به بینش‌های عمیق‌تر، کمپین‌های بازاریابی مستقیم ای‌ام‌سی موفق‌تر از قبل عمل می‌کنند. در یک مثال جدید، بخش‌بندی اطلاعات و مدل‌سازی موارد مشابه به شرکت کمک کرده است تا بینندگان فعلی و جدیدی را هدف قرار دهد که بدین ترتیب تعداد تراکنش‌های ویدئو بر حسب درخواست ای‌ام‌سی بسیار فراتر از حد انتظار بوده است.




این توانایی جدید برای دستیابی به بینندگان جدید بر اساس اولویت‌ها و نیازهای فردی آنها فقط برای شرکت ای‌ام‌سی ارزشمند نیست بلکه ارزش بالقوه‌ی عظیمی برای شرکای تبلیغاتی شرکت نیز به همراه دارد. ای‌ام‌سی در حال حاضر به دنبال ایجاد دسترسی به مجموعه داده‌های غنی و ابزارهای تحلیلی خود به عنوان یک سرویس برای تبلیغ‌کنندگان خود می‌باشد تا به آنها در تعریف دقیق کمپین‌هایی که مخاطبان بیشتری را در سرتاسر کانال‌های خطی و دیجیتالی جذب خود کند کمک کند.

تسیوین اینگونه نتیجه‌گیری می‌کند که «اکنون که ما واقعاً می‌توانیم ارزش کلان‌داده‌ها را مهار کرده و در اختیار بگیریم می‌توانیم پیشنهادات جذاب‌تری را برای مشتریان و تبلیغ‌کنندگان خود ارائه کنیم و حتی محتوای بهتری را ایجاد کرده، بازاریابی مؤثرتری برای آن انجام داده و با بهره‌گیری کامل از مزایای قابلیت‌های چندکاناله به طیف گسترده‌تری از مخاطبان دسترسی پیدا کنیم.»

سوالاتی برای مورد کاربردی ۷.۲

۱. شرکت‌های پخش برنامه‌های رادیویی و تلویزیونی امروزه با چه چالش‌های متداولی مواجه هستند؟ تحلیل‌ها چگونه می‌توانند به کاهش این چالش‌ها کمک کنند؟
۲. شرکت ای‌ام‌سی چگونه از تحلیل‌ها برای ارتقاء عملکرد کسب‌وکار خود بهره گرفت؟
۳. انواع تحلیل متن‌ها و راه‌حل‌های متنی که شبکه‌های ای‌ام‌سی توسعه داد چه بودند؟ آیا می‌توانید کاربردهای بالقوه‌ی دیگری از متن‌کاوی را در صنعت پخش برنامه‌های رادیویی و تلویزیونی نام ببرید؟

Sources: IBM Customer Case Study. "Using Analytics to Capture New Viewers, Predict Ratings and Add Value for Advertisers in a Multichannel World." <http://www-03.ibm.com/software/businesscasestudies/us/en/corp?synkey=A023603A76220M60> (accessed July 2016); www.ibm.com; www.amcnetworks.com

ای‌ام‌سی داده‌های رتبه‌بندی را با اطلاعات بینندگان که از طیف گسترده‌ای از کانال‌های دیجیتال بدست آمده است ترکیب می‌کند: سرویس‌های پخش زنده و ویدئو بر حسب درخواست، خرده فروش‌ها و خدمات تلویزیونی آنلاین	ابزارمند	
یک موتور تحلیل و کلان‌داده‌ی قدرتمند و جامع که داده‌ها را متمرکز نموده و آنها را در دسترس طیف گسترده‌ای از ابزارهای تحلیلی توصیفی و پیشگویانه برای تسریع مدل‌سازی، گزارش‌گیری و تجزیه و تحلیل قرار می‌دهد.	متصل	
ای‌ام‌سی می‌تواند پیش‌بینی کند که کدام نمایش‌ها موفق خواهند بود، چگونه باید جدول زمانبندی پخش آنها را برنامه‌ریزی کرد، چه آگهی و تبلیغاتی باید برای آن	هوشمند	

پردازش زبان طبیعی در حوزه‌های مختلف و برای طیف گسترده‌ای از وظایف از طریق برنامه‌های رایانه‌ای برای خودکارسازی پردازش زبان طبیعی انسان که قبلاً فقط توسط انسان انجام می‌شدند با موفقیت بکار گرفته شده‌اند. در ادامه معروف‌ترین این وظایف توضیح داده شده است:

- **پاسخ به سوالات.** وظیفه‌ی پاسخگویی خودکار به سوالی که به زبان طبیعی پرسیده شده یعنی بتوان زمانیکه سوالی به زبان انسانی پرسیده می‌شود بتوان به زبان انسان پاسخ داد. برای یافتن پاسخ برای یک سوال، برنامه رایانه‌ای می‌تواند از یک پایگاه داده ساخت‌یافته یا مجموعه‌ای از سندهای موجود به زبان طبیعی استفاده کند (یک پیکره‌ی متنی مثل شبکه گسترده‌ی جهانی).
 - **خلاصه‌سازی خودکار.** ایجاد نسخه‌ی کوتاهی از یک سند متنی توسط یک برنامه‌ی رایانه‌ای که حاوی مهم‌ترین نکات موجود در سند اصلی است.
 - **تولید زبان طبیعی.** تبدیل اطلاعات از پایگاه داده‌های رایانه‌ای به زبان انسانی قابل خواندن.
 - **درک زبان طبیعی.** تبدیل نمونه‌هایی از زبان انسانی به بازنمودهای رسمی‌تر که کار با آنها برای برنامه‌های رایانه‌ای آسان‌تر باشد.
 - **ترجمه‌ی ماشینی.** ترجمه‌ی خودکار از یک زبان انسانی به زبان دیگر.
 - **خواندن زبان خارجی.** یک برنامه‌ی رایانه‌ای که به سخنوران زبان‌های غیربومی کمک می‌کند تا یک زبان خارجی را با تلفظ صحیح و تکیه کلام بر روی بخش‌های مختلف کلمات بخوانند.
 - **نوشتن به یک زبان خارجی.** یک برنامه‌ی رایانه‌ای که به سخنوران زبان‌های غیربومی کمک می‌کند تا به یک زبان خارجی بنویسند.
 - **تشخیص گفتار.** تبدیل کلمات گفتاری به یک ورودی قابل خواندن برای ماشین. سیستم می‌تواند یک کلیپ صوتی از صحبت‌های یک شخص را به یک متن نوشتاری تبدیل کند.
 - **متن به گفتار.** یک برنامه رایانه‌ای که متن نوشته شده استاندارد را به صورت خودکار به گفتار انسانی تبدیل می‌کند.
 - **تصحیح متن.** یک برنامه‌ی رایانه‌ای که نسخه‌ای از یک متن را برای شناسایی و تصحیح خطاها می‌خواند.
 - **تشخیص نوری کاراکترها.** ترجمه‌ی خودکار تصاویر دست‌نویست، تایپ شده یا متن چاپ شده (که معمولاً اسکن شده‌اند) به اسناد متنی قابل ویرایش توسط ماشین.
- موفقیت و محبوبیت متن کاوی به میزان زیادی به دستاوردهای زبان طبیعی هم در تولید و هم درک زبان‌های انسانی بستگی دارد و پردازش زبان طبیعی امکان استخراج ویژگی‌ها از یک متن ساخت‌نیافته را فراهم می‌کند که بدین ترتیب از طیف گسترده‌ای از تکنیک‌های داده‌کاوی برای استخراج دانش (الگوها و ارتباطات نوین و مفید) از آن استفاده کرد. بدین ترتیب، به بیان ساده، متن کاوی ترکیبی از پردازش زبان طبیعی و داده کاوی است.

سوالاتی برای مرور بخش ۷.۳

۱۰. پردازش زبان طبیعی چیست؟

۱۱. پردازش زبان طبیعی چه ارتباطی با متن کاوی دارد؟

۱۲. برخی از مزایا و چالش‌های پردازش زبان طبیعی چه هستند؟

۱۳. متداول‌ترین وظایفی که توسط پردازش زبان طبیعی انجام می‌شوند چه هستند؟

با افزایش حجم داده‌های ساخت‌نیافته‌ی گردآوری شده توسط سازمان‌ها، گزاره‌ی ارزشی و محبوبیت ابزارهای متن کاوی نیز افزایش می‌یابد. بسیاری از سازمان‌ها اکنون به اهمیت استخراج دانش از مخازن داده‌های متنی و سندی خود از طریق بکارگیری ابزارهای متن کاوی پی برده‌اند. در ادامه به معرفی فقط بخش کوچکی از دسته بندی کاربردهای متن کاوی پرداخته‌ایم:

از متن کاوی می‌توان برای افزایش بیش‌فروشی^۱ و فروش مکمل^۲ از طریق تجزیه و تحلیل داده‌های ساخت‌نیافته‌ی تولید شده توسط مراکز تماس استفاده کرد. یادداشت‌های ایجاد شده از متن‌ها از مراکز تماس و همچنین رونوشت گفتگوها و مکالمات صوتی با مشتریان را می‌توان با استفاده از الگوریتم‌های متن کاوی برای استخراج اطلاعات بدیع و اقدام‌پذیر در مورد ادراک مشتری از محصولات و خدمات یک شرکت مورد تجزیه و تحلیل قرار داد. علاوه بر این، وبلاگ‌ها، نظرات کاربران در خصوص محصولات در یک وبسایت مستقل و مطالب ارسال شده در بخش‌های بحث و گفتگو، معادن غنی از احساسات مشتریان هستند. این مجموعه غنی از اطلاعات را زمانیکه به درستی تحلیل شوند می‌توان برای افزایش رضایت مشتریان و افزایش ارزش طول عمر مشتری بکار برد (کوسمان آند وان‌دن پُل ۲۰۰۸).

متن کاوی به ابزاری فوق‌العاده ارزشمند برای مدیریت ارتباط با مشتریان تبدیل شده است. شرکت‌ها می‌توانند از متن کاوی برای تحلیل مجموعه غنی از داده‌های متنی ساخت‌نیافته به همراه داده‌های ساخت‌یافته‌ی مرتبط که از پایگاه‌های داده سازمانی استخراج شده‌اند برای پیش‌بینی ادراک مشتریان و رفتارهای خرید بعدی آنها استفاده کنند. کوسمان آند وان‌دن پُل (۲۰۰۹) از متن کاوی برای بهبود چشمگیر توانایی یک مدل در پیش‌بینی نگهداشت مشتری استفاده کرده‌اند به نحویکه مشتریانی که به احتمال بسیار زیاد مایل به ترک شرکت بوده‌اند دقیقاً برای اتخاذ تاکتیک‌های حفظ و نگهداشت شناسایی شده‌اند.

قانی و همکاران (۲۰۰۶) از متن کاوی برای توسعه‌ی سیستمی که قادر به استنتاج صفات و ویژگی‌های ضمنی و آشکار محصولات است برای ارتقاء توانایی خرده‌فروش‌ها در تحلیل پایگاه داده‌های محصول استفاده کرده‌اند. اینکده محصولات را به جای موجودیت‌ها و نهادهای اتمی و تجزیه‌ناپذیر به صورت زوج‌های صفت-مقدار ببینیم می‌تواند به طور بالقوه موجب افزایش اثربخشی بسیاری از کاربردها در کسب‌وکار از جمله پیش‌بینی تقاضا، بهینه‌سازی سبد دسته‌بندی کالا، توصیه محصول، مقایسه‌ی سبد کالا بین خرده‌فروش‌ها و تولیدکنندگان و گزینش تأمین‌کننده‌ی محصول شود. سیستم پیشنهادی به یک شرکت یا کسب‌وکار اجازه می‌دهد تا محصولاتش را با کمترین تلاش دستی بر حسب صفات و مقادیر صفت بیان کند. این سیستم با بکارگیری تکنیک‌های یادگیری نظارت شده و نیمه نظارت شده بر روی شرح محصولات که در وبسایت‌های خرده‌فروش‌ها یافت می‌شود به یادگیری می‌پردازد.

¹ Up-selling

² Cross-selling

یکی از بزرگ‌ترین و مهم‌ترین کاربردهای متن‌کاوی در حوزه امنیت را شاید بتوان سیستم نظارتی کاملاً محرمانه‌ی ECHELON دانست. طبق شایعات، ECHELON قادر به شناسایی محتوای تماس‌های تلفنی، فکس‌ها، نامه‌های الکترونیکی و سایر انواع داده‌ها، ردگیری اطلاعات ارسالی از طریق ماهواره‌ها، شبکه‌های تلفن سوئیچ عمومی و لینک‌های میکروویو می‌باشد.

در سال ۲۰۰۷، آژانس اتحادیه اروپا برای همکاری در اجرای قانون (یورپول)^۱ یک سیستم یکپارچه را توسعه داد که قادر بود تا برای ردیابی جرائم مالی سازماندهی شده به حجم انبوهی از منابع داده‌ی ساخت‌یافته و ساخت‌نیافته دسترسی داشته و آنها را ذخیره و تحلیل کند. این سیستم که سیستم تحلیل کلی برای پشتیبانی اطلاعات^۲ (OASIS) نامیده می‌شود با دنبال یکپارچه‌سازی پیشرفته‌ترین فناوری‌های داده‌کاوی و متن‌کاوی موجود در بازار امروز است. این سیستم، یورپول را قادر ساخته تا در پشتیبانی و حمایت از اهداف اجرای قانون در سطح بین‌المللی به پیشرفت قابل توجهی برسد.

اداره تحقیقات فدرال ایالات متحده (افبی‌آی) و آژانس اطلاعات مرکزی (سی‌آی‌آی) تحت مدیریت وزارت امنیت ملی به طور مشترک به توسعه‌ی یک سیستم ابررایانه‌ای متن‌کاوی و داده‌کاوی می‌پردازد. این سیستم قرار است تا یک انبار داده‌ی غول‌آسا را به همراه انواع گوناگون ماژول‌های متن‌کاوی و داده‌کاوی برای برآوردن کردن نیازهای کشف دانش نهادهای اعمال قانون محلی، ایالتی و فدرال ایجاد کند. تا پیش از این پروژه، افبی‌آی و سی‌آی‌آی هر کدام از پایگاه داده‌های جداگانه‌ی خودشان استفاده می‌کردند که تقریباً هیچگونه ارتباطی بین آنها برقرار نبود.

یک کاربرد دیگر مرتبط با امنیت در متن‌کاوی، حوزه‌ی تشخیص فریبکاری است. فولر، بیروس و دیلن (۲۰۰۸) با بکارگیری متن‌کاوی بر روی مجموعه بزرگی از اظهارات و گزارشات جنایی جهان، مدل‌های پیش‌بینی را برای متمایز کردن اظهارات فریبکارانه از اظهارات حقیقی توسعه دادند. این مدل با استفاده از یک مجموعه غنی از سرنخ‌هایی که از اظهارات متنی استخراج شده‌اند به پیش‌بینی نمونه‌ها با ۷۰ درصد دقت پرداخته است که با توجه به اینکه سرنخ‌ها فقط از اظهارات متنی (و نه شفاهی یا دیداری) استخراج شده‌اند موفقیت بزرگی محسوب می‌شود. علاوه بر این، این روش در مقایسه با سایر تکنیک‌های دروغ‌سنجی، بدون نفوذ امواج بوده و به طور گسترده‌ای می‌توان آن را نه فقط بر روی داده‌های متنی بلکه بر روی رونوشت‌های صوت‌های ضبط شده بکار گرفت. یک شرح مفصل‌تر از تشخیص فریبکاری بر اساس متن در مورد کاربردی ۷.۳ آورده شده است.

متن‌کاوی به چند دلیل از پتانسیل بالایی در رشته‌ی پزشکی به طور کلی و در زیست‌پزشکی به طور خاص برخوردار است. ابتدا اینکه ادبیات و کتابچه‌های منتشر شده (به ویژه با پیدایش مجلات منبع باز) در این رشته با نرخ نمایی در حال گسترش هستند. دوم اینکه، در مقایسه با اکثر رشته‌های دیگر، ادبیات موضوع پزشکی از نظم و استاندارد بیشتری برخوردار بوده که آن را به یک منبع اطلاعاتی «استخراج‌شدنی» تبدیل کرده است. نهایتاً، مجموعه واژگان و اصطلاحات بکار رفته در این ادبیات به دلیل ماهیت استاندارد آن نسبتاً بدون تغییر و ثابت هستند. در ادامه به معرفی چند مطالعه‌ی انجام شده در خصوص بکارگیری موفقیت‌آمیز تکنیک‌های متن‌کاوی در استخراج الگوهای نوین از ادبیات زیست‌پزشکی می‌پردازیم.

¹ European Union Agency for Law Enforcement

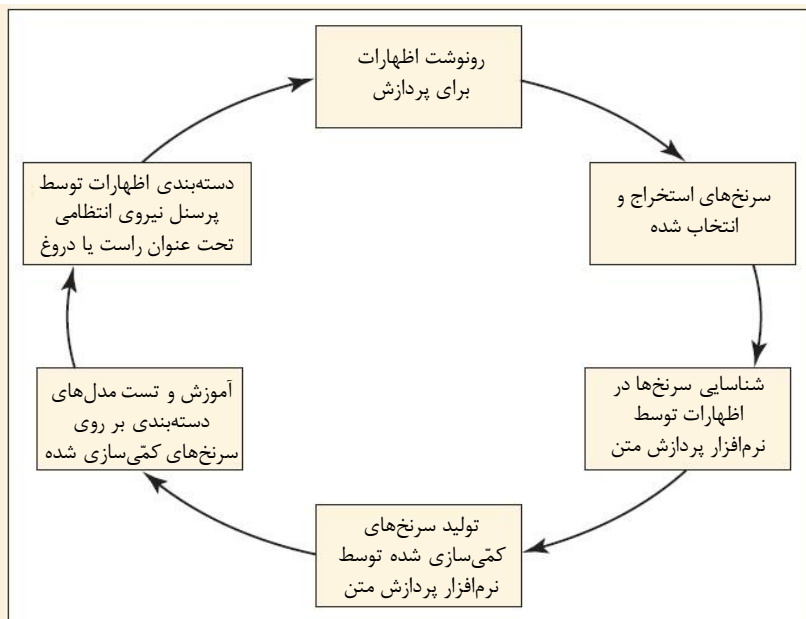
² Overall Analysis System for Intelligence Support (OASIS)

ارتباطات رایانه‌ای به واسطه‌ی پیشرفت‌های به دست آمده در فناوری‌های اطلاعاتی مبتنی بر وب و افزایش جهانی‌سازی همچنان به نفوذ به تمامی جنبه‌های زندگی روزمره ما ادامه داده و آوردگاه‌های جدیدی را برای فریبکاری و دغل‌بازی با خود به همراه می‌آورند. حجم چت‌های متنی، پیام‌های کوتاه، پیام‌های متنی و متون تولید شده توسط جوامع آنلاین به سرعت در حال افزایش است. با رشد انبوه ارتباطات متنی، توان بالقوه‌ی فریب دادن مردم از طریق ارتباطات رایانه‌ای نیز افزایش یافته و اینگونه فریب‌کاری‌ها می‌تواند نتایج فاجعه‌باری را به همراه داشته باشد.

متأسفانه، به طور کلی انسان‌ها در انجام وظایف تشخیص و شناسایی فریب‌ها ضعیف عمل می‌کنند. این پدیده در ارتباطات متنی تشدید می‌شود. بخش بزرگی از پژوهش‌های انجام شده در خصوص تشخیص فریب‌کاری (که آن را اعتبارسنجی نیز می‌نامند) شامل مصاحبه‌ها و ملاقات‌های رو در رو بوده است. اما با رشد ارتباطات متنی، استفاده از تکنیک‌های تشخیص فریب‌کاری مبتنی بر متن امری ضروری است.

تکنیک‌هایی که برای تشخیص موفقیت‌آمیز فریب‌کاری‌ها - یعنی دروغ‌ها - وجود دارند را در حوزه‌های متنوع و گسترده‌ای می‌توان بکار گرفت. در اعمال قانون می‌توان از تکنیک‌ها و ابزارهای پشتیبان تصمیم برای تحقیق و تفحص در خصوص جرائم، غربال‌گری امنیتی در فرودگاه‌ها و پایش ارتباطات تروریست‌های مشکوک استفاده کرد. متخصصان و کارشناسان منابع انسانی می‌توانند از ابزارهای تشخیص فریب‌کاری برای غربال نمودن متقاضیان استفاده کنند. این ابزارها و تکنیک‌ها همچنین قادر به غربال نمودن نامه‌های الکترونیکی برای کشف کلاهبرداری‌ها یا سایر خلاف‌کاری‌هایی که توسط افسران سازمانی ممکن است انجام شود هستند. اگرچه برخی معتقدند که به آسانی می‌توانند افرادی که دروغگو هستند را تشخیص دهند ولی چکیده‌ای از پژوهش‌های انجام شده نشان می‌دهد که به طور متوسط فقط در ۵۴ درصد موارد می‌توانند چنین تشخیصی را به درستی انجام دهند (باند آند دی پائولو ۲۰۰۶). این رقم ممکن است زمانیکه به تشخیص فریب‌کاری در متن نوشتاری می‌پردازیم پایین‌تر نیز باشد.

فولر و همکاران (۲۰۰۸) با استفاده از ترکیبی از تکنیک‌های متن‌کاوی و داده‌کاوی به تحلیل اظهارات افرادی که درگیر جرائم پایگاه‌های نظامی بوده‌اند پرداختند. در این اظهارات، از مضمونین و شاهدها خواسته شده تا آنچه از وقایع و رویدادها به خاطر دارند را از زبان خودشان بنویسند. پرسنل نیروی انتظامی در اظهارات بایگانی شده به جستجوی داده‌هایی پرداختند که بتوانند به آنها در شناسایی افراد فریبکار یا حقیقت‌گو کمک کند. این تصمیمات بر اساس پرونده‌ی مختومه یا شواهد تأییدکننده اتخاذ شده‌اند. به محض اینکه داده‌ی مذکور تحت عنوان فریبنده یا راستگو شناسایی شد، پرسنل نیروی انتظامی آن اطلاعات را حذف نموده و اظهارات را به تیم پژوهش می‌دهند. در مجموع ۳۷۱ اظهارنامه قابل استفاده برای تحلیل بدست آمد. روشی که فولر و همکاران برای تشخیص فریب‌کاری استفاده کردند مبتنی بر فرایندی به نام کاوش ویژگی پیام بوده است که متکی بر تکنیک‌های متن‌کاوی و داده‌کاوی است. تصویر ساده شده‌ای از این فرایند در شکل ۷.۳ نشان داده شده است.



شکل ۷.۳ فرایند تشخیص فریبکاری مبتنی بر متن.

Source: Fuller, C. M., D. Biros, & D. Delen. (2008, January).

Exploration of Feature Selection and Advanced Classification Models for High-Stakes Deception Detection. *Proceedings of the Forty-First Annual Hawaii International Conference on System Sciences (HICSS)*, Big Island, HI: IEEE Press, pp. 80–99.

اول، پژوهشگران داده‌ها را برای پردازش آماده کردند. اظهارات دست‌نویس اصلی می‌بایست در یک فایل پردازش متن وارد می‌شد. دوم، ویژگی‌ها (یعنی سرنخ‌ها) شناسایی شدند. پژوهشگران توانستند ۳۱ ویژگی را شناسایی کنند که بیانگر دسته یا انواع زبان‌هایی بود که تقریباً مستقل از محتوای متنی بودند و به آسانی می‌توان آنها را به روش‌های خودکار تحلیل نمود. به عنوان مثال، ضمایر اول شخص مثل « من » یا « به من » را می‌توان بدون تحلیل متن حواشی آن شناسایی کرد. جدول ۷.۱ فهرستی از گروه‌ها و مثال‌های ویژگی‌های بکار رفته در این مطالعه را نشان می‌دهد.

جدول ۷.۱ گروه‌ها و مثال‌هایی از ویژگی‌های زبان‌شناختی بکار رفته در تشخیص فریب‌کاری

ردیف	گروه (دسته)	سرنخ‌های نمونه
۱	کمیت	شمارش فعل، شمارش جملات اسمی، ...
۲	پیچیدگی	متوسط تعداد گزاره‌ها، متوسط طول جمله، ...
۳	عدم قطعیت	توصیف‌گرها، افعال وجهی، ...
۴	غیرمستقیم و با واسطه بودن	وجه مجهول، مفعول‌سازی، ...
۵	رسایی	احساسی بودن
۶	تنوع	تنوع واژگانی، افزونگی، ...
۷	غیررسمیت	نسبت خطای چاپی
۸	حساسیت و ویژگی	اطلاعات فضایی-زمانی، اطلاعات ادراکی، ...
۹	تأثیر	تأثیر مثبت، تأثیر منفی، ...

این ویژگی‌ها از اظهارات متنی استخراج شده و برای پردازش‌های بعدی وارد یک فایل تخت (مسطح) شدند. پژوهشگران با استفاده از چندین روش انتخاب ویژگی و اعتبارسنجی متقابل ۱۰ مرتبه‌ای، دقت پیش‌بینی سه روش داده‌کاوی معروف را مقایسه کردند. نتایج آنها نشان داد که مدل‌های شبکه عصبی با دقت پیش‌بینی ۷۳.۴۶ درصد بر روی نمونه داده‌های تستی بهترین عملکرد را دارند و پس از آن روش درخت تصمیم با دقت پیش‌بینی ۷۱.۶۰ درصد و رگرسیون لجستیک با دقت ۶۵.۲۸ درصد قرار دارند. نتایج نشان می‌دهند که تشخیص فریب‌کاری خودکار و مبتنی بر متن این پتانسیل را دارد تا به آنهایی که وظیفه‌شان شناسایی دروغ در متن است کمک کنند و می‌توان این روش را با موفقیت بر روی داده‌های واقعی بکار گرفت. دقت این تکنیک‌ها بیشتر از دقت اکثر تکنیک‌های تشخیص فریب‌کاری بوده اگرچه محدود به سرخ‌های متنی بوده است.

سوالاتی برای مورد کاربردی ۷.۳

۱. چرا تشخیص فریب‌کاری دشوار است؟
۲. چگونه از متن‌کاوی/داده‌کاوی می‌توان برای تشخیص فریب‌کاری استفاده کرد؟
۳. به نظر شما چالش‌های اصلی اینگونه سیستم‌های خودکار چیست؟

Sources: Fuller, C. M., D. Biro, & D. Delen. (2008, January).

"Exploration of Feature Selection and Advanced Classification Models for High-Stakes Deception Detection." *Proceedings of the Forty-First Annual Hawaii International Conference on System Sciences (HICSS)*, Big Island, HI: IEEE Press, pp. 80-99; Bond, C. F., & B. M. DePaulo. (2006).

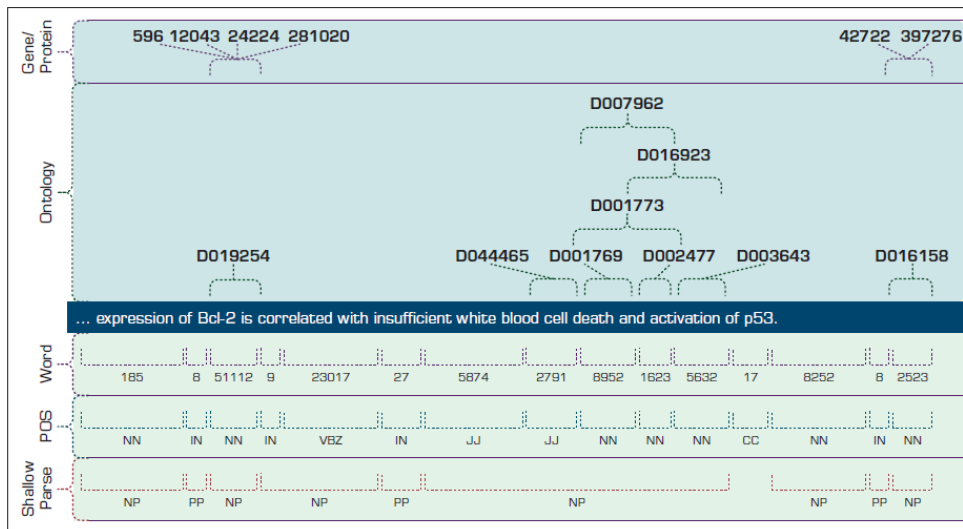
"Accuracy of Deception Judgments." *Personality and Social Psychology Reports*, 10(3), pp. 214-234.

تکنیک‌های آزمایشی از قبیل تحلیل ریزآرایه‌های دی‌ان‌ای، تحلیل متوالی تظاهر ژن و طیف‌سنجی جرمی، حجم انبوهی از داده‌های مربوط به ژن‌ها و پروتئین‌ها را تولید می‌کنند. همانند هر رویکرد آزمایشی دیگر، لازم است که این حجم انبوه داده‌ها در بافتار اطلاعات از پیش مشخص شده در مورد موجودیت‌های بیولوژیکی تحت مطالعه تحلیل شوند. ادبیات موضوع، منبعی ارزشمند از اطلاعات برای اعتبارسنجی و تفسیر آزمایش است. از این رو، توسعه ابزارهای متن‌کاوی خودکار برای کمک به اینگونه تفسیرها یکی از چالش‌های اساسی در پژوهش بیوانفورماتیک جاری است.

آگاهی از مکان یک پروتئین در داخل یک سلول می‌تواند به روشن کردن نقش آن در فرایندهای بیولوژیکی و تعیین توان بالقوه‌ی آن کمک کند. در ادبیات موضوع به توضیح تعداد زیادی از سیستم‌های پیش‌بینی موقعیت و مکان پرداخته شده است که برخی از آنها بر ارگانسیم‌های خاصی متمرکز هستند در حالیکه برخی دیگر سعی دارند تا طیف گسترده‌ای از ارگانسیم‌ها را تحلیل کنند. شاتکای و همکاران (۲۰۰۷) سیستم جامعی را پیشنهاد کردند که از انواع ویژگی‌های مبتنی بر متن و توالی برای پیش‌بینی مکان پروتئین‌ها استفاده می‌کند. نوآوری اصلی سیستم‌ها آنها در روشی است که منابع و ویژگی‌های متنی را انتخاب کرده و آنها را با ویژگی‌های مبتنی بر توالی‌ها ادغام و ترکیب می‌کند. آنها سیستم را بر روی مجموعه داده‌های جدید و قدیمی که قبلاً استفاده شده بود اجرا کردند تا قدرت پیش‌بینی آن را تست کنند. نتایج نشان داد که نتایج سیستم آنها بسیار بهتر از نتایج گزارش شده قبلی است.

چان و همکاران (۲۰۰۶) سیستمی را توضیح داده‌اند که ارتباطات بیماری-ژن را از ادبیات موضوع که از طریق MEDLINE به آنها دسترسی داشته استخراج می‌کند. آنها با استفاده از شش پایگاه داده عمومی یک فرهنگ واژگان از اسامی بیماری‌ها و ژن‌ها ایجاد کرده و از طریق تطبیق با فرهنگ واژگان، روابط را استخراج کردند. از آنجائیکه تطبیق با فرهنگ واژگان، حجم زیادی از مثبت‌های کاذب را تولید می‌کند آنها یک روش سیستم مبتنی بر یادگیری ماشین به نام بازشناسی موجودیت نام‌دار را برای فیلتر کردن تشخیص‌های اشتباه اسامی بیماری/ژن‌ها توسعه دادند. آنها متوجه شدند که موفقیت در استخراج روابط به میزان زیادی

به عملکرد فیلترینگ بازشناسی موجودیت نام‌دار بستگی دارد و این فیلترینگ موجب بهبود دقت استخراج روابط به میزان ۲۶.۷ درصد شده است در حالیکه فراخوانی‌ها به میزان کمی کاهش یافته است. شکل ۷.۴ یک تصویر ساده‌سازی شده از فرایند تحلیل متن چندسطحی را برای کشف ارتباطات ژن-پروتئین در ادبیات موضوع زیست‌پزشکی (ناکوف، شوارتز، وولف، هرست، ۲۰۰۵) نشان می‌دهد.



شکل ۷.۴ تحلیل چندسطحی متن برای شناسایی تعامل بین ژن و پروتئین.

Source: Used with permission of Nakov, P., Schwartz, A., Wolf, B., & Hearst, M. A. (2005). Supporting annotation layers for natural language processing. *Proceedings of the Association for Computational Linguistics (ACL)*, Interactive Poster and Demonstration Sessions, Ann Arbor, MI. Association for Computational Linguistics, 65–68

همانطور که در این مثال ساده شده که از یک جمله‌ی ساده از یک متن زیست‌پزشکی استفاده می‌کند مشاهده می‌کنید، ابتدا (در سه سطح پایینی)، متن با استفاده از برجسب‌گذاری جزء کلمه^۱ و تجزیه و تحلیل سطحی جملات (پارسر کم عمق)^۲ علامت‌گذاری (توکن) می‌شود. عبارات (کلمات) علامت‌گذاری شده سپس با نمایش سلسله‌مراتبی هستی‌شناسی دامنه تطبیق داده می‌شوند تا ارتباطات بین ژن و پروتئین مشخص شود. بکارگیری این روش (با شکل‌های دیگر آن) در ادبیات زیست‌پزشکی، پتانسیل بسیار بالایی را برای رمزگشایی پیچیدگی‌ها در پروژه‌ی ژنوم انسانی^۳ به همراه دارد.

مسئله‌ی متن‌کاوی از اهمیت زیادی برای ناشرانی که پایگاه‌های داده بزرگی از اطلاعاتی را که نیازمند اندیس‌گذاری برای بازیابی ساده‌تر و سریع‌تر هستند نگهداری می‌کنند برخوردار است. این مسئله به ویژه در مورد رشته‌های علمی که حجم زیادی از اطلاعات

¹ Part-of-speech tagging

² Shallow Parsing

³ Human Genome Project

تخصصی و خاص در قالب یک متن نوشتاری ارائه می‌شود صادق است. راهکارهایی از قبیل پروپوزال نیچر^۱ برای یک رابط متن کاوی باز^۲ و تعریف نوع سند انتشار ژورنال^۳ متعلق به مؤسسه ملی سلامت^۴ راه‌اندازی شده‌اند که سرنخ‌های معنایی را برای پاسخگویی پرس و جوهای خاص در یک متن در اختیار ماشین‌ها قرار می‌دهند بدون اینکه موانع ناشر برای دسترسی عمومی به متون را از میان بردارند.

موسسات دانشگاهی نیز راهکارهای متن‌کاوی را راه‌اندازی کرده‌اند. به عنوان مثال، مرکز ملی متن‌کاوی که یک فعالیت مشارکتی بین دانشگاه‌های منچستر و لیورپول است ابزارهای سفارشی‌سازی شده، تسهیلات پژوهشی و مشاوره‌هایی را در خصوص متن‌کاوی به جامعه‌ی دانشگاهی ارائه می‌کند. پژوهش‌های متن‌کاوی که در ابتدا بر علوم زیست‌شناسی و زیست‌پزشکی متمرکز بوده اکنون به علوم اجتماعی نیز گسترش یافته است. در ایالات متحده، دانشکده‌ی اطلاعات در دانشگاه برکلی کالیفرنیا در حال توسعه‌ی برنامه‌ای به نام بیوتکست^۵ است تا به پژوهشگران علوم‌زیستی در تحلیل و کاوش متن کمک کند.

همانطور که در این بخش توضیح داده شد، متن‌کاوی کاربردهای بسیار متنوع و گسترده‌ای در رشته‌های مختلف دارد. برای یک مثال از بکارگیری متن‌کاوی توسط یک تولیدکننده‌ی پیشگام محصولات رایانه‌ای برای درک بهتر نیازها و خواسته‌های مشتریان فعلی و بالقوه خود در خصوص کیفیت و طراحی محصول به مورد کاربردی ۷.۴ مراجعه کنید. تکنیک‌های تحلیلی پیشرفته که از داده‌های ساخت‌یافته و ساخت‌نیافته استفاده می‌کنند در بسیاری از حوزه‌های کاربردی با موفقیت بکار گرفته شده‌اند. مورد کاربردی ۷.۴ یک مثال جالب از بکارگیری طیف گسترده‌ای از قابلیت‌های تحلیلی برای مدیریت سازمان اورلاندو مجیک^۶ در داخل و بیرون از زمین‌های بازی این‌بی‌ای ارائه کرده است.

جادوی پشت مجیک: دسترسی آنی به اطلاعات به اورلاندو مجیک کمک می‌کند تا بازی و تجربه‌ی طرفداران را ارتقاء دهد

مورد کاربردی ۷.۴

اورلاندو مجیک از فروش بلیت تا آرایش و چیدمان آغاز بازی از فصل افتتاحیه‌شان در سال ۱۹۸۹ تا کنون راه درازی را طی کرده است. در آن سالهای ابتدایی این تیم پیروزی‌هایی زیادی نداشت ولی این فرانشیز موجب شد تا فراز و نشیب‌ها برای رقابت در بالاترین سطوح این‌بی‌ای را پشت سر بگذارد.

تیم‌های ورزشی حرفه‌ای در بازارهای کوچکتر اغلب برای ایجاد یک پایگاه درآمدی کافی برای رقابت با رقبایشان در بازارهای بزرگتر، تلاش می‌کنند. اورلاندو مجیک علیرغم اینکه در بیستمین بازار بزرگ حضور داشت توانست با استفاده از اس‌ای‌اس آنالیتیکز و مدیریت داده‌های اس‌ای‌اس جزو پردرآمدترین تیم‌ها در این‌بی‌ای قرار بگیرد.

اورلاندو مجیک اینکار را با مطالعه‌ی بازار فروش مجدد بلیت برای قیمت‌گذاری بهتر بلیت‌ها، پیش‌بینی دارندگان بلیت‌های فصلی و تحلیل امتیازهای انحصاری و و فروش محصولات برای حصول اطمینان از اینکه طرفدارها هر زمان که وارد ورزشگاه می‌شوند هر

¹ Nature

² Open Text Mining Interface

³ Journal Publishing Document Type Definition

⁴ National Institute of Health

⁵ BioText

⁶ Orlando Magic

آنچه که می‌خواهند را می‌توانند تهیه کنند انجام داده است. این کلوپ حتی از اس‌ای‌اس برای کمک به مربیان برای تعیین بهترین آرایش و چیدمان (ترکیب تیم) استفاده کرده است.

الکس مارتینز مدیرعامل اورلاندو مجیک می‌گوید: «بزرگ‌ترین چالش ما این است که تجربه‌ی سفرهای طرفداران خلق کنیم و اس‌ای‌اس به ما کمک می‌کند تا کل اینکار را به شکلی محکم و استوار مدیریت کنیم.» مارتینز که از ابتدای تشکیل اورلاندو مجیک با این باشگاه بوده است همه چیز را دیده و می‌داند که تحلیل‌ها چه ارزش افزوده‌ای برایش به همراه دارد. تحت رهبری مارتینز، فروش بلیت در فصل تا ۱۴،۲۰۰ رسیده و دپارتمان فروش سازمانی شاهد رشد چشمگیری بوده است.

چالش: پر کردن تمام جایگاه‌ها

اما همانند تمامی تیم‌های ورزشی حرفه‌ای، مجیک دائماً به دنبال استراتژی‌های جدیدی بود تا بتواند تمام جایگاه‌های خود را در هر کدام از بازی‌های خانگی خود در ۴۱ سال گذشته پر کند. آنتونی پِرز، معاون استراتژی کسب‌وکار می‌گوید: «ایجاد جریان‌های درآمدی جدید در این دوران که حقوق بازیکنان و مخارج و هزینه‌ها به طور سرسام‌آوری بالا رفته است بسیار اهمیت دارد.» اما با ظهور یک بازار بلیت ثانویه‌ی آنلاین، دستیابی به تراز ۹۰ درصدی تجدید بلیت‌های فصل کار دشوارتری شده است. آنتونی پِرز می‌گوید: «در سال اول، ما شاهد افزایش درآمد ۵۰ درصدی از فروش بلیت بودیم. طی سه سال گذشته این رشد به ۷۵ درصد رسیده است و این تأثیر بسزایی داشته است.»

گروه پِرز از یک رویکرد کل‌نگر و ترکیب داده‌های بدست آمده از تمامی جریان‌های درآمدی (حق امتیاز انحصاری، فروش کالا و فروش بلیت) با داده‌های بیرونی (بازار فروش بلیت ثانویه) برای توسعه‌ی مدل‌هایی استفاده کرده‌اند که به نفع کل سازمان باشد. او توضیح می‌دهد: «ما مثل یک گروه مشاوره‌ی درون‌سازمانی هستیم.»

در خصوص دارندگان بلیت فصلی این تیم از داده‌های خرید و الگوهای تمدید بلیت‌های فصلی برای ساخت مدل‌های درخت تصمیم استفاده کرده است که ثبت‌نام کنندگان را در سه گروه قرار می‌دهند: آنهایی که به احتمال زیاد بلیت‌هایشان را تمدید خواهند کرد، آنهایی که به احتمال زیاد تمدید نخواهند کرد و آنهایی که نظری ندارند. در اینجاست که توجه دپارتمان خدمات مشتری معطوف افرادی که نظری ندارند می‌شوند.

«اس‌ای‌اس به رشد کسب‌وکار ما کمک کرده است. شاید این یکی از بزرگ‌ترین سرمایه‌گذاری‌هایی باشد که ما طی ۶ سال گذشته انجام داده‌ایم چونکه می‌توانیم به رشد درآمدی بالایی که اس‌ای‌اس به ما کمک کرده تا خلق کنیم اشاره کنیم که برای هر کدام از گروه مشتریان مان می‌توانیم پیام‌های خاصی را ارسال کنیم.»

آنها چگونه تمدید بلیت‌های فصلی را پیش‌بینی می‌کنند؟

زمانیکه تحلیل‌ها نشان داد که ۸۰ درصد از درآمد تیم از دارندگان بلیت فصلی کسب می‌شود، تصمیم گرفت تا رویکرد کنشگرانه‌ای را برای تمدید و حساب‌های کاربری پرمخاطره اتخاذ کند. مجیک گوی بلورین ندارد ولی اس‌ای‌اس ماینر اینترپرایز را دارد که به آنها اجازه می‌دهد تا داده‌هایشان را بهتر درک کرده و مدل‌های تحلیلی را توسعه دهند که سه رکن اساسی را برای پیش‌بینی تمدید دارندگان بلیت‌های فصلی با یکدیگر ترکیب کند:

- تصدی (مشتری برای چه مدت دارنده‌ی بلیت بوده است؟)
- استفاده از بلیت (آیا مشتری واقعاً در بازی‌ها حضور داشته است؟)
- فعالیت در بازار ثانویه (آیا بلیت‌های استفاده نشده با موفقیت در سایت‌های ثانویه فروخته شده‌اند؟)

ابزارهای داده‌کاوی به تیم اجازه داد تا امتیازدهی دقیق‌تری برای بازاریابی و حفظ و نگهداشت مشتری انجام دهند.

سهولت استفاده به انتشار پیام ابزارهای تحلیلی کمک می‌کند

پرز از سهولت بکارگیری اس‌ای‌اس بسیار خرسند است و عاملی است که به جای برون‌سپاری کار ما را ترغیب به انجام کار در داخل سازمان می‌کند. تیم پرز فرایندهای تکرارپذیری را برقرار کرده و آنها را خودکارسازی نموده‌اند. دستکاری داده‌ها به حداقل رسیده و به ما زمان بیشتری را برای تفسیر داده‌ها می‌دهد. کاربران در سرتاسر سازمان از جمله مدیران اجرایی از طریق اس‌ای‌اس و بیژوال آنالیتیکز دسترسی آنی به اطلاعات دارند. پرز می‌گوید: «فقط این نیست که ما از این ابزارها به صورت روزانه استفاده می‌کنیم بلکه در سرتاسر روز از آنها برای تصمیم‌گیری استفاده می‌کنیم.»

داده‌محور بودن

مارتینز می‌گوید: «ما چندین سال پیش از یک رویکرد تحلیلی استفاده کردیم و شاهد این هستیم که کل سازمان ما را دگرگون کرده است. تحلیل‌ها به ما در درک بهتر مشتریان، برنامه‌ریزی کسب‌وکار (قیمت‌گذاری بلیت) و ارائه‌ی داده‌ها به تفکیک بازی‌ها و سال‌ها و حتی بر حسب جایگاه بر حسب تقاضا کمک می‌کند.» تحلیل‌ها موجب شده تا خود بازی‌ها نیز دگرگون شوند. تیم‌های تحلیل و GM بر تمامی جنبه‌های بازی از جمله حرکت و جابجایی بازیکن‌ها در میدان بازی نظارت دارند تا بتوانند داده‌ها را به گونه‌ای تغییرشکل دهند که قادر به پیش‌بینی دفاع در برابر برخی تیم‌های خاص باشند. ما اکنون می‌توانیم از خودمان سوال کنیم "مؤثرترین چیدمان و آرایش در یک بازی کدام است؟ کدام تیم می‌تواند در مقابل یک آرایش تیمی دیگر امتیاز بیشتری کسب کند؟ کدام تیم عملکرد دفاعی بهتری نسبت به دیگری دارد؟" مارتینز می‌گوید: «ما یک مجموعه از گزارشات را به صورت دستی تهیه می‌کردیم ولی اکنون می‌توانیم اینکار را با پنج کلیک یک ماوس (به جای پنج ساعت شب‌کاری برای پیش‌بینی نتیجه‌ی بازی فردا) انجام دهیم. ما در عرض چند دقیقه می‌توانیم ده‌ها گزارش را در اختیار پرسنل و کارکنان قرار دهیم. ابزارهای تحلیلی ما را هوشمندتر کرده است.»

بعد چه اتفاقی می‌افتد؟

مارتینز می‌گوید: «قدم بعدی ما در مسیر رشد تحلیلی‌مان این است که به داده‌های آنی و بلادرنگ دسترسی پیدا کنیم. در روز بازی، دسترسی به داده‌های آنی برای ردیابی اینکه چه بلیت‌هایی موجود است و چگونه حداکثر بهره را از آن بلیت‌ها ببریم اهمیت حیاتی برای ما دارد. علاوه بر این، شما شاهد تغییرات تکنولوژیکی مهم و پذیرش فناوری بر روی نیمکت هستید تا ببینید بازی‌ها چگونه رو به جلو جریان دارند. ممکن است همین فصل بعدی دستیار مربیان ما را ببینید که با تبلت‌های آی‌پد به داده‌های آنی دسترسی دارند و از تاکتیک‌ها و فعالیت‌های تیم حریف و تیم خودشان مطلع می‌شوند. این امر در آینده ضروری است. مارتینز می‌گوید: «ما خودمان را برای موفقیت آماده می‌کنیم و در آینده‌ای بسیار نزدیک مجدداً در جایگاهی خواهیم بود که برای قهرمانی کنفرانس و قهرمانی ان‌بی‌ای رقابت کنیم. تمام حرکت‌هایی که امسال کرده‌ایم و سال آینده انجام خواهیم داد برای کسب موفقیت در میدان بازی و خارج از آن است.»

سوالاتی برای مورد کاربردی ۷.۴

۱. بر اساس این مطالعه موردی، اورلاندو مجیک با چه چالش‌های اساسی روبرو بود؟

۲. تحلیل‌ها چگونه به اورلاندو مجیک کمک کردند تا بر برخی از بزرگ‌ترین چالش‌هایشان در داخل و خارج میدان بازی غلبه کنند؟
۳. آیا می‌توانید کاربردهای دیگری از تحلیل‌ها را در ورزش و بویژه در خصوص اورلاندو مجیک نام ببرید؟ برای یافتن پاسخ این سوال می‌توانید در وب به جستجو بپردازید.

Source: SAS Customer Story, "The magic behind the Magic: Instant access to information helps the Orlando Magic up their game and the fan's experience" at https://www.sas.com/en_us/customers/orlando-magic.html and <https://www.nba.com/magic/news/denton-25-years-magic-history> (accessed November 2018).

سوالاتی برای مرور بخش ۷.۴

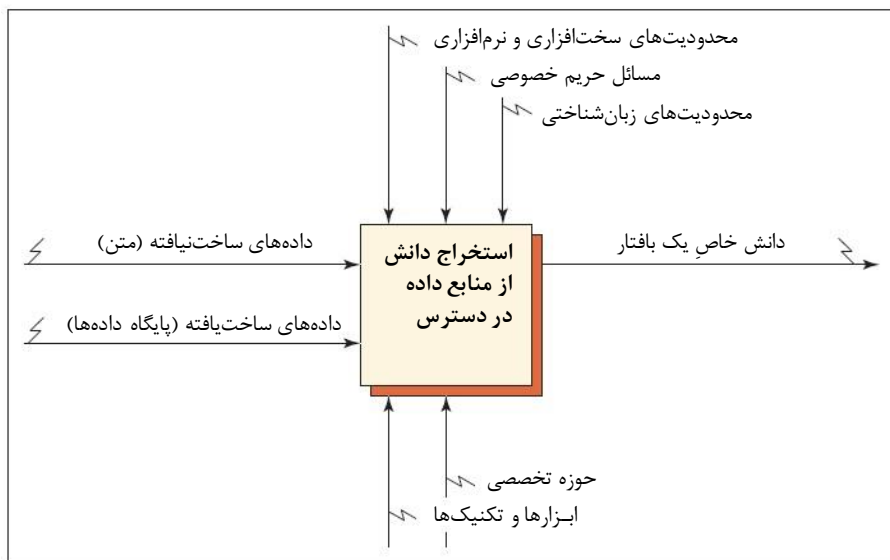
۷. برخی از کاربردهای متن‌کاوی در بازاریابی را نام برده و به اختصار توضیح دهید.
۸. چگونه می‌توان از متن‌کاوی در امنیت و ضدتروریسم استفاده کرد؟
۹. برخی از کاربردهای متن‌کاوی در زیست‌پزشکی چه هستند؟

برای موفقیت‌آمیز بودن مطالعات متن‌کاوی باید از یک متدولوژی درست مبتنی بر بهترین تجارب استفاده کنیم. ما به یک مدل فرایندی استانداردسازی شده مشابه متدولوژی کریسپ که یک استاندارد صنعتی برای پروژه‌های داده‌کاوی است نیاز داریم. اگرچه اکثر بخش‌های متدولوژی کریسپ را می‌توان در پروژه‌های متن‌کاوی نیز بکار برد ولی یک مدل فرایندی خاص برای متن‌کاوی شامل فعالیت‌های مبسوط‌تری برای پیش‌پردازش داده‌ها است. شکل ۷.۵ یک نمودار سطح صفر سطح بالا از یک فرایند متن‌کاوی معمول را نشان می‌دهد (دیلن آند کراس‌لند ۲۰۰۸). این نمودار سطح صفر، دامنه‌ی فرایند را نشان می‌دهد که رابط‌های آن با محیط بزرگ‌تری در ارتباط هستند. در واقع، مرزها و محدوده‌ی حول یک فرایند خاص را ترسیم کرده تا آنچه به طور خاص در فرایند متن‌کاوی وجود دارد یا ندارد را شناسایی کند.

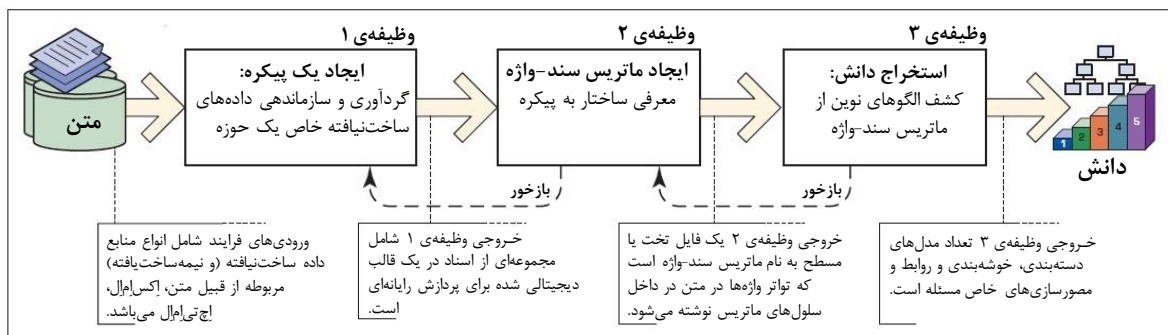
همانطور که در این نمودار مشاهده می‌کنید، ورودی (ارتباطات رو به داخل به سمت چپ کادر) فرایند کشف دانش مبتنی بر متن شامل داده‌های ساخت‌یافته و ساخت‌نیافته‌ای است که گردآوری، انبارش و در اختیار فرایند قرار گرفته‌اند. خروجی (ارتباطات به سمت خارج از سمت راست کادر) فرایند شامل دانش خاص آن بافتار است که می‌توان از آن برای تصمیم‌گیری استفاده کرد. کنترل‌ها که محدودیت‌ها نیز نامیده می‌شوند (ارتباطات به سمت داخل بالای کادر) شامل محدودیت‌های نرم‌افزاری و سخت‌افزاری، مسائل حریم خصوصی و مشکلات مربوط به پردازش متن هستند که به شکل زبان طبیعی بیان می‌شوند. مکانیزم‌های فرایند (ارتباطات رو به داخل به سمت پایین کادر) شامل تکنیک‌های مناسب، ابزارهای نرم‌افزاری و تخصص دامنه هستند. هدف اصلی متن‌کاوی (در مقوله‌ی کشف دانش) این است که داده‌های ساخت‌نیافته (متنی) (به همراه داده‌های ساخت‌یافته در صورتیکه مرتبط با مسئله باشند) را برای استخراج الگوهای معنادار و اقدام‌پذیر برای تصمیم‌گیری بهتر پردازش کند. در یک سطح بسیار بالا، فرایند متن‌کاوی را می‌توان به سه وظیفه‌ی متوالی تجزیه کرد که هر کدام ورودی‌ها و خروجی‌های خاص خودشان را دارند (شکل ۷.۶). اگر به هر دلیلی خروجی یک وظیفه، آن چیزی نباشد که مورد انتظار بوده است باید مجدداً به اجرای وظیفه‌ی قبلی بازگشت.

هدف اصلی این فعالیت، گردآوری تمامی اسناد مرتبط با بافتار (دامنه‌ی مورد نظر) مورد مطالعه است. این گردآوری ممکن است شامل اسناد متنی، فایل‌های اِکس‌اِم‌اِل، نامه‌های الکترونیکی، صفحات وب و یادداشت‌های کوتاه باشد. علاوه بر داده‌های متنی که به آسانی در دسترس هستند ممکن است متن فایل‌های صوتی ضبط شده نیز با استفاده از الگوریتم‌های تشخیص گفتار استخراج شده و بخشی از این مجموعه متنی شوند.

پس از گردآوری، مستندات متنی به نحوی تبدیل و سازماندهی می‌شوند که تمامی آنها به یک فرم یکسان قابل ارائه (مثلاً فایل‌های متنی اَسکی) برای پردازش رایانه‌ها درآیند. سازماندهی اسناد می‌تواند به سادگی گردآوری گزیده‌های متنی دیجیتالی شده که در یک فایل ذخیره شده‌اند یا پیوندهایی به مجموعه‌ای از صفحات وب در یک حوزه‌ی خاص باشد. بسیاری از ابزارهای نرم‌افزاری تجاری متن‌کاوی موجود در بازار می‌توانند این‌ها را به عنوان ورودی پذیرفته و آنها را برای پردازش به یک فایل تخت (مسطح) تبدیل کنند. به نحوی دیگر می‌توان فایل تخت یا مسطح را در خارج از محیط نرم‌افزار متن‌کاوی آماده کرده و سپس آن را به عنوان ورودی برنامه‌ی متن‌کاوی ارائه کرد.



شکل ۷.۵ نمودار سطح صفر برای فرایند متن‌کاوی



شکل ۷.۶ فرایند متن‌کاوی سه مرحله‌ای

در این مرحله، اسناد دیجیتالی شده و سازماندهی شده (پیکره) برای ایجاد یک ماتریس سند-واژه بکار می‌روند. در این ماتریس، ردیف‌ها بیانگر اسناد و ستون‌ها بیانگر واژه‌ها هستند. ارتباطات بین اسناد و واژه‌ها توسط اندیس‌ها نشان داده می‌شوند (یعنی یک سنجی رابطه‌ای که می‌تواند تعداد دفعات وقوع یا مشاهده‌ی یک واژه در اسناد باشد). شکل ۷.۷ یک مثال ساده از یک ماتریس سند-واژه را نشان می‌دهد.

واژه / سند	ریسک سرمایه‌گذاری	مدیریت پروژه	مهندسی نرم‌افزار	برنامه‌نویسی	SAP
Document 1	1			1	
Document 2		1			
Document 3			3		1
Document 4		1			
Document 5			2	1	
Document 6	1			1	
...					

شکل ۷.۷ یک ماتریس سند-واژه ساده

هدف در اینجا این است که فهرست اسناد سازماندهی شده (پیکره) را به یک ماتریس سند-واژه تبدیل کنیم که سلول‌های این ماتریس با مناسب‌ترین اندیس‌ها پر می‌شوند. فرض بر این است که ذات و اساس یک سند را می‌توان بر حسب یک فهرست و فراوانی عبارات بکار رفته در آن سند نمایش داد. با وجود این، آیا همه‌ی عبارات در این اسناد از اهمیت یکسانی برخوردارند؟ مشخصاً پاسخ منفی است. برخی عبارات از قبیل حروف تعریف، افعال کمکی و عبارات بکار رفته در تقریباً تمامی اسناد پیکره هیچگونه ویژگی و توان متمایزکننده‌ای ندارند و از این رو باید از فرایند اندیس‌گذاری کنار گذاشته شوند. این فهرست از عبارات که آنها را معمولاً/بسیار واژه‌ها یا واژه‌های پالایشی یا واژه‌های تصفیه‌شونده می‌نامند مختص هر حوزه مطالعه بوده و باید توسط کارشناسان و خبرگان آن حوزه تعیین و شناسایی شود. از طرف دیگر، می‌توان مجموعه‌ای از عبارات از پیش تعیین شده را انتخاب کرد که

اسناد را بر اساس آن اندیس‌گذاری نمود (این فهرست عبارات را فرهنگ واژگان می‌نامند. علاوه بر این، مترادف‌ها (زوج عباراتی که یکسان در نظر گرفته می‌شوند) و جملات خاص (مثلاً «برج ایفل») را نیز می‌توان تعیین کرد تا وارد کردن اندیس‌ها دقیق‌تر انجام شود.

یک عملیات فیلترسازی دیگر که باید برای ایجاد هرچه دقیق‌تر اندیس‌ها انجام شود ^۱ بن‌سازی نام دارد که تقلیل و کاهش واژگان تا سطح ریشه یا بن کلمه می‌باشد به عنوان مثال، شکل‌های مختلف دست‌ورزبانی یا افعال صرف شده به عنوان کلمه‌ی یکسان شناسایی و اندیس‌گذاری می‌شوند. به عنوان مثال، با انجام بن‌سازی مطمئن می‌شویم که *modeled* و *modeling* به عنوان یک کلمه‌ی *model* شناسایی می‌شوند.

نسل اول ماتریس‌های سند-واژه شامل تمامی عبارات منحصر‌بفرد موجود در پیکره بودند (در داخل ستون‌ها) به استثنای عباراتی که در فهرست ایست‌واژه‌ها آورده شده بودند و تمامی اسناد (ردیف‌ها) و تعداد دفعات تکرار هر عبارت در هر سند (مقادیر داخل سلول‌ها). اگر پیکره شامل تعداد اسناد زیادی باشد که معمولاً نیز این چنین است، آنگاه این شانس وجود دارد که تعداد عبارات بسیار زیادی در ماتریس سند-واژه وجود داشته باشد. پردازش چنین ماتریس بزرگی ممکن است زمان‌بر بوده و مهمتر اینکه ممکن است منجر به استخراج الگوهای نادقیق شود. در این حالت، یکی از دو گزینه زیر را باید انتخاب کنیم: (۱) بهترین روش نمایش اندیس‌ها چیست؟ (۲) چگونه می‌توانیم ابعاد این ماتریس را به یک اندازه‌ی قابل مدیریت کاهش دهیم؟

طرز نمایش اندیس‌ها پس از اینکه اسناد ورودی اندیس‌گذاری شده و فراوانی کلمات (در هر سند) شمارش شد، چندین عملیات تبدیل دیگر را می‌توان برای خلاصه‌سازی و تجمیع اطلاعات استخراج شده انجام داد. فراوانی‌های کلمات در حالت اولیه معمولاً بیانگر این است که یک واژه در هر سند تا چه اندازه مهم یا برجسته است. به بیان خاص، کلماتی که تعداد دفعات بیشتری در یک سند نمایان می‌شوند توصیف‌گرهای بهتری از محتوای آن سند می‌باشند. با وجود این، منطقی نیست که فرض کنیم شمارش تعداد کلمات به خودی خود نشان دهنده‌ی اهمیت آنها و اسناد می‌باشد. به عنوان مثال، اگر کلمه‌ای یک مرتبه در سند A ولی سه مرتبه در سند B نمایان شود، لزوماً به این معنی نیست که این کلمه در سند B سه برابر مهمتر از کلمه‌ی مذکور در سند A است. برای اینکه ماتریس سند-واژه‌ی سازگارتری را برای تحلیل‌های بعدی داشته باشیم این اندیس‌های اولیه را باید نرمال‌سازی کنیم. برای اینکار به جای اینکه تعداد دفعات تکرار کلمه یا فراوانی آن را در متن نشان دهیم، باز نمود عددی بین عبارات و اسناد را می‌توانیم با استفاده از چند روش مثل فراوانی لگاریتمی^۲، فراوانی دودویی^۳ و فراوانی معکوس سند^۴ نرمال‌سازی کنیم.

کاهش تعداد ابعاد ماتریس چونکه ماتریس سند-واژه در اغلب مواقع بسیار بزرگ و پراکنده است (اکثر سلول‌ها با عدد صفر پر شده‌اند)، سوال مهم دیگری که مطرح است این است که « چگونه می‌توانیم تعداد ابعاد این ماتریس را تا اندازه‌ی قابل مدیریت کاهش دهیم؟ » برای مدیریت اندازه و حجم ماتریس چندین گزینه وجود دارد:

- یک خبره در حوزه، فهرست عبارات را بررسی کرده و آنهایی را که چندان ارتباطی به بافتار و حوزه‌ی مطالعه ندارند حذف می‌کند (این یک فرایند دستی و بسیار پرزحمت است).

¹ Stemming

² Log Frequency

³ Binary Frequency

⁴ Inverse Document Frequency

- عباراتی که در تعداد اسناد بسیار معدود و به دفعات بسیار کمی تکرار شده‌اند را حذف کنید.
- ماتریس را با استفاده از SVD تغییر دهید.

تجزیه مقدارهای منفرد^۱ که ارتباط نزدیکی با تحلیل مؤلفه‌های اصلی دارد می‌تواند تعداد ابعاد ماتریس ورودی (تعداد اسناد ورودی بر حسب تعداد عبارات یا کلمات استخراج شده) را به یک فضای کم‌بعد کاهش دهد که در آن هر بُعد متوالی بیانگر بزرگترین درجه‌ی تغییرپذیری (بین کلمات و اسناد) ممکن است (مانینگ آند شوتز ۱۹۹۹). به طور ایده‌آل، تحلیلگر ممکن است دو یا سه بُعد بسیار مهم را شناسایی کند که بیشترین تغییرپذیری (تفاوت‌ها) بین کلمات و اسناد مربوط به آنها باشد و بدین ترتیب فضای معنایی نهفته^۲ برای سازماندهی کلمات و اسناد را در این تحلیل شناسایی کند. پس از شناسایی این ابعاد، «معنای» زیربنایی آنچه در اسناد مستتر شده را می‌توان استخراج نمود.

با استفاده از ماتریس ساختارمند سند-واژه که با عناصر داده‌ای ساخت‌یافته‌ی دیگری ارتقاء یافته است، الگوهای نوینی در بافتار مسئله‌ی خاصی که مورد بررسی است استخراج می‌شود. گروه‌های اصلی روش‌های استخراج دانش عبارتند از دسته‌بندی، خوشه‌بندی، ارتباط و تحلیل روند. در ادامه به شرح مختصر هر کدام می‌پردازیم.

دسته‌بندی به طور قطع می‌توان دسته‌بندی اشیاء مشخص را متداول‌ترین موضوع کشف دانش در تحلیل منابع داده پیچیده دانست. وظیفه دسته‌بندی این است که یک نمونه داده مفروض را در قالب دسته‌های (یا کلاس‌های) از پیش تعیین شده دسته‌بندی کند. در حوزه متن‌کاوی، این وظیفه را دسته‌بندی متن در دسته یا گروه‌های (موضوع، عنوان یا مفهوم) و مجموعه‌ای از اسناد متنی می‌نامند که هدف آن یافتن عنوان صحیح (موضوع یا مفهوم) برای هر سند با استفاده از مدل‌های توسعه یافته با مجموعه داده‌های آموزشی است که شامل اسناد و گروه یا دسته‌های اسناد می‌باشد. امروزه، از دسته‌بندی خودکار متن در حوزه‌های مختلف از جمله اندیس‌گذاری خودکار یا نیمه‌خودکار (تعاملی) متن، فیلترسازی هرزنامه‌ها، دسته‌بندی صفحات وب تحت کاتالوگ‌های سلسله‌مراتبی، تولید خودکار فراداده‌ها و تشخیص سبک متن استفاده می‌شود.

دو روش اصلی برای دسته‌بندی متن عبارتند از مهندسی دانش^۳ و یادگیری ماشینی^۴ (فلدمن آند سانجر ۲۰۰۷). با روش مهندسی دانش، دانش یک فرد خبره در مورد گروه و دسته‌ها یا به شکل اخباری و اعلانی^۵ یا به شکل قوانین دسته‌بندی رویه‌ای^۶ در داخل سیستم کدگذاری می‌شود. در روش یادگیری ماشینی، با استفاده از یک فرایند استنتاجی عمومی یک دسته‌بند از طریق یادگیری از مجموعه‌ای از مثال‌های از پیش دسته‌بندی شده ایجاد می‌شود. با افزایش تعداد اسناد با نرخ‌ی نمایی و دشوارتر شدن کار افراد خبره در حوزه دانشی، محبوبیت روش یادگیری ماشینی بیشتر افزایش یافته است.

¹ Singular Value Decomposition (SVD)

² Latent semantic space

³ Knowledge Engineering

⁴ Machine-Learning

⁵ Declaratively

⁶ Procedural Classification Rules

خوشه‌بندی خوشه‌بندی یک فرایند نظارت‌نشده است که در آن اشیاء در گروه‌های « طبیعی » به نام خوشه‌ها دسته‌بندی می‌شوند. در مقایسه با دسته‌بندی یا طبقه‌بندی که از مجموعه‌ای از مثال‌های آموزشی از پیش دسته‌بندی شده برای توسعه و ایجاد یک مدل بر اساس ویژگی‌های توصیفی دسته‌ها برای دسته‌بندی یک مثال جدید و نام‌گذاری نشده استفاده می‌کند در خوشه‌بندی، مسئله این است که مجموعه‌ای از اشیاء برچسب‌گذاری نشده (مثلاً اسناد، اظهارات مشتریان، صفحات وب) را بدون هرگونه دانش قبلی در قالب خوشه‌های معنادار گروه‌بندی کنیم.

خوشه‌بندی در طیف وسیعی از کاربردها از بازیابی سند تا جستجوی بهتر محتوای وب کاربرد دارد. در واقع، یکی از کاربردهای بارز خوشه‌بندی، تحلیل و یافتن مسیر از بین مجموعه متن‌های بسیار بزرگ مثل صفحات وب است. فرضیه‌ی زیربنایی اساسی این است که اسناد مرتبط در مقایسه با اسناد نامرتبط شباهت بیشتری بهم دارند. اگر این فرضیه درست باشد، خوشه‌بندی اسناد بر اساس تشابه محتوای آنها موجب بهبود اثربخشی جستجو می‌شود (فلدمن آند سانجر ۲۰۰۷):

- **بهبود فراخوانی جستجو.** از آنجائیکه این روش مبتنی بر تشابه کلی به جای وجود یک کلمه یا واژه در متن است، خوشه‌بندی می‌تواند موجب بهبود فراخوانی جستجوهای مبتنی بر پرس و جو شود به نحویکه زمانیکه یک پرس و جو با یک سند انطباق دارد، کل خوشه‌ی مربوط به آن فراخوانی می‌شود.
- **بهبود دقت جستجو.** خوشه‌بندی همچنین می‌تواند دقت جستجو را افزایش دهد. با افزایش تعداد اسناد در یک مجموعه، جستجو در یک فهرست برای یافتن اسناد انطباق یافته کار دشواری خواهد بود. خوشه‌بندی از طریق گروه‌بندی اسناد در تعدادی از گروه‌های کوچکتر اسناد مرتبط، مرتب‌سازی آنها بر اساس میزان ارتباط و فقط با بازگرداندن اسنادی که بیشتری ارتباط را در گروه دارند می‌تواند به این موضوع کمک کند.

دو روش متداول خوشه‌بندی عبارتند از خوشه‌بندی تجمعی و پراکندگی و خوشه‌بندی بر اساس پرس و جو:

- **تجمعی/پراکندگی:** این روش جستجوی اسناد از خوشه‌بندی برای افزایش کارایی جستجوی اسناد توسط انسان‌ها و زمانیکه نتوان یک جستجو بر اساس پرس و جو را انجام داد استفاده می‌کند. به بیان دیگر، در این روش جدولی از محتوای مورد نظر برای گردآوری به شکل پویا تولید شده و خودش را متناسب با انتخاب کاربر، وفق داده و تغییر می‌دهد.
- **خوشه‌بندی بر اساس پرس و جو:** در اینجا از یک روش خوشه‌بندی سلسله‌مراتبی استفاده می‌شود که مرتبط‌ترین اسناد برای یک پرس و جو در خوشه‌های محدودتر و کوچکتری ظاهر می‌شود که این خوشه‌ها در دل خوشه‌های بزرگتری که حاوی اسنادی با شباهت کمتر هستند قرار دارند و بدین ترتیب طیفی از اسناد بهم مرتبط را ایجاد می‌کند. این روش به خوبی برای مجموعه اسنادی که حجم بالایی دارند جواب داده است.

ارتباط (وابستگی) وابستگی یا یادگیری قانون وابستگی در داده‌کاوی، یک تکنیک معروف برای کشف ارتباطات جالب بین متغیرها در پایگاه داده‌های بزرگ است. ایده‌ی اصلی در اینجا تولید قوانین وابستگی (یا حل مسائل تحلیل سبد بازار) برای شناسایی مجموعه‌هایی است که با یکدیگر همراه می‌شوند.

در متن کاوی، وابستگی‌ها به طور خاص به ارتباطات مستقیم بین مفاهیم (عبارات) یا مجموعه مفاهیم اشاره دارند. قانون وابستگی مجموعه مفهوم $A + C$ که دو مجموعه مفهوم تکرار شونده‌ی A و C را بهم مرتبط می‌کند را می‌توان با استفاده از دو سنجهی

پایه‌ی پشتیبانی و اطمینان کمی‌سازی کرد. در این مورد، اطمینان عبارتست از درصد اسنادی که شامل تمامی مفاهیم C در همان زیرمجموعه‌ای از اسناد هستند که شامل تمامی مفاهیم A و C هستند. به عنوان مثال، در یک مجموعه از اسناد که مفهوم « شکست در اجرای نرم‌افزار » اغلب ممکن است با مفاهیم « برنامه‌ریزی منابع انسانی » و « مدیریت ارتباط با مشتریان » با پشتیبانی (۰.۴٪) و اطمینان (۰.۵۵٪) ظاهر شود که بدین معناست که ۴ درصد از اسناد هر سه مفهوم نشان داده شده در یک سند و اسنادی که شامل « برنامه‌ریزی منابع سازمانی » هستند را شامل می‌شوند و ۵۵ درصد از آنها همچنین شامل « برنامه‌ریزی منابع سازمانی » و « مدیریت ارتباط با مشتری » هستند. متن کاوی به روش قوانین وابستگی برای تحلیل ادبیات منتشر شده (مقالات دانشگاهی و اخبار منتشر شده بر روی وب) برای ترسیم نمودار شیوع آنفولانزای مرغی استفاده شد (مهمقوب و همکاران ۲۰۰۸). هدف از اینکار، شناسایی خودکار وابستگی بین مناطق جغرافیایی، گسترش بین گونه‌ها و اقدامات متقابل (درمان) بوده است.

تحلیل روند روش‌های اخیر تحلیل روند در متن کاوی مبتنی بر این مفهوم بوده‌اند که انواع مختلف توزیع‌های مفاهیم، تابعی از مجموعه اسناد هستند یعنی اینکه مجموعه‌های مختلف برای مجموعه یکسانی از مفاهیم منجر به توزیع‌های مختلفی از مفاهیم می‌شوند. از این رو، می‌توان دو توزیع را که مشابه یکدیگر نیستند با هم مقایسه کرد مگر اینکه از زیرمجموعه‌های متفاوتی باشند. یک جهت‌گیری بارز از این نوع تحلیل این است که دو مجموعه از یک منبع ولی در دو نقطه‌ی زمانی متفاوت داشته باشیم. دلیل آند کراسلند (۲۰۰۸) از تحلیل روند بر روی تعداد زیادی از مقالات دانشگاهی برای شناسایی تکامل مفاهیم کلیدی در رشته‌ی سیستم‌های اطلاعاتی استفاده کردند.

همانطور که در این بخش توضیح داده شد، روش‌های متعددی برای متن کاوی وجود دارد. مورد کاربردی ۷.۵ استفاده از چندین تکنیک مختلف برای تحلیل مجموعه بزرگی از ادبیات موضوع را توضیح می‌دهد.

بررسی دقیق ادبیات پژوهشی با متن کاوی

مورد کاربردی ۷.۵

پژوهشگرانی که در ادبیات موضوع به جستجو و نقد می‌پردازند با وظیفه‌ای سنگین و بیش از پیش پیچیده مواجه هستند. در توسعه و گسترش بدنه‌ی دانش، تلاش برای گردآوری، سازماندهی، تحلیل و تلفیق اطلاعات موجود در ادبیات به ویژه از رشته‌ی تحصیلی شخص بسیار حائز اهمیت بوده است. با افزایش فراوانی پژوهش‌های مهمی که در رشته‌های مربوطه گزارش می‌شود و حتی در رشته‌های مطالعاتی که نامرتب دانسته می‌شدند، وظیفه‌ی پژوهشگران بیش از پیش سنگین شده است.

در جریانهای پژوهشی جدید، وظیفه‌ی پژوهشگر می‌تواند خسته کننده‌تر و پیچیده‌تر باشد. تلاش برای کشف کارهای مربوطه که دیگران گزارش کرده‌اند، در بهترین حالت و در صورت نیاز به بررسی مقالات انتقادی نوشته شده بر روی ادبیات منتشر شده حتی غیرممکن باشد. حتی با لژیونی از دانش آموختگان فارغ التحصیل اختصاصی یا همکاران مفید، پوشش دادن همه موضوعات منتشر شده کار دشواری است.

هرساله کنفرانس‌های دانشگاهی متعددی برگزار می‌شود. علاوه بر توسعه‌ی بدنه‌ی دانش‌ تمرکز جاری کنفرانس‌ها، سازماندهی‌کنندگان کنفرانس‌ها اغلب تمایل دارند تا کارگاه‌های بیشتری را برگزار کنند. در بسیاری از موارد، این رویدادهای تکمیلی با هدف معرفی جریان‌های پژوهشی مهم در رشته‌های مورد نظر و تلاش برای شناسایی «بزرگ‌ترین رویدادهای آینده» در حوزه‌های تمرکز و علائق پژوهشی به شرکت‌کنندگان برگزار می‌شوند. شناسایی موضوعات منطقی برای اینگونه کارگروه‌ها اغلب به صورت ذهنی انجام می‌شود تا به صورت عینی که از پژوهش‌های موجود و نوظهور استخراج شود.

در مطالعه‌ی اخیر که توسط دیلن و کراسلند (۲۰۰۸) انجام شده، روشی برای کمک و ارتقاء تلاش‌های پژوهشگران پیشنهاد شده که امکان یک تجزیه و تحلیل نیمه‌خودکار حجم انبوهی از ادبیات موضوع منتشر شده را با بکارگیری متن‌کاوی فراهم می‌کند. با استفاده از کتابخانه‌های دیجیتالی استاندارد و موتورهای جستجوی نشریات آنلاین، مؤلفان می‌توانند کلیه‌ی مقالات موجود در سه ژورنال اصلی در رشته‌ی سیستم‌های اطلاعات مدیریت را بارگذاری و گردآوری کنند: *فصلنامه‌ی سیستم‌های اطلاعات مدیریت*^۱، *پژوهش سیستم‌های اطلاعاتی*^۲ و *ژورنال سیستم‌های اطلاعات مدیریت*^۳. برای اینکه برای هر سه ژورنال یک بازه‌ی زمانی در نظر گرفته شود، ژورنالی که آخرین مقاله را به صورت دیجیتالی منتشر کرده است به عنوان زمان شروع این مطالعه در نظر گرفته شده است (یعنی مقالات JMIS از سال ۱۹۹۴ به صورت دیجیتالی در دسترس بوده است). برای هر مقاله، دیلن و کراسلند عنوان، چکیده، فهرست مؤلفان، کلیدواژه‌ها، شماره نشریه، شماره جلد و سال نشر را استخراج کردند. سپس تمامی این اطلاعات را در یک فایل پایگاه داده‌ی ساده وارد کردند. آنها همچنین فیلدی را در مجموعه داده‌های ترکیبی گنجاندند که نوع ژورنال برای هر مقاله را مشخص می‌نمود. یادداشت‌های ویراستارها، یادداشت‌های پژوهشی و اظهارنظرهای اجرایی از مجموعه داده‌ها حذف شدند. جدول ۷.۲ نشان می‌دهد که داده‌ها چگونه در یک قالب جدولی ارائه شده‌اند.

در فاز تحلیل، پژوهشگران تصمیم گرفتند که فقط از چکیده یک مقاله به عنوان منبع استخراج اطلاعات استفاده کنند. آنها به دو دلیل تصمیم گرفتند که از کلیدواژه‌های منتشر شده استفاده نکنند: (۱) تحت شرایط عادی، چکیده مقالات شامل کلیدواژه‌ها نیز می‌شود به همین خاطر در نظر گرفتن کلیدواژه‌ها برای تحلیل به معنای تکرار اطلاعات یکسان و در نظر گرفتن وزن نامناسب برای آنها است و (۲) کلیدواژه‌های فهرست شده می‌توانند عباراتی انتخاب شده باشند که مؤلفان دوست دارند تا مقاله آنها به این کلیدواژه‌ها مرتبط شود (در صورتیکه ممکن است محتوای مقاله کاملاً خلاف این باشد) و بدین ترتیب موجب بروز خطا یا انحراف در تحلیل محتوا شود.

در اولین مطالعه‌ی اکتشافی، هدف این بود که نگاهی به چشم‌انداز طولی این سه ژورنال (یعنی تکامل عناوین و موضوعات پژوهشی در طول زمان) داشته باشیم. برای انجام یک مطالعه‌ی طولی، دیلن و کراسلند یک بازه‌ی زمانی ۱۲ ساله را (از ۱۹۹۴ تا ۲۰۰۵) به چهار دوره‌ی زمانی ۳ ساله برای هر یک از سه ژورنال تقسیم کردند. این چارچوب منجر به ۱۲ آزمایش متن‌کاوی با ۱۲ مجموعه داده‌ی دو به دو ناسازگار شد. در این مرحله، برای هر یک از ۱۲ مجموعه داده، پژوهشگران از متن‌کاوی برای استخراج توصیفی‌ترین عبارات از این مجموعه مقالات و بر اساس چکیده‌ها نمودند. نتایج در قالب جدول درآورده شده و برای دوره‌های زمانی متغیر بر حسب عباراتی که در این سه ژورنال منتشر شده است مورد ارزیابی و بررسی قرار گرفتند.

¹ MIS Quarterly (MISQ)

² Information Systems Research (ISR)

³ Journal of Management Information Systems (JMIS)

جدول ۷.۲ نمایش جدولی فیلهای موجود در مجموعه داده‌های ترکیب شده

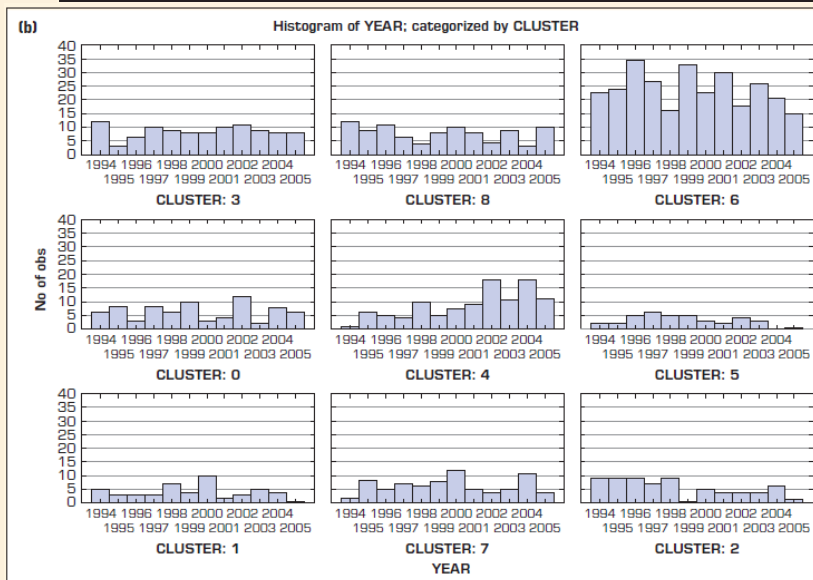
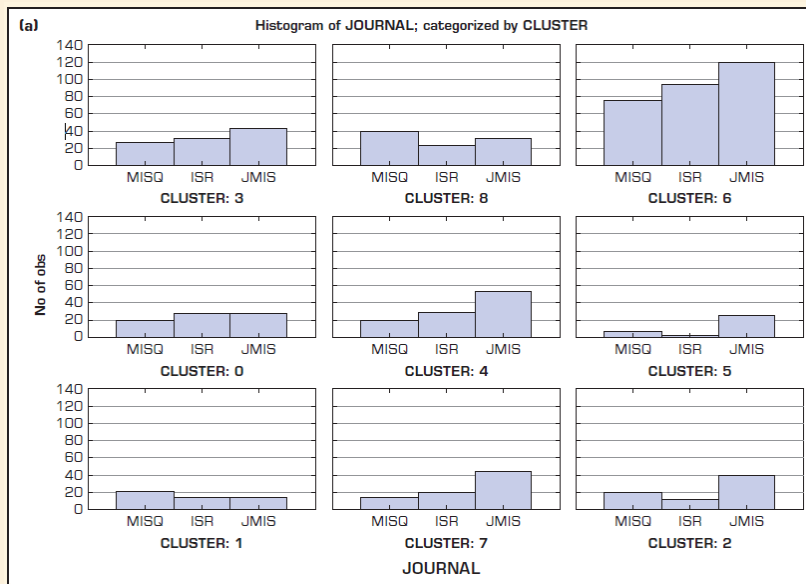
	A	B	C	D	E	F	G	H	I	J
1	ID	YEAR	JOURNAL	ABSTRACT						
2	PID001	2005	MISQ	The need for continual value innovation is driving supply chains to evolve from						
3	PID002	1999	ISR	Although much contemporary thought considers advanced information techn						
4	PID003	2001	JMIS	When producers of goods (or services) are confronted by a situation in which						
5	PID004	1995	ISR	Preservation of organizational memory becomes increasingly important to org						
6	PID005	1994	ISR	The research reported here is an adaptation of a model developed to measure						
7	PID006	1995	MISQ	This study evaluates the extent to which the added value to customers from						
8	PID007	2003	MISQ	This paper reports the results(-) of a field-study of six medical project teams t						
9	PID008	1999	JMIS	Researchers and managers are beginning to realize that the full advantages c						
10	PID009	2000	JMIS	The Internet commerce technologies have significantly reduced sellers' costs						
11	PID010	1997	ISR	Adaptive Structuration Theory (AST) is rapidly becoming an influential theoret						
12	PID011	1995	JMIS	Research shows that group support systems (GSS) have dramatically increa						
13	PID012	2000	MISQ	Increasingly, business leaders are demanding that IT play the role of a busine						
14	PID013	2001	ISR	Alignment between business strategy and IS strategy is widely believed to in						
15	PID014	1999	JMIS	A framework is outlined that includes the planning of and setting goals for IT,						
16	PID015	1999	JMIS	The continuously growing importance of information technology (IT) requires c						
17	PID016	1994	MISQ	Identifying the best way to organize the IS functions within an enterprise has b						
18	PID017	1996	ISR	Reasons for the mixed reactions to todays electronic off-exchange trading sy						
19	PID018	1996	JMIS	The performance impacts of information technology investments in organizati						
20	PID019	1997	JMIS	Anonymity is a fundamental concept in group support systems (GSS) resear						
21	PID020	2002	ISR	Although electronic commerce (EC) has created new opportunities for busine						
22	PID021	2005	JMIS	Understanding the successful adoption of information technology is largely be						
23	PID022	2005	MISQ	Enterprise resource planning (ERP) systems and other complex information						
24	PID023	1994	JMIS	Model management systems support modelers in various phases of the mod						
25	PID024	1995	ISR	While computer training is widely recognized as an essential contributor to th						

به عنوان دومین مطالعه‌ی اکتشافی، دیلن و کراسلند با استفاده از کل مجموعه داده (مشمول بر هر سه ژورنال و هر چهار بازه‌ی زمانی) یک تحلیل خوشه‌بندی انجام دادند. خوشه‌بندی را بی‌شک می‌توان پرکاربردترین تکنیک متن‌کاوی دانست. خوشه‌بندی در این مطالعه برای شناسایی گروه‌بندی‌های طبیعی مقالات (با قرار دادن آنها در خوشه‌های مجزا) و سپس فهرست نمودن توصیفی‌ترین عباراتی که بیانگر آن خوشه‌ها بودند بکار گرفته شد. آنها از روش SVD برای کاهش ابعاد ماتریس سند-واژه و سپس اجرای الگوریتم امید ریاضی-بیشینه‌سازی برای ایجاد خوشه‌ها استفاده کردند. آنها برای تعیین تعداد بهینه‌ی خوشه‌ها چندین آزمایش انجام دادند که نهایتاً ۹ خوشه ایجاد شد. پس از ایجاد این ۹ خوشه، آنها محتوای این خوشه‌ها را از دو منظر تحلیل کردند: (۱) ارائه و نمایش نوع ژورنال (شکل 7.8a) و (۲) ارائه و نمایش زمان (شکل 7.8b).

هدف از این کار بررسی تفاوت‌های بالقوه و/یا تشابهات بین این سه ژورنال و تغییرات بالقوه با تأکید بر این خوشه‌ها بوده است یعنی به این سوال جواب بدهیم که «آیا خوشه‌هایی وجود دارند که بیانگر مضامین پژوهشی متفاوتی در یک ژورنال خاص باشند؟» و «آیا این خوشه‌ها در بازه‌های زمانی متغیر ویژگی‌های متفاوتی را از خود نشان می‌دهند؟» پژوهشگران چندین الگوی جالب را با استفاده از نمایش جدولی و گرافیکی یافته‌هایشان کشف کرده و مورد بحث و بررسی قرار داده‌اند (برای اطلاعات بیشتر مراجعه کنید به دیلن آند کراسلند ۲۰۰۸).

سوالاتی برای مورد کاربردی ۷.۵

۱. چگونه می‌توان از متن‌کاوی برای تسهیل مرور ادبیات استفاده کرد؟
۲. نتایج معمول یک پروژه‌ی متن‌کاوی بر روی مجموعه‌ی خاصی از مقالات یک ژورنال چه هستند؟ آیا می‌توانید نتایج دیگری را که در این مورد کاربردی بیان نشده است ذکر کنید؟



شکل ۷.۸ (آ) توزیع تعداد مقالات برای سه ژورنال در ۹ خوشه. (ب) ایجاد ۹ خوشه در طول سالها

Source: Used with permission of Delen, D., & M. Crossland. (2008). "Seeding the Survey and Analysis of Research Literature with Text Mining." *Expert Systems with Applications*, 34(3), pp. 1707-1720.

سوالاتی برای مرور بخش ۷.۵

۹. گام‌های اصلی در فرایند متن‌کاوی چه هستند؟
۱۰. دلیل نرمال‌سازی تواتر کلمات چیست؟ روش‌های متداول نرمال‌سازی تواتر کلمات چه هستند؟
۱۱. روش SVD چیست؟ چگونه از آن در متن‌کاوی استفاده می‌شود؟
۱۲. روش‌های اصلی استخراج دانش از پیکره متنی چه هستند؟

ما انسان‌ها موجوداتی اجتماعی هستیم. ما در بکارگیری انواع ابزارهای ارتباطی مهارت داریم. ما اغلب پیش از آنکه در مورد یک سرمایه‌گذاری تصمیم بگیریم از نشست‌های بحث و گفتگوی مالی مشورت می‌گیریم؛ در مورد رستورانی که به تازگی افتتاح شده یا فیلمی که به تازگی نمایش داده شده از دوستانمان نظر می‌خواهیم و پیش از آنکه خرید مهمی مثل خانه، ماشین یا لوازم خانگی انجام بدهیم در اینترنت جستجو کرده و نظرات کارشناسان و مصرف‌کنندگان را می‌خوانیم. ما برای تصمیم‌گیری بهتر به ویژه در زمینه‌هایی که دانش یا تجربه‌ی چندانی نداریم به نظرات دیگران اتکا می‌کنیم. به لطف محبوبیت و دسترسی‌پذیری روزافزون منابع اینترنتی که مملو از نظرات و عقاید دیگران هستند مثل رسانه‌های اجتماعی (توییتر، فیسبوک)، سایت‌های آنلاین مرور و بازنگری و وبلاگ‌های شخصی در حال حاضر بسیار آسان‌تر از هر وقت دیگری می‌توانیم نظرات افراد دیگر را در خصوص هر چیزی از آخرین گجت‌ها گرفته تا چهره‌های سیاسی و اجتماعی پیدا کنیم. اگرچه همه نظرها را در اینترنت ابراز نمی‌کنند – عمدتاً به دلیل تعداد روزافزون قابلیت‌های کانال‌های ارتباطی اجتماعی – ولی تعداد آنها با یک روند نمایی در حال رشد است.

واژه‌ی *احساسات*^۱ را به سختی می‌توان تعریف کرد و اغلب با عباراتی همچون *باور*، *دیدگاه*، *نظر* و *اعتقاد* *راسخ* هم‌معنی یا اشتباه گرفته می‌شود. احساسات بیانگر عقیده یا باور ثابت شده‌ای هستند که بیانگر احساسات یک شخص می‌باشد (مجووا ۲۰۰۹). احساسات دارای ویژگی‌های منحصر‌فردی است که آن را از سایر مفاهیمی که دوست داریم در یک متن تعیین یا شناسایی کنیم جدا می‌کند. ما اغلب به دنبال این هستیم که متن را بر اساس موضوع دسته‌بندی کنیم که می‌تواند شامل تمام طبقه‌بندی‌های موضوعات باشد. دسته‌بندی احساسات از طرف دیگر معمولاً با دو گروه یا دسته‌ی مثبت و منفی، طیفی از قطبیت‌ها (رتبه‌بندی ستاره‌ای برای فیلم‌ها) یا حتی یک طیف برای میزان قدرت عقیده سر و کار دارد (پانگ آند لی ۲۰۰۸). این دسته یا گروه‌ها موضوعات، کاربران و اسناد بسیاری را در بر می‌گیرند. اگرچه کار کردن فقط با چند دسته یا گروه خاص کار راحت‌تری نسبت به تحلیل متن استاندارد به نظر می‌رسد ولی حقیقت چیزی فراتر از این است.

تحلیل احساسات به عنوان یک رشته‌ی پژوهشی ارتباط نزدیکی با زبان‌شناسی رایانشی، پردازش زبان طبیعی و متن‌کاوی دارد. تحلیل احساسات، نام‌های بسیاری دارد و اغلب از آن با عناوین *عقیده‌کاوی*^۲، *تحلیل عینیت*^۳ و *استخراج دانش از منابع اطلاعاتی مناسب*^۴ که ارتباطی نیز با رایانش عاطفی^۵ (تشخیص و ابراز عواطف و احساسات به صورت رایانه‌ای) نام برده می‌شود. توجه و فعالیت ناگهانی به حوزه‌ی تحلیل احساسات (یعنی عقیده‌کاوی) که با استخراج خودکار عقاید، احساسات و فردیت در متن سر

¹ Sentiment

² Opinion Mining

³ Subjectivity

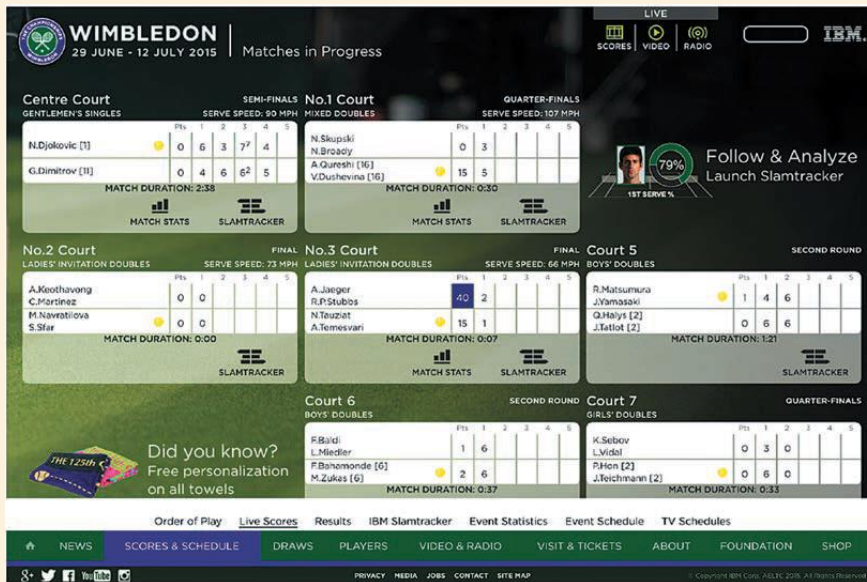
⁴ Appraisal Extraction

⁵ Affective Computing

و کار دارد، فرصت‌ها و تهدیدهایی را برای کسب‌وکارها و همچنین افراد ایجاد کرده است. آنهایی که از این فرصت‌ها بهره‌برداری می‌کنند از آن بسیار نفع خواهند برد. هر عقیده یا نظری که توسط فرد یا شرکتی بر روی اینترنت قرار داده می‌شود (خوب یا بد) به اعتبار آن شخص یا شرکت بوده و توسط دیگران بازبایی و کاوش خواهد شد (اغلب به صورت خودکار توسط برنامه‌های رایانه‌ای).

تحلیل احساسات سعی دارد تا با کاوش در عقاید و نظرات دیگران با استفاده از ابزارهای متعدد خودکار به این سوال پاسخ دهد که « مردم چه احساسی در مورد یک موضوع خاص دارند؟ » تحلیل احساسات با گردهم آوردن پژوهشگران و کارشناسان متعدد از حوزه‌های کسب‌وکار، علوم رایانه، زبانشناسی رایانشی، داده‌کاوی، روانشناسی و حتی جامعه‌شناسی سعی دارد تا تحلیل‌های متنی مبتنی بر حقایق را به مرزهای جدیدتری کشانده و سیستم‌های اطلاعاتی عقیده‌محور را محقق سازد. در یک محیط کسب‌وکار، به ویژه در بازاریابی و مدیریت ارتباط با مشتریان، تحلیل احساسات به دنبال شناسایی عقاید مطلوب و نامطلوب در خصوص محصولات و/یا خدمات با استفاده از تعداد زیادی از منابع داده‌های متنی است (بازخورد مشتری به شکل پست‌های وب، توئیتهای، وبلاگ‌ها و غیره).

احساساتی که در متون یافت می‌شود به دو شکل است: صریح و آشکار که در آن یک جمله‌ی ذهنی مستقیماً عقیده یا نظر شخص را ابراز می‌کند (« چه روز فوق‌العاده‌ای!») و ضمنی یا پنهان که در آن متن به یک عقیده یا نظر به طور ضمنی اشاره می‌کند. اکثر کارهایی که تا در حوزه‌ی تحلیل احساسات انجام شده بر نوع اول احساسات متمرکز بوده‌اند زیرا تحلیل آن بسیار آسان‌تر است. روندهای جاری در حوزه تحلیل احساسات شامل پیاده‌سازی و اجرای روش‌های تحلیلی برای احساسات صریح (آشکار) و ضمنی (پنهان) هستند. این دو قطبی بودن احساسات یکی از ویژگی‌های خاص متن است که تحلیل احساسات عمدتاً بر آن متمرکز است. این دو قطبی بودن معمولاً به دو صورت - مثبت و منفی - بوده ولی می‌تواند به صورت یک طیف نیز بیان شود. یک سند که حاوی نقد و بررسی‌های متعددی است یک حالت قطبیت آمیخته است که متفاوت از حالتی است که به هیچ عنوان قطبیت وجود نداشته باشد. گردآوری و تحلیل به موقع داده‌های متنی که از منابع مختلف به دست می‌آیند - نوشته‌های مرکز تماس مشتریان تا پست‌های رسانه‌های اجتماعی - بخش حیاتی از قابلیت‌های شرکت‌های کنشگر و مشتری‌محور امروزی به شمار می‌روند. این تحلیل‌های آنی و بلادرنگ داده‌های متنی اغلب در قالب داشبوردهایی که به آسانی قابل درک هستند مصورسازی می‌شوند. مورد کاربردی ۷.۶ یک داستان موفقیت مشتری را ارائه می‌کند که در آن مجموعه‌ای از راه‌حل‌های تحلیلی به طور جمعی برای ارتقاء تجربه‌ی تماشاچیان در مسابقات تنیس ویمبلدون استفاده شده است.



نمایش زنده‌ی امتیازها در سایت Wimbledon.com سایت رسمی مسابقات قهرمانی ویمبلدون
Copyright AELTC and IBM. Used with permission.

مسابقات قهرمانی «ویمبلدون» قدیمی‌ترین مسابقه از چهار گرند اسلم و یکی از بزرگ‌ترین رویدادهای ورزشی جهان است. این مسابقات که توسط باشگاه تنیس انگلستان سازماندهی می‌شود فعالیت خود به عنوان یک مؤسسه‌ی فرهنگی و ورزشی را از سال ۱۸۷۷ آغاز کرده است.

قهرمان قهرمانان

سازمان دهندگان مسابقات قهرمانی ویمبلدون و باشگاه تنیس انگلستان یک هدف ساده را دنبال می‌کنند: هر ساله، آنها می‌خواهند میزبان بهترین مسابقات قهرمانی تنیس در جهان از هر لحاظ باشند.

انگیزه‌ی پشت این تعهد فقط افتخارآفرینی نیست بلکه مبنای تجاری نیز دارد. برند ویمبلدون بر اساس وضعیت تراز اولش ساخته شده است یعنی هم طرفداران و هم شرکا را جذب می‌کند. بهترین و بزرگترین سازمان‌ها و شرکت‌های رسانه‌های جهان از جمله آی‌بی‌ام به دلیل شهرت و اعتبار ویمبلدون در تعالی خواهان ارتباط با آن هستند.

به همین دلیل، حذف اعتبار مسابقات قهرمانی یکی از اولویتهای بالای باشگاه تنیس انگلستان است ولی فقط دو راه وجود دارد که سازمان‌ها بتوانند این موضوع را مستقیماً کنترل کنند.

نخست، و مهمترین روش، ارائه‌ی یک تجربه‌ی منحصر بفرد برای بازیکنان، روزنامه‌نگاران و تماشاچیان است که به قدر کافی خوش شانس بوده‌اند تا مسابقات تنیس را از نزدیک تماشا کنند. باشگاه تنیس انگلستان تجربه‌ی گسترده‌ای در این حوزه دارد. از سال

۱۸۷۷، دو هفته‌ی به یادماندنی و هیجان‌انگیز از رقابت‌ها را در محیطی باصفا و دلنشین ارائه کرده است: تنیس در یک باغ روستایی انگلیسی.

دوم، حضور آنلاین است که از طریق وب سایت **Wimbledon.com**، اپلیکیشن‌های موبایل و کانال‌های رسانه‌های اجتماعی انجام می‌شود. رشد و تکامل مستمر این پلتفرم‌های دیجیتالی، نتیجه‌ی یک شراکت ۲۶ ساله بین باشگاه تنیس انگلستان و آی‌بی‌ام است. میک دزmond، مدیر رسانه و بازرگانی باشگاه تنیس انگلستان توضیح می‌دهد «زمانیکه شما مسابقات ویمبلدون را از تلویزیون تماشا می‌کنید، همه چیز را از دریچه‌ی دید گوینده‌ی رادیو و تلویزیون می‌بینید. ما هر کاری می‌کنیم تا به شرکای رسانه‌ای خود کمک کنیم تا بهترین نمایش را ارائه دهند ولی در انتهای روز، آنچه آنها از مسابقات قهرمانی پخش کرده‌اند مبنا خواهد بود.»

وی می‌افزاید، «دنایای دیجیتال، متفاوت است: این پلتفرم ماست که می‌توانیم مستقیماً با طرفداران مان صحبت کنیم - به همین خاطر ارائه‌ی بهترین تجربه‌ی ممکن به آنها ضروری و حیاتی است. هیچ رویداد یا کانال رسانه‌ی ورزشی این حق را ندارد که توجه بینندگان را طلب کند بنابراین اگر می‌خواهیم برندمان را تقویت کنیم باید کاری کنیم که مردم، تجربه‌ی دیجیتال ما را به عنوان اولین مکان برای پیگیری آنلاین مسابقات قهرمانی به رسمیت بشناسند.»

تا بدینجا، باشگاه تنیس انگلستان، جذب ۷۰ میلیون بازدید، ۲۰ میلیون دستگاه و ۸ میلیون دنبال‌کننده در رسانه‌های اجتماعی را طی ۲ هفته مسابقه در سال ۲۰۱۵ هدف‌گذاری کرده بود. آی‌بی‌ام و باشگاه تنیس انگلستان باید راهی برای تحقق این هدف پیدا می‌کردند.

ارائه‌ی یک تجربه‌ی دیجیتالی منحصر بفرد

آی‌بی‌ام و باشگاه تنیس انگلستان، پلتفرم دیجیتالی خود را با استفاده از دانش کسب شده از مخاطبان مسابقات قهرمانی به طور کامل مورد طراحی مجدد قرار دادند تا تجربه‌ای خاص و منحصر بفرد را برای جذب و حفظ طرفداران تنیس از سرتاسر جهان ارائه کنند.

به گفته‌ی الگسندرا ویلیس، رئیس بخش دیجیتال و محتوا در باشگاه تنیس انگلستان، «ما متوجه شدیم که اگرچه استفاده از پلتفرم‌های سیار به شدت در حال افزایش است ولی ۸۰٪ از بازدیدکنندگان ما از رایانه‌های رومیزی برای دسترسی به وب سایت ما استفاده می‌کنند. چالش ما برای سال ۲۰۱۵ این بود که چگونه دارایی‌های دیجیتال خودمان را برای سازگاری با دنیای فناوری‌های سیار ارتقاء بدهیم و در عین حال بهترین تجربه را برای کاربران رایانه‌های رومیزی فراهم کنیم. ما می‌خواستیم که سایت جدید ما از آن صفحه نمایش بزرگ، حداکثر بهره را برده و غنی‌ترین تجربه‌ی ممکن را از لحاظ محتوای ویدئویی و دیداری با بالاترین کیفیت در اختیار کاربران رایانه‌های رومیزی قرار دهیم و در عین حال این تجربه را برای فرمت‌های کوچکتر تبلت و موبایل نیز فراهم کنیم. دوم اینکه ما تأکید ویژه‌ای بر این داشتیم که محتوا را در بافتار ارائه کنیم - یکپارچه‌سازی مقالات با تصاویر، ویدئوهای مربوطه و ساده‌سازی گشت و گذار کاربران برای یافتن جالب‌ترین محتوا برایشان.

از منظر کاربران فناوری‌های سیار، این تیم متوجه شد که دسترسی بیشتر به ارتباطات پهنای باند 4G بدین معناست که وب سایت موبایل بیش از پیش محبوبیت خواهد یافت و دسترسی آسان‌تری را به تمامی محتواهای رسانه‌ای فراهم می‌کند. در عین حال، اپلیکیشن‌های موبایل مسابقات قهرمانی به نحوی ارتقاء یافتند که رویدادها و امتیازهای مسابقات را به صورت اعلان‌های آنی اطلاع‌رسانی می‌کند و حتی زمانیکه بازدیدکنندگان در حال عبور از ایستگاه‌ها به سمت زمین مسابقات هستند به آنها خوشامد می‌گویند. این تیم همچنین مجموعه ویژه‌ای از وب‌سایت‌ها را برای مهمترین طرفداران ورزش تنیس یعنی خود بازیکنان ایجاد کردند. این تیم با استفاده از فناوری بلومیکس آی‌بی‌ام یک برنامه وبی امن را ایجاد کردند که یک دید سفارشی و شخصی از رزرو

زمین‌های مسابقات، ترانسپورت و زمان‌های حضور در زمین مسابقه را برایشان نمایش داده و همچنین به آنها کمک می‌کرد تا با دسترسی به آمارهای هر مسابقه‌ای که در آن شرکت داشته‌اند به مرور و ارزیابی عملکردشان بپردازند.

تبدیل داده‌ها به بینش و تبدیل بینش به روایت (داستان)

این تیم به منظور تأمین بهترین محتوای ممکن برای پلتفرم‌های دیجیتالی خود از یک فرصت منحصر بفرد بهره‌برداری کرد و آن دسترسی به داده‌های بلادرنگ هر مسابقه در زمان بازی بوده است. در طی دو هفته‌ای که مسابقات ویمبلدون برگزار می‌شود، ۴۸ کارشناس خبره در کنار زمین، تقریباً ۳.۴ میلیون داده را ثبت نموده و نوع ضربات، استراتژی‌ها و نتیجه‌ی هر امتیاز را ثبت می‌کردند.

این داده‌ها به صورت آنی گردآوری و تحلیل می‌شدند و آمارهای لازم برای روزنامه‌نگاران و منتقدان تلویزیونی – و همچنین تیم ویراستاری خود پلتفرم دیجیتالی را فراهم می‌کردند.

ویلیس همچنان ادامه می‌دهد که: امسال آی‌بی‌ام مزیتی را برای ما فراهم کرد که تا پیش از این از آن برخوردار نبودیم – استفاده از فناوری جاری‌سازی داده‌ها^۱ برای ارائه‌ی بینش‌های آنی و لحظه‌ای در مورد اخبار فوری به تیم ویراستاری. این سیستم به طور خودکار جریان داده‌هایی که از کل ۱۹ زمین بازی می‌آمد را پایش کرده و زمانیکه اتفاق مهمی رخ می‌داد – مثلاً زمانیکه سَم گرات^۲ دومین سرویس سریع را در تاریخ مسابقات زده بود – بلافاصله ما را مطلع می‌کرد. در عرض چند ثانیه، می‌توانستیم آن اخبار را به مخاطبان دیجیتالی داده و آن را برای جذب ترافیک بیشتر به سمت وبسایت‌مان بر روی رسانه‌های اجتماعی به اشتراک بگذاریم.

توانایی ثبت لحظات مهم و کشف روایت‌های شگفت‌انگیز از درون داده‌ها با سرعتی بی‌مانند، کلید کار بود. اگر می‌خواستید حس و حال مسابقات قهرمانی را به صورت زنده تجربه کنید، به غیر از حضور در آنجا این بود که مسابقه را به طور زنده از wimbledon.com دنبال کنید.

کنترل و مهار قدرت زبان طبیعی

یک قابلیت جدید دیگر که در سال ۲۰۱۵ آزمایش شد، استفاده از فناوری‌های پردازش زبان طبیعی آی‌بی‌ام برای کمک به کاوش در کتابخانه‌ی عظیم تاریخچه‌ی تنیس باشگاه تنیس انگلستان برای یافتن اطلاعات بافتاری جالب بود. این تیم، مشاور واتسون آی‌بی‌ام^۳ را برای پردازش این حجم از داده‌های ساختنیافته‌ی غنی و استفاده از آن برای پاسخگویی به پرس و جوها آموزش دادند.

از همان فرانت‌اند پردازش زبان طبیعی برای اتصال به یک پایگاه داده‌ی ساختنیافته‌ی جامع از آمارهای تطبیقی که قدمت آن به اولین مسابقات قهرمانی در ۱۸۷۷ باز می‌گردد استفاده شد تا به سوالات پایه و استعمال‌های پیچیده‌تر پاسخ داده شود.

دزموند می‌گوید: «آزمون واتسون نشان داد که از توان بالقوه‌ی بسیار زیادی برخوردار است. سپس بعد، ما به عنوان بخشی از فرایند برنامه‌ریزی نوآوری سالانه‌مان می‌خواهیم استفاده گسترده‌تری از آن داشته باشیم و نهایتاً طرفداران دسترسی بیشتری به این منبع غنی از دانش ورزش تنیس داشته باشند.»

¹ Data Streaming

² Sam Groth

³ IBM Watson Engagement Advisor

حرکت به سمت فضای «آبر»

آی‌بی‌ام کل محیط دیجیتال خود را در فضای ابر هیبریدی میزبانی می‌کند. آی‌بی‌ام از تکنیک‌های مدل سازی پیچیده برای پیش بینی نقاط اوج تقاضا بر اساس برنامه زمانبندی، محبوبیت هر بازیکن، وقت از روز و بسیاری عوامل دیگر استفاده کرده است - این امکان را می‌دهد تا منابع ابری را به طور پویا به هر قسمت از محتوای دیجیتالی اختصاص دهد و تجربه ای یکپارچه را برای میلیون ها بازدید کننده در سراسر جهان تضمین کند.

علاوه بر پلتفرم ابری خصوصی قدرتمند که برای چندین سال از مسابقات قهرمانی پشتیبانی کرده است، آی‌بی‌ام از یک فضای ابری مجزا به نام سافت لیر نیز برای میزبانی مرکز فرماندهی اجتماعی ویمبلدون^۱ و فراهم کردن ظرفیت اضافی برای پشتیبانی از محیط ابری اصلی در زمان‌های اوج تقاضا استفاده کرده است.

حفظ ایمنی و امنیت ویمبلدون

امنیت آنلاین یکی از دغدغه ها و نگرانی های اصلی برای تمامی سازمان های امروزی است. شهرت و اعتبار برند به ویژه برای رویدادهای ورزشی مهم، همه چیز محسوب می شود و با توجه به اینکه دنیا در حال تماشای این رویدادها است بسیار مهم است که قربانی جرائم سایبری نشوند. به همین دلایل، امنیت نقش بسیار مهمی را در شراکت آی‌بی‌ام با باشگاه تنیس انگلستان ایفا می‌کند.

طی پنج ماهه ابتدای سال ۲۰۱۵، سیستم‌های امنیتی آی‌بی‌ام یک افزایش ۹۴ درصدی را نسبت به دوره مشابه سال ۲۰۱۴ در رویدادهای امنیتی زیرساخت‌های **wimbledon.com** شناسایی کردند. با فراگیر شدن تهدیدهای امنیتی - به ویژه حملات منع سرویس - آی‌بی‌ام به طور مستمر تمرکز خود را بر ارائه سطوح امنیتی سطح بالا برای پلتفرم کاملاً دیجیتالی باشگاه تنیس انگلستان نمود.

مجموعه کاملی از محصولات امنیتی آی‌بی‌ام از جمله IBM QRadar SIEM و IBM Preventia Intrusion Prevention به برگزاری امن و روان مسابقات قهرمانی ۲۰۱۵ و ارائه ی تجربه ی با کیفیت بالا به کاربران از طریق پلتفرم دیجیتالی کمک کردند.

تصاحب قلبها و ذهنها

موفقیت پلتفرم دیجیتالی جدید ۲۰۱۵ به پشتیبانی فناوری های ابری، تحلیلی، موبایل، اجتماعی و امنیتی آی‌بی‌ام بسیار کامل و آنی بود. اهداف تعداد بازدیدها و تعداد بازدیدکنندگان غیر تکراری نه تنها برآورده شد بلکه فراتر از آن نیز رفت. دستیابی به ۷۱ میلیون بازدید و ۵۴۲ میلیون بازدید از صفحات وب از طریق ۲۱.۱ میلیون دستگاه نشان داد که این پلتفرم در جذب مخاطبان بیشتر نسبت به قبل موفق تر بوده و آن بینندگان را در طول مسابقات قهرمانی سرگرم نگهداشته است.

ویلیس می گوید « در مجموع تعداد بازدیدهای ما ۱۳٪ و دستگاههایی که از آن دسترسی داشته‌اند ۲۳٪ نسبت به سال گذشته افزایش داشته است و رشد استفاده از سایت **wimbledon.com** از دستگاه‌های موبایل حتی بیشتر از این بوده است. به گفته‌ی دزموند: « ما شاهد رشد ۱۲۵ درصدی دستگاه‌های موبایل، رشد ۹۸ درصدی بازدیدها و رشد ۷۹ درصدی بازدید از صفحات وب بوده‌ایم. نتایج نشان می‌دهد که در سال ۲۰۱۵ ما پیروز جنگ تصاحب قلبها و ذهن‌های طرفداران بودیم. مردم ممکن است برای

¹ Wimbledon Social Command Centre

۵۰ هفته در سال به روزنامه‌ها و وبسایت‌های ورزشی محبوب خودشان رجوع کنند ولی برای این دو هفته به ما مراجعه کردند.» او ادامه می‌دهد: «این گواهی بر کیفیت تجربه‌ای است که می‌توانیم با بهره‌گیری از مزیت‌های منحصر‌فردمان به آنها ارائه دهیم و آنها را به کانال‌های رسانه‌ای خودمان نزدیک‌تر کنیم. توانایی ثبت و ارائه‌ی محتوای مرتبط به صورت بلادرنگ به طرفداران ما کمک کرد تا تجربه‌ی زنده و واضح‌تری نسبت به قبل از مسابقات قهرمانی داشته باشند.»

سوالاتی برای مورد کاربردی ۷.۶

۱. ویمبلدون چگونه از قابلیت‌های تحلیلی برای ارتقاء تجربه‌ی بینندگان خود استفاده کرد؟
۲. چالش‌ها، راه‌حل پیشنهادی و نتایج بدست آمده چه بودند؟

Source: IBM Case Study. "Creating a Unique Digital Experience to Capture the Moments That Matter.

" <http://www-03.ibm.com/software/businesscasestudies/us/en/corp?synkey=D140192K15783Q68> (accessed May 2016).

در مقایسه با روش‌های سنتی تحلیل احساسات که مبتنی بر نظرسنجی یا گروه‌های تمرکز، پرهزینه و زمان‌بر بودند (به همین خاطر از یک نمونه کوچک از شرکت کنندگان استفاده می‌شد)، چهره‌ی جدید تحلیل احساسات مبتنی بر تحلیل‌های متنی یک خط‌شکن است. راه‌حل‌های جاری به خودکارسازی روش‌های گردآوری؛ فیلترینگ، دسته‌بندی و خوشه‌بندی داده‌ها در مقیاس زیاد از طریق فناوری‌های پردازش زبان طبیعی و داده‌کاوی پرداخته‌اند که شامل اطلاعات واقعی و ذهنی می‌شود. تحلیل احساسات را شاید بتوان محبوب‌ترین رویکرد تحلیل متن دانست که از منابع داده متعدد مثل توییت‌ها، پست‌های فیسبوک، جوامع آنلاین، بوردهای بحث و گفتگو، وبلاگ‌ها، نقد و بررسی محصول، سوابق و فایل‌های صوتی ضبط شده‌ی مراکز تماس، سایت‌های رتبه‌بندی محصول، اتاق‌های گپ و گفتگو، پورتال‌های مقایسه‌ی قیمت، سوابق موتورهای جستجو و گروه‌های خبری استفاده می‌کنند. در ادامه برخی از کاربردهای تحلیل احساسات را برای به تصویر کشیدن قدرت و پوشش‌دهی گسترده‌ی این فناوری ارائه می‌کنیم.

صدای مشتری صدای مشتری^۱ جزء لاینفک سیستم‌های مدیریت تجربه‌ی مشتری و مدیریت ارتباط با مشتریان است. تحلیل احساسات به عنوان یک توانمندساز صدای مشتری می‌تواند برای درک بهتر و مدیریت بهتر شکایات و تعریف و تمجیدهای مشتریان به نقد و بررسی آنها در مورد محصولات و خدمات شرکت کمک کند. به عنوان مثال، یک شرکت بازاریابی/تبلیغات فیلم‌ها می‌تواند احساسات و عقاید منفی در مورد یک فیلم را به محض نمایش در سالن‌ها تشخیص داده (بر اساس پیش‌پرده) و سریعاً استراتژی تبلیغات و ترکیب پیش‌پرده‌ها را برای پیشگیری از تأثیر منفی بر فیلم تغییر دهند. به همین منوال، یک شرکت نرم‌افزاری می‌تواند شایعات منفی در خصوص ایرادات محصول جدیدشان را شناسایی کرده و بسته‌های اصلاحی را برای رفع سریع مشکل منتشر کند.

اغلب، تمرکز صدای مشتری بر تک‌تک مشتریان، نیازهای خدماتی و پشتیبانی، خواسته‌ها و مسائل آنهاست. صدای مشتری، داده‌ها را از مجموعه کاملی از نقاط تماس با مشتریان از جمله نامه‌های الکترونیکی، نظرسنجی‌ها، یادداشت‌ها/صداها ضبط شده مرکز تماس و پست‌های رسانه‌های اجتماعی گرفته و صداهای مشتریان را با تراکنش‌ها (استعلام‌ها، خریدها و مرجوعی‌ها) و

¹ Voice of Customer (VOC)

پروفایل‌های تک تک مشتریان که در سیستم‌های عملیاتی سازمانی ذخیره شده‌اند تطبیق می‌دهد. صدای مشتری که عامل محرک آن اکثراً تحلیل احساسات است یک عنصر کلیدی در راهکارهای مدیریت تجربه مشتری است که هدف از آن ایجاد یک رابطه‌ی صمیمی و نزدیک با مشتری است.

صدای بازار صدای بازار^۱ عبارتست از درک و فهم عقاید و روندهای تجمیعی و دانستن این موضوع که ذینفعان – مشتریان، مشتریان بالقوه، اینفلوئنسرها در مورد محصولات و خدمات شما چه می‌گویند. یک تحلیل صدای بازار که به خوبی انجام شده باشد به شرکت‌ها در دستیابی به اطلاعات رقابتی، توسعه و موقعیت‌یابی محصول کمک کند.

صدای کارکنان صدای کارکنان^۲ تا پیش از این محدود به نظرسنجی رضایت کارکنان بود. تحلیل متن به طور کلی (و تحلیل احساسات به طور خاص) یک عامل توانمندساز عظیم برای ارزیابی صدای کارکنان است. استفاده از داده‌های متنی غنی، روشی کارا و موثر را برای گوش دادن به حرف‌های کارکنان فراهم می‌کند. همانطور که می‌دانیم، کارکنان راضی و خوشنود به ارتقاء تجربه‌ی مشتری و بهبود رضایت آنها کمک می‌کنند.

مدیریت برند مدیریت برند^۳ بر گوش دادن به رسانه‌های اجتماعی تمرکز دارد که هر کسی (مشتریان گذشته، جاری، آینده، کارشناسان و خبرگان صنعت و سایرین) می‌توانند عقاید و نظریاتشان را که می‌تواند به شهرت و اعتبار سازمان آسیب زده یا موجب ارتقاء آن شود منتشر کنند. چندین شرکت نوپا که به تازگی راه اندازی شده اند خدمات مدیریت برند تحلیل محور را به دیگران عرضه می‌کنند. مدیریت برند بر محصول و شرکت متمرکز است تا بر مشتری و تلاش می‌کند تا به جای مدیریت تجارب با استفاده از تکنیک‌های تحلیل احساسات به ادراک و برداشت مشتریان شکل دهد.

بازارهای مالی پیش بینی ارزش آینده‌ی بازارهای سهام به یک مسئله‌ی جالب و در عین حال لاینحل تبدیل شده است. آنچه ارزش یک سهام (یا یک گروه از سهام‌ها) را بالا و پایین می‌برد هر چیزی می‌تواند باشد به غیر از یک علم دقیق. بسیاری بر این باورند که بازار سهام عمدتاً مبتنی بر احساسات است و موجب شده تا هیچگونه رفتار منطقی و عقلانی نداشته باشد (به ویژه برای تحركات کوتاه مدت بازار سهام). از این رو، استفاده از تحلیل احساسات در بازارهای مالی محبوبیت چشمگیری یافته است. تحلیل خودکار احساسات بازار با استفاده از رسانه‌های اجتماعی، اخبار، وبلاگ‌ها و گروه‌های بحث و گفتگو، روش مناسبی برای محاسبه‌ی تحركات بازار به نظر می‌رسد. تحلیل احساسات در صورتیکه به درستی انجام شود می‌تواند تحركات کوتاه مدت بازار سهام را بر اساس شایعات بازار که ممکن است بر نقدینگی و تجارت تأثیر بگذارند شناسایی کند.

سیاست همانطور که می‌دانیم، عقاید و نظرات در سیاست اهمیت زیادی دارند. از آنجائیکه جلسات بحث و گفتگوی سیاسی با نقل قول‌ها، کنایه‌ها و ارجاعات پیچیده به اشخاص، سازمان‌ها و ایده‌ها همراه هستند سیاست یکی از دشوارترین و در عین حال پربارترین حوزه‌ها برای تحلیل احساسات است. با تحلیل احساسات نشست‌های انتخاباتی می‌توان پیش‌بینی کرد که چه کسی پیروز خواهد شد یا شکست خواهد خورد. تحلیل احساسات می‌تواند کمک کنند که درک کنیم رأی‌دهندگان به چه چیزی فکر

¹ Voice of Market

² Voice of Employee

³ Brand Management

می‌کنند و دیدگاه یک نامزد انتخاباتی را در خصوص مسائل درک کنیم. تحلیل احساسات می‌تواند به سازمان‌های سیاسی، کمپین‌ها و تحلیلگران خبری برای درک بهتر مسائل و موضوعات مهم برای رأی دهندگان کمک کند. هر دو حزب در کمپین‌های انتخابات ریاست جمهوری ایالات متحده در سال‌های ۲۰۰۸ و ۲۰۱۲ با موفقیت از فناوری استفاده کردند.

اطلاعات دولتی اطلاعات دولت یکی دیگر از کاربردهای تحلیل احساسات توسط نهادهای اطلاعاتی است. به عنوان مثال، می‌توان منابع را برای افزایش ارتباطات منفی یا خصمانه پایش نمود. تحلیل احساسات می‌تواند امکان تحلیل خودکار عقاید و نظراتی که مردم در مورد سیاست‌های بلا تکلیف یا پروپوزال‌های قانون‌گذاری دولت بیان می‌کنند را فراهم کند. علاوه بر این، نهادهایی از قبیل وزارت امنیت می‌توانند از پایش ارتباطات برای کشف احساسات منفی بهره بگیرند.

سایر حوزه‌های کاربردی جالب از احساسات مشتریان می‌توان برای طراحی بهتر سایت‌های تجارت الکترونیک (توصیه‌ی محصول، تبلیغات)، جایگیری بهتر تبلیغات (مثلاً جای‌گذاری تبلیغات به صورت پویا برای محصولات و خدمات) و مدیریت موتورهای جستجوی عقیده‌محور (یعنی یک وب‌سایت تجمیع عقاید، یک جایگزین برای سایت‌های مشابه اپینون^۱ که نظرات کاربران را خلاصه و تجمیع می‌کند). تحلیل احساسات می‌تواند با دسته بندی و اولویت بندی نامه‌های الکترونیکی دریافتی به فیلتراسیون نامه‌های الکترونیکی کمک کند (نامه‌های الکترونیکی بسیار منفی را شناسایی کرده و به یک پوشه‌ی مناسب ارسال کند) و تحلیل استنادی می‌تواند تعیین کند که آیا یک نویسنده از نوشته یا متن استنادی به عنوان شواهد پشتیبان استفاده کرده است یا خیر.

۲.۱.۱ فرایند تحلیل احساسات

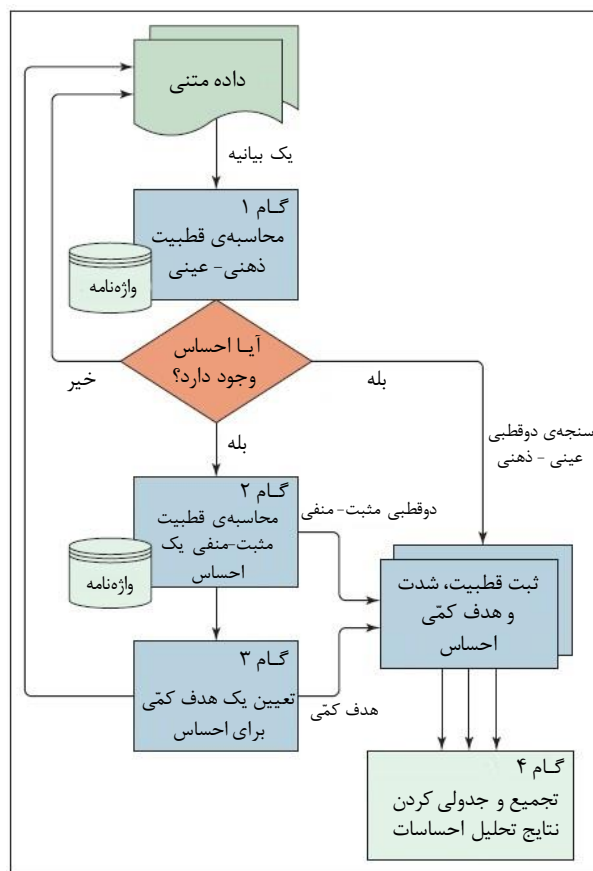
به دلیل پیچیدگی مسئله (مفاهیم بنیادی، بافتاری که متن در آن بیان شده و ...)، هیچ فرایند استاندارد آماده‌ای در حال حاضر برای تحلیل احساسات وجود ندارد. با وجود این، بر اساس کارهای منتشر شده در رشته‌ی تحلیل حساسیت (هم بر روی روش‌های پژوهشی و هم طیف کاربردها)، یک فرایند منطقی چندمرحله‌ای ساده در شکل ۷.۹ نشان داده شده است که به نظر می‌رسد متدولوژی مناسبی برای تحلیل احساسات باشد. این گام‌های منطقی تکرارپذیر بوده (یعنی بازخورد، تصحیح و تکرار بخشی از فرایند کشف هستند) و دارای ماهیت آزمایشی و تجربی هستند و زمانیکه تکمیل و ترکیب شوند قادر هستند تا بینش‌های مطلوبی را در مورد عقاید و نظرات مطرح شده در متن مورد نظر تولید کنند.

گام ۱ - تشخیص احساسات اولین وظیفه‌ی اصلی در تحلیل حساسیت پس از بازیابی و آماده‌سازی اسناد متنی، تشخیص یا شناسایی عینیت یا واقعیت است. هدف در اینجا این است که یک واقعیت و یک عقیده را از یکدیگر متمایز کنیم که می‌توان اینکار را دسته بندی یک متن در گروه عینی یا ذهنی دانست. اینکار را می‌توان به صورت محاسبه‌ی عینیت-ذهنیت نیز دانست. اگر مقدار عینیت نزدیک به ۱ باشد آنگاه هیچ عقیده‌ای وجود ندارد که بخواهیم به آن رسیدگی کنیم (یعنی یک واقعیت است) از این رو فرایند تکرار شده و داده‌های متنی بعدی برای تحلیل برداشته می‌شود. معمولاً شناسایی یا تشخیص عقیده بر اساس ارزیابی صفت‌ها در متن انجام می‌شود. به عنوان مثال، قطبیت «چه کاری فوق‌العاده‌ای!» را به آسانی می‌توان با نگاه به یک صفت تعیین کرد.

¹ Epinions

گام ۲ - دسته‌بندی دوقطبی (مثبت یا منفی) دومین وظیفه‌ای اصلی در اینجا تعیین قطبیت برای دسته‌بندی است. هدف در اینجا این است که عقیده یا نظری که در یک متن شناسایی شده است را در یکی از دو گروه مثبت یا منفی قرار دهیم تا موقعیت آن را بر روی طیف بین این دو قطب تعیین کنیم (پانگ آند لی ۲۰۰۸). زمانیکه این دسته‌بندی را از دیدگاه دودویی نگاه کنیم، دسته‌بندی دوقطبی به شکل یک دسته‌بندی دودویی خواهد بود که متن یا سند مورد نظر را به صورت کلی یک عقیده مثبت یا عقیده منفی در نظر می‌گیرد. علاوه بر شناسایی دو قطبی مثبت-منفی می‌بایست شدت احساسات (به جای صرفاً مثبت می‌توان آن را به صورت خفیف، متوسط، قوی یا بسیار قوی نیز بیان کرد). اکثر این پژوهش‌ها بر روی نقد و بررسی محصولات یا فیلم‌ها انجام شده که تعاریف «مثبت» و «منفی» کاملاً واضح و روشن هستند. سایر وظایف از قبیل دسته‌بندی اخبار در قالب «خوب» یا «بد» مشکلاتی را به همراه دارند. به عنوان مثال، یک مقاله ممکن است حاوی اخبار منفی باشد بدون اینکه در آن صراحتاً از عبارات یا کلمات ذهنی استفاده شده باشد. علاوه بر این، این گروه یا دسته‌ها معمولاً زمانیکه سند و متن مورد نظر هر دو احساس مثبت و منفی را ابراز کند در هم آمیخته می‌شوند آنگاه باید به دنبال شناسایی احساس اصلی یا غالب در سند باشیم.

هنوز برای متن‌های طولانی، دسته‌بندی ممکن است در چندین سطح انجام شود: عبارت، جمله و شاید سطح سند. برای آنها، معمولاً از خروجی یک سطح به عنوان ورودی لایه سطح بعدی استفاده می‌شود. در بخش بعدی به توضیح چندین روش بکار رفته برای تشخیص قطبیت اسناد می‌پردازیم.



شکل ۷.۹ فرایند چندمرحله‌ای تحلیل احساسات

گام ۳ - شناسایی هدف کمی هدف از این گام این است که هدف از احساسات بیان شده (مثلاً یک شخص، یک محصول، یک رویداد) را به دقت شناسایی کنیم. دشواری این کار به میزان زیادی به دامن‌های تحلیل بستگی دارد. اگرچه شناسایی دقیق هدف برای نقد و نظر در خصوص محصول یا فیلم معمولاً آسان است زیرا نقد و بررسی، مستقیماً به هدف متصل است ولی در سایر حوزه‌ها می‌تواند چالش‌برانگیز باشد. برای مثال، متون طولانی و عمومی مثل صفحات وب، مقالات خبری و وبلاگ‌ها همیشه یک عنوان از پیش تعیین شده ندارند و اغلب به موضوعات متعددی اشاره دارند که هر یک از آنها را می‌توان به عنوان یک هدف در نظر گرفت. برخی اوقات حتی بیش از یک هدف در یک جمله احساسی وجود دارد که در خصوص متون تفضیلی به این شکل است. در یک جمله تفضیلی ذهنی، موضوعات به ترتیب اولویت چیده می‌شوند. به عنوان مثال: «این رایانه‌ی لپ‌تاپ بهتر از رایانه‌ی رومیزی من است.» این جملات را می‌توان با استفاده از صفت‌ها و قیده‌های تفضیلی (بیشتر، کمتر، بهتر، طولانی‌تر)، صفت‌های برترین (بیشترین، کمترین، بهترین) و سایر کلمات مشخص کرد. به محض بازیابی جملات، موضوعات را می‌توان به بهترین ترتیبی که در متن توضیح داده شده مرتب نمود.

گام ۴ - گردآوری و تجمیع پس از شناسایی و محاسبه‌ی تمامی نقاط داده‌ی متنی در سند در این مرحله تمامی آنها تجمیع شده و به یک سنج‌ی احساسات واحد برای کل سند تبدیل می‌شوند. این تجمیع می‌تواند به سادگی یک جمع زدن قطبیت‌ها و نقاط قوت تمامی متون یا به پیچیدگی استفاده از تکنیک‌های تجمیع معنایی از پردازش زبان طبیعی برای شناسایی احساسات نهایی باشد.

همانطور که در بخش قبل گفته شد، شناسایی و تعیین قطبیت می‌تواند در سطح کلمه، عبارت، جمله یا کل سند باشد. جزءترین سطح قطبیت در سطح کلمه است که پس از شناسایی و تعیین قطبیت در سطح کلمه، می‌توان آن را تجمیع نموده و به سطح بالاتری دست پیدا کرد و سپس اینکار را آنقدر انجام می‌دهیم تا به سطح مطلوبی از تحلیل احساسات می‌رسید. دو تکنیک اصلی برای شناسایی و تعیین قطبیت در سطح کلمه/عبارت وجود دارد که هر کدام مزایا و معایب خاص خودشان را دارند:

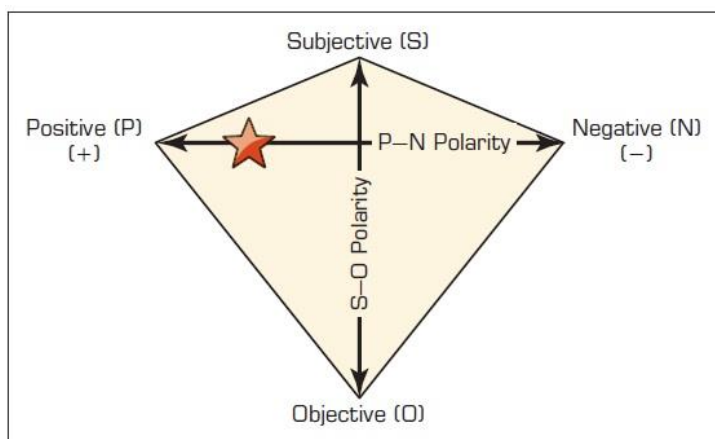
۱. استفاده از یک واژه‌نامه به عنوان یک کتابخانه‌ی مرجع (که به صورت دستی یا خودکار توسط فردی برای یک وظیفه‌ای خاص ایجاد شده یا مؤسسه‌ای برای استفاده عمومی آن را توسعه داده باشد).
۲. استفاده از مجموعه‌ای از اسناد آموزشی به عنوان منبع دانش در خصوص قطبیت عبارات در یک حوزه‌های خاص (از جمله مدل‌های پیشگویانه از اسناد متنی).

یک *واژه‌نامه*^۱ در واقع یک کاتالوگ از کلمات، مترادف‌ها و معانی آنها در یک زبان خاص است. علاوه بر واژه‌نامه‌ها در بسیاری از زبان‌های دیگر، واژه‌نامه‌های عمومی برای زبان انگلیسی تهیه شده است. واژه‌نامه‌های عمومی اغلب برای تهیه‌ی واژه‌نامه‌های تخصصی در پروژه‌های تحلیل احساسات بکار می‌روند. شاید محبوب‌ترین واژه‌نامه‌ی عمومی، *وردنت*^۲ باشد که در دانشگاه پرینستون توسعه یافته و بسیاری از پژوهشگران و کارشناسان برای تحلیل احساسات از آن استفاده کرده و آن را توسعه داده‌اند. همانطور

¹ Lexicon

² WordNet

که در وب سایت وردنت توضیح داده شده است (wordnet.princeton.edu) این یک پایگاه داده‌ی واژه‌نامه‌ای بزرگ زبان انگلیسی است که شامل اسامی، افعال، صفات و قیده‌ها است که در قالب مجموعه مترادف‌های شناختی دسته‌بندی شده‌اند و هر یک مفهوم متمایزی را بیان می‌کنند. این مجموعه مترادف‌ها از طریق روابط واژه‌ای مفهومی-معنایی به یکدیگر مرتبط هستند. یک نسخه‌ی جالب از وردنت توسط اسولی آند سیاستیانی (۲۰۰۶) ایجاد شد که آنها برچسب‌های قطبیت (منفی- مثبت) و عینیت (عینی- ذهنی) برای هر عبارت در واژه‌نامه افزودند. برای برچسب‌گذاری هر عبارت، آنها گروه مترادف‌ها را با استفاده از مجموعه‌ای از دسته‌بندی‌های سه‌گانه دسته‌بندی کردند (سنجه‌ای که دقیقاً یکی از سه برچسب را به هر شیء یا موضوع متصل می‌کند) و هرکدام قادر هستند تصمیم بگیرند که مجموعه مترادف‌ها آیا مثبت، منفی یا عینی است یا خیر. امتیازهای بدست آمده بین ۰ تا ۱ است که یک ارزیابی از ویژگی‌های احساسی یا عقیده‌ای از عبارات را به ما می‌دهد. شکل ۷.۱۰ نمایش دیداری کل این فرایند است. اضلاع مثلث بیانگر یکی از سه دسته‌بندی‌ها می‌باشد (مثبت، منفی، عینی). یک عبارت را می‌توان در این فضا به صورت یک نقطه پیدا کرد که بیانگر این است که تا چه اندازه به هر یک از این دسته‌بندی‌ها تعلق دارد.



شکل ۷.۱۰ نمایش گرافیکی ارتباطات قطبیت مثبت- منفی و عینی- ذهنی

یک متدولوژی انشعابی مشابه دیگر برای ایجاد سنتی‌وردنت^۱ استفاده شده که یک واژه‌نامه عمومی است به طور خاص برای اهداف عقیده‌کاوی (تحلیل احساسات) توسعه یافته است. سنتی‌وردنت سه امتیاز احساسی در وردنت را به هر مجموعه مترادف تخصیص می‌دهد: مثبت بودن، منفی بودن و عینیت. برای اطلاعات بیشتر در مورد سنتی‌وردنت به sentiwordnet.isti.cnr.it مراجعه کنید.

یک انشعاب دیگر از وردنت، وردنت افکت^۲ نام دارد که توسط استراپاراوا آند والیتوتی^۳ (۲۰۰۴) توسعه یافته است. آنها مجموعه مترادف‌های وردنت را با استفاده از برچسب‌های مؤثری که بیانگر گروه یا دسته‌های احساسی متفاوتی هستند برچسب‌گذاری می‌گذارند (عاطفه، حالت شناختی، نگرش و احساسات). وردنت همچنین مستقیماً در تحلیل احساسات بکار رفته است. به عنوان مثال، کیم آند هووی^۴ (۲۰۰۴) و لیو، هو و چنگ (۲۰۰۵) واژه‌نامه‌هایی از عبارات مثبت و منفی را تهیه کردند و برای اینکار از

¹ SentiWorNet

² WordNet-Affect

³ Strapparava & Valitutti

⁴ Kim & Hovy

فهرست کوچکی از عبارات دوقطبی شناخته شده (مثلاً عشق، دوست داشتن، ...) شروع کرده و با استفاده از ویژگی‌های مترادف و متضادی عبارات آنها را در هر دسته یا گروه دوقطبی قرار می‌دهد.

دسته‌بندی احساسات را می‌توان با استفاده از ابزارهای تحلیل آماری و یادگیری ماشینی که از منابع گسترده‌ای از اسناد برچسب‌گذاری شده‌ی موجود بهره می‌برند انجام داد. وبسایت‌های نقد و بررسی محصولات از قبیل آمازون، سی‌نت^۱، ای‌بی، راتن تومیتوز^۲ و پایگاه داده‌های اینترنتی فیلم‌ها^۳ همگی به طور گسترده‌ای به عنوان منابع داده‌های حاشیه‌نویسی^۴ شده مورد استفاده قرار گرفته‌اند. سیستم ستاره‌ای یک برچسب صریح برای نقد و بررسی‌های قطبی ارائه کرده و اغلب به عنوان یک استاندارد طلایی در ارزیابی الگوریتم‌ها مورد استفاده قرار می‌گیرد.

انواع گوناگونی از داده‌های متنی برچسب‌گذاری شده به صورت دستی از طریق ارزیابی‌هایی از قبیل کنفرانس بازیابی متن^۵، مجموعه تستی انستیتو ملی انفورماتیک^۶ برای سیستم‌های آی آر و نشست ارزشیابی چندزبانی^۷ در دسترس قرار گرفته است. مجموعه داده‌هایی که از این طریق تولید می‌شوند اغلب به عنوان یک استاندارد در جوامع متن‌کاوی از جمله پژوهشگران تحلیل احساسات بکار می‌روند. پژوهشگران و گروه‌های پژوهشی نیز مجموعه داده‌های بسیار جالبی را تولید کرده‌اند. بینش فناوری^{۷.۲} برخی از متداول‌ترین آنها را فهرست کرده است. به محض دستیابی به یک مجموعه داده‌های متنی که قبلاً برچسب‌گذاری شده است از انواع الگوریتم‌های مدلسازی پیشگویانه و یادگیری ماشینی می‌توان برای آموزش دسته‌بندی احساسات استفاده کرد. برخی از متداول‌ترین الگوریتم‌های بکار رفته برای این منظور شامل شبکه‌های عصبی مصنوعی، ماشین‌های برداری پشتیبان، همسایگی نزدیک‌ترین کی، بیز ساده، درخت تصمیم و خوشه‌بندی امید ریاضی - بیشینه کردن هستند.

بینش فناوری ۷.۲ مجموعه‌های بزرگ داده‌های متنی برای متن کاوی پیشگویانه و تحلیل احساسات

در ادامه فهرستی از برخی از پرکاربردترین مثال‌ها در مجموعه‌های بزرگ داده‌های متنی ارائه شده است:

- **نسخه‌های نوشتاری مناظرات کنگره:** این مقاله که توسط توماس، پانگ و لی (۲۰۰۶) منتشر شده است حاوی برخی از سخنرانی‌های سیاسی است که برای اینکه نشان دهند آیا سخنران از یک قانون‌گذاری حمایت کرده یا مخالف آن بوده برچسب‌گذاری شده‌اند.
- **اکنونا پیننگ:** توسط دانشکده‌ی استرن در دانشگاه نیویورک منتشر شده و حاوی بازخوردهای تجار و بازرگانان در وب سایت آمازون دات کام می‌باشد.

¹ C-Net

² Rotten Tomatoes

³ IMDB

⁴ Annotated data

⁵ Text Retrieval Conference

⁶ NII Test Collection

⁷ Cross Language Evaluation Forum

- **مجموعه داده‌های نقد و بررسی فیلم کورنل!** توسط پانگ آند لی (۲۰۰۸) معرفی شده که حاوی ۱۰,۰۰۰ برچسب مثبت و ۱,۰۰۰ برچسب منفی است که به طور خودکار در سطح سند استخراج شده‌اند و حاوی ۵,۳۳۱ جمله‌ی مثبت و ۵,۳۳۱ جمله‌ی منفی است.
- **مجموعه داده‌های نقد و بررسی فیلم استانفورد:** مجموعه‌ای از ۲۵,۰۰۰ نقد و بررسی قطبی فیلم‌ها برای آموزش و ۲۵,۰۰۰ نقد و بررسی برای تست. همچنین داده‌های فاقد برچسب بیشتری برای استفاده وجود دارد. متون اولیه و فرمت‌های کلمات پردازش شده نیز در سایت <http://ai.stanford.edu/amaas/data/sentiment> موجود است.
- **پیکره‌ی ام‌پی‌کی‌وای!** سیستم تشخیص عقیده و پیکره متنی حاوی ۵۳۵ مقاله‌ی خبری حاشیه‌نویسی شده شبه صورت دستی است که از منابع خبری گوناگون که حاوی برچسب‌ها برای عقاید، احساسات و حالت‌های خصوصی هستند (باورها، عواطف و غیره) می‌باشد.
- **نقد و بررسی چندوجهی رستوران‌ها:** توسط سیندر آند بارزیلای^۳ (۲۰۰۷) معرفی شده که حاوی ۴,۴۸۸ نقد و بررسی و رتبه‌بندی صریح بین ۱ تا ۵ برای پنج جنبه‌ی مختلف است: غذا، فضا، سرویس، ارزش و تجربه‌ی کلی.

پس از تعیین جهت‌گیری معنایی تک تک کلمات، اغلب بهتر است اینکار را به سطح جمله یا عبارتی که این کلمه در آن وجود دارد بسط بدهیم. ساده‌ترین روش برای اینکار این است که از یک نوع میانگین‌گیری قطبیت کلمات در جملات یا عبارات استفاده کنیم. اگرچه این روش به ندرت مورد استفاده قرار می‌گیرد ولی بکارگیری یک یا چند تکنیک یادگیری ماشینی و برقراری یک رابطه‌ی پیشگویانه بین کلمات (و مقادیر قطبی آنها) و عبارات یا جملات می‌تواند این تکنیک را پیچیده‌تر کند.

اگرچه اکثریت کارهایی که در این حوزه انجام شده‌اند به منظور تعیین جهت‌گیری معنایی کلمات و/یا عبارات و جملات بوده است ولی کارهایی از قبیل خلاصه‌سازی و بازیابی اطلاعات نیازمند این است که ما برچسب‌گذاری معنایی را برای کل سند انجام دهیم (راماج و همکاران ۲۰۰۹). مشابه تجمیع مقادیر قطبی احساسات از سطح کلمه تا سطح عبارت یا جمله در اینجا نیز برای تجمیع مقادیر قطبی تا سطح سند از یک روش میانگین‌گیری استفاده می‌شود. جهت‌گیری احساسات در یک سند ممکن است در خصوص اسناد بسیار بزرگ چندان معنا نداشته باشد از این رو اغلب از این روش برای اسناد کوچک یا متوسط که در اینترنت منتشر می‌شوند استفاده می‌شود.

سوالاتی برای مرور بخش ۷.۶

۶. تحلیل احساسات چیست؟ چه ارتباطی با متن کاوی دارد؟
۷. متداول‌ترین کاربردهای تحلیل احساسات چه هستند؟ چرا؟
۸. مزایا و منافع مورد انتظار از تحلیل احساسات در سیاست چه هستند؟
۹. گام‌های اصلی در اجرای پروژه‌های تحلیل احساسات چه هستند؟

¹ Cornell Movie-Review Data Set

² MPQA Corpus

³ Synder & Barzilay

۱۰. دو روش متداول برای شناسایی و تعیین قطبیت چه هستند؟ توضیح دهید.

اینترنت موجب شده تا چشم‌انداز کسب‌وکارها برای همیشه تغییر کند. با توجه به گسترش رقابت و تبدیل دنیای ما به دنیایی کاملاً متصل و تخت، شرکت‌های امروزی با فرصت‌ها (توانایی دسترسی به مشتریان و بازارهایی که تا پیش از این امکان‌پذیر نبود) و تهدیدهای روزافزونی (یک بازار رقابتی در سطح جهانی و دائماً در حال تغییر) مواجه هستند. شرکت‌هایی که برای مواجهه با چنین محیط ناپایداری برای خودشان چشم‌انداز و قابلیت‌هایی را در نظر گرفته‌اند بسیار از آن بهره می‌برند در حالیکه آنهایی در برابر آن مقاومت می‌کنند برای بقاء خود دچار مشکل شده‌اند. حضور مستمر و فعال در اینترنت امروزه یک حق انتخاب به شمار نمی‌رود بلکه یک الزام کسب‌وکار است. مشتریان از شرکت‌ها انتظار دارند تا محصولات و/یا خدمات خود را از طریق اینترنت عرضه کنند. مشتریان نه تنها محصولات و خدمات را از اینترنت خریداری می‌کنند بلکه در مورد شرکت‌ها و به اشتراک‌گذاری تجاربشان با دیگران نیز از طریق اینترنت صحبت می‌کنند.

رشد اینترنت و فناوری‌های آن موجب شده تا ایجاد، جمع‌آوری و تبادل داده‌ها/اطلاعات/عقاید راحت‌تر شود. تأخیر در ارائه خدمت، تولید، ارسال، تحویل و استعلام مشتریان دیگر به هیچ عنوان پذیرفته نیستند. در حال حاضر، به لطف ابزارهای رسانه‌های اجتماعی و فناوری‌های موجود در اینترنت، هرکس از همه چیز اطلاع دارد. شرکت‌های موفق آنهایی هستند که با آغوش باز پذیرای این فناوری‌های اینترنتی باشند و از آنها برای بهبود فرایند کسب‌وکارشان به منظور برقراری ارتباط بهتر با مشتریان، درک خواسته‌ها و نیازهای آنها و خدمت‌رسانی تمام و کمال بهره ببرند. با توجه به اینکه کسب‌وکار امروزه در عصر اینترنت و رسانه‌های اجتماعی هستند تمرکز بر مشتریان و شاد و راضی نگهداشتن آنها هیچگاه تا این اندازه اهمیت نداشته است.

شبکه گسترده جهانی (یا به اختصار وب) به عنوان یک مخزن عظیم از داده‌ها و اطلاعات در مورد هر چیزی که به ذهن انسان برسد عمل می‌کند. وب را شاید بتوان بزرگ‌ترین مخزن داده و متن در جهان دانست و میزان اطلاعاتی که در وب موجود است به سرعت در حال رشد است. اطلاعات جالب بسیاری را می‌توان به صورت آنلاین پیدا کرد: صفحه اصلی آنها به کدام صفحات وصل است، چه تعداد افراد به یک صفحه‌ی وب خاص لینک زده‌اند و سازماندهی یک سایت خاص به چه شکل است؟ علاوه بر این، هر بازدید کننده از یک سایت وب، هر جستجو در یک موتور جستجو، هر کلیک بر روی یک پیوند و هر تراکنش در یک سایت تجارت الکترونیک می‌تواند داده‌های بیشتری را تولید کند. اگرچه داده‌های متنی ساخت‌نیافته به شکل صفحات وب که به زبان‌های اچ‌تی‌ام‌ال یا اکس‌ام‌ال کدگذاری شده‌اند محتوای غالب وب را تشکیل می‌دهند ولی زیرساخت‌های وبی حاوی اطلاعات هایپرلینک (اتصال به سایر صفحات وب) و اطلاعات کاربرد (سوابق تعاملات بازدیدکنندگان با وب‌سایت) هستند که تمامی آنها داده‌های غنی را برای کشف دانش در اختیار ما قرار می‌دهند. تحلیل این اطلاعات می‌تواند به ما کمک کند تا از وب سایت‌ها بهتر استفاده کنیم و به ما در ارتقاء و بهبود روابط و ارزشی که وب سایت‌ها به بازدیدکنندگان عرضه می‌کنند کمک می‌کند.

کاوش در وب به دلیل حجم و پیچیدگی بسیار بالا به هیچ عنوان کار آسانی نیست. وب چالش‌هایی را نیز برای کشف دانش به نحوی کارا و اثربخش به همراه دارد (هان و کامبر ۲۰۰۶):

- **وب برای داده‌کاوی مؤثر بیش از اندازه بزرگ است:** وب به اندازه‌ای بزرگ است و به اندازه‌ای سریع رشد می‌کند که حتی تعیین حجم آن کاری دشواری است. به دلیل اندازه و حجم بسیار بالای وب، ایجاد و راه‌اندازی یک انبار داده برای تکثیر، ذخیره‌سازی و یکپارچه‌سازی تمامی داده‌های موجود در وب امکان‌پذیر نبوده و گردآوری و جمع‌آوری داده‌ها را به یک چالش بدل کرده است.
- **وب بیش از اندازه پیچیده است:** پیچیدگی یک صفحه‌ی وب بسیار بیشتر از یک صفحه از یک سند متنی معمولی است. صفحات وب یک ساختار متداول شکل و یکسان ندارند. تنوع و گوناگونی سبک نوشتاری و محتوا در آنها از هر کتاب، مقاله یا سایر اسناد متنی دیگر بیشتر است.
- **وب بیش از اندازه پویا است:** وب یک منبع اطلاعاتی بسیار پویا می‌باشد و نه فقط به سرعت در حال رشد و گسترش است بلکه محتوای آن نیز دائماً به‌روزآوری می‌شود. وبلاگ‌ها، داستان‌های خبری، نتایج بازار سهام و بورس، گزارشات هواشناسی، نتایج مسابقات ورزشی، قیمت‌ها، تبلیغات شرکت‌ها و انواع دیگر اطلاعات به‌طور منظم بر روی وب به‌روزآوری می‌شوند.
- **وب به دامنه‌ی مشخصی تعلق ندارد.** وب میلیاردها ایستگاه کاری و جوامع بسیار گوناگون و متنوعی را به یکدیگر متصل می‌کند. کاربران وب دارای سوابق و تاریخچه‌ها، سلیق و اهداف کاربردی بسیار متفاوتی هستند. اکثر کاربران ممکن است دانش مناسبی در خصوص ساختار و شبکه اطلاعاتی نداشته باشند و از هزینه‌های سنگین جستجوی خاصی که در وب انجام می‌دهند مطلع نباشند.
- **همه چیز در وب وجود دارد.** فقط بخش کوچکی از اطلاعات موجود بر روی وب مفید و مرتبط است. گفته می‌شود که ۹۹ درصد از اطلاعات موجود بر روی وب برای ۹۹ درصد از کاربران وب هیچ فایده‌ای ندارد. اگرچه این موضوع ممکن است چندان واضح و آشکار نباشد ولی حقیقت دارد که یک شخص خاص به‌طور کلی فقط به بخش کوچکی از اطلاعات وب علاقمند است در حالیکه مابقی وب حاوی اطلاعاتی است که برای کاربر جذابیت ندارد. یافتن آن بخشی از وب که واقعاً برای یک کاربر جذاب و مفید است یکی از مسائل اساسی در جستجو و پژوهش‌های مرتبط با وب محسوب می‌شود.

این چالش‌ها موجب شده تا بسیاری از تلاش‌ها و فعالیت‌های پژوهشی به دنبال ارتقاء کارایی و اثربخشی کشف و بکارگیری دارایی‌های داده‌ای بر روی وب باشند. چندین موتور جستجوی وبی نمایه‌دار دائماً در وب به جستجو پرداخته و صفحات وب را با کلیدواژه‌های خاصی نمایه‌گذاری می‌کنند. یک کاربر مجرب با استفاده از این موتورهای جستجو قادر خواهد بود تا با وارد کردن عبارات یا کلیدواژه‌های محدود، اسناد مورد نظر خود را پیدا کند. با وجود این، یک موتور جستجوی ساده مبتنی بر کلیدواژه‌ها نواقص و کمبودهای متعددی دارد. اول اینکه یک موضوع بسیار کلی می‌تواند صدها یا هزاران سند را در بر بگیرد. بدین ترتیب، موتور جستجو ممکن است تعداد بسیار زیادی سند را پیدا کند که بسیاری از آنها چندان ارتباطی با موضوع ما نداشته باشند. دوم، بسیاری از اسنادی که به موضوع ما کاملاً مربوط هستند ممکن است حاوی کلیدواژه‌های دقیقی که معرف آنها باشند نباشند. وب‌کاوی در مقایسه با جستجوی وب مبتنی بر کلیدواژه‌ها یک روش شایع است که می‌توان از آن برای بهبود توان موتورهای جستجوی وبی استفاده کرد زیرا وب‌کاوی قادر به شناسایی صفحات وب، دسته‌بندی اسناد وب و رفع ابهامات و ظرافت‌های موتورهای جستجوی وب مبتنی بر کلیدواژه‌ها می‌باشد.

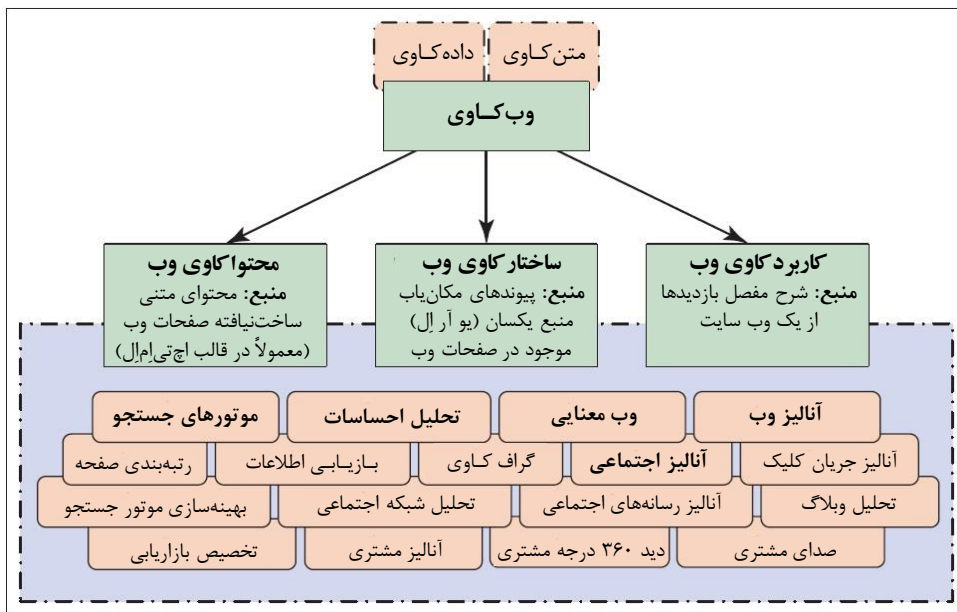
وب‌کاوی^۱ (یا داده‌کاوی وب) یک فرایند کشف ارتباطات درونی و ذاتی (یعنی اطلاعات مفید و جذاب) از داده‌های وب است که به شکل اطلاعات متنی، پیوندها یا اطلاعات کاربردها بیان می‌شوند. عبارت وب‌کاوی اولین بار توسط اِتسیونی^۲ (۱۹۹۶) بکار برده شد و امروزه بسیاری از کنفرانس‌ها، مجلات و کتاب‌ها بر داده‌کاوی وب تمرکز دارند. وب‌کاوی یک حوزه‌ی روش کسب‌وکار و فناوری است که در حال رشد و تکامل مستمر است. وب‌کاوی لزوماً همانند داده‌کاوی است که از داده‌های تولید شده بر روی وب استفاده می‌کند. هدف از وب‌کاوی، تبدیل مخازن عظیم و گسترده‌ی داده‌های تراکنش‌های کسب‌وکارها، تعاملات با مشتریان و داده‌های استفاده از وب‌سایت‌ها به اطلاعات اقدام‌پذیر (یعنی دانش) است تا بتوان تصمیم‌گیری بهتری را در سرتاسر سازمان انجام داد. به دلیل افزایش محبوبیت عبارت تحلیل (آنالیز)، امروزه بسیاری از مردم از وب‌کاوی با نام *آنالیز وب*^۳ یاد می‌کنند. با وجود این، این دو عبارت یکسان نیستند. در حالیکه آنالیز وب عمدتاً بر داده‌های مربوط به استفاده از وب‌سایت متمرکز است، وب‌کاوی شامل کلیه‌ی داده‌های تولید شده از طریق اینترنت از جمله تراکنش‌ها، اجتماعی و داده‌های کاربری است. هدف آنالیز وب این است که توضیح دهد که اتفاقی بر روی وب‌سایت افتاده است (بکارگیری یک متدولوژی آنالیز توصیفی از پیش تعریف شده) در حالیکه هدف وب‌کاوی، کشف الگوها و روابطی است که تا پیش از این ناشناخته بوده‌اند (بکارگیری یک متدولوژی تحلیلی تجویزی یا توصیفی نوین). از یک منظر بالاتر، آنالیز وب را می‌توان بخشی از وب‌کاوی در نظر گرفت. شکل ۷.۱۱ یک طبقه‌بندی ساده از وب‌کاوی را نشان می‌دهد که به سه حوزه‌ی اصلی تقسیم شده‌اند: محتواکاوی وب، ساختارکاوی وب و کاربردکاوی وب. در این تصویر، منابع داده‌ی بکار رفته در این سه حوزه نیز مشخص شده‌اند. اگرچه این سه حوزه به طور جداگانه نشان داده شده‌اند اما همانطور که در بخش بعدی خواهیم دید، اغلب در کنار یکدیگر برای رسیدگی به مسائل و مشکلات کسب‌وکار و بهره‌برداری از فرصت‌ها بکار برده می‌شوند.

همانطور که در شکل ۷.۱۱ مشاهده می‌کنید، وب‌کاوی عمدتاً متکی بر داده‌کاوی و متن‌کاوی و ابزارها و تکنیک‌های آنها است که در اوایل این فصل و فصل قبل به تفصیل شرح داده شدند. این شکل همچنین نشان می‌دهد که این سه حوزه‌های عمومی نیز به چندین حوزه‌های کاربردی بسیار معروف تقسیم شده‌اند. برخی از این حوزه‌ها در فصل‌های قبل توضیح داده شده‌اند و برخی نیز به تفصیل در این فصل شرح داده خواهند شد.

¹ Web mining

² Etsioni

³ Web Analytics



شکل ۷.۱۱ طبقه‌بندی ساده‌ی وب‌کاوی

محتواکاوی وب^۱ به استخراج اطلاعات مفید از صفحات وب اشاره دارد. اسناد را می‌تواند در یک قالب و شکل قابل خواندن توسط ماشین استخراج نمود به نحویکه تکنیک‌های خودکار می‌توانند برخی از این اطلاعات را از این صفحات وب استخراج کنند. خزنده‌های وب^۲ (یا اسپایدرها^۳) برای خواندن خودکار محتوای یک وب سایت بکار می‌روند. اطلاعات گردآوری شده می‌تواند مشابه آنچه در متن‌کاوی مورد استفاده قرار می‌گیرد شامل مشخصه‌های سند باشد ولی همچنین ممکن است شامل مفاهیم اضافی از قبیل ساختار سلسله‌مراتبی سند باشد. چنین فرایند خودکاری (یا نیمه خودکار) برای گردآوری و محتواکاوی را می‌توان برای کسب اطلاعات رقابتی (گردآوری اطلاعات در مورد محصولات، خدمات و مشتریان رقیب) بکار برد. همچنین می‌توان از آن برای گردآوری اطلاعات/اخبار/عقاید و خلاصه‌سازی، تحلیل احساسات و گردآوری و ساختاردهی خودکار داده‌ها برای مدل‌سازی پیشگویانه بکار برد. یک مثال بارز از بکارگیری محتواکاوی به عنوان یک ابزار خودکار گردآوری داده‌ها در ادامه آمده است. برای بیش از ۱۰ سال است که دو نفر از سه نویسنده‌ی این کتاب (دکتر شادرا و دیلن) به توسعه‌ی مدل‌هایی برای پیشگویی موفقیت مالی فیلم‌های هالیوودی پیش از پخش و انتشار آنها در سالن‌های سینما پرداخته‌اند. آنها از داده‌های چندین وب سایت برای آموزش مدل‌هایشان استفاده می‌کنند که هرکدام از این وب سایت‌ها ساختار صفحه‌بندی متفاوتی را دارند. گردآوری مجموعه بزرگی از متغیرها برای هزاران فیلم از این وب‌سایت‌ها فرایند بسیار زمان‌بر و دارای خطا است. از این رو، شادرا و دیلن از اسپایدرها و محتواکاوی وب به عنوان یک فناوری قدرتمند برای گردآوری، صحت‌گذاری، اعتبارسنجی و ذخیره‌سازی خودکار این مقادیر در یک پایگاه داده‌ی رابطه‌ای استفاده می‌کنند (در صورتیکه برخی داده‌های مشخص بر روی بیش از یک وب سایت موجود باشند آنگاه مقادیر آنها با

¹ Web Content Mining

² Web crawlers

³ Spiders

یکدیگر مقایسه و اعتبارسنجی شده و آنومالی‌ها ثابت و ضبط می‌شوند). بدین ترتیب، از کیفیت داده‌ها اطمینان حاصل شده و در عین حال در وقت ارزشمند ما نیز صرفه‌جویی می‌شود (چندین روز یا هفته).

علاوه بر متن، صفحات وب همچنین حاوی هایپرلینک‌هایی هستند که یک صفحه را به صفحه‌ی دیگری متصل می‌کند. هایپرلینک‌ها حاوی حجم زیادی از نمایه‌های انسانی مخفی هستند که می‌توانند به ما کمک کنند تا به صورت خودکار به مفهوم مرکزیت یا/اعتبار^۱ پی ببریم. زمانیکه یک برنامه‌نویس صفحات وب یک پیوند را بر روی صفحه قرار می‌دهد، این می‌تواند به معنای تأیید آن صفحه و وب توسط برنامه‌نویس باشد. تأیید یک صفحه‌ی وب توسط برنامه‌نویسان و توسعه‌دهندگان مختلف صفحات وب می‌تواند نشان دهنده‌ی اهمیت آن صفحه بوده و طبیعتاً منجر به کشف صفحات وب معتبر شود (میلر ۲۰۰۵). از این رو، حجم انبوهی از اطلاعات پیوندهای وبی می‌تواند مجموعه‌ای غنی از اطلاعات را در مورد مرتبط بودن، کیفیت و ساختار محتوای وب و منبعی غنی را برای وب‌کاوی فراهم کند.

از محتواکوی وب همچنین می‌توان برای ارتقاء نتایج بدست آمده از موتورهای جستجو استفاده کرد. در واقع، جستجو را شاید بتوان شایع‌ترین کاربرد محتواکوی وب و ساختارکوی وب دانست. با جستجوی یک موضوع خاص برای دستیابی به اطلاعات در وب (در قالب مجموعه‌ای کلیدواژه‌ها یا یک جمله) معمولاً چندین صفحه‌ی وب مرتبط و با کیفیت بالا و تعداد زیادی صفحات وب غیرقابل استفاده بازگردانده می‌شود. استفاده از نمایه‌های مرتبط بر روی کلیدواژه‌ها و صفحات معتبر موجب بهبود نتایج جستجو و رتبه‌بندی صفحات مربوطه می‌شود. ایده‌ی اعتبار از کارهای قبلی که در خصوص بازیابی اطلاعات با استفاده از استنادها در بین مقالات ژورنال‌ها برای ارزشیابی تأثیر مقالات پژوهشی انجام شده نشأت گرفته است (میلر ۲۰۰۵). اگرچه این سرمنشأ این ایده بود ولی تفاوت‌های اساسی بین استنادها در مقالات پژوهشی و هایپرلینک‌ها در صفحات وب وجود دارد. اول اینکه همه هایپرلینک‌ها به معنای تأیید نیستند (برخی پیوندها برای مقاصد مسیریابی و برخی دیگر به خاطر تبلیغات پولی ایجاد می‌شوند).

اگرچه این موضوع حقیقت دارد، اگر اکثریت هایپرلینک‌ها واقعاً به معنای تأیید باشند آنگاه عقیده‌ی جمعی همچنان پیروز و صادق است. دوم، به خاطر منافع تجاری و رقابتی، یک شخص دینفوذ به ندرت صفحه وب خودش را به رقبایی که در همان حوزه یا دامنه فعالیت دارند مرتبط می‌کند. به عنوان مثال، مایکروسافت ممکن است ترجیح می‌دهد که پیوندهای مربوط به سایت‌های اپل را بر روی صفحات وب خود قرار ندهد زیرا این موضوع می‌تواند به معنای تأیید اعتبار رقیب خودش باشد. سوم، صفحات معتبر به ندرت دارای ماهیت و ویژگی توصیفی هستند. به عنوان مثال، صفحه‌ی اصلی یاهو! ممکن است حاوی خودتوصیفی صریح نباشد که در واقع یک موتور جستجوی وبی است.

ساختار هایپرلینک‌های وب منجر به شکل‌گیری یک گروه مهم دیگر از صفحات وب به نام هاب^۲ شده است که به مجموعه‌ی یک یا چند صفحه‌ی وب گفته می‌شود که مجموعه‌ای از پیوندها را به صفحات معتبر ارائه می‌کنند. صفحات هاب ممکن است چندان برجسته و آشکار نباشند و فقط چند پیوند محدود به آنها وجود داشته باشد ولی هاب‌ها پیوندهایی را به مجموعه‌ای از سایت‌های برجسته و تراز اول در خصوص یک موضوع جالب ایجاد می‌کنند. یک هاب می‌تواند فهرستی از پیوندهای توصیه‌شده در خصوص صفحه اصلی یک وب سایت، سایت‌های مرجع توصیه شده بر روی یک صفحه وب یا فهرست تجمیع‌شده‌ای از منابع در خصوص یک موضوع خاص باشد. در واقع یک رابطه‌ی همزیستی بسیار نزدیک بین هاب‌های خوب و صفحات معتبر وجود دارد. یک هاب خوب از این لحاظ خوب است که به تعداد زیادی صفحه‌ی معتبر خوب اشاره می‌کند و یک صفحه‌ی معتبر خوب به این لحاظ خوب است که تعداد زیادی از هاب‌های خوب به آن ارجاع داده‌اند. اینگونه روابط بین هاب‌ها و صفحات معتبر موجب شده تا بتوان محتوای با کیفیت بالا را به صورت خودکار از وب بازیابی کرد.

¹ Authority

² Hub

شناخته شده‌ترین الگوریتم مرجع که برای محاسبه‌ی هاب‌ها و صفحات معتبر بکار می‌رود، جستجوی موضوعی القا شده با هایپرلینک (الگوریتم هیتس)^۱ نام دارد. این الگوریتم در ابتدا توسط کلاینبِرگ (۱۹۹۹) توسعه داده شده و پس از آن توسط پژوهشگران بسیاری بهبود داده شده است. الگوریتم هیتس یک الگوریتم تحلیل پیوند است که با استفاده از اطلاعات هایپرلینک‌هایی که درون صفحات وب قرار دارند به رتبه‌بندی آنها می‌پردازد. در مقوله‌ی جستجوی وب، الگوریتم هیتس یک مجموعه سند پایه را برای یک پرس و جوی خاص گردآوری می‌کند. سپس مقادیر هاب و صفحات معتبر را برای هر سند محاسبه می‌کند. برای گردآوری مجموعه سند پایه، یک مجموعه‌ی ریشه که با پرس و جو تطابق داشته باشد از یک موتور جستجو گرفته می‌شود. برای هر سند بازپایی شده، مجموعه اسنادی که به سند اصلی اشاره دارند و مجموعه دیگری از اسناد که سند اصلی به آنها اشاره دارد به عنوان همسایگی سند اصلی به مجموعه افزوده می‌شوند. فرایند رفت و برگشتی شناسایی سند و تحلیل پیوند تا زمانی ادامه می‌یابد که مقادیر هاب و صفحات معتبر به یکدیگر نزدیک شوند. سپس از این مقادیر برای نمایه‌گذاری و اولویت‌بندی مجموعه سند تولید شده برای یک پرس و جوی خاص استفاده می‌شود.

کاوش ساختار وب فرایند استخراج اطلاعات مفید از پیوندهایی است که درون اسناد وب کار گذاشته شده‌اند و از آن برای شناسایی صفحات معتبر و هاب‌ها استفاده می‌شود که ارکان اصلی الگوریتم‌های امروزی رتبه‌بندی صفحات هستند که نقش اساسی را در موتورهای جستجوی معروف از قبیل گوگل و یاهو! ایفا می‌کنند. همانطور که پیوندهایی که به یک صفحه وب اشاره دارند می‌توانند نشان دهنده‌ی محبوبیت سایت باشند (یا اعتبار آن)، پیوندهایی که درون یک صفحه وب (یا کل وب سایت) قرار دارند می‌توانند نشان دهنده‌ی عمق پوشش دهی یک موضوع خاص باشند. تحلیل پیوندها برای درک ارتباطات درونی بین تعداد زیادی از صفحات وب بسیار حائز اهمیت است که منجر به درک بهتر یک جامعه، دسته یا گروه وبی خاص می‌شود.

سوالاتی برای مرور بخش ۷.۷

۱۴. برخی چالش‌های اصلی صفحات وب برای کشف دانش چه هستند؟
۱۵. وب‌کاوی چیست؟ چه تفاوتی با متن‌کاوی و داده‌کاوی معمول دارد؟
۱۶. سه حوزه‌ی اصلی وب‌کاوی چه هستند؟
۱۷. محتواکاوی وب چیست؟ چگونه از آن می‌توان برای مزیت رقابتی استفاده کرد؟
۱۸. ساختار کاوی وب چیست؟ چه تفاوتی با محتواکاوی وب دارد؟

در این عصر و دوران، اهمیت موتورهای جستجوی اینترنتی را نمی‌توان به هیچ وجه انکار کرد. با افزایش حجم و پیچیدگی شبکه گسترده‌ی جهانی، یافتن آنچه به دنبال آن هستید بسیار پیچیده و پرزحمت شده است. مردم از موتورهای جستجو به دلایل گوناگونی استفاده می‌کنند. ما از آنها برای اطلاع از یک محصول یا خدمت پیش از آنکه تصمیم به خرید آن بگیریم (چه کس دیگری آن را می‌فروشد، فروشندگان/مغازه‌های مختلف به چه قیمت‌هایی می‌فروشند، مسائل متداولی که خریداران با آنها مواجه می‌شوند، خریداران قبلی تا چه اندازه رضایت داشته‌اند و چه محصولات یا خدمات بهتری ممکن است وجود داشته باشد) و همچنین جستجوی مکان‌هایی که می‌خواهیم برویم، افرادی که می‌خواهیم ملاقات کنیم و کارهایی که می‌خواهیم انجام دهیم استفاده

¹ Hyperlink-induced Topic Search (HITS)

می‌کنیم. موتورهای جستجو تبدیل به بخش اصلی اکثر تراکنش‌ها و سایر فعالیت‌های اینترنتی شده‌اند. موفقیت و محبوبیت خارق‌العاده‌ی گوگل که محبوب‌ترین شرکت موتور جستجو می‌باشد گواهی خوبی بر این ادعاست. آنچه برای بسیاری از مردم همچنان در حاله‌ای از رمز و ابهام قرار دارد این است که یک موتور جستجو واقعاً چگونه کار می‌کند. به ساده‌ترین بیان، یک موتور جستجو^۱ یک برنامه‌ی نرم‌افزاری است که بر اساس کلیدواژه‌ها (کلمات، عبارات چندکلمه‌ای یا یک جمله‌ی کامل) که کاربران در مورد یک موضوع یا سوژه‌ی خاص وارد می‌کنند به جستجوی اسناد (سایت‌های اینترنتی یا فایل‌ها) می‌پردازد. موتورهای جستجو در واقع نیروهای پیشران اینترنت هستند که روزانه پاسخگوی میلیاردها پرس و جو به صدها زبان مختلف می‌باشند.

به بیان فنی، «موتور جستجو» یک عبارت متداول برای سیستم‌های بازیابی اطلاعات است. اگرچه موتورهای جستجوی وبی، متداول‌ترین نوع موتورهای جستجو هستند ولی موتورهای جستجو اغلب در مقوله‌هایی به غیر از وب بکار می‌روند از جمله موتورهای جستجوی رومیزی و موتورهای جستجوی اسناد. همانطور که در این بخش خواهید دید، بسیاری از مفاهیم و تکنیک‌هایی که در بخش تحلیل متن و متن‌کاوی ارائه کردیم در اینجا نیز صادق هستند. هدف کلی یک موتور جستجو این است که یا چند سند/صفحه را که بیش از همه با پرس و جوی کاربر انطباق دارد بازیابی کند (اگر بیش از یک سند/صفحه وجود داشته باشد آنگاه یک فهرست رتبه‌بندی شده را ارائه می‌کند). به منظور ارزیابی موتورهای جستجو از دو سنجه استفاده می‌شود: ^۲ اثربخشی (کیفیت-یافتن سند/صفحات درست) و ^۳ کارایی (سرعت - به سرعت بازگرداندن نتیجه‌ی جستجو). این دو سنجه در جهت معکوس کار می‌کنند یعنی بهبود یک سنجه موجب بدتر شدن سنجه‌ی دیگر می‌شود. اغلب موتورهای جستجو بر اساس انتظارات کاربران بر یکی از این دو سنجه تمرکز می‌کند. موتورهای جستجوی بهتر آنهایی هستند که به طور همزمان هر دو سنجه را بهبود بخشند. از آنجائیکه موتورهای جستجو نه فقط به جستجو بلکه به یافتن و بازگرداندن صفحات/اسناد می‌پردازند شاید مناسب‌ترین نام برای آنها موتورهای یابنده^۴ باشد.

اکنون اجازه بدهد یک موتور جستجو را کالبدشکافی کرده و به درون آن نگاه کنیم. در بالاترین سطح، یک سیستم موتور جستجو از دو چرخه‌ی اصلی تشکیل شده است: یک چرخه‌ی توسعه و یک چرخه‌ی پاسخگویی (نگاه کنید به ساختار یک موتور جستجوی اینترنتی معمولی در شکل ۷.۱۲). در حالیکه یک چرخه با شبکه گسترده‌ی جهانی رابطه دارد، چرخه‌ی دیگر با کاربر در ارتباط است. چرخه‌ی توسعه را می‌توان به صورت یک فرایند تولید (ساخت و تولید اسناد/صفحات) و چرخه‌ی پاسخگویی را به صورت یک فرایند خرده‌فروشی (ارائه آنچه مشتریان/کاربران می‌خواهند) در نظر گرفت. در بخش بعدی به توضیح بیشتر این دو چرخه خواهیم پرداخت.

۱.

دو مؤلفه‌ی اصلی چرخه‌ی توسعه عبارتند از خزنده‌های وب و نمایه‌گذار سند. هدف این چرخه این است که یک پایگاه داده عظیم از اسناد/صفحات را که بر اساس محتوا و ارزش اطلاعاتی‌شان سازماندهی و نمایه‌گذاری شده‌اند ایجاد کند. دلیل ایجاد چنین مخزنی

¹ Search Engine

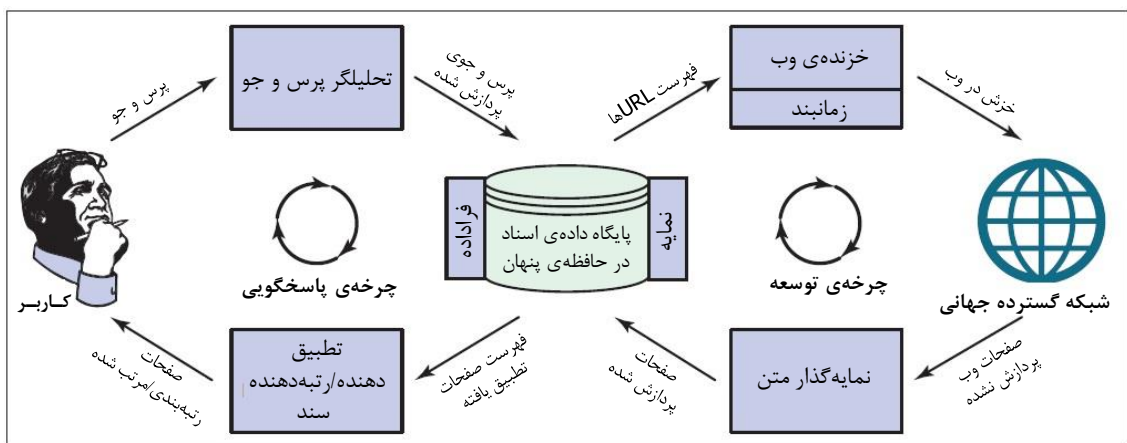
² Effectiveness

³ Efficiency

⁴ Finding Engines

از اسناد/صفحات کاملاً واضح است: به دلیل حجم و پیچیدگی بسیار بالای وب، جستجو در وب برای یافتن صفحات مورد نظر کاربر امکان‌پذیر نیست (یا حداقل اینکه در یک بازه‌ی زمانی معقول امکان‌پذیر نیست) و از این رو موتورهای جستجو، وب را به صورت حافظه‌ی موقت در پایگاه‌های داده خود قرار داده و از این نسخه‌ی کپی شده در حافظه موقت برای جستجو و یافتن استفاده می‌کنند. پس از ایجاد این پایگاه داده، موتورهای جستجو می‌توانند به سرعت و دقت به پرس و جوی کاربر پاسخ دهند.

خزنده‌های وب - یک خزنده‌ی وب (که آن را اسپایدار یا وب اسپایدار نیز می‌نامند) یک بخش نرم‌افزاری است که به صورت سیستماتیک برای یافتن و بازبایی صفحات وب در شبکه گسترده جهانی به جستجو می‌پردازد. خزنده‌های وب اغلب تمامی صفحاتی که بازدید می‌کنند را برای پردازش‌های بعدی توسط موتور جستجو کپی می‌کنند.



شکل ۷.۱۲ ساختار یک موتور جستجوی اینترنتی معمولی

یک خزنده‌ی وب کارش را با فهرستی از یوآرال‌ها که در زمان‌بند آورده شده‌اند و اغلب هسته نامیده می‌شوند شروع می‌کند. این یوآرال‌ها ممکن است از فرم‌های پذیرش که توسط وب‌مستر ایجاد شده‌اند یا از آبرپیوندهای داخلی از صفحات/اسناد قبلی آمده باشند. خزنده با بازدید از این یوآرال‌ها تمامی ابرپیوندهای موجود در آن صفحه را شناسایی کرده و آنها به فهرست یوآرال‌هایی که باید بازدید شوند می‌افزاید. یوآرال‌ها در زمان‌بند بر اساس مجموعه‌ای از سیاست‌های تعیین شده توسط موتور جستجوهای خاص مکرراً مورد بازدید قرار می‌گیرند. با توجه به حجم بسیار بالای صفحات وب، خزنده می‌تواند فقط تعداد محدودی از آنها را در یک بازه زمانی مشخص بارگذاری کند و از این رو ممکن است نیاز به اولویت‌بندی بارگذاری‌ها باشد.

نمایه‌گذار سند^۱ - با شناسایی و استخراج اسناد توسط خزنده، آنها در یک فضای موقت ذخیره می‌شوند تا توسط نمایه‌گذار سند مورد پردازش قرار گیرند. نمایه‌گذار سند مسئولیت پردازش اسناد (صفحات وب یا فایل‌های سند) و قراردادن آنها در پایگاه داده سند را بر عهده دارد. برای تبدیل اسناد/صفحات به قالب مطلوب که به آسانی قابل جستجو باشد، نمایه‌گذار شند وظایف زیر را انجام می‌دهد.

¹ Document Indexer

بر روی پایگاه داده سند کار می کنند؛ یکی اسناد/صفحات را با استفاده از یک ساختار نمایه خاص وارد کرده و دیگری، یک رشته‌ی پرس و جو را به همان ساختار تبدیل می کند که می توان آن را برای موقعیت یابی سریع مربوطترین صفحات/اسناد بکار برد.

رتبه‌بند/انطباق دهنده سند^۱ – این جایی است که داده‌های پرس و جوی ساختیافته برای یافتن مرتبطترین اسناد/صفحات با پایگاه داده سند تطبیق داده شده و آنها را بر حسب اهمیت/ارتباط رتبه‌بندی می کنند. اثربخشی این گام شاید مهمترین مؤلفه در زمان مقایسه‌ی موتورهای جستجوی مختلف باشد. هر موتور جستجو الگوریتم خودش را دارد که از آن برای اجرای این گام مهم استفاده می کند.

موتورهای جستجوی اولیه از یک تطبیق ساده‌ی کلیدواژه با پایگاه داده سند استفاده کرده و فهرستی از صفحات/اسناد مرتب شده را نتیجه می دادند که مخرج آن تابعی بود که از تعداد واژه‌ها/عبارات تطبیق یافته بین پرس و جو و سند به همراه اوزان آنها واژه‌ها/عبارات استفاده می کرد. کیفیت و میزان مفید بودن نتایج جستجو به آن خوبی که انتظار می رفت نبود. آنگاه در سال ۱۹۹۷، خالقان گوگل به الگوریتم جدیدی دست یافتند و آن را پیج رنک^۲ نامیدند. همانگونه که از نامش پیداست، پیج رنک روشی الگوریتمی برای رتبه‌بندی اسناد/صفحات بر اساس میزان مرتبط بودن و اهمیت/ارزش آنها بوده است. اگرچه پیج رنک روشی نوآورانه برای رتبه‌بندی صفحات/اسناد است یک حالت افزوده برای فرایند بازبایی اسناد مرتبط از پایگاه داده و رتبه‌بندی آنها بر اساس اوزان واژه‌ها/عبارات است. گوگل تمامی این وظایف را برای شناسایی مرتبطترین فهرست از اسناد/صفحات برای یک درخواست جستجوی خاص به طور یکجا انجام می دهد. پس از اینکه فهرس مرتب شده‌ای از اسناد/صفحات ایجاد شد، در یک قالب ساده و قابل فهم به کاربر بازگردانده می شود. در این مقطع، کاربران ممکن است بر روی هر کدام از اسناد داخل فهرست کلیک کنند و ممکن است لزوماً سندی در بالای فهرست آمده نباشد. اگر بر روی سندی که در بالای فهرست آمده کلیک نکنند آنگاه آیا می توانیم نتیجه بگیریم که موتور جستجو کارش را به خوبی انجام نداده است؟ شاید اینطور باشد. موتورهای جستجوی پیشگام مثل گوگل با ثبت، ضبط و تحلیل تجربیات و اقدامات کاربران به پایش عملکرد نتایج جستجویشان می پردازند. این تحلیل ها اغلب منجر به قواعد بیشتر و بیشتری برای بازتعریف و اصلاح رتبه‌بندی صفحات/اسناد می شوند تا پیوندهایی که در قسمت بالاتر فهرست نمایان می شوند برای کاربران نهایی ارجح تر باشند.

بهینه‌سازی موتور جستجو^۳ یک فعالیت‌های هدفمند برای تأثیرگذاری بر رؤیت پذیری یک سایت یا وب سایت تجارت الکترونیک در نتایج جستجوی طبیعی یک موتور جستجو است. به طور کلی، هرچه در رتبه بالاتری در صفحه قرار بگیرد و به دفعات بیشتری در نتایج جستجو پدیدار شود، کاربران بیشتری از طریق موتورهای جستجو از آن سایت بازدید خواهند کرد. بهینه‌سازی موتور جستجو به عنوان یک استراتژی بازاریابی اینترنتی، نحوه عملکرد موتورهای جستجو، آنچه افراد به دنبال آن هستند، عبارات یا کلیدواژه‌های تایپ شده در موتورهای جستجو و موتورهای جستجوی مورد ترجیح مخاطبان هدف را مد نظر قرار می دهد. بهینه‌سازی یک وب سایت می تواند شامل ویرایش محتوای آن، اچ تی ام ال و کدگذاری مربوطه برای افزایش میزان مربوط بودن آن با کلیدواژه‌های مشخص و حذف موانع فعالیت‌های نمایه گذاری موتورهای جستجو باشد.

ارتقاء یک سایت برای افزایش تعداد بکلینک و باندلینک‌ها نیز یک تاکتیک دیگر بهینه‌سازی موتور جستجو می باشد.

¹ Document Matcher/Ranker

² PageRank

³ Search Engine Optimization

اوایل، تنها چیزی که وب‌مسترها برای نمایه‌گذاری لازم بود انجام دهند این بود که آدرس یک صفحه یا یوآرال را به موتورهای جستجوی مختلف ارسال می‌کردند و سپس یک «اسپایدر» را برای «خزش» در آن صفحه می‌فرستادند که پیوندها به صفحات دیگر را از داخل آن استخراج کرده و اطلاعات یافت شده بر روی آن صفحه را برای نمایه‌گذاری به آن سرور می‌فرستاد. این فرایند که قبلاً توضیح داده شد شامل یک اسپایدر موتور جستجو است که یک صفحه را بارگذاری کرده و آن را بر روی سرور خود موتور جستجو ذخیره می‌کند و یک برنامه دوم که تحت عنوان نمایه‌گذار شناخته می‌شود، اطلاعات مختلف در مورد آن صفحه از قبیل تعداد واژه‌ها، محل قرار گرفتن آنها در متن و وزن واژه‌ها و کلمات خاص را استخراج می‌کند. امروزه، موتورهای جستجو برای ارسال یوآرال‌ها متکی به وب‌مسترها نیستند در عوض، موتورهای جستجو به طور مستمر و پیشگراانه در وب به کاوش، اخذ و نمایه‌گذاری هرگونه اطلاعات می‌پردازند.

نمایه‌گذاری توسط موتورهای جستجویی از قبیل گوگل، بینگ و یاهو! برای کسب‌وکارها کافی نیست بلکه رتبه‌بندی بر روی معروف‌ترین و پرکاربردترین موتورهای جستجو و دستیابی به رتبه بالاتر نسبت به رقباست که تفاوت را در چشم مشتریان و سایر ذینفعان بوجود می‌آورد. روش‌های گوناگون می‌توانند رتبه‌بندی یک صفحه وب را در محدوده‌ی نتایج جستجو بالا ببرند. ایجاد پیوندها بین صفحات یک وب سایت برای برقراری پیوند به مهم‌ترین صفحات می‌تواند رؤیت‌پذیری آن صفحه وب را افزایش دهد. نوشتن محتوایی که شامل عبارات و جملاتی باشد که مکرراً جستجو شده‌اند و بتواند پاسخگوی انواع پرس و جوها باشد موجب افزایش ترافیک وب می‌شود. به‌روزآوری محتوا برای اینکه موتورهای جستجو به طور مستمر به خزش ادامه دهند می‌تواند وزن بیشتری به وب سایت‌ها بدهد. افزودن کلیدواژه‌های مرتبط به فراداده‌های یک صفحه وب از جمله تگ عنوان^۱ و متادیسکریپشن^۲ موجب بهبود مرتبط بودن فهرست جستجوی یک سایت و در نتیجه، افزایش ترافیک به آن سایت می‌شود. نرمال‌سازی یوآرال صفحات وب (که از طریق چندین یوآرال قابل دسترس باشند) و استفاده از عناصر پیوند کانونیکال^۳ و مسیردهی‌های مجدد^۴ می‌تواند به کمک کند که مطمئن شویم پیوندها به نسخه‌های مختلف صفحات وب و یوآرال‌های آنها تماماً به امتیاز محبوبیت پیوند وب سایت می‌افزاید.

به طور کلی، تکنیک‌های بهینه‌سازی موتورهای جستجو را می‌توان به دو گروه اصلی تقسیم کرد: تکنیک‌هایی که موتورهای جستجو به عنوان بخشی از طراحی خوب سایت توصیه می‌کنند و تکنیک‌هایی که موتورهای جستجو تأیید نمی‌کنند. موتورهای جستجو سعی دارند تا تأثیر تکنیک‌هایی که مورد تأیید نیستند را به حداقل برسانند که اغلب این را هرزآگهی یا پیام بازرگانی هرز^۵ می‌نامند. مفسران صنعت و کارشناسانی که این تکنیک‌ها را بکار می‌گیرند این روش‌ها را به دو گروه بهینه‌سازی موتور جستجوی کلاه سفید^۶ و بهینه‌سازی موتور جستجوی کلاه سیاه^۷ تقسیم کرده‌اند (گودمن ۲۰۰۵). کلاه سفیدها نتایجی را تولید می‌کنند که مدت زمان طولانی باقی می‌مانند در حالیکه کلاه سیاه‌ها پیش‌بینی می‌کنند که به محض اینکه موتورهای جستجو کشف کنند که سایت‌های آنها چه کاری انجام می‌دهند آنها را نهایتاً به طور موقت یا دائم ممنوع کنند.

¹ Title tag

² Metadescription

³ Canonical link elements

⁴ Redirects

⁵ Spamdexing, search spam, search engine spam, search engine poisoning

⁶ White-hat SEO

⁷ Black-hat SEO

یک تکنیک بهینه‌سازی موتور جستجو در صورتی کلاه سفید در نظر گرفته می‌شود که با اصول راهنمای موتور جستجو تطابق داشته و هیچگونه فریب‌کاری و دغل‌کاری نداشته باشد. از آنجائیکه اصول راهنمای موتورهای جستجو به صورت مجموعه‌ای از قوانین مکتوب نشده‌اند باید به این وجه تمایز بسیار توجه نمود. کلاه سفیدها فقط شامل پیروی از اصول راهنما نیستند بلکه باید اطمینان حاصل شود که محتوایی که یک موتور جستجو آن را نمایه‌گذاری و سپس رتبه‌بندی می‌کند دقیقاً همان محتوایی است که کاربر خواهد دید. توصیه‌ی کلاه سفیدها این است که محتوا برای کاربران ایجاد شود نه برای موتورهای جستجو و سپس این محتوا به آسانی برای اسپایدرها قابل دسترسی باشد به جای اینکه سعی شود تا الگوریتم را از هدفی که برایش در نظر گرفته شده منحرف کنیم. بهینه‌سازی کلاه سفید به طرق مختلف مشابه توسعه‌ی وب است که دسترسی پذیری را ترویج می‌کند اگرچه این دو یکسان نیستند.

بینش فناوری ۷.۳ ۱۵ موتور جستجوی برتر (آگوست ۲۰۱۶)

اینها ۱۵ موتور جستجوی برتر بر اساس رتبه‌بندی (ebizmba.com/articles/search-engines) eBizMBA هستند که میانگین رتبه‌ی آکسا وب سایت است که به طور دائم به‌روزرآوری می‌شود.

رتبه	نام	پیش‌بینی تعداد بازدیدکنندگان ماهانه
۱	Google	۱,۶۰۰,۰۰۰,۰۰۰
۲	Bing	۴۰۰,۰۰۰,۰۰۰
۳	Yahoo! Search	۳۰۰,۰۰۰,۰۰۰
۴	Ask	۲۴۵,۰۰۰,۰۰۰
۵	AOL Search	۱۲۵,۰۰۰,۰۰۰
۶	Wow	۱۰۰,۰۰۰,۰۰۰
۷	WebCrawler	۶۵,۰۰۰,۰۰۰
۸	MyWebSearch	۶۰,۰۰۰,۰۰۰
۹	Infospace	۲۴,۰۰۰,۰۰۰
۱۰	Info	۱۳,۵۰۰,۰۰۰
۱۱	DuckDuckGo	۱۱,۰۰۰,۰۰۰
۱۲	Contentko	۱۰,۵۰۰,۰۰۰
۱۳	Dogpile	۷,۵۰۰,۰۰۰
۱۴	Alhea	۴,۰۰۰,۰۰۰
۱۵	ixQuick	۱,۰۰۰,۰۰۰

بهینه‌سازی موتور جستجوی کلاه سیاه سعی دارد تا رتبه‌بندی‌ها را به نحوی که مورد تأیید موتورهای جستجو نیست یا از طریق فریب‌کاری بهبود بخشد. یکی از تکنیک‌های کلاه سیاه از یک متن پنهان یا مشابه یک متن رنگی در پس زمینه در داخل یک **div tag** نامرئی (که بخش یا یک قسمت در یک سند اچ‌تی‌ام‌ال را مشخص می‌کند) یا جایی خارج از صفحه استفاده می‌کند. در یک روش دیگر، بسته به اینکه آیا یک بازدیدکننده‌ی انسانی یا یک موتور جستجو صفحه‌ای را درخواست کرده باشد، صفحه‌ی

متفاوتی نمایش داده می‌شود که این تکنیک را پنهان‌کاری^۱ می‌نامند. موتورهای جستجو می‌توانند سایت‌هایی که از روش‌های کلاه سیاه استفاده می‌کنند را یا از طریق کاهش رتبه‌بندی آنها یا حذف فهرست آنها از پایگاه‌های داده جریمه کنند. اینگونه جریمه‌ها ممکن است به صورت خودکار توسط الگوریتم‌های موتورهای جستجو یا یک مرور دستی سایت اعمال شوند. یک مثال از آن حذف بی‌ام‌دبلیو آلمان و ریکو آلمان از گوگل در فوریه‌ی ۲۰۰۶ به خاطر استفاده از روش‌های تأیید نشده بود (کاتز ۲۰۰۶). هر دو شرکت سریعاً عذرخواهی کرده و روش‌هایشان را اصلاح کردند و مجدداً در فهرست گوگل قرار گرفتند. برای برخی از کسب‌وکارها، بهینه‌سازی موتور جستجو می‌تواند بازگشت سرمایه‌ی بسیار زیادی به همراه داشته باشد. با وجود این، باید در نظر داشت که موتورهای جستجو بابت ترافیک جستجو پولی دریافت نمی‌کنند، الگوریتم‌های آنها دائماً تغییر می‌کند و هیچ تضمین برای تداوم ارجاعات وجود ندارد. به دلیل این عدم اطمینان و پایداری، کسب و کاری که شدیداً به ترافیک موتور جستجو متکی است در صورتیکه موتورهای جستجو الگوریتم‌هایشان را تغییر دهند دچار ضرر و زیان‌های سنگینی می‌شوند. طبق اظهارات مدیرعامل اجرایی گوگل، اریک اشمیت در سال ۲۰۱۰، گوگل بیش از ۵۰۰ بار الگوریتم خود را تغییر داد یعنی ۱.۵ مرتبه در روز. به خاطر دشواری همگامی با تغییرات قوانین موتورهای جستجو، شرکت‌هایی که بر روش ترافیک جستجو متکی هستند یک یا چند روش زیر را بکار می‌گیرند: (۱) شرکتی را که در زمینه‌ی بهینه‌سازی موتور جستجو تخصص دارد برای بهبود مستمر ظاهر وب سایت متناسب با روش‌های متغیر موتورهای جستجو استخدام می‌کنند (۲) به شرکت‌های ارائه‌دهنده‌ی موتورهای جستجو پول می‌دهند تا در فهرست بخش حامیان و اسپانسرهای پولی قرار بگیرند و (۳) خودشان را از وابستگی به ترافیک موتور جستجو رها می‌کنند.

مهمترین چیز برای یک سایت تجارت الکترونیک این است که احتمال انجام تراکنش‌ها توسط مشتریان را چه از طریق یک موتور جستجو یا از سایت‌های دیگر به حداکثر برساند. اینکه تعداد زیادی بازدید کننده داشته باشیم بدون اینکه فروش اتفاق بیوفتد هدف اصلی ایجاد یک وب سایت تجارت الکترونیک نیست. مورد کاربردی ۷.۷ در مورد یک مرکز خرید اینترنتی بزرگ بحث می‌کند که از تحلیل رفتار مشتری (با استفاده از کلیک استریم و سایر منابع داده) برای بهبود چشمگیر نرخ تبدیل استفاده می‌شود. آنچه برای یک وب سایت تجارت الکترونیک مهم است این است که تراکنش‌های فروش مشتریان خود را به حداکثر برساند صرف‌نظر از اینکه از طریق یک موتور جستجو یا پاسخگویی به کمپین‌های بازاریابی پست الکترونیکی یا سایت‌های رسانه‌های اجتماعی باشد. مورد کاربردی ۷.۷ نشان می‌دهد که یک شرکت تولید پوشاک و لباس‌های مد روز که یک قرن سابقه دارد چگونه از کمپین‌های مبتنی بر نامه‌های لکترونیکی برای تولید مشتریان راغب (سرنخ) برای کسب‌وکار تجارت الکترونیک خود چگونه استفاده می‌کند.

¹ Cloaking

ارائه‌ی محتوای اختصاصی (فردی) و پیشبرد درگیرسازی دیجیتال: چگونه باربور بیش از ۴۹۰,۰۰۰ مشتری راغب جدید را در عرض ۱ ماه با استفاده از ترادیتا اینتراکتیو گردآوری کرد

مورد کاربردی ۷.۷

تاریخچه

باربور^۱ که در سال ۱۸۹۴ تأسیس شد یک برند انگلیسی سبک زندگی است که به خاطر پوشاک ضدآب و به ویژه ژاکت کلاسیک از جنس پشم واکس زده خود معروف است. باربور با سفارش و تولید بیش از ۱۰,۰۰۰ ژاکت در سال به مدت بیش از یک قرن است که موقعیت بسیار قوی و ویژه‌ای در صنعت کالاهای لوکس داشته و رابطه‌ی قوی را با زنان و مردان اهل مد از سرزمین انگلستان برقرار کرده است. در سال ۲۰۰۰، باربور ارائه‌ی محصولات خود را توسعه داده و طیف کاملی از البسه و زیورآلات زندگی روزمره را به مشتریان ارائه کرد. بازارهای اصلی این شرکت در انگلستان، ایالات متحده و آلمان است ولی باربور در بیش از ۴۰ کشور در سرتاسر جهان از جمله اتریش، زلاندنو و ژاپن حضور دارد. باربور با استفاده از بینش‌های اختصاصی و فردی که از خدمات و قابلیت‌های بازاریابی دیجیتال ترادیتا اینتراکتیو^۲ استخراج کرده است یک کمپین یک ماهه را اجرا کرد که منجر به بیش از ۴۹,۷۰۰ مشتری راغب جدید و ۴۵۰,۰۰۰ کلیک در وب سایت خود شد.

چالش: در تملک گرفتن ارتباطات مشتریان

باربور در طول عمر فعالیت خود رشد پایدار و چشمگیری را داشته و در آگوست ۲۰۱۳ اولین سایت تجارت الکترونیک خود را برای بهره‌گیری از یک حضور آنلاین قوی‌تر راه‌اندازی کرد. با وجود این، اگرچه کار خود را در دنیای تجارت الکترونیک با تأخیر شروع کرد، جا افتادن در این عرصه‌ی دیجیتال اشباع شده برای باربور یک چالش محسوب می‌شد. این شرکت که قبلاً محصولاتش را فقط از طریق عمده‌فروشی‌ها و خرده‌فروشی‌های مستقل به بازار عرضه می‌کرد به دنبال این بود که مالکیت ارتباطات با مصرف‌کنندگان و کاربران نهایی خود را در اختیار بگیرد. اگرچه این برند در سرتاسر جهان بسیار سرشناس و مورد احترام است باربور از اهمیت برقراری ارتباطات مستقیم با مخاطبان هدف آگاه است به ویژه زمانیکه کاربران را ترغیب می‌کند تا با پلتفرم جدید تجارت الکترونیک شرکت ارتباط بگیرند. شرکت همچنین متوجه شد که شکل دادن به سفر مشتری را باید بیشتر در کنترل بگیرد. بدین ترتیب باربور توانست به همان سطح کیفی استثنایی که در تولید محصولاتش دست پیدا کرده بود در تجربه‌ی مشتری و کاربر نیز دست پیدا کند. برای اینکار، شرکت باید رفتار آنلاین بازار هدف خود را درک می‌کرد. باربور با هدف دستیابی به مخاطبان هدف برای برقراری روابط معنادار با مشتریان از ترادیتا یاری جست. دپارتمان بازاریابی باربور از ترادیتا اینتراکتیو برای ارائه‌ی راه‌حلی کمک خواست تا بتواند دانش خود را در خصوص ویژگی‌ها و نیازهای منحصر بفرد تک تک مشتریان و همچنین پشتیبانی از راه‌اندازی وب سایت جدید تجارت الکترونیک در انگلستان ارتقاء دهد.

¹ Barbour

² Teradata Interactive

راه‌حل: پیاده‌سازی یک برنامه‌ی پرورش مشتریان راغب

شیفت روزافزون به سمت تجارت الکترونیک جهانی و رشد کنسرسیوم‌های دیجیتال موجب شد تا برندها مجبور باشند تا حضور آنلاین پررنگی باشند. این همچنین بدان معناست که خرده‌فروشی‌ها باید استراتژی‌هایی را اجرا کنند که از نیازها و خواسته‌های روزافزون مشتریان چه به صورت آنلاین و چه آفلاین پشتیبانی کند.

باربور و ترادیتا اینتراکتیو به طراحی و ساخت یک برنامه‌ی پرورش مشتریان راغب برای یک دوره‌ی ۱ ماهه پرداختند. هدف این کمپین نه تنها افزایش سطح آگاهی و ایجاد تقاضا برای فروش بی‌واسطه بلکه ایجاد یک مکانیزم مشارکت و درگیرسازی بلندمدت‌تر بود که منجر به فروش بیشتر در یک دوره زمانی طولانی‌تر شود. از همان ابتدا روشن بود که برقراری روابط قوی بین باربور و مشتریانش یک عامل حیاتی بود که آن را از سایر رقبای خرده‌فروشی لوکس مجزا می‌کرد. ترادیتا اینتراکتیو مشتاق بود اطمینان حاصل کند که این رابطه از طریق یک فرایند تولید مشتریان راغب جدید حفظ و تقویت می‌شود.

اجرای یک کمپین تبلیغاتی خاص باربور بود. کمپین‌های معمول برای تولید مشتریان راغب اغلب به صورت رویدادهای تک‌ثبت‌نامی با یک تبلیغ فروش در ذهن اجرا می‌شدند. این داده‌ها معمولاً محدود به آدرس پست الکترونیک و شرح حال بود که بدون در نظر گرفتن نیازهای شخصی ثبت‌نام‌کنندگان ایجاد شده و فقط برای کمپین‌های خبرنامه‌ای معمول بکار می‌رفتند. این استراتژی اغلب موجب می‌شد تا برندهای مطرح در زمان مطلع شدن از نقطه نظرات و دیدگاه‌های مشتریان، فرصت‌های بزرگ را از دست بدهند که اغلب منجر به فروش ضعیف می‌شد. ترادیتا اینتراکتیو متوجه شد که ارزش واقعی تولید مشتریان راغب، دو برابر است. اول از همه، با استفاده از یک رویکرد نام‌نویسی برای گردآوری حداکثر اطلاعات ممکن است، به یک درک از نیات خرید آتی و عوامل تأثیرگذار آن دست یافتند. سپس، با حصول اطمینان از اینکه داده‌های گردآوری شده به طور مؤثری برای حلق محتوای ارزشمند و اختصاصی بکار رفته‌اند، فرصت‌های فروش زمانیکه مشتریان برای خرید به بازار مراجعه می‌کنند فراهم می‌شوند. برای حصول اطمینان از اینکه این استراتژی به فروش بلندمدت منجر می‌شود، ترادیتا اینتراکتیو یک برنامه‌ی چرخه طول عمر مشتری را ایجاد کرد که محتوا را از طریق پست الکترونیکی و نمایشگرهای آنلاین ارسال می‌کرد.

محتوای برنامه همراه با تبلیغات نمایشگرها بوده و به اشتراک‌گذاری رسانه‌های اجتماعی را ترویج می‌نمود. باربور از طریق تگ‌گذاری هوشمند ترادیتا اینتراکتیو توانست مخاطبان را بر اساس اولویت‌های محصول بخش‌بندی کرده و بنرهای هدف‌گیری مجدد را راه‌اندازی کند. از ثبت‌نام‌کنندگان دعوت شد تا محتوای خود را در شبکه اجتماعی به اشتراک بگذارند که ترادیتا اینتراکتیو را قادر می‌کرد تا کاربران را برای طرح‌های وفاداری آینده و فعالیت‌های «به دوست خود بگویند» شناسایی کند. علاوه بر تمرکز بر افزایش خبرنامه‌های باربور، ترادیتا یک ممیزی داده برای تحلیل تمامی داده‌های گردآوری شده و درک بهتر عواملی که بر رفتار درگیرسازی کاربران تأثیر دارد اجرا نمود.

نتایج

همکاری قوی بین ترادیتا و باربور بدین معنا بود که در طی یک کمپین یک ماهه، باربور توانست روش‌های جدید و نوآورانه‌ای را برای برقراری ارتباط با مشتریان خود ایجاد کند. بیش از ۴۹,۷۰۰ مشتری راغب در مناطق انگلستان و آلمان، اتریش و سوئیس گردآوری شدند و برنامه تولید مشتریان راغب نشان داد که نرخ کلیک بین ۴ و ۱۱ درصد بوده است. طی این کمپین همچنین بیش از ۴۵۰,۰۰۰ کلیک بر روی وب سایت باربور شده و به قدری بین بلاگرهای مد و نشریات ملی محبوبیت یافت که به صورت یک داستان در *The Daily Mirror* چاپ شد. اگرچه این کمپین فقط یک ماه به طور انجامید، تمرکز اصلی، کمک به استراتژی بازاریابی آینده‌ی باربور بود. یک نظرسنجی در خصوص طراحی کمپین انجام شد که منجر به نرخ تکمیل مشوقی ۶۵ درصدی گردید. داده‌های کاربران شامل موارد زیر بوده است:

- مشارکت در شبکه‌های اجتماعی
- مشارکت و درگیری‌سازی دستگاه‌ها
- موقعیت تا نزدیک‌ترین فروشگاه
- ملاحظات مهم در نظر مشتری

باربور به واسطه‌ی سطح عمیقی از بینشی که بدست آورد به یک توانایی عظیم برای ارائه‌ی محتوای شخصی به کاربرانش دست پیدا کرد.

سوالاتی برای مورد کاربردی ۷.۷

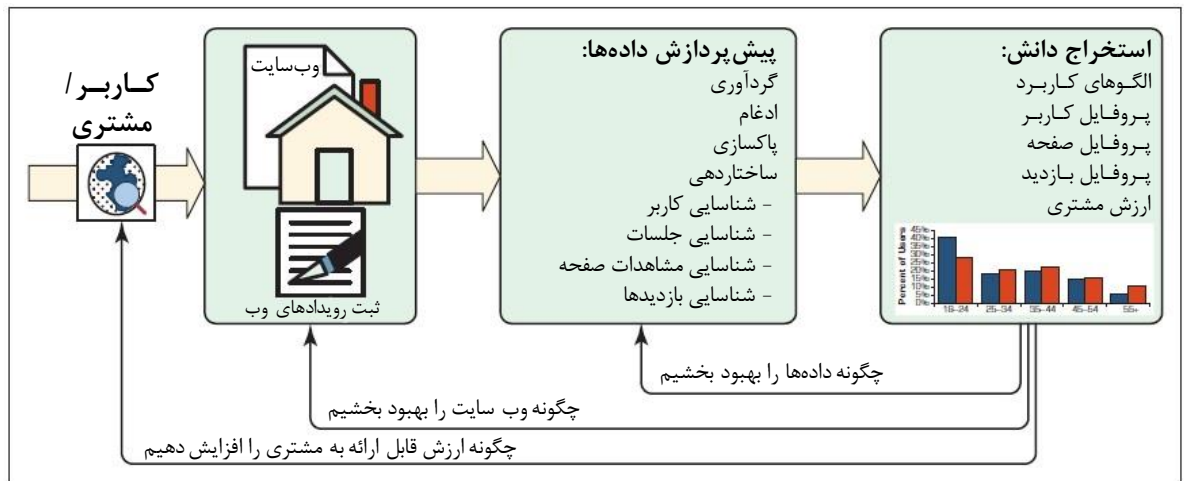
۱. باربور به چه کسب‌وکاری مشغول است؟ باربور با چه چالشی مواجه بود؟
۲. راه‌حل تحلیلی پیشنهادی چه بود؟
۳. نتایج آن چه بود؟

Source: Teradata Case Study, "How Barbour Collected More Than 49,000 New Leads in One Month with Teradata Interactive"
http://assets.teradata.com/resourceCenter/downloads/CaseStudies/EB-8791_Interactive-Case-Study_Barbour.pdf

(accessed November 2018).

سوالاتی برای مرور بخش ۷.۸

۱۱. موتور جستجو چیست؟ چرا موتورهای جستجو اهمیت حیاتی برای کسب‌وکارهای امروزی دارند؟
۱۲. خزنده‌ی وب چیست؟ به چه کاری می‌آید؟ چگونه کار می‌کند؟
۱۳. «بهینه‌سازی موتور جستجو» به چه معناست؟ چه کسانی از آن نفع می‌برند؟
۱۴. چه چیزهایی می‌توانند به یک صفحه وب کمک کننده تا در نتایج موتور جستجو در رتبه بالاتری قرار بگیرد؟



شکل ۷.۱۳ استخراج دانش از داده‌های کاربرد وب

الگو به این ترتیب در دسترس قرار می‌گیرد. در این روش، کاربر می‌تواند به راحتی به محتوای مورد نیاز خود دسترسی داشته باشد. این روش به دلیل سادگی و سهولت استفاده، در بسیاری از سیستم‌های وب رایج به کار می‌رود.

۱. بازدید صفحه وب! **از آن برای مشاهده و به‌روزرسانی داده‌ها استفاده می‌کنیم.**

در این روش، کاربر می‌تواند به راحتی به محتوای مورد نیاز خود دسترسی داشته باشد. این روش به دلیل سادگی و سهولت استفاده، در بسیاری از سیستم‌های وب رایج به کار می‌رود.

۲. **زمان ماندن کاربر در سایت.**

این روش به کاربر اجازه می‌دهد تا به راحتی به محتوای مورد نیاز خود دسترسی داشته باشد. این روش به دلیل سادگی و سهولت استفاده، در بسیاری از سیستم‌های وب رایج به کار می‌رود.

۳. **بارگذاری‌ها.**

این روش به کاربر اجازه می‌دهد تا به راحتی به محتوای مورد نیاز خود دسترسی داشته باشد. این روش به دلیل سادگی و سهولت استفاده، در بسیاری از سیستم‌های وب رایج به کار می‌رود.

۴. **نقشه کلیک.**

این روش به کاربر اجازه می‌دهد تا به راحتی به محتوای مورد نیاز خود دسترسی داشته باشد. این روش به دلیل سادگی و سهولت استفاده، در بسیاری از سیستم‌های وب رایج به کار می‌رود.

۵. **مسیرهای کلیک.**

این روش به کاربر اجازه می‌دهد تا به راحتی به محتوای مورد نیاز خود دسترسی داشته باشد. این روش به دلیل سادگی و سهولت استفاده، در بسیاری از سیستم‌های وب رایج به کار می‌رود.

در این مقاله، ما به بررسی روش‌های مختلف برای جذب ترافیک از وبسایت‌های دیگر می‌پردازیم. این روش‌ها شامل تبلیغات، لینک‌سازی، سئو و شبکه‌های اجتماعی است. ما همچنین به شما خواهیم گفت که چگونه می‌توانید این روش‌ها را به نفع خود استفاده کنید.

تبلیغات

تبلیغات یکی از سریع‌ترین و ساده‌ترین روش‌ها برای جذب ترافیک است. شما می‌توانید تبلیغات را در وبسایت‌های دیگر، شبکه‌های اجتماعی و موتورهای جستجو قرار دهید. با تبلیغات، شما می‌توانید به مخاطبان هدف خود دسترسی پیدا کنید و آنها را به وبسایت خود هدایت کنید.

1. وب سایت‌های ارجاع! ¹

وبسایت‌های ارجاع، وبسایت‌هایی هستند که به شما کمک می‌کنند تا ترافیک خود را از وبسایت‌های دیگر جذب کنید. این وبسایت‌ها معمولاً دارای لینک‌های تبلیغاتی هستند که می‌توانید در وبسایت خود قرار دهید. با کلیک بر روی این لینک‌ها، بازدیدکنندگان به وبسایت شما هدایت می‌شوند.

2. موتورهای جستجو

موتورهای جستجو، وبسایت‌هایی هستند که به شما کمک می‌کنند تا ترافیک خود را از موتورهای جستجو جذب کنید. این موتورها معمولاً دارای لینک‌های تبلیغاتی هستند که می‌توانید در وبسایت خود قرار دهید. با کلیک بر روی این لینک‌ها، بازدیدکنندگان به وبسایت شما هدایت می‌شوند.

3. مستقیم

مستقیم، روشی است که به شما کمک می‌کند تا ترافیک خود را مستقیماً از وبسایت‌های دیگر جذب کنید. این روش معمولاً شامل لینک‌های تبلیغاتی در وبسایت‌های دیگر، شبکه‌های اجتماعی و ایمیل‌ها است.

4. کمپین‌های آفلاین

کمپین‌های آفلاین، روشی است که به شما کمک می‌کند تا ترافیک خود را از طریق روش‌های آفلاین جذب کنید. این روش‌ها شامل تبلیغات در روزنامه‌ها، مجله‌ها، تلویزیون و رادیو است. با این روش‌ها، شما می‌توانید به مخاطبان هدف خود دسترسی پیدا کنید و آنها را به وبسایت خود هدایت کنید.

www.mycompany.com/offer50

¹ Referral Web Sites

۵. **کمپین‌های آنلاین.** این کتاب به شما می‌آموزد که چگونه کمپین‌های آنلاین را طراحی و اجرا کنید. این بخش شامل موارد زیر است:

• **هدف‌گذاری:** تعیین اهداف مشخص و قابل اندازه‌گیری برای کمپین شما.

• **شناسایی مخاطب:** شناسایی مخاطبان هدف و نحوه دسترسی به آن‌ها.

• **انتخاب کانال‌ها:** انتخاب کانال‌های تبلیغاتی مناسب برای کمپین شما.

۱. **واژه‌های کلیدی.** این بخش به شما می‌آموزد که چگونه واژه‌های کلیدی مناسب برای کمپین خود را انتخاب کنید. این بخش شامل موارد زیر است:

• **تحقیق واژه‌های کلیدی:** استفاده از ابزارهای تحقیق واژه‌های کلیدی برای شناسایی واژه‌های پربازدید.

• **انتخاب واژه‌های کلیدی:** انتخاب واژه‌های کلیدی مناسب برای کمپین شما.

• **تنظیم کمپین:** تنظیم کمپین تبلیغاتی خود بر اساس واژه‌های کلیدی انتخاب شده.

۲. **گروه‌بندی محتوا.** این بخش به شما می‌آموزد که چگونه محتوا را به گروه‌های مختلف تقسیم کنید. این بخش شامل موارد زیر است:

• **شناسایی موضوعات:** شناسایی موضوعات مرتبط با کمپین شما.

• **گروه‌بندی محتوا:** گروه‌بندی محتوا بر اساس موضوعات شناسایی شده.

• **تولید محتوا:** تولید محتوا برای هر یک از گروه‌های مختلف.

۳. **جغرافیا.** این بخش به شما می‌آموزد که چگونه کمپین خود را به مناطق مختلف جغرافیایی محدود کنید. این بخش شامل موارد زیر است:

• **شناسایی مناطق هدف:** شناسایی مناطق جغرافیایی هدف برای کمپین شما.

• **تنظیم کمپین:** تنظیم کمپین تبلیغاتی خود بر اساس مناطق هدف.

۴. **زمان.** این بخش به شما می‌آموزد که چگونه کمپین خود را در زمان مناسب اجرا کنید. این بخش شامل موارد زیر است:

• **شناسایی زمان‌های پربازدید:** شناسایی زمان‌های پربازدید برای کمپین شما.

• **تنظیم کمپین:** تنظیم کمپین تبلیغاتی خود بر اساس زمان‌های پربازدید.

۵. **پروفایل صفحه هدف.** این بخش به شما می‌آموزد که چگونه صفحه هدف خود را به گونه‌ای طراحی کنید که بتواند مخاطبان را به اقدام مورد نیاز ترغیب کند. این بخش شامل موارد زیر است:

این روش با توجه به اینکه در این روش، هر یک از اعضای هیئت مدیره، به صورت جداگانه و به صورت مستقل، در مورد تصویب یا ردی که در هیئت مدیره صورت می‌گیرد، رأی می‌دهند، به این روش رأی‌دهی «رای‌دهی جداگانه» می‌گویند.

این روش به این ترتیب:

در هیئت مدیره، هر یک از اعضای هیئت مدیره، به صورت جداگانه و به صورت مستقل، در مورد تصویب یا ردی که در هیئت مدیره صورت می‌گیرد، رأی می‌دهند. این روش به این ترتیب:

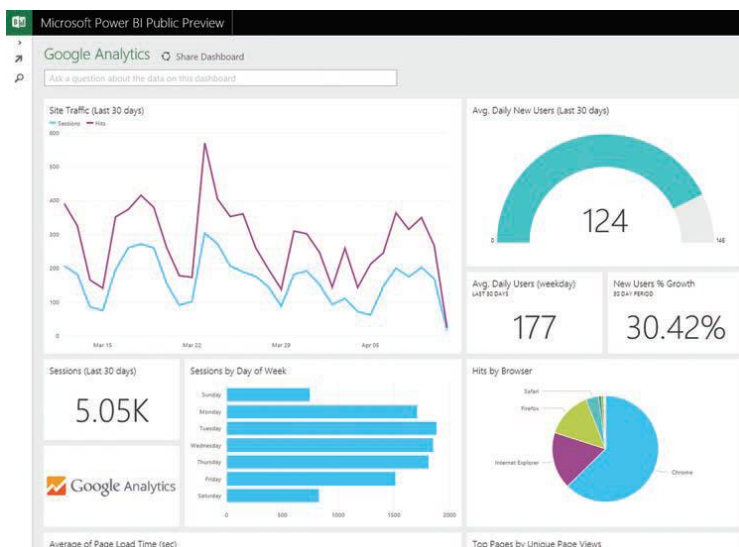
۱. **بازدیدکنندگان جدید.** این روش شامل الزامات زیر است: هر یک از اعضای هیئت مدیره، به صورت جداگانه و به صورت مستقل، در مورد تصویب یا ردی که در هیئت مدیره صورت می‌گیرد، رأی می‌دهند.

۲. **بازدیدکنندگان قبلی.** این روش شامل الزامات زیر است: هر یک از اعضای هیئت مدیره، به صورت جداگانه و به صورت مستقل، در مورد تصویب یا ردی که در هیئت مدیره صورت می‌گیرد، رأی می‌دهند.

۳. **مشتریان راغب.** این روش شامل الزامات زیر است: هر یک از اعضای هیئت مدیره، به صورت جداگانه و به صورت مستقل، در مورد تصویب یا ردی که در هیئت مدیره صورت می‌گیرد، رأی می‌دهند.

۴. **فروش/تبدیل.** این روش شامل الزامات زیر است: هر یک از اعضای هیئت مدیره، به صورت جداگانه و به صورت مستقل، در مورد تصویب یا ردی که در هیئت مدیره صورت می‌گیرد، رأی می‌دهند.

۵. **نرخ ترک/خروج.** این روش شامل الزامات زیر است: هر یک از اعضای هیئت مدیره، به صورت جداگانه و به صورت مستقل، در مورد تصویب یا ردی که در هیئت مدیره صورت می‌گیرد، رأی می‌دهند.



شکل ۷.۱۴ یک نمونه داشبورد تحلیل وب

سوالاتی برای مرور بخش ۷.۹

۱۰. در تحلیل اجتماعی، تعریفی که در فرهنگ لغات برای تحلیل اجتماعی^۱ آورده شده به یک دیدگاه فلسفی اشاره دارد که توسط تاریخ‌شناس و فیلسوف دانمارکی لارس-هنریک اشمیت در دهه ۸۰ بکار گرفته شد. موضوع نظری این دیدگاه، سکویوس^۲ می‌باشد که نوعی از «مشارکت و یاری»^۳ است که نه یک دلیل جهان شمول و نه یک همگانی است که توسط تمامی اعضای یک جامعه به اشتراک گذاشته شده باشد (اشمیت ۱۹۹۶). بنابراین، تحلیل اجتماعی با فلسفه ی رایج و جامعه‌شناسی تفاوت دارد. ممکن است آن را به صورت دیدگاهی دانست که سعی دارد تا مباحثه و مجادلاتی را بین فلسفه و جامعه‌شناسی بیان کند.
۱۱. تعریف ما از تحلیل اجتماعی کمی متفاوت است: برخلاف تمرکز بر جنبه‌ی «اجتماعی» (آنگونه که در تعریف فلسفی آن انجام می‌شود)، ما بیشتر علاقمند به جنبه‌ی «تحلیلی» آن هستیم. گارتتر (یک شرکت بسیار معروف بین‌المللی مشاوره فناوری
۱۲. ...
۱۳. ...

تحلیل‌های اجتماعی برای افراد مختلف و بسته به چشم‌انداز و رشته‌ی مطالعه‌شان می‌تواند معانی متفاوتی داشته باشد. به عنوان مثال، تعریفی که در فرهنگ لغات برای تحلیل اجتماعی^۱ آورده شده به یک دیدگاه فلسفی اشاره دارد که توسط تاریخ‌شناس و فیلسوف دانمارکی لارس-هنریک اشمیت در دهه ۸۰ بکار گرفته شد. موضوع نظری این دیدگاه، سکویوس^۲ می‌باشد که نوعی از «مشارکت و یاری»^۳ است که نه یک دلیل جهان شمول و نه یک همگانی است که توسط تمامی اعضای یک جامعه به اشتراک گذاشته شده باشد (اشمیت ۱۹۹۶). بنابراین، تحلیل اجتماعی با فلسفه ی رایج و جامعه‌شناسی تفاوت دارد. ممکن است آن را به صورت دیدگاهی دانست که سعی دارد تا مباحثه و مجادلاتی را بین فلسفه و جامعه‌شناسی بیان کند.

تعریف ما از تحلیل اجتماعی کمی متفاوت است: برخلاف تمرکز بر جنبه‌ی «اجتماعی» (آنگونه که در تعریف فلسفی آن انجام می‌شود)، ما بیشتر علاقمند به جنبه‌ی «تحلیلی» آن هستیم. گارتتر (یک شرکت بسیار معروف بین‌المللی مشاوره فناوری

¹ Social Analytics

² Socius

³ Commonness

⁴ Schmidt

اطلاعات) تحلیل اجتماعی را به صورت « پایش، تحلیل، سنجش و تفسیر تعاملات و روابط دیجیتالی بین افراد، موضوعات، ایده‌ها و محتوا » تعریف می‌کند (gartner.com/it-glossary/social-analytics/). تحلیل اجتماعی شامل کاوش محتوای متنی ایجاد شده از طریق رسانه‌های اجتماعی (یعنی آنالیز احساسات، ان‌ال‌پی) و تحلیل شبکه‌های اجتماعی (مثلاً شناسایی اینفلوئنسرها، پیش بینی) با هدف دستیابی به بینش در خصوص رفتارهای جاری و آتی مشتریان موجود و بالقوه و در مورد علایق و سلیق آنها در خصوص محصولات و خدمات شرکت است. بر اساس این تعریف و روش‌های جاری، تحلیل‌های اجتماعی را می‌توان به دو شاخه‌ی متفاوت ولی نه لزوماً کاملاً مجزا از یکدیگر تفکیک کرد: تحلیل شبکه اجتماعی و تحلیل رسانه‌های اجتماعی.

یک ساختار اجتماعی متشکل از افراد/مردم (یا گروه‌هایی از افراد یا سازمان‌ها) است که از طریق نوعی روابط/اتصالات به یکدیگر پیوند خورده‌اند. چشم‌انداز شبکه اجتماعی یک رویکرد کل‌نگر را برای تحلیل ساختار و دینامیک‌های موجودیت‌های اجتماعی فراهم می‌کند. در مطالعه‌ی این ساختارها از تحلیل شبکه اجتماعی برای شناسایی الگوهای محلی و جهانی، مکان‌یابی موجودیت‌های تأثیرگذار و ارزیابی دینامیک‌های شبکه استفاده می‌شود. شبکه‌های اجتماعی و تحلیل آنها اساساً یک رشته‌ی بینارشته‌ای است که از رشته‌های روانشناسی اجتماعی، جامعه‌شناسی، آمار و نظریه‌ی گراف‌ها پدیدار شده است. توسعه و رسمی‌سازی گسترده‌ی ریاضی تحلیل شبکه اجتماعی به دهه ۵۰ باز می‌گردد؛ توسعه‌ی نظریه‌ها و روش‌های اساسی شبکه‌های اجتماعی به دهه ۸۰ باز می‌گردد (اسکات آند دیویس ۲۰۰۳). تحلیل شبکه اجتماعی در حال حاضر یکی از پارادایم‌های مهم در تحلیل کسب‌وکار، اطلاعات مصرف‌کننده و روانشناسی امروزی بوده و در چند رشته علمی اجتماعی دیگر نیز کاربرد دارد.

یک شبکه اجتماعی یک ساختار نظری مفید در علوم اجتماعی است که برای مطالعه‌ی روابط بین افراد، گروه‌ها، سازمان‌ها یا حتی کل جوامع (واحدهای اجتماعی) بکار می‌رود. این عبارت برای توصیف یک ساختار اجتماعی که به واسطه‌ی اینگونه تعاملات تعیین شده است بکار می‌رود. پیوندهایی که هر واحد اجتماعی را به دیگر متصل می‌سازد بیانگر همگرایی آحاد مختلف در داخل آن واحد اجتماعی است. به طور کلی، شبکه‌های اجتماعی خودسازمانده، نوظهور و پیچیده هستند به نحوی که یک الگوی منسجم جهانی از تعامل محلی این اجزا پدیدار می‌شود که کل سیستم را تشکیل می‌دهد. در ادامه به معرفی چند نوع شبکه اجتماعی معمول که با فعالیت‌های کسب‌وکار مرتبط هستند می‌پردازیم.

شبکه‌های ارتباطی - مطالعات ارتباطات اغلب به عنوان بخشی از علوم اجتماعی و علوم انسانی شناخته می‌شوند که کاملاً به رشته‌هایی از قبیل جامعه‌شناسی، روانشناسی، انسان‌شناسی، علم اطلاعات، زیست‌شناسی، علوم سیاسی و اقتصاد وابسته و متکی هستند. بسیاری از مفاهیم ارتباطی به توضیح انتقال اطلاعات از یک منبع به منبع دیگر پرداخته و می‌توان آن را به صورت یک شبکه اجتماعی نمایش داد. شرکت‌های مخابراتی از این منبع غنی اطلاعاتی برای بهینه‌سازی روش‌های کسب‌وکار و بهبود ارتباطات با مشتریان استفاده می‌کنند.

¹ Social Network

² Scott & Davis

شبکه‌های اجتماعی - از قدیم، جامعه^۱ به یک مکان جغرافیایی خاص اطلاق می‌شد و مطالعات اجتماعی یا جامعه به افرادی که در فعالیت‌های اجتماعی حضور داشتند و با آنها صحبت، داد و ستد و ... می‌شد سر و کار داشت. اما امروزه جوامع "آنلاین" از طریق ابزارهای شبکه‌بندی اجتماعی و دستگاه‌ها و لوازم مخابراتی بوجود آمده‌اند. اینگونه ابزارها و دستگاه‌ها دائماً در حال تولید حجم انبوهی از داده‌هایی هستند که شرکت‌ها می‌توانند از آنها برای کشف اطلاعات اقدام پذیر و فوق‌العاده ارزشمند استفاده کنند.

شبکه‌های جنایی - در بحث جرم‌شناسی و جامعه‌شناسی شهری^۲، توجه بسیاری به تأثیر شبکه‌های اجتماعی در بین عوامل جنایی شده است. به عنوان مثال، مطالعه‌ی قتل‌ها و جنایت‌های گروهی و سایر فعالیت‌های غیرقانونی به صورت مجموعه‌ای از تبادلات بین گروه‌های تبهکار می‌تواند منجر به درک بهتر و پیشگیری از اینگونه فعالیت‌های جنایی شود. امروزه که ما در یک دنیای کاملاً بهم متصل و یکپارچه زندگی می‌کنیم (به لطف اینترنت)، اکثر شکل‌گیری‌های شبکه‌های جنایی و فعالیت‌های آنها توسط ارگان‌های امنیتی و با استفاده از ابزارها و تاکتیک‌های اینترنتی در حال نظارت و پیگیری است. اگرچه اینترنت چشم‌انداز شبکه‌های جنایی و ارگان‌های انتظامی را تغییر داده است ولی نظریه‌های فلسفی و اجتماعی رایج و قدیمی همچنان به میزان زیادی اعتبار دارند.

شبکه‌های نوآوری - مطالعات کسب‌وکار بر روی انتشار ایده‌ها و نوآوری‌ها در یک محیط شبکه‌ای بر گسترش و استفاده از ایده‌ها و نظرات بین اعضای یک شبکه اجتماعی تمرکز دارند. ایده اصلی در اینجا درک این موضوع است که چرا برخی شبکه‌ها نوآورتر هستند و برخی جوامع به سرعت ایده‌ها و نظرات نوآورانه را بکار می‌گیرند (یعنی ارزیابی تأثیر ساختار شبکه اجتماعی بر تأثیرگذاری و نفوذ گسترش یک رفتار نوآورانه).

تحلیل شبکه اجتماعی یا ارزیابی سیستماتیک شبکه‌های اجتماعی، روابط اجتماعی را بر اساس نظریه‌ی شبکه‌ها متشکل از گره‌ها (بیانگر افراد یا سازمان‌ها در درون شبکه) و ارتباطات بین آنها از قبیل دوستی یا جایگاه سازمانی می‌بیند. این شبکه‌ها اغلب با استفاده از نمودارهای شبکه اجتماعی نمایش داده می‌شوند که گره‌ها به صورت نقاط و ارتباطات به صورت خطوط ترسیم می‌شوند. مورد کاربردی ۷.۸ یک مثال جالب از تحلیل‌های اجتماعی چندکاناله را ارائه می‌کند.

مورد کاربردی ۷.۸

وفاداری برند با استفاده از یک استراتژی اجتماعی مطمئن و درست برای ودکای تیتو

اگر بخواهیم مأموریت ودکای دست‌ساز تیتو را تنها با یک سنجه‌ی رسانه اجتماعی نشان دهیم، این سنجه میزان درگیر شدن مشتری خواهد بود. متصل شدن به دوست‌داران ودکا به روشی مطمئن و درست چیزی است که تیتو آن را کاملاً جدی گرفته و استراتژی اجتماعی برند این شرکت بیانگر این چشم‌انداز می‌باشد.

¹ Community

² Urban Sociology

تیتو که تقریباً دو دهه پیش تأسیس شد برای رسانه‌های اجتماعی ارزش خاصی قائل شده و نقش حیاتی را در مشارکت دادن و درگیرسازی طرفداران و ارتقاء میزان آگاهی از برند ایفا می‌کند. بنیان‌گذار تیتو یعنی برت تیتو بوردیج در مصاحبه‌ای با/ینترپرنور^۱ از نقش رسانه‌های اجتماعی به خاطر این توانمندسازی تیتو در رقابت با برندهای سرشناس تر لیکور قدردانی نمود. او می‌گوید: «رسانه‌های اجتماعی، یک پلتفرم بسیار مهم برای تبلیغات شفاهی برند محسوب می‌شود.»

با افزایش سطح بلوغ برند تیتو، تیم رسانه‌های اجتماعی نسبت به ارزش‌های بنیادین این برند صادق و وفادار مانده و فعالانه از توثیق و اینستاگرام برای برقراری مکالمات یک به یک و اتصال به مشتاقان و علاقمندان به برند استفاده می‌کنند. کتی جلهاوزن، هماهنگ‌کننده وب و رسانه‌های اجتماعی می‌گوید: «ما هیچگاه رسانه‌های اجتماعی را به عنوان روشی جایگزین برای تبلیغات ندیدیم» (رسانه‌های اجتماعی همواره روش اصلی و اول ما برای تبلیغات بوده است). «ما از رسانه‌های اجتماعی استفاده می‌کنیم تا مشتریان با ما صحبت کنند.»

تا بدینجا، تیتو از اسپراوت سوشیال^۲ برای درک فضای صنعت، توسعه‌ی یک برند اجتماعی منسجم و برقراری گفت‌وگو با مخاطبانش استفاده کرده است. در نتیجه، جامعه‌ی مخاطبان تیتو در توثیق و اینستاگرام در عرض چهار ماه به میزان ۴۳.۵ درصد و ۱۲.۶ درصد افزایش یافته است.

اطلاع رسانی یک استراتژی بازاریابی یکپارچه فصلی

طرح کوکتیل فصلی تیتو، بخشی مهم از استراتژی بازاریابی یکپارچه برند آن است. هر فصل، یک دستور تهیه کوکتیل از طریق فعالیت‌های بازاریابی آنلاین و آفلاین تیتو تهیه و توزیع می‌شود.

حصول اطمینان از اینکه دستور تهیه با مرکز توجه برند و همچنین جهت‌گیری صنعت بالادست خود همسو باشد بسیار حائز اهمیت است. از این رو، جلهاوزن از عبارات کلیدی برند اسپراوت برای پایش روندهای صنعت و پروفایل‌های طعم کوکتیل‌ها استفاده می‌کند. وی می‌گوید: «اسپراوت ابزار بسیار مهمی برای پایش رسانه‌های اجتماعی بوده است. صندوق نامه روش مناسبی برای پیگیری هشتک‌ها و مشاهده‌ی روندهای عمومی در یکجا می‌باشد.»

اطلاعات بدست آمده در اختیار تیم متصدی آمیختن مشروبات قرار گرفته و برای حصول اطمینان از اینکه دستور تهیه‌ی یکسانی در اختیار تیم فروش برند و در سرتاسر کانال‌های بازاریابی قرار می‌گیرد اطلاع‌رسانی می‌گردد. جلهاوزن می‌گوید: «فرقی نمی‌کند که شما مشروبات تیتو را در یک میکده می‌نوشید یا آن را از مشروب‌فروشی می‌خرید یا ما را در رسانه‌های اجتماعی دنبال می‌کنید، کوکتیل فصلی یکسانی را دریافت می‌کنید.»

طبق این طرح، شخص در هر نقطه تماس با مصرف‌کننده تجربه‌ی سازگار و یکپارچه‌ای از برند را خواهد داشت و این سازگاری بسیار حیاتی است. در واقع، طبق مطالعه‌ای که توسط اینفوسیس بر روی تجربه خرید تک‌کاناله انجام شده است، ۳۴ درصد اظهار کرده‌اند که سازگاری بین چندین کانال بازاریابی و فروش موجب شده تا پول بیشتری را برای خرید کردن هزینه کنند در حالی که ۳۹ درصد، ناسازگاری را دلیل اصلی برای خرید کمتر عنوان کرده‌اند. در شرکت تیتو، گردآوری بینش در مورد صنعت از طریق پایش اجتماعی توثیق و اینستاگرام با استفاده از پلتفرم اسپراوت آغاز می‌شود. ولی استراتژی اجتماعی برند شرکت به اینجا ختم نمی‌شود. تیتو که به ارزش‌های اصلی و زیربنایی خود وفادار مانده روزانه از این پلتفرم برای برقراری ارتباطات مؤثر و درست با مشتریان خود استفاده می‌کند.

¹ Entrepreneur

² Sprout Social



صندوق نام‌هی هوشمند اسپراوت، حساب‌های توئیتر و اینستاگرام را در قالب یک فید یا خوراک وب منسجم و تکی نمایش می‌دهد. این به جله‌اوزن کمک می‌کند تا پیام‌های ورودی را مدیریت کرده و آنهایی که نیازمند پاسخ هستند را سریعاً شناسایی کند. او می‌گوید: « اسپراوت به ما اجازه می‌دهد تا مکالمات و گفتگوهایی که با دنبال‌کنندگان داریم را پیگیری کنیم. تعامل ساده با محتوای چندین حساب در یکجا کار بسیار لذت‌بخشی است.»

تبلیغات در توئیتر

روش تیتو در توئیتر بسیار ساده اتس: درگیرسازی شخصی و مکالمات یک به یک با طرفداران. گفتمان، عامل محرک هر برند بوده و در یک دوره‌ی زمانی چهار ماهه، ۸۸ درصد از توئیتهای ارسالی، پاسخ‌هایی به پیام‌های ورودی بوده‌اند. استفاده از توئیتر به عنوان یک خط ارتباطی آزاد بین تیتو و طرفدارانش منجر به یک افزایش ۱۶۲.۲ درصدی در مشارکت و یک افزایش ۴۳.۵ درصدی در تعداد دنبال‌کنندگان شد. از استراتژی مشابهی در اینستاگرام نیز استفاده می‌شود که تیتو برای تقویت و سرعت‌بخشی به روابط خود با طرفدارانش به انتشار تصاویر و ویدئوهای رویدادهای جدید، ایده‌ها و نظرات در مورد دستور تهیه‌های جدید و ... می‌پردازد.

در اختیار گرفتن قدرت در اینستاگرام

در اینستاگرام، تیتو عمدتاً به انتشار محتوای سبک‌زندگی پرداخته و دنبال‌کنندگان را ترغیب می‌کند تا از این برند در مناسبت‌های روزمره استفاده کنند. تیتو همچنین از پلتفرم برای ترویج و تبلیغ و انتشار داستان برند خود استفاده می‌کند. تیم تیتو در گزارشات شرح‌حال اینستاگرام اسپراوت به شناسایی ارزش می‌پردازد که به آنها در شناسایی اینکه کدام رسانه بیش از بقیه مشارکت افراد را جلب کرده است، تحلیل جمعیت‌شناختی مخاطبان و رشد آنها، شناسایی عمیق‌تر الگوهای انتشار محتوا و کمی‌سازی عملکرد هشتک‌های **outband** کمک می‌کند.

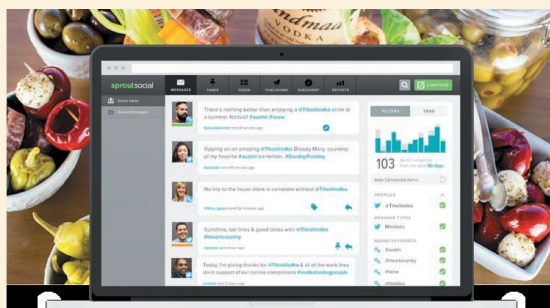
جله‌اوزن می‌گوید: « با توجه به خوراک (فید) شخصی‌سازی شده‌ی جدید اینستاگرام، بسیار مهم است که به آنچه واقعاً طنین انداخته است توجه کنیم.»

تیتو با استفاده از گزارش پروفایل اینستاگرام قادر به اندازه‌گیری تأثیر استراتژی بازاریابی اینستاگرام خود و بازنگری رویکرد بر آن اساس بوده است. برند تیتو با بهره‌گیری از شبکه به عنوان روش دیگری برای درگیرسازی طرفداران، جامعه مخاطبان خود را دائماً توسعه داده است. در عرض چهار ماه، تعداد دنبال‌کنندگان @TitoVodka به میزان ۱۲.۶ درصد و میزان مشارکت به میزان ۳۷.۱

درصد افزایش یافته است. به طور متوسط، هر محتوای منتشر شده دارای ۵۳۴ تراکنش بوده و میزان اشاره به هشتک نام برند #titoshansmadevodka به میزان ۳۳ درصد افزایش یافته است.

از اینجا به کجا می‌رویم؟

پرداختن به رسانه‌های اجتماعی یک سرمایه‌گذاری مستمر در زمان و توجه است. تیتو با ادامه‌ی روند ارائه‌ی کمپین‌های تبلیغاتی فصلی به این نیروی شتاب‌دهنده ادامه می‌دهد. جله‌اوزن می‌گوید: «ما با اتخاذ استراتژی‌های اجتماعی مان همواره هوشمندتر شده و اطمینان حاصل می‌کنیم که محتوای مرتبط و مناسب را منتشر می‌کنیم.» استفاده از رسانه‌های اجتماعی برای برقراری ارتباط مناسب، هوشمندانه و به‌یادماندنی با طرفداران همواره به عنوان رکن اصلی فعالیت‌های بازاریابی دیجیتال تیتو باقی خواهد ماند.



تیتو با استفاده از مجموعه ابزارهای مدیریت رسانه‌های اجتماعی اسپراوت به افزایش وفاداری جامعه‌ی مشتریان خود ادامه خواهد داد. برخی از موفقیت‌های مهم تیتو به شرح زیر می‌باشند:

- افزایش ۱۶۲ درصدی مشارکت در توئیت
- افزایش ۸۱ درصدی تأثیرگذاری در توئیت
- افزایش ۳۷ درصدی مشارکت در اینستاگرام

سوالاتی برای مورد کاربردی ۷.۸

۱. چگونه می‌توان از تحلیل‌های رسانه‌های اجتماعی در صنعت محصولات مصرفی استفاده کرد؟
۲. به نظر شما چالش‌های اصلی، راه‌حل‌های بالقوه و نتایج احتمالی بکارگیری تحلیل‌های رسانه‌های اجتماعی در شرکت‌های خدماتی و محصولات مصرفی چه می‌توانند باشند؟

Source: SproutSocial Case Study, "Tito's Vodka Establishes Brand Loyalty with an Authentic Social Strategy."

<http://sproutsocial.com/insights/case-studies/titos/> (accessed July 2016). Used with permission.

طی سالیان متمادی، سنجه‌های مختلفی برای تحلیل ساختار شبکه‌های اجتماعی از دیدگاه‌های متفاوتی ایجاد شده است. این سنجه‌ها اغلب در قالب سه گروه دسته‌بندی می‌شوند: ارتباطات^۱، توزیع‌ها^۲ و بخش‌بندی^۳.

گروه ارتباطات شامل سنجه‌های زیر هستند:

هوموفیلی: فعالان همسان و ناهمسان تا چه میزان با یکدیگر ارتباط برقرار می‌کنند. میزان همسانی را می‌توان بر حسب جنسیت، نژاد، سن، شغل، سطح تحصیلات، ارزش‌ها، وضعیت یا هر مشخصه‌ی بارز دیگر تعریف کرد.

چندشقی^۴: تعداد شکل‌های محتوایی موجود در یک رابطه. به عنوان مثال، دو نفر که با یکدیگر دوست و همکار هستند دارای چندشقی دوتایی هستند. چندشقی با قدرت رابطه مرتبط است.

رابطه متقابل: به چه میزان دو فعال با یکدیگر رابطه‌ی دوستی یا هرگونه تعامل دیگری را به صورت دوطرفه و متقابل برقرار می‌کنند.

وابستگی شبکه‌ای^۵: یک سنجه که میزان کامل شدن روابط نسبی سه‌گانه را اندازه‌گیری می‌کند. فرض یک نفر از وابستگی شبکه‌ای را حالت گذرا (انتقال‌پذیری) می‌نامند. این حالت، نتیجه و پیامد نیاز وابستگی شناختی فرد یا سازمان است.

مجاورت^۶: تمایل فعالان برای برقراری ارتباطات بیشتر با آنهایی که از لحاظ جغرافیایی در نزدیکی آنها قرار دارند.

موارد زیر مربوط به دسته‌ی توزیع‌ها می‌باشد:

پل^۷: فردی که روابط ضعیف او یک حفره‌ی ساختاری را پر می‌کند و تنها پیوند بین دو فرد یا خوشه است. پل‌ها همچنین شامل کوتاهترین مسیر در زمانی هستند که استفاده از مسیر طولانی‌تر به دلیل ریسک بالای بروز انحراف در پیام یا عدم ارسال پیام امکان‌پذیر نباشد.

¹ Connections

² Distributions

³ Segmentation

⁴ Multiplexity

⁵ Network closure

⁶ Propinquity

⁷ Bridge

مرکزیت! گروهی از سنجه‌ها که هدف آنها کمی‌سازی اهمیت یا نفوذ یک گرهی خاص (یا گروهی از گره‌ها) در یک شبکه است. مثال‌هایی از روش‌های متداول اندازه‌گیری مرکزیت شامل بینابینی مرکزی^۱، نزدیکی مرکزی^۲، بردارویژه مرکزی^۳، آلفای مرکزی و درجه‌ی مرکزی می‌باشند.

چگالی: سهم ارتباطات مستقیم در یک شبکه نسبت به تعداد کل ارتباطات امکان‌پذیر.

فاصله: حداقل تعداد ارتباطات مورد نیاز برای متصل کردن دو فعال خاص.

حفره‌های ساختاری! عدم وجود ارتباطات بین دو بخش از یک شبکه. یافتن و بهره‌برداری از یک حفره‌ی ساختاری می‌تواند مزیت رقابتی برای یک کارآفرین به همراه داشته باشد. این مفهوم توسط رولند بارت جامعه‌شناس مطرح شده و برخی اوقات از آن تحت عنوان مفهوم جایگزین سرمایه اجتماعی یاد می‌شود.

قدرت رابطه! بر اساس ترکیب خطی زمان، شدت عاطفی، صمیمیت و رفت و برگشت تعریف می‌شود. روابط قوی با هوموفیلی، مجاورت و حالت تعدی یا گذرا مرتبط هستند در حالیکه روابط ضعیف با پل‌ها مرتبط هستند.

۱۱-۱۱-۱۱

این دسته شامل موارد زیر است:

گروهک‌ها و حلقه‌های اجتماعی: در صورتیکه هر فرد مستقیماً با تک تک افراد دیگر مرتبط باشد این گروه‌ها را گروهک^۶ می‌نامند ولی اگر تعداد ارتباطات و تماس‌های مستقیم با افراد دیگر کمتر باشد آن را حلقه اجتماعی^۷ می‌نامند که از دقت پایینی برخوردار هستند.

ضریب خوشه‌بندی! یک سنجه‌ی احتمال اینکه دو عضو یک گره با یکدیگر در ارتباط باشند. هرچه ضریب خوشه‌بندی بالاتر باشد به این معناست که احتمال تشکیل گروهک‌ها بالاتر خواهد بود.

چسبندگی یا انسجام! فعالان تا چه اندازه بر اساس پیوندهای منسجم و قوی مستقیماً به یکدیگر متصل شده‌اند. چسبندگی یا انسجام ساختاری به حداقل تعداد اعضای اشاره دارد که اگر از یک گروه حذف شوند کل گروه از هم می‌پاشد.

۱۱-۱۱-۱۱

رسانه‌های اجتماعی به فناوری‌های توانمندساز برای برقراری تعاملات اجتماعی بین افراد اشاره دارد که به دنبال ایجاد، به اشتراک‌گذاری و تبادل اطلاعات، عقاید، نظرات در شبکه‌ها و جوامع مجازی می‌باشند. رسانه‌های اجتماعی، گروهی از برنامه‌های

¹ Centrality

² Betweenness centrality

³ Closeness centrality

⁴ Eigenvector centrality

⁵ Structural Holes

⁶ Tie Strength

⁷ Clique

⁸ Social Circle

⁹ Clustering Coefficient

¹⁰ Cohesion

نرم‌افزاری اینترنتی هستند که بر مبنای ایدئولوژیکی و تکنولوژیکی وب ۲۰۰ ایجاد شده‌اند و امکان خلق و تبادل محتوای تولید شده توسط کاربر را فراهم می‌آورند (کاپلان و هنلین ۲۰۱۰). رسانه‌های اجتماعی برای ایجاد پلتفرم‌های تعاملی سطح بالا چه در سطح فردی و چه عمومی برای به اشتراک‌گذاری، هم‌آفرینی، بحث و اصلاح محتوای تولید شده توسط کاربر به فناوری‌های موبایل و سایر فناوری‌های مبتنی بر وب وابسته هستند.

فناوری‌های رسانه‌های اجتماعی مبتنی بر وب از بدو پیدایش‌شان در اوایل دهه ۹۰ از لحاظ کمی و کیفی بهبودهای بسیاری داشته‌اند. این فناوری‌ها به شکل‌های بسیار متفاوتی از جمله مجله‌های آنلاین، نشست‌های اینترنتی، وبلاگ‌ها، بلاگ‌های اجتماعی، میکرووبلاگ‌نویسی، ویکی‌ها، شبکه‌های اجتماعی، پادکست‌ها، عکس‌ها، ویدئوها و ارزیابی/رتبه‌بندی محصول/خدمت وجود دارد. کاپلان و هنلین با بکارگیری مجموعه‌ای از نظریه‌ها در رشته‌های پژوهش رسانه‌ای (حضور اجتماعی، غنای رسانه‌ای) و فرایندهای اجتماعی (خودابرازی، خوداظهاری) یک الگوی دسته‌بندی را با شش نوع رسانه‌ای اجتماعی مختلف ارائه کردند: پروژه‌های مشارکتی (مثلاً ویکیپدیا)، بلاگ‌ها و میکرووبلاگ‌ها (مثلاً توئیتر)، جوامع محتوایی (مثلاً یوتیوب)، سایت‌های شبکه‌های اجتماعی (مثلاً فیسبوک)، دنیاهای بازی‌های مجازی (مثلاً وُرد آو وارکرفت^۲) و دنیای جوامع مجازی (مثلاً زندگی دوم^۳).

رسانه‌های اجتماعی مبتنی بر وب با رسانه‌های صنعتی/سنتی مثل روزنامه‌ها، تلویزیون و فیلم تفاوت دارند زیرا انتشار یا دسترسی به اطلاعات برای همگان بسیار ارزان بوده و قابل دسترس می‌باشند.

رسانه‌های صنعتی معمولاً برای انتشار اطلاعات به منابع بسیار زیادی نیاز دارند زیرا در اکثر موارد، مقالات یا کتاب‌ها پیش از انتشار تحت بازنگری‌های بسیاری قرار می‌گیرند. در ادامه به معرفی برخی از ویژگی‌های مهم متمایزکننده رسانه‌های اجتماعی و صنعتی می‌پردازیم (مورگان، جونز و هاج ۲۰۱۰).

کیفیت: در نشر صنعتی - که ناشر نقش واسطه را در اینجا دارد- طیف متداول برای کیفیت نشر معمولاً محدودتر از بازارهای نیچ و بی‌واسطه است. چالش اصلی محتوای ارائه شده در سایت‌های رسانه‌های اجتماعی این است که توزیع کیفیت از بسیار باکیفیت تا شدیداً بی‌کیفیت متغیر است.

دامنه دسترسی: فناوری‌های صنعتی و رسانه‌های اجتماعی هر دو مقیاس‌پذیر بوده و به مخاطبان جهانی دسترسی دارند. ولی رسانه‌های صنعتی از یک چارچوب متمرکز برای سازماندهی، تولید و انتشار استفاده می‌کنند در حالیکه رسانه‌های اجتماعی به دلیل ماهیت‌شان غیرمتمرکزتر بوده و دارای چندین نقطه‌ی تولید و بهره‌برداری است.

تواتر: در مقایسه با رسانه‌های صنعتی، به‌روزآوری و ارسال از طریق پلتفرم‌های رسانه‌های اجتماعی آسان‌تر، سریع‌تر و ارزان‌تر است و از این رو بیشتر مورد استفاده قرار می‌گیرند.

دسترسی‌پذیری: تولید در رسانه‌های صنعتی معمولاً از طریق دولت و/یا سازمان‌ها (خصوصی) انجام شده و پرهزینه است در حالیکه ابزارهای رسانه‌های اجتماعی معمولاً با هزینه‌ی بسیار ناچیز در دسترس عموم قرار دارد.

کاربردپذیری: تولید رسانه‌های صنعتی معمولاً نیازمند مهارت‌ها و آموزش‌های تخصصی است. برعکس، اکثر تولیدها توسط رسانه‌های اجتماعی نیازمند مهارت‌های ساده‌ی موجود است و به لحاظ نظری، هر کسی که به رسانه‌های اجتماعی دسترسی داشته باشد قادر به تولید خواهد بود.

¹ Kaplan & Haenlein

² World of Warcraft

³ Second Life

فوریت (بی‌واسطگی): فاصله زمانی بین ارتباطات ایجاد شده توسط رسانه‌های صنعتی در مقایسه با رسانه‌های اجتماعی ممکن است طولانی باشد (هفته‌ها، ماه‌ها یا حتی سال‌ها).

قابلیت به‌روز‌آوری: رسانه‌های صنعتی را پس از ایجاد نمی‌توان تغییر داد (زمانیکه یک مجله چاپ و توزیع می‌شود نمی‌توان همان مجله را تغییر داد) در حالیکه رسانه‌های اجتماعی را بلافاصله با استفاده از اظهارنظرات یا ویرایش می‌توان تغییر داده و اصلاح کرد.

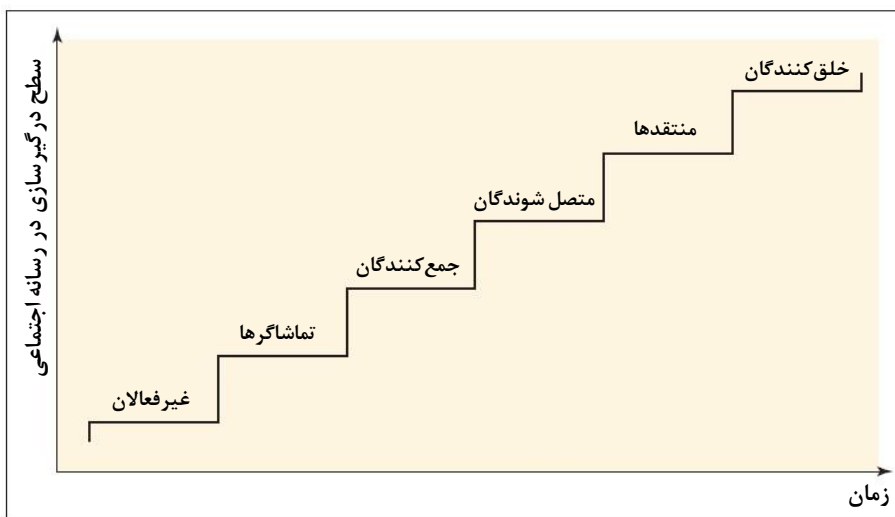
نه تنها تعداد سایت‌های شبکه‌های اجتماعی در حال افزایش است بلکه میزان درگیر شدن آنها با کانال‌ها نیز افزایش یافته است. بروگان و باستون^۱ (۲۰۱۱) نتایج پژوهشی استفاده‌ی کاربران از رسانه‌های اجتماعی را پرزنت کرده و شکل‌گیری و تکامل این بخش از کاربران در طول زمان را پیگیری کرده‌اند. آنها شش سطح درگیرسازی متفاوت را ارائه کرده‌اند (شکل ۷.۱۵). بر اساس نتایج این پژوهش، جامعه‌ی کاربران آنلاین در این ساختار سلسله‌مراتبی درگیرسازی به طور مداوم به سمت بالا حرکت کرده‌اند. قابل توجه‌ترین تغییر در بین غیرفعال‌ها اتفاق افتاده است.

از این جمعیت آنلاین، ۴۴ درصد در سال ۲۰۰۸ در این دسته قرار گرفته‌اند. دو سال بعد، بیش از نیمی از این غیرفعالان به شکل‌های مختلف وارد رسانه‌های اجتماعی شده‌اند. به گفته‌ی باستون: «امروزه تقریباً ۸۲٪ از جمعیت جوان آنلاین در یکی از دسته‌های بالایی قرار دارد». «رسانه‌های اجتماعی حقیقتاً به یک وضعیت استفاده‌ی انبوه رسیده است.» (بروگان و باستون ۲۰۱۱). تحلیل‌های رسانه‌های اجتماعی به روش‌های سیستماتیک و علمی برای استفاده از حجم انبوه محتوای تولید شده توسط ابزارها و تکنیک‌های رسانه‌های اجتماعی مبتنی بر وب برای افزایش رقابت‌پذیری یک سازمان اشاره دارند. تحلیل‌های رسانه‌های اجتماعی به سرعت به یک نیروی جدید و محرک در سازمان‌های سرتاسر جهان تبدیل شده است که به آنها اجازه می‌دهد تا به شکلی که تا پیش از این وجود نداشته است به مصرف‌کنندگان دسترسی یافته و آنها را درک کنند. در بسیاری از شرکت‌ها، تحلیل‌های رسانه‌های اجتماعی به ابزاری برای راهبردهای بازاریابی و ارتباطات یکپارچه بدل شده است.

رشد نمایی آوت‌لت‌های رسانه‌های اجتماعی از بلاگ‌ها، فیسبوک و توئیتر گرفته تا لینکدین و یوتیوب و ابزارهای تحلیلی که از این منابع داده غنی بهره می‌برند به سازمان‌ها این فرصت را می‌دهد تا روزانه با میلیون‌های مصرف‌کننده در سرتاسر جهان وارد گفتگو و مکالمه شوند. به دلیل همین توانایی است که قریب به دو سوم ۲۱۰۰ شرکتی که در نظرسنجی اخیر خدمات تحلیلی هاروارد بیزینس ریویو^۲ شرکت داشته‌اند گفته‌اند که در حال حاضر نه از کانال‌های رسانه‌های اجتماعی استفاده می‌کنند و نه اینکه در کارهایشان طرحی برای استفاده از رسانه‌های اجتماعی داشته‌اند (هاروارد بیزینس ریویو ۲۰۱۰). اما هنوز بسیاری معتقدند که رسانه‌های اجتماعی در فاز آزمایش است و سعی دارند تا بهترین روش استفاده از کانال‌های مختلف را یافته، اثربخشی آنها را سنجیده و تنظیم کرده و رسانه‌های اجتماعی را در راهبردهایشان بکار بگیرند.

¹ Brogan & Bastone

² Harvard Business Review Analytics Service



شکل ۷.۱۵ رشد و تکامل میزان درگیرسازی در رسانه های اجتماعی

برای سازمان‌های بزرگ و کوچک، محتوای تولید شده توسط کاربر در سایت‌های رسانه‌های اجتماعی حاوی بینش‌های پنهان بسیار ارزشمندی است. اما چگونه می‌توانیم از میان ده‌ها سایت رتبه‌بندی، هزاران وبلاگ، میلیون‌ها پست فیسبوک و میلیاردها توییت بهره‌برداری کنیم؟ حتی پس از اینکه اینکار را انجام دادیم چگونه تأثیر تلاش‌هایمان را بسنجیم؟ این پرسش‌ها را به کمک تحلیل‌های فناوری‌های رسانه‌های اجتماعی می‌توانیم پاسخ دهیم. پس از اینکه در مورد هدف‌تان از رسانه‌های اجتماعی تصمیم گرفتید (چه کاری را می‌خواهید انجام دهید)، ابزارهای بسیاری می‌توانند به شما کمک کنند تا به آن هدف برسید. این ابزارهای تحلیلی معمولاً در سه گروه یا دسته‌ی بزرگ قرار می‌گیرند:

- **تحلیل‌های توصیفی:** از آمارهای ساده برای شناسایی ویژگی‌ها و روندهای فعالیت استفاده می‌کنند مثل تعداد دنبال‌کنندگان، تعداد اظهارنظرهای ایجاد شده در فیسبوک و اینکه بیشتر از چه کانال‌هایی استفاده می‌شود.
- **تحلیل شبکه‌های اجتماعی:** پیوندهای بین دوستان، طرفداران و دنبال‌کنندگان را برای شناسایی ارتباطات تأثیرگذار و همچنین بزرگ‌ترین منابع تأثیرگذار دنبال می‌کنند.
- **تحلیل‌های پیشرفته:** شامل تحلیل‌های پیشگویانه و تحلیل متن است که به ارزیابی محتوا در مکالمات آنلاین برای شناسایی الگوها، احساسات، عقاید و ارتباطاتی می‌پردازد که با روش‌های معمول قابل شناسایی نیستند. راه‌حل‌ها و ابزارهای پیچیده‌ی تحلیل رسانه‌های اجتماعی از هر سه گروه تحلیل‌ها (یعنی توصیفی، تجویزی و پیشگویانه) به نحوی پیشرو و مترقی استفاده می‌کنند.

بسیاری از شرکت‌ها از تحلیل‌های رسانه‌های اجتماعی به عنوان ابزاری نوظهور به نحوی اتفاقی استفاده می‌کنند. از آنجائیکه متدولوژی‌های استقرار یافته‌ای وجود ندارد هرکس سعی دارد تا روش خودش را از طریق آزمون و خطا ایجاد کند. در ادامه بهترین تجارب تحلیل‌های رسانه‌های اجتماعی که توسط پین و چاوز (۲۰۱۲) پیشنهاد شده است را ارائه می‌کنیم.

سنجش را یک نظام راهنما ببینید نه یک نظام رتبه‌بندی – از سنجش‌ها اغلب برای تنبیه و تشویق استفاده می‌شود ولی اینگونه نیستند. نظام سنجش به دنبال این است که مؤثرترین ابزارها و روش‌ها، آنچه را که فایده و کاربردی ندارد و آنچه را که باید به دلیل اثربخشی، بیشتر انجام داد را پیدا کند. یک سیستم تحلیلی خوب باید به شما بگوید که تمرکزتان را به کجا معطوف کنید. شاید نیازی نباشد که تمام تمرکز شما بر روی فیسبوک باشد چون مخاطبان شما در فیسبوک نیستند. شاید همه مخاطبان شما در توئیتر باشند یا برعکس. طبق پین و چاوز (۲۰۱۲)، تعیین اولویت در بکارگیری کانال‌ها لزوماً شهودی نیست: «ما با یک هتل همکاری داشتیم که برای یکی از برندهایش تقریباً هیچ فعالیتی در توئیتر نداشت ولی برای یک برند سطح بالای دیگرش فعالیت زیادی در توئیتر داشت. بدون یک ابزار اندازه‌گیری و سنجش دقیق واقعاً نمی‌توانید تشخیص دهد کدام درست است.

احساسات زیرکانه را دنبال کنید – مشتریان دوست دارند تا آنچه را که از مکالمات آنلاین می‌شنوند یاد گرفته و بر اساس آن اقدام کنند. کلید این کار، دقت در استخراج و علامت گذاری قصد و نیت آنها از طریق سنجش احساسات و عقاید آنها است. همانطور که پیشتر در این فصل گفته شد، ابزارهای تحلیلی می‌توانند محتوای آنلاین را دسته بندی کرده، مفاهیم بهم مرتبط را کشف کرده و احساسات را در یک مکالمه بر اساس کلماتی که افراد بکار می‌برند به صورت «مثبت»، «منفی» یا «خنثی» شناسایی کنند. به شکل ایده آل، ما به دنبال این هستیم که یک احساس را به یک محصول، خدمت یا واحد کسب و کار خاص اختصاص دهیم. هرچه دقیق تر بتوانیم ادراک و حالت ابراز و بیان گفتار شاخص را درک کنیم، اطلاعاتی که بدست می‌آورم اقدام پذیرتر خواهد شد زیرا از دغدغه‌های مربوط به دوقطبی شدن پیشگیری می‌کنید. یک عبارت دوقطبی-آمیخته مثل «یک هتل در مکانی فوق‌العاده ولی با دستشویی بدبو» را نباید خنثی دانست زیرا شما نکات مثبت و منفی دارید که یکدیگر را خنثی می‌کنند. این نوع عبارات برای اینکه اقدام‌پذیر باشند باید به صورت جداگانه مورد بررسی قرار بگیرند؛ «دستشویی بدبو» چیزی است که می‌توان بر روی آن کاری انجام داده و آن را بهبود بخشید. می‌توانیم این احساسات و عقاید را طبقه‌بندی و دسته‌بندی کنیم، به روندها در طول زمان نگاه کنیم و تفاوت‌های مهم در روش صحبت کردن افراد چه به صورت مثبت و چه منفی را ببینیم. علاوه بر این، می‌توانیم احساسات و نظرات در مورد برند خودمان را با رقبایمان مقایسه کنیم.

دقت تحلیل متن را به طور مستمر بهبود دهید – یک بسته‌ی نرم‌افزاری تحلیل متن برای یک صنعت خاص دارای فرهنگ واژگان مرتبط با کسب‌وکار شماست. این سیستم دارای قواعد زبانشناسی می‌باشد ولی در طول زمان قادر به یادگیری و بهتر شدن است. همانگونه که برای تنظیم یک مدل آماری، داده‌های بیشتر و پارامترهای بهتر یا تکنیک‌های جدیدی را برای ارائه‌ی نتایج بهتر فراهم می‌کنید همین کار را نیز با این‌ال‌پی در تحلیل احساسات انجام می‌دهید. شما قواعد، طبقه بندی‌ها، دسته بندی‌ها و معنا و مفهوم کلمات را تعیین می‌کنید؛ نتایج را مشاهده کرده و سپس بازگشته و مجدداً اینکار را تکرار می‌کنید.

به اثر موجی توجه کنید - اینکه بتوانید در یک سایت سطح بالا و معروف در صدر باشید فقط شروع کار است. بین یک برند معروف که در صدر قرار دارد و سپس محو می‌شود و یک برند تراز اول که وبلاگ‌نویسان بانفوذ بارها و بارها نام آن را توثیق می‌کنند تفاوت وجود دارد. تحلیل‌ها باید نشان دهنده که کدام فعالیت‌های رسانه‌های اجتماعی «ویروسی» خواهند شد و به چه علت.

فراتر از برندگان را ببینید - یکی از بزرگ‌ترین اشتباهاتی که افراد مرتکب می‌شوند این است که فقط به برند خودشان فکر می‌کنند. برای تحلیل و اقدام موفق در خصوص رسانه‌های اجتماعی نه تنها باید آنچه در مورد برند گفته می‌شود را درک کنیم بلکه باید به مکالمات و گفتگوهای گسترده‌تری که در خصوص مسائل حول محصولات و خدمات انجام می‌شود نیز توجه کنیم. مشتریان معمولاً به پیام یا برند یک سازمان توجه نمی‌کنند بلکه به خودشان توجه می‌کنند. بنابراین، شما باید به آنچه مشتریان در موردش صحبت می‌کنند، جاییکه صحبت می‌کنند و اینکه منافع آنها در چه چیزی است توجه کنید.

قدرتمندترین اینفلوئنسرهایتان را شناسایی کنید - سازمان‌ها برای شناسایی افرادی که بیشترین قدرت را در شکل دادن به افکار عمومی دارند در تلاش هستند. اینگونه به نظر می‌رسد که مهمترین افراد تأثیرگذار لزوماً آنهایی نیستند که به طور خاص طرفدار و حامی برند آنها هستند بلکه آنهایی هستند که بر قلمروی مکالمات و گفتگوها در خصوص موضوع مورد بحث آنها نفوذ دارند. سازمان‌ها باید درک کنند که آیا اینفلوئنسرهای آنها حرف‌های خوبی می‌زنند، حمایت خود را ابراز می‌کنند یا فقط به مشاهده یا انتقاد می‌پردازند. ماهیت مکالمات آنها چیست؟ موقعیت برند سازمان شما نسبت به رقبایتان در آن فضای کسب‌وکار چگونه است؟

نگاه نزدیک‌تری به دقت ابزارهای تحلیلی بکار رفته داشته باشید - تا همین اواخر، دقت ابزارهای خودکار رایانه‌ای برای واریسی و غربال کردن محتوای آنلاین به اندازه‌ی انسان نبود. حتی امروزه نیز دقت آنها بسته به رسانه‌های بکار رفته متغیر است. در خصوص سایت‌های اظهارنظر و رتبه‌بندی محصولات، سایت‌های رتبه‌بندی هتل‌ها و توئیتر، دقت می‌تواند بین ۸۰ تا ۹۰ درصد باشد زیرا در این خصوص بسیار کار شده است. زمانیکه سازمانی شروع به نگاه کردن به بلاگ‌ها و فروم‌های بحث و گفتگو می‌کند که موضوع مکالمات در آنها بسیار گسترده است، دقت نرم‌افزار بین ۶۰ تا ۷۰ درصد خواهد بود (پین و چاوز ۲۰۱۲). این ارقام به دلیل به‌روزرسانی مستمر ابزارهای تحلیلی با قواعد جدید و الگوریتم‌های بهبود یافته برای محصولات جدید، شرایط متغیر بازار و الگوهای گفتاری نوظهور افزایش خواهد یافت.

از اطلاعات رسانه‌های اجتماعی در برنامه‌ریزی استفاده کنید - زمانیکه سازمان به یک چشم‌انداز تصویر بزرگ و بینش مشروح دست یافت می‌تواند شروع به بکارگیری این اطلاعات در چرخه‌ی برنامه‌ریزی خود کند. اما گفتن این حرف آسان‌تر از انجام دادنش است. یک نظرسنجی سریع از مخاطبان نشان داده است که تعداد افراد معدودی هستند که در حال حاضر از مکالمات آنلاین خود در چرخه برنامه‌ریزی‌شان استفاده می‌کنند (پین و چاوز ۲۰۱۲). یک روش برای اینکار این است که موقعیت‌های بهم مرتبط از لحاظ زمانی را بین سنج‌های رسانه‌های اجتماعی و سایر فعالیت‌های کسب‌وکار یا رویدادهای بازار پیدا کنیم. رسانه‌های اجتماعی یا به صورت ارگانی یا توسط کاری یک سازمان انجام می‌دهد مورد استناد قرار می‌گیرند و از این رو؛ در صورتیکه نوسان شدیدی در یک مقطع زمانی در فعالیتی مشاهده شود، می‌خواهد ببیند که چه اتفاقی در پس آن افتاده است.

سوالاتی برای مرور بخش ۷.۱۰

۱۰. منظور از تحلیل‌های اجتماعی چیست؟ چرا موضوع مهمی برای کسب‌وکارها محسوب می‌شود؟
۱۱. منظور از یک شبکه‌ی اجتماعی چیست؟ چه نیازی به تحلیل شبکه‌های اجتماعی داریم؟
۱۲. رسانه‌های اجتماعی چه هستند؟ چه ارتباطی با وب ۲.۰ دارند؟
۱۳. تحلیل‌های رسانه‌های اجتماعی چه هستند؟ دلایل افزایش محبوبیت آنها چیست؟
۱۴. چگونه می‌توانید تأثیر تحلیل‌های رسانه‌های اجتماعی را اندازه‌گیری کنید؟

- در این بخش، شما با مفاهیم اساسی تحلیل‌های اجتماعی آشنا خواهید شد. این شامل بررسی انواع مختلف شبکه‌های اجتماعی و نحوه عملکرد آنها می‌گردد. همچنین، شما یاد خواهید گرفت که چگونه می‌توانید از این ابزارها برای اهداف بازاریابی خود استفاده کنید.
- در این بخش، شما با مفاهیم اساسی تحلیل‌های اجتماعی آشنا خواهید شد. این شامل بررسی انواع مختلف شبکه‌های اجتماعی و نحوه عملکرد آنها می‌گردد. همچنین، شما یاد خواهید گرفت که چگونه می‌توانید از این ابزارها برای اهداف بازاریابی خود استفاده کنید.
- در این بخش، شما با مفاهیم اساسی تحلیل‌های اجتماعی آشنا خواهید شد. این شامل بررسی انواع مختلف شبکه‌های اجتماعی و نحوه عملکرد آنها می‌گردد. همچنین، شما یاد خواهید گرفت که چگونه می‌توانید از این ابزارها برای اهداف بازاریابی خود استفاده کنید.
- در این بخش، شما با مفاهیم اساسی تحلیل‌های اجتماعی آشنا خواهید شد. این شامل بررسی انواع مختلف شبکه‌های اجتماعی و نحوه عملکرد آنها می‌گردد. همچنین، شما یاد خواهید گرفت که چگونه می‌توانید از این ابزارها برای اهداف بازاریابی خود استفاده کنید.
- در این بخش، شما با مفاهیم اساسی تحلیل‌های اجتماعی آشنا خواهید شد. این شامل بررسی انواع مختلف شبکه‌های اجتماعی و نحوه عملکرد آنها می‌گردد. همچنین، شما یاد خواهید گرفت که چگونه می‌توانید از این ابزارها برای اهداف بازاریابی خود استفاده کنید.
- در این بخش، شما با مفاهیم اساسی تحلیل‌های اجتماعی آشنا خواهید شد. این شامل بررسی انواع مختلف شبکه‌های اجتماعی و نحوه عملکرد آنها می‌گردد. همچنین، شما یاد خواهید گرفت که چگونه می‌توانید از این ابزارها برای اهداف بازاریابی خود استفاده کنید.
- در این بخش، شما با مفاهیم اساسی تحلیل‌های اجتماعی آشنا خواهید شد. این شامل بررسی انواع مختلف شبکه‌های اجتماعی و نحوه عملکرد آنها می‌گردد. همچنین، شما یاد خواهید گرفت که چگونه می‌توانید از این ابزارها برای اهداف بازاریابی خود استفاده کنید.
- در این بخش، شما با مفاهیم اساسی تحلیل‌های اجتماعی آشنا خواهید شد. این شامل بررسی انواع مختلف شبکه‌های اجتماعی و نحوه عملکرد آنها می‌گردد. همچنین، شما یاد خواهید گرفت که چگونه می‌توانید از این ابزارها برای اهداف بازاریابی خود استفاده کنید.
- در این بخش، شما با مفاهیم اساسی تحلیل‌های اجتماعی آشنا خواهید شد. این شامل بررسی انواع مختلف شبکه‌های اجتماعی و نحوه عملکرد آنها می‌گردد. همچنین، شما یاد خواهید گرفت که چگونه می‌توانید از این ابزارها برای اهداف بازاریابی خود استفاده کنید.
- در این بخش، شما با مفاهیم اساسی تحلیل‌های اجتماعی آشنا خواهید شد. این شامل بررسی انواع مختلف شبکه‌های اجتماعی و نحوه عملکرد آنها می‌گردد. همچنین، شما یاد خواهید گرفت که چگونه می‌توانید از این ابزارها برای اهداف بازاریابی خود استفاده کنید.
- صدای مشتری جزو لاینفک سیستم‌های مدیریت تجربه مشتری و مدیریت ارتباط با مشتری بوده و اغلب از تحلیل احساسات در آن استفاده می‌شود.
- VOM با درک عقاید و نظرات و روندهای تجمیعی در سطح بازار سر و کار دارد.
- برای تشخیص دوقطبی بودن در تحلیل احساسات یا از یک واژه‌نامه به عنوان یک کتابخانه‌ی مرجع یا مجموعه‌ای از اسناد آموزشی استفاده می‌شود.
- وُردنِت یک واژه‌نامه‌ی عمومی محبوب است که توسط دانشگاه پرینستون ایجاد شده است.
- سِنِتی وُردنِت یک انشعاب از وُردنِت است که برای شناسایی احساسات بکار می‌رود.
- تحلیل گفتار یک رشته‌ی علمی در حال رشد و گسترش است که به کاربران اجازه می‌دهد تا اطلاعات را از مکالمات زنده و ضبط شده تحلیل و استخراج کنند.
- وب‌کاوی را می‌توان کشف و تحلیل اطلاعات جالب و مفید از وب، در مورد وب و معمولاً با استفاده از ابزارهای مبتنی بر وب تعریف کرد.
- وب‌کاوی را می‌توان متشکل از سه حوزه دانست: محتواکاوی، ساختارکاوی و کاربردکاوی.

	نام خانوادگی	واژگان پالایشی (ایستاده شده یا تصفیه‌شونده)	ردیف
--	--------------	---	------

			۹۳
			۹۴
			۹۵
			۹۶
			۹۷
			۹۸
			۹۹
			۱۰۰
			۱۰۱
			۱۰۲
			۱۰۳
			۱۰۴
			۱۰۵
			۱۰۶
			۱۰۷
			۱۰۸
			۱۰۹
			۱۱۰

	teradatauniversitynetwork.com	.15
	teradatauniversitynetwork.com	.16
eBay	teradatauniversitynetwork.com	.17
	Analytics	
	teradatauniversitynetwork.com	.18
	How Do We Fix an App Like That	
	teradatauniversitynetwork.com	.19
		.20

		.1
		.2
		.3
		.4
	Text Analytics	

...
...
... .۵

...
...

... .۱
...
...

... Downloads statsoft.com .۲

... sas.com .۳

... ibm.com .۴

... teradata.com .۵

... clarabridge.com .۶

... kdnuggets.com .۷

... .۸

... attensity.com .۹



تحلیل‌های تجویزی: بهینه‌سازی و شبیه‌سازی

- در این فصل به بررسی کاربردهای تحلیل‌ها فراتر از گزارش‌گیری و تحلیل‌های پیشگویانه خواهیم پرداخت. این فصل شامل تکنیک‌های منتخبی است که می‌توان به همراه مدل‌های پیشگویانه برای کمک به پشتیبانی تصمیمات بکار گرفت. تمرکز ما بر تکنیک‌هایی است که بتوان آنها را به آسانی یا با استفاده از ابزارهای صفحه گسترده یا با استفاده از ابزارهای نرم‌افزاری مستقل پیاده‌سازی کرد. البته، آشنایی با مدل‌های علم مدیریت نیازمند یادگیری جزئیات بسیار زیادی است ولی هدف ما در این فضا صرفاً این است که نشان دهیم در شرایط واقعی چه چیزی امکان‌پذیر است و چگونه پیاده‌سازی و اجرا می‌شود. مطالبی که در این فصل مطرح می‌کنیم با یک احتیاط همراه است: مدلسازی می‌تواند موضوع دشواری باشد و بیش از آنکه یک هنر باشد یک علم محسوب می‌شود. هدف از این فصل تنها این نیست که شما در موضوعات مدلسازی و تحلیل به استادی برسید بلکه هدف در اینجا آشنایی با مفاهیم مهمی است که با تحلیل تجویزی و کاربرد آنها در تصمیم‌گیری مرتبط است. بسیار مهم است توجه داشته باشید که مدلسازی مورد نظر ما در این فصل با مدلسازی داده‌ها مرتبط است. شما نباید این دو مفهوم را با یکدیگر اشتباه بگیرید. ما ابتدا به ارائه تعاریف و مفاهیم پایه‌ی مدلسازی تصمیم پرداخته و سپس به معرفی مستقیم ایده‌ی بکارگیری مدلسازی در تصمیم‌گیری می‌پردازیم. سپس در مورد ساختار و کاربردهای دو متدولوژی و مدل موفق خواهیم پرداخت: برنامه ریزی خطی و شبیه‌سازی رویداد گسسته. این فصل از بخش‌های زیر تشکیل شده است:

در این فصل به بررسی کاربردهای تحلیل‌ها فراتر از گزارش‌گیری و تحلیل‌های پیشگویانه خواهیم پرداخت. این فصل شامل تکنیک‌های منتخبی است که می‌توان به همراه مدل‌های پیشگویانه برای کمک به پشتیبانی تصمیمات بکار گرفت. تمرکز ما بر تکنیک‌هایی است که بتوان آنها را به آسانی یا با استفاده از ابزارهای صفحه گسترده یا با استفاده از ابزارهای نرم‌افزاری مستقل پیاده‌سازی کرد. البته، آشنایی با مدل‌های علم مدیریت نیازمند یادگیری جزئیات بسیار زیادی است ولی هدف ما در این فضا صرفاً این است که نشان دهیم در شرایط واقعی چه چیزی امکان‌پذیر است و چگونه پیاده‌سازی و اجرا می‌شود. مطالبی که در این فصل مطرح می‌کنیم با یک احتیاط همراه است: مدلسازی می‌تواند موضوع دشواری باشد و بیش از آنکه یک هنر باشد یک علم محسوب می‌شود. هدف از این فصل تنها این نیست که شما در موضوعات مدلسازی و تحلیل به استادی برسید بلکه هدف در اینجا آشنایی با مفاهیم مهمی است که با تحلیل تجویزی و کاربرد آنها در تصمیم‌گیری مرتبط است. بسیار مهم است توجه داشته باشید که مدلسازی مورد نظر ما در این فصل با مدلسازی داده‌ها مرتبط است. شما نباید این دو مفهوم را با یکدیگر اشتباه بگیرید. ما ابتدا به ارائه تعاریف و مفاهیم پایه‌ی مدلسازی تصمیم پرداخته و سپس به معرفی مستقیم ایده‌ی بکارگیری مدلسازی در تصمیم‌گیری می‌پردازیم. سپس در مورد ساختار و کاربردهای دو متدولوژی و مدل موفق خواهیم پرداخت: برنامه ریزی خطی و شبیه‌سازی رویداد گسسته. این فصل از بخش‌های زیر تشکیل شده است:

۸.۱. پرونده‌ی آغازین: بخش آموزش و پرورشی فیلادلفیا از تحلیل تجویزی برای یافتن راه‌حل بهینه برای عقد قراردادهای اتوبوسرانی استفاده می‌کند

۸.۲. تصمیم‌گیری بر اساس مدل

۸.۳. ساختار مدل‌های ریاضی برای پشتیبانی تصمیم

۸.۴. قطعیت، عدم قطعیت، ریسک

۸.۵. مدل‌سازی تصمیم با صفحات گسترده

۸.۶. بهینه‌سازی برنامه‌ریزی ریاضی

۸.۷. اهداف چندگانه، تحلیل حساسیت، تحلیل شرطی و جستجوی هدف

۸.۸. تحلیل تصمیم با جداول تصمیم و درخت تصمیم

۸.۹. مقدمه‌ای بر شبیه‌سازی

۸.۱۰. شبیه‌سازی تعاملی دیداری

انتخاب بهترین تأمین‌کننده یا پیمانکار برای همکاری دشوارترین و در عین حال مهمترین کار برای شرکت‌ها و سازمان‌های دولتی می‌باشد. پس از اینکه یک پیمانکار پروپوزال خود را برای یک کار مشخص از طریق فرایند مزایده ارسال می‌کند، شرکت یا سازمان به ارزیابی پروپوزال پرداخته و بهترین پیمانکار را متناسب با نیازهای خود انتخاب می‌کند. معمولاً سازمان‌های دولتی باید از یک فرایند مزایده برای انتخاب یک یا چند پیمانکار استفاده کنند. بخش آموزش و پرورشی فیلادلفیا به دنبال برون‌سپاری بخشی از سرویس‌های مدارس خود به اتوبوسرانی بخش خصوصی بود. این بخش، تعداد معدودی اتوبوس در اختیار داشت ولی برای خدمت‌رسانی به دانش‌آموزان به تعداد بیشتری نیاز داشت. آنها می‌خواستند از اتوبوس‌های خودشان برای ۳۰ تا ۴۰ درصد از مسیرها استفاده کرده و مابقی مسیرها را به این بخش خصوصی واگذار کنند. چارلز لوویتز که هماهنگ‌کننده‌ی مالی اداره‌ی حمل و نقل بود مأموریت یافت تا تعیین کند که چگونه به بیشترین نرخ بازگشت سرمایه دست پیدا کنند و نحوه‌ی تخصیص مسیرها به تأمین‌کنندگان مختلف را مشخص کنند. از لحاظ تاریخی، فرایند تصمیم‌گیری در خصوص عقد قراردادهای اتوبوسرانی با توجه به محدودیت‌های بودجه و زمان کاری بسیار طاقت‌فرسا بود که عمدتاً به صورت دستی انجام میشد. علاوه بر این، عوامل و متغیرهای مختلفی که باید در نظر گرفته می‌شدند به پیچیدگی مسئله می‌افزودند. پیمانکارها بر اساس پنج متغیر مورد ارزیابی قرار می‌گرفتند: هزینه، قابلیت‌ها، اتکا، پایداری مالی و هوشمندی تجاری. هر پیمانکار یک پروپوزال را با قیمتی متفاوت برای مسیرهای متفاوت ارسال می‌کند. برخی پیمانکارها حداقل تعداد مسیر را مشخص کرده‌اند که اگر این حداقل تعداد برآورده نشود، هزینه‌ی آنها افزایش خواهد یافت. لوویتز باید متوجه می‌شد که چگونه اطلاعات اخذ شده از هر پروپوزال را ترکیب کند تا تعیین کند که کدام مسیر را به کدام پیمانکار تخصیص دهد تا کلیه‌ی الزامات مسیرها با کمترین هزینه برای آن ناحیه برآورده شود.

لوویتر ابتدا به دنبال نرم‌افزاری بود که بتواند از آن به همراه مدل قراردادش در نرم‌افزار اکسل استفاده کند. او شروع به استفاده از پرمیوم سالور پلتفرم^۱ از شرکت فرانتلاین سیستمز^۲ نمود که می‌توانست سودمندترین پیمانکارها را از نقطه نظر مالی و عملیاتی برای آن منطقه پیدا کند. او یک مدل بهینه‌سازی را ایجاد کرد که متغیرهای ذکر شده برای هر پیمانکار را در نظر می‌گرفت. این مدل شامل متغیرهای عدد صحیح دودویی (بله/خیر) برای هر مسیری که قرار بود به شرکت‌کنندگان در مزایده واگذار شود بود. این مدل شامل حدوداً ۱۶۰۰ متغیر دودویی بود. این مدل همچنین شامل محدودیت‌هایی بود که هر مسیر به یک پیمانکار تخصیص داده میشد و البته هر مسیری که باید سرویس داده شود. محدودیت‌های دیگر شامل حداقل تعداد مسیرهایی است که یک پیمانکار قبول می‌کند می‌باشد. تمامی این محدودیت‌ها را می‌توان به صورت معادلات نوشته و در یک مدل برنامه‌ریزی خطی عدد صحیح وارد نمود. اینگونه مدل‌ها را می‌توان توسط ابزارهای نرم‌افزاری متعددی فرموله و حل نمود ولی درک این مدل با استفاده از نرم‌افزار مایکروسافت اکسل راحت‌تر است. نرم‌افزار سالور فرانتلاین سیستمز در داخل مایکروسافت اکسل و برای حل مسائل کوچکتر به صورت رایگان تعبیه شده است. نسخه‌ی بزرگتر آن را می‌توان برای حل مدل‌های بزرگتر و پیچیده‌تر خریداری کرد که لوویتر از آن استفاده کرد.

نتیجه‌گیری

علاوه بر تعیین تعداد پیمانکارانی که می‌بایست با آنها قرارداد بست، این مدل به تعیین اندازه و حجم هر قرارداد نیز کمک کرد. اندازه و حجم هر قرارداد بسته به پیمانکاری که چهار مسیر را انتخاب کرده تا پیمانکاری که ۹۷ مسیر را انتخاب کرده متغیر است. نهایتاً، بخش آموزش و پرورشی فیلادلفیا توانست طرحی را با تعداد بهینه‌ی پیمانکاران اتوبوسرانی و با استفاده از نرم‌افزار اکسل به جای یک فرایند دست‌نویس تهیه کند. بخش آموزش و پرورش با استفاده از ابزارهای تحلیلی پرمیوم سالور پلتفرم برای ایجاد یک مدل بهینه‌سازی با متغیرهای مختلف توانست در زمان و هزینه صرفه‌جویی کند.

سوالاتی برای مرور پرونده‌ی آغازین

۳۳. در این پرونده چه تصمیمی قرار بود گرفته شود؟

۳۴. برای اتخاذ بهترین تصمیم در این سناریو از چه نوع داده‌هایی (توصیفی و پیشگویانه) ممکن است استفاده شود؟

۳۵. برای عقد قرارداد برای اینگونه مسیرها چه هزینه‌ها یا محدودیت‌های دیگری را باید در نظر گرفت؟

۳۶. این اینگونه مدل‌ها در چه موقعیت‌های دیگری می‌توان استفاده کرد؟

اکثر سازمان‌ها با مسئله‌ی تصمیم‌گیری و انتخاب یک گزینه از بین چندین گزینه مواجه هستند. هر گزینه، هزینه‌ها و قابلیت‌های خاص خودش را دارد. هدف از تعریف و بکارگیری اینگونه مدل‌ها انتخاب یک ترکیب بهینه است که کلیه الزامات را برآورده کرده و در عین حال هزینه‌ها را بهینه نماید. تحلیل‌های تجویزی به طور خاص برای اینگونه مسائل تصمیم‌گیری کاربرد دارند. و اینگونه ابزارها مثل پرمیوم سالور برای اکسل بکارگیری این تکنیک‌ها را آسان‌تر می‌کند.

¹ Premium Solver Platform

² Frontline Systems

همانطور که در پرونده‌ی آغازین دیدیم، تصمیم‌گیری با استفاده از یک مدل تحلیلی را تحلیل تجویزی می‌نامیم. در چند فصل گذشته در خصوص ارزش و فرایند اطلاع از تاریخچه‌ی آنچه رخ داده است و بکارگیری آن اطلاعات برای پیش‌بینی آنچه احتمالاً رخ خواهد داد آموختیم. با وجود این، باید تعیین کنیم که قدم بعدی ما چیست. این تصمیم ممکن است شامل تعیین این باشد که کدام مشتریان به احتمال از ما خرید خواهند کرد و قیمتی را تعیین کنیم که این احتمال را به حداکثر برساند و به سود بهینه دست پیدا کنیم. برعکس، ممکن است تصمیم شامل این باشد که پیش‌بینی کنیم کدام مشتریان احتمالاً به جای دیگری خواهند رفت یک پیشنهاد تخفیف برایشان در نظر بگیریم که مشتری را حفظ کرده و در عین حال ارزش کسب شده را بهینه کنیم. ممکن است لازم باشد برای اعطای قرارداد به پیمانکاران تصمیم‌گیری کنیم تا مطمئن شویم همه نیازهای ما تامین شده و هزینه‌ها به حداقل می‌رسد. ما ممکن است در شرایطی قرار بگیریم که تصمیم بگیریم چه مشتریان احتمالی باید چه نوع تبلیغاتی را دریافت کنند تا هزینه تبلیغات ما مبهوت کننده نباشد و در حین مدیریت بودجه، میزان پاسخگویی را به حداکثر برسانیم. ممکن است تصمیم بگیریم که چقدر برای کلمات کلیدی جستجو هزینه پرداخت کنیم تا بازده سرمایه‌گذاری بودجه تبلیغاتی ما حداکثر شود. در شرایط دیگر، ممکن است مجبور شویم تاریخچه الگوهای ورود مشتریان خود را مطالعه کرده و از آن اطلاعات برای پیش‌بینی نرخ ورود آینده استفاده کنیم و از آن برای برنامه‌ریزی تعداد مناسب کارکنان فروشگاه برای حداکثر کردن پاسخ به مشتری و بهینه‌سازی هزینه‌های نیروی انسانی خود استفاده کنیم. ما بر اساس تجزیه و تحلیل و پیش‌بینی تقاضا برای محصولات و هزینه‌های زنجیره تأمین، می‌توانیم محل انبارهای خود را تعیین کنیم. ما می‌توانیم مسیرهای تحویل روزانه را بر اساس حجم محصول که در مکان‌های مختلف تحویل داده می‌شود تعیین کنیم هزینه‌های تحویل و در دسترس بودن خودرو می‌توان صدها نمونه از موقعیت‌هایی را یافت که تصمیمات مبتنی بر داده با ارزش هستند. در واقع، بزرگترین فرصت برای حرفه‌ی در حال رشد تجزیه و تحلیل، توانایی استفاده از بینش‌های توصیفی و پیشگویانه برای کمک به تصمیم‌گیرنده در اخذ تصمیمات بهتری می‌باشد اگرچه شرایطی وجود دارد که می‌توان از تجربه و شهود برای تصمیم‌گیری استفاده کرد، اما به احتمال زیاد تصمیمی که به پشتیبانی یک مدل اتخاذ می‌شود به تصمیم‌گیرنده در تصمیم‌گیری بهتر کمک می‌کند. بنابراین تجزیه و تحلیل تجویزی به عنوان خط مقدم بعدی در حوزه‌ی تجزیه و تحلیل مطرح شده است. این اساساً شامل استفاده از یک مدل تحلیلی برای هدایت و راهنمایی تصمیم‌گیرنده در تصمیم‌گیری یا خودکارسازی آن است به گونه‌ای که یک مدل بتواند توصیه کرده یا تصمیم‌گیری کند. از آنجا که تمرکز تجزیه و تحلیل تجویزی بر ارائه توصیه‌ها یا تصمیم‌گیری است برخی این دسته از تجزیه و تحلیل تصمیم‌ها را *تحلیل تصمیم* می‌نامند.

نشریات اینفورمز مثل *اینترفیس*^۱، *آرام/اس تودی*^۲ و *مجله‌ی آنالیتیکز*^۳ همگی حاوی داستان‌هایی از کاربردهای موفق مدل‌های تصمیم‌گیری در محیط‌های دنیای واقعی هستند. این فصل شامل نمونه‌های بسیاری از اینگونه کاربردهای تحلیل تجویزی است. استفاده از مدل‌ها در شرایط واقعی می‌تواند میلیون‌ها دلار صرفه‌جویی یا میلیون‌ها دلار درآمد به همراه داشته باشد. کریستینسن

¹ Interface

² ORMS Today

³ Analytics

و همکاران^۱ (۲۰۰۹) کاربردهای این مدلها را در عملیات شرکت حمل و نقل با استفاده از توربو روتر^۲ که یک سیستم پشتیبانی تصمیم برای مسیریابی و برنامه‌ریزی کشتی‌ها است توضیح داده‌اند. آنها مدعی هستند که در طول یک دوره ۳ هفته‌ای، یک شرکت از این مدل برای استفاده بهتر از ناوگان خود استفاده کرده و در چنین بازه زمانی کوتاهی ۱-۲ میلیون دلار سود بیشتر کسب کرده است. ما یک مثال دیگر از کاربرد مدل‌ها در ورزش را در مورد کاربردی ۸.۱ ارائه می‌دهیم.

مورد کاربردی ۸.۱

لیگ فوتبال کانادایی، جداول زمانی بازی‌ها را بهینه‌سازی می‌کند

لیگ فوتبال کانادایی در واقع معادل لیگ ملی فوتبال ایالات متحده است. این لیگ با چالش بزرگی مواجه است که سازماندهی ۸۱ بازی فوتبال برای ۹ تیم در یک بازه زمانی ۵ ماهه است که در عین حال نیازمند این است که اولویت‌های درآمد حاصل از فروش، رتبه‌بندی و امتیازدهی تلویزیونی و روزهای استراحت تیم‌ها را نیز در نظر گرفت. از ملاحظات دیگر می‌توان به سازماندهی تیم‌ها در مناطق زمانی متفاوت (اختلاف زمانی) و همچنین برگزاری بازی‌های مهم در روزهای تعطیل عمومی اشاره کرد. هر لیگ برای همکاری‌های مختلف خود از قبیل هماهنگی با کانال‌های پخش تلویزیونی و سازماندهی فروش بلیت بازی به یک جدول زمانبندی قرص و محکم نیاز دارد. اگر این جدول بهینه‌سازی نشده باشد مستقیماً بر کاهش درآمد و رتبه‌بندی‌های ضعیف در کانال‌ها منجر خواهد شد. لیگ فوتبال کانادایی این جداول زمانی را به صورت دستی تهیه می‌کرد و به همین خاطر باید روش‌های بهتر و آسان‌تری را برای بهبود جداول زمانی و با در نظر گرفتن تمامی محدودیت‌ها پیدا می‌کرد. آنها سعی کردند تا برای ساخت یک مدل جامع زمانبندی از یک مشاور کمک بگیرند ولی پیاده‌سازی و اجرای این مدل همچنان یک چالش باقی ماند. لیگ تصمیم گرفت تا مسئله را با بکارگیری سالور موجود در مایکروسافت اکسل حل کند. برای برخی از اولویت‌ها باید تعادل برقرار می‌شد و حالیکه بهینه‌سازی جدول زمانی شامل موارد زیر بود:

۱. درآمد فروش – تنظیم جدول زمانی مسابقات و بازه‌های زمانی برای کلوپ‌هایی که درآمد بیشتری ایجاد می‌کنند.
۲. رتبه‌بندی کانال‌ها – تنظیم یک جدول زمانی که رتبه‌بندی کانال را برای آن شرکت پخش تلویزیونی بهبود بخشد.
۳. روزهای استراحت تیم – تنظیم یک جدول زمانی با دو تیم که در مقابل یکدیگر بازی می‌کنند و روزهای استراحت کافی برایشان در نظر گرفته شده است.

لیگ تصمیم گرفت تا جداول زمانی مسابقات را به این شکل بهبود بخشد که روزهای استراحت بازیکنان از اولویت بالاتری برخوردار است و پس از آن درآمد فروش و امتیاز کانال برای شرکت پخش رادیو و تلویزیونی در نظر گرفته می‌شود. این عمدتاً بدان خاطر است که درآمد فروش و امتیاز کانال، محصول جانبی عملکرد بازیکنان تیم در زمین هستند که مستقیماً با روزهای استراحت تیم‌ها مرتبط است.

متدولوژی/راه‌حل

در ابتدا، سازماندهی جداول زمانی در نرم‌افزار اکسل و توسط قابلیت سالور کار بسیار عظیمی بود. فرانتلین سیستمز یک نسخه‌ی پولی برای سالور ارائه کرده است که این امکان را می‌دهد تا تعداد تصمیم‌ها را از ۲۰۰ تصمیم به ۸۰۰۰ تصمیم افزایش دهیم. لیگ حتی مجبور بود تا محدودیت‌های خاص صنعت از قبیل پخش تلویزیونی در مناطق زمانی مختلف و برنامه‌ریزی انجام مسابقات رقابتی مهم در روز ملی کار را در نظر بگیرد. افزودن این محدودیت‌ها به هیچ عنوان کار ساده‌ای نبود تا زمانیکه مشاوران فرانتلین سیستمز به کمک لیگ شتافته و این مسئله‌ی غیرخطی را به یک مسئله‌ی خطی تبدیل کردند. «موتور» برنامه‌ریزی خطی به عنوان قوه‌ی محرکه‌ی این مدل عمل کرد. نرم‌افزار پرمیوم سالور توانست کمک بزرگی به بهبود و بهینه‌سازی جدول زمانی بازی‌ها کند.

¹ Christiansen et al.

² TurboRouter

نتایج/منافع

استفاده از یک جدول زمانی بهینه منجر به افزایش درآمد از طریق افزایش فروش بلیت و کسب امتیازهای تلویزیونی بالاتر برای کانال‌های پخش تلویزیونی خواهد شد. این امر محقق شد زیرا این ابزار قادر بود تا با راحتی از محدودیت‌های افزوده شده برای پیمانکاران پشتیبانی کند. جدول زمانبندی بهینه‌سازی شده برای اکثر ذینفعان لیگ مطلوبیت داشت. این یک فرایند تکرارپذیر است ولی این جداول زمانی مسابقات، پیشرفته‌ترین جداول زمانی مسابقات فصلی لیگ تا به امروز بوده‌اند.

سوالاتی برای مورد کاربردی ۸.۱

۲۵. سه روشی که در آن بهینه‌سازی جدول زمانی بازی‌ها به کمک سالور در مقایسه با زمانبندی دستی می‌تواند منجر به افزایش درآمد شود را نام ببرید.
۲۶. لیگ فوتبال کانادایی به چه روش‌های دیگری می‌تواند از نرم‌افزار سالور برای توسعه و ارتقاء عملیات کسب‌وکار خود استفاده کند؟
۲۷. در زمانبندی اینگونه بازی‌ها چه ملاحظات دیگری حائز اهمیت است؟

از این مورد کاربردی چه می‌توانیم بیاموزیم؟

با استفاده از افزونه‌ی سالور برای اکسل، لیگ فوتبال کانادایی با در نظر گرفتن ذینفعان و محدودیت‌های صنعت تصمیمات بهتری در زمانبندی بازی‌های خود گرفت که منجر به درآمدزایی و رتبه بندی بهتر کانال‌های تلویزیونی شد. بنابراین، یک برنامه بهینه شده، یکی از حوزه‌های اجرایی تحلیل تجویزی، ارزش قابل توجهی را به دست آورده است. با توجه به مطالعه موردی، مدلساز، آقای ترور هاردی، یک کاربر متخصص اکسل بود، اما در مدلسازی متخصص نبود. با این حال، سهولت استفاده از اکسل به او اجازه داد تا یک برنامه‌ی کاربردی از تحلیل تجویزی را توسعه دهد.

Compiled from "Canadian Football League Uses Frontline Solvers to Optimize Scheduling in 2016." Solver, September 7 2016, www.solver.com/news/canadian-football-league-uses-frontlinesolvers-optimize-scheduling-2016 (accessed September 2018); Kostuk, Kent J., and Keith A. Willoughby. "A Decision Support System for Scheduling the Canadian Football League." *Interfaces*, vol. 42, no. 3, 2012, pp. 286-295; Dilkina, Bistra N., and William S. Havens. The U.S. National Football League Scheduling Problem. Intelligent Systems Lab, www.cs.cornell.edu/~bistra/papers/NFLsched1.pdf (accessed September 2018).

مدلسازی یک عنصر کلیدی در تحلیل تجویزی به شمار می‌آید. در مثال‌هایی که در بخش مقدمه و موارد کاربردی ذکر شد، برای اینکه بتوانیم یک تصمیم را برای هر مسئله‌ی واقع‌بینانه توصیه کنیم باید از یک مدل ریاضی استفاده کنیم. به عنوان مثال، تصمیم‌گیری در خصوص اینکه چه پیشنهادی را به کدام مشتریان بدهیم که ارزش پاسخ کلی را به حداکثر رسانده ولی در عین حال از بودجه فراتر نرود چیزی نیست که بتوان به صورت دستی انجام داد. ساخت یک مدل پیشینه‌سازی پاسخ بر اساس احتمالات که محدودیت بودجه‌ای دارد، اطلاعاتی که به دنبال آن هستیم را در اختیارمان قرار می‌دهد. بسته به نوع مسئله‌ای که داریم، گروه‌های بسیاری از مدل‌ها و تکنیک‌های تخصصی بسیاری برای حل هر یک از آنها وجود دارد. ما در این فصل با دو روش مدلسازی مختلف آشنا خواهیم شد. اکثر دانشگاه‌ها چندین واحد درسی دارند این که موضوعات را تحت عناوینی از قبیل پژوهش عملیاتی، علم مدیریت، سیستم‌های پشتیبان تصمیم و شبیه‌سازی پوشش می‌دهند که می‌توانند به شما در دستیابی به تخصص بیشتر در این موضوعات کمک کنند. از آنجائیکه تحلیل‌های تجویزی عموماً شامل بکارگیری مدل‌های ریاضی است برخی اوقات عبارت علم داده‌ها برای بکارگیری اینگونه مدل‌های ریاضی مورد استفاده قرار می‌گیرد. پیش از آنکه بحث را در خصوص پشتیبانی مدلسازی ریاضی در تحلیل تجویزی آغاز کنیم، ابتدا به درک برخی از مسائل و مشکلات در مدلسازی می‌پردازیم.

هیچ تصمیمی در فضای خلاء گرفته نمی‌شود. تحلیل دامنه و نیروها و دینامیک‌های محیطی بسیار اهمیت دارد. یک تصمیم‌گیرنده باید فرهنگ سازمانی و فرایندهای تصمیم‌گیری سازمانی (مثلاً چه کسی تصمیم‌گیرنده است، درجه‌ی تمرکز سازمانی) را شناسایی کند. این امکان کاملاً وجود دارد که عوامل محیطی منجر به بروز مسئله‌ی جاری شده باشند. اینکار را **تحلیل و پیمایش محیطی**^۱ می‌نامند که عبارتست از پایش، پیمایش و تفسیر اطلاعات گردآوری شده. ابزارهای تحلیل کسب‌وکار/ هوش کسب‌وکار می‌توانند با پیمایش این عوامل به شناسایی مسئله کمک کنند. مسئله باید درک شود و هرکسی که درگیر آن است باید به درک مشترک و یکسانی از مسئله برسد زیرا مسئله نهایتاً در یک قالب از طریق مدل ارائه می‌شود. در غیراینصورت، این مدل هیچ کمکی به تصمیم‌گیرنده نخواهد کرد.

شناسایی متغیر شناسایی متغیرهای یک مدل (مثلاً تصمیم، نتیجه، کنترل‌ناپذیر) و همچنین روابط بین متغیرها کاری حساس و بسیار مهم است. نمودارهای اثرگذاری که مدل‌های گرافیکی از مدل‌های ریاضی هستند می‌توانند فرایند شناسایی را تسهیل کنند. یک شکل عمومی‌تر از یک نمودار اثرگذاری، نقشه‌ی شناختی نام دارد که می‌تواند به تصمیم‌گیرنده در درک بهتر مسئله و به ویژه متغیرها و تعاملات بین آنها کمک کند.

پیش‌بینی (تحلیل پیشگویانه) همانطور که قبلاً گفته شد، یک پیشنیاز مهم در تحلیل تجویزی، اطلاع و آگاهی از آنچه اتفاق افتاده و آنچه اتفاق خواهد افتاد است. این شکل از تحلیل پیشگویانه برای ساخت و بکارگیری مدل‌ها بسیار ضروری است زیرا زمانیکه یک تصمیم اجرا می‌شود، نتایج آن معمولاً در آینده رخ می‌دهند. اجرای یک تحلیل شرطی (حساسیت) در گذشته هیچ فایده‌ای ندارد زیرا تصمیماتی که در آن زمان گرفته شده‌اند هیچ تأثیری بر آینده نخواهند داشت. تجارت و ارتباطات آنلاین موجب شده تا نیاز مبرمی به پیش‌بینی و وفور اطلاعات موجود برای انجام **پیش‌بینی**^۲ بوجود آید. این فعالیت‌ها به سرعت اتفاق می‌افتند ولی اطلاعات مربوط به اینگونه خریدها گردآوری شده و باید برای پیش‌بینی‌های بهتر تحلیل شوند. بخشی از تحلیل صرفاً شامل پیش‌بینی تقاضا است ولی مدل‌های پیش‌بینی می‌توانند از نیازهای چرخه عمر محصول و اطلاعات مربوط به بازار و مشتریان برای تحلیل کل موقعیت استفاده کرده و به فروش بیشتر محصولات و خدمات کمک کنند. ما در اینجا به توضیح یک مثال مؤثر از اینگونه پیش‌بینی‌ها و بکارگیری آن در تصمیم‌گیری در اینگرام میکرو در مورد کاربردی ۸.۲ می‌پردازیم.

**اینگرام میکرو از برنامه‌های کاربردی هوش تجاری برای تصمیمات
قیمت‌گذاری استفاده می‌کند.**

مورد کاربردی ۸.۲

اینگرام میکرو^۳ بزرگترین توزیع‌کننده محصولات فناوری در جهان است. از طریق یک سیستم توزیع دو لایه، این شرکت محصولات را از تولیدکنندگان خریداری کرده و آنها را به خرده‌فروشان که آنها نیز به نوبه خود این محصولات را به کاربران نهایی می‌فروشند توزیع می‌کند. به عنوان مثال، می‌توانید یک بسته‌ی مایکروسافت آفیس ۳۶۵ را به جای اینکه مستقیماً از

¹ Environmental scanning and analysis

² Forecasting

³ Ingram Micro

سایت مایکروسافت خریداری کنید از اینگرام میکرو بخريد. اینگرام با بست‌بای^۱، بوفالو^۲، گوگل، هانی‌پول^۳، لیبراتون^۴ و شارپر ایمج^۵ همکاری دارد. این شرکت محصولات خود را به ۲۰۰,۰۰۰ شرکت تأمین‌کننده در سراسر جهان تحویل می‌دهد و بنابراین حجم زیادی از داده‌های تراکنش را در اختیار دارد. اینگرام به دنبال این بود تا از اطلاعات موجود در این داده‌ها برای شناسایی فرصت‌های فروش مکمل و تعیین قیمت‌ها برای پیشنهاد به مشتریان خاص به همراه بسته‌های نرم‌افزاری استفاده کند. این امر مستلزم راه اندازی یک مرکز هوش تجاری برای گردآوری و تجزیه و تحلیل داده‌ها بود. در راه اندازی این مرکز، اینگرام با مشکلات مختلفی روبرو شد.

۱. در فرایند ثبت و ضبط داده‌های خود با چندین مشکل مانند فقدان داده‌های از دست رفته، اطمینان از صحت اطلاعات کاربر نهایی مواجه شد و مرتبط کردن استعلام قیمت‌ها به سفارشات مواجه شد.
۲. اینگرام در پیاده سازی یک سیستم مدیریت ارتباط با مشتری که بتواند عملیات خود را در سراسر جهان اداره کند با مشکلات فنی روبرو بود.
۳. آنها در برابر ایده قیمت‌گذاری تقاضا (تعیین قیمت بر اساس تقاضای محصول) با مقاومت روبرو شدند.

متدولوژی/راه‌حل

اینگرام مستقیماً از طریق نامه‌های الکترونیکی با مشتریان (واسطه‌ها) خود تماس برقرار کرده و تخفیفاتی را برای خرید فناوری‌های پشتیبان در رابطه با محصولاتی که سفارش می‌دهند پیشنهاد کرد. آنها از طریق تحلیل سبد بخش‌بندی شده‌ی بازار این فرصت‌ها را شناسایی کردند و کاربردهای هوش تجاری زیر را که در تعیین قیمت بهینه به آنها کمک می‌کرد توسعه دادند. اینگرام یک ابزار جدید بهینه‌سازی قیمت به نام ایمپرایم^۶ را توسعه داد که قادر به تعیین قیمت‌ها بر اساس داده‌ها و ارائه‌ی مشاوره و راهنمایی برای مذاکره بر سر قیمت بر اساس داده‌ها می‌باشد. ایمپرایم برای هر سطحی از ساختار سلسله‌مراتب محصولات یک قیمت بهینه تعیین می‌کند (یعنی سطح مشتری، سطح فروشنده-مشتری، سطح مشتری-بخش، سطح فروشنده-مشتری-بخش). این ابزار این کار را با در نظر گرفتن توازن بین سیگنال تقاضا و قیمت‌گذاری در آن سطح انجام می‌دهد. شرکت همچنین یک پلتفرم بازاریابی دیجیتالی به نام اینتلیجنس اینگرام^۷ را توسعه داده است. این پلتفرم از امتیازدهی پیشگویانه مشتریان راغب (بالفوه) استفاده می‌کند که کاربران نهایی را انتخاب کرده و آنها را از طریق برنامه‌های بازاریابی خاص هدف قرار می‌دهد. امتیازدهی پیشگویانه مشتریان راغب برای شرکت‌هایی است که هیچ ارتباط مستقیمی با کاربران نهایی ندارند. این پلتفرم برای اجرای برنامه‌های فضای سفید بکار می‌رود که با ارائه‌ی تخفیفات، واسطه‌گر را به خرید محصولات مرتبط ترغیب می‌کند. به عنوان مثال، اگر یک واسطه‌گر، از اینگرام یک سرور را خریداری کند آنگاه اینگرام برای خرید دیسک‌های ذخیره‌سازی تخفیف خواهد داد زیرا هر دو محصول با یکدیگر کار می‌کنند. به همین منوال، از اینتلیجنس اینگرام برای اجرای کمپین‌های تشویقی (ارائه‌ی جوایز نقدی به واسطه‌گرها در صورتیکه اهداف فروش فصلی را زده باشند) و کمپین‌های فروش مکمل (ارسال نامه الکترونیکی به کاربران نهایی در مورد محصولات مرتبط با محصولی که اخیراً خریداری کرده‌اند) بکار رفته است.

نتایج/منافع

سود حاصل از بکارگیری ایمپرایم با استفاده از یک متدولوژی سنجش لیفت اندازه‌گیری شد. این متدولوژی، دوره‌های زمانی قبل و بعد از تغییر قیمت و همچنین گروه‌های تست در مقابل گروه‌های کنترل را مقایسه می‌کند. سنجش لیفت بر روی متوسط فروش روزانه، حاشیه ناخالص و حاشیه (یادگیری ماشین) انجام می‌شود. استفاده از ایمپرایم منجر به یک رشد درآمد ۷۵۷ میلیون دلاری و افزایش سود ناخالص به میزان ۱۸/۸ میلیون دلار شد.

¹ Best Buy

² Buffalo

³ Honeywell

⁴ Libratone

⁵ Sharper Image

⁶ IMPRIME

⁷ Intelligence INGRAM

سوالاتی برای بحث و گفتگو

۱. اینگرام میکرو در توسعه‌ی یک مرکز هوش تجاری با چه چالش‌های اساسی مواجه بود؟
۲. تمامی راه‌حل‌های هوش تجاری که اینگرام برای بهینه‌سازی قیمت محصولات خود توسعه داده است را نام ببرید.
۳. اینگرام پس از استفاده از برنامه‌های کاربردی هوش تجاری خود به چه مزایای دست یافت؟

از این مورد کاربردی چه می‌توانیم بیاموزیم؟

یک شرکت ابتدا با ساخت یک مرکز هوش تجاری شروع به درک بهتر محصولات، مشتریان و الگوهای خرید آنها می‌کند. این بینش از تحلیل‌های توصیفی و پیشگویانه حاصل می‌شود. از طریق بهینه‌سازی قیمت‌گذاری که یکی از حوزه‌های کاربرد تحلیل تجویزی است می‌توان به ارزش‌های بیشتری در آینده دست یافت.

Sources: R. Mookherjee, J. Martineau, L. Xu, M. Gullo, K. Zhou, A. Hazlewood, X. Zhang, F. Griarte, & N. Li. (2016). "End-to-End Predictive Analytics and Optimization in Ingram Micro's Two-Tier Distribution Business." *Interfaces*, 46(1), 49-73; ingrammicrocommerce.com, "CUSTOMERS," <https://www.ingrammicrocommerce.com/customers/> (accessed July 2016).

جدول ۸.۱ برخی از مدل‌های تصمیم‌گیری را در قالب هفت گروه دسته‌بندی کرده و چندین تکنیک بارز از هر گروه را نام برده است. هر تکنیک را می‌توان برای یک مدل ایستا یا پویا بکار گرفت که این مدل‌ها را می‌توان تحت شرایط قطعیت، عدم قطعیت یا ریسک ایجاد کرد. برای تسریع ساخت مدل می‌توانیم از سیستم‌های تحلیل تصمیم ویژه که قابلیت‌ها و زبان‌های مدلسازی در داخل آنها تعبیه شده است استفاده کنیم. اینها شامل صفحات گسترده، سیستم‌های داده‌کاوی، سیستم‌های پردازش تحلیلی آنلاین و زبان‌های مدلسازی هستند که می‌توانند به یک تحلیلگر در ساخت یک مدل کمک کنند. ما در ادامه‌ی این فصل به معرفی یکی از این سیستم‌ها خواهیم پرداخت.

مدیریت مدل مدل‌ها نیز همانند داده‌ها باید برای حفظ انسجام و یکپارچگی و کاربردپذیری‌شان مدیریت شوند. چنین مدیریتی به کمک سیستم‌های مدیریت مبتنی بر مدل‌ها که مشابه سیستم‌های مدیریت پایگاه داده است انجام می‌پذیرد.

مدلسازی مبتنی بر دانش سیستم‌های پشتیبان تصمیم عمدتاً از مدل‌های کمی استفاده می‌کنند در حالیکه سیستم‌های خبره از مدل‌های کیفی و مبتنی بر دانش در برنامه‌های کاربردی خود استفاده می‌کنند. ساخت مدل‌های قابل حل و قابل استفاده نیازمند کمی دانش است. از بسیاری از تکنیک‌های تحلیلی پیشگویانه از قبیل دسته‌بندی و خوشه‌بندی می‌توان برای ساخت مدل‌های مبتنی بر دانش استفاده کرد.

جدول ۸.۱ دسته‌بندی مدل‌ها

دسته	فرایند و هدف	تکنیک‌های بارز
بهینه‌سازی مسائل با آلترناتیوهای محدود	یافتن بهترین جواب از بین تعداد معدودی از آلترناتیوها	جداول تصمیم، درخت تصمیم، فرایند سلسله‌مراتب تحلیلی

مدل‌های برنامه‌ریزی خطی و سایر مدل‌های برنامه‌ریزی ریاضی، مدل‌های شبکه‌ای	یافتن بهترین جواب از بین تعداد زیادی از آلترناتیوها با استفاده از یک فرایند بهبود گام به گام	بهینه‌سازی از طریق الگوریتم
برخی از مدل‌های موجودی	یافتن بهترین جواب در یک گام با استفاده از یک فرمول	بهینه‌سازی از طریق یک فرمول تحلیلی
انواع مختلف شبیه‌سازی	یافتن یک جواب مناسب یا بهترین جواب از بین آلترناتیوهای بررسی شده از طریق آزمایش	شبیه‌سازی
برنامه‌ریزی هیوریستیک، سیستم‌های خبره	یافتن یک جواب مناسب با استفاده از قواعد	هیوریستیک
مدل‌های پیش‌بینی، آنالیز مارکوف	پیش‌بینی آینده برای یک سناریوی مفروض	مدل‌های پیشگویانه
مدلسازی مالی، مدل‌های صف	حل یک پرونده یا مسئله شرطی با استفاده از یک فرمول	سایر مدل‌ها

روندهای جاری در مدلسازی یک روند اخیر در مدلسازی شامل توسعه‌ی کتابخانه‌هایی از مدل‌ها و کتابخانه‌های تکنیک جواب مسئله است. برخی از این کدها را می‌توان مستقیماً به صورت رایگان بر روی وب سرور اجرا کرده و برخی دیگر را می‌توان بر روی یک رایانه محل بارگذاری و اجرا نمود. در دسترس بودن این کدها بدین معناست که بسته‌های نرم‌افزاری شبیه‌سازی و بهینه‌سازی قدرتمندی در اختیار تصمیم‌گیرندگان قرار دارد که ممکن است با این ابزارها فقط در کلاس درس و به صورت نظری آشنا شده باشند. به عنوان مثال، بخش علوم رایانه‌ای و ریاضی^۱ در آزمایشگاه ملی آرگون^۲، سرور NEOS را برای بهینه‌سازی در مسیر <https://neos-server.org/neos/index.html> نگهداری می‌کند. با کلیک بر روی لینک Resources در آدرس <https://neos-server.org/neos/index.html> که وب سایت مؤسسه‌ی پژوهش عملیاتی و علوم مدیریت (اینفورمز) است می‌توانید به سایت‌های دیگری هدایت شوید. مجموعه‌ی بسیار غنی از اطلاعات مدلسازی و جواب‌ها در اینفورمز موجود است. وب سایت یکی از نشریات اینفورمز به نام *آر/ام/اس تودی*^۳ در آدرس <http://www.orms-today.org/ormsmain.shtml> شامل پیوندهایی به گروه‌های بسیاری از نرم‌افزارهای مدلسازی می‌باشد.

این روندی واضح و روشن در جهت توسعه و بکارگیری ابزارها و نرم‌افزارهای مبتنی بر ابر برای دسترسی و حتی اجرای نرم‌افزار به منظور مدلسازی، بهینه‌سازی و شبیه‌سازی است. این روند به طرق گوناگونی موجب شده تا بکارگیری تعداد زیادی از مدل‌ها در مسائل دنیای واقعی ساده‌تر شود. با وجود این، برای استفاده مؤثر از مدل‌ها و تکنیک‌های حل مسئله، لازم است تا از طریق توسعه و حل مسائل ساده‌تر به تجربه کافی دست پیدا کنیم. این جنبه اغلب نادیده گرفته می‌شود. سازمان‌هایی که تحلیلگران کلیدی را در اختیار دارند که نحوه‌ی بکارگیری این مدل‌ها را می‌دانند، آنها را به نحوی مؤثر نیز بکار می‌گیرند. این روند به طرز قابل توجهی در حوزه‌ی مدیریت درآمد رخ می‌دهد که از حیطة شرکت‌های هواپیمایی، هتل‌ها و اجاره‌ی خودرو به خرده‌فروشی، بیمه، سرگرمی و بسیاری حوزه‌های دیگر نیز سرایت کرده است. مدیریت ارتباط با مشتریان نیز از مدل‌هایی که اغلب برای کاربران بسیار شفاف هستند استفاده می‌کنند. در مدل‌های مدیریت، حجم داده‌ها و اندازه‌ی مدل نیز بسیار بزرگ می‌شود که نیازمند بکارگیری انبارهای داده برای تأمین داده‌ها و سخت‌افزارهای رایانش موازی برای دستیابی به جواب در یک بازه زمانی قابل قبول است.

¹ Mathematics and Computer Science Division

² Argonne National Laboratory

³ OR/MS Today

این یک روند مستمر در جهت شفافیت هرچه بیشتر مدل‌های تحلیلی برای تصمیم‌گیرندگان است. به عنوان مثال، تحلیل (مدلسازی) چندبُعدی شامل تحلیل داده در چندین بُعد است. در تحلیل (مدلسازی) چندبُعدی، داده‌ها معمولاً در یک قالب صفحه گسترده که اکثر تصمیم‌گیرندگان با آن آشنا هستند نمایش داده می‌شود. بسیاری از تصمیم‌گیرندگان که به تفکیک معکب داده‌ها عادت دارند امروزه در سیستم‌های OLAP که به انبار داده‌ها دسترسی دارند استفاده می‌کنند. اگرچه این روش‌ها موجب شده‌اند تا فرایند مدلسازی رضایت‌بخش شود ولی بسیاری از گروه مدل‌های مهم و قابل استفاده و در نتیجه برخی جنبه‌های مهم و ظریف در تفسیر جواب مسئله که باید مد نظر قرار دارد را نیز حذف می‌کنند. در مدلسازی از داده‌های بسیار بیشتری در مقایسه با تحلیل داده‌ها استفاده می‌شوند که خطوط روند و روابط بین داده‌ها به کمک روش‌های آماری تعیین می‌شود. همچنین برای کمک به تحلیل مدل‌ها، یک مدل از روی یک مدل ساخته می‌شود. یک نمودار تأثیرگذاری در واقع نمایش گرافیکی یک مدل است. برخی بسته‌های نرم‌افزاری ترسیم نمودارهای تأثیرگذاری قادر به ایجاد و حل یک مدل هستند.

سوالاتی برای مرور بخش ۸.۲

۱۸. سه درسی که از مدلسازی آموخته‌اید را نام ببرید.
۱۹. سه مشکل اصلی در مدلسازی را نام برده و توضیح دهید.
۲۰. انواع اصلی مدل‌های بکار رفته در سیستم‌های پشتیبان تصمیم چه هستند؟
۲۱. چرا برخی از مدل‌ها به آن اندازه که لازم است در صنعت بکار گرفته نمی‌شوند؟
۲۲. روندهای جاری در مدلسازی چه هستند؟

تمامی مدل‌های کمی معمولاً از چهار جزء اصلی تشکیل شده‌اند (شکل ۸.۱): متغیرهای نتیجه (یا پیامد)، متغیرهای تصمیم، متغیرهای کنترل‌ناپذیر (و/یا پارامترها) و متغیرهای نتیجه‌ی بینابینی. این اجزا از طریق روابط ریاضی بهم پیوند خورده‌اند. در مدل‌های غیرکمی، روابط به صورت سمبولیک یا کیفی نمایش داده می‌شوند. نتایج تصمیمات بر اساس تصمیم گرفته شده (یعنی مقادیر متغیرهای تصمیم)، عواملی تحت کنترل تصمیم‌گیر نیستند (در محیط) و روابط بین متغیرها تعیین می‌شود. فرایند مدلسازی شامل شناسایی متغیرها و روابط بین آنها است. با حل یک مدل، مقادیر این متغیرها و متغیرهای نتیجه مشخص می‌شود.

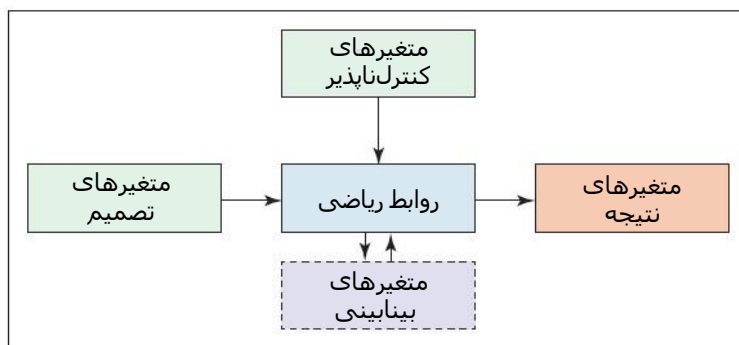
متغیرهای نتیجه (پیامد) متغیرهای نتیجه (پیامد)^۱ بیانگر سطح اثربخشی یک سیستم هستند یعنی نشان می‌دهند که یک سیستم تا چه اندازه خوب کار می‌کند یا با اهداف خود دست پیدا می‌کند. این متغیرها؛ خروجی هستند. مثال‌هایی از متغیرهای

¹ Result (outcome) variables

نتیجه در جدول ۸.۲ نشان داده شده‌اند. متغیرهای نتیجه را *متغیرهای وابسته*^۱ می‌نامند. متغیرهای نتیجه‌ی بینابینی برخی اوقات در مدلسازی برای شناسایی نتایج واسطه یا بینابینی بکار می‌روند. در خصوص یک متغیر وابسته، پیش از آنکه رویدادی که این متغیر توضیح می‌دهد اتفاق بیوفتد می‌بایست یک رویداد دیگر رخ بدهد. متغیرهای نتیجه به وقوع متغیرهای تصمیم و متغیرهای کنترل‌ناپذیر بستگی دارند.

متغیرهای تصمیم متغیرهای تصمیم^۲ به توضیح حالت‌های آلترناتیو برای اقدام می‌پردازند. تصمیم‌گیرنده قادر به کنترل کردن متغیرهای تصمیم است. به عنوان مثال، برای یک مسئله‌ی سرمایه‌گذاری، میزان سرمایه‌گذاری در اوراق قرضه یک متغیر تصمیم محسوب می‌شود. در یک مسئله‌ی زمانبندی، متغیرهای تصمیم عبارتند از نیروی انسانی، زمان و جداول زمانی. مثال‌های دیگر در جدول ۸.۲ آورده شده است.

پارامترها یا متغیرهای کنترل‌ناپذیر در هر موقعیت تصمیم‌گیری، عواملی وجود دارند که بر متغیرهای تصمیم تأثیر می‌گذارند ولی تحت کنترل شخص تصمیم‌گیرنده نیستند. هر یک از این عوامل را می‌توان تثبیت نمود که در این حالت آنها را **پارامترها**^۳ یا **متغیرهای کنترل‌ناپذیر**^۴ می‌نامند یا اینکه ممکن است متغیر باشند که در این صورت آنها را متغیر می‌نامند. مثال‌هایی از این عوامل شامل نرخ بهره، ضوابط ساخت و ساز ساختمان در شهرها، قوانین و مقررات مالیاتی و هزینه‌های بهره‌برداری (آب، برق، تلفن و ...) می‌باشد. اکثر این عوامل کنترل‌ناپذیر هستند زیرا توسط عناصری از محیطی که تصمیم‌گیرنده در آن کار می‌کند تعیین می‌شوند. برخی از این متغیرها موجب محدود شدن تصمیم‌گیر شده و آنچه را که ما محدودیت‌های مسئله می‌نامیم تشکیل می‌دهند.



شکل ۸.۱ ساختار عمومی یک مدل کمی

جدول ۸.۲ مثال‌هایی از اجزای مدل‌ها

حوزه	متغیرهای تصمیم	متغیرهای نتیجه	پارامترها و متغیرهای کنترل‌ناپذیر
سرمایه‌گذاری مالی	حجم سرمایه‌گذاری و آلترناتیوها	سود کل، ریسک	نرخ تورم

¹ Dependant variables

² Decision variables

³ Parameters

⁴ Uncontrollable Parameters

نرخ بهره پایه رقابت	نرخ بازگشت سرمایه سود هر سهم میزان نقدینگی		
درآمد حاصل از مشتری اقدامات رقبا	سهم بازار رضایت مشتری	بودجه تبلیغات محل تبلیغات	بازاریابی
ظرفیت ماشین‌آلات فناوری قیمت مواد	هزینه کل سطح کیفیت رضایت کارکنان	چه چیزی و چه میزان تولید شود سطح موجودی برنامه‌های جبران خدمت	ساخت و تولید
فناوری رایانه‌ای نرخ مالیات الزامات قانونی	هزینه پردازش داده نرخ خطا	استفاده از رایانه‌ها جدول زمانی ممیزی	حسابداری
مسافت تحویل محموله قوانین و مقررات	هزینه‌ی کل حمل و نقل زمان شناوری پرداخت	جدول زمانی ارسال استفاده از سبدهای هوشمند	حمل و نقل
تقاضا برای خدمات	رضایت مشتریان	Staffing level	خدمات

متغیرهای نتیجه بینابینی متغیرهای نتیجه‌ی بینابینی¹ بیانگر نتایج یا پیامدهای میانی در مدل‌های ریاضی هستند. به عنوان مثال، در مسئله‌ی زمانبندی ماشین‌آلات، ضایعات به عنوان یک متغیر نتیجه‌ی بینابین و سود کل به عنوان یک متغیر نتیجه تعریف می‌شود (یعنی ضایعات یکی از عوامل تعیین کننده‌ی سود کل است). یک مثال دیگر، حقوق و دستمزد کارکنان است که یک متغیر تصمیم برای مدیریت به شمار می‌آید و تعیین کننده‌ی رضایت کارکنان (یک نتیجه‌ی بینابین) است که به نوبه‌ی خود تعیین کننده‌ی میزان بهره‌وری کارکنان خواهد بود (یک نتیجه‌ی نهایی).

اجزا و مؤلفه‌های یک مدل کمی به وسیله‌ی معادلات یا نامعادلات ریاضی (جبری) به یکدیگر متصل می‌شوند. یک مدل مالی بسیار ساده به شکل زیر است:

$$P = R - C$$

که P = سود، R = درآمد و C = هزینه است. این معادله رابطه‌ی بین متغیرها را توضیح می‌دهد. یک مدل مالی معروف دیگر، مدل ساده‌ی جریان نقدی ارزش فعلی است که P = ارزش فعلی، F = پرداخت آتی بر حسب دلار، i = نرخ بهره (درصد) و n = تعداد سال است. با این مدل می‌توان ارزش فعلی یک پرداخت ۱۰۰,۰۰۰ دلاری را در بازه زمانی ۵ سال از اکنون با نرخ بهره‌ی ۱۰٪ تعیین کرد.

$$P = 100,000 / (1 + 0.1)^5 = 62,092$$

در بخش‌های بعدی به معرفی مدل‌های ریاضی جالب‌تر و پیچیده‌تر خواهیم پرداخت.

¹ Intermediate

سوالاتی برای مرور بخش ۸.۳

۱۴. یک متغیر تصمیم چیست؟

۱۵. اجزای یک مدل کمی را نام برده و به اختصار توضیح دهید.

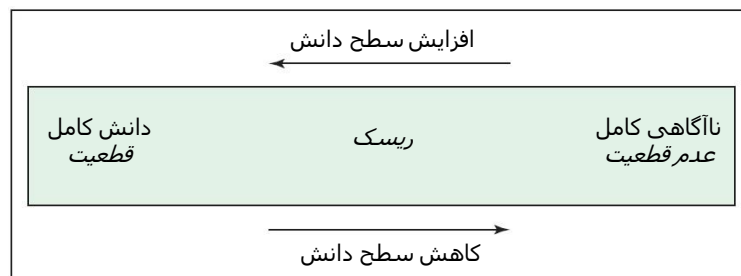
۱۶. نقش متغیرهای نتیجه‌ی بینابین را توضیح دهید.

فرایند تصمیم‌گیری شامل ارزیابی و مقایسه‌ی آلترناتیوها می‌باشد. در طی این فرایند، لازم است که نتیجه و پیامدی آتی هر آلترناتیو پیشنهادی پیش‌بینی شود. موقعیت‌های تصمیم را اغلب بر اساس آنچه تصمیم‌گیرنده در مورد نتایج پیش‌بینی شده می‌داند دسته‌بندی می‌کنند. ما به طور معمول این دانش را در قالب سه گروه از اطلاع کامل تا بی‌اطلاعی کامل دسته‌بندی می‌کنیم (شکل ۸.۲):

- قطعیت
- عدم قطعیت
- ریسک

در زمان توسعه‌ی مدل‌ها هر یک از این شرایط ممکن است رخ دهد و برای هر یک از شرایط ممکن است انواع مختلفی از مدل‌ها مناسب باشند. در ادامه ما به بحث در خصوص تعاریف پایه این عبارات و برخی مسائل مهم در مدلسازی برای هر کدام از این شرایط می‌پردازیم.

در تصمیم‌گیری تحت شرایط قطعیت فرض بر این است که اطلاع کامل از شرایط داریم به نحویکه تصمیم‌گیرنده دقیقاً می‌داند که نتیجه یا پیامد هر اقدام چه خواهد بود (همانند یک محیط قطعی). ممکن است این فرض که ما ۱۰۰٪ از نتایج و پیامدها آگاه هستیم درست نباشد یا اینکه لازم نباشد که تمامی نتایج و پیامدهای را واقعاً ارزشیابی کنیم ولی این فرض اغلب موجب ساده‌سازی مدل می‌شود. تصمیم‌گیرنده در اینجا یک پیش‌بینی‌کننده‌ی بی‌عیب و نقص آینده است زیرا فرض بر این است که برای هر آلترناتیو فقط یک نتیجه وجود دارد. به عنوان مثال، یک آلترناتیو برای سرمایه‌گذاری در اوراق خزانه‌داری ایالات متحده داشتن اطلاعات کامل در مورد نرخ بازگشت سرمایه‌ی آینده‌ی آن است. یکی از موقعیت‌هایی که تصمیم‌گیری تحت قطعیت صورت می‌گیرد در مسائل ساختارمند و بازه‌های زمانی کوتاه (حداکثر ۱ ساله) است. مدل‌های قطعی را به سادگی می‌توان توسعه داده و حل کرد و می‌توانند جواب‌های بهینه‌ای داشته باشند. بسیاری از مدل‌های مالی تحت فرضیه‌ی قطعیت ساخته می‌شوند اگرچه هیچ‌چیز در بازار ۱۰۰٪ قطعی نیست.



شکل ۸.۲ منطقه‌بندی تصمیم‌گیری

در تصمیم‌گیری تحت شرایط عدم قطعیت، تصمیم‌گیرنده موقعیت‌هایی را در نظر می‌گیرد که برای هر اقدام چندین نتیجه یا پیامد امکان‌پذیر است. برخلاف موقعیت ریسک، در این حالت، تصمیم‌گیرنده از احتمال وقوع نتایج ممکن مطلع نبوده یا قادر به تخمین آن نیست. تصمیم‌گیری تحت شرایط عدم قطعیت بسیار دشوارتر از تصمیم‌گیری تحت قطعیت است زیرا اطلاعات کافی در اختیار تصمیم‌گیرنده نیست. مدلسازی اینگونه موقعیت‌ها شامل ارزیابی نگرش تصمیم‌گیرنده (یا سازمان) نسبت به ریسک است. مدیران سعی دارند تا حد امکان از شرایط عدم قطعیت اجتناب کنند حتی تا جائیکه حاضر هستند آن را نادیده بگیرند. به جای سر و کله زدن با شرایط عدم قطعیت آنها تلاش می‌کنند تا اطلاعات بیشتری را بدست بیاورند تا بتوانند مسئله را تحت شرایط قطعیت یا ریسک محاسبه شده حل و فصل کنند. اگر اطلاعات بیشتری موجود نباشد، مسئله را باید تحت یک شرط عدم قطعیت حل نمود.

تصمیمی که تحت ریسک گرفته می‌شود (موقعیت تصمیم‌گیری احتمالی یا تصادفی) موقعیتی است که تصمیم‌گیرنده باید چندین نتیجه یا پیامد ممکن را برای هر آلترناتیو که هر یک احتمال وقوع متفاوتی دارند در نظر بگیرد. فرض بر این است که احتمالات طولانی مدتی که نتایج و پیامدهای مفروض تحت آنها رخ خواهند داد، معلوم هستند یا اینکه می‌توان آنها را تخمین زد. تحت این مفروضات، تصمیم‌گیرنده می‌تواند میزان ریسک هر آلترناتیو را ارزیابی کند (ریسک محاسبه شده). اکثر تصمیمات کسب‌وکار تحت ریسک مفروض یا محاسبه شده اتخاذ می‌شوند. **تحلیل ریسک**^۱ (یعنی ریسک محاسبه شده) یک روش تصمیم‌گیری است که به تجزیه و تحلیل ریسک (بر اساس احتمالات معلوم و مفروض) مرتبط با آلترناتیوهای مختلف می‌پردازد. تحلیل ریسک را می‌توان با محاسبه‌ی ارزش مورد انتظار هر آلترناتیو و انتخاب بهترین ارزش مورد انتظار انجام داد. مورد کاربردی ۸.۳ یکی از کاربردهای تحلیل ریسک را برای کاهش عدم قطعیت نشان می‌دهد.

آمریکن ایرلاینز از مدلسازی Should-Cost برای ارزیابی عدم قطعیت پیشنهادات (مزایده) مسیرهای حمل مرسولات استفاده می‌کند.

مورد کاربردی ۸.۳

شرکت آمریکن ایرلاینز^۲ یکی از بزرگترین شرکت‌های هوایمایی است که کسب‌وکار اصلی آن حمل و نقل مسافر است ولی فعالیت‌های فرعی مهم دیگری از قبیل حمل و نقل کامیون درستی^۳ تجهیزات نگهداری و تعمیرات و ارسال حین پرواز اقلام خدمت‌رسانی به مسافران^۴ است که می‌تواند ارزش موجودی این شرکت را تا ۱ میلیارد دلار افزایش دهد. آمریکن ایرلاینز پیشنهادات متعددی را در پاسخگویی به درخواست‌های پیشنهاد قیمت موجودی‌های خود از تأمین‌کنندگان دریافت می‌کند. تعداد این درخواست‌های پیشنهاد قیمت در هر سال ممکن است به بیش از ۵۰۰ مورد برسد. مضمّن قیمت‌ها به دلیل تعداد بسیار بالای پیشنهادات قیمت و پیچیدگی فرایند ارسال پیشنهادات می‌تواند بسیار متغیر باشد. انحراف قیمت برای پیشنهاد قیمت یک قرارداد

¹ Risk Analysis

² American Airlines

³ Full-Truckload Fright Shipment

⁴ In-flight shipment of passenger service items

ممکن است برخی اوقات تا ۲۰٪ انحراف داشته باشد. به دلیل پیچیدگی فرایند، معمولاً پرداخت‌های انجام شده بابت خدمات تأمین‌کنندگان، بیشتر یا کمتر از میزان است. تا به اینجا، امریکن ایرلاینز به یک مدل *should-cost* برای ارزیابی قیمت‌های پیشنهادی ارسالی توسط تأمین‌کنندگان نیاز داشت تا بتواند بهترین پیشنهاد را که برای هر دو طرف منصفانه باشد انتخاب نماید.

متدولوژی/راه‌حل

تعیین هزینه منصفانه برای خدمات و محصولات تأمین‌کننده در سه مرحله انجام شده است:

۱. منابع اولیه (مصاحبه‌ها) و ثانویه (اینترنت) به عنوان موارد اصلی مورد بررسی قرار گرفتند و متغیرهای هزینه‌ای تأثیرگذار بر قیمت پیشنهادی برای حمل و نقل کامیون درستی بر اساس طیف داده‌ها تعیین گردید.
۲. متغیرهای هزینه‌ای به نحوی انتخاب شدند که دو به دو ناسازگار و جمعی جامع باشند.
۳. از نرم‌افزار تحلیل تصمیم DPL برای مدلسازی عدم قطعیت استفاده شد.

علاوه بر این از روش تقریب بسطیافته‌ی سوانسون-مگیل^۱ برای مدلسازی توزیع احتمال حساس‌ترین متغیرهای هزینه استفاده شده که اینکار به دلیل تغییرپذیری بالای پیشنهادات قیمت در مدل اولیه بوده است.

نتایج/منافع

یک آزمون پایلوت بر روی یک درخواست قیمت پیشنهادی که از شش شرکت حمل FTL اخذ شده بود انجام شد. از بین شش قیمت پیشنهادی ارسالی، پنج مورد آنها در محدوده‌ی سه انحراف استاندارد از میانگین بودند در حالیکه یک مورد داده‌ی پرت بود. در نتیجه، امریکن ایرلاینز از مدل *should-cost* FTL بر روی بیش از ۲۰ درخواست قیمت پیشنهادی استفاده کرد تا هزینه‌ی دقیق و منصفانه‌ی خدمات و محصولات را تعیین کند. انتظار می‌رفت که این مدل به کاهش ریسک پرداخت بیشتر یا کمتر از حد انتظار به تأمین‌کنندگان بکاهد.

سوالاتی برای بحث و گفتگو

۱. علاوه بر کاهش ریسک پرداخت کمتر یا بیشتر از حد به تأمین‌کنندگان، امریکن ایرلاینز چه منافعی می‌تواند از مدل *should-be* بدست آورد؟
۲. آیا به غیر از حمل و نقل هوایی حوزه‌ی دیگری برای بکارگیری این مدل می‌شناسید؟
۳. در مورد روش‌های دیگری که امریکن ایرلاینز می‌توانست برای حل مشکل اضافه پرداخت یا کم پرداخت به تأمین‌کنندگان بکار بگیرد بحث کنید.

Source: Based on Bailey, M. J., Snapp, J., Yetur, S., Stonebraker, J. S., Edwards, S. A., Davis, A., & Cox, R. (2011). Practice summaries: American Airlines uses should-cost modeling to assess the uncertainty of bids for its full-truckload shipment routes. *Interfaces*, 41(2), 194-196.

سوالاتی برای مرور بخش ۸.۴

۱۰. تصمیم‌گیری تحت قطعیت مفروض، ریسک و عدم قطعیت را توضیح دهید.
۱۱. چگونه می‌توان مسائل تصمیم‌گیری تحت قطعیت مفروض را حل کرد؟
۱۲. چگونه می‌توان مسائل تصمیم‌گیری تحت عدم قطعیت مفروض را حل کرد؟

¹ Extended Swanson-Megill

۱۳. چگونه می‌توان مسائل تصمیم‌گیری تحت ریسک مفروض را حل کرد؟

مدل‌ها را می‌توان با استفاده از سیستم‌ها و زبان‌های مختلف برنامه‌نویسی توسعه داده و پیاده‌سازی کرد. تمرکز ما عمدتاً بر صفحات گسترده^۱ (به همراه افزونه‌های آنها)، زبان‌های مدل‌سازی و ابزارهای تحلیل داده‌ی شفاف است. بسته‌های نرم‌افزاری صفحه گسترده با قابلیت و انعطاف‌پذیری خود سریعاً به عنوان نرم‌افزارهای پیاده‌سازی آسان برای توسعه‌ی طیف وسیعی از برنامه‌های کاربردی در کسب‌وکار، مهندسی، ریاضی و علوم دیگر محبوبیت یافتند. صفحات گسترده شامل قابلیت‌ها، توابع و روتین‌های گسترده‌ی آماری، پیش‌بینی و مدیریت پایگاه‌های داده هستند. یک بسته‌ی نرم‌افزار صفحه گسترده شامل افزونه‌هایی برای ایجاد ساختار و حل گروه‌های خاصی از مدل‌ها می‌باشد. از بین بسته‌های افزونه‌ها تعداد بسیاری برای توسعه‌ی سیستم‌های پشتیبان تصمیم توسعه یافته‌اند. این افزونه‌های مرتبط با سیستم‌های پشتیبان تصمیم شامل سالور^۲ (فرانتلاین سیستمز، solver.com) و واتز بست^۳ (نسخه‌ای از لیندو از شرکت لیندو سیستمز^۴، lindo.com) برای بهینه‌سازی‌های خطی و غیرخطی؛ برینسل^۵ (جوریک ریسرچ سافت‌ور^۶، jurikres.com) و نیورال تولز^۷ (پاليساد کورپ^۸، palisade.com) برای شبکه‌های عصبی مصنوعی؛ ایوالور^۹ (پاليساد کورپ) برای الگوریتم‌های ژنتیک و @RISK (پاليساد کورپ) برای مطالعات شبیه‌سازی هستند. افزونه‌های رایگان یا به قیمت بسیار پایین نیز وجود دارد که با قابلیت‌های این افزونه‌ها برابری می‌کنند.

صفحه گسترده مشخصاً محبوب‌ترین ابزار مدل‌سازی کاربر نهایی است زیرا شامل تعداد زیادی از توابع قدرتمند مالی، آماری، ریاضی و غیره می‌باشد. صفحات گسترده می‌توانند وظایفی از قبیل برنامه‌ریزی خطی و تحلیل رگرسیون را برای حل مدل انجام دهند. صفحات گسترده به ابزار مهمی برای تحلیل، برنامه‌ریزی و مدل‌سازی تبدیل شده‌اند (نگاه کنید به فاراسین، پرکوز و وان د وِده، ۲۰۰۸؛ هِری و بالز، ۲۰۰۸؛ اُفچینیکوف و میلنر ۲۰۰۸). مورد کاربردی ۸.۴ و ۸.۵ کاربردهای جالب مدل‌های مبتنی بر صفحات گسترده را در یک سازمان غیرانتفاعی نشان می‌دهد.

مرکز فرزندپذیری پنسیلوانیا از مدل صفحه گسترده برای تطبیق بهتر کودکان با خانواده‌ها استفاده می‌کند

مورد کاربردی ۸.۴

مرکز فرزندپذیری پنسیلوانیا توسط ایالت پنسیلوانیا در سال ۱۹۷۹ به منظور کمک به ارگان‌های استانی و غیرانتفاعی برای یافتن خانواده‌های احتمالی کودکان بی‌سرپرست که به دلایل سنی یا نیازهای ویژه به فرزندخواندگی پذیرفته نشده‌اند تأسیس شد.

¹ Spreadsheets

² Solver

³ What's Best!

⁴ Lindo Systems Inc.

⁵ Brancel

⁶ Jurik Research Software

⁷ NeuralTools

⁸ Palisade Corp.

⁹ Evolver

این مرکز پرونده‌های کاملی از کودکان و اولویت‌های خانواده‌هایی که به دنبال فرزندپذیری هستند را نگهداری می‌کند. این مرکز به دنبال خانواده‌هایی برای کودکان سرتاسر ۶۷ بخش ایالت پنسیلوانیا می‌باشد. مرکز فرزندپذیری ایالتی پنسیلوانیا و شبکه‌ی خانواده مسئول یافتن منزل دائمی برای کودکان بی‌سرپرست هستند. اگر پس از چندین بار تلاش، این شبکه موفق به یافتن مکان مناسبی برای کودک نشود، آنها از مرکز فرزندپذیری پنسیلوانیا کمک می‌گیرند. این مرکز از یک ابزار ارزیابی خودکار برای انطباق دادن کودکان با خانواده‌ها استفاده می‌کند. این ابزار با محاسبه‌ی یک امتیاز بین ۰ تا ۱۰۰٪ برای یک کودک بر روی ۷۸ زوج مقدار صفات کودک و اولویت‌های خانواده‌ها، موارد انطباقی را توصیه می‌کند. این مرکز چندین سال است که با ارائه‌ی توصیه‌های تطبیق به مددکاران درگیر است. مدیریت یک پایگاه داده‌ی گسترده از کودکان که در طول زمان از ۶۷ بخش ایالتی گردآوری شده باشد کار دشواری است. الگوریتم جستجوی پایه، توصیه‌هایی را ارائه می‌کرد که برای مددکاران چندان مفید نبودند. در نتیجه، تعداد کودکانی که به فرزندخواندگی پذیرفته نمی‌شدند به میزان زیادی افزایش یافته و نیاز مبرمی به یافتن خانواده‌ها برای این کودکان بی‌سرپرست احساس شد.

متدولوژی/راه حل

مرکز فرزندپذیری پنسیلوانیا از طریق نظرسنجی‌های آنلاین که شامل مجموعه‌ی جدیدی از پرسش‌ها بودند شروع به گردآوری اطلاعات در مورد کودکان بی‌سرپرست و خانواده‌ها کرد. این پرسش‌ها اطلاعاتی در مورد سرگرمی‌های کودکان، اولویت‌های مددکاران-کودک برای خانواده‌ها و ترجیح محدوده سنی کودکان برای خانواده‌ها را گردآوری می‌کردند. مرکز و مشاوران یک ابزار انطباق صفحه‌گسترده را ایجاد کردند که در مقایسه با ابزارهای خودکار قبلی شامل ویژگی‌های بیشتری بود. در این مدل، مددکاران می‌توانند صفات و ویژگی‌ها را برای انتخاب یک خانواده برای یک کودک وزن‌دهی کنند. به عنوان مثال، اگر خانواده‌ی در خصوص جنسیت، سن و نژاد کودک سخت‌گیر باشد آنگاه این عوامل وزن بالاتری به خود می‌گیرند. همچنین، مددکاران می‌توانند به محل سکونت خانواده اولویت بالاتری دهند زیرا روابط اجتماعی یک عامل مهم برای کودک محسوب می‌شود. کمیته‌ی تطبیق با استفاده از این ابزار می‌تواند کودک و خانواده را بر اساس هر صفت و ویژگی مقایسه کرده و تصمیم دقیق‌تری را برای تطبیق بین خانواده و کودک اتخاذ کند.

نتایج/منافع

از زمانیکه مرکز فرزندپذیری پنسیلوانیا شروع به استفاده از مدل صفحه‌گسترده برای تطبیق کودک و خانواده کرد آنها توانسته‌اند تصمیمات بهتری را در این خصوص اتخاذ کنند. در نتیجه، درصد کودکانی که وارد یک خانواده می‌شوند افزایش یافته است. این مورد کاربرد کوتاهی یکی از چندین مثال بکارگیری صفحات گسترده به عنوان یک ابزار پشتیبانی تصمیم است. با ایجاد یک سیستم امتیازدهی ساده برای صفات و ویژگی‌های کودکان و تمایل خانواده یک سیستم تطبیق بهتر تولید شده و تعداد پرونده‌های مردودی از هر دو طرف کاهش می‌یابد.

سوالاتی برای بحث و گفتگو

۱. مرکز فرزندپذیری پنسیلوانیا برای اتخاذ تصمیمات تطبیق کودک و خانواده با چه چالش‌هایی مواجه بود؟
۲. کدام ویژگی‌های این ابزار صفحه گسترده‌ی جدید به مرکز فرزندپذیری در حل مشکل تطبیق کودک و خانواده کمک کرد؟

Source: Based on Slaugh, V. W., Akan, M., Kesten, O., & Unver, M. U. (2016). The Pennsylvania Adoption Exchange improves its matching process. *Interfaces*, 46(2), 133-154.

مترو میلز آن ویلز از اکسل برای یافتن بهترین مسیرهای تحویل استفاده می‌کند

مورد کاربردی ۸.۵

انجمن میلز آن ویلز آمریکا^۱ یک سازمان غیرانتفاعی است که حدوداً ۱ میلیون بسته‌ی غذایی را به منازل سالمندان نیازمند در سرتاسر ایالات متحده ارسال می‌کند. مترو میلز آن ویلز ترژر ولی^۲ یک شاخه‌ی محلی از میلز آن ویلز آمریکا است که در ایالت آیداهو فعالیت می‌کند. این شاخه تیمی از رانندگان دواطلب را برای هدایت روزانه‌ی وسایل نقلیه‌ی شخصی برای تحویل بسته‌های غذایی به ۸۰۰ مشتری در ۲۱ مسیر بکار گرفته و ناحیه‌ای با مساحت ۲,۷۴۵ کیلومتر مربع را پوشش می‌دهد. این سازمان با مسائل بسیاری روبرو بود. اول اینکه، آنها به دنبال این بودند که زمان تحویل را به کمترین میزان ممکن برسانند زیرا غذاهای طبخ شده (گرم) نسبت به دما حساس بودند و ممکن بود به آسانی از بین بروند. آنها می‌خواستند تا غذای طبخ شده (گرم) را در عرض ۹۰ دقیقه پس از اینکه راننده به راه افتاد تحویل دهند. دوم، فرایند زمانبندی بسیار زمان‌بر بود. دو کارمند بخش زیادی از زمان خود را صرف تهیه‌ی جداول زمانی مسیرهای تحویل می‌کردند. یک هماهنگ‌کننده‌ی مسیر، ایستگاه‌ها را بر اساس تعداد دریافت‌کنندگان غذا در هر روز تعیین می‌کرد. پس از تعیین ایستگاه‌ها، هماهنگ‌کننده یک توالی از ایستگاه‌ها را تعیین می‌کرد که زمان سفر دواطلبان را به حداقل میزان برساند. سپس این جدول زمانبندی مسیرها را وارد یک ابزار آنلاین می‌کردند تا دستورالعمل‌های نوبه به نوبه‌ی رانندگان مشخص شود. کل این فرایند دستی برای تصمیم‌گیری مسیرها زمان اضافی زیادی را می‌گرفت. مترو میلز آن ویلز به دنبال یک ابزار مسیریابی بود که بتواند سیستم تحویل آنها را بهینه کرده و راه‌حل‌های مسیریابی را برای مسیرهای یک طرفه و رفت و برگشت برای تحویل غذاها تهیه کند. آنهایی که راننده ثابت هستند می‌توانند روز بعد غذای گرم‌تر یا سردتر را تحویل دهند ولی آنهایی که هر از گاهی رانندگی می‌کنند برای گرفتن غذاهای گرم‌تر یا سردتر باید به آشپزخانه بازگردند.

متدولوژی/راه‌حل

برای حل مسئله‌ی مسیریابی یک ابزار مبتنی بر صفحه‌گسترده تولید شد که دارای یک رابط کاربری بود که گیرنده به راحتی می‌توانست اطلاعاتی از قبیل نام، نیاز غذایی و آدرس تحویل را وارد کند. این اطلاعات باید برای هر ایستگاه در مسیر در داخل صفحه گسترده وارد میشد. سپس، از کارکرد ویزوال بیسیک در اکسل برای دسترسی به رابط برنامه‌نویسی کاربردی نقشه‌ی شبکه به نام مپ‌کوئست^۳ استفاده شد. این رابط برنامه‌نویسی برای ایجاد یک ماتریس سفرها استفاده شد که اطلاعات زمان و فاصله‌ی تحویل را برای ۵,۰۰۰ زوج مکان در هر روز و بدون هزینه محاسبه می‌کند. زمانیکه برنامه شروع می‌شود، ابتدا مپ‌کوئست ای‌پی‌آی آدرس‌های وارد شده برای گیرندگان را صحه‌گذاری می‌کند. سپس این برنامه از رابط برنامه‌نویسی برای بازیابی مسافت، زمان تخمینی تحویل و دستورالعمل‌های نوبت به نوبت برای رانندگی بین تمامی ایستگاه‌ها در مسیر استفاده می‌کند. سپس این ابزار می‌تواند مسیر بهینه را برای حداکثر ۳۰ ایستگاه در یک محدوده‌ی زمانی امکان‌پذیر پیدا کند.

نتایج/منافع

در نتیجه‌ی بکارگیری این ابزار، کل مسافت رانندگی در سال به میزان ۱۰,۰۰۰ مایل کاهش یافت در حالیکه زمان سفر نیز به میزان ۵۳۰ ساعت کاهش یافته است. مترو میلز آن ویلز ترژر ولی در سال ۲۰۱۵ مبلغ ۵,۸۰۰ دلار را بر اساس نرخ صرفه‌جویی تخمین شده‌ی ۰/۵۸ دلار در هر مایل صرفه‌جویی کرد. این ابزار همچنین زمان صرف شده برای برنامه‌ریزی مسیرهای تحویل غذاها را کاهش داد. از مزایای دیگر این ابزار می‌توان به افزایش رضایت دواطلبان و نگهداشت بیشتر (کاهش ترک دواطلبان) آنان اشاره کرد.

سوالاتی برای بحث و گفتگو

۱. مترو میلز آن ویلز ترژر ولی پیش از بکارگیری ابزار مبتنی بر صفحه‌گسترده برای تحویل بسته‌های غذایی با چه چالش‌هایی مواجه بود؟
۲. طراحی مدل مبتنی بر صفحه گسترده را توضیح دهید.

¹ Meals on Wheels Association of America

² Metro Meals on Wheels Treasure Valley

³ MapQuest

۳. مزایای نامشهود بکارگیری مدل مبتنی بر اکسل برای مترو میلز آن ویلز چه بوده است؟

Source: Based on Manikas, A. S., Kroes, J. R., & Gattiker, T. F. (2016). Metro Meals on Wheels Treasure Valley employs a lowcost routing tool to improve deliveries. *Interfaces*, 46(2), 154-167.

از ویژگی‌های مهم دیگر صفحه‌گسترده‌ها می‌توان به تحلیل شرطی، جستجوی هدف، مدیریت داده‌ها و قابلیت برنامه‌نویسی (یعنی ماکروها) اشاره کرد. به کمک یک صفحه‌گسترده به آسانی می‌توان مقدار یک سلول را تغییر داده و نتیجه را بلافاصله مشاهده کرد. جستجوی هدف با نشان دادن یک سلول هدف، مقدار مطلوب آن و یک سلول متغیر انجام می‌شود. مدیریت گسترده‌ی پایگاه‌های داده را می‌توان با مجموعه داده‌های کوچک انجام داده یا بخش‌هایی از یک پایگاه داده را برای تحلیل وارد نمود (دقیقاً روشی است که OLAP با مکعب داده‌های چندبُعدی کار می‌کند و در واقع اکثر سیستم‌های OLAP پس از بارگذاری داده‌ها شکل و شمایل‌ی مانند نرم‌افزارهای صفحه‌گسترده‌ی پیشرفته دارد). الگوها، ماکروها و سایر ابزارها موجب افزایش بهره‌وری در ساخت سیستم‌های پشتیبان تصمیم می‌شوند.

اکثر بسته‌های صفحه‌گسترده امکان یکپارچه‌سازی بسیار کاملی را فراهم می‌آورند زیرا قادر به خواندن و نوشتن ساختارهای فایلی متداول و برقراری آسان ارتباط با پایگاه‌های داده و سایر ابزارها می‌باشد. مایکروسافت اکسل محبوب‌ترین بسته‌ی نرم‌افزاری صفحه‌گسترده است. در شکل ۸.۳ ما یک مدل ساده‌ی محاسبه‌ی وام را نشان داده‌ایم که باکس‌های صفحه‌گسترده نشان دهنده‌ی محتوای سلول‌ها هستند که حاوی فرمول می‌باشند. هر تغییری در نرخ بهره در سلول E7 بلافاصله در پرداخت ماهانه در سلول E13 نشان داده می‌شود. نتایج را می‌توان بلافاصله مشاهده و تحلیل کرد. اگر بخواهیم مقدار پرداخت ماهانه‌ی مشخصی را داشته باشیم، می‌توانیم از جستجوی هدف برای تعیین یک نرخ بهره‌ی مناسب یا مبلغ وام استفاده کرد.

مدل‌های ایستا یا پویا را می‌توان در یک صفحه‌گسترده ایجاد کرد. به عنوان مثال، صفحه‌گسترده‌ی محاسبه‌ی ماهانه وام در شکل ۸.۳ یک مدل ایستا است. اگرچه نوع مسئله در طول زمان بر وام‌گیرنده تأثیر می‌گذارد این مدل یک عملکرد تک ماهه را نشان می‌دهد که تکرار شده است. برعکس، یک مدل پویا بیانگر رفتار در طول زمان است. محاسبات وام در صفحه‌گسترده‌ی شکل ۸.۴ نشان دهنده‌ی تأثیر پیش‌پرداخت بر اصل وام در طول زمان است. با استفاده از تولیدکننده‌های اعداد تصادفی می‌توان تحلیل ریسک را در صفحه‌گسترده‌ها انجام داد تا بتوان مدل‌های شبیه‌سازی را توسعه داد.

کاربردهای صفحات گسترده در مدل‌ها بسیار زیاد است که در بخش بعدی در مورد نحوه استفاده از یک مدل بهینه‌سازی مبتنی بر صفحه‌گسترده صحبت خواهیم کرد.

	A	B	C	D	E	F	G	H	
1									
2									
3		Simple Loan Calculation Model in Excel							
4									
5									
6		Loan Amount			\$150,000				
7		Interest Rate			8.00%				
8		Number of Years			30				
9									
10		Number of Months			360			=E8*12	
11		Interest Rate/Month			0.67%			=E7/12	
12									
13		Monthly Loan Payment			\$1,100.68			=PMT (E11, E10, E6, 0)	
14									
15									
16									
17		Excel Spreadsheet Static Model Example of a Simple Loan							
18									
19									
20									
21									
22									

شکل ۸.۳ یک مثال از مدل ایستای صفحه گسترده اکسل برای محاسبه پرداخت‌های ماهانه وام

سوالاتی برای مرور بخش ۸.۵

۱۳. صفحه گسترده چیست؟

۱۴. یک افزونه‌ی صفحه گسترده چیست؟ این افزونه‌ها چگونه می‌توانند به ایجاد و استفاده از سیستم‌های پشتیبان تصمیم کمک کنند؟

۱۵. توضیح دهید چرا یک صفحه گسترده برای توسعه‌ی سیستم‌های پشتیبان تصمیم بسیار سودمند است.

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3		Dynamic Loan Calculation Model with Prepayment in Excel									
4											
5											
6		Loan Amount			\$150,000						
7		Interest Rate			8.00%						
8		Number of Years			30						
9											
10		Number of Months			360						
11		Interest Rate/Month			0.67%						
12											
13		Monthly Loan Payment			\$1,100.68						
14											
15											
16											
17		Excel Spreadsheet Dynamic Model Example of a Simple Loan									
18											
19											
20											
21											
22		Month	Normal Payment	Prepay Amount	Total Payment	Principle Owed					
23		0				\$150,000					
24		1	\$1,100.65	\$100.00	\$1,200.65	\$149,795					
25		2	\$1,100.65	\$100.00	\$1,200.65	\$149,597					
26		3	\$1,100.65	\$100.00	\$1,200.65	\$149,394					
27		4	\$1,100.65	\$100.00	\$1,200.65	\$149,189					
28		5	\$1,100.65	\$100.00	\$1,200.65	\$148,983					
29											
30											

شکل ۸.۴ یک مثال از مدل پویای صفحه گسترده اکسل برای محاسبه پرداخت‌های ماهانه وام و اثرات پیش پرداخت

برنامه‌ریزی ریاضی^۱ خانواده‌ای از ابزارهایی هستند که برای کمک به حل مسائل مدیریتی طراحی شده که تصمیم‌گیرنده باید منابع محدود را بین فعالیت‌های رقابتی برای بهینه‌سازی یک هدف کمی پذیر تخصیص دهد. به عنوان مثال، توزیع زمان ماشین (منبع) بین محصولات مختلف (فعالیت‌ها) یکی از مسائل متداول تخصیص منابع است. **برنامه‌ریزی خطی^۲** بهترین تکنیک از خانواده ابزارهای بهینه‌سازی به نام **برنامه‌ریزی ریاضی** است که تمامی روابط بین متغیرها از نوع خطی می‌باشد. از آن به طور گسترده‌ای در سیستم‌های پشتیبانی تصمیم استفاده می‌شود (مورد کاربردی ۸.۶). مدل‌های برنامه‌ریزی خطی کاربردهای عملی بسیار مهمی دارند که شامل مدیریت زنجیره تأمین، تصمیمات آمیزه‌ی محصول، مسیریابی و غیره هستند. شکل‌های خاصی از این مدل‌ها را می‌توان برای کاربردهای خاص بکار برد. به عنوان مثال، مورد کاربردی ۸.۶ یک مدل صفحه گسترده را توضیح می‌دهد که برای ایجاد یک جدول زمانی برای پزشکان بکار رفته است.

مسائل تخصیص برنامه‌ریزی خطی معمولاً دارای ویژگی‌های زیر هستند:

- تعداد محدود منابع اقتصادی موجود برای تخصیص.
- منابع برای تولید محصولات یا خدمات بکار می‌روند.
- منابع را به دو یا چند روش می‌توان بکار برد که هر یک را *جواب (راه‌حل)* یا یک *برنامه* می‌گویند.
- هر فعالیت (محصول یا خدمت) که منابع در آن مورد استفاده قرار می‌گیرند نتیجه‌ای را بر حسب هدف تعیین شده باز می‌گرداند.
- تخصیص معمولاً با الزامات و محدودیت‌های متعددی به نام *محدودیت^۳* همراه است.

مدل برنامه‌ریزی عدد صحیح مختلط به مرکز پزشکی دانشگاه تنسی برای زمانبندی پزشکان کمک می‌کند

مورد کاربردی ۸.۶

اعضای وابسته‌ی منطقه‌ای نوزادان نارس^۴ یک گروه ۹ نفری از پزشکان هستند که برای بخش مراقبت‌های ویژه‌ی نوزادان نارس در مرکز پزشکی دانشگاه تنسی در ناکسویل^۵ ایالت تنسی فعالیت می‌کنند. این گروه همچنین به ارائه‌ی خدمات اضطراری در دو بیمارستان محلی ناکسویل نیز می‌پردازند. برای سالیان متمادی، یکی از اعضای گروه مسئول تهیه‌ی جدول زمانی پزشکان به صورت دستی بوده است. با وجود این، با توجه به بازنشستگی این شخص، به یک سیستم خودکار برای زمانبندی پزشکان نیاز بود. پزشکان می‌خواستند این سیستم یک توازن در بار کاری آنها برقرار کند زیرا در روش قبلی چنین توازنی بین بار کاری آنها برقرار نبود. علاوه بر این، زمانبندی باید به شکلی می‌بود که پزشکان حتماً ۷ روز هفته و ۲۴ ساعته در بخش مراقبت‌های ویژه‌ی نوزادان نارس حضور می‌داشتند و در صورت امکان این زمانبندی باید اولویت‌ها و ترجیحات تک تک پزشکان از لحاظ نوع شیفت را نیز در نظر می‌گرفت. برای رسیدگی به این مسئله، پزشکان با دانشکده‌ی علوم مدیریت دانشگاه تنسی تماس گرفتند.

¹ Mathematical Programming

² Linear Programming

³ Constraints

⁴ Regional Neonatal Associates

⁵ Neonatal Intensive Care Unit (NICU) at the University of Tennessee Medical Center in Knoxville

مسئله‌ی زمانبندی پزشکان در شیفت‌ها دارای محدودیت‌هایی از لحاظ بار کاری و انتخاب‌های سبک زندگی بود. اولین گام برای حل مسئله‌ی زمانبندی این بود که شیفت‌ها را بر اساس نوع آنها (شب و روز) گروه‌بندی کنند. گام بعدی، تعیین محدودیت‌های مسئله بود. این مدل باید یک بازه‌ی زمانی $\frac{7}{24}$ هفته‌ای را برای $\frac{7}{24}$ پزشک پوشش می‌داد که دو پزشک در روزهای کاری و یک پزشک در طول شب و تعطیلات آخر هفته کار می‌کردند. علاوه بر این، یک پزشک باید به طور انحصاری برای پوشش‌دهی دو بیمارستان محلی به صورت $\frac{7}{24}$ تخصیص داده می‌شد. محدودیت‌های واضح دیگری نیز می‌بایست در نظر گرفته می‌شد مثل اینکه به پزشکی که شیفت شب داشته نباید شیفت روز بعد را تخصیص داد.

متدولوژی/راه‌حل

مسئله با ایجاد یک مدل بهینه‌سازی عدد صحیح-متخلط دودویی تدوین شد. این مدل ابتدا بار کاری را به صورت مساوی بین $\frac{7}{24}$ پزشک تقسیم کرد ولی نمی‌توانست عدد یکسانی از شیفت‌های روز و شب را بین آنها تخصیص دهد. بدین ترتیب مسئله‌ی توزیع عادلانه مطرح شد. علاوه بر این، پزشکان در مورد بار کاری تخصیص داده شده نظرات متفاوتی داشتند. شش پزشک می‌خواستند هر پزشک در بازه‌ی زمانی $\frac{7}{24}$ هفته‌ی شیفت‌های روز و شب یکسانی داشته باشند در حالیکه دیگران خواهان یک جدول زمانی بر اساس اولویت‌های فردی در مورد شیفت‌ها بودند. برای برآوردن نیازمندی‌های هر دو گروه، مدل جدیدی به نام مدل زمانبندی اولویت‌های هیبریدی (اچ‌پی‌اس‌ام)¹ ایجاد شد. برای برآوردن الزام مساوات شش پزشک، این مدل ابتدا بار کاری یک هفته را محاسبه کرده و آن را بین $\frac{7}{24}$ هفته تقسیم نمود. بدین ترتیب، کار به صورت عادلانه و مساوی بین تمامی شش پزشک تقسیم شده بود. بار کاری سه پزشک باقیمانده بر اساس اولویت‌ها و ترجیحات آنها در یک زمانبندی $\frac{7}{24}$ هفته‌ای توزیع شد. جدول زمانی تولید شده توسط پزشکان بررسی شده و این جدول برایشان قابل‌قبول‌تر بوده است.

مزایا و منافع

روش اچ‌پی‌اس‌ام هر دو الزام تساوی و عدالت و هم ترجیح تک تک پزشکان را در نظر گرفت. علاوه بر این، جداول زمانی بدست آمده از این مدل، در مقایسه با جداول زمانی دستی زمان‌های استراحت بهتری را برای پزشکان فراهم آورده و امکان ثبت درخواست‌های تعطیلات در آن نیز وجود داشت. مدل اچ‌پی‌اس‌ام قادر به حل مسائل زمانبندی مشابه که نیازمند ترجیحات و اولویت‌های نسبی بین انواع شیفت‌ها است می‌باشد. تکنیک‌هایی از قبیل مدل‌های برنامه‌ریزی عدد صحیح-مختلط می‌توانند زمانبندی‌های بهینه‌ای را ایجاد کرده و به عملیات کمک کنند. مدت زمان طولانی است که در سازمان‌های بزرگ از این تکنیک‌ها استفاده می‌شود. امروزه می‌توان اینگونه مدل‌های تحلیلی تجویزی را در صفحات گسترده و سایر نرم‌افزارهای موجود به آسانی اجرا کرد.

سوالاتی برای مورد کاربردی ۷/۶

۱. گروه اعضای وابسته‌ی منطقه‌ای نوزادان نارس با چه مشکلی مواجه بودند؟
۲. مدل اچ‌پی‌اس‌ام چگونه الزامات و نیازمندی‌های تمامی پزشکان را برآورده کرد؟

Source: Adapted from Bowers, M. R., Noon, C. E., Wu, W., & Bass, J. K. (2016). Neonatal physician scheduling at the University of Tennessee Medical Center. *Interfaces*, 46(2), 168–182.

¹ Hybrid Preference Scheduling Model (HPSM)

مدل تخصیص برنامه‌ریزی خطی مبتنی بر مفروضات اقتصادی منطقی زیر است:

- بازگشت‌ها از منابع مختلف تخصیص یافته را می‌توان مقایسه کرد؛ یعنی اینکه آنها را با یک واحد مشترک می‌تواند سنجید (مثلاً دلار، بهره‌وری).
- بازگشت از هر منبع تخصیص یافته مستقل از سایر تخصیص‌ها است.
- کل بازگشت‌ها عبارتست از حاصلجمع بازگشت‌های فعالیت‌های مختلف.
- تمامی داده‌ها با قطعیت مشخص هستند.
- منابع به اقتصادی‌ترین شکل ممکن استفاده می‌شوند.

مسائل تخصیص معمولاً جواب‌های ممکن بسیار زیادی دارند. بسته به مفروضات اساسی، تعداد جواب‌ها می‌تواند محدود یا نامحدود باشد. معمولاً، جواب‌های متفاوت، پاداش‌های متفاوتی نیز دارند. از بین جواب‌های موجود، حداقل یک جواب از لحاظ میزان دستیابی به هدف، بهترین جواب است (یعنی پاداش کل در بیشترین حد ممکن است) که آن را **جواب بهینه**¹ می‌نامند و با استفاده از یک الگوریتم ویژه قابل دستیابی است.

هر مدل برنامه‌ریزی خطی از متغیرهای تصمیم (که مقادیر آنها نامعلوم است و به دنبال آن هستیم)، یک تابع هدف (یک تابع ریاضی خطی که بین متغیرهای تصمیم و هدف رابطه برقرار کرده، میزان تحقق را سنجیده و باید بهینه‌سازی شود)، ضرایب تابع هدف (ضرایب هزینه یا سود که نشانگر میزان تأثیر بر هدف یک واحد از یک متغیر تصمیم است)، محدودیت‌ها (که به شکل نامعادلات خطی یا نامعادلاتی که بیانگر محدودیت‌های منابع و/یا نیازمندی‌ها هستند و از طریق روابط خطی متغیرها را بهم مرتبط می‌کنند)، ظرفیت‌ها (که بیانگر حدود بالایی و برخی اوقات پایینی محدودیت‌ها و متغیرها هستند) و ضرایب ورودی/خروجی (فناوری) (بیانگر میزان بهره‌برداری از منابع برای یک متغیر تصمیم است) تشکیل شده است.

اکنون به یک مثال توجه کنید. شرکت ام‌بی‌آی که تولیدکننده‌ی رایانه‌های تخصصی است باید تصمیمی بگیرد: ماه آینده در سایت بوستون چه تعداد رایانه باید تولید کند؟ ام‌بی‌آی دو نوع رایانه را در نظر دارد: CC@7 که به ۳۰۰ روز کاری و ۱۰,۰۰۰ دلار مواد نیاز دارد و CC@8 که نیازمند ۵۰۰ روز کاری و ۱۵,۰۰۰ دلار مواد است. سهم سود هر CC@7 معادل ۸,۰۰۰ دلار است در حالیکه این رقم برای CC@8 معادل ۱۲,۰۰۰ دلار می‌باشد. ظرفیت ماهانه کارخانه ۲۰۰,۰۰۰ روز کاری است و بودجه ماهانه خرید مواد اولیه نیز ۸ میلیون دلار می‌باشد. نیاز بازاریابی، تولید حداقل ۱۰۰ واحد از CC@7 و حداقل ۲۰۰ واحد از CC@8 در هر ماه می‌باشد. مسئله در اینجا بیشینه کردن سود شرکت با تعیین تعداد تولید ماهانه هر کدام از مدل‌های CC@7 و CC@8 می‌باشد. توجه داشته باشید که در دنیای واقعی، گردآوری داده‌های مورد نیاز مسئله و در عین حال گردآوری داده‌های مورد نیاز تصمیم‌گیر برای نحوه‌ی ساخت مدل برای حل مسئله ممکن است چندین ماه به طول بینجامد. ابزارهای مبتنی بر وب می‌توانند به گردآوری داده‌ها کمک کنند.

پیش فناوری ۸.۱ برنامه‌ریزی خطی

¹ Optimal solution

برنامه‌ریزی خطی را شاید بتوان شناخته‌شده‌ترین مدل بهینه‌سازی دانست که با تخصیص بهینه‌ی منابع بین فعالیت‌ها سر و کار دارد. مسئله‌ی تخصیص را با یک مدل در اینجا توضیح می‌دهیم.

مسئله در اینجا یافتن مقادیر متغیرهای تصمیم X_1, X_2, \dots به نحوی است که مقدار متغیر نتیجه‌ی Z بیشینه گردد که در اینجا مجموعه‌ای محدودیت‌های خطی که بیانگر فناوری، شرایط بازار و سایر متغیرهای کنترل‌ناپذیر هستند نیز وجود دارند. روابط ریاضی همگی به شکل معادلات و معادلات خطی هستند. از لحاظ نظری، هر مسئله‌ی تخصیص از این نوع دارای تعداد نامحدودی جواب امکان‌پذیر است. با استفاده از روال‌های ریاضی ویژه، روش برنامه‌ریزی خطی یک روال جستجوی رایانه‌ای منحصرفرد را برای یافتن بهترین جواب(ها) در عرض چند ثانیه بکار می‌گیرد. همچنین، روش یافتن جواب به طور خودکار به تحلیل حساسیت نیز می‌پردازد.

یک مدل برنامه‌ریزی خطی استاندارد را می‌توان برای مسئله‌ی شرکت ام‌بی‌آی ایجاد کرد. همانطور که در بینش فناوری ۸.۱ بحث شد، مدل برنامه‌ریزی خطی دارای سه مؤلفه است: متغیرهای تصمیم، متغیرهای نتیجه و متغیرهای کنترل‌ناپذیر (محدودیت‌ها). متغیرهای تصمیم به شرح زیر می‌باشند:

$$X_1 = \text{CC@7}$$

$$X_2 = \text{CC@8}$$

$$Z =$$

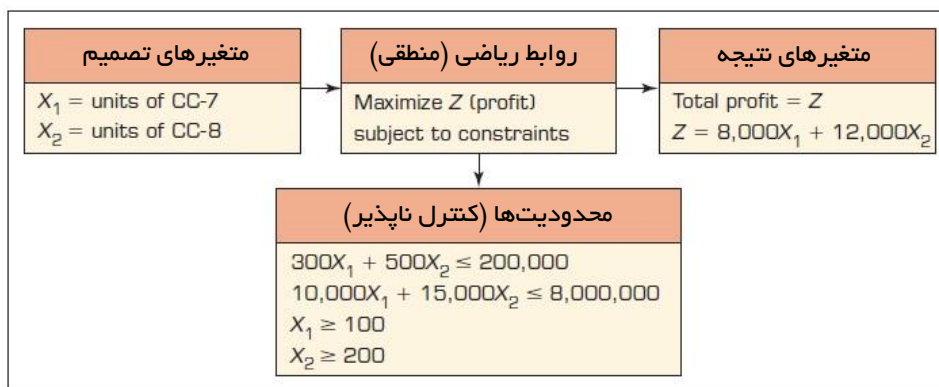
$$Z = 8,000X_1 + 12,000X_2$$

$$300X_1 + 500X_2 \leq 200,000$$

$$10,000X_1 + 15,000X_2 \leq 8,000,000$$

$$X_1 \geq 100 \quad \text{CC-7}$$

$$X_1 \geq 200 \quad \text{CC-8}$$



شکل ۸.۵ مدل ریاضی برای یک مثال آمیزه-محصول

Options Menu	Analysis	Data	Add-ins
Target Cell	By Changing Cells	Max	
Total Consumed	Add-in	D7 C7	E7
		D8 C8	
E	D C	X_2 X_1	
F	SUMPRODUCT		
Solver Add-in	Add-in		

Options

Limits Answer, Sensitivity

$$\text{Profit} = 5,066,667 \quad X_2 = 200 \quad X_1 = 333.33$$

F1

Product-Mix Model				
	X1	X2	Total Consumed	Limit
Decision Variables:	333.3333	200.00		
Total Profit:	8	12	5066.6667	
Labor:	0.3	0.5	200.00	200
Budget:	10	15	6333.33	8000
X1 Lower:	1	0	333.33	100
X2 Lower:	0	1	200.00	200

(Profit and Constraints Scaled by 1000)

Solver Parameters

Set Objective:

To: Max Min Value Of:

By Changing Variable Cells:

Subject to the Constraints:

- Add
- Change
- Delete
- Load/Save

Make Unconstrained Variables Non-Negative

Select a Solving Method: Options

Solving Method
 Select the GRG Nonlinear engine for Solver Problems that are smooth nonlinear. Select the LP Simplex engine for linear Solver Problems, and select the Evolutionary engine for Solver problems that are non-smooth.

Close Solve

شکل ۸.۶ جواب سالور اکسل برای مثال آمیزه‌ی محصول

	Electoral			Influence	
	State	Votes	W/E	N/S	Function
4	NV	6	West		F1
5	CO	9	West		F2
6	IA	6	West	North	F3
7	WI	10	West	North	F1
8	OH	18	East	North	F2
9	VA	13	East	South	F2
10	NC	15	East	South	F1
11	FL	29	East	South	F3
12	NH	4	East		F3
14	F1	Young	Old		
15	Men	3	1	4	
16	Women	3	3	6	
17		6	4	10	Total
19	F2	Young	Old		
20	Men	1.5	2.5	4	
21	Women	2.5	1	3.5	
22		4	3.5	7.5	Total
24	F3	Young	Old		
25	Men	2.5	2.5	5	
26	Women	1	2	3	
27		3.5	4.5	8	Total

شکل ۸.۷ داده‌های مثال تخصیص منابع برای انتخابات

NV, CO, IA, WI, OH, VA, NC, FL

NH

EV

EV

$$\text{Max } (6NV + 9CO + 6IA + 10WI + 18OH + 13VA + 15NC + 29FL + 4NH)$$

$$NV + CO + IA + WI + OH + VA + NC + FL + NH \leq 1,050$$

$$NV \geq 0.05 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$CO \geq 0.05 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$IA \geq 0.05 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$WI \geq 0.05 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$OH \geq 0.05 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$VA \geq 0.05 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$NC \geq 0.05 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$FL \geq 0.05 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$NH \geq 0.05 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

"عمومی" بنویسیم.

$$NV \leq 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$CO \leq 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$IA \leq 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$WI \leq 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$OH \leq 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$VA \leq 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$NC \leq 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$FL \leq 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$NH \leq 0.25 (NV + CO + IA + WI + OH + VA + NC + FL + NH)$$

$$NV + CO + IA + WI =$$

$$OH + VA + NC + FL + NH =$$

$$(NV + CO + IA + WI) \geq 0.60 (OH + VA + NC + FL + NH)$$

$$(10NV + 7.5CO + 8IA + 10WI + 7.5OH + 7.5VA + 10NC + 8FL + 8NH) \geq 9,200$$

... ..

-
-
-
-
-
-
-

analytics-magazine.org/spring-2009/205-software-solutions-the-abcs-of-simulationpractice.html

... ..

¹ Banks & Gibson

شبیه‌سازی احتمالی^۱ در شبیه‌سازی احتمالی، یک یا چند متغیر مستقل (مثلاً تقاضا در یک مسئله‌ی موجودی) دارای توزیع احتمالی است که از یک توزیع احتمالی خاص که می‌تواند توزیع گسسته باشد یا پیوسته تبعیت می‌کند:

- توزیع‌های گسسته به موقعیت‌هایی اطلاق می‌شود که تعداد محدودی رویداد (یا متغیر) می‌توانند تعداد محدودی مقدار را به خود بگیرند.

شبیه‌سازی وابسته به زمان و مستقل از زمان شبیه‌سازی مستقل از زمان به موقعیتی اشاره دارد که در آن برای ما مهم نیست بدانیم یک رویداد دقیقاً چه زمانی رخ داده است. به عنوان مثال، ممکن است بدانیم که تقاضای یک محصول حاصل در روز سه واحد است ولی برایمان مهم است که در چه زمانی از روز این محصول تقاضا دارد. در برخی موقعیت‌ها، زمان ممکن است به هیچ عنوان یک عامل شبیه‌سازی محسوب نشود مثل طراحی کنترل حالت ثابت یک کارخانه. با وجود این، در مسائلی که صف انتظار در تجارت الکترونیک مطرح باشد، اطلاع از زمان دقیق رسید کالا بسیار اهمیت دارد که آن را موقعیت وابسته به زمان می‌گویند.

Palisade.com (http://www.palisade.com/risk/monte_carlo_simulation.asp)

@RISK

@Risk

¹ Probabilistic Simulation

² Monte Carlo Simulation

³ Crisall Ball

⁴ Lilly

شبیه‌سازی رویداد گسسته به ایجاد و ساخت یک مدل از یک سیستم اشاره دارد که تعامل بین موجودیت‌های مختلف در آن مورد مطالعه و بررسی قرار می‌گیرد. ساده‌ترین مثال از این نوع، فروشگاه‌های متشکل از یک سرویس دهنده و مشتریان است. اگر مشتریان را به گونه‌ای مدلسازی کنیم که با سرعت‌های مختلف وارد فروشگاه می‌شوند و سرویس دهنده نیز با سرعت‌های مختلفی سرویس می‌دهد آنگاه می‌توانیم میانگین عملکرد سیستم، زمان انتظار، تعداد مشتریان در حال انتظار و غیره را تخمین بزنیم. اینگونه سیستم‌ها را مجموعه‌هایی از مشتریان، صف‌ها و سرویس‌دهنده‌ها در نظر می‌گیرند. هزاران کاربرد از مدل‌های شبیه‌سازی رویداد گسسته در حوزه‌های مهندسی، کسب‌وکار و غیره مستندسازی شده است. ابزارهای ساخت مدل‌های شبیه‌سازی رویداد گسسته برای مدت زمان طولانی وجود داشته‌اند ولی قابلیت‌های گرافیکی آنها برای ایجاد و درک نتایج اینگونه مدل‌های شبیه‌سازی در طول زمان تکامل یافته است. ما در بخش بعدی به بحث و بررسی بیشتر در خصوص این روش مدلسازی خواهیم پرداخت. مورد کاربردی ۸.۸ یک مثال از کاربرد اینگونه شبیه‌سازی در تحلیل پیچیدگی‌های زنجیره تأمین که از شبیه‌سازی دیداری استفاده می‌کند ارائه می‌کند.

جدول ۸.۵ توزیع‌های احتمال گسسته و پیوسته

تقاضای روزانه	احتمال گسسته	احتمال پیوسته
۵	۰.۱	تقاضای روزانه دارای توزیع نرمال با میانگین ۷ و انحراف استاندارد ۱.۲ است.
۶	۰.۱۵	
۷	۰.۳۰	
۸	۰.۲۵	
۹	۰.۲۰	

کوزان با استفاده از شبیه‌سازی، زنجیره‌ی تأمین انرژی تجدیدپذیر خود را بهبود می‌بخشد

مورد کاربردی ۸.۷

مقدمه

کوزان^۱ یک شرکت خوشه‌ای در برزیل است که در سرتاسر جهان فعالیت دارد. یکی از فعالیت‌های اصلی آن پرورش و فراوری نیشکر است. نیشکر علاوه بر اینکه منبع مهم شکر محسوب می‌شود در حال حاضر منبع اصلی اتانول که ماده تشکیل دهنده اصلی انرژی تجدیدپذیر نیز هست محسوب می‌شود. به دلیل تقاضای روزافزون انرژی تجدیدپذیر، تولید اتانول به فعالیت اصلی کوزان تبدیل شده است که در حال حاضر دو تصفیه‌خانه و ۱۸ کارخانه تولیدی و البته میلیون‌های هکتار مزرعه‌ی نیشکر را تحت فعالیت دارد. بر اساس داده‌های جدید، این شرکت بیش از ۴۴ میلیون تن نیشکر را فراوری کرده و بیش از ۱/۳ میلیارد لیتر اتانول و ۳/۳ میلیون تن شرکت را تولید کرده است. همانطور که تصور می‌کنید، عملیاتی در این مقیاس دارای زنجیره‌ی تأمین پیچیده‌ای خواهد بود. به همین دلیل از تیم لجستیک خواسته شد تا توصیه‌هایی را در خصوص موارد زیر به مدیریت ارشد ارائه دهد:

- تعداد بهینه‌ی وسایل نقلیه‌ی مورد نیاز ناوگان برای حمل و نقل نیشکر به مراکز فراوری برای حفظ سرمایه را تعیین کنند.
- پیشنهادات در خصوص نحوه‌ی افزایش ظرفیت واقعی نیشکر ورودی در مراکز فراوری را ارائه دهند.
- مشکلات گلوگاهی تولید را برای بهبود جریان نیشکر شناسایی کنند.

متدولوژی/راه‌حل

تیم لجستیک با استفاده از نرم‌افزار سیمیو یک مدل شبیه‌سازی پیچیده از زنجیره‌ی تأمین کوزان را برای رسیدگی به این مشکلات ایجاد کرد. طبق خلاصه گزارش سیمیو، «در یک دوره‌ی زمانی سه ماهه، مهندسان جدیدالاستخدام به گردآوری داده‌های میدانی پرداخته و مساعدت‌های مدل‌سازی و آموزشی را از گروه مشاوره‌ی پاراگون سان پالو^۲ دریافت کردند.» برای مدل‌سازی عملیات کشاورزی به منظور تحلیل مسیر حرکت نیشکرهای درون شده تا مراکز تولید و فراوری، اهداف مدل شامل جزئیات ناوگان حمل و نقل محصول نیشکر تا یونیتی کوستا پینتو، ظرفیت واقعی دریافت مراکز، گلوگاه‌ها و نقاط بهیو جریان CCT نیشکر بوده است.

پارامترهای مدل به شرح زیر هستند:

متغیرهای ورودی: ۳۲

متغیرهای خروجی: ۳۹

متغیرهای کمکی: ۹۲

موجودیت‌های متغیر: ۸

جداول ورودی: ۱۹

روزهای شبیه‌سازی شده: ۲۴۰ (فصل اول)

تعداد موجودیت‌ها: ۱۲

مزایا و منافع

تحلیل‌های انجام شده توسط این مدل‌های سیمیو، دید خوبی از ریسک تولید در یک دوره‌ی زمانی ۲۴۰ روزه را تحت شرایط مختلف عدم قطعیت ارائه کرد. با تحلیل گلوگاه‌های مختلف و روش‌های پیشگیری از این سناریوها، شرکت توانست تصمیمات بهتری را اتخاذ کرده و بیش از ۵۰۰,۰۰۰ دلار فقط از محل این مدل‌سازی صرفه‌جویی کند.

از این مطالعه‌ی موردی چه می‌توانیم بیاموزیم؟

این داستان کاربردی کوتاه، ارزش بکارگیری شبیه‌سازی در مسئله‌ای را نشان می‌دهد که ساخت یک مدل بهینه‌سازی در آن می‌تواند دشوار باشد. با بکارگیری یک مدل شبیه‌سازی رویداد گسسته و شبیه‌سازی تعاملی دیداری می‌توان تأثیر وقفه‌های ناشی از مشکلات ناوگان، خرابی‌های غیرمنتظره در کارخانه و غیره در زنجیره‌های تأمین را تجسم کرده و راهکارهای اصلاحی را در نظر گرفت.

¹ Cosan

² Paragon, San Palo

سوالاتی برای مورد بحث و بررسی

۱. در جابجایی نیشکر از مزارع تا کارخانجات تولید برای تولید شکر و اتانول چه وقفه‌هایی ممکن است در زنجیره‌ی تأمین رخ دهد؟
۲. چه نوع برنامه‌ریزی‌ها و پیش‌بینی‌های پیشرفته‌ای می‌توانند در پیشگیری از اینگونه وقفه‌ها مفید باشند؟

Sources: Compiled from Wikipedia contributors, Cosan, *Wikipedia, The Free Encyclopedia*, <https://en.wikipedia.org/w/index.php?title=Cosan&oldid=713298536> (accessed July 10, 2016); Agricultural Operations Simulation Case Study: Cosan, <http://www.simio.com/case-studies/Cosan-agricultural-logistics-simulation-software-case-study/agricultural-simulation-software-case-study-video-cosan.php> (accessed July 2016); Cosan Case Study: Optimizing agricultural logistics operations, <http://www.simio.com/case-studies/Cosan-agricultural-logistics-simulation-software-case-study/index.php> (accessed July 2016).

تصمیم‌گیرندگانی که از شبیه‌سازی تعاملی دیداری استفاده می‌کنند معمولاً از نتایج‌شان پشتیبانی کرده و به آن اطمینان دارند. شبیه‌سازی تعاملی دیداری معمولاً از نمایشگرهای گرافیکی رایانه‌ای متحرک برای نمایش تأثیر تصمیمات مدیریتی مختلف استفاده می‌کنند که با گرافیک‌های معمولی از این نظر تفاوت دارند که کاربر می‌تواند فرایند تصمیم‌گیری را تغییر داده و نتایج این کار را مشاهده کند. یک مدل دیداری، یک گرافیک است که نه فقط به عنوان یک وسیله‌ی ارتباطی، بلکه به عنوان جزء لاینفک تصمیم‌گیری یا حل مسئله بکار می‌رود. برخی افراد بهتر از دیگران به نمایشگرهای گرافیکی واکنش نشان می‌دهند و این نوع تعامل می‌تواند به مدیران در اطلاع از موقعیت تصمیم‌گیری کمک کند. شبیه‌سازی تعاملی دیداری می‌تواند بیانر سیستم‌های ایستا یا پویا باشد. مدل‌های ایستا یک تصویر دیداری از نتیجه‌ی یک آلترناتیو تصمیم را در یک زمان نشان می‌دهند. مدل‌های پویا سیستم‌هایی را نشان می‌دهند که در طول زمان تکامل پیدا می‌کنند و این تکامل از طریق پویانمایی نشان داده می‌شود. آخرین فناوری‌های شبیه‌سازی دیداری با مفهوم واقعیت مجازی در هم آمیخته‌اند که در آن یک دنیای مجازی برای اهداف متعددی از آموزش گرفته تا سرگرمی برای مشاهده‌ی داده‌ها در یک چشم‌انداز مصنوعی ایجاد می‌شود. به عنوان مثال، ارتش ایالات متحده از سیستم‌های شبیه‌سازی تعاملی دیداری به گونه‌ای استفاده می‌کند که نیروهای زمینی بتوانند با زمین یا شهر سریعاً آشنا شده و خودشان را وفق دهند. خلبان‌ها نیز از شبیه‌سازی تعاملی دیداری برای آشنایی با اهداف و از طریق شبیه‌سازی حملات استفاده می‌کنند. نرم‌افزار شبیه‌سازی تعاملی دیداری می‌تواند شامل مختصات جی‌آی‌اس نیز باشد.

مدل‌های تعاملی دیداری در تصمیمات متعدد مدیریت عملیات برای پشتیبانی تصمیم بکار گرفته شده‌اند. این روش شامل پرایمینگ (مثل پرایمینگ پمپ آب) یک مدل تعاملی دیداری از یک کارخانه (یا شرکت) در وضعیت فعلی آن است. سپس مدل به سرعت بر روی رایانه اجرا شده و به مدیران اجازه می‌دهد تا نحوه‌ی عملکرد کارخانه را در آینده مشاهده کنند.

مدیریت خط انتظار (صف) مثال خوبی از مدل‌های تعاملی دیداری است. چنین سیستم پشتیبانی تصمیم معمولاً چندین سنج‌های عملکردی را برای آلترناتیوهای تصمیم مختلف (مثلاً زمان انتظار در سیستم) محاسبه می‌کند. مسائل صف پیچیده نیازمند شبیه‌سازی هستند. مدل‌های تعاملی دیداری می‌توانند اندازه‌ی صف را که در طول دفعات اجرای شبیه‌سازی تغییر می‌کند نشان داده و پاسخ‌های سوالات شرطی را با توجه به تغییر متغیرهای ورودی به صورت گرافیکی نمایش دهند. مورد کاربرد ۸.۹ یک مثال از یک شبیه‌سازی دیداری را که برای بررسی کاربردهای فناوری شناسایی فرکانس رادیویی در توسعه‌ی قواعد زمانبندی جدید در یک شرکت تولیدی بکار رفته است نشان می‌دهد.

روش مدلسازی تعاملی دیداری را می‌توان به همراه هوش مصنوعی نیز بکار برد. یکپارچه‌سازی دو تکنیک، چندین قابلیت را از توانایی ایجاد سیستم‌های گرافیکی تا یادگیری در مورد دینامیک‌های سیستم می‌افزاید. این سیستم‌ها، به ویژه آنهایی که برای مقاصد نظامی و بازی‌های ویدیویی ساخته شده‌اند دارای شخصیت‌های «متفکر» هستند که می‌توانند با یک سطح هوشمندی نسبتاً بالا در تعامل با کاربران رفتار کنند.

شبیه‌سازی دیداری

صدها بسته‌ی نرم‌افزاری شبیه‌سازی برای موقعیت‌های مختلف تصمیم‌گیری وجود دارد. بسیاری از آنها به صورت سیستم‌های مبتنی بر وب اجرا می‌شوند. *آرام‌اس تودی*^۱ یک نشریه‌ی دوره‌ای از نرم‌افزارهای شبیه‌سازی را منتشر می‌کند. یکی از شماره‌های اخیر آن (اکتبر ۲۰۱۸) در آدرس <https://www.informs.org/ORMS-Today/Public-Articles/October-Volume-44-Number-5/Simulation-Software-Survey-Simulation-new-and-improved-reality-show> رایانه‌های شخصی شامل آنالیتیکا^۲ (لومینا دیسیژن سیستمز^۳، lumina.com) و افزونه‌ی کریستال بال اکسل (که اکنون توسط شرکت اوراکل تحت عنوان اوراکل کریستال بال، oracle.com فروخته می‌شود) و *RISK@* (شرکت پالیساد، palisade.com) هستند. یکی از نرم‌افزارهای تجاری مهم برای شبیه‌سازی رویداد گسسته، آرینا (شرکت راکول اینتل^۴، arenasimulation.com) است. توسعه‌دهندگان اصلی آرینا^۵ در حال حاضر نرم‌افزار سیمو (simio.com) که یک نرم‌افزار شبیه‌سازی تعاملی دیداری کاربر پسند می‌باشد را توسعه داده‌اند. یک نرم‌افزار محبوب دیگر شبیه‌سازی تعاملی دیداری رویداد گسسته، اکستندسیم^۶ (extendsim.com) نام دارد. *اس‌ای‌اس* یک بسته‌ی نرم‌افزار تحلیلی گرافیکی به نام *جی‌ام‌پی*^۷ را که شامل یک کامپوننت شبیه‌سازی است را ارائه کرده است.

بهبود تصمیمات زمانبندی کارگاهی از طریق شناسایی فرکانس رادیویی: یک ارزیابی مبتنی بر شبیه‌سازی

مورد کاربرد ۸.۹

یک ارائه دهنده‌ی خدمات تولیدی قطعات الکترومکانیکی و آپتیکال به دنبال افزایش کارایی تصمیمات زمانبندی کارگاهی خود به دلیل مشکلات و مسائل زیر است:

¹ ORMS Today

² Analytica

³ Lumina Decision Systems

⁴ Rockwell Intl.

⁵ Arena

⁶ ExtendSim

⁷ JMP

- هیچ سیستمی برای ثبت زمان ورود اقلام در جریان ساخت یا خروج آنها از ایستگاه‌های عملیاتی و مدت زمانی که اقلام در جریان ساخت در هر ایستگاه توقف می‌کنند وجود ندارد.
- سیستم جاری قادر به پایش یا ردیابی لحظه‌ای حرکت و جابجایی هر قلم در جریان ساخت در خط تولید نمی‌باشد. در نتیجه شرکت با دو مسئله‌ی اساسی در خط تولیدش مواجه است: کارهای ناتمام زیاد و هزینه‌های بالای اضافه‌کاری برای برآورده کردن تقاضا. علاوه بر این، بخش بالادستی قادر به پاسخگویی سریع به مشکلات غیرمنتظره از قبیل تغییر تقاضا یا کمبود مواد و بازرنگری جدول زمانبندی به شکل اقتصادی نیست. شرکت به فکر پیاده‌سازی شناسایی فرکانس رادیویی در خط تولید خود است با وجود این، شرکت نمی‌داند که آیا چنین سرمایه‌گذاری سنگینی برای افزودن تراشه‌های آر اف آی دی بر روی جعبه‌های تولید، نصب بارکدخوان‌های آر اف آی دی در سرتاسر خط تولید و البته سیستم‌هایی برای پردازش این اطلاعات منجر به منافع بزرگی خواهد شد یا خیر. بنابراین یک مسئله این است که به دنبال هرگونه تغییری در زمانبندی تولید جدید باشیم که از سرمایه‌گذاری در زیرساخت‌های آر اف آی دی حاصل می‌شود.

متدولوژی

با توجه به اینکه بکارگیری هرگونه سیستم جدیدی در سیستم تولید فیزیکی ممکن است بسیار هزینه‌بر یا حتی اخلال‌گر باشد، یک مدل شبیه‌سازی رویداد گسسته برای ارزیابی نحوه‌ی ردیابی و قابلیت ردیابی از طریق آر اف آی دی که موجب تسهیل فعالیت‌های زمانبندی تولید کارگاهی شود توسعه داده شد. یک قاعده‌ی زمانبندی مبتنی بر شفافیت^۱ که از سیستم‌های قابلیت ردیابی بلادرنگ برای ردیابی اقلام در حال ساخت، قطعات و اجزا و مواد اولیه در عملیات کارگاهی بهره می‌برد پیشنهاد شد. یک روش شبیه‌سازی برای بررسی مزایای بکارگیری قاعده‌ی VBS در برابر قواعد زمانبندی کلاسیک بکار گرفته شد: قواعد خروج به ترتیب ورود و زودترین زمان ارسال. یک مدل شبیه‌سازی با استفاده از سیمو توسعه یافت. سیمو یک بسته نرم‌افزاری مدل‌سازی شبیه‌سازی سه بُعدی است که از یک رویکرد شی‌گرا برای مدل‌سازی استفاده کرده و در بسیاری از حوزه‌ها از قبیل کارخانجات، زنجیره‌های تأمین، مراقبت سلامت، فرودگاه‌ها و سیستم‌های خدماتی بکار گرفته شده است.

شکل ۸/۱۳ یک اسکرین‌شات از پنل رابط کاربری این خط تولید را در سیمو نشان می‌دهد. تخمین‌های پارامتری بکار رفته برای وضعیت اولیه‌ی این مدل شبیه‌سازی شامل پیش‌بینی و تقاضای هفتگی، جریان فرایند، تعداد ایستگاه‌های کاری و تعداد اپراتورهای کارگاه و زمان عملیات در هر ایستگاه می‌باشند. علاوه بر این، پارامترهای برخی از داده‌های ورودی از قبیل زمان تگ‌گذاری آر اف آی دی، زمان بازبینی اطلاعات یا زمان به‌روزرسانی سیستم نیز برای مطالعه‌ی پایلوت و بر اساس نظرات خبرگان تخمین زده می‌شود. شکل ۸/۱۴ فرایند مدل شبیه‌سازی را نشان می‌دهد که دستورات خاص شبیه‌سازی در آن پیاده‌سازی و کدگذاری شده‌اند. شکل ۸/۱۵ و ۸/۱۶ یک گزارش استاندارد و گزارش شبیه پیووت که روشی برای بسیار سریع برای یافتن نتایج آماری خاص مثل مقادیر میانگین، درصد، کل، بیشینه یا کمینه‌ی متغیرهای تخصیص داده شده و ثبت شده به عنوان خروجی یک مدل شبیه‌سازی را نشان می‌دهد.

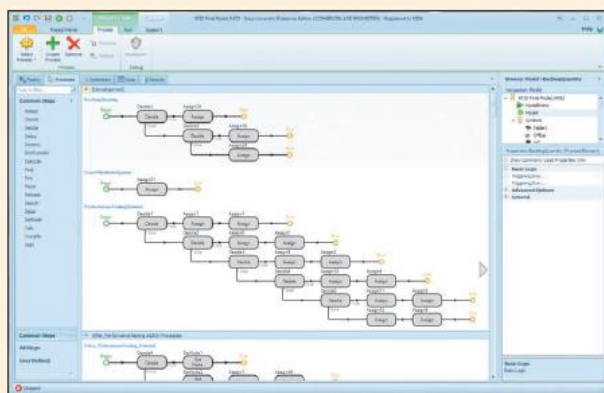
نتایج

نتایج شبیه‌سازی بیانگر این هستند که یک قاعده‌ی زمانبندی مبتنی بر آر اف آی دی در مقایسه با قواعد زمانبندی متداول از لحاظ زمان پردازش، زمان تولید، بهره‌وری منابع، کارهای ناتمام و بهره‌وری عملکرد بهتری دارد. شرکت می‌تواند این مزایای بهره‌وری را گرفته و برای تصمیمات سرمایه‌گذاری نهایی تحلیل‌های سود/هزینه را انجام دهد.

¹ Visibility-based Scheduling (VBS)



شکل ۸.۱۳ رابط کاربری سیستم شبیه‌سازی سیمو
 منبع: Used with permission from Simio LLC.



شکل ۸.۱۴ ویوی فرایندی مدل شبیه‌سازی
 منبع: Used with permission from Simio LLC.

Scenario	Date Source	Category	Run Date	Run Status	Maximum	Minimum
Scenario: Interactive Run						
Scenario: Scenario 1	Scenario 1	Scenario 1	1/1/2020	Run	100	100
Scenario: Scenario 2	Scenario 2	Scenario 2	1/1/2020	Run	100	100
Scenario: Scenario 3	Scenario 3	Scenario 3	1/1/2020	Run	100	100
Scenario: Scenario 4	Scenario 4	Scenario 4	1/1/2020	Run	100	100
Scenario: Scenario 5	Scenario 5	Scenario 5	1/1/2020	Run	100	100
Scenario: Scenario 6	Scenario 6	Scenario 6	1/1/2020	Run	100	100
Scenario: Scenario 7	Scenario 7	Scenario 7	1/1/2020	Run	100	100
Scenario: Scenario 8	Scenario 8	Scenario 8	1/1/2020	Run	100	100
Scenario: Scenario 9	Scenario 9	Scenario 9	1/1/2020	Run	100	100
Scenario: Scenario 10	Scenario 10	Scenario 10	1/1/2020	Run	100	100
Scenario: Scenario 11	Scenario 11	Scenario 11	1/1/2020	Run	100	100
Scenario: Scenario 12	Scenario 12	Scenario 12	1/1/2020	Run	100	100
Scenario: Scenario 13	Scenario 13	Scenario 13	1/1/2020	Run	100	100
Scenario: Scenario 14	Scenario 14	Scenario 14	1/1/2020	Run	100	100
Scenario: Scenario 15	Scenario 15	Scenario 15	1/1/2020	Run	100	100
Scenario: Scenario 16	Scenario 16	Scenario 16	1/1/2020	Run	100	100
Scenario: Scenario 17	Scenario 17	Scenario 17	1/1/2020	Run	100	100
Scenario: Scenario 18	Scenario 18	Scenario 18	1/1/2020	Run	100	100
Scenario: Scenario 19	Scenario 19	Scenario 19	1/1/2020	Run	100	100
Scenario: Scenario 20	Scenario 20	Scenario 20	1/1/2020	Run	100	100

شکل ۸.۱۵ نمای گزارش استاندارد
 منبع: Used with permission from Simio LLC.

...
...
...
...
...
...
...
...
...

...

- ... ۱۱۱
- ... ۱۱۲
- ... ۱۱۳
- ... ۱۱۴
- ... ۱۱۵
- ... ۱۱۶
- ... ۱۱۷
- ... ۱۱۸
- ... ۱۱۹
- ... ۱۲۰
- ... ۱۲۱
- ... ۱۲۲
- ... ۱۲۳
- ... ۱۲۴
- ... ۱۲۵
- ... ۱۲۶

...

...
...
...

... **teradatauniversitynetwork.com** ... ۲۱.

... ۲۲.

... a.

... b.

... c.

... ۲۳.

... **lindo.com** ...

... CC-8 ...

... CC-8 ...

... ۲۴.

... کرده است ...

... ۲۵.

...
...
...

...
...
...

1 Chocolate Lucky Charms
2 Butterfinger bars
3 Milk Duds
4 Cap'n Crunch
5 Chips Ahoy

--	--	--	--	--	--	--	--

Text Analytics

- .۶
- .۷
- .۸
- .۹
- Text Analytics**
- .۱۰

Downloads

- .۱۰
- statsoft.com**
- .۱۱
- sas.com**
- .۱۲
- ibm.com**
- .۱۳
- teradata.com**
- .۱۴

clarabridge.com	۱۵
kdnuggets.com	۱۶
	۱۷
attensity.com	۱۸

کلان داده‌ها، رایانش ابری و تحلیل مکانی: مفاهیم و ابزارها

کلان‌داده‌ها که برای بسیاری از مردم معانی بسیاری دارد یک مُد تکنولوژیکی زودگذر نیست و تبدیل به یک اولویت کسب‌وکار شده است که توان بالقوه‌ی بالایی برای تغییر عمیق چشم‌انداز رقابتی در اقتصاد جهانی امروز را دارد. علاوه بر ارائه‌ی راه‌حل‌های نوآورانه برای چالش‌های مداوم کسب‌وکارها، کلان‌داده‌ها و تحلیل‌ها روش‌های نوینی را برای تغییر شکل فرایندها، سازمان‌ها، کل صنایع و حتی جامعه ارائه می‌کنند. اما پوشش رسانه‌ای گسترده موجب شده است تا به راحتی نتوانیم واقعیت را از مجاز تشخیص دهیم. هدف ما در این فصل، پوشش جامع موضوع کلان‌داده‌ها، فناوری‌های توانمندساز آن و مفاهیم تحلیلی مرتبط برای کمک به درک قابلیت‌ها و محدودیت‌های این فناوری نوظهور است. این فصل را با یک تعریف و مفاهیم مربوط به کلان‌داده‌ها آغاز کرده و سپس

- **۹.۱. پرونده‌ی آغازین:** تحلیل ریزش مشتری در یک شرکت مخابراتی با استفاده از روش‌های کلان‌داده‌ها
- **۹.۲. تعریف کلان‌داده‌ها**
- **۹.۳. اصول و مبانی تحلیل کلان‌داده‌ها**
- **۹.۴. فناوری‌های کلان‌داده‌ها**
- **۹.۵. کلان‌داده‌ها و انبارسازی داده‌ها**
- **۹.۶. تحلیل درون حافظه‌ای و آپاچی اسپارک**
- **۹.۷. کلان‌داده‌ها و تحلیل جریان**
- **۹.۸. پلتفرم‌ها و عرضه‌کنندگان کلان‌داده‌ها**
- **۹.۹. رایانش ابری و تحلیل کسب‌وکار**
- **۹.۱۰. تحلیل مکانی برای سازمان‌ها**

جزئیات فنی فناوری‌های توانمندساز از جمله هَدوپ، مَپ‌رِدیوس و نوَاس کیو_ال را توضیح می‌دهیم. همچنین یک تحلیل مقایسه‌ای بین انبارسازی داده‌ها و تحلیل‌های کلان‌داده‌ها ارائه می‌کنیم. بخش آخر این فصل نیز به تحلیل استریم اختصاص دارد که یکی از امیدبخش‌ترین گزاره‌های ارزشی تحلیل کلان‌داده‌ها است. این فصل از بخش‌های زیر تشکیل شده است:

- ۹.۱. پرونده‌ی آغازین:** تحلیل ریزش مشتری در یک شرکت مخابراتی با استفاده از روش‌های کلان‌داده‌ها
- ۹.۲. تعریف کلان‌داده‌ها**
- ۹.۳. اصول و مبانی تحلیل کلان‌داده‌ها**
- ۹.۴. فناوری‌های کلان‌داده‌ها**
- ۹.۵. کلان‌داده‌ها و انبارسازی داده‌ها**
- ۹.۶. تحلیل درون حافظه‌ای و آپاچی اسپارک**
- ۹.۷. کلان‌داده‌ها و تحلیل جریان**
- ۹.۸. پلتفرم‌ها و عرضه‌کنندگان کلان‌داده‌ها**
- ۹.۹. رایانش ابری و تحلیل کسب‌وکار**
- ۹.۱۰. تحلیل مکانی برای سازمان‌ها**

یک شرکت مخابراتی به دنبال ریشه‌یابی علل ریزش مشتریان و عدم تمایل آنها به دریافت خدمات مخابراتی است. ریزش مشتریان در صنعت مخابرات امری متداول است. با وجود این، این شرکت مشتریان خود را با نرخ بسیار بالایی از دست می‌داد. این پدیده دلایل و راه‌حل‌های بالقوه‌ی متعددی داشت. مدیریت شرکت متوجه شد که بسیاری از لغوها شامل ارتباطات بین دپارتمان خدمات مشتری و مشتریان بوده است. تا به اینجا، یک تیم ضربتی متشکل از اعضای دفتر روابط مشتریان و دپارتمان فناوری اطلاعات برای بررسی دقیق‌تر مسئله تشکیل شد. وظیفه‌ی آنها بررسی این موضوع بود که چگونه ریزش مشتریان را بر اساس تحلیل الگوهای ارتباطی مشتریان بتوان کاهش داد (آساموا، شادرا، زاده و کالگوترا ۲۰۱۶).

هر زمان که یک مشتری مشکل در مورد مسائلی از قبیل صورتحساب، برنامه و کیفیت تماس داشته باشد به چندین روش می‌تواند با شرکت تماس برقرار کند از جمله مرکز تماس، وب سایت شرکت و مراکز خدمات. مشتریان می‌توانند حساب‌شان را از یکی از این روش‌ها لغو کنند. شرکت می‌خواهد ببیند که آیا از تحلیل این تعاملات با مشتریان می‌تواند به بینش و درک مناسبی از سوالاتی که مشتریان پرسیده‌اند یا کانال‌های تماس که پیش از لغو حساب‌شان استفاده کرده‌اند دست یابد. داده‌های تولید شده از این تعاملات هم به صورت متنی و هم صوتی هستند. بنابراین، شرکت باید تمامی این داده‌ها در یک مکان گردآوری و ترکیب کند. شرکت از پلتفرم‌های رایج برای مدیریت داده‌ها استفاده کرده ولی به زودی متوجه شد که از تنوع کافی برای تحلیل‌های داده‌ی پیشرفته در سناریوهایی که چندین فرمت داده از چندین منبع وجود دارد برخوردار نیستند (توسو، شائو و آنتونی ۲۰۱۰).

در تحلیل این داده‌ها دو نوع چالش اساسی وجود دارد: تعدد منابع داده که منجر به تنوع داده‌ها می‌شود و حجم زیاد داده‌ها.

۱. **تعدد منابع داده‌ها:** مشتریان می‌توانند با دسترسی به حساب‌هایشان بر روی وب سایت شرکت به شرکت متصل شوند که به شرکت اجازه می‌دهد تا اطلاعات سوابق فعالیت مشتریان در وب‌سایت را تولید کند. ردیابی سوابق وب سایت به شرکت اجازه داد تا ببیند که آیا مشتری طرح جاری خود را بازنگری کرده، شکایتی ارسال کرده یا صورتحساب خود را به صورت آنلاین چک کرده است یا خیر. در مرکز خدمات مشتریان، مشتریان می‌توانند یک شکایت خدمات را ثبت کنند، یک تغییر در طرح را درخواست دهند یا سرویس خود را لغو کنند. این فعالیت‌ها در سیستم تراکنش شرکت ثبت شده و سپس وارد پایگاه داده‌های تحلیلی می‌شوند. نهایتاً، یک مشتری می‌تواند با مرکز خدمات مشتری به صورت تلفنی تماس گرفته و کارشان را به همان روش حضوری در مرکز خدمات مشتریان انجام دهند. اینگونه تراکنش‌ها می‌توانند شامل استعلام تراز حساب یا اقدام به لغو به طرح باشد. سوابق تماس در داخل یک سیستم به همراه پرونده دلایل تماس مشتری در دسترس هستند. برای تحلیل معنادار، مجموعه تک تک داده‌ها را باید به فرمت‌های ساخت یافته مشابه تبدیل کرد.

¹ Asamoah, Shadra, Zadeh, Kalgotra

² Tusso, Shao, Anthony

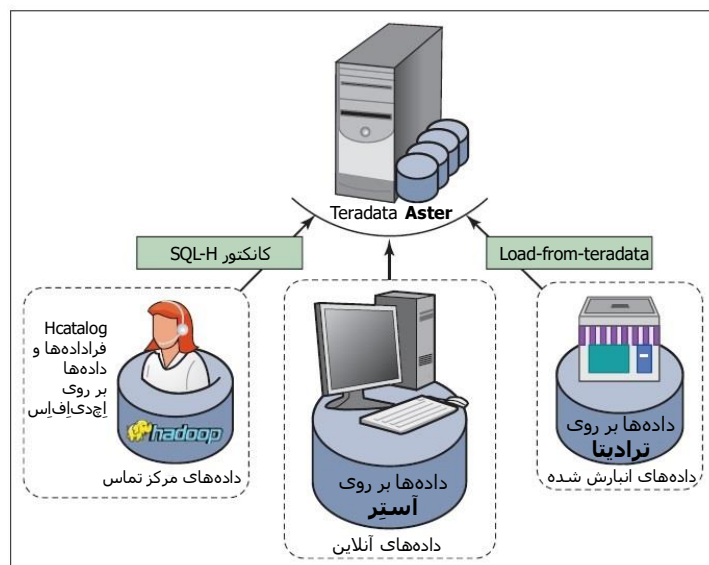
۲. **حجم داده‌ها.** چالش دوم، حجم بسیار زیاد داده‌ها از منبع است که باید استخراج، پاکسازی، ساخت‌یافته و تحلیل شوند. اگرچه پروژه‌های تحلیل داده قبلی اکثراً از مجموعه داده‌های کوچک برای تحلیل استفاده می‌کنند شرکت تصمیم گرفت از چندین منبع داده‌ی گوناگون و حجم زیادی از داده‌های ثبت شده برای تولید بینش‌های متعدد استفاده کند.

یک روش تحلیلی که می‌تواند از تمامی کانال‌ها و منبع داده‌های هرچند بزرگ استفاده کند، از پتانسیل لازم برای تولید بینش‌های غنی و عمیق از داده‌ها برای کمک به کاهش ریزش مشتریان برخوردار است.

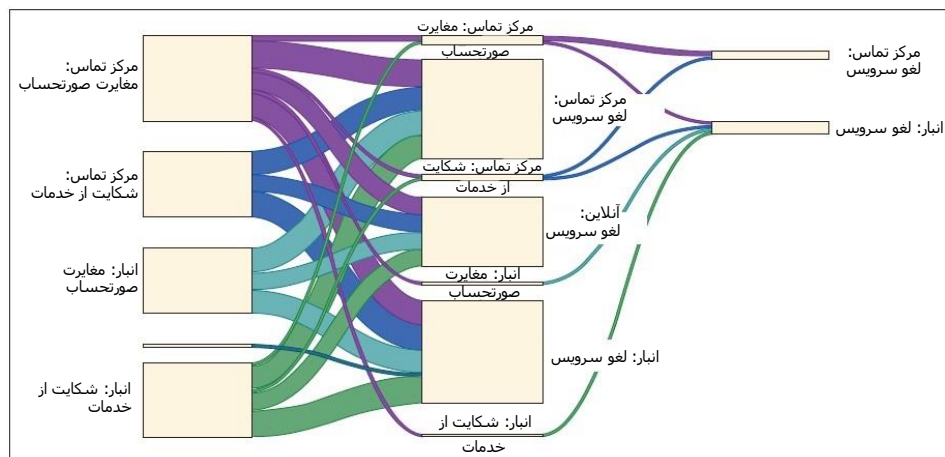
از معماری یکپارچه‌ی کلان‌داده ترادیتا و نتج^۱ (که قبلاً تحت عنوان ترادیتا آستر عرضه می‌شد) برای مدیریت و تحلیل داده‌های چندساختاری بزرگ و حجیم بکار گرفته شد. ما در بخش ۹.۸ به معرفی ترادیتا و نتج خواهیم پرداخت. یک نمای شماتیک از داده‌های ترکیب شده در شکل ۹.۱ نشان داده شده است. سه جدول بر اساس هر منبع داده به همراه هر جدول که حاوی متغیر زیر است ایجاد شد: شماره شناسه مشتری، کانال ارتباطی، تاریخ/زمان و اقدام اتخاذ شده. پیش از لغو نهایی یک سرویس، متغیر اقدام اتخاذ شده می‌تواند یک یا چند مورد از ۱۱ گزینه‌ی موجود باشد: ارائه‌ی یک مغایرت در صورت‌حساب، درخواست ارتقاء یک طرح، درخواست تقلیل یک طرح، به‌روزرآوری پروفایل، مشاهده‌ی خلاصه حساب، دسترسی به پشتیبانی مشتری، مشاهده‌ی صورت‌حساب، بازنگری قرارداد، دسترسی به کارکرد مکان‌یاب انبار بر روی وب‌سایت، دسترسی به بخش پرسشگان بر روی وب‌سایت یا جستجوی دستگاه‌ها و ادوات. هدف از تحلیل، یافتن متداول‌ترین مسیری بود که منجر به لغو نهایی یک سرویس می‌شد. داده‌ها به شکلی بخش‌بندی^۲ شدند تا رشته‌ای از رویدادها را شکل دهند که یک مشتری خاص را در یک دوره زمانی تعریف شده شامل می‌شد (دوره زمانی ۵ روزه از طریق تمامی کانال‌های ارتباطی). نهایتاً، تابع توالی زمانی مسیر n و نتج (که در چارچوب اس‌کیوال-مپ‌دیوس عملیاتی شده است) برای تحلیل روندهای متداول که منجر به لغو شدن یک طرح می‌شود بکار گرفته شد.

¹ Teradata Vantage

² Sessionize



شکل ۹.۱ تلفیق چندین منبع داده در ترادیتا و نتج



شکل ۹.۲ ۲۰ مصورسازی بر تر مسیرهها منبع: Teradata Corp

نتایج اولیه نشان داد که چندین مسیر می‌توانند منجر به یک درخواست لغو سرویس شوند. شرکت هزاران مسیر را شناسایی کرد که مشتری ممکن است برای لغو یک سرویس انتخاب کند. یک تحلیل فالوآپ برای شناسایی پرتکرارترین مسیرهای درخواست‌های لغو شناسایی شد که آن را مسیر طلایی نامیدند. بیست مسیر مهم که منجر به لغو سرویس در کوتاه مدت و بلند مدت می‌شوند شناسایی شدند. یک نمونه در شکل ۹.۲ نشان داده شده است. این تحلیل به شرکت کمک کرد تا یک مشتری را پیش از آنکه بخواهد سرویس خود را لغو کند شناسایی کرده و مشوق‌هایی را به او ارائه کرده یا حداقل برای رفع مشکل وی به نحوی اقدام کند که نهایتاً منجر به لغو سرویس مشتری نشود.

سوالاتی برای مرور پرونده‌ی آغازین

۳۷. لغو کردن سرویس‌های مشتریان چه مشکلی برای بقاء شرکت مخابراتی بوجود آورده بود؟
۳۸. موانع فنی ناشی از ماهیت و ویژگی‌های داده‌های این شرکت مخابراتی را شناسایی کرده و توضیح دهید.
۳۹. بخش‌بندی داده‌ها به چه معناست؟ چرا اینکار برای شرکت مخابراتی ضروری است؟
۴۰. در خصوص سایر مطالعات انجام شده که از مدل‌های ریزش مشتریان استفاده شده است تحقیق کنید. از چه نوع متغیرهایی در آن مطالعات استفاده شده است؟ چه تفاوتی با این پرونده آغازین دارند؟
۴۱. علاوه بر ترادیتا و نتج، پلتفرم‌های محبوب دیگر برای تحلیل کلان‌داده‌ها که می‌توانند موانع مطرح شده در این مطالعه را رفع کنند شناسایی کنید.

در تمامی مسائل کسب‌وکار نمی‌توان از پلتفرم تحلیل کلان‌داده‌ها استفاده کرد. این موقعیت بیانگر موردی است که استفاده از یک پلتفرم کلان‌داده را تضمین می‌کند. چالش اصلی، حول ویژگی‌های داده‌های مورد نظر است. سه نوع مجموعه داده‌های متفاوت از تعاملات مشتریان موجب بروز چالش در تحلیل شد. فرمت‌ها و فیلدهای داده‌های تولید شده در هر یک از این سیستم‌ها بسیار زیاد بود. و حجم داده‌ها نیز بسیار زیاد بود. این شرایط استفاده از یک پلتفرم را که از فناوری‌ها برای تحلیل حجم زیادی از داده‌ها در فرمت‌های گوناگون اجتناب‌ناپذیر می‌کرد.

اخیراً، ترادیتا بازاریابی آستر به عنوان یک محصول مجزا را متوقف کرده و تمامی امکانات و قابلیت‌های آستر را در محصول جدیدی به نام ترادیتا و نتج عرضه کرده است. اگرچه این تغییر تا حدودی بر نحوه‌ی توسعه‌ی نرم‌افزار تأثیر می‌گذارد ولی هنوز مثال بارزی از تلفیق و ترکیب انواع گوناگون داده‌ها برای تصمیم‌گیری کسب‌وکار است. شایان ذکر است که شرکت مخابراتی تمامی سوالات مطرح شده در مورد داده‌ها را با استراتژی کسب‌وکار سازمان همسو کرده است. این سوالات همچنین به نوع تحلیلی که انجام می‌شود نیز اشاره دارند. درک این نکته بسیار مهم است که برای هر کاربرد معماری کلان‌داده‌ها، استراتژی کسب‌وکار سازمان و تولید سوالات مرتبط از عوامل اساسی برای شناسایی نوع تحلیلی که باید انجام شوند محسوب می‌شوند.

Sources: D. Asamoah, R. Sharda, A. Zadeh, & P. Kalgotra. (2016). "Preparing a Big Data Analytics Professional: A Pedagogic Experience." In *DSI 2016 Conference*, Austin, TX. A. Thusoo, Z. Shao, & S. Anthony. (2010). "Data Warehousing and Analytics Infrastructure at Facebook." In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data* (p. 1013). doi: 10.1145/1807167.1807278.

استفاده از داده‌ها برای درک مشتریان/ارباب رجوع و عملیات کسب‌وکار برای حفظ (پیشبرد) رشد و سودآوری یکی از وظایف شدیداً چالش‌برانگیز در بنگاه‌ها و سازمان‌های امروزی محسوب می‌شود. با افزایش هرچه بیشتر داده‌های موجود در قالب‌ها و شکل‌های مختلف، پردازش به موقع داده‌ها به روش‌های سنتی و رایج تقریباً ناممکن شده است. امروز این پدیده را کلان‌داده می‌نامند که پوشش رسانه‌ای بسیار گسترده و توجه ویژه‌ای از جانب کاربران و متخصصان فناوری اطلاعات را جلب نموده است. نتیجه اینکه، کلان‌داده‌ها تبدیل به یک لغت باب روز در حوزه بازاریابی شده است به نحویکه بسیاری از کارشناسان و خبرگان این صنعت در مورد بکار نگرفتن این عبارت به طور کلی بحث کرده‌اند.

کلان داده‌ها برای افراد مختلف با پس‌زمینه‌ها و علائق مختلف معانی متفاوتی دارد. به لحاظ سنتی، عبارت کلان‌داده برای توضیح حجم انبوه داده‌های تحلیل شده توسط سازمان‌های بزرگ مثل گوگل یا پروژه‌های علوم پژوهشی در ناسا اطلاق می‌شد. اما برای اکثر کسب‌وکارها، این یک عبارت نسبی است: «کلان» به اندازه‌ی سازمان بستگی دارد. نکته در اینجا بیشتر به یافتن یک ارزش جدید در درون و بیرون منبع داده‌های رایج است. فراتر رفتن از مرزهای تحلیل داده‌ها منجر به کشف بینش‌ها و فرصت‌های جدید می‌شود و «کلان» به این بستگی دارد که کارتان را از کجا آغاز کرده و چگونه پیش ببرید. این توضیح معروف از کلان‌داده را در نظر بگیرید: کلان‌داده‌ها از محیط دسترسی پرکاربرد سخت‌افزاری و/یا قابلیت‌های ابزارهای نرم‌افزاری برای ضبط، مدیریت و پردازش آن در یک محدوده‌ی زمانی قابل قبول برای جامعه‌ی کاربران فراتر می‌رود. کلان‌داده‌ها^۱ به عبارتی محبوب و معروف برای توضیح رشد نمایی، در دسترس بودن و بکارگیری اطلاعات ساخت‌یافته و ساخت‌نیافته اطلاق می‌شود. مطالب بسیار زیادی در خصوص روند کلان‌داده‌ها و نقش آنها به عنوان پایه و اساس نوآوری، تمایز و رشد نوشته شده است. به دلیل چالش‌های فناوری در مدیریت حجم زیاد داده‌های بدست آمده از منابع متعدد و برخی اوقات با سرعت بالا، فناوری‌های نوین دیگری نیز برای غلبه بر این چالش‌های فناوری توسعه یافته‌اند. بکارگیری عبارت کلان‌داده معمولاً با اینگونه فناوری‌ها مرتبط است. به دلیل استفاده‌ی اولیه‌ی انبارش، اینگونه داده‌ها از طریق تحلیل به بینش‌هایی دست می‌یابند که به همین دلیل عبارت کلان‌داده به تحلیل کلان‌داده گسترش یافته است. اما این عبارت تبدیل به یک عبارت محتوا-ناسته شده است بدین معنی که برای افراد مختلف معانی متفاوتی دارد. چون هدف ما در اینجا معرفی مجموعه داده‌های بزرگ و توان بالقوه‌ی آنها در تولید بینش است ما از عبارت اصلی در این فصل استفاده می‌کنیم.

کلان‌داده‌ها از کجا می‌آیند؟ یک پاسخ ساده این است که «از همه جا». منابعی که زمانی که به خاطر محدودیت‌های فنی نادیده گرفته می‌شدند اکنون به عنوان معادن طلا شناخته شده‌اند. کلان‌داده‌ها ممکن است وبلاگ‌ها، شناسایی فرکانس رادیویی، سیستم‌های مکان‌یاب جغرافیایی، شبکه حسگرها، شبکه‌های اجتماعی، اسناد متنی اینترنتی، ایندکس‌های جستجوی اینترنتی، جزئیات پرونده‌های تماس، اخترشناسی، علوم جوی، زیست‌شناسی، ژنومیک، فیزیک هسته‌ای، آزمایشات بیوشیمیایی، پرونده‌های پزشکی، پژوهش‌های علمی، نظارت‌های نظامی، آرشیو عکس‌ها، آرشیو ویدئوها و رویکردهای تجارت الکترونیک در مقیاس بزرگ بدست آیند. کلان‌داده‌ها موضوع جدیدی نیستند. آنچه جدید است این است که تعریف و ساختار کلان‌داده‌ها دائماً تغییر می‌کند. شرکت‌ها از اوایل دهه ۹۰ و با پیدایش پایگاه داده‌های تحلیلی به انبارسازی و تحلیل حجم انبوهی از داده‌ها پرداخته‌اند. اگرچه زمانی پایگاه داده‌های تحلیلی مترادف با ترابایت بود امروزه با اگزابایت تعریف می‌شوند و نرخ رشد حجم داده‌ها نیز همراه با تلاش سازمان‌ها برای انبارش و تحلیل جزئیات بیشتری از تراکنش‌ها و همچنین داده‌های تولید شده توسط وب و ماشین‌ها برای دستیابی به درک بهتری از رفتار مشتری و محرک‌های کسب‌وکار همچنان افزایش می‌یابد.

بسیاری (دانشگاهیان و تحلیلگران صنعت) معتقدند که «کلان‌داده‌ها» یک نام بی‌مسمی است و نام و معنای آن دقیقاً یکسان نیستند. بدین معنا که «کلان‌داده» فقط به معنای «کلان» نیست. حجم انبوه داده‌ها فقط یکی از چندین ویژگی کلان‌داده‌ها است که از آن جمله می‌توان به تنوع، سرعت، درستی، تغییرپذیری و گزاره ارزشی اشاره کرد.

¹ Big Data

کلان داده‌ها معمولاً با سه «V» تعریف می‌شوند: حجم، تنوع و سرعت ولی علاوه بر اینها ما شاهد این هستیم که برخی از ارائه‌دهندگان خدمات کلان داده‌ها موارد دیگری از قبیل درستی (آی‌بی‌ام)، تغییرپذیری (ساس) و گزاره ارزشی نیز بکار می‌برند. **حجم** مشخصاً متداول‌ترین ویژگی کلان داده‌ها است. عوامل بسیاری از قبیل داده‌های تراکنشی که طی چندین سال انبارش شده‌اند، داده‌های متنی که دائماً از طریق رسانه‌های اجتماعی جریان می‌یابند، حجم فزاینده‌ی داده‌های حسگرها که گردآوری می‌شوند، داده‌های آراف‌آی‌دی و جی‌پی‌اس که به صورت خودکار تولید می‌شوند و ... موجب افزایش نمایی حجم داده‌ها شده‌اند. در گذشته، حجم بسیار زیاد داده‌ها مشکلاتی را چه به لحاظ فنی و چه مالی در انبارش به وجود می‌آورد؛ اما با فناوری‌های پیشرفته‌ی امروزی همراه با کاهش هزینه‌های انبارش، این موارد در حال حاضر مشکل محسوب نمی‌شوند؛ در عوض، مسائل دیگری پدیدار شده است از جمله اینکه چگونه مرتبط بودن بین حجم انبوهی از داده‌ها را مشخص کنیم و چگونه از داده‌هایی که به نظر نامربوط می‌رسند خلق ارزش کنیم.

همانگونه که گفته شد، کلان یک عبارت نامربوط است که در طول زمان تغییر می‌کند و سازمان‌های مختلف تفسیرهای متفاوتی از آن دارند. با افزایش سرسام آور حجم داده‌ها، حتی نام‌گذاری پله‌ی بعدی کلان داده‌ها نیز کار دشواری است. بالاترین حجم از داده‌ها را پتابایت (PB) می‌نامیدند که جای خود را به زتابایت (ZB) داده است که یک تریلیون گیگابایت یا یک میلیارد ترابایت است. بینش فناوری ۹.۱ یک نگاه کلی بر اندازه و نام‌گذاری حجم کلان داده‌ها ارائه کرده است. از یک منظر تاریخی کوتاه، در سال ۲۰۰۹ حجم داده‌ها در جهان حدود ۰.۸ زتابایت بود؛ در سال ۲۰۱۰ به بیش از ۱ زتابایت رسیده و در پایان سال ۲۰۱۱ این رقم در حدود ۱.۸ زتابایت بود. انتظار می‌رود که حجم داده‌ها در سال ۲۰۲۰ در کل جهان به ۴۴ زتابایت برسد (آدشید ۲۰۱۴). با رشد حسگرها و اینترنت اشیا، این پیش‌بینی‌ها ممکن است کاملاً غلط از آب درآید. اگرچه اینها ارقام حیرت‌آوری هستند، چالش‌ها و فرصت‌هایی که به همراه دارند نیز همینگونه است.

تنوع امروزه داده‌ها در انواع قالب‌ها وجود دارند از پایگاه‌های داده سنتی گرفته تا پایگاه داده‌های تحلیلی سلسله‌مراتبی که توسط کاربران نهایی و سیستم‌های OLAP برای اسناد متنم، نامه‌های الکترونیکی، اکس‌ام‌ال، داده‌های حسگرها، داده‌های ویدئویی، صوتی و ... ایجاد می‌شوند. بر اساس تخمین‌ها، ۸۰ تا ۸۵٪ از کل داده‌های سازمانی به نوعی در قالب ساخت‌نیافته و نیمه‌ساخت‌یافته وجود دارند. اما در ارزش آنها هیچ شکی نیست و باید برای پشتیبانی تصمیم‌گیری در تحلیل‌ها از آنها استفاده کرد.

سرعت طبق گارتنر، سرعت هم به معنای سرعت تولید داده‌ها و هم سرعت پردازش داده‌ها (یعنی ضبط، انبارش و تحلیل) برای برآورده کردن یک نیاز یا تقاضا است. تگ‌های آراف‌آی‌دی، حسگرهای خودکار، دستگاه‌های جی‌پی‌اس و دستگاه‌های اندازه‌گیری هوشمند عامل محرکی برای نیاز روزافزون پردازش داده‌ها به صورت بلادرنگ هستند. شاید سرعت بیش از هر ویژگی دیگری در بحث کلان داده‌ها نادیده گرفته شده است. واکنش سریع نسبت به تغییرات پرسرعت برای اکثر سازمان‌ها یک چالش محسوب می‌شود. برای محیط‌های حساس نسبت به زمان، هزینه‌ی فرصت داده‌ها از آن لحظه‌ای که داده‌ها ایجاد می‌شوند آغاز می‌شود. با گذشت زمان، گزاره ارزشی داده‌ها کاهش یافته و نهایتاً بی‌ارزش می‌شوند. صرفنظر از اینکه موضوع بحث، سلامت یک بیمار، یک سیستم ترافیک یا سلامت یک پورفولیوی سرمایه‌گذاری باشد، دسترسی به داده‌ها و واکنش سریع‌تر به شرایط همواره نتایج سودمندی را به همراه خواهد داشت.

بینش فناوری ۹.۱ داده‌ها بزرگ، بزرگ‌تر و بزرگ‌تر می‌شوند

شاخص اندازه داده‌ها و نام‌های جدید آنها به زحمت همگام با هم پیش می‌روند. همه‌ی ما با کیلوبایت (KB که معادل ۱,۰۰۰ بایت است)، مگابایت (MB که معادل ۱,۰۰۰,۰۰۰ بایت است)، گیگابایت (GB که معادل ۱,۰۰۰,۰۰۰,۰۰۰ بایت است) و ترابایت (TB که معادل ۱,۰۰۰,۰۰۰,۰۰۰,۰۰۰ بایت است) آشنا هستیم. فراتر از اینها، بسیاری از نام‌هایی که به اندازه داده‌ها اطلاق شده‌اند برای اکثر ما جدید هستند. در جدول زیر برخی از نام‌هایی که بعد از ترابایت وجود دارند را آورده‌ایم:

مقدار	سمبول	نام
۱۰ ^۳	KB	کیلوبایت
۱۰ ^۶	MB	مگابایت
۱۰ ^۹	GB	گیگابایت
۱۰ ^{۱۲}	TB	ترابایت
۱۰ ^{۱۵}	PB	پتابایت
۱۰ ^{۱۸}	EB	اگزابایت
۱۰ ^{۲۱}	ZB	زتابایت
۱۰ ^{۲۴}	YB	یوتابایت
۱۰ ^{۲۷}	BB	برونوبایت
۱۰ ^{۳۰}	GeB	ژئوبایت

در نظر بگیرید که یک اگزابایت از داده‌ها هر روزه در فضای اینترنت ایجاد می‌شود که معادل ۲۵۰ میلیون دی‌وی‌دی اطلاعات است. زمانیکه حجم اطلاعاتی که در طول یکسال در اینترنت ایجاد می‌شود ایده‌ی حجم بزرگتری از داده‌ها در اندازه‌ی یک زتابایت چیز دور از ذهنی نیست. در واقع، کارشناسان صنعت برآورد کرده‌اند که ما در سال ۲۰۱۶ سالانه ترافیک داده‌ای در حدود ۱.۳ زتابایت را در فضای اینترنت خواهیم داشت که ممکن است تا سال ۲۰۲۰ به ۲.۳ زتابایت برسد. تا سال ۲۰۲۰، انتظار می‌رود که ترافیک اینترنت به طور سرانه در سال به ۳۰۰ گیگابایت برسد. زمانیکه به یوتابایت اشاره کنیم، برخی از دانشمندان علم کلان‌داده‌ها از اینکه ان‌اس‌ای یا اف‌بی‌آی چه حجم عظیمی از داده‌ها را در مورد مردم در اختیار دارند تعجب می‌کنند. اگر بخواهیم بر حسب حجم اطلاعات دی‌وی‌دی صحبت کنیم، یک یوتابایت نیازمند ۲۵۰ تریلیون دی‌وی‌دی است. یک برونوبایت که یک پیشوند رسمی اس‌آی نیست معادل ۱ به همراه ۲۷ صفر است. چنین حجم و اندازه‌ی بزرگی را می‌توان برای توضیح میزان داده‌های حسگرها که از دهه‌ی آتی از اینترنت بدست می‌آید بکار برد. یک ژئوبایت برابر ۱۰ به توان ۳۰ است. به توجه به منبع کلان‌داده‌ها، موارد زیر را در نظر بگیرید:

- برخورد دهنده‌ی هادرونی بزرگ آزمایشگاه سرن^۱ ۱ پتابایت در ثانیه تولید می‌کند.
- حسگرهای موتور جت بوئینگ هر ساعت ۲۰ ترابایت داده تولید می‌کنند.
- روزانه ۶۰۰ ترابایت داده‌ی جدید در پایگاه داده‌های فیسبوک تولید می‌شود.
- در یوتیوب در هر دقیقه ۳۰۰ ساعت ویدئو آپلود می‌شود که معادل ۱ ترابایت در دقیقه است.
- تلسکوپ آرایه کیلومتر مربعی^۲ (بزرگ‌ترین تلسکوپ جهان) روزانه یک اگزابایت داده تولید خواهد کرد.

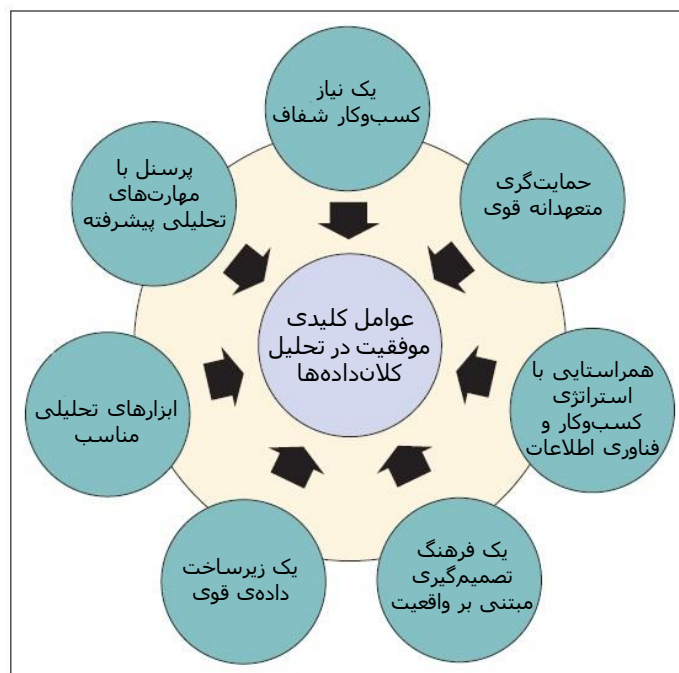
Sources: S. Higginbotham. (2012). "As Data Gets Bigger, What Comes after a Yottabyte?" gigaom.com/2012/10/30/as-data-gets-bigger-what-comes-after-a-yottabyte (accessed October 2018). Cisco. (2016).

"The Zettabyte Era: Trends and Analysis." cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.pdf (accessed October 2018).

¹ CERN Large Hadron Collider

² Square Kilometer Array Telescope

در این طوفان کلان‌داده‌ها ما در حال حاضر شاهد این هستیم که تمرکز همه بر استفاده از تحلیل‌های در حال استراحت و سیستم‌های نرم‌افزاری و سخت‌افزاری بهینه‌سازی شده برای استخراج مقادیر انبوهی از منابع مختلف داده‌ها می‌باشد. اگرچه این موضوع از اهمیت اساسی و ارزش بالایی برخوردار است ولی گروه دیگری از تحلیل‌ها که از ویژگی سرعت کلان‌داده‌ها مشتق شده‌اند وجود دارد که «تحلیل استریم داده‌ها» یا «تحلیل در حال حرکت» نام دارند که به سرعت در حال رشد و تکامل است. تحلیل استریم داده‌ها در صورتیکه به درستی انجام شود می‌تواند به اندازه‌ی تحلیل‌های در حال استراحت ارزشمند بوده یا حتی در برخی محیط‌های کسب‌وکار از ارزش بالاتری برخوردار باشند. در ادامه‌ی این فصل به تفصیل به بررسی این موضوع خواهیم پرداخت.



شکل ۹.۴ عوامل کلیدی موفقیت در تحلیل کلان‌داده‌ها

• **اسباب و وسایل:** سخت‌افزار و نرم‌افزار را در قالب یک واحد فیزیکی گرد هم می‌آورد که نه تنها سریع بلکه مقیاس‌پذیر بر اساس نیاز می‌باشد.

الزامات محاسباتی فقط بخش کوچکی از فهرست چالش‌هایی هستند که کلان‌داده بر سازمان‌های امروزی تحمیل می‌کند. در ادامه فهرستی از چالش‌هایی که مدیران اجرایی برای موفقیت در پیاده‌سازی و اجرای تحلیل‌های کلان‌داده‌ها با آنها مواجه شده‌اند را آورده‌ایم. در زمان طراحی پروژه‌ها و معماری کلان‌داده‌ها، توجه و در نظر داشتن این چالش‌ها، مسیر دستیابی به کفایت و شایستگی در تحلیل‌ها را برایمان آسان‌تر خواهد کرد.

حجم داده‌ها: توانایی ضبط، انبارش و پردازش حجم انبوهی از داده‌ها با سرعتی قابل قبول به نحویکه آخرین و جدیدترین اطلاعات در زمان لازم در اختیار تصمیم‌گیرندگان قرار گیرد.

یکپارچگی داده‌ها: توانایی تلفیق هرچه سریع‌تر و کم‌هزینه‌تر داده‌هایی که ساختار یا منبع مشابه و یکسانی ندارند.

قابلیت‌های پردازشی: توانایی پردازش سریع داده‌ها به محض ضبط. روش قدیمی برای گردآوری و پردازش داده‌ها ممکن است کارایی نداشته باشد. در بسیاری از موقعیت‌ها، داده‌ها را باید هرچه سریع‌تر و به محض ضبط و گردآوری تحلیل نمود تا بیشترین ارزش را برای ما فراهم کنند. (این را تحلیل/استریم می‌نامند که در فصل بعدی به بررسی آن خواهیم پرداخت).

حکمرانی داده‌ها: توانایی رسیدگی به مسائل امنیتی، حریم خصوصی، مالکیت و کیفیت کلان‌داده‌ها. همانند حجم، تنوع (قالب و منبع) و سرعت تغییر داده‌ها، قابلیت‌های روش‌های حکمرانی نیز تغییر می‌کند.

دسترسی‌پذیری مهارت‌ها: کلان‌داده‌ها با ابزارهای جدیدی به کار گرفته می‌شوند و به روش‌های متفاوتی به آنها نگاه می‌شود. تعداد افرادی که از مهارت کافی برای کار با کلان‌داده‌ها برخوردارند اندک است (دانشمندان داده).

هزینه راه‌حل: از آنجائیکه کلان‌داده‌ها دنیایی از بهبودهای امکان‌پذیر در کسب‌وکار را به روی ما گشوده‌اند، فعالیت‌های بسیار زیادی برای آزمایش و کشف الگوهای مهم و بینش‌هایی که منجر به ارزش می‌شوند در حال رخ دادن است. برای حصول اطمینان از نرخ بازگشت مثبت سرمایه‌گذاری در پروژه‌های کلان‌داده، کاهش هزینه‌ی راه‌حل‌های نرم‌افزاری بکار رفته برای یافتن آن ارزش‌ها اهمیت زیادی دارد.

اگرچه این چالش‌ها واقعیت دارند ولی گزاره ارزشی تحلیل‌های کلان‌داده نیز واقعیت دارد. هرآنچه که بتوانید به عنوان رهبر تحلیل کسب‌وکارتان برای ارتقاء ارزش منابع داده جدید انجام دهید، سازمان‌تان را فراتر از مرزهای آزمایش و اکتشاف کلان‌داده‌ها برده و تبدیل به یک عامل تمایز برای سازمان خواهد شد. اکتشاف در کلان‌داده‌ها هیچ ایرادی ندارد ولی نهایتاً بکارگیری آن بینش‌ها است که برای ما خلق ارزش می‌کند.

مهمترین مسائلی که در کسب‌وکار مطرح بوده و به کمک کلان‌داده‌ها می‌توان آنها را حل نمود به طور کلی بحث کارایی فرایند، کاهش هزینه و همچنین تجربه‌ی مشتری هستند ولی اولویت‌های متفاوتی نیز بسته به نوع صنعت مطرح می‌شود. کارایی فرایند و کاهش هزینه شاید جزو مهمترین مسائلی باشند که به کمک تحلیل کلان‌داده‌ها بتوان در بخش‌های تولید، دولت، انرژی، ارتباطات و رسانه‌ها، حمل و نقل و مراقب سلامت به آنها رسیدگی کرد. بهبود تجربه‌ی مشتری جزو برترین مشکلات شرکت‌های بیمه و خرده‌فروشی‌ها است. مدیریت ریسک معمولاً در اول فهرست شرکت‌های بانک‌داری و آموزشی قرار دارد. در ادامه فهرست کوچکی از مسائل و مشکلاتی که به کمک تحلیل کلان‌داده‌ها می‌توان به آنها رسیدگی کرد آورده شده است:

کارایی فرایند و کاهش هزینه

مدیریت برند

حداکثرسازی درآمد، فروش مکمل و بیش فروشی

شناسایی ریزش مشتری و جذب مشتری

بهبود خدمات مشتریان

مورد کاربردی ۹.۲ یک مثال عالی در صنعت خرده‌فروشی را به تصویر می‌کشد که منابع پراکنده از داده‌ها با هدف درک سفر مشتری در یک زیرساخت کلان‌داده تجمیع و یکپارچه شده‌اند.

اُورِاستاک. کام چندین پایگاه داده را برای درک سفر مشتری ادغام می‌کند

مورد کاربردی ۹.۲

سازمان‌های خرده‌فروشی بزرگ از قبیل اُورِاستاک. کام^۱ برای رشد درآمدشان در کمپین‌های بازاریابی بسیاری سرمایه‌گذاری می‌کنند. این کمپین‌ها می‌توانند شامل کمپین‌های آنلاین هدفمند، کمپین‌های پستی مستقیم، تبلیغات از طریق کانال‌های مختلف، برنامه‌های وفاداری از طریق ارائه مشوق‌های مختلف به مشتریان و غیره باشد. هرکدام از اینها هزینه‌های هنگفت بازاریابی را به همراه داشته و نرخ بازگشت سرمایه آنها نیز متغیر است. یک چالش برای هر شرکت این است که تمامی این کمپین‌ها را به نحوی تحلیل کند که داده‌ها را در یک مکان گرد آورده و بتواند سفر مشتری را درک کند. کدام ترکیب از کمپین‌ها یا تعاملات نهایتاً منجر به خرید کالا یا خدمات توسط مشتری و به چه میزان می‌شود؟ این منبع داده‌ها می‌توانند شامل داده‌های ترافیک وبسایت‌ها باشند که در لاگ فایل‌های ساخت‌نیافته نگهداری می‌شوند؛ داده‌های عملکرد کمپین‌های الکترونیکی در قالب نیمه‌ساخت‌یافته که از طریق شرکت‌های برگزار کننده کمپین‌های نامه‌های الکترونیکی ارسال می‌شوند؛ داده‌های رسانه‌های اجتماعی از قبیل پست‌های فیسبوک و غیره باشند. بهم مرتبط کردن تمامی این داده‌ها به داده‌های درونی شرکت با این هدف که ارزشی برای مشتری به منظور خرید کالاها و محصولات خلق شود و بتوان نرخ بازگشت سرمایه حاصل از ترکیب‌های کمپین‌های مختلف را محاسبه نمود یک چالش دیگر در یکپارچه سازی داده‌ها است. اما ترکیب کردن اینگونه منابع داده تحت یک چارچوب کلان‌داده عملی‌تر است. سپس با استفاده از قابلیت‌های تحلیل مسیر که در پرونده‌ی آغازین نشان دادیم، یک کاربر غیرفنی نیز می‌تواند سفرهای مختلف مشتری را مشاهده کرده و آنهایی که منجر به فروش کارآمدتر و نرخ بازگشت سرمایه‌ی بالاتری برای فعالیت‌های بازاریابی می‌شوند را شناسایی کند. هدف از اینکار، برقراری رابطه‌ی بلندمدت با مشتریان از طریق درک الگوهای جستجو، رفتارهای خرید، پاسخ‌ها در وبسایت و غیره است. اُورِاستاک. کام با بکارگیری کارکرد تحلیل مسیر نرم‌افزار ترادیتا و نتج ولی با ادغام منابع داده بسیار متفاوت تحت چارچوب کلان‌داده توانست با موفقیت به این امر دست پیدا کند.

سوالاتی برای بحث و گفتگو

۴. برخی از کمپین‌های بازاریابی متفاوت که یک شرکت می‌تواند برای ترغیب و جذب مشتریان بکار بگیرد چه هستند؟
۵. با مصورسازی رایج‌ترین مسیرهایی که منجر به خرید مشتری می‌شوند چگونه می‌توان از آن اطلاعات به منظور تصمیم‌گیری در خصوص کمپین‌های بازاریابی آینده استفاده کرد؟
۶. چه کاربردهای دیگری از اینگونه تکنیک‌های تحلیل مسیر را می‌توانید نام ببرید؟

Compiled from: "Overstock.com Uses Teradata Path Analysis to Boost Its Customer Journey Analytics," March 27, 2018, at www.retailintights.com/doc/overstock-com-uses-teradatapath-analysis-boost-customer-journey-analytics-0001 (accessed October 2018), and "Overstock.com: Revolutionizing Data and Analytics to Connect Soulfully with Their Customers," at www.teradata.com/Resources/Videos/Overstock-com-Revolutionizing-data-and-analy (accessed October 2018).

¹ Overstock.com

در این بخش به معرفی اصول و مبانی کلان داده و برخی کاربردهای بالقوه‌ی آن پرداختیم. در بخش بعدی در مورد چند اصطلاح و فناوری نوظهور در حوزه کلان داده توضیح خواهیم داد.

سوالاتی برای مرور بخش ۹.۳

۲۳. تحلیل کلان داده چیست؟ چه تفاوتی با سایر تحلیل‌های معمول دارد؟
۲۴. عوامل کلیدی موفقیت تحلیل‌های کلان داده چه هستند؟
۲۵. چالش‌های بزرگی که در پیاده‌سازی و اجرای تحلیل کلان داده‌ها باید در نظر داشت کدامها هستند؟
۲۶. مسائل متداول کسب‌وکار که به کمک تحلیل کلان داده قابل حل هستند چه هستند؟

مَپ‌ردیوس^۱ یک تکنیک معروف از شرکت گوگل است که پردازش فایل داده‌های چندساختاری بسیار بزرگ را بین خوشه‌ی عظیمی از ماشین‌ها توزیع می‌کند. این تکنیک با تجزیه‌ی پردازش به واحدهای کاری کوچک که امکان اجرای آن به طور موازی بین صدها یا هزاران گروه در یک خوشه وجود دارد به عملکرد بالایی دست می‌یابد. مطلب زیر از مقاله‌ای در خصوص مَپ‌ردیوس نقل شده است:

مَپ‌ردیوس یک مدل برنامه‌نویسی و پیاده‌سازی مرتبط با آن برای پردازش و تولید مجموعه داده‌های بزرگ است. برنامه‌هایی که به این سبک عملیاتی نوشته می‌شوند به طور خودکار بر روی خوشه‌ی بزرگی از ماشین‌ها موازی‌سازی و اجرا می‌شوند. برنامه‌نویسان بدون هرگونه تجربه با سیستم‌های موازی و توزیع‌یافته به آسانی می‌توانند از منابع یک سیستم توزیع‌یافته بزرگ بهره‌برداری کنند (دین و گماوات ۲۰۰۴).

نکته اصلی از این نقل قول این است که مَپ‌ردیوس یک مدل برنامه‌نویسی است نه یک زبان برنامه‌نویسی بدین معنی که برای استفاده‌ی برنامه‌نویسات طراحی شده نه کاربران کسب‌وکار. آسان‌ترین روش برای توضیح نحوه‌ی کارکرد مَپ‌ردیوس از طریق یک مثال است (شکل ۹.۵).

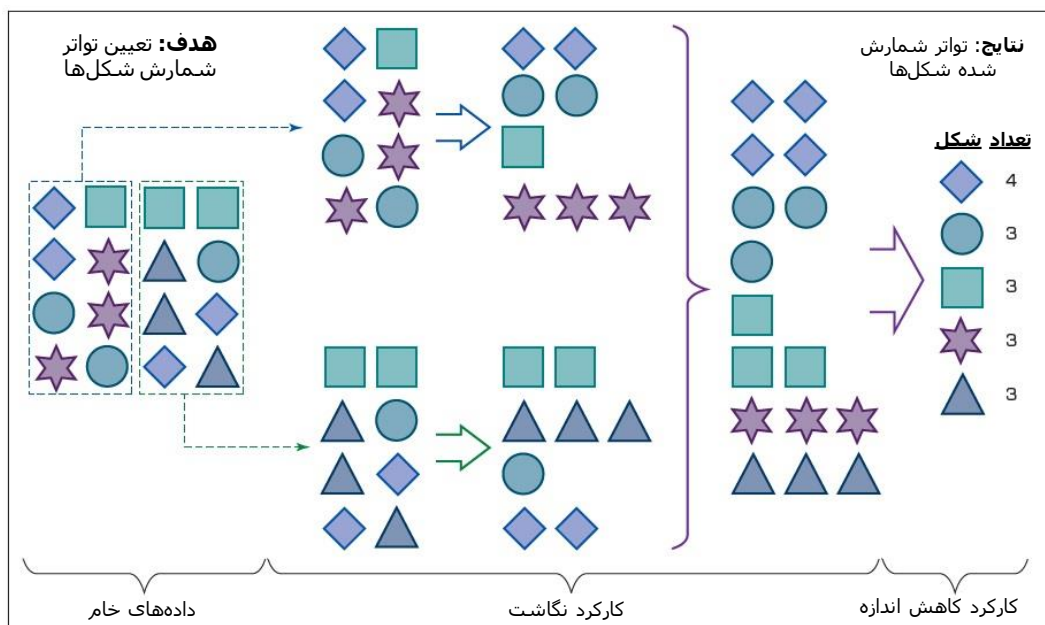
¹ MapReduce

ورودی فرایند مَپِردیوس در شکل ۹.۵ یک مجموعه از مربع‌های رنگی هستند. هدف در اینجا شمارش تعداد مربع‌ها به تفکیک هر رنگ است. برنامه‌نویس در این مثال مسئول کدگذاری نقشه و کاهش حجم برنامه‌ها است؛ مابقی فرایند پردازش از طریق سیستم نرم‌افزاری که مدل برنامه‌نویسی مَپِردیوس را اجرا می‌کند انجام می‌شود.

سیستم مَپِردیوس ابتدا فایل ورودی را خوانده و آن را به چندین بخش تقسیم می‌کند. در این مثال، به دو بخش تقسیم شده است ولی در سناریوهای واقعی تعداد بخش‌ها معمولاً بسیار بیشتر است. سپس این بخش‌ها توسط چندین برنامه‌ی نگاشت که به طور موازی بر روی گره‌های یک خوشه اجرا می‌شوند پردازش می‌شوند. نقش هر برنامه‌ی نگاشت در این حالت این است که داده‌های درون هر بخش را به تفکیک رنگ گروه‌بندی کند. سیستم مَپِردیوس سپس خروجی هر برنامه‌ی نگاشت را گرفته و نتایج را به عنوان ورودی سیستم کاهش تعداد ادغام می‌کند که در اینجا مجموع تعداد مربعات هر رنگ را محاسبه می‌کند. در این مثال فقط یک کپی از برنامه‌ی کاهش تعداد استفاده شده است ولی در عمل بیشتر از این می‌باشد. برای بهینه‌سازی عملکرد، برنامه‌نویسان می‌توانند برنامه‌های کاهش تعداد خودشان را نوشته و همچنین از یک کامپایلر برای تلفیق فایل‌های خروجی نگاشت‌ها برای کاهش تعداد فایل‌های خروجی استفاده کنند.

۹.۵.۱.۱. فرایند مَپِردیوس

مَپِردیوس به سازمان‌ها در پردازش و تحلیل حجم بزرگی از داده‌های چندساختاری کمک می‌کند. مثال‌هایی از کاربرد آن شامل اندیس‌گذاری و جستجو، تحلیل گراف، تحلیل متن، یادگیری ماشین، تبدیل داده‌ها و غیره می‌باشد. پیاده‌سازی این نوع کاربردها با استفاده از اس‌کیوال استاندارد که در سیستم‌های مدیریت پایگاه داده رابطه‌ای استفاده می‌شود اغلب دشوار است. ماهیت رویه‌ای مَپِردیوس موجب شده تا برنامه‌نویسان ماهر به راحتی آن را درک کنند. مزیت دیگر آن این است که برنامه‌نویسان نیازی نیست که نگران پیاده‌سازی رایانش موازی باشد زیرا اینکار توسط سیستم انجام می‌شود.



شکل ۹.۵ نمایش گرافیکی یک فرایند مَپِردیوس

مَپِردیوس برای برنامه‌نویسان طراحی شده است و افرادی که تخصص برنامه‌نویسی ندارند می‌توانند از برنامه‌های کاربردی از پیش ساخته شده مَپِردیوس و کتابخانه‌ی توابع آن استفاده کنند. کتابخانه‌های تجاری و منبع باز مَپِردیوس موجود بوده و طیف وسیعی از قابلیت‌های تحلیلی را ارائه می‌کنند. به عنوان مثال، آپاچی ماهوت یک کتابخانه‌ی یادگیری ماشین منبع باز از «الگوریتم‌های خوشه‌بندی، دسته‌بندی و فیلترینگ مشارکتی مبتنی بَچ» می‌باشد که با استفاده از مَپِردیوس اجرا می‌شود.

هَدوپ یک چارچوب منبع باز برای پردازش، انبارش و تحلیل حجم انبوهی از داده‌های توزیع‌شده و ساخت‌نیافته است. این چارچوب که در اصل توسط داگ کاتینگ در یاهو! و با الهام از مَپِردیوس که یک کارکرد توسعه یافته در گوگل در اوایل سال‌های ۲۰۰۰ برای اندیس‌گذاری وب بود ایجاد شد. هدف از طراحی آن، مدیریت موازی حجم‌های پتابایت و اِگزابایت از داده‌های توزیع شده بین گره‌های متعدد بود. خوشه‌های هَدوپ بر روی سخت‌افزارهای ارزان قیمت اجرا می‌شوند تا هزینه‌های پروژه‌ها کاهش یابد.



Courtesy of The Apache Software Foundation

هَدوپ در حال حاضر یک پروژه از بنیاد نرم‌افزار آپاچی است که صدها نفر به طور مستمر به بهبود فناوری اصلی آن می‌پردازند. مفهوم بنیادی: هَدوپ به جای اینکه یک بلوک حجیم و عظیم از داده‌ها را بر روی یک ماشین پردازش کند، کلان‌داده‌ها را به چندین بخش تقسیم می‌کند تا هر بخش را بتوان به طور همزمان پردازش و تحلیل نمود.

یک کارگزار^۱ به داده‌های ساخت‌نیافته و نیمه‌ساخت‌نیافته از منابع متعدد از جمله لاگ فایل‌ها، فیدهای رسانه‌های اجتماعی و پایگاه داده‌های تحلیلی درونی دسترسی پیدا می‌کند. داده‌ها به «بخش»هایی تقسیم می‌کند که سپس بر روی یک فایل سیستم متشکل از چندین گره که بر روی سخت‌افزارهای ارزان قیمت اجرا می‌شوند بارگذاری می‌شوند. انبار فایل پیش‌گزیده در هَدوپ، سیستم فایل توزیع‌شده هَدوپ^۲ یا اِچ‌دی‌اِف‌اس نام دارد. سیستم‌های فایلی همچون اِچ‌دی‌اِف‌اس قادر به انبارش حجم انبوهی از داده‌های ساخت‌نیافته و نیمه‌ساخت‌نیافته هستند زیرا نیازی به سازمان‌دهی داده‌ها در قالب سطر و ستون‌های رابطه‌ای ندارند. هر «بخش» چندین مرتبه تکثیر شده و در داخل سیستم فایل بارگذاری می‌شود به نحویکه اگر یک گره موفق به پردازش نشد، گره‌ی دیگر یک نسخه از داده‌ها را در اختیار داشته باشد. یک Name Node به عنوان یک تسهیل‌گر عمل کرده و اطلاعاتی از

¹ Client

² Hadoop Distributed File System (HDFS)

قبیل اینکه کدام گره در حال حاضر در دسترس است، داده‌ها در کدام خوشه قرار دارند و کدام گره‌ها موفق نبوده‌اند را به کارگزار می‌دهد.

پس از بارگذاری داده‌ها در داخل خوشه، آماده‌ی تحلیل توسط چارچوب مَپ‌ردیوس می‌باشند. کارگزار یک کار «نگاشت» که معمولاً به صورت یک پرس و جو به زبان جاوا نوشته می‌شود را به یکی از گره‌ها در خوشه‌ای به نام دنبال‌کننده کار^۱ ارسال می‌کند. این گره برای تعیین اینکه برای تکمیل کار و اینکه داده‌ها در کجای خوشه قرار دارند به Name Node رجوع می‌کند. پس از اینکار، دنبال‌کننده کار پرس و جو را به گره‌های مربوطه ارسال می‌کند. به جای اینکه تمامی داده‌ها به یک محل مرکزی برای پردازش بازگردانده شوند، پردازش به طور همزمان یا موازی در هر گره اتفاق می‌افتد. این ویژگی اساسی هَدوپ است. زمانیکه کار پردازش هر گره به اتمام رسید، نتایج را ذخیره می‌کند. کارگزار یک کار Reduce را از طریق دنبال‌کننده کار آغاز می‌کند که در آن نتایج فاز نگاشت که به صورت محلی بر روی تک تک گروه‌ها ذخیره شده‌اند برای تعیین «پاسخ» به پرس و جوی اصلی تجمیع می‌شوند و سپس بر روی گره‌ی دیگری در داخل خوشه بارگذاری می‌شوند. سرویس‌گیرنده به این نتایج دسترسی دارد که سپس می‌تواند آنها را برای آنالیز بیشتر بر روی یکی از چندین محیط تحلیلی بارگذاری کند. کار مَپ‌ردیوس در اینجا به اتمام رسیده است.

به محض اتمام فاز مَپ‌ردیوس، داده‌های پردازش شده برای تحلیل بیشتر توسط دانشمندان علم داده‌ها و افرادی که از مهارت‌های تحلیل داده‌ی پیشرفته برخوردارند آماده است. دانشمندان علم داده‌ها^۲ می‌توانند داده‌ها را با استفاده از هر کدام از ابزارها برای هر کاربردی از جمله جستجوی الگوها و بینش‌های پنهان یا مبنای ساخت برنامه‌های کاربردی تحلیلی مورد تحلیل قرار دهند. داده‌ها را همچنین می‌توان از خوشه‌های هَدوپ خوشه‌سازی کرده و به پایگاه‌های داده رابطه‌ای، پایگاه داده‌ها تحلیلی و سایر سیستم‌های فناوری اطلاعات معمول برای تحلیل بیشتر و/یا پشتیبانی از پردازش‌های تراکنشی منتقل کرد.

هَدوپ از چندین کامپوننت تشکیل شده که شامل :

سیستم فایل توزیع شده هَدوپ (اچ‌دی‌اف‌اس): لایه‌ی انبارش پیش‌گزیده در هر خوشه‌ی هَدوپ.

گره‌ی نام: گره‌ای در خوشه‌ی هَدوپ که اطلاعاتی در مورد اینکه داده‌های خاص در کجای خوشه قرار گرفته‌اند و آیا خوشه‌ای از کار افتاده است یا خیر به کارگزار می‌دهد.

گره‌ی ثانویه: یک پشتیبان از گره‌ی نام که به صورت دوره‌ای داده‌های گره‌ی نام را در صورت از کار افتادن تکثیر می‌کند.

دنبال‌کننده کار: یک گره در خوشه‌ی هَدوپ که کارهای مَپ‌ردیوس یا پردازش داده‌ها را شروع و هماهنگ می‌کند.

گره‌های اسلیو^۳: این گره‌ها داده‌ها را ذخیره کرده و برای پردازش آنها از دنبال‌کننده‌ی کار دستور می‌گیرند.

علاوه بر این کامپوننت‌ها، اکوسیستم هَدوپ از چند زیرپروژه‌ی مکمل نیز تشکیل شده است. پایگاه داده‌های تحلیلی نواس کیوال مثل کاساندر و اچ‌پیس نیز برای انبارش نتایج کارهای مَپ‌ردیوس در هَدوپ استفاده می‌شوند. علاوه بر جاوا، برخی از کارهای

¹ Job Tracker

² Data Scientists

³ Slave Node

مَپردیوس و برخی توابع دیگر هَدوپ به زبان پیگ که یک زبان منبع باز که به طور خاص برای هَدوپ طراحی شده است نیز نوشته می‌شوند. هاپو یک پایگاه داده‌ی تحلیلی منبع باز است که در اصل توسط فیسبوک توسعه یافت که امکان مدلسازی تحلیلی در چارچوب هَدوپ را فراهم می‌کند. برخی از معروف‌ترین زیرپروژه‌های هَدوپ به شرح زیر می‌باشند:

هاپو^۱ یک چارچوب انبارش داده‌ی مبتنی بر هَدوپ است که در اصل توسط فیسبوک توسعه یافت و به کاربران اجازه می‌دهد تا پرس و جوها را در زبانی مشابه اس‌کیوال^۲ به نام هاپوکیوال^۳ بنویسند که سپس به مَپردیوس تبدیل می‌شوند. بدین ترتیب برنامه‌نویسان اس‌کیوال که هیچگونه تجربه‌ای در مَپردیوس ندارند می‌توانند از پایگاه داده تحلیلی استفاده کنند و یکپارچه‌سازی آن با ابزارهای هوش تجاری و مصورسازی مثل میکرواستراتژی، تابلو، رولوشنز آنالیتیکز^۴ و غیره نیز آسان تر می‌شود.

پیگ^۵ یک زبان پرس و جوی مبتنی بر هَدوپ است که توسط یاهو! توسعه یافت و یادگیری آن نسبتاً آسان بوده و در لایه‌های بسیار عمیق و طولانی از داده‌ها کار می‌کند (یکی از محدودیت‌های اس‌کیوال).

اچ‌پیس^۶ یک پایگاه داده‌ی غیررابطه‌ای است که امکان جستجوهای با تأخیر کم و سریع را در هَدوپ فراهم می‌آورد و با افزودن قابلیت‌های تراکنشی به هَدوپ اجازه‌ی به‌روزرسانی، درج و حذف را به کاربران می‌دهد. ای‌پی و فیسبوک از اچ‌هایو استفاده می‌کنند.

فلوم^۷ یک چارچوب برای قراردادن داده‌ها در هَدوپ است. عوامل برای گردآوری و ادغام داده‌ها در هَدوپ در سرتاسر زیرساخت فناوری اطلاعات از جمله وب سرورها، اپلیکیشن سرورها و دستگاه‌های سیار پخش می‌شوند.

اوزی^۸ اوزی یک سیستم پردازش جریان کار است که به کاربران اجازه می‌دهد تا مجموعه‌ای از کارها که به زبان‌های مختلف نوشته‌اند – مثل مَپردیوس، پیگ و هاپو – را تعریف کرده و آنها را به شکل هوشمندانه‌ای به یکدیگر متصل کنند. اوزی به کاربران اجازه می‌دهد تا مشخص کنند که مثلاً یک پرس و جوی خاص فقط باید پس از انجام و تکمیل کارهای مشخص شده‌ی قبلی آغاز شود.

آمباری^۹ آمباری یک مجموعه ابزار مبتنی بر وب برای جاری‌سازی، مدیریت و پایش خوشه‌های آپاچی هَدوپ است. توسعه‌ی آن توسط مهندسانی از هورتون‌وُرکز آغاز شد که آمباری را در پلتفرم داده‌های هورتون‌وُرکز^۹ ارائه کرده‌اند.

¹ HIVE

² HiveQL

³ Revolutions Analytics

⁴ PIG

⁵ HBase

⁶ FLUME

⁷ OOZIE

⁸ AMBARI

⁹ Hortonworks Data Platform

آروو^۱ آروو یک سیستم سریالی سازی داده است که اجازه می‌دهد تا اسکیمای فایل‌های هَدوپ را رمزنگاری کنیم. آروو در تجزیه‌ی داده‌ها و فراخوانی رویه‌های حذف شده مهارت دارد.

ماهوت^۲ ماهوت یک کتابخانه‌ی داده‌کاوی است که از معروف‌ترین الگوریتم‌های داده‌کاوی برای خوشه‌بندی، آزمون رگرسیون و مدل سازی آماری استفاده کرده و آنها را با استفاده از مدل مَپر دیوس پیاده‌سازی می‌کند.

اسکوپ^۳ اسکوپ یک ابزار برقراری ارتباط^۴ برای جابجایی داده‌ها از پایگاه داده‌های تحلیلی غیرهَدوپ مثل پایگاه داده‌های رابطه‌ای و پایگاه داده تحلیلی به هَدوپ است. اسکوپ به کاربران اجازه می‌دهد تا مکان هدف در داخل هَدوپ را مشخص کند و به اسکوپ دستور می‌دهد تا داده‌های را از ارواکل، ترادیتا یا هر پایگاه داده رابطه‌ای دیگری به مکان مورد نظر انتقال دهد.

اچ کاتالوگ^۵ اچ کاتالوگ یک سرویس به اشتراک‌گذاری و مدیریت فراداده‌های متمرکز برای آپاچی هَدوپ است که یک دید یکپارچه از کل داده‌ها در خوشه‌های هَدوپ ایجاد کرده و به ابزارهای گوناگون از قبیل پیگ و هایو اجازه می‌دهد تا هر جزء داده را بدون نیاز به آگاهی از مکان فیزیکی آن در خوشه پردازش کند.

مزیت اصلی هَدوپ این است که به سازمان‌ها اجازه می‌دهد تا حجم بسیار زیادی از داده‌های ساخت‌نیافته و نیمه‌ساخت‌یافته را به شکلی اقتصادی از لحاظ زمان و هزینه پردازش و تحلیل کنند. از آنجائیکه خوشه‌های هَدوپ می‌توانند تا حجم‌های پتابایت و حتی اگزابایت داده‌ها را پردازش کنند، سازمان‌ها دیگر نیازی نیست از داده‌های نمونه استفاده کنند بلکه می‌توانند تمامی داده‌های مرتبط را پردازش و تحلیل کنند. دانشمندان علم داده‌ها می‌توانند یک روش تکرارپذیر را برای تحلیل، پالایش و آزمون مستمر پرس و جوها برای کشف بینش‌هایی که تا پیش از این ناشناخته بوده‌اند بکار بگیرند. شروع کار با هَدوپ نیز بسیار کم هزینه است. برنامه‌نویسان می‌توانند نرم‌افزار آپاچی هَدوپ را به صورت رایگان بارگذاری کرده و در کمتر از یک روز شروع به کار با آن بکنند.

جنبه‌ی منفی هَدوپ و کامپوننت‌های آن این است که هنوز به بلوغ نرسیده و در حال توسعه هستند. پیاده‌سازی و مدیریت خوشه‌های هَدوپ و انجام تحلیل‌های پیشرفته بر روی حجم انبوهی از داده‌های ساخت‌نیافته همانند هر فناوری جدید و نابالغ دیگری نیازمند تخصص، مهارت و آموزش‌های بسیاری است. متأسفانه، در حال حاضر تعداد برنامه‌نویسان هَدوپ و دانشمندان

¹ ARVO

² MAHOUT

³ SQOOP

⁴ Connectivity tool

⁵ HCatalog

علم داده‌ها بسیار اندک است که موجب شده تا بسیاری از سازمان‌ها قادر به نگهداری و بهره‌برداری از خوشه‌های پیچیده‌ی هَدوپ نباشند. در پایان اینکه هَدوپ یک چارچوب دسته‌گرا^۱ است بدین معنا که از پردازش و تحلیل بلادرنگ داده‌ها پشتیبانی نمی‌کند. خبر خوب اینکه برخی از خبرگان فناوری اطلاعات در پروژه‌ی آپاچی هَدوپ مشارکت دارند و نسل جدیدی از برنامه‌نویسان هَدوپ و دانشمندان علم داده‌ها در حال رشد و بلوغ می‌باشند. در نتیجه، این فناوری به سرعت در حال رشد است و به ابزاری قدرتمندتر و ساده‌تر برای پیاده‌سازی و مدیریت تبدیل شده است. یک اکوسیستم از تأمین‌کنندگان خدمات نرم‌افزاری چه شرکت‌های نوپای متمرکز بر هَدوپ مثل کلاودار^۲ و هورتون‌وُرکز و شرکت‌های با سابقه مثل آی‌بی‌ام، میکروسافت، ترادیتا و اوراکل برای عرضه‌ی ابزارها و سرویس‌های تجاری و سازمانی هَدوپ و پیاده‌سازی و مدیریت این فناوری برای سازمان‌های سنتی همکاری دارند. شرکت‌های نوپای دیگری برای رفع نواقص و کمبودهای پایگاه داده‌های نواس کیوال (نه فقط اس کیوال) که قادر به ارائه‌ی بینش‌های تقریباً لحظه‌ای و بلادرنگ در کنار هَدوپ هستند تلاش می‌کنند. بینش فناوری ۹.۲ چند واقعیت را برای شفاف کردن برخی تصورات اشتباه در مورد هَدوپ ارائه کرده است.

بینش فناوری ۹.۲ چند واقعیت آشکارساز در مورد هَدوپ

اگرچه هَدوپ و فناوری‌های مرتبط با آن چیزی در حدود ۵ سال است که پدیده آمده‌اند ولی اکثر مردم همچنان تصورات نادرستی در مورد هَدوپ و فناوری‌های مربوطه از قبیل مَپر دیوس و هایو دارند. در ادامه فهرستی از ۱۰ واقعیت برای شفاف‌سازی این تصورات نادرست در مورد هَدوپ و اینکه هوش تجاری مبتنی بر هَدوپ، انبارش داده‌ها و تحلیل‌ها در چه موقعیت‌های کسب‌وکار و فناوری می‌توانند مفید باشند ارائه شده است (راسوم ۲۰۱۳).

واقعیت ۱. هَدوپ از چندین محصول تشکیل شده است. ما طوری از هَدوپ صحبت می‌کنیم گویی یک نرم‌افزار یک تکه است در حالیکه از گروهی از محصولات و فناوری‌های منبع باز که تحت نظارت بنیاد نرم‌افزار آپاچی است تشکیل شده است. (فروشنندگان دیگری نیز محصولات هَدوپ را عرضه می‌کنند).

کتابخانه‌ی آپاچی هَدوپ (به ترتیب اولویت هوش تجاری) شامل اچ‌دی‌اف‌اس، مَپر دیوس، هایو، اچ‌بیس، پیگ، زوکپیر، فوم، اسکوپ، اوزی، هیو و غیره است. شما می‌توانید آنها را به روش‌های مختلفی ادغام کنید ولی اچ‌دی‌اف‌اس و مَپر دیوس (شاید با اچ‌بیس و هایو) یک مجموعه فناوری مفید را برای کاربردهای هوش تجاری، انبارش داده و تحلیل‌ها فراهم می‌کنند.

واقعیت ۲. هَدوپ اگرچه منبع باز است ولی فروشنندگان دیگری نیز آن را عرضه می‌کنند. کتابخانه‌ی نرم‌افزار منبع باز آپاچی هَدوپ توسط ای‌اس‌اف در سایت apache.org در دسترس است. برای کاربرانی که به بسته‌های سازمانی آماده‌تر نیاز دارند چند فروشنده در حال حاضر نسخه‌ای از هَدوپ را که شامل ابزارهای اداری/اجرایی اضافی و پشتیبانی فنی است عرضه می‌کنند.

واقعیت ۳. هَدوپ یک محصول تکی نیست بلکه یک اکوسیستم است. علاوه بر محصولات شرکت آپاچی، اکوسیستم توسعه یافته‌ی هَدوپ شامل فهرستی فزاینده‌ای از محصولات سایر عرضه‌کنندگان است که فناوری‌های هَدوپ ادغام شده‌اند. با یک دقیقه جستجو در اینترنت می‌توانید آنها را شناسایی کنید.

واقعیت ۴. اچ‌دی‌اف‌اس یک سیستم فایل است نه یک سیستم مدیریت پایگاه داده. هَدوپ در وهله‌ی اول یک سیستم فایل توزیع‌شده بوده و فاقد قابلیت‌هایی است که بتوان با سیستم‌های مدیریت پایگاه داده مثل اندیس‌گذاری، دسترسی تصادفی به داده‌ها و پشتیبانی اس‌کیوال ارتباط برقرار کرد. البته اشکالی ندارد چون اچ‌دی‌اف‌اس کارهایی می‌تواند انجام دهد که سیستم‌های مدیریت پایگاه داده نمی‌توانند.

واقعیت ۵. هایو مشابه اس‌کیوال است ولی یک اس‌کیوال استاندارد نیست. بسیاری از ما معتاد به اس‌کیوال شده‌ایم چون به خوبی آن را بلذیم و ابزارهای ما به آن وابسته هستند. افرادی که با اس‌کیوال آشنایی دارند سریعاً می‌توانند کدنویسی هایو را بیاموزند ولی این مشکل سازگاری با ابزارهای مبتنی بر اس‌کیوال را رفع نمی‌کند.

واقعیت ۶. هَدوپ و مَپر دیوس بهم مرتبط هستند ولی به یکدیگر وابسته نیستند. برنامه‌نویسان شرکت گوگل مَپر دیوس را پیش از آنکه اچ‌دی‌اف‌اس وجود داشته باشد نوشتند و برخی نسخه‌های مَپر دیوس با انواع فناوری‌های انبارش از جمله اچ‌دی‌اف‌اس، سیستم‌های فایلی دیگر و برخی سیستم‌های مدیریت پایگاه داده کار می‌کنند.

¹ Batch-oriented framework

² Cloudera

واقعیت ۷. مَپر دیوس به خودی خود یک ابزار تحلیل نیست بلکه کنترل‌هایی را برای تحلیل‌ها ارائه می‌کند. مَپر دیوس یک موتور اجرای چندمنظوره است که قادر است تا پیچیدگی‌های ارتباطات شبکه‌ای، برنامه‌ریزی موازی و تولرانس خطای هر نوع برنامه کاربردی که بتوانید کدنویسی کنید را مدیریت کند.

واقعیت ۸. هَدوپ با تنوع داده‌ها سر و کار دارد نه فقط با حجم داده‌ها. از لحاظ نظری، اِچ‌دی‌اف‌اس می‌تواند انبارش و دسترسی به هر نوع داده را تا زمانی‌که داده‌هایتان را در یک فایل قرار داده و آن فایل را در اِچ‌دی‌اف‌اس کپی کنید مدیریت کند. به خاطر همین سادگی است بسیاری از کاربران روی به استفاده از آپاچی اِچ‌دی‌اف‌اس آورده‌اند.

واقعیت ۹. هَدوپ یک فناوری مکمل به همراه یک پایگاه داده تحلیلی است و به ندرت پیش می‌آید که جایگزین آن باشد. اکثر سازمان‌ها انبار داده‌های خودشان را برای داده‌های ساخت‌یافته رابطه‌ای طراحی کرده‌اند که جلب ارزش هوش تجاری از داده‌های ساخت‌نیافته یا نیمه‌ساخت‌یافته را دشوار می‌کند. هَدوپ با مدیریت انواع داده‌های چندساختاری که پایگاه داده‌های تحلیلی قادر به مدیریت آنها نیستند به عنوان یک مکمل برای پایگاه داده‌های تحلیلی عمل می‌کند.

واقعیت ۱۰. هَدوپ امکان انجام انواع تحلیل‌ها و نه فقط تحلیل‌های وب را فراهم می‌کند. هَدوپ اطلاعات بسیار زیادی را در مورد اینکه شرکت‌ها چگونه از اینترنت برای تحلیل لاگ‌های وب و سایر داده‌های وبی استفاده می‌کنند ارائه می‌کند ولی این تمام ماجرا نیست. به عنوان مثال، یک کلان‌داده که از دستگاه‌های حسگر مثل روباتیک در ساخت‌تولید، آر‌اف‌آی‌دی در خرده‌فروشی یا پایش شبکه‌ای در خدمات رفاهی را در نظر بگیرید. در کاربردهای تحلیلی قدیمی‌تر که نیازمند نمونه داده‌های بزرگ هستند – مثل بخش‌بندی مشتریان، شناسایی کلاهبرداری و تحلیل ریسک – می‌توان از کلان‌داده‌های بیشتری که توسط هَدوپ مدیریت می‌شود بهره برد. به همین منوال، داده‌های اضافی هَدوپ می‌توانند یک دید ۳۶۰ درجه به ما بدهند.

یک سبک نسبتاً جدید از پایگاه‌های داده به نام نواس کیوال^۱ برای پردازش حجم زیادی از داده‌های چندساختاری بوجود آمده است. با وجود این، اگرچه هَدوپ از تحلیل تاریخی *batch-style* در مقیاس بزرگ پشتیبانی می‌کند، پایگاه داده‌های نواس کیوال عمده‌تاً برای داده‌های گسسته که در بین حجم انبوهی از داده‌های چندساختاری و برای کاربردهای کلان‌داده کاربران نهایی انبارش شده‌اند در نظر گرفته شده است. این قابلیت در فناوری پایگاه داده رابطه‌ای وجود ندارد و به همین خاطر است که نمی‌توانند از کارایی مناسبی در یک مقیاس کلان‌داده برخوردار باشند.

در برخی موارد، نواس کیوال و هَدوپ در کنار یکدیگر کار می‌کنند. به عنوان مثال، اِچ‌بی‌سی یک پایگاه داده‌ی معروف نواس کیوال است که پس از بیگ‌تیبیل گوگل مدلسازی که شد که اغلب بر روی اِچ‌دی‌اف‌اس، سیستم فایل توزیع‌شده هَدوپ اجرا می‌شود تا جستجوی سریع و بدون تأخیر را در هَدوپ انجام دهد. نقطه ضعف اکثر پایگاه داده‌های نواس کیوال این است که انطباق اسید (تجزیه‌ناپذیری، هم‌خوانی، انزوا و پایایی^۲) را قربانی کارایی و مقیاس‌پذیری می‌کنند. بسیاری نیز فاقد ابزارهای پایش و مدیریت بالغ هستند. هر دوی این نواقص توسط جوامع نواس کیوال متن‌باز و بسیاری از عرضه‌کنندگان تجاری پایگاه داده‌های مختلف نواس کیوال در حال رفع هستند. پایگاه داده‌های نواس کیوال موجود در حال حاضر شامل اِچ‌بی‌سی، کاساندر^۳، مونگودی‌بی^۴، آکومولو^۵، ریاک^۶، کوچ‌دی‌بی^۷ و دینامودی‌بی^۸ می‌باشند. مورد کاربردی ۹.۳ کاربرد پایگاه داده‌های نواس کیوال را در ای‌بی‌بی نشان می‌دهد. اگرچه این مورد کاربردی تا حدودی قدیمی است ولی سعی کردیم تا با طرح این مورد نشان دهیم چگونه چندین مجموعه داده در کنار هم آورده می‌شوند. مورد کاربردی ۹.۴ یک کاربرد زیرساخت هَدوپ در رسانه اجتماعی

¹ NoSQL

² ACID (Atomicity, Consistency, Isolation, Durability)

³ Cassandra

⁴ MongoDB

⁵ Accumulo

⁶ Riak

⁷ CouchDB

⁸ DynamoDB

را برای کامپایل کردن پیکره‌های پیام‌های متنی در توئیتر برای درک و پشتیبانی انواع کاربرانی که به دنبال اطلاعاتی در خصوص بیماری‌های مزمن روحی هستند به تصویر کشیده است.

مورد کاربردی ۹.۳

راه‌حل کلان‌داده در ای‌پی

ای‌پی یکی از بزرگترین بازارهای آنلاین جهان است که امکان خرید و فروش تقریباً هر چیزی در آن وجود دارد. یکی عوامل کلیدی موفقیت خارق‌العاده‌ی ای‌پی، توانایی آن در تبدیل حجم‌های انبوه و عظیمی از داده‌ها برای تولید بینش‌های مفید است که مشتریان آن بتوانند مستقیماً در صفحات وب سایت ای‌پی که در آنها به گشت و گذار می‌پردازند استفاده کنند. ای‌پی برای اینکه بتواند رشد انفجاری داده‌هایش را مدیریت کند – روزانه میلیاردها فعالیت خواندن و نوشتن در مرکز داده‌های ای‌پی اتفاق می‌افتد – و به دلیل افزایش تقاضا برای پردازش داده‌ها با سرعتی سرسام‌آور، ای‌پی به راه‌حلی نیاز داشت که گلوگاه‌های معمول، مسائل مقیاس‌پذیری و محدودیت‌های تراکنشی مربوط به رویکردهای پایگاه‌های داده رابطه‌ای متداول نداشته باشند. شرکت همچنین نیازمند تحلیل سریع داده‌های ساخت‌یافته و ساخت‌نیافته‌ای که ضبط کرده بود.

راه‌حل: تحلیل‌ها و داده‌های بلادرنگ یکپارچه

نیازمندی‌های کلان‌داده‌ای ای‌پی موجب شد تا این شرکت به فناوری‌های نواس‌کیوال و به ویژه آپاچی کاساندر و دیتا‌استک اینترپرایز^۱ روی آورد. ای‌پی در کنار کاساندر و قابلیت‌های داده‌های پرسرعت آن به سمت تحلیل‌های آپاچی هُدوپ که دیتا‌استکس اینترپرایز عرضه می‌شد رفت. این راه‌حل شامل یک معماری مقیاس افقی بود که ای‌پی را قادر می‌ساخت تا چندین خوشه‌ی دیتا‌استکس اینترپرایز بر روی چندین مرکز داده مختلف و با استفاده از سخت‌افزارهای ضعیف و ارزان قیمت استفاده کند. نتیجه‌ی نهایی این‌که ای‌پی اکنون قادر است حجم انبوهی از داده‌ها را با هزینه‌ی بسیار پایین‌تر و سرعت بسی بالاتر پردازش کرده و به چیزی بیش از آنچه قبلاً با استفاده از سیستم‌های پرهزینه‌ی خود دست می‌یافت برسد. در حال حاضر ای‌پی بخش قابل توجهی از نیازهای مرکز داده‌های خود را – بیش از ۲۵۰ ترابایت انبارش – در خوشه‌های آپاچی کاساندر و دیتا‌استکس اینترپرایز مدیریت می‌کند. عوامل فنی دیگری که در این تصمیم ای‌پی برای استقرار دیتا‌استکس اینترپرایز نقش داشتند شامل مقیاس‌پذیری خطی دسترس‌پذیری بسیار بالا بدون حتی یک نقطه‌ی خرابی و کارایی نوشتاری بسیار بالای آن بود.

مدیریت موردهای کاربرد متنوع

ای‌پی از دیتا‌استکس اینترپرایز برای موردهای کاربرد بسیار مختلفی استفاده می‌کند. در ادامه مثال‌هایی از برخی روش‌ها که شرکت برای برآورده کردن نیازهای کلان‌داده‌ای خود با استفاده از قابلیت‌های تحلیل و مدیریت بسیار سریع داده‌ها بک می‌گیرد آورده شده است. طبیعتاً، ای‌پی از ترافیک نوشتاری بسیار بالایی برخوردار است که پیاده‌سازی کاساندر در دیتا‌استکس اینترپرایز موجب شده تا شرکت بتواند اینکار را با کارایی بسیار بالاتری نسبت به سایر راه‌حل‌های سیستم‌های مدیریت پایگاه داده رابطه‌ای یا نواس‌کیوال انجام دهد. ای‌پی در حال حاضر روزانه بیش از ۶ میلیارد فعالیت نوشتن در چندین خوشه کاساندر و بیش از ۵ میلیارد فعالیت خواندن (اکثراً آفلاین) دارد. یکی از موردهای کاربرد که توسط دیتا‌استکس اینترپرایز پشتیبانی می‌شود شامل کمی‌سازی داده‌های اجتماعی است که ای‌پی بر روی صفحات محصول خود نمایش می‌دهد. توزیع و گسترش کاساندر در دیتا‌استکس اینترپرایز، تمامی اطلاعات مورد نیاز برای شمارش داده‌های مربوط به «like»، «own» و «want» بر روی صفحات محصول ای‌پی را ذخیره می‌کند. همچنین داده‌های یکسانی را برای صفحه‌ی «Your Favorites» که حاوی تمامی اقلامی که کاربر دوست داشته، خریداری کرده

¹ DataStax Enterprise

خواسته است می‌باشد و کاساندر را کلیه‌ی خدمات و اطلاعات مورد نیاز صفحه‌ی «Your Favorites» را تأمین می‌کند. ای‌پی این داده‌ها را از طریق ویژگی شمارشگرهای مقیاس‌پذیر کاساندر ارائه می‌کند.

برقراری توازن در بارگذاری و در دسترس بودن برنامه‌های کاربردی از جنبه‌های مهم این مورد کاربرد خاص هستند. دیتاستکس اینترپرایز توانست انعطاف‌پذیری مورد نیاز معماران ای‌پی برای طراحی سیستمی که بتواند درخواست هر کاربر را به هر یک از مرکز داده‌ها ارسال کند و هر مرکز داده یک خوشه‌ی دیتاستکس اینترپرایز خاص خودش را داشته باشد که به آن خوشه‌ها گسترش می‌یابند فراهم کرد. این ویژگی در طراحی به ایجاد توازن در بار داده‌های ورودی از کاربر کمک کرده و هرگونه تهدید برای خرابی و از کار افتادگی برنامه کاربردی را از بین می‌برد. علاوه بر تأمین داده‌هایی که موجب تقویت صفحات وبی که مشتریان از آنها بازدید می‌کنند می‌شود، ای‌پی توانست تحلیل‌های سریع را با توانایی حفظ یک مرکز داده‌های مجزا برای گره‌های هدوپ در همان حلقه‌ی دیتاستکس اینترپرایز انجام دهد (شکل ۹/۶).

یک مورد کاربرد دیگر شامل «گراف سلیقه^۱» هانچ^۲ (یک شرکت خواهرخوانده‌ی ای‌پی) برای کاربران و اقلام ای‌پی است که توصیه‌هایی را بر اساس علایق و سلیق مشتریان ارائه می‌کند. وب سایت ای‌پی در واقع یک گراف بین تمامی کاربران و اقلامی که برای فروش موجود است می‌باشد. تمامی رویدادها (مزایده، خرید، فروش) توسط سیستم‌های ای‌پی ضبط و به صورت یک گراف در کاساندر ذخیره می‌شوند. در این کاربرد روزانه بیش از ۲۰۰ میلیون فعل نوشتن انجام شده و بیش از ۴۰ میلیارد تکه داده نگهداری می‌شود.

ای‌پی از دیتاستکس اینترپرایز همچنین برای تعداد زیادی از موردهای کاربرد سری‌های زمانی که پردازش آنی و لحظه‌ای حجم زیادی از داده‌ها در آنها بالاترین اهمیت را دارد استفاده می‌کند. این موردهای کاربرد شامل لاگینگ و ردیابی نوتیفیکیشن موبایل (هر زمان که ای‌پی یک نوتیفیکیشن به تلفن همراه ارسال می‌کند در کاساندر ثبت می‌شود)، تشخیص کلاهبرداری، لاگینگ ظرفیت بار درخواست/پاسخ SOA و لاگ‌ها و تحلیل‌های سرور ردلیزر^۳ (یک شرکت دیگر خواهرخوانده‌ی ای‌پی) می‌باشد.

نقطه‌ی مشترک تمامی این موردهای کاربرد، زمان فعال بودن سیستم است. ای‌پی به خوبی آگاه است که کسب‌وکارش باید همواره فعال باشد و دیتاستکس اینترپرایز نقش اساسی را با پشتیبانی از خوشه‌های با دسترسی‌پذیری بالا در این امر ایفا می‌کند. به گفته‌ی چی پاتل از معماران ای‌پی: «ما همیشه باید برای بازیابی فاجعه آماده باشیم. خیلی خوب است که کاساندر به ما اجازه می‌دهد تا چندین مرکز داده‌ی فعال داشته باشیم که بتوانیم داده‌ها را در هر کجا و هر زمان که می‌خواهیم بخوانیم و بنویسیم.»

سوالاتی برای بحث و گفتگو

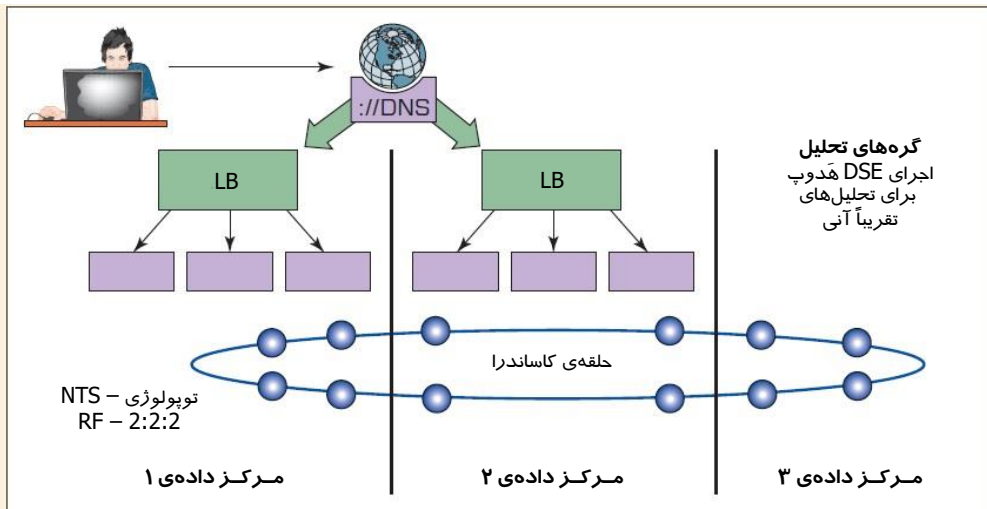
۴. چرا ای‌پی به یک راه‌حل کلان‌داده نیاز داشت؟
۵. چالش‌ها، راه‌حل پیشنهادی و نتایج بدست آمده چه بودند؟
۶. در مورد روش‌های دیگری که امریکن ایرلاینز می‌توانست برای حل مشکل اضافه پرداخت یا کم پرداخت به تأمین‌کنندگان بکار بگیرد بحث کنید.

Source: DataStax. Customer case studies. datastax.com/resources/casestudies/eBay (accessed October 2018).

¹ Taste Graph

² Hunch

³ RedLaser



شکل ۹.۶ استقرار چندین مرکز داده در ای‌پی Source: DataStax

درک کیفیت و قابلیت اطمینان اطلاعات پشتیبانی مراقبت سلامت در توئیتر

مورد کاربردی ۹.۴

در دنیای اینترنت امروزی، تمامی کاربران این قدرت را دارند تا به همان اندازه که از اطلاعات استفاده می‌کنند در تولید آن سهمی باشند. از این قدرت به روش‌های بسیاری استفاده می‌شود. در پلتفرم‌های شبکه‌های اجتماعی از قبیل توئیتر، کاربر می‌تواند اطلاعاتی را در مورد وضعیت سلامت‌شان ارسال کرده و همچنین اطلاعاتی در خصوص چگونگی مدیریت شرایط سلامت‌شان دریافت کند. بسیاری از کاربران از کیفیت اطلاعات منتشر شده در پلتفرم‌های شبکه‌های اجتماعی شگفت‌زده شده‌اند اگرچه توانایی نوشتن و انتشار اطلاعات سلامت در توئیتر برای بسیاری از کاربرانی که به دنبال رفع مشکل و بیماری هستند موضوع ارزشمندی محسوب می‌شود ولی صحت و اعتبار این اطلاعات به ویژه زمانی که این اطلاعات از جانب افراد که جزو گروه یا جامعه خاصی نیستند همواره مورد شک و تردید بوده است. بسیاری از کاربران پرسیده‌اند «چگونه صحت اعتبار اطلاعات مربوط به چگونگی مدیریت مسائل حیاتی از قبیل وضعیت سلامت‌تان که توسط افراد غیرخبره ارسال می‌شود را تأیید کنیم؟»

چه نوع کاربرانی، چه نوع اطلاعاتی را به اشتراک گذاشته و در مورد آن بحث می‌کنند؟ آیا کاربرانی که تعداد دنبال‌کنندگان زیادی دارند همان نوع اطلاعاتی را به اشتراک گذاشته و در مورد آن بحث می‌کنند که کاربرانی با تعداد دنبال‌کنندگان کم انجام می‌دهند؟ تعداد دنبال‌کنندگان یک کاربر با میزان تأثیرگذاری و نفوذ آن کاربر ارتباط دارد. ویژگی‌های اطلاعات بر حسب کیفیت و بی‌طرفی توئیتر ارسال شده سنجیده می‌شود. یک تیم از دانشمندان علم داده‌ها شروع به کشف روابط بین تعداد دنبال‌کنندگان یک کاربر و ویژگی‌های اطلاعاتی که آن کاربر منتشر کرده است نموده‌اند (آساموا و شاردا ۲۰۱۵).

راه‌حل

داده‌ها با استفاده از ای‌پی‌آی توئیتر از پلتفرم توئیتر استخراج شد. دانشمندان علم داده‌ها از مدل مدیریت داده‌ها و کشف دانش برای مدیریت و تحلیل این مجموعه‌ی بزرگ از داده‌ها استفاده کردند. این مدل برای مدیریت و تحلیل کلان‌داده‌های بدست آمده از یک پلتفرم شبکه اجتماعی بهینه‌سازی شده و شامل فازهای دستیابی به دانش حوزه، توسعه‌ی یک پلتفرم کلان‌داده‌ی مناسب ضبط و انبارش داده‌ها، پاکسازی داده‌ها، اعتبارسنجی داده‌ها، تحلیل داده‌ها و نتایج و استقرار بوده است.

فناوری بکار رفته

توئیت‌ها با استفاده از نسخه‌ی کلاویرا آپاچی هَدوپ استخراج، مدیریت و تحلیل شدند. چارچوب آپاچی هَدوپ دارای چندین زیرپروژه است که از انواع مختلف فعالیت‌های مدیریت داده پشتیبانی می‌کنند. به عنوان مثال، زیرپروژه‌ی آپاچی هابو از خواندن، نوشتن و مدیریت داده‌های بزرگ توئیت شده پشتیبانی می‌کند. ابزارهای تحلیل داده از قبیل جفی^۱ برای تحلیل شبکه اجتماعی و ابزار آر^۲ برای مدلسازی پیشگویانه بکار برده شده‌اند. آنها دو تحلیل را به طور موازی انجام دادند؛ تحلیل شبکه اجتماعی برای درک میزان تأثیر شبکه بر پلتفرم و متن‌کاوی برای درک محتوای پست‌های توئیت شده توسط کاربران.

چه چیزی یافت شد؟

همانطور که قبلاً گفته شد، توئیت‌های کاربران تأثیرگذار و غیرتأثیرگذار گردآوری و تحلیل شدند. نتایج نشان داد که کیفیت و بی‌طرفی اطلاعات منتشر شده توسط کاربران تأثیرگذار بیشتر از اطلاعات منتشر شده توسط کاربران غیرتأثیرگذار بوده است. آنها همچنین متوجه شدند که کاربران تأثیرگذار، جریان اطلاعات در شبکه را کنترل می‌کنند و سایر کاربران از عقیده و نظر آنها در مورد یک موضوع خاص پیروی می‌کنند. اختلاف واضحی بین نوع پشتیبانی اطلاعات از جانب کاربران تأثیرگذار و سایرین بوده است. کاربران تأثیرگذار اطلاعات بی‌طرفانه‌تری را در خصوص مدیریت بیماری‌ها – مسائلی از قبیل تشخیص بیماری، داروها و درمان‌های متداول – منتشر می‌کردند. کاربران غیرتأثیرگذار بیشتر اطلاعات احساسی و روش‌های مختلف برای کنار آمدن با اینگونه بیماری‌ها را مطرح می‌کردند. بنابراین، یک تفاوت فاحش بین کاربران تأثیرگذار و دیگران مشهود بوده است.

از منظر غیرکارشناسان، دانشمندان علم داده‌ها نشان می‌دهند با کمک به بیماران در شناسایی و بکارگیری منابع ارزشمند در وب برای مدیریت شرایط بیماری‌شان چگونه می‌توان تدارکات مراقبت سلامت را بهبود و ارتقاء بخشید. اینکار همچنین به افراد غیرکارشناس کمک می‌کند تا اطلاعات مراقبت سلامت را که ممکن است برای مدیریت شرایط سلامتی‌شان مناسب نباشند شناسایی و فیلتر کنند.

سوالاتی برای بحث و گفتگو

۱. نگرانی اصلی دانشمندان علم داده‌ها در خصوص اطلاعات سلامت که در پلتفرم توئیت منتشر می‌شود چه بود؟
۲. دانشمندان علم داده‌ها چگونه اطمینان حاصل کردند که اطلاعات غیرکارشناسی منتشر شده در رسانه‌های اجتماعی می‌تواند حاوی اطلاعات ارزشمندی نیز باشد؟
۳. آیا منطقی است که کاربران تأثیرگذار، اطلاعات بی‌طرفانه‌تری را نسبت به کاربران غیرتأثیرگذار منتشر می‌کنند؟ چرا؟

Sources: D. Asamoah & R. Sharda. (2015). "Adapting CRISP-DM Process for Social Network Analytics: Application to Healthcare." In *AMCIS 2015 Proceedings*. aisel.aisnet.org/amcis2015/BizAnalytics/GeneralPresentations/33/ (accessed October 2018). Sarasoohn-Kahn, J. (2008). *The Wisdom of Patients: Health Care Meets Online Social Media*. Oakland, CA: California HealthCare Foundation.

سوالاتی برای مرور بخش ۹.۴

۱۴. ویژگی‌های مشترک فناوری‌های نوظهور کلان‌داده‌ها چه می‌باشند؟
۱۵. مپ‌ردیوس چیست؟ چه کاری انجام می‌دهد؟ چگونه اینکار را انجام می‌دهد؟
۱۶. هَدوپ چیست؟ چگونه کار می‌کند؟
۱۷. کامپوننت‌های اصلی هَدوپ چه هستند؟ چه کاری انجام می‌دهند؟
۱۸. نواس کیوال چیست؟ چه جایگاهی در بین تحلیل کلان‌داده‌ها دارد؟

¹ Gephi

² R

شکی نیست که ظهور کلان داده‌ها به نحوی چشمگیر انبارسازی داده‌ها را دستخوش تغییر کرده و خواهد کرد. تا همین اواخر، پایگاه داده‌های تحلیلی سازمانی (فصل ۳)، مرکز ثقل تمامی فناوری‌های پشتیبان تصمیم بودند. امروزه، باید جایگاه‌شان را با فناوری جدیدالورود کلان داده‌ها شریک شوند. مسئله‌ای که همه جا مطرح است این است که آیا کلان داده و فناوری‌های توانمندساز آن مثل هُدوپ جایگزین انبارسازی داده‌ها و فناوری اصلی آن یعنی سیستم‌های مدیریت پایگاه داده رابطه‌ای خواهند شد؟ آیا ما شاهد یک چالش بین یک پایگاه داده‌ی تحلیلی و کلان داده هستیم (یا از نقطه نظر فناوری، هُدوپ در برابر سیستم‌های مدیریت پایگاه داده‌ی رابطه‌ای)؟ در این بخش ما توضیح خواهیم داد که چرا این سوالات هیچ پایه و اساسی ندارند و توجیه خواهیم کرد که چرا اینگونه انتخاب بین این یا آن در این برهه از زمان هیچگونه بازتابی از واقعیت ندارد.

در یک دهه‌ی گذشته یا بیشتر، ما شاهد بهبودهای چشمگیر در حوزه‌ی سیستم‌های پشتیبان تصمیم رایانه‌ای بوده‌ایم که به میزان زیادی به اعتبار پیشرفت‌های تکنولوژیکی و انبارسازی داده‌ها هم در بخش سخت‌افزار و هم نرم‌افزار برای ضبط، انبارش و تحلیل داده‌ها بوده است. قابلیت‌های پایگاه داده‌های تحلیلی نیز افزایش می‌یابد. برخی از این پیشرفت‌ها در حوزه پایگاه داده‌های تحلیلی شامل پردازش موازی انبوه (از یک یا چند پردازنده‌ی موازی تا تعداد بسیار زیاد)، شبکه‌های انبارسازی، ذخیره‌سازی حالت جامد^۱، پردازش درون پایگاه داده‌ای^۲، پردازش درون حافظه‌ای و پایگاه داده‌های ستونی^۳ می‌باشند. این پیشرفت‌ها به ما کمک کرده‌اند تا افزایش حجم و اندازه‌ی داده‌ها را تحت کنترل داشته باشیم و به شکلی مؤثر پاسخگوی نیازهای تحلیلی تصمیم‌گیرندگان نیز باشیم. آنچه موجب تغییر چشم‌انداز در سال‌های اخیر شده، تنوع و پیچیدگی داده‌ها است که موجب شده تا پایگاه داده‌های تحلیلی قادر نباشند یا به پای این تغییرات پیش بیانند. در اینجا مسئله‌ی حجم داده‌ها مطرح نیست بلکه سرعت و تنوع داده‌ها است که دنیای فناوری اطلاعات را ملزم کرده تا پارادایم جدیدی را که ما امروزه «کلان داده» می‌نامیم توسعه دهند. اکنون که ما این دو پارادایم - انبارسازی داده‌ها و کلان داده - که به نظر می‌رسد برای یک هدف یکسان رقابت می‌کنند - تبدیل داده‌ها به اطلاعات اقدام پذیر- را داریم کدامیک پیروز خواهند شد؟ آیا این سوال درست و منصفانه‌ای است؟ یا اینکه ما تصویر بزرگ را نادیده گرفته‌ایم؟ در این بخش ما سعی داریم تا بخشی از این مسئله‌ی پیچیده را روشن کنیم.

همانند بسیاری از نوآوری‌های پیشین در حوزه فناوری، تبلیغات مبالغه‌آمیز در مورد کلان داده‌ها و فناوری‌های توانمندسازی آن مثل هُدوپ و مَپر دیوس نیز بسیار شایع است. آماتورها و همچنین کارشناسان به واسطه‌ی این عقاید و نظرات گوناگون مستأصل شده‌اند. اما برخی دیگر متوجه شدند که مردم با این ادعا که هُدوپ جایگزین پایگاه داده‌های رابطه‌ای می‌شود و تبدیل به یک پایگاه داده‌ی تحلیلی جدید شده است متوجه اصل مطلب نشده‌اند. به آسانی متوجه می‌شویم که این ادعاها از کجا نشأت می‌گیرد زیرا هم هُدوپ و هم سیستم‌های پایگاه داده‌های تحلیلی می‌توانند به موازات، با حجم عظیم داده‌ها کار کرده و دارای معماری‌های غیراشتراکی^۴ باشند. در یک سطح مفهومی، می‌توان آنها را جایگزین یکدیگر دانست. واقعیت این است که آنها جایگزین هم نیستند و تفاوت‌های بین آنها بسیار بیشتر از تشابهات آنها است. اگر این دو جایگزین یکدیگر نیستند پس چگونه تصمیم بگیریم که کجا از هُدوپ و کجا از یک پایگاه داده‌ی تحلیلی استفاده کنیم؟

¹ Solid-State Storage

² In-database Processing

³ Columnar Databases

⁴ Shared-nothing architecture

همانطور که پیشتر این فصل گفته شد، هُدوپ نتیجه‌ی دستاوردهای جدید در فناوری‌های رایانه‌ای و شبکه‌ی انبارسازی است. هُدوپ با بکارگیری سخت‌افزارهای ضعیف و ارزان قیمت به عنوان زیربنای خود یک لایه‌ی نرم‌افزاری را ارائه می‌کند که در کل سطح شبکه گسترش یافته و همه‌ی آنها را به یک سیستم تبدیل می‌کند. در نتیجه، این معماری دارای عوامل متمایزکننده‌ی اساسی است:

هَدوپ یک مخزن و پالایشگاه برای داده‌های خام است.
هَدوپ یک آرشيو قدرتمند، مقرون به صرفه و فعال است.
بدین ترتیب، هَدوپ در هر دو انتهای چرخه‌ی طول عمر داده‌های بزرگ قرار می‌گیرد - ابتدا زمانیکه داده‌های خام تولید می‌شوند و در پایان زمانیکه داده‌ها بازنشسته می‌شوند ولی همچنان به طور موردی به آنها نیاز داریم.

۱. **هَدوپ به عنوان یک مخزن و پالایشگاه.** با ورود حجم عظیمی از کلان‌داده‌ها از منابع مختلف مثل حسگرها، ماشین‌ها، رسانه‌های اجتماعی و تعاملات کلیک استریم، اولین گام این است که تمامی داده‌ها را به نحوی مقرون به صرفه و مطمئن ضبط کنیم. زمانیکه حجم داده‌ها بالا باشد، استراتژی استفاده از یک سرور برای طولانی مدت جوابگو نخواهد بود. ریختن داده‌ها به داخل اِچ‌دی‌اِف‌اس، انعطاف‌پذیری لازم را به معمارها می‌دهد. آنها نه تنها می‌توانند صدها ترابایت را در یک روز ضبط کنند بلکه می‌توانند پیکربندی هَدوپ را برای شرایط افزایش یا کاهش مصرف داده‌ها تنظیم کنند. اینکار به دلیل استفاده از سخت‌افزارهای ارزان قیمت و مزایای اقتصادی متن‌باز با کمترین هزینه‌ی ممکن به ازای هر گیگابایت انجام می‌شود.

از آنجائیکه داده‌ها به جای شبکه‌های محلی انبارسازی بر روی یک انبار محلی ذخیره می‌شوند، دسترسی به داده‌های هَدوپ اغلب سریع‌تر است و شبکه را با جابجایی چندین ترابایت داده درگیر نمی‌کند. به محض ضبط داده‌های خام، از هَدوپ برای تصفیه و پالایش آن استفاده می‌شود. هَدوپ می‌تواند با بهره‌گیری از فناوری‌های تجاری یا دست‌نویس تبدیل داده‌ها به صورت یک موتور موازی «استخراج، تبدیل و بارگذاری بر روی استروئیدها» عمل کند. بسیاری از این تبدیل‌های داده‌های خام نیازمند تبدیل داده‌های بدون فرم پیچیده به قالب‌های ساخت‌یافته هستند. این به ویژه در مورد قالب داده‌های پیچیده‌ی حسگرها و کلیک استریم‌ها صادق است. در نتیجه، یک برنامه‌نویس باید کاه را از گندم جدا کند و سیگنال‌های ارزشمند را در بین نویزها شناسایی کند.

۲. **هَدوپ به عنوان یک بایگانی فعال.** در مصاحبه‌ای با ای‌سی‌ام در سال ۲۰۰۳، جیم گری مدعی شد که با دیسک‌های سخت می‌توان همچون نوارها کار کرد. اگرچه چندین سال دیگر طول می‌کشد تا بایگانی‌های نوارهای مغناطیسی کنار گذاشته شوند ولی امروزه هنوز بخش‌هایی از اطلاعات نوارهای مغناطیسی به سمت خوشه‌های هَدوپ هدایت می‌شوند. این تغییر به دو دلیل اساسی اتفاق می‌افتد. اول اینکه اگرچه ممکن است ذخیره‌سازی داده‌ها بر روی نوارهای مغناطیسی ارزان به نظر برسد ولی هزینه‌ی واقعی خودش را با دشواری بازیابی داده‌ها نشان می‌دهد. داده‌ها نه تنها به صورت آفلاین ذخیره می‌شوند بازیابی آنها ممکن است چندین ساعت به طور بینجامد ولی کارتریج‌های نوار مغناطیسی مستعد خرابی در طول زمان هستند که موجب از بین رفتن داده‌ها و تحمیل هزینه‌های

بیشتر به شرکت‌ها می‌شود. بدتر اینکه فرمت نوارها هر چند سال یکبار تغییر می‌کند که سازمان‌ها باید داده‌های خود را یا بر روی جدیدترین فرمت نوارها انتقال دهند یا اینکه ریسک این را بپذیرند که نتوانند داده‌ها را از روی نوارهای مهجور و منسوخ شده بازیابی کنند.

دوم اینکه همانطور که نشان داده شد نگهداری داده‌های قدیمی به صورت آنلاین و قابل دسترس کار ارزشمندی است. در مثال کلیک استریم، نگهداری داده‌های خام بر روی یک دیسک گِردام برای مدت زمان طولانی‌تر موجب می‌شود تا شرکت‌ها در صورت نیاز به اعمال محدودیت‌های جدید راحت‌تر بتوانند به داده‌ها رجوع کنند. جستجوی هزاران دیسک با هُدوپ بسیار سریع‌تر و آسان‌تر از چرخاندن صدها نوار مغناطیسی است. علاوه بر این، با دو برابر شدن چگالی دیسک‌ها در هر ۱۸ ماه، برای سازمان‌ها بسیار مقرون به صرفه‌تر است که داده‌های خام یا پالایش شده به ارزش چندین سال را بر روی اِچ‌دی‌اِف‌اس نگهداری کنند. بنابراین، شبکه انبارسازی هُدوپ هم برای پیش‌پردازش داده‌های خام و هم انبارسازی طولانی مدت داده‌ها مفید است. هُدوپ حقیقتاً یک «بایگانی فعال» است زیرا نه تنها داده‌ها را ذخیره کرده و از آنها محافظت می‌کند بلکه کاربران را قادر می‌سازد تا به آسانی و به سرعت از آنها داده‌ها ارزش استخراج کنند.

پس از نزدیک به ۳۰ سال سرمایه‌گذاری، اصلاح و رشد، فهرست ویژگی‌های موجود در یک پایگاه داده‌ی تحلیلی بسیار سرسام‌آور است. تفاوت‌های اصلی در این معماری که بر اساس فناوری پایگاه داده رابطه‌ای که از اسکیمایها و ابزارهای یکپارچه‌سازی هوش تجاری استفاده می‌کند عبارتند از

کارایی پایگاه داده تحلیلی
داده‌های یکپارچه که برای کسب‌وکار خلق ارزش می‌کنند
ابزارهای هوش تجاری تعاملی برای کاربران نهایی

۱. **کارایی پایگاه داده‌ی تحلیلی.** اندیس‌گذاری پایه که در پایگاه داده‌های متن‌باز از قبیل مای‌اس‌کیو‌ال^۱ یا پُست‌گرس^۲ وجود دارد یک ویژگی استاندارد است که برای بهبود زمان پاسخگویی به پرس و جویها یا اعمال محدودیت‌ها بر روی داده‌ها بکار می‌رود. فرم‌های پیشرفته‌تر از قبیل نماهای از پیش تولید شده^۳، اندیس‌های پیوند تجمعی^۴، اندیس‌های مکعبی^۵ و اندیس‌های پیوند پراکنده^۶ مزایای عملکردی متعددی را در پایگاه داده‌های تحلیلی فراهم می‌کنند. با وجود این، مهمترین بهبود کارایی تا به امروز، بهینه‌ساز هزینه^۷ می‌باشد. بهینه‌ساز به بررسی پرس و جویهای اس‌کیو‌ال ورودی پرداخته و چندین طرح را برای اجرای هر پرس و جو در سریع‌ترین زمان ممکن در نظر می‌گیرد. بهینه‌ساز با مقایسه‌ی درخواست اس‌کیو‌ال با طراحی پایگاه

¹ MySQL

² Postgres

³ Materialized view

⁴ Aggregate join indexes

⁵ Cube indexes

⁶ Sparse join indexes

⁷ Cost-based optimizer

داده و آمارهای گسترده‌ی داده‌ای که به شناسایی بهترین ترکیب گام‌های اجرایی کمک می‌کنند اینکار را انجام می‌دهد. در واقع، مثل این است که بهینه‌ساز دارای یک برنامه‌نویس نابغه است که هر پرس و جو را بررسی کرده و آن را برای دستیابی به بهترین عملکرد تنظیم می‌کند. در صورت نبود یک بهینه‌ساز یا آمارهای جمعیت‌شناختی داده‌ها، انجام یک پرس و جو حتی با وجود اندیس‌های بسیار ممکن است به جای چند دقیقه چندین ساعت به طول بینجامد. به همین دلیل، ارائه‌دهندگان پایگاه‌های داده دائماً انواع جدید ویژگی‌های اندیس‌ها، تقسیم‌بندی، آمار و بهینه‌سازها را می‌افزایند. برای ۳۰ سال گذشته، هر نسخه‌ی نرم‌افزار منتشر شده یک نسخه‌ی کارایی (عملکردی) بوده است. همانطور که در انتهای این بخش خواهیم دید، هُدوپ از لحاظ کارایی پرس و جو بر پایگاه داده‌های تحلیلی سنتی برتری جسته است.

۲. **یکپارچه‌سازی داده‌هایی که برای کسب‌وکار خلق ارزش می‌کنند.** هدف غائی هر پایگاه داده‌ی تحلیلی، پاسخگویی به سوالات اساسی کسب‌وکار است. داده‌های یکپارچه، تنها زیربنای لازم برای دستیابی به این هدف است. بیرون کشیدن داده‌ها از چندین حوزه کاربردی و وارد کردن آنها در یک مخزن دلیل وجودی پایگاه داده‌های تحلیلی است. طراحان مدل داده‌ها و معماران استخراج، تبدیل و بارگذاری مجهز به فراداده‌ها، ابزارهای پاکسازی داده‌ها باید فرمت داده‌ها، سیستم‌های منبع و معناری داده‌ها را به شکلی منطقی‌سازی کنند که قابل درک و قابل اطمینان باشند. بدین ترتیب یک فرهنگ واژگان مشترک در سازمان ایجاد می‌شود که اندازه‌گیری و درک یکسانی از مفاهیم اساسی از قبیل «مشتری»، «انتهای ماه» و «کشش قیمتی» بوجود می‌آید. در هیچ کجای یک مرکز داده‌های فناوری اطلاعات، داده‌ها همانند پایگاه داده‌های تحلیلی گردآوری، پاکسازی و ادغام نمی‌شوند.

۳. **ابزارهای هوش تجاری تعاملی.** ابزارهای هوش تجاری از قبیل میکرواستراتژی، تابلو، آی‌بی‌ام کونگوس و غیره دسترسی مستقیم به بینش‌های پایگاه داده‌های تحلیلی را برای کاربران کسب‌وکارها فراهم می‌کنند. اول اینکه کاربر می‌تواند گزارشات و تحلیل‌های پیچیده را به سرعت و به آسانی با استفاده از این ابزارها ایجاد کند. در نتیجه، در بسیاری از سایت‌های پایگاه داده‌های تحلیلی روند به سمت سلف‌سرویس کاربر نهایی است. کاربران می‌توانند گزارشات بیشتری را از نسبت به آنچه دپارتمان فناوری اطلاعات می‌تواند ارائه کنند تقاضا کنند. ولی مهمتری از سلف‌سرویس این است که کاربران کاملاً با داده‌ها آشنا شوند. آنها می‌توانند در عرض چند دقیقه یک گزارش را ایجاد کرده، متوجه شوند که یک سنج یا فیلتر را فراموش کرده‌اند، اصلاحات لازم را به عمل آورده و گزارش را مجدداً ایجاد کنند. این فرایند منجر به تغییرات اساسی در درک کاربران از کسب‌وکار و فرایند تصمیم‌گیری آنها می‌شود. اول اینکه کاربران از پرسیدن سوالات بدیهی خودداری کرده و شروع به طرح سوالات استراتژیک پیچیده‌تر می‌کنند. معمولاً، هرچه یک گزارش پیچیده‌تر و استراتژیک‌تر باشد، درآمد و صرفه‌جویی هزینه‌ی بیشتری برای کاربر دارد. این موجب می‌شود که برخی کاربران تبدیل به «کاربران قدرتمند» در یک شرکت شوند. این افراد به جادوگرانی تبدیل می‌شوند که ارزش کسب‌وکار را از داده‌ها نشان می‌دهند و اطلاعات استراتژیک ارزشمندی را به کارکنان اجرایی ارائه می‌کنند. هر پایگاه داده‌ی تحلیلی چیزی در حدود ۲ تا ۲۰ کاربر قدرتمند دارد. همانطور که در بخش ۹.۸ اشاره شد، تمامی این ابزارهای هوش تجاری برای اینکه بتوانند در حد و اندازه‌های انبار داده‌های بزرگ سرویس بدهند از هُدوپ استفاده می‌کنند.

اگرچه هَدوپ و پایگاه داده‌های تحلیلی در چندین حوزه با یکدیگر تفاوت دارند ولی حوزه‌هایی نیز وجود دارد که این دو را نمی‌توان مشخصاً از یکدیگر متمایز کرد. در این حوزه‌ها هر کدام از ابزارها می‌تواند راه‌حل درست و مناسبی باشد. انتخاب یکی از این دو ابزار به نیازمندی‌ها و اولویت‌های سازمان بستگی دارد. در بسیاری موارد، هَدوپ و پایگاه داده‌های تحلیلی در یک زنجیره تأمین اطلاعات با یکدیگر کار می‌کنند و به همان اندازه نیز اغلب یکی از ابزارها برای یک کاربرد خاص بهتر عمل می‌کند (آوادالا و گراهام ۲۰۱۲). جدول ۹.۱ پلتفرم‌های ترجیحی را برای تعدادی از نیازمندی‌های متداول مشاهده شده ارائه کرده است.

جدول ۹.۱ چه زمان از کدام پلتفرم استفاده کنیم – هَدوپ یا پایگاه داده تحلیلی

هَدوپ	پایگاه داده تحلیلی	نیازمندی
	<input checked="" type="checkbox"/>	گزارشات کم تأخیر تعاملی با OLAP
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	الزام انطباق با ANSI 2003 SQL
<input checked="" type="checkbox"/>		پیش‌پردازش یا اکتشاف از داده‌های ساخت‌نیافته خام
<input checked="" type="checkbox"/>		بایگانی آنلاین جایگزین نوارهای مغناطیسی
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	داده‌های منسجم و پاکسازی شده با کیفیت بالا
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	۱۰۰ تا ۱۰۰۰ کاربر همزمان
<input checked="" type="checkbox"/>		کشف ارتباطات ناشناخته در داده‌ها
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	منطق فرایند پیچیده‌ی موازی
	<input checked="" type="checkbox"/>	تحلیل سنگین CPU
<input checked="" type="checkbox"/>		سیستم، کاربران و حکمرانی داده‌ها
<input checked="" type="checkbox"/>		اجرای موازی تعداد زیادی از زبان‌های برنامه‌نویسی
<input checked="" type="checkbox"/>		اکتشاف نامحدود و آزاد در جعبه‌ی شن
	<input checked="" type="checkbox"/>	تحلیل داده‌های موقت
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	انطباق با قوانین و امنیت بالا

چندین سناریو برای استفاده‌ی ترکیبی از فناوری‌های انبارسازی داده‌ها بر اساس سیستم‌های مدیریت پایگاه داده‌های رابطه‌ای و هَدوپ وجود دارد که در اینجا به برخی از آنها اشاره می‌کنیم (وایت ۲۰۱۲):

۱. **استفاده از هَدوپ برای انبارسازی و بایگانی داده‌های چندساختاری.** برای استخراج داده‌های مورد نیاز از هَدوپ به منظور تحلیل آنها توسط سیستم‌های مدیریت پایگاه داده‌های رابطه‌ای می‌توان از یک کانکتور که به سیستم مدیریت پایگاه داده رابطه‌ای وصل می‌شود استفاده کرد. در صورتیکه سیستم‌های مدیریت پایگاه داده‌های رابطه‌ای از کارکردهای مَپر دیوس پشتیبانی کنند، از این کارکردها می‌توان برای استخراج داده‌ها استفاده کرد. آداپتور وِنِتِج-هَدوپ به عنوان مثال از کارکردهای

اس کیوال-مپردیوس برای بارگذاری سریع و دوطرفه‌ی داده‌ها بین اچ‌دی‌اف‌اس و ونتج دیتابیس^۱ استفاده می‌کند. سپس داده‌های بارگذاری شده در ونتج دیتابیس را می‌توان با استفاده از اس کیوال و مپردیوس تحلیل کرد.

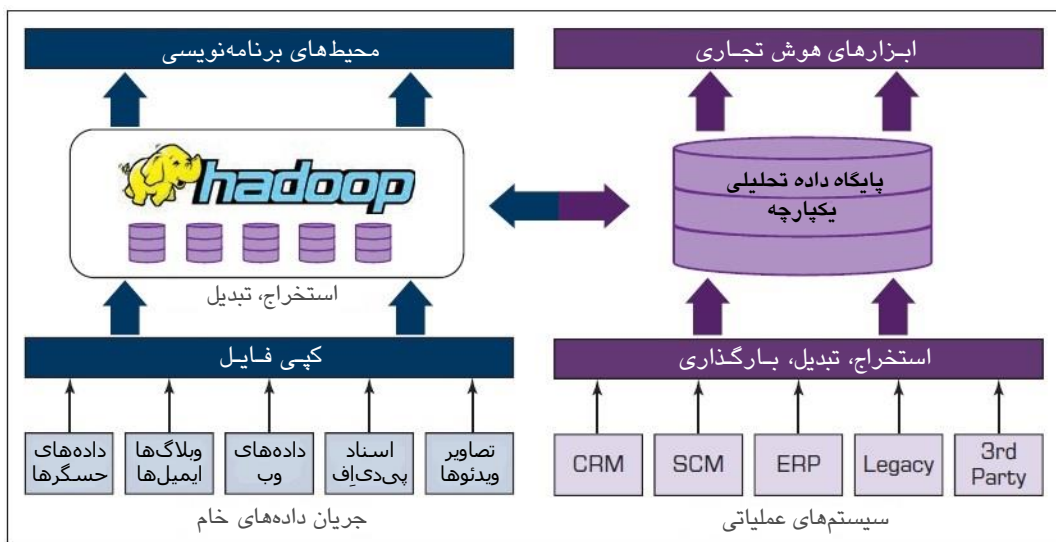
۲. **استفاده از هُدوپ برای فیلترینگ، تبدیل و/یا ادغام داده‌های چندساختاری.** یک کانکتور مثل آداپتور ونتج-هَدوپ را می‌توان برای استخراج نتایج پردازش هَدوپ از سیستم‌های مدیریت پایگاه داده رابطه‌ای برای تحلیل‌های بیشتر بکار برد.

۳. **استفاده از هَدوپ برای تحلیل حجم بزرگی از داده‌های چندساختاری و انتشار نتایج تحلیلی.** در این کاربرد، هَدوپ به عنوان پلتفرم تحلیلی عمل می‌کند ولی نتایج را می‌توان در محیط‌های متداول انبارسازی داده‌ها، یک انبار کارگروهی مشترک یا یک رابط کاربری معمول منتشر کرد.

۴. **استفاده از یک سیستم مدیریت پایگاه داده‌ی رابطه‌ای که قابلیت‌های مپردیوس را به عنوان یک پلتفرم رایانش تحقیقاتی ارائه می‌کند.** دانشمندان علم داده‌ها می‌توانند از سیستم‌های مدیریت پایگاه داده‌های رابطه‌ای (به عنوان مثال سیستم ونتج دیتابیس) برای تحلیل ترکیبی از داده‌های ساخت‌یافته و چندساختاری (که از هَدوپ بارگذاری شده) با استفاده از آمیزه‌ای از کارکردهای تحلیلی مپردیوس و کارکردهای پردازشی اس کیوال استفاده کنند.

۵. **استفاده از یک ابزار پرس و جوی فرانت‌اند برای دسترسی و تحلیل داده‌ها.** در اینجا داده‌ها هم در هَدوپ و هم در سیستم مدیریت پایگاه داده رابطه‌ای ذخیره می‌شوند.

این سناریوها از محیطی پشتیبانی می‌کنند که هَدوپ و سیستم مدیریت پایگاه داده رابطه‌ای به طور جداگانه از یکدیگر کار می‌کنند و از این نرم‌افزار ارتباطی برای تبادل داده‌ها بین این دو سیستم استفاده می‌شود (شکل ۹.۷). جهت‌گیری صنعت طی چندین سال آینده احتمالاً به سمت وابستگی بیشتر بین فناوری‌های انبار داده‌های مبتنی بر سیستم‌های مدیریت پایگاه داده‌های رابطه‌ای و هَدوپ هم از لحاظ سخت‌افزاری و هم نرم‌افزاری خواهد بود.



¹ Vantage Database

شکل ۹.۷ همزیستی هَدوپ و پایگاه داده تحلیلی

Source: "Hadoop and the Data Warehouse: When to Use Which, teradata, 2012." Used with permission from Teradata Corporation

اینگونه یکپارچه‌سازی‌ها مزایای متعددی را به همراه دارند از جمله حذف نیاز به نصب و نگهداری سیستم‌های متعدد، کاهش جابجایی داده‌ها، ارائه‌ی یک انبار فراداده برای توسعه‌ی برنامه‌های کاربردی و ارائه‌ی یک رابط برای کاربرانی کسب‌وکار و ابزارهای تحلیلی. پرونده‌ی آغازین یک مثال از یکپارچه‌سازی داده‌های بدست آمده از یک پایگاه داده‌ی تحلیلی معمول و دو مجموعه داده‌ی ساخت‌نیافته‌ی مختلف در هَدوپ را برای کاربردهای تحلیلی و دستیابی به بینش در خصوص تعاملات مشتریان شرکت پیش از آنکه حساب‌هایشان را ببندند به تصویر کشید. به عنوان یک مدیر، بینش‌هایی که می‌توانید از داده‌ها استخراج کنید برایتان بسیار اهمیت دارد صرف‌نظر از اینکه داده‌ها در یک پایگاه داده‌ی تحلیلی ساخت‌یافته یا یک خوشه‌ی هَدوپ ذخیره شده باشند.

سوالاتی برای مرور بخش ۹.۵

۱. چالش‌های پایگاه داده‌های تحلیلی و کلان‌داده‌ها چه هستند؟ آیا ما شاهد پایان عصر پایگاه داده‌های تحلیلی هستیم؟
۲. موارد کاربرد کلان‌داده و هَدوپ چه هستند؟
۳. موارد کاربرد انبارسازی داده‌ها و سیستم‌های مدیریت پایگاه داده‌های رابطه‌ای چه هستند؟
۴. تحت چه سناریوهایی هَدوپ و سیستم‌های مدیریت پایگاه داده‌های رابطه‌ای با یکدیگر همزیستی دارند؟

هَدوپ از چارچوب پردازش دسته‌ای بهره برده و فاقد قابلیت‌های پردازش بلادرنگ و آنی می‌باشد. با رشد و تکامل رایانش کلان‌داده‌ها، تحلیل‌های درون‌حافظه به عنوان یک تکنیک پردازش نوظهور برای تحلیل داده‌های ذخیره شده در پایگاه داده‌های درون‌حافظه مطرح شده است. از آنجائیکه دسترسی به داده‌های ذخیره شده در داخل حافظه بسیار سریع‌تر از داده‌های موجود در دیسک سخت می‌باشد، پردازش درون‌حافظه بسیار کارآمدتر از پردازش دسته‌ای است. این همچنین امکان تحلیل آنی داده‌های استریمینگ را فراهم می‌کند.

تحلیل‌های درون‌حافظه‌ای دارای کاربردهای متعددی در مواردی است که نیاز به اجرای کم تأخیر و سریع باشد و می‌تواند به ساخت داشبوردهای بلادرنگ برای دستیابی به بینش بهتر و تصمیم‌گیری سریع‌تر کمک کند. کاربردهای بلادرنگ شامل درک رفتار و مشارکت‌دادن مشتری، پیش‌بینی قسمت سهام، بهینه‌سازی قیمت بلیت هواپیما، پیش‌بینی کلاهبرداری و موارد دیگر است.

متداول‌ترین ابزار پشتیبانی در پردازش درون‌حافظه‌ای، آپاچی اسپارک می‌باشد که یک موتور تحلیل یکپارچه است که قادر به پردازش داده‌های دسته‌ای و استریمینگ است. آپاچی اسپارک که در دانشگاه برکلی کالیفرنیا در سال ۲۰۰۹ توسعه یافت از

محاسبات درون حافظه‌ای برای دستیابی به کارایی بالا در پردازش داده‌ها در مقیاس بزرگ استفاده می‌کند. آپاچی اسپارک با بکارگیری یک رویکرد پردازش درون حافظه‌ای بسیار سریع‌تر از آپاچی هُدوپ اجرا می‌شود. علاوه بر این، می‌توان از آن به عنوان جایگزین پوسته‌های جاوا، اسکالا، پایتون، آر و اس کیوال برای نوشتن برنامه‌های کاربردی مدیریت داده‌ها و یادگیری ماشینی استفاده کرد. آپاچی اسپارک قابلیت اجرا بر روی آپاچی هُدوپ، آپاچی مِزوس^۱، کابرنِتس^۲، به صورت مستقل یا در فضای ابر را دارد. همچنین، می‌تواند به منبع داده‌های بیرونی مختلف از قبیل اِچ‌دی‌اف‌اس، آلوکسیو^۳، آپاچی کاساندر، آپاچی اِچ‌بیس، آپاچی هایو و ... نیز متصل شود.

از آپاچی اسپارک می‌توان برای ایجاد برنامه‌های کاربردی یادگیری ماشین، رایانش مه^۴، گراف، استریمینگ و تحلیل‌های بلادرنگ استفاده کرد. چندین فعال بزرگ در بازار در بخش تحلیل‌ها از آپاچی اسپارک استفاده می‌کنند که مثال‌هایی از آنها شامل اوبر، پینترست، نتفلیکس، یاهو و ای‌بی‌بی هستند. اوبر از آپاچی اسپارک برای تشخیص سفرهای قلابی استفاده می‌کند. پینترست از آپاچی اسپارک برای سنجش بلادرنگ میزان درگیرسازی کاربران استفاده می‌کند. موتور توصیه‌ی نتفلیکس نیز از قابلیت‌های آپاچی اسپارک بهره می‌برد. یاهو که یکی از اولین استفاده‌کنندگان آپاچی اسپارک بود از آن برای ایجاد برنامه‌های کاربردی هوش تجاری استفاده کرده است. نهایتاً، ای‌بی‌بی از آپاچی اسپارک برای مدیریت داده‌ها و پردازش استریم استفاده می‌کند.

استفاده از پردازش زبان طبیعی برای تحلیل بازخورد مشتری در اظهارنظرات

تریپ‌آدوایزر

مورد کاربردی ۹.۵

پلتفرم وب تریپ‌آدوایزر حاوی اطلاعاتی در مورد هتل‌ها، رستوران‌ها و دیگر محتوای مرتبط با سفر است. این پلتفرم همچنین شامل فرم‌های تعاملی در مورد سفر است که اظهارنظرات مشتریان و مدیران در مورد هتل‌ها و رستوران‌ها را ثبت می‌کند. تریپ‌آدوایزر برای بهبود محتوای این فرم تصمیم گرفت تا تگ‌هایی را برای هر کدام از جاذبه‌های مسافرتی از جمله هتل‌ها و رستوران‌ها در نظر بگیرد. تریپ‌آدوایزر با ارسال یک فرم نظرسنجی سوالاتی «بلی» و «خیر» برای هر یک از آنها به گردآوری اظهارنظرها پرداخت. پاسخ‌های بدست آمده از مشتریان موجب شد تا تگ‌های متفاوتی بدست آید. شرکت با استفاده از اطلاعات گذشته تصمیم گرفت تا یک مدل رگرسیون لجستیک را برای پیش‌بینی پاسخ‌های «بلی» و «خیر» از مشتریان آینده و پیش‌بینی تگ‌ها ایجاد کند. مسئله بسیار پیچیده است زیرا هر مکان ویژگی‌های خاص خودش را دارد. با استفاده از تجارب مشتریان قبلی در فرم اظهارنظر، اطلاعات متنی برای آموزش مدل بکار گرفته شد. در آموزش مدل از اظهارنظرها در خصوص مکان‌هایی که دارای تگ بودند و همچنین اظهارنظرهایی که فاقد تگ بودند استفاده شد.

¹ Apache Mesos

² Kubernetes

³ Alluxio

⁴ Fog Computing

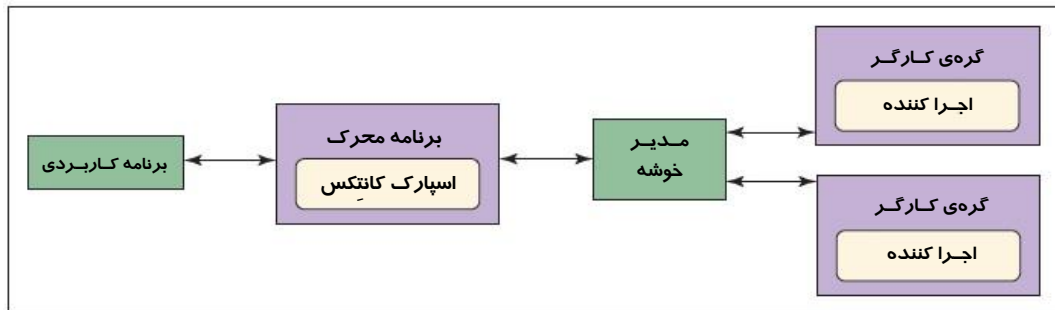
برای ساخت این مدل بر روی کلان‌داده‌های حاوی میلیون‌ها اظهار نظر و صدها تگ، شرکت از آپاچی اسپارک استفاده کرد. با استفاده از پردازش موازی و پردازش درون حافظه‌ی اسپارک، مدل برای هر تگ در هر مکان آموزش داده شد. داده‌ها به تفکیک مکان بخش‌بندی و تفکیک شدند تا میزان ارتباطات بین گره‌ها به حداقل برسد. کل فرایند به نحوی کارآمد پیاده‌سازی و اجرا شد.

سوالاتی برای بحث و گفتگو

۱. مدلسازی پیشگویانه چگونه به تریپ‌آوایزر کمک کرد؟
۲. چرا از اسپارک استفاده شد؟

Compiled from: Palmucci, J., "Using Apache Spark for Massively Parallel NLP," at <http://engineering.tripadvisor.com/usingapache-spark-for-massively-parallel-nlp/> (accessed October 2018) and Dalinina, R., "Using Natural Language Processing to Analyze Customer Feedback in Hotel Reviews," at www.datascience.com/resources/notebooks/data-science-summarize-hotelreviews (accessed October 2018)

آپاچی اسپارک یک چارچوب ارباب/برده است. در این چارچوب یک برنامه محرک وجود دارد که با گره‌ی ارباب به نام مدیر خوشه ارتباط برقرار می‌کند گره‌های کارگر را مدیریت می‌کند. اجزای وظایف در گره‌های کارگر اتفاق می‌افتد که مجری‌ها در آنجا اجرا می‌شوند. نقطه‌ی ورودی موتور را اسپارک کانتیکس می‌نامند که به عنوان یک پل ارتباطی بین برنامه کاربردی و محیط اجرایی اسپارک عمل می‌کند (شکل ۹.۸). همانطور که گفته شد اسپارک در حالات مختلفی می‌تواند اجرا شود. در حالت مستقل، یک برنامه کاربردی بر روی گره‌های مختلف در خوشه‌ای که توسط خود اسپارک مدیریت می‌شود اجرا می‌شود. با وجود این، در حالت هَدوپ، اسپارک از خوشه‌ی هَدوپ برای اجرای کارها استفاده کرده و از چارچوب اِچ‌دی‌اِف‌اِس و مَپر‌دیوس بهره می‌گیرد.



شکل ۹.۸ معماری آپاچی اسپارک

یک کامپوننت بسیار مهم آپاچی اسپارک، مجموعه داده‌ی توزیع‌شده‌ی انعطاف‌پذیر^۱ یا آردی‌دی نام دارد که لینچ (سطر‌بندی)، مدیریت حافظه، تحمل‌پذیری در برابر خطا و بخش‌بندی داده‌ها را در تمامی گره‌های یک خوشه مدیریت می‌کند. آردی‌دی چندین کارکرد تبدیل از قبیل نگاشت، فیلتر و پیوند را که بر روی آردی‌دی‌های موجود انجام می‌شوند ارائه می‌کند تا یک آردی‌دی جدید ایجاد کند. تمامی تبدیل‌ها در اسپارک با تأخیر انجام می‌شوند بدین معنی که اسپارک این عملیات را تا زمانیکه هیچ کارکردی بر روی داده‌ها انجام نشود اجرا نمی‌کند. هر کارکردی اعم از شمارش یا کاهش، پس از اجرا یک مقدار را چاپ

¹ Resilient Distributed Datasets (RDD)

کرده یا باز می‌گرداند. این روش را **سنجش موردی**^۱ می‌نامند. در اسپارک استریمینگ^۲، مجموعه‌هایی از آر‌دی‌ها به نام دی‌استریم^۳ برای پردازش داده‌های استریمینگ بکار گرفته می‌شوند.

در این بخش، ما توضیح خواهیم داد که چگونه کار با آپاچی اسپارک را بر روی نسخه‌ی کوئیک استارت کلاودرا هدوپ شروع کنید. اینکار ابتدا با بارگذاری آخرین نسخه‌ی ماشین مجازی کلاودرا کوئیک استارت آغاز شده و به اجرای پرس و جوی اسپارک خاتمه می‌یابد.

بررسی نیازمندی‌های سخت‌افزاری و نرم‌افزاری

- یک رایانه با سیستم عامل میزبان ۶۴ بیتی (ویندوز یا لینوکس) و حداقل ۱۲ گیگابایت حافظه‌ی داخلی برای عملکرد بهتر
- برنامه‌ی مجازی‌ساز وُرك‌استیشن پلیبر: آخرین نسخه (رایگان) مجازی‌ساز را از www.vmware.com/products/workstation-player/workstation-player-evaluation.html بارگذاری کنید.
- ۸ گیگابایت حافظه برای ماشین مجازی / ۲۰ گیگابایت فضای دیسک آزاد
- برنامه 7-Zip: بسته‌ی کوئیک استارت کلاودرا را با استفاده از برنامه‌ی (7-Zip) (www.7zip.org) بارگذاری کنید.

گام‌هایی که برای شروع به کار با اسپارک بر روی ماشین مجازی کوئیک استارت کلاودرا باید برداشت:

۱. ماشین مجازی کوئیک استارت کلاودرا را از آدرس www.cloudera.com/downloads/quickstart_vms/5-13.html بارگذاری کنید.
۲. فایل بارگذاری شده که حاوی یک ماشین مجازی‌ساز است را با استفاده از نرم‌افزار 7-Zip باز کنید.
۳. فایل مجازی‌ساز پلیبر وُرك‌استیشن را نصب و اجرا کنید. اکنون، تصاویر ماشین مجازی کلاودرا را از طریق پلیبر مجازی‌ساز
۴. پیش از روشن کردن ماشین مجازی، باید تنظیمات حافظه و پردازنده را تعیین کنید. تنظیم پیش فرض حافظه برای ماشین مجازی معادل ۴ گیگابایت است. برای تغییر تنظیمات بر روی گزینه‌ی "Edit virtual machine setting" کلیک کنید. حتماً میزان حافظه را بیش از ۸ گیگابایت و تعداد هسته‌ی پردازنده را ۲ قرار دهید.
۵. ماشین را اجرا کنید. هدوپ و کامپوننت‌های آن بر روی سیستم عامل لینوکس سنت‌اَس نصب شده است.
۶. یک نام کاربری پیش‌فرض "cloudera" با کلمه‌ی عبور "cloudera" ایجاد شده است.
۷. بر روی فضای دسکتاپ ماشین مجازی، "Launch Cloudera Express" را باز کنید. موتور در عرض چند دقیقه شروع به کار می‌کند.

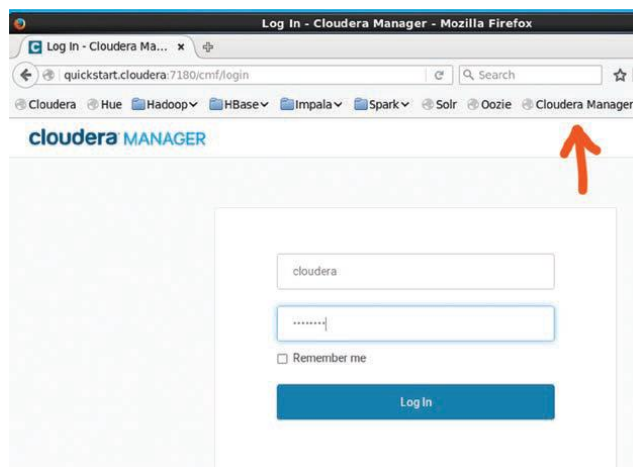
¹ Lazy Evaluation

² Spark Streaming

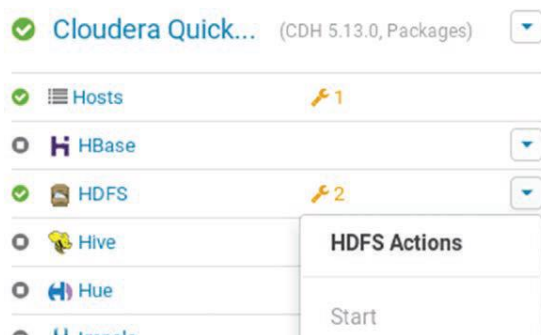
³ Dstream



۸. به محض شروع، مرورگر وب داخل ماشین مجازی را باز کنید. یک آیکون در قسمت بالای کلاودرا دسکتاپ مشاهده می کنید.



۹. با استفاده از نام کاربری "cloudera" و کلمه عبور "cloudera" وارد بخش Cloudera Manager شوید.
 ۱۰. برای استفاده از اِچ دی اِف اِس و مَپ ردیوس باید دو سرویس را راه اندازی کنیم: اِچ دی اِف اِس و یارن^۱ با استفاده از منوی پایین افتادنی در جلوی آنها.



۱۱. برای روشن کردن اسپارک، سرویس اسپارک را اجرا کنید.
 ۱۲. برای اجرای پرس و جوها در اسپارک، می توانیم از برنامه نویسی پایتون یا اسکالا استفاده کنیم. با کلیک راست بر روی دسکتاپ ماشین مجازی، ترمینال را باز کنید.
 ۱۳. برای ورود به پوسته پایتون طبق تصویر زیر، تایپ کنید *pyspark* و برای خروج از پوسته پایتون تایپ کنید *exit()*.

¹ YARN

```

cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart Desktop]$ pyspark
Python 2.6.6 (r266:84292, Jul 23 2015, 15:22:56)
[GCC 4.4.7 20120313 (Red Hat 4.4.7-11)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel).
18/08/13 15:45:34 WARN util.Utils: Your hostname, quickstart.cloudera resolves to
a loopback address: 127.0.0.1; using 192.168.60.131 instead (on interface eth1
)
18/08/13 15:45:34 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to ano
ther address
Welcome to

 version 1.6.0

Using Python version 2.6.6 (r266:84292, Jul 23 2015 15:22:56)
SparkContext available as sc, HiveContext available as sqlContext.
>>> exit()
[cloudera@quickstart Desktop]$

```

۱۴. برای ورود به پوسته‌ی اسکالا اسپارک مطابق تصویر زیر تایپ کنید *spark-shell* و برای خروج از پوسته‌ی اسکالا اسپارک تایپ کنید *exit*.

```

[cloudera@quickstart Desktop]$ spark-shell
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel).

```

۱۵. از اینجا به بعد، ما گام‌های اجرای یک برنامه کاربردی اسپارک برای شمارش تعداد واژگان را توضیح می‌دهیم که شمارش واژگان به صورت تعاملی انجام می‌شود. ما این برنامه را در محیط پوسته‌ی اسکالا اسپارک اجرا می‌کنیم. برای استفاده تعاملی از اسپارک استریمینگ، باید پوسته‌ی اسکالا اسپارک را با حداقل دو رشته اجرا کنیم. برای اینکه تایپ می‌کنیم *spark-shell-master local[2]*

```

cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart Desktop]$ spark-shell --master local[2]
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel).

```

ا. سپس برای اجرای یک برنامه‌ی استریمینگ، باید این سه کلاس را یک به یک همانند شکل زیر وارد کنیم.

```

scala> import org.apache.spark.streaming.StreamingContext
import org.apache.spark.streaming.StreamingContext

scala> import org.apache.spark.streaming.StreamingContext._
import org.apache.spark.streaming.StreamingContext._

scala> import org.apache.spark.streaming.Seconds
import org.apache.spark.streaming.Seconds

scala>

```

ب. پس از وارد کردن کلاس‌های مورد نیاز، یک اسپارک استریمینگ کانکتکس *SSS* را با دسته‌های زمانی ۱۰ ثانیه‌ای مشابه شکل زیر ایجاد کنید.

```

scala> val sss = new StreamingContext(sc,Seconds(10))
sss: org.apache.spark.streaming.StreamingContext = org.apache.spark.streaming.StreamingContext@40eaf1e

```

ج. یک استریم گسسته (دی استریم) که یک مفهوم انتزاعی پایه در اسپارک استریمینگ است را برای خواندن متن از پورت ۱۱۱۱ ایجاد کنید.

```
scala> val firststream = sss.socketTextStream("localhost",1111)
firststream: org.apache.spark.streaming.dstream.ReceiverInputDStream[String] = org.ap
ache.spark.streaming.dstream.SocketInputDStream@7fa649a
```

د. برای شمارش دفعات تکرار واژگان در یک استریم، کدهای نوشته شده در مپردیوس که در تصویر زیر نشان داده شده‌اند را اجرا می‌کنیم. سپس دستور `count.print()` را برای چاپ شمارش واژگان در دسته‌ی زمانی ۱۰ ثانیه‌ای بکار می‌گیریم.

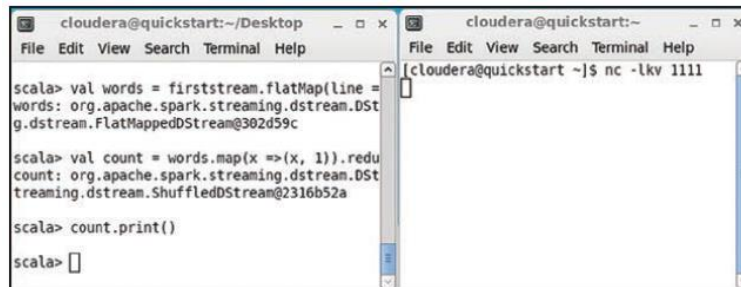
```
scala> val words = firststream.flatMap(_.split(" "))
words: org.apache.spark.streaming.dstream.DStream[String] = org.apache.spark.str
eaming.dstream.FlatMappedDStream@7d9f5f12

scala> val pairs = words.map(word => (word, 1))
pairs: org.apache.spark.streaming.dstream.DStream[(String, Int)] = org.apache.sp
ark.streaming.dstream.MappedDStream@61614d4e

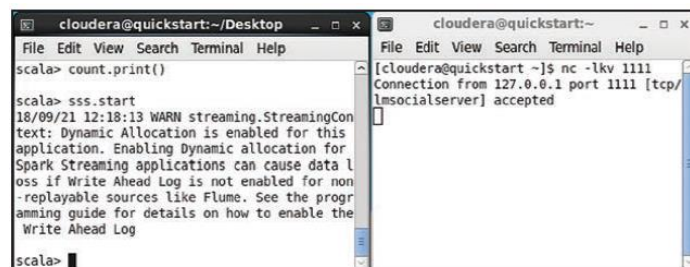
scala> val count = pairs.reduceByKey(+)
count: org.apache.spark.streaming.dstream.DStream[(String, Int)] = org.apache.sp
ark.streaming.dstream.ShuffledDStream@60519731

scala> count.print()
```

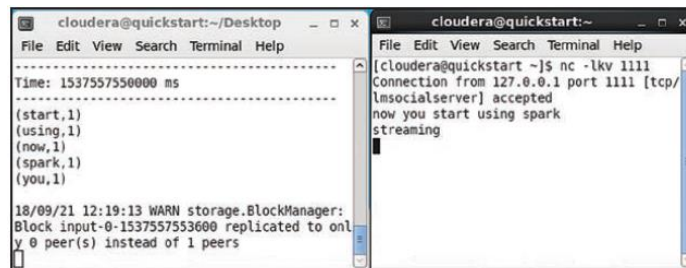
ه. در این مقطع، یک ترمینال جدید باز کرده و دستور `nc -lkv 1111` را مشابه ترمینال سمت راست تصویر زیر اجرا کنید.



و. برای شروع کانتکس استریمینگ، دستور `sss.start()` را در پوسته‌ی اسپارک اجرا کنید. اینکار موجب برقراری اتصال بین دی استریم `sss` با سوکت (ترمینال سمت راست) می‌شود.



ز. در گام آخر، دستور `sss.awaitTermination()` را در پوسته‌ی اسپارک اجرا کرده و شروع به تایپ کردن همان کلمات در ترمینال سمت راست کنید. پس از هر ۱۰ ثانیه، زوج کلمات شمارش شده در پوسته‌ی اسپارک محاسبه خواهند شد.



ح. برای توقف فرایند، ترمینال سمت راست را بسته و کلید `CTRL + C` را در ترمینال سمت چپ فشار دهید.

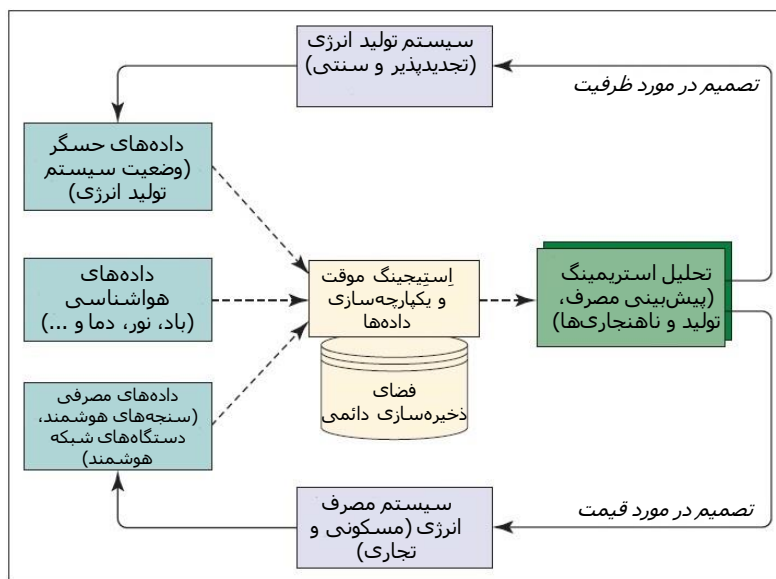
ط. از آنجائیکه ممکن است نیاز باشد تا برنامه را مجدداً اجرا کنید، تمامی دستورها را در اینجا آورده‌ایم.

```
spark-shell --master local[2]
import org.apache.spark.streaming.StreamingContext
import org.apache.spark.streaming.StreamingContext._
import org.apache.spark.streaming.Seconds
val sss = new StreamingContext(sc,Seconds(10))
val firststream = sss.socketTextStream("localhost",1111)
val words = firststream.flatMap(_.split(" "))
val pairs = words.map(word => (word, 1))
val count = pairs.reduceByKey(_+_ )
count.print()
sss.start()
sss.awaitTermination()
```

سوالاتی برای مرور بخش ۹.۶

۱۶. برخی از ویژگی‌های منحصر بفرد اسپارک در مقایسه با هادوپ چه هستند؟
۱۷. مثال‌هایی از شرکت‌هایی که از آپاچی اسپارک استفاده کرده‌اند ارائه کنید. چند مثال آنلاین ذکر کنید.
۱۸. تمرین این بخش را اجرا کنید. از این تمرین چه آموختید؟

عبارات استریمینگ و پیوسته^۱ برای اکثر مردم یک معنا دارند و در بسیاری از موارد مترادف یکدیگر بکار برده می‌شوند. با وجود این، در مقوله‌ی سیستم‌های هوشمند، این دو با یکدیگر تفاوت دارند (جوناس ۲۰۰۷). تحلیل استریمینگ شامل بکارگیری منطق سطح تراکنش بر روی مشاهدات بلادرنگ است. قواعدی که برای این مشاهدات بکار گرفته می‌شوند شامل مشاهدات قبلی تا زمانی که در پنجره‌ی قبلی رخ دهند می‌باشد؛ این پنجره‌ها دارای اندازه‌ی اختیاری هستند (مثلاً ۵ ثانیه برای ۱۰,۰۰۰ مشاهده). تحلیل پیوسته از طرف دیگر به ارزشیابی هر مشاهده‌ی ورودی در برابر تمامی مشاهدات قبلی می‌پردازد و بحث اندازه‌ی پنجره در اینجا مطرح نیست. پی بردن به رابطه‌ای بین مشاهدات جدید و تمامی مشاهدات قبلی ما را قادر به کشف آنی و لحظه‌ای بینش‌های جدید می‌کند. اما تحلیل‌های استریمینگ و پیوسته مزایا و معایب و جایگاه خاص خودشان را در دنیای تحلیل کسب‌وکار دارند.



شکل ۹.۹ یک مورد کاربرد تحلیل استریمینگ در صنعت انرژی

¹ Perpetual

کاوش جریان داده‌ای^۲ به عنوان یک فناوری توانمندساز برای تحلیل جریان‌ها عبارتست از فرایند استخراج الگوها و ساختارهای دانشی جدید برای رکوردهای پیوسته و سریع داده‌ها. همانگونه که در فصل داده‌کاوی دیدیم، در روش‌های رایج داده‌کاوی، داده‌ها باید در یک قالب فایل مناسب گردآوری و سازماندهی شده و سپس به شکل بازگشتی برای یادگیری الگوهای زیرین پردازش شوند. برعکس، یک جریان داده‌ای یک جریان پیوسته از یک توالی منظم از نمونه‌ها^۳ یا مثال‌ها است که در بسیاری از کاربردهای کاوش جریان داده‌ای می‌توان آنها را فقط یکبار یا برای دفعات معدودی با استفاده از قابلیت‌های انبارش و رایانش خوانده و پردازش نمود. مثال‌هایی از جریان‌های داده‌ای شامل داده‌های حسگرها، ترافیک شبکه‌ی رایانه‌ای، مکالمات تلفنی، تعاملات دستگاه‌های خودپرداز، جستجوهای وب و داده‌های مالی هستند. کاوش جریان داده‌ای یک رشته‌ی فرعی از داده‌کاوی، یادگیری ماشینی و اکتشاف دانش محسوب می‌شود. هدف در بسیاری از کاربردهای کاوش جریان داده‌ای، پیش‌بینی دسته یا ارزش نمونه‌های جدید در جریان داده با توجه به دانش موجود در مورد عضویت دسته یا مقادیر نمونه‌های قبلی در جریان داده‌ای است. تکنیک‌های تخصصی یادگیری ماشینی (عمدتاً مشتق از تکنیک‌های یادگیری ماشینی متداول هستند) را می‌توان برای یادگیری این وظیفه‌ی پیش‌بینی از روی مثال‌های برجسب‌گذاری شده به صورت خودکار بکار برد. یک مثال از اینگونه روش پیش‌بینی توسط دیلن، کلتک و کیم (۲۰۰۵) توسعه یافت که یک مدل درخت تصمیم را با استفاده از یک زیرمجموعه از داده‌ها در یک زمان به تدریج ساخته و اصلاح کردند.

¹ Critical Events Processing

² Data Stream Mining

³ Instance

استفاده از تحلیل جریان داده‌ای (استریمینگ) به دلیل قدرت ایجاد بینش در لحظه و کمک به تصمیم‌گیرندگان در مواجهه با رویدادها به محض وقوع و کمک به سازمان‌ها در رسیدگی به مسأله پیش از آنکه تبدیل به مشکل شوند از یک روند روزافزون با رشد نمایی برخوردار است. در ادامه برخی از حوزه‌های کاربردی تحلیل‌های جریان داده‌ای را بیان می‌کنیم.

شرکت‌هایی مثل آمازون و ای‌بی‌سی سعی دارند تا بیشترین استفاده را از داده‌هایی که در زمان حضور مشتریان در وبسایت‌هایشان گردآوری می‌کنند بدست آورند. هر بازدید صفحه، هر محصولی که مشاهده می‌شود، هر جستجویی که انجام می‌شود و هر کلیک برای دستیابی به بیشترین ارزش از بازدیدهای کاربران ضبط و تحلیل می‌شود. تحلیل اینگونه جریان داده‌ها در صورتیکه به سرعت انجام شود می‌تواند جستجوگرها را به خریداران و خریداران را به معناداران به خرید تبدیل کند. زمانیکه ما از یک وبسایت تجارت الکترونیک بازدید می‌کنیم، حتی از وبسایت‌هایی که ما در آنها عضو نیستیم، پس از چند کلیک در اینجا و آنجا، پیشنهادات بسیار جذابی برای خرید کالاها به ما عرضه می‌شود. در پشت صحنه، تحلیل‌های پیشرفته در حال متراکم کردن داده‌های بلادرنگی که از کلیک‌های ما و هزاران نفر دیگر بدست آمده‌اند هستند تا «متوجه شوند» که ما به چه چیزی علاقه داریم و با ارائه‌ی پیشنهادات خلاقانه، از آن اطلاعات حداکثر بهره را ببرند.

حجم داده‌هایی که از رکوردهای جزئیات تماس شرکت‌های مخابراتی بدست می‌آید مبهوت کننده است. اگرچه از این اطلاعات اخیراً برای صدور صورتحساب استفاده شده ولی دانش بسیار زیادی در درون این کلان‌داده‌ها نهفته است که شرکت‌های مخابراتی به تازگی متوجه آن شده و سعی در بهره‌برداری از آنها دارند. به عنوان مثال، داده‌های رکوردهای جزئیات تماس را می‌توان برای پیشگیری از ریزش کاربران از طریق شناسایی شبکه‌های تماس‌گیرندگان، اینفلوئنسرها، لیدرها و دنبال‌کنندگان در آن شبکه‌ها و اتخاذ اقدامات کنشگرانه بر روی آن اطلاعات مورد تحلیل قرار داد. همانگونه که می‌دانیم، اینفلوئنسرها و لیدرها بر تغییر ادراک و برداشت دنبال‌کنندگان نسبت به سرویس‌دهندگان در شبکه‌هایشان تأثیر می‌گذارند و شرکت‌های مخابراتی برای مدیریت بهتر پایگاه مشتریان‌شان سعی در شناسایی اینفلوئنسرها و لیدرها و مشترکین شبکه‌هایشان دارند. علاوه بر تحلیل ریزش کاربران، از اینگونه اطلاعات می‌توان برای عضوگیری جدید و پیشینه کردن ارزش اعضای فعلی استفاده کرد.

سینلر فورس از جریان داده‌ای برای ارتقاء ارزش مشتریان استفاده می‌کند

مورد کاربردی ۹.۶

سینلر فورس^۱ خدمات بازاریابی ابری خود را توسعه و گسترش داده و شامل ویژگی‌هایی از قبیل «امتیاز پیشگویانه^۲» و «مخاطبان پیشگویانه^۳» با نام «سفر پیشگویانه‌ی بازاریابی ابری^۴» می‌باشد. این ویژگی از جریان داده‌های بلادرنگ برای ارتقاء تعامل و درگیرسازی مشتریان آنلاین استفاده می‌کند. ابتدا، یک پردیکتیو اسکور منحصر بفرد به هر مشتری داده می‌شود. این امتیاز بر اساس عوامل متعددی از جمله تاریخچه‌ی مرور آنها در وبسایت، کلیک بر روی یک پیوند نامه الکترونیکی،

¹ Salesforce

² Predictive Score

³ Predictive Audience

⁴ Marketing Cloud Predictive Journey

سابقه‌ی خرید، میزان پول پرداختی، آخرین زمان خرید آنها و پاسخگویی به یک کمپین از طریق نامه الکترونیکی یا آگهی تبلیغاتی محاسبه می‌شود. پس از محاسبه‌ی امتیاز مشتریان، آنها به گروه‌های متفاوتی تقسیم می‌شوند. این گروه‌ها بر اساس رفتارهای پیشگویانه‌ای که به آنها تخصیص یافته است طرح‌ها و اهداف متفاوتی برایشان در نظر گرفته می‌شود. امتیازها و بخش‌بندی‌ها به صورت روزانه تغییر کرده و به‌روز‌آوری می‌شوند و نقشه راه بهتری را در اختیار شرکت‌ها قرار می‌دهند تا به هدف و پاسخ مورد نظرشان دست پیدا کنند. این راه‌حل‌های بازاریابی بسیار دقیق‌تر بوده و روش‌های شخصی‌سازی شده‌تری را برای شرکت‌ها به منظور حفظ و نگهداشت مشتریان‌شان ارائه می‌کنند.

سوالاتی برای بحث و گفتگو

۳. آیا ممکن است جریان‌های داده‌ای در یک صنعت نامربوط باشد؟
۴. علاوه بر نگهداشت مشتری، استفاده از تحلیل‌های پیشگویانه چه مزایای دیگری دارد؟

از این مورد چه می‌توانیم بیاموزیم؟

شرکت‌ها با تحلیل داده‌های کسب شده در اینجا و اکنون می‌توانند سریع‌تر در مورد مشتریان‌شان پیش‌بینی کرده و تصمیم بگیرند. بدین ترتیب اطمینان حاصل می‌شود که شرکت‌ها مشتریان درست را هدف قرار داده، جذب و نگهداری کرده و بیشترین ارزش را از آنها کسب می‌کنند. داده‌هایی که هفته‌ی گذشته بدست آمده‌اند به اندازه‌ی داده‌هایی که امروز به دست می‌آیند مفید و سودمند نیستند. استفاده از داده‌های مرتبط موجب می‌شود تا تحلیل پیشگویانه‌ی دقیق‌تر و کارآمدتری داشته باشیم.

Sources: M. Amodio. (2015). "Salesforce Adds Predictive Analytics to Marketing Cloud. Cloud Contact Center."

www.cloudcontactcenterzone.com/topics/cloud-contact-center/articles/413611-salesforce-adds-predictive-analytics-marketing-cloud.htm (accessed

October 2018). J. Davis. (2015). "Salesforce Adds New Predictive Analytics to Marketing Cloud." *Information Week*. www.informationweek.com/big-data/big-data-analytics/salesforce-adds-new-predictive-analytics-to-marketing-cloud/d/d-id/1323201 (accessed October 2018). D. Henschen. (2016).

"Salesforce Reboots Wave Analytics, Preps IoT Cloud." *ZD Net*. www.zdnet.com/article/salesforce-reboots-wave-analytics-preps-iot-cloud/ (accessed October 2018).

برای ارزیابی اثربخشی کمپین‌های بازاریابی می‌توان جریان‌های پیوسته‌ی داده‌هایی که از رکوردهای جزئیات تماس‌ها بدست می‌آیند را با داده‌های رسانه‌های اجتماعی (تحلیل احساسات) ترکیب کرد. بینش بدست آمده از این جریان داده‌ها را می‌توان برای واکنش سریع در برابر تأثیرات نامساعد (که می‌توانند منجر به از دست رفتن مشتریان شوند) یا تقویت اثرات مثبت (که ممکن است منجر به بیشترین میزان خرید مشتریان فعلی یا جذب مشتریان جدید شود) بکار برد. علاوه بر این، فرایند دستیابی به بینش از رکوردهای جزئیات تماس را می‌توان با استفاده از رکوردهای جزئیات پروتکل اینترنت برای شبکه‌های داده‌ای نیز تکرار کرد. از آنجائیکه اکثر شرکت‌های مخابراتی هر دو نوع این خدمات را ارائه می‌دهند، یک بهینه‌سازی کل‌نگر از تمامی این کمپین‌های بازاریابی می‌تواند به منافع خارق‌العاده‌ای در بازار منتهی شود. مورد کاربردی ۹.۶ یک مثال از دستیابی سیل‌زفوس. کام به درک بهتری از مشتریان‌شان بر اساس تحلیل کلیک‌استریم‌ها را ارائه کرده است.

جریان کلان‌داده‌ها فرصت‌های عالی را برای بهبود پیشگیری از وقوع جرائم، اعمال قانون و ارتقاء امنیت ارائه می‌کند. آنها پتانسیل بی‌نظیری را در مورد کاربردهای امنیتی که در فضا ساخته شوند ارائه می‌کنند، مانند آگاهی موقعیتی بلادرنگ^۱، نظارت چندوجهی^۲،

¹ Real-time situational awareness

² Multimodal Surveillance

تشخیص امنیت سایبری^۱، استراق سمع قانونی^۲، نظارت ویدئویی و تشخیص چهره (زیکوپولوس و همکاران ۲۰۱۳). به عنوان یکی از کاربردهای اطمینان اطلاعات، سازمان‌ها می‌توانند از تحلیل‌های استریمینگ (جریان داده‌ای) برای تشخیص و پیشگیری از نفوذ به شبکه‌ها، حملات سایبری و فعالیت‌های خرابکارانه با استریمینگ و تحلیل لاگ‌های شبکه و سایر منابع پایش فعالیت‌های اینترنتی استفاده کنند.

به دلیل استفاده‌ی روزافزون از سنجه‌های هوشمند، حجم داده‌های بلادرنگ که توسط تأسیسات برق گردآوری می‌شود با یک رشد نمایی در حال افزایش است. سنجه‌هایی که اطلاعات را هر ما یکبار می‌خواندند امروزه هر ۱۵ دقیقه یکبار (یا حتی بیشتر) حجم انبوهی از داده‌های بسیار ارزشمند را برای تأسیسات برق گردآوری و جمع می‌کنند. این سنجه‌های هوشمند و سایر حسگرهایی که در سرتاسر شبکه‌ی برق قرار دارند اطلاعات را برای تحلیل بلادرنگ به مراکز کنترل ارسال می‌کنند. اینگونه تحلیل‌ها به شرکت‌های خدمات عمومی (آب، برق، گاز) در بهینه‌سازی تصمیمات زنجیره‌ی تأمین (مثلاً تنظیم ظرفیت، گزینه‌های شبکه‌ی توزیع، خرید یا فروش بلادرنگ) بر اساس به‌روزترین و آخرین داده‌های مصرف مشتری و الگوهای تقاضا کمک می‌کنند. علاوه بر این، شرکت‌های خدمات عمومی می‌توانند داده‌های شرایط آب و هوایی و سایر داده‌های شرایط طبیعی را برای بهینه‌سازی تولید برق از منابع جایگزین و پیش‌بینی بهتر تقاضای انرژی در بخش‌های مختلف جغرافیایی در ابزارهای تحلیلی خود ترکیب کرده و بکار بگیرند. در شرکت‌های خدمات عمومی دیگر از قبیل سازمان آب و گاز نیز مزایای مشابهی وجود دارد.

شرکت‌های خدمات مالی از بارزترین مثال‌هایی هستند که تحلیل جریان کلان‌داده‌ای در آنها می‌تواند منجر به تصمیمات بهتری و سریع‌تر، مزیت رقابتی و نظارت بر مقررات شود. توانایی تحلیل سریع حجم بالایی از داده‌ها با تأخیر بسیار کم در بازارها و کشورها، مزیت فوق‌العاده‌ای برای اتخاذ تصمیمات خرید/فروش است که نهایتاً منجر به بهره و منفعت مالی زیادی می‌شود. علاوه بر تصمیمات بهینه‌ی خرید/فروش، تحلیل جریان داده‌ای می‌تواند به شرکت‌های خدمات مالی در پایش بلادرنگ تجارت به منظور تشخیص کلاهبرداری و سایر فعالیت‌های غیرقانونی کمک کند.

دستگاه‌های پزشکی عصر مدرن (مثلاً الکتروکاردیوگرام‌ها و تجهیزات اندازه‌گیری فشار خون، سطح اکسیژن خون، سطح قند خون و دمای بدن) قادر به تولید جریان بی‌نهایت ارزشمندی از داده‌های حسگرها/تشخیص بیماری با سرعت بسیار بالا هستند. کنترل و بهره‌برداری از این داده‌ها و تحلیل بلادرنگ آنها بر خلاف هر رشته‌ی دیگری مزایای بسیاری به همراه دارد – آنچه که ما آن را «هرگ و زندگی» می‌نامیم. علاوه بر کمک به ارتقاء کارایی و اثربخشی شرکت‌های مراقبت سلامت (و سودآورتر و رقابتی‌تر شدن آنها) تحلیل جریان داده‌ای همچنین به بهبود شرایط بیمار و نجات زندگی آنها نیز کمک می‌کند.

بسیاری از سیستم‌های بیمارستانی در سرتاسر جهان در حال توسعه‌ی سیستم‌های سلامت و زیرساخت‌های مراقبت آینده‌نگر هستند. هدف این سیستم‌ها بهره‌گیری کامل از آن چیزی است که فناوری برای ارائه دارد. استفاده از دستگاه‌های سخت‌افزاری که قادر به تولید داده‌های با کیفیت بسیار بالا با سرعت بالا هستند به همراه رایانه‌های ابرسریع که قادر به تحلیل چندین جریان داده‌ای هستند، احتمال

¹ Cyber-Security detection

² Legal Wiretapping

زنده ماندن بیمار را با تشخیص سریع آنومالی‌ها افزایش می‌دهد. هدف این سیستم‌ها، کمک به تصمیم‌گیرندگان در اتخاذ تصمیمات سریع‌تر و بهتر می‌باشد.

دولت‌ها در سرتاسر جهان سعی در یافتن روش‌هایی برای افزایش کارایی (استفاده‌ی بهینه از منابع محدود) و اثربخشی (ارائه‌ی خدمات مورد نیاز مردم) هستند. با تبدیل رویکرد دولت الکترونیک به یک جریان اساسی به همراه استفاده و دسترسی گسترده به رسانه‌های اجتماعی، حجم بسیار انبوهی از داده‌ها (ساخت‌یافته و ساخت‌نیافته) در دسترس ارگان‌های دولتی هستند. استفاده‌ی درست و به موقع از این جریان‌های کلان‌داده‌ای، عامل تمایز ارگان‌های کنشگر و بسیار کارآمد از ارگان‌هایی است که هنوز از روش‌های سنتی برای واکنش به موقعیت‌هایی که رخ می‌دهند استفاده می‌کنند. یک روش دیگر که ارگان‌های دولتی می‌توانند از قابلیت‌های تحلیلی بلادرنگ بهره ببرند، مدیریت بلایای طبیعی از قبیل کولاک‌ها، گردبادها و آتش‌سوزی‌ها از طریق نظارت داده‌های استریمینگ بدست آمده از رادارها، حسگرها و دیگر لوازم و دستگاه‌های تشخیص هوشمند است. آنها همچنین می‌توانند از رویکردهای مشابهی برای پایش کیفیت آب، هوا و الگوهای مصرف و تشخیص ناهنجاری‌ها پیش از آنکه تبدیل به مشکلات اساسی شوند استفاده کنند. حوزه‌ی دیگری که ارگان‌های دولتی می‌توانند از تحلیل جریان داده‌ای در آن استفاده کنند، مدیریت ترافیک در شهرهای شلوغ و پر ازدحام است. با استفاده از داده‌های بدست آمده از دوربین‌های کنترل جریان ترافیک، داده‌های جی‌پی‌اس بدست آمده از وسائل نقلیه‌ی تجاری و حسگرهای ترافیک که در بزرگراه‌ها تعبیه شده‌اند، ارگان‌ها می‌توانند توالی چراغ راهنمایی و خطوط جریان ترافیک را تغییر دهند.

سوالاتی برای مرور بخش ۹.۷

۱۹. منظور از یک جریان یا استریم (در دنیای کلان‌داده) چیست؟
۲۰. انگیزه‌های اصلی برای استفاده از تحلیل جریان داده‌ای چه هستند؟
۲۱. منظور از تحلیل جریان داده‌ای یا استریم چیست؟ چه تفاوتی با تحلیل‌های معمول دارد؟
۲۲. پردازش رویداد بحرانی چیست؟ چه ارتباطی به تحلیل استریم یا جریان داده‌ای دارد؟
۲۳. کاوش جریان داده‌ای را تعریف کنید. کاوش جریان داده‌ای چه چالش‌های دیگری را به همراه دارد؟
۲۴. سودمندترین صنایع برای استفاده از تحلیل جریان داده‌ای یا استریم کدام‌ها هستند؟
۲۵. چگونه می‌توان از تحلیل جریان داده در تجارت الکترونیک استفاده کرد؟
۲۶. علاوه بر مواردی که در این بخش ذکر شد آیا می‌توانید صنایع دیگر و/یا موارد کاربرد دیگری را برای تحلیل جریان داده‌ای ذکر کنید؟
۲۷. در مقایسه با تحلیل‌های متداول، آیا به نظر شما تحلیل جریان داده‌ای می‌تواند موارد کاربرد دیگری در عرصه‌ی تحلیل‌های کلان‌داده داشته باشد؟ چرا؟

<http://mattturck.com/wp->

[content/uploads/2018/07/Matt_Turck_FirstMark_Big_Data_Landscape_2018_Final_reduced-768x539.png](http://mattturck.com/wp-content/uploads/2018/07/Matt_Turck_FirstMark_Big_Data_Landscape_2018_Final_reduced-768x539.png)

با در نظر گرفتن این پیش‌بینی‌های احتیاطی، یک روش فشرده برای گروه‌بندی عرضه‌کنندگان کلان‌داده‌ها این است که از دسته‌بندی کلی زیر استفاده کنیم:

- عرضه‌کنندگان خدمات زیرساخت
- عرضه‌کنندگان راه‌حل‌های تحلیلی
- عرضه‌کنندگان هوش تجاری که به سمت کلان‌داده‌ها حرکت کرده‌اند

hortonworks.com

cloudera.com

mapr.com

¹ Jeff Hammerbacher

² MapR

1. Datastax
2. Nutanix
3. VMWare
4. Amazon Web Services
5. Microsoft Azure
6. Google Cloud Platform
7. IBM Cloud Platform
8. Amazon AWS
9. Amazon Neptune
10. AutoML
11. Dell EMC
12. IBM Big Insight
13. SAP's Hanna
14. Oracle Big Data
15. Netezza
16. InfoSphere Warehouse

1. Datastax
2. Nutanix
3. VMWare
4. Amazon Web Services
5. Microsoft Azure
6. Google Cloud Platform
7. IBM Cloud Platform
8. Amazon AWS
9. Amazon Neptune
10. AutoML
11. Dell EMC
12. IBM Big Insight
13. SAP's Hanna
14. Oracle Big Data
15. Netezza
16. InfoSphere Warehouse

-
- 1 Datastax
 - 2 Nutanix
 - 3 VMWare
 - 4 Amazon Web Services
 - 5 Microsoft Azure
 - 6 Google Cloud Platform
 - 7 IBM Cloud Platform
 - 8 Amazon AWS
 - 9 Amazon Neptune
 - 10 AutoML
 - 11 Dell EMC
 - 12 IBM Big Insight
 - 13 SAP's Hanna
 - 14 Oracle Big Data
 - 15 Netezza
 - 16 InfoSphere Warehouse

مورد کاربردی ۹.۷

استفاده از رسانه‌های اجتماعی برای پیش‌بینی شیوع آنفولانزا

بیماری‌های عفونی بار سنگینی را بر دوش نظام سلامت عمومی ایالات متحده تحمیل می‌کنند. ظهور اچ‌آی‌وی/ایدز در اواخر دهه ۷۰، بیماری هم‌گیر آنفولانزای خوک در سال ۲۰۰۹ و بیماری اپیدمی آنفولانزای H3N2 در زمستان ۲۰۱۲-۲۰۱۳، شیوع بیماری ویروس ایبولا در سال ۲۰۱۵ و ویروس زیکا در سال ۲۰۱۶ نشان داد که انسان تا چه اندازه در برابر بیماری‌های مسری آسیب‌پذیر است. ما تقریباً هر ساله شاهد شیوع آنفولانزا به شکل‌های مختلف هستیم و تأثیرات گوناگونی نیز به همراه دارند. تأثیر سالانه‌ی شیوع آنفولانزای فصلی در ایالات متحده به طور متوسط ۶۶۰،۶۱۰ سال از عمر مردم را کم کرده، ۳/۱ میلیون روز بستری، ۳۱/۴ میلیون ویزیت سرپایی و در مجموع ۸۷/۱ میلیارد دلار فشار اقتصادی را تحمیل می‌کند. در نتیجه‌ی این رشد فزاینده، تکنیک‌ها و فناوری‌های جدید تحلیل داده‌ها که قادر به تشخیص، ردیابی، نگاشت و مدیریت اینگونه بیماری‌ها باشند در سال‌های اخیر وارد صحنه شده‌اند. به طور خاص، سیستم‌های نظارتی دیجیتالی به واسطه‌ی ظرفیت کشف الگوهای جستجوی سلامت عمومی و تبدیل این اکتشافات به استراتژی‌های اقدام‌پذیر بسیار امیدبخش بوده‌اند.

این پروژه نشان داد که از رسانه‌های اجتماعی می‌توان به عنوان روشی مؤثر برای تشخیص زودهنگام شیوع آنفولانزا استفاده کرد. ما از یک پلتفرم کلان‌داده برای استفاده از داده‌های توئیتر به منظور پایش فعالیت آنفولانزا در ایالات متحده استفاده کرده‌ایم. روش‌های تحلیل کلان‌داده‌ی ما شامل متن‌کاوی، کاوش زمانی و مکانی است. در تحلیل زمانی ما بررسی کردیم که آیا از داده‌های توئیتر نیز می‌توان برای کنون‌بینی شیوع آنفولانزا استفاده کرد یا خیر. در تحلیل مکانی، ما شیوع آنفولانزا را بر روی

¹ Cognos

² InfoSphere Streams

³ SAS Viya

⁴ Qlik

⁵ Spotfire

ویژگی مکانی جغرافیایی داده‌های توییتر به منظور شناسایی نقاط حساس شیوع آنفلانزا نگاشت کردیم. از تحلیل متن برای شناسایی نشانه‌ها و روش‌های درمان بارز آنفلانزا که در توییترها ذکر شده‌اند استفاده کردیم. از پلتفرم آی‌پی‌ام اینفواسفیر بیگ اینسایت برای تحلیل دو مجموعه از داده‌های آنفلانزا استفاده شد: از داده‌های توییتر برای پایش شیوع آنفلانزا در ایالات متحده و از پایگاه داده‌ی تحلیلی سِریر هِلْت‌فَکْتز^۱ برای ردیابی بی‌درنگ مراجعات بالینی. حجم عظیمی از توییترها مرتبط با آنفلانزا با استفاده از توییتر استریمینگ آی‌پی‌آی استخراج شده و سپس وارد یک خوشه‌ی هندوپ شدند. به محض ورود موفق داده‌ها از ابزار زبان پرس و جوی چیسان^۲ برای دستکاری و بخش‌بندی داده‌های نیمه‌ساخت‌یافته‌ی نشانه‌گذاری شیء جاوااسکریپت^۳ استفاده شد. سپس، از هاپو برای جدولی‌سازی داده‌های متنی و جداسازی اطلاعات برای تحلیل موقعیت مکانی-زمانی و مصورسازی در نرم‌افزار آر استفاده شد. کل فرایند داده کاوی با استفاده از توابع مَپر دیوس انجام شد. ما از بسته‌ی نرم‌افزاری بیگ‌آر^۴ برای ارسال اسکریپت‌های آر بر روی داده‌های ذخیره شده در اچ‌دی‌اف‌اس استفاده کردیم. این بسته ما را قادر ساخت تا از قابلیت محاسبات موازی اچ‌دی‌اف‌اس استفاده کرده و عملیات مَپر دیوس را بر روی آنها انجام دهیم. کتابخانه‌های آی‌پی‌آی گوگل مَپس به عنوان یک ابزار نگاشت پایه برای مصورسازی مکان‌های ارسال توییترها بکار گرفته شد. یافته‌های ما نشان داد که یکپارچه‌سازی پرونده‌های پزشکی و رسانه‌های اجتماعی می‌تواند یک تکنیک مکمل ارزشمند برای سیستم‌های نظارتی موجود باشد. نتایج ما تأیید کرد که ازدحام مربوط به شیوع آنفلانزا در رسانه‌های اجتماعی رابطه‌ی نزدیکی با شیوع حقیقی آنفلانزا دارد. پژوهشگران دیگری از قبیل سنت لونیس و زورلو، ۲۰۱۲؛ برونیاتوفسکی، پُل و دروزه، ۲۰۱۳ نیز این نتیجه‌گیری را داشته‌اند. ما برای دستیابی به همبستگی مکانی-زمانی بین دو روند یک تحلیل سری‌های زمانی انجام دادیم و مشاهده کردیم که مراجعات بالینی برای آنفلانزا عقب‌تر از پست‌های آنلاین است که در مورد آنفلانزا منتشر می‌شود. علاوه بر این، تحلیل مکانی ما نشان داد که اکثریت توییترها از چند مکان عمومی منتشر و پخش می‌شود. این یافته‌ها به مقامات و دولت‌ها در توسعه‌ی مدل‌های پیش‌بینی دقیق‌تر و به موقع‌تر در طی شیوع بیماری‌ها و اطلاع‌رسانی به مردم در مورد مکان‌هایی که در طی آن دوره زمانی باید از آنها دوری کنند کمک می‌کند.

سوالاتی برای بحث و گفتگو

۴. چرا رسانه‌های اجتماعی باید به عنوان یک پیش‌بینی‌کننده زود هنگام شیوع آنفلانزا عمل کنند؟
۵. چه متغیرهای دیگری می‌توانند به پیش‌بینی اینگونه شیوع‌ها کمک کنند؟
۶. چرا این مسئله را به خوبی می‌توان با استفاده از فناوری‌های کلان داده که در این فصل توضیح داده شد حل کرد؟

Sources: A. H. Zadeh, H. M. Zolbanin, R. Sharda, & D. Delen. (2015). "Social Media for Nowcasting the Flu Activity: Spatial-Temporal and Text Analysis." *Business Analytics Congress, Pre-ICIS Conference*, Fort Worth, TX. D. A. Broniatowski, M. J. Paul, & M. Dredze. (2013). "National and Local Influenza Surveillance through Twitter: An Analysis of the 2012–2013 Influenza Epidemic." *PLoS One*, 8(12), e83672. P. A. Moran. (1950). "Notes on Continuous Stochastic Phenomena." *Biometrika*, 17–23.

پیش فناوری ۹.۳ یک پلتفرم نمایانگر فناوری کلان داده: تِرادیتا و نتج

مقدمه

این توضیحات از محتوای ارائه شده توسط تِرادیتا و به ویژه سری راگوان اقتباس شده است. تِرادیتا و نتج یک پلتفرم تحلیل پیشرفته است که موتورها و توابع تحلیلی پیشرفته در آن تعبیه شده‌اند که امکان پیاده‌سازی و اجرای آنها توسط زبان‌های علم داده‌ها (مثلاً اسکریپت، پایتون، آر) و ابزارهای علم داده‌ها (مثلاً تِرادیتا استودیو، تِرادیتا آپ‌سینتر، آر استودیو، ژوپیتِر نوتبوک) بر روی حجم انبوهی از انواع مختلف داده‌ها با پرسوناهای تحلیلی گوناگون (مثلاً دانشمندان علم داده، تحلیلگر کسب‌وکار) در چندین محیط گوناگون (بازارهای ابری عمومی، ابرهای اختصاصی) وجود دارد. پنج بخش مفهومی مهم برای درک و نتج وجود دارد: توابع و موتورهای تحلیل، دسترسی و انبارش داده‌ها، ابزارها و زبان‌های تحلیلی، استقرار و بکارگیری. شکل ۹/۱۰ معماری کلی و نتج و روابط آن با سایر ابزارها را به تصویر کشیده است.

¹ Cerner HealthFacts

² JSON Query Language (JAQL)

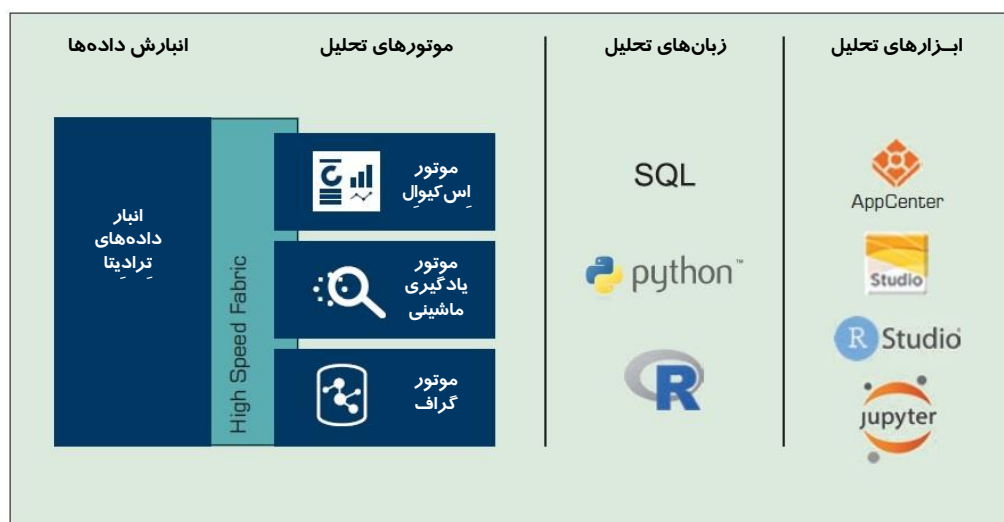
³ JavaScript Object Notation (JSON)

⁴ BigR

توابع و موتورهای تحلیل

یک موتور تحلیل یک چارچوب جامع است که شامل تمامی کامپوننت‌های نرم‌افزاری یکپارچه در قالب یک محفظه برای ارائه‌ی کارکردهای تحلیلی پیشرفته است که گروه‌های مشخص از کاربران قادر به پیاده‌سازی و اجرای آنها هستند. یک کامپوننت موتور تحلیل شامل موارد زیر است:

- توابع تحلیلی پیشرفته
 - نقاط دسترسی به انبار داده‌ها با انواع مختلف
 - یکپارچه‌سازی با ابزارهای مصورسازی و جریان کار تحلیلی
 - ابزارهای مدیریت و پایش تعبیه شده
 - محیط کارآمد و کاملاً مقیاس‌پذیر با آستانه‌های از پیش تعیین شده
- برخورداری از یک موتور تحلیل به دلیل ارائه‌ی یک محیط محاسباتی محفوظ که بتوان آن را از فضای انبارش داده تفکیک کرد بسیار سودمند است. علاوه بر این، موتورهای تحلیل را می‌توان برای دسترسی و استفاده‌ی کاربران خاص سفارشی‌سازی نمود (مثلاً تحلیلگر کسب‌وکار). در نسخه‌ی اول و نتایج سه موتور تحلیل وجود دارد که عبارتند از موتور نیواس‌کیوال، موتور یادگیری ماشین و موتور گراف.



شکل ۹.۱۰ معماری تِرادیتا و نتج

موتور نیواس‌کیوال شامل توابع و کارکردهای تحلیلی توکار است. تِرادیتا برای افزایش سرعت پردازش تحلیلی بالا همچنان به افزودن توابع و کارکردهای بیشتر ادامه خواهد داد. کارکردهای جدید در موتور نیواس‌کیوال عبارتند از:

- nPath
- Sessionization
- Attribution
- سری‌های زمانی
- تحلیل‌های چهار بُعدی
- توابع امتیازدهی (مثلاً بیژ ساده، جنگل‌های تصمیم)

موتور یادگیری ماشین بیش از ۱۲۰ کارکرد تحلیل از پیش ساخته شده را برای تحلیل مسیر، الگو، آماری و متن برای حل طیف وسیعی از مسائل کسب‌وکار ارائه می‌کند. کارکردها از درک احساسات تا تحلیل پیش‌گویانه‌ی خرابی قطعات را شامل می‌شوند.

موتور گراف مجموعه‌ای از کارکردها را که قادر به کشف ارتباطات بین افراد، محصولات و فرایندها در یک شبکه هستند ارائه می‌کند. تحلیل گراف قادر به حل مسائل پیچیده از قبیل ارتباطات شبکه‌های اجتماعی، ارتباطات اینفلوئنسرها، تشخیص کلاهبرداری و شناسایی تهدیدها است. و نتایج موتورهای تحلیل را نزدیک به داده‌ها قرار می‌دهد که دیگر نیاز به جابجایی داده‌ها نبوده و به کاربران اجازه می‌دهد تا تحلیل‌هایشان را بدون نمونه‌برداری بر روی مجموعه داده‌های بزرگتر انجام داده و مدل‌ها را با سرعت و تواتر بیشتری اجرا کنند. این امر با بکارگیری محفظه‌ها یا کانتینرهایی که توسط کوبرنیتز مدیریت می‌شوند میسر می‌شود که به کسب‌وکارها اجازه می‌دهد تا به راحتی موتورهای تحلیلی پیشرفته و جدید از قبیل اسپارک و تنسورفلو که هر دوی آنها در آینده‌ی نزدیک در دسترس خواهند بود را مدیریت کرده و جاری‌سازی کنند. یک مزیت دیگر محفظه‌ها توانایی تغییر مقیاس است. از دیدگاه یک کاربر، و نتایج یک چارچوب یکپارچه از داده‌ها و تحلیل‌ها ارائه می‌کند. در بخش زیرین، دارای یک لایه‌ی سازمانده‌ی بین موتورها است که داده‌های درست و درخواست‌های تحلیل را به موتورهای تحلیل درست انتقال می‌دهد. بدین ترتیب یک تحلیلگر کسب‌وکار یا دانشمند علم داده‌ها می‌تواند بدون اینکه نگران جابجایی از یک سرور یا برنامه کاربردی به برنامه یا سرور دیگر باشد توابع تحلیل را از موتورهای مختلف در یک برنامه کاربردی مقل ژوپیتر نوت‌بوک فراخوانی کند. نتیجه‌ی اینکار، یکپارچگی در پیاده‌سازی تحلیلی‌هایی است محدود به سیلوهای داده‌ای یا وظیفه‌ای نیستند.

انبارش و دسترسی داده‌ها: ترادیتا و نتایج همراه با پایگاه‌داده‌ی ترادیتا امپی‌پی که در داخل آن تعبیه شده عرضه می‌شود. همچنین، یک فابریک داده‌های پرسرعت (ترادیتا کوئری‌گرید^۱ و پرستو^۲) پلتفرم را به منبع داده‌های بیرونی متصل می‌کند که شامل پایگاه داده‌های تحلیلی سازمانی شخص ثالث (مثلاً اوراکل)، پلتفرم‌های داده‌های متن‌باز (مثلاً هادوپ)، پایگاه داده‌های نواس‌کیوال (مثلاً کاساندر) و غیره است. انواع پشتیبانی‌ها از داده‌ها از رابطه‌ای، مکانی و زمانی گرفته تا اکس‌امال، چی‌سان، آژور^۳ و فرمت‌های سری‌های زمانی را شامل می‌شود.

ابزارها و زبان‌های تحلیلی: ترادیتا و نتایج برای پاسخگویی به این نیاز ساخته شد که متخصصان و کارشناسان تحلیل مثل دانشمندان علم داده‌ها و تحلیلگران کسب‌وکارها به مجموعه متنوعی از ابزارها و زبان‌ها برای پردازش حجم انبوهی از داده‌ها برای دستیابی به بینش‌های تحلیلی نیاز داشتند. و نتایج شامل زبان‌هایی از قبیل اسکوال، آر و پایتون است که می‌توان توابع و کارکردهای تحلیلی را از طریق ترادیتا استودیو، آر استودیو و ژوپیتر نوت‌بوک اجرا کرد.

استقرار: پلتفرم و نتایج قابلیت‌های پردازشی یکسانی را از طریق گزینه‌های جاری‌سازی و استقرار از جمله ابر ترادیتا و ابر عمومی ارائه کرده و همچنین می‌توان آن را بر روی سخت‌افزارهای ترادیتا یا سخت‌افزارهای ضعیف‌تر نصب نمود. و نتایج همچنین به عنوان یک سرویس نیز موجود است.

کاربرد: ترادیتا و نتایج قرار است توسط طیف متنوعی از افراد تحلیلگر مورد استفاده قرار بگیرد. سهولت استفاده از اسکوال موجب شده تا دانشمندان علم داده‌ها و تحلیلگران کسب‌وکارها بتوانند کارکردها و توابع تحلیلی از پیش ساخته شده را که به صورت یکپارچه با موتورهای تحلیلی درآمده‌اند اجرا کنند. توانایی فراخوانی بسته‌های نرم‌افزاری تحت پشتیبانی ترادیتا از قبیل *dplyr* و *teradatamal* که دانشمندان علم داده‌ها که با زبان‌های آر و پایتون آشنا هستند این امکان را می‌دهد تا بتوانند بسته‌های نرم‌افزاری تحلیلی را از طریق آر استودیو و ژوپیتر نوت‌بوک بر روی پلتفرم اجرا کرد. کاربرانی که مهارت کافی در اجرای برنامه‌ها ندارند می‌توانند توابع تحلیلی کدگذاری شده در برنامه‌های کاربردی تعبیه شده در ترادیتا آپ‌سنتر^۴ که یک چارچوب برنامه‌سازی موجود در و نتایج است را برای مصورسازی‌ها از قبیل سان‌کی^۵، تری^۶، دیگرام‌های سیگما و ابرهای واژگان^۷ فراخوانی کنند.

یک مثال از کاربردها: یک خرده‌فروش بین‌المللی وب‌سایتی داشت که نتایج جستجو را برای خریداران بالقوه با سطحی نه چندان بهینه ارائه می‌کرد. در خصوص خریدهای آنلاین که حدود ۲۵٪ از کل فروش را به خود اختصاص داده بودند، نتایج جستجوی نادقیق تأثیر منفی بر تجربه‌ی مشتری و درآمد کل گذاشته بود. این شرکت از الگوریتم‌های یادگیری ماشینی ترادیتا که در ترادیتا

¹ Teradata QueryGrid

² Presto

³ Avro

⁴ Teradata AppCenter

⁵ Sankey

⁶ Tree

⁷ Word clouds

و نتایج موجود است برای تجمیع، تفکیک و دسته‌بندی عبارات و جملات جستجو استفاده کرد. این الگوریتم‌ها پاسخ‌های مورد نیاز برای شناسایی نتایجی که تطابق زیادی با نیازهای مشتریان آنلاین داشتند را ارائه می‌کردند. بدین ترتیب بیش از ۱/۳ میلیون دلار افزایش تدریجی درآمد از مشتریان بر اساس حجم خرید و در یک دوره زمانی تعطیلات دو ماهه حاصل شد.

مورد کاربردی ۹.۸ یک کاربرد دیگر از ترادیتا ونتج را نشان می‌دهد که از قابلیت‌های پیشرفته‌ی تحلیل شبکه برای تحلیل داده‌های بدست آمده از پایگاه داده‌های رکوردهای پزشکی الکترونیکی بزرگ استفاده می‌کند.

تحلیل الگوهای بیماری‌ها از یک پایگاه داده‌ی سوابق پزشکی الکترونیکی

مورد کاربردی ۹.۸

یک پایگاه داده‌های تحلیلی انبوه توسط سرنر کورپوریشن^۱ که یک عرضه‌کننده‌ی مهم رکوردهای پزشکی الکترونیکی است برای کمک به توسعه‌ی برنامه‌های تحلیلی در اختیار مرکز نوآوری سیستم‌های سلامت در دانشگاه ایالتی اکلاهما قرار گرفته است. این انبار داده‌های تحلیلی شامل سوابق پزشکی بیش از ۵۰ میلیون بیمار در بیمارستان‌های ایالات متحده است (۲۰۱۵-۲۰۰۰). این بزرگ‌ترین و تنها پایگاه داده‌ی رابطه‌ای در این صنعت است که سوابق جامع و کاملی از داده‌های مربوط به دارو، آزمایشات، رویدادهای بالینی، پذیرش‌ها و صورتحساب‌ها را شامل می‌شود. این پایگاه داده همچنین شامل نتایج بیش از ۲/۴ میلیارد آزمایش و بیش از ۲۹۵ سفارش نزدیک به ۴،۵۰۰ دارو بر اساس نام و برند است. این پایگاه یکی از بزرگ‌ترین مجموعه داده‌های بی‌نام واقعی و منطبق با HIPAA از این نوع می‌باشد. سوابق الکترونیکی پزشکی را می‌توان برای توسعه‌ی چندین برنامه‌ی کاربردی تحلیلی بکار برد. یکی از این کاربردها، درک روابط بین بیماری‌ها بر اساس اطلاعات مربوط به بیماری‌هایی است که به طور همزمان بین بیماران وجود داشته است. زمانیکه بیمار به چندین بیماری مبتلا باشد این شرایط را چندابتلایی می‌نامند. چندابتلایی می‌تواند بین گروه‌های جمعیتی مختلف متفاوت باشد. در یک کاربرد (کالگو ترا، شاردا و کراف ۲۰۱۷) به مطالعه‌ی ناهمگونی‌های سلامتی بر حسب چندابتلایی‌ها به تفکیک جنسیت پرداخته‌اند.

برای مقایسه‌ی چندابتلایی‌ها، از یک رویکرد تحلیل شبکه استفاده شده است. یک شبکه از مجموعه‌ی معینی از آیتم‌ها به نام گره‌ها تشکیل شده که از طریق لبه‌ها به یکدیگر متصل شده‌اند. یک لبه بیانگر یک رابطه‌ی معین و تعریف شده بین گره‌ها است. یک مثال بسیار بارز از یک شبکه، شبکه دوستی است که در آن افراد بر اساس رابطه‌ی دوستی با یکدیگر مرتبط هستند. از شبکه‌های بارز دیگر می‌توان به شبکه‌های رایانه‌ای، شبکه صفحات وب، شبکه راه‌ها و جاده‌ها و شبکه‌های فرودگامی اشاره کرد. برای مقایسه‌ی چندابتلایی‌ها، شبکه‌هایی از تشخیص بیماری‌های ایجاد شده توسط مردان و زنان ایجاد می‌شود. از اطلاعات مربوط به بیماری‌هایی که در هر بیمار در تاریخچه طول زندگی وی وجود داشته برای ایجاد یک شبکه‌ی چندابتلایی استفاده می‌شود. برای تحلیل از اطلاعات مربوط به ۱۲ میلیون بیمار زن و ۹/۹ بیمار مرد استفاده می‌شود. برای مدیریت چنین مجموعه داده‌های عظیمی، از پلتفرم کلان داده‌ی ترادیتا آستر^۲ استفاده شد. برای استخراج و آماده‌سازی داده‌های شبکه، از چارچوب‌های اسکوال، اسکوال-ام آر^۳ و اسکوال-جی آر^۴ که توسط آستر پشتیبانی می‌شدند استفاده شد. برای مصورسازی شبکه‌ها از آستر آپ سنتر^۵ و جفی^۶ استفاده شد. شکل ۹/۱۱ شبکه‌های چندابتلایی زنان و مردان را نشان می‌دهد. در این شبکه‌ها، گره‌ها بیانگر بیماری‌های مختلفی هستند که بر اساس دسته‌بندی بین‌المللی بیماری‌ها^۷، بازننگری نهم، اصلاح بالینی دسته‌بندی شده و در یک سطح سه رقمی تجمیع گردیده‌اند.

¹ Cerner Corporation

² Teradata Aster Big Data Platform

³ SQL-MR

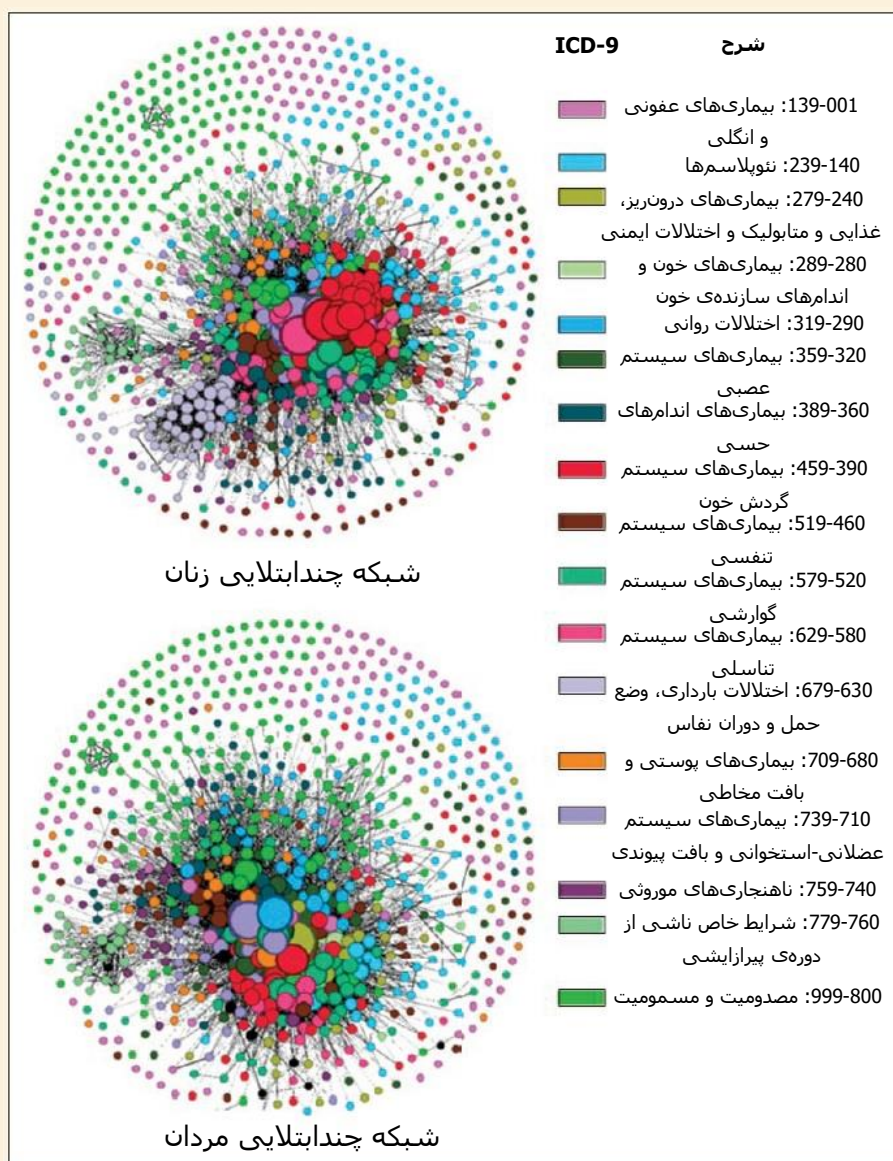
⁴ SQL-GR

⁵ Aster AppCenter

⁶ Gephi

⁷ International Classification of Diseases

دو بیماری بر اساس تشابه محاسبه شده با استفاده شاخص کسینوسی سالتون به یکدیگر مرتبط شده‌اند. هرچه اندازه‌ی یک گروه بزرگتر باشد، چندابتلایی آن بیماری نیز بیشتر خواهد بود. شبکه‌ی چندابتلایی زنان چگالی بیشتری نسبت به مردان دارد. تعداد گره‌ها و لبه‌ها در شبکه زنان به ترتیب ۸۹۹ و ۱۴،۸۱۰ است در حالیکه این ارقام برای شبکه مردان ۸۳۹ و ۱۲،۴۹۸ می‌باشد. مصورسازی‌ها بیانگر تفاوت بین الگوی شکل‌گیری بیماری‌ها بین بیماران مرد و زن است. به طور خاص، چندابتلایی اختلالات روانی در زنان بیشتر از مردان است. از طرف دیگر، شدت برخی بیماری‌های ارتباطی بین متابولیسم لیپید و اختلالات مزمن قلبی در مردان بیشتر از زنان است. اینگونه تفاوت‌ها، سوالاتی را برای پژوهش‌های زیست‌شناسی، رفتاری، بالینی و خطمشی مطرح می‌کند. سیستم‌های پایگاه داده‌ی سنتی در پردازش مؤثر چنین مجموعه داده‌ی عظیمی با مشکل مواجه می‌شوند. با استفاده از ترادیتا آستر تحلیل داده‌های حاوی اطلاعات میلیون‌ها پرونده به آسانی و سرعت نسبتاً بالایی انجام شد. تحلیل شبکه‌ها اغلب به عنوان روشی برای تحلیل مجموعه داده‌های کلان توصیه می‌شود که به درک داده‌ها در یک تصویر کمک می‌کند. در این کاربرد، شبکه‌ی چندابتلایی به توضیح روابط بین بیماری‌ها در یکجا می‌پردازد.



شکل ۹.۱۱ شبکه‌های چندابتلایی مردان و زنان

سوالاتی برای بحث و گفتگو

۳. علل استفاده از تفاوت‌های سلامت بین مردان و زنان چه می‌تواند باشد؟
۴. مؤلفه‌های اصلی یک شبکه چه هستند؟
۵. در این کاربرد از چه نوع تحلیل‌هایی استفاده شد؟

Source: Kalgotra, P., Sharda, R., & Croff, J. M. (2017). Examining health disparities by gender: A multimorbidity network analysis of electronic medical record. *International Journal of Medical Informatics*, 108, 22–28.

همانطور که قبلاً نیز اشاره شد، هدف ما در این بخش تأکید بر برخی از فعالان در فضای فناوری کلان‌داده‌ها می‌باشد. علاوه بر عرضه‌کنندگانی که در بالا ذکر شدند، در کنار صدها عرضه‌کننده که در این دسته‌ها وجود دارند و قبلاً به آنها اشاره شد، کاربردهای خاص هر صنعت نیز وجود دارد. به جای این نام آنها را در اینجا بیاوریم از شما دعوت می‌کنیم تا به آخرین نسخه‌ی اکوسیستم تحلیل‌های کلان‌داده‌ها در آدرس <http://matttruck.com/bigdata2018/> مراجعه کنید (دسترسی در اکتبر ۲۰۱۸). دیاگرام اکوسیستم به‌روزآوری شده‌ی مت تراک به شناسایی شرکت‌ها در هر یک از خوشه‌ها پرداخته است.

سوالاتی برای مرور بخش ۹.۸

۱۹. برخی از عرضه‌کنندگان مهم فناوری کلان‌داده‌ها که تمرکز اصلی آنها بر پلتفرم‌های هَدوپ درون‌سازمانی است را نام ببرید.
۲۰. چشم‌انداز عرضه‌کنندگان کلان‌داده‌ها چه ویژگی خاصی دارد؟ فعالان مهم آن کدامها هستند؟
۲۱. تفاوت‌ها و تشابهات مهم بین عرضه‌کنندگان ابری محصولات تحلیلی و حضور عرضه‌کنندگان تحلیلی در پلتفرم‌های ابری خاص را جستجو و شناسایی کنید.
۲۲. برخی از ویژگی‌های پلتفرم‌هایی از قبیل ترادیتا ونتج چه هستند؟

یک فناوری نوظهور دیگر که کاربران تحلیل‌های کسب‌وکار باید از آن مطلع باشند، رایانش ابری است. مؤسسه‌ی ملی استاندارد و فناوری^۱ رایانش ابری^۲ را اینگونه تعریف می‌کند «یک مدل برای دسترسی راحت و بر حسب تقاضای شبکه به مجموعه‌ای از منابع رایانشی قابل پیکربندی (مثلاً شبکه‌ها، سرورها، انبارش و خدمات) که بتوان سریعاً آنها را با کمترین تلاش مدیریتی یا تعامل با سرویس‌دهنده فراهم کرده و منتشر نمود.» ویکیپدیا رایانش ابری را اینگونه تعریف می‌کند: «یک سبک رایانشی که در آن منابع مقیاس‌پذیر و اغلب مجازی‌سازی شده بر روی اینترنت عرضه می‌شوند. کاربران نیازی به دانش یا تجربه یا کنترل نسبت به زیرساخت‌های فناوری در ابری که از آنها پشتیبانی می‌کند ندارند.» این تعریف بسیار جامع و گسترده است. به طرق دیگر، رایانش ابری نام جدیدی برای بسیاری از روندهای مرتبط قبلی است: رایانش همگانی، رایانش شبکه‌ی ارائه‌دهنده‌ی اپلیکیشن سرورها، رایانش بر حسب تقاضا، نرم‌افزار به عنوان یک سرویس^۳ و حتی موارد قدیمی‌تر مثل رایانش متمرکز با ترمینال‌های

¹ The National Institute of Standards and Technology (NIST)

² Cloud Computing

³ Software as a Service (SaaS)

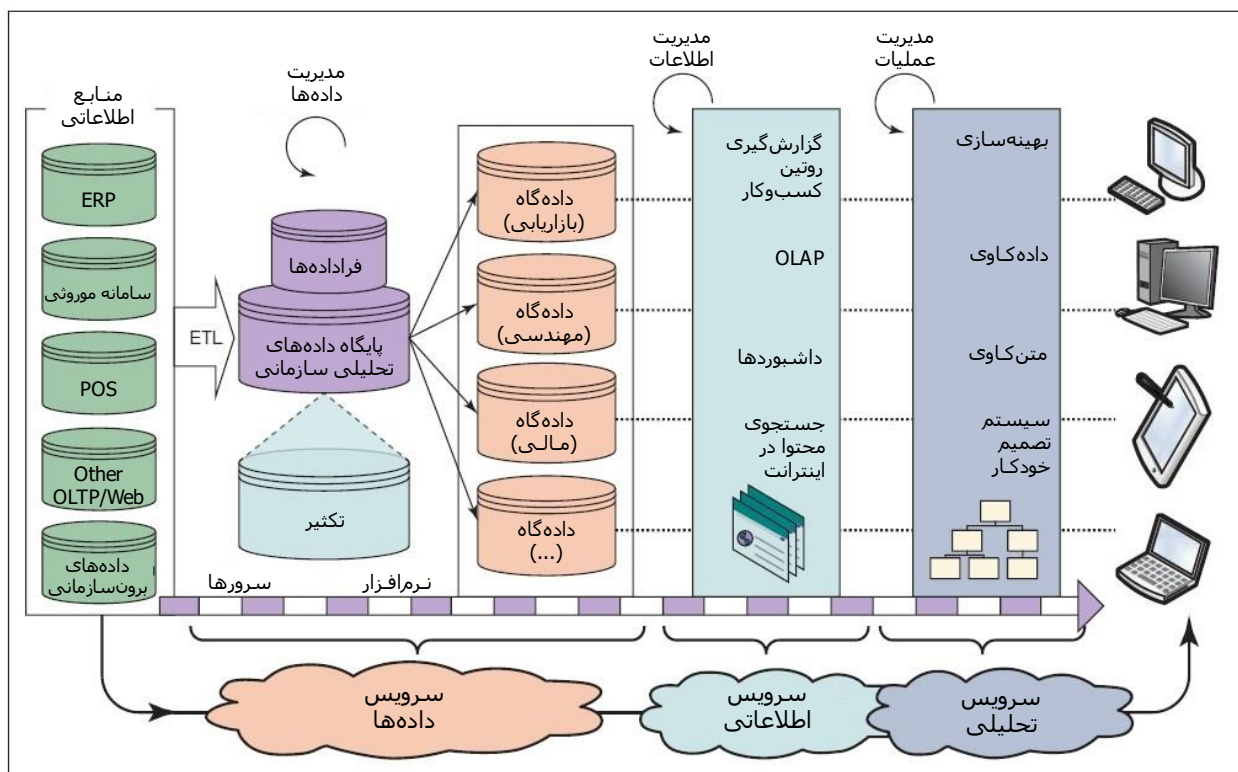
رایانه‌ای است. ولی عبارت *رایانش/ابری* از یک ارجاع به اینترنت به عنوان یک «ابر» نشأت گرفته و بیانگر تکامل تمامی روندهای رایانشی متمرکز/اشتراکی قبلی است. همچنین بر اساس تعریف ویکی‌پدیا، رایانش ابری ترکیبی از چندین کامپوننت فناوری اطلاعات در قالب سرویس‌ها است. به عنوان مثال، زیرساخت به عنوان یک سرویس¹ به ارائه‌ی رایانش پلتفرم‌ها به عنوان یک سرویس² و همچنین تمامی تدارکات پلتفرم‌های پایه از قبیل مدیریت، امنیت و غیره اشاره دارد. این شامل نرم‌افزار به عنوان یک سرویس نیز می‌باشد که شامل اپلیکیشن‌هایی است که می‌بایست از طریق یک مرورگر وب ارائه شوند در حالیکه داده‌ها و برنامه‌های کاربردی بر روی سرور دیگری قرار دارند.

اگرچه ما نامه‌های الکترونیکی مبتنی بر وب را یک مثال از رایانش ابری نمی‌دانیم ولی می‌توان آن را یک اپلیکیشن ابری پایه در نظر گرفت. معمولاً اپلیکیشن نامه الکترونیکی، داده‌ها (پیام‌های نامه الکترونیکی) و نرم‌افزار (برنامه‌های نامه الکترونیکی که امکان پردازش و مدیریت نامه‌های الکترونیکی را به ما می‌دهند) را ذخیره می‌کند. ارائه‌دهنده‌ی نامه الکترونیکی همچنین سخت‌افزار/نرم‌افزار و تمامی زیرساخت‌های پایه‌ی دیگر را تأمین می‌کند. تا زمانیکه اینترنت در دسترس است، از هر کجا از فضای ابر می‌توان به نامه‌های الکترونیکی دسترسی یافت

¹ Infrastructure as a Service (IaaS)

² Platforms as a Service (PaaS)

³ Pay-as-you-go

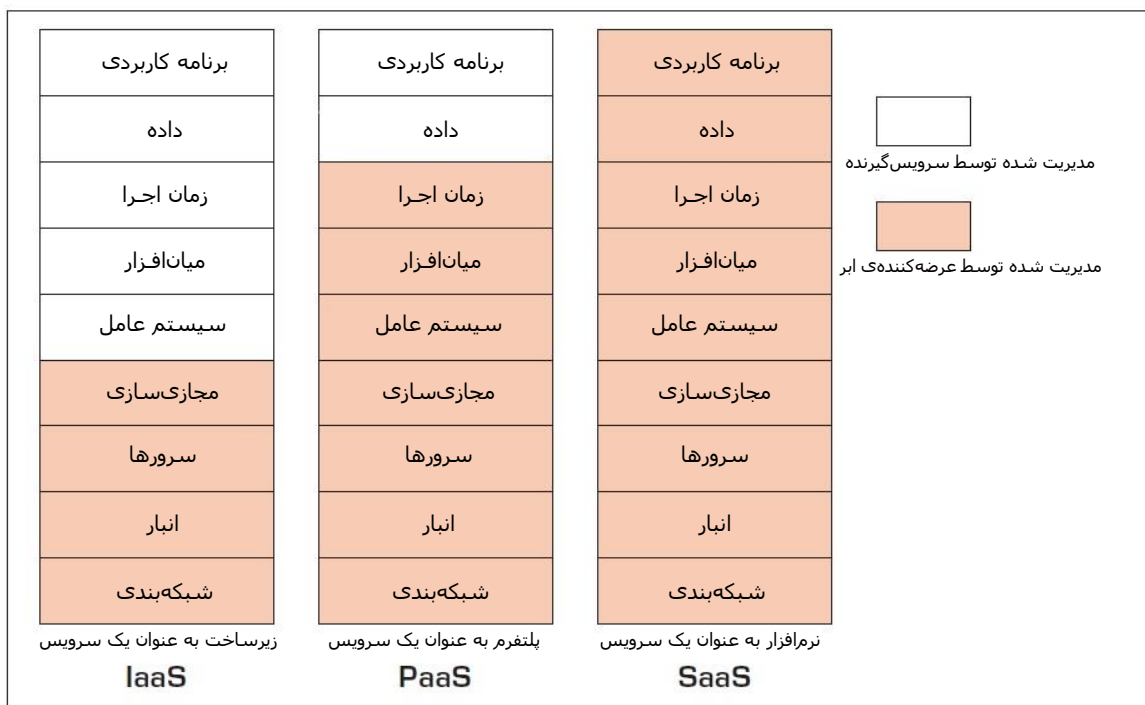


شکل ۹.۱۱ معماری مفهومی یک سیستم پشتیبان مبتنی بر ابر

Source: Based on Demirkan, H., & Delen, D. (2013, April). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and Big Data in cloud. *Decision Support Systems*, 55(1), 412-421.

¹ Google App Engine

² Amazon AWS



شکل ۹.۱۳ انباشت فناوری به عنوان یک خدمت برای انواع مختلف محصولات قابل عرضه ابری

ارتباطات تاریخی بین شرکت های خدمات عمومی و نیروهای امدادی توسط تماس های تلفنی یا رادیوهای دو طرفه انجام می شد. برخی از این ارتباطات با نیروهای امدادی در صحنه و برخی دیگر با واحدهای اعزامی یا دیگر واحدهای سازمان های امدادی انجام می شوند. زمانیکه فردی از جامعه شاهد یک حادثه در محل است معمولاً فقط با شماره ۹۱۱ که به نیروهای امدادی وصل می شود تماس می گیرند. مراکز اعزام نیرو، نزدیکترین نیروی امداد را به محل حادثه اعزام می کنند که آنها نیز با از طریق رادیو یا تلفن های همراه با مرکز تماس برقرار کرده و وضعیت محل حادثه را توضیح می دهند. سپس مراکز تماس با شرکت های خدمات عمومی مربوطه برای ارسال تیم هایشان به محل حادثه تماس برقرار می کنند. اینکار نیازمند آگاهی و ارسال محل دقیق حادثه برای مرکز اعزام و از مرکز اعزام به شرکت های خدمات عمومی است که معمولاً در شرایطی که آدرس پستی محل حادثه مشخص نباشد کار بسیار چالش برانگیزی خواهد بود (مثلاً در کنار بزرگراه، در دشت یا فضای آزاد و ...). شرکت خدمات عمومی نیز مرکز تماس را از وضعیت واقعی نیروهای خود مطلع می کند. این اطلاعات باید برای نیروهای امدادی حاضر در صحنه ارسال شود. اکثر این فرایندها متکی به این است که اطلاعات به صورت شفاهی ارسال شده و سپس به یک یا چند گیرنده انتقال یابد و این اطلاعات باید در طول زنجیره ای تماس رفت و برگشت داشته باشد. کل این فرایند می تواند منجر به آشفتگی ارتباطات و پیام های ناقص شود که موجب از دست رفتن ساعت ها و حتی دقایق ارزشمند گردد.

یک شرکت مهم خدمات عمومی در کرانه غربی ایالات متحده که پیشگام در بکارگیری فناوری برای رسیدگی به مسائل و مشکلات متداول است به این قطعیت رسید که بسیاری از این چالش ها را می توان از طریق به اشتراک گذاری بهتر و به موقع اطلاعات با استفاده از فناوری موبایل-ابر رفع نمود. حوزه استحقاقی آنها شامل شهرهای بسیار متراکم و پرجمعیت تا مناطق روستایی دور افتاده و کیلومترها مناطق صحرایی، پارک های طبیعی و غیره بود.

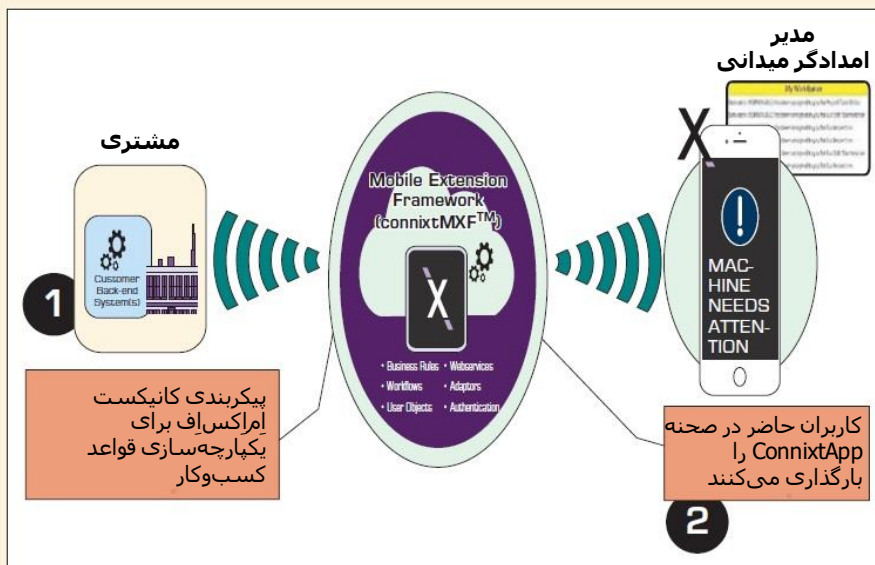
با توجه به اینکه اکثر نیروهای امدادی دارای گوشی هوشمند یا تبلت هستند، این شرکت خدمات عمومی مجموعه موبایل کانیکست آی مارک^۱ را به عنوان آپلیکیشن موبایل خود انتخاب نمود که به نیروهای امدادی اجازه می داد تا در خصوص هرگونه حادثه در سر صحنه به این شرکت مشاوره دهند. این فناوری همچنین نیروهای امدادی را در جریان آخرین وضعیت پاسخگویی شرکت خدماتی به حادثه قرار می داد. با گسترش بیش از ۲۰,۰۰۰ نیروی امدادی در سرتاسر این منطقه، کاهش موانع پذیرش این فناوری یک عامل بسیار مهم محسوب می شد. به گفته جی. ساتیش، یکی از بنیان گذاران و مدیرعامل شرکت کانیکست، «بهبود ارتباطات با گروه هایی که خارج از سازمان هستند یک مشکل تاریخی است و برای این استقرار، تمرکز بر ساده سازی کلی اصلی موفقیت است.»

از نیروهای امدادی دعوت به عمل آمد تا آپلیکیشن را بازگذاری و سلف رجیستر کنند و به محض ارائه ای حقوق دسترسی توسط شرکت خدمات عمومی، آنها می توانند حوادث را با استفاده از تبلت ها یا گوشی های هوشمند خود گزارش دهند. نیروهای امدادی به سادگی از یک منوی پایین افتادنی برای انتخاب از یک لیست از پیش تعریف شده از حوادث استفاده کرده و در صورتیکه در سر صحنه منتظر بمانند می توانند از طریق یک گزینه این را نمایش دهند و تصاویر را به همراه توضیحات پیوست گزارش کنند و تمامی اینها تنها با چند کلیک بر روی دستگاه انجام می شود. شرکت خدمات عمومی، اعلان ها و نوتیفیکیشن های مربوط با حادثه را دریافت کرده، اطلاعات زمانی و جغرافیایی را بررسی نمود و واکنش خود را به روزآوری می کند. این واکنش (که ممکن است یک کامیون باریبری باشد) برای نیروهای امدادی ارسال شده و در آپلیکیشن نیز نگهداری می شود. سادگی این آپلیکیشن موجب شده تا نیروهای امدادی به راحتی بتوانند از آن استفاده کنند. آنها از تلفن یا تبلت خودشان استفاده کرده و به همان نحوی که عادت دارند تماس برقرار می کنند و اطلاعات مورد نیاز را به آسانی و به نحوی اثربخش ارائه می کنند. آنها قادر به مشاهده به روزآوری های شرکت خدمات عمومی هستند (مثل وضعیت جاری کامیون ارسال). پیام های

¹ Connixt iMarq

تلفنی آشفته یا از دست رفته به حداقل می‌رسد. گزینه‌هایی از قبیل یادداشت‌های پیام‌های صوتی ضبط شده، استفاده از قابلیت گفتار به متن و غیره نیز وجود دارد.

فناوری ابر به ویژه در این مورد مفید بوده است – استقرار بدون وجود مسائل مربوط به تأمین سخت‌افزار، نصب و پشتیبانی‌های مناسب سریع‌تر انجام می‌شود. چارچوب اِکس‌تِنشن موبایل (اِم‌اِکس‌اِف)^۱ مبتنی بر ابر کانیکست برای پی‌کر بندی و استقرار سریع طرح‌ریزی شده است – پی‌کر بندی در فضای ابر انجام شده و پس از آن، اپلیکیشن‌ها برای بارگذاری و استقرار آماده هستند. مهمتر اینکه، به کمک اِم‌اِکس‌اِف می‌توان فرم‌ها و فرایندها را به آسانی تغییر داد – به عنوان مثال اگر شرکت خدمات عمومی بخواهد گزینه‌های بیشتری را به فهرست پایین‌افتادنی خود اضافه کند اینکار به سادگی در اِم‌اِکس‌اِف امکان‌پذیر است. در عرض چند دقیقه این گزینه برای تمامی کاربران در دسترس خواهد بود. شکل ۹/۱۴ این معماری را به تصویر کشیده است. سیستمی که از فناوری‌های سیار و ابری هم‌جا حاضر بهره می‌برد دارای مزایای بیشتری نیز می‌باشد. از آنجائیکه تمامی این پی‌کر بندی‌ها و منطق کسب‌وکار در فضای ابر ذخیره می‌شوند، خود نرم‌افزار می‌تواند به عنوان یک سیستم مستقل برای مشتریانی که از سیستم‌های بک‌اند برخوردار نیستند عمل کند که این موضوع در کسب‌وکارهای کوچک و متوسط بسیار اهمیت دارد. و مشتریان که از سیستم‌های بک‌اند برخوردارند، ارتباطات پیوسته و مداوم از طریق خدمات وب و سیستم بک‌اند به عنوان یک سیستم ضبط و ثبت عمل می‌کند. این به کسب‌وکارها کمک می‌کند تا این فناوری را به شکل فزاینده‌تری بکار بگیرند که با یک سیستم مستقل با حداقل تأثیر درون سازمانی در حوزه فناوری اطلاعات آغاز شده و با خودکارسازی عملیات به سمت یکپارچه‌سازی سیستم بک‌اند حرکت می‌کند.



شکل ۹.۱۴ روابط بین کارگران و فناوری در یک برنامه کاربردی تحلیلی مبتنی بر ابر

از طرف دیگر، اپلیکیشن‌های موبایل خودشان نیز سیستم‌ناپاور هستند و از طریق خدمات وب استاندارد برای برقراری ارتباطات استفاده می‌کنند و دستگاه نهایی کاربر می‌تواند اندروئید، آی‌اواس و گوشی هوشمند یا تبلت باشد. بنابراین، صرف‌نظر از اینکه از چه دستگاهی استفاده می‌شود تمامی ارتباطات، منطق کسب‌وکار و الگوریتم‌ها بین تمامی دستگاه‌ها/پلتفرم‌ها به صورت استاندارد درآمده است. آی‌مارک به عنوان یک اپلیکیشن بومی بر روی تمامی دستگاه‌ها از فناوری استاندارد که توسط سازندگان تولیدکنندگان دستگاه‌ها و عرضه‌کنندگان سیستم‌های عامل ارائه شده‌اند بهره می‌برد. به عنوان مثال، استفاده از برنامه‌های کاربردی نقشه‌های بومی به اپلیکیشن‌ها اجازه می‌دهد تا از بهبودهای ایجاد شده توسط عرضه‌کنندگان پلتفرم‌ها بهره ببرند و بدین ترتیب با افزایش دقت نقشه‌ها، کاربران نهایی اپلیکیشن‌های موبایل نیز از این پیشرفت‌ها بهره‌مند می‌شوند.

¹ Mobile Extension Framework (MXF)

نهایتاً، برای جاری‌سازی موفق، فناوری ابر-موبایل سازمانی باید کاملاً کاربرمحور باشد. ظاهر و حس آن باید متناسب با راحتی کاربر باشد که عمدتاً چیزی است که کاربران خبره از هر اپلیکیشن موبایل که استفاده می‌کنند انتظار دارند. رفتار با کاربران تجاری همانند یک کاربر اپلیکیشن که انتظارات استاندارد آنها را برآورده می‌کند در زمان و انرژی آنها صرفه‌جویی می‌کند. این روش برای حصول اطمینان از بکارگیری موفق اپلیکیشن‌ها ضروری است.

شرکت خدمات عمومی اکنون اطلاعات بهتری از نیروهای امداد در اختیار دارد زیرا اطلاعات مستقیماً از محل برایشان ارسال می‌شود (نه از طریق اعزامگر یا هر شخص ثالث دیگری)، تصاویر در اختیارشان قرار می‌گیرد و نشانه‌گذاری زمانی و مکانی وجود دارد. از ارسال پیام‌های تلفنی آشفته اجتناب می‌شود. ارتباطات دوطرفه بین شرکت خدمات عمومی و نیروی امداد در محل حادثه نیز بهبود می‌یابد. سوابق تاریخی حوادث نگهداری می‌شوند. شرکت خدمات عمومی و نیروهای امداد اکنون در پاسخگویی سریع و کامل به حوادث و بهبود خدمات به عموم مردم یکپارچه‌تر عمل می‌کنند. با برقراری پیوندهای محکم‌تر با نیروهای امداد (پرسنل دپارتمان پلیس و آتش‌نشانی)، پاسخ‌های بهتر و هماهنگ‌تری در خصوص حوادثی که توسط نیروهای امداد کشف شده است به مردم داده می‌شود.

سوالاتی برای بحث و گفتگو

۱. فناوری ابری چگونه می‌تواند بر نرم‌افزارهای سازمانی کسب‌وکار کوچک و متوسط تأثیر بگذارد؟
۲. برخی از حوزه‌هایی که می‌توانند از فناوری موبایل استفاده کدامها هستند؟
۳. چه نوع کسب‌وکارهایی پیشگامان بکارگیری فناوری ابر-موبایل هستند؟
۴. مزایای نرم‌افزار سازمانی مبتنی بر ابر به جای مدل‌های درون سازمانی متداول چیست؟
۵. ریسک‌های احتمالی بکارگیری اپلیکیشن‌های ابر در مقایسه با اپلیکیشن‌های درون سازمانی رایج چیست؟

از این مورد چه می‌توانیم بیاموزیم؟

شرکت‌ها با تحلیل داده‌های کسب شده در اینجا و اکنون می‌توانند سریع‌تر در مورد مشتریان‌شان پیش‌بینی کرده و تصمیم بگیرند. بدین ترتیب اطمینان حاصل می‌شود که شرکت‌ها مشتریان درست را هدف قرار داده، جذب و نگهداری کرده و بیشترین ارزش را از آنها کسب می‌کنند. داده‌هایی که هفته‌ی گذشته بدست آمده‌اند به اندازه‌ی داده‌هایی که امروز به دست می‌آیند مفید و سودمند نیستند. استفاده از داده‌های مرتبط موجب می‌شود تا تحلیل پیشگویانه‌ی دقیق‌تر و کارآمدتری داشته باشیم.

Source: Used with permission from G Satish, Connixit, Inc.

ابر آی بی ام آی بی ام تمامی محصولات تحلیلی خود را از طریق فناوری ابر خود عرضه می کند. ابر آی بی ام چنین گروه از ابزارهای تحلیلی و هوش مصنوعی را عرضه می کند. به عنوان مثال، آی بی ام واتسون آنالیتیکز اکثر قابلیت ها و ویژگی های تحلیلی که می تواند از طریق فناوری ابر ایجاد کرده و استقرار دهد را به صورت یکپارچه عرضه می کند. آی بی ام واتسون کوانتیتو¹ یکی از محصولات مهم مبتنی بر ابر این شرکت بوده که متن کاوی و یادگیری عمیق را در سطح بسیار بالایی بکار گرفته است که پیشتر در بخش متن کاوی به آن اشاره شد.

ماین مای تکست. کام² یکی از حوزه های مهم در تحلیل ها که رشد داشته، متن کاوی می باشد. متن کاوی به شناسایی عناوین و موضوعات سطح بالا در اسناد، استنتاج احساسات از اظهار نظرات و مصورسازی سند یا ارتباطات بین واژه/مفهوم می پردازد. یک شرکت نوپا به نام ماین مای تکست. کام این قابلیت ها را از طریق وب سایت خود در فضای ابر عرضه می کند.

¹ IBM Watson Cognitvie

² MineMyText.com

ساس ویا^۱ یکی از نرم‌افزارهای تحلیلی خود را بر اساس تقاضا از طریق فضای ابر عرضه می‌کند. در حال حاضر، ساس ویژوال استاتستیکز^۲ فقط به عنوان یک سرویس ابری در دسترس بوده و رقبل تابلو محسوب می‌شود.

تابلو^۳ تابلو یک نرم‌افزار مصورسازی مهم است که در مقوله‌ی تحلیل‌های توصیفی عرضه شد و از طریق فضای ابر نیز در دسترس است.

اسنو فلیک^۴ اسنو فلیک یک پایگاه داده‌ی تحلیلی مبتنی بر ابر است. کاربران می‌توانند داده‌هایشان را از منابع متعدد در یکجا گردهم آورده و با استفاده از اسنو فلیک آنها را تحلیل کنند.

در این بخش ما به معرفی چند برنامه کاربردی تحلیلی مهم در فضای ابر می‌پردازیم.

مردم به طور فزاینده‌ای از اپلیکیشن‌های موبایل برای ردیابی کارها و فعالیت‌های روزانه و همچنین حفظ و نگهداری تاریخچه‌ی سلامت خود استفاده می‌کنند. زایان چاینا که یک عرضه‌کننده‌ی خدمات مراقبت سلامت سیار است یک ابزار نوآورانه برای پایش سلامت ارائه کرده است که داده‌ای مربوط به مشکلات سلامتی از قبیل سطح گلوکوز، فشار خون، رژیم، داروها را گردآوری کرده و به آنها در بهبود کیفیت زندگی‌شان را با ارائه‌ی توصیه‌هایی را به صورت روزانه برای مدیریت سلامت‌شان و پیشگیری یا درمان بیماری‌ها ارائه می‌کند.

حجم انبوهی از داده‌های بلادرنگ که مشکلات مقیاس‌پذیری و مدیریت داده‌ها را به وجود می‌آورد موجب شد تا این شرکت با همکاری مایکروسافت از تحلیل جریان، یادگیری ماشین و راه‌حل‌های پی‌اس و پاور بی‌آی بهره‌برد که موجب افزایش امنیت داده‌ها و بهبود تحلیل‌ها گردید. زایان چاینا کاملاً به هوش تجاری سنتی وابسته بود که داده‌ها را از دستگاه‌های مختلف یا ابر گردآوری می‌کرد. زایان با استفاده از یک معماری تحلیلی مبتنی بر ابر توانست سرعت، امنیت و ویژگی‌های متعددی را بیفزاید. آنها یک هاب اینترنت اشیا را برای انتقال بهتر داده‌ها از دستگاه به فضای ابر به فرانت‌اند افزودند. داده‌ها ابتدا از طریق بلوتوث از دستگاه به یک اپلیکیشن موبایل ارسال شده و سپس از طریق اچ‌تی‌تی‌پی‌اس و ای‌ام‌کیوپی به یک هاب اینترنت اشیا منتقل می‌شوند. تحلیل جریان به پردازش بلادرنگ اطلاعات گردآوری شده در هاب اینترنت اشیا کمک کرده و اطلاعات و بینش‌های مفیدی را تولید می‌کند که سپس به یک پایگاه داده‌ی اس‌کیوال منتقل می‌شوند. آنها از یادگیری ماشینی آזור برای تولید مدل‌های پیشگویانه بر روی داده‌های بیماران دیابتی و بهبود تحلیل و پیش‌بینی استفاده می‌کنند. پاور بی‌آی مصورسازی ساده و آسان بینش‌های بدست آمده از تحلیل‌ها را برای کاربران فراهم می‌کند.

¹ SAS Viya

² SAS Visual Statistics

³ Tableau

⁴ Snowflake

⁵ IOT solution

Sources: "Zion China Uses Azure IoT, Stream Analytics, and Machine Learning to Evolve Its Intelligent Diabetes Management Solution" at www.codeproject.com/Articles/1194824/Zion-China-uses-Azure-IoTStream-Analytics-and-M (accessed October 2018) and <https://microsoft.github.io/techcasestudies/iot/2016/12/02/IoT-ZionChina.html> (accessed October 2018).

گلف ایر^۱ یک شرکت حمل و نقل ملی متعلق به بحرین است که با بیش از ۳,۰۰۰ کارمند به ۴۵ شهر در ۲۴ کشور سه قاره خدمت‌رسانی می‌کند. گلف ایر یک پیشگام صنعت در ارائه‌ی خدمات میزبانی عربی به مشتریان است. برای اطلاعات بیشتر در مورد نظرات مشتریان در خصوص این خدمات، این شرکت هواپیمایی به دنبال این بود که بداند مشتریان در مورد خدمات میزبانی این شرکت چه نظراتی در رسانه‌های اجتماعی بیان می‌کنند. چالش آنها تجزیه و تحلیل تمامی نظرات و پست‌های مشتریان بود که شامل صدها هزار پست روزانه بود. پایش دستی این پست‌ها کار بسیار زمان‌بر و طاقت‌فرسای بود و مستعد خطای انسانی بود.

گلف ایر می‌خواست این فعالیت را به صورت خودکار درآورده و برای آگاهی از روندهای بازارهای نوظهور این داده‌ها را تحلیل نماید. در کنار آن، شرکت می‌خواست یک زیرساخت محکم و مقاوم برای میزبانی چنین نرم‌افزار پایش رسانه‌های اجتماعی داشته باشد که در خارج از مرزهای جغرافیایی نیز در دسترس باشد.

گلف ایر یک نرم‌افزار تحلیل احساسات به نام «تحلیل احساسات عربی» را توسعه داد که قادر به تحلیل پست‌های رسانه‌های اجتماعی به زبان‌های انگلیسی و عربی بود. ابزار «تحلیل احساسات عربی» مبتنی بر توزیع کلاود را از چارچوب کلان‌داده‌ی هُدوپ است که بر روی محیط ابر خصوصی گلف ایر اجرا شده و از پلتفرم اپلیکیشن اینترپرایز رِدِهت جی‌باس^۲ نیز استفاده می‌کند. این ابر خصوصی قادر به نگهداری ۵۰ ترابایت داده است و ابزار تحلیل احساسات عربی می‌تواند هزاران پست را بر روی رسانه‌های اجتماعی تحلیل نموده و نتایج را در عرض چند دقیقه ارائه کند.

گلف ایر با قرار دادن اپلیکیشن «تحلیل احساسات عربی»^۳ بر روی محیط ابر خصوصی فعلی شرکت توانست صرفه‌جویی قابل توجهی در هزینه‌ها به عمل آورد زیرا نیازی به سرمایه‌گذاری برای راه‌اندازی یک زیرساخت جهت استقرار برنامه کاربردی نبود. «تحلیل احساسات عربی» به گلف ایر کمک کرد تا در مورد آفرها و تبلیغات برای مسافران تصمیم به موقع و بهنگام بگیرند و جلوتر از رقبایشان باشند. در خصوص خرابی سرور اصلی نیز شرکت هواپیمایی «تصاویر دوگانه‌ای»^۴ از سرور را ایجاد کرد که می‌توانست بلافاصله پس از کار افتادن سرور اصلی به جای آن فعال شود. این راه‌حل کلان‌داده با سرعت و کارآمدی بسیار بالایی پست‌ها را به صورت دوره‌ای اخذ کرده و آنها را به گزارشات تبدیل می‌کرد و بدین ترتیب گلف ایر می‌توانست دیدهای به‌روزی از تغییرات در عقاید و نظرات یا تقاضا بدست آورده و به سرعت به آنها پاسخ دهد. بینش‌های بدست آمده از این راه‌حل کلان‌داده تأثیر مثبتی بر کار کارمندان گلف ایر داشت.

Sources: RedHat.com. (2016). "Gulf Air Builds Private Cloud for Big Data Innovation with Red Hat Technologies." www.redhat.com/en/about/press-releases/gulf-air-builds-private-cloud-big-data-innovation-redhat-technologies (accessed October 2018); RedHat.com. (2016). "Gulf Air's Big Data Innovation Delivers Deeper Customer Insight." www.redhat.com/en/success-stories (accessed October 2018); ComputerWeekly.com.(2016). "Big-Data

¹ Gulf Air

² Red Hat JBoss Enterprise Application

³ Arabian Sentiment Analysis

⁴ Ghost Images

چایم^۱ یک گزینه‌ی بانکداری است که به ارائه‌ی یک کارت نقدی ویزا، حساب پس‌انداز با تضمین اِف‌دی‌آی‌سی^۲ و یک اپلیکیشن موبایل می‌پردازد که بانکداری را برای عموم مردم آسان‌تر کرده است. چایم به دنبال این بود که اطلاعات بیشتری در خصوص مشارکت و درگیرسازی مشتریان خود بدست آورد. آنها می‌خواستند داده‌های بدست آمده از پلتفرم‌های موبایل، وب و بک‌اند خود را برای کمک به ارتقاء تجربه‌ی کاربران تحلیل نمایند. با وجود این، اخذ و جمع‌آوری داده‌ها از منابع متعدد از قبیل سرویس‌های آگهی‌های تبلیغاتی فیسبوک و گوگل و رویدادهای ابزارهای تحلیلی شخص ثالث مثل اسناد جی‌سان (جاوااسکرپت آبجکت نوتیفیکیشن^۳) کار بسیار دشواری بود. آنها به نرم‌افزاری نیاز داشتند تا بتواند داده‌های بدست آمده از این منابع را جمع و تحلیل کند. چایم به راه‌حلی نیاز داشت که بتواند منابع داده‌ی جی‌سان را پردازش کرده و با استفاده از جداول پایگاه داده‌های اس‌کیوال استاندارد، پرس و جوهای لازم را به عمل آورد.

چایم شروع به استفاده از راه‌حل پایگاه داده‌ی تحلیلی الاستیک^۴ شرکت اسنو فلیک نمود. اسنو فلیک داده‌ها را از تمامی ۱۴ منبع داده‌ی شرکت مثل اسناد جی‌سان بیرون می‌کشید. اسنو فلیک به چایم در تحلیل سریع داده‌های جی‌سان به منظور ارتقاء خدمات اعضا و ارائه‌ی تجربه‌ی بانکداری شخصی‌تر به مشتریان کمک کرد.

Source: Based on Snowflake.net. (n.d.). Chime delivers personalized customer experience using Chime. <http://www.snowflake.net/product> (accessed Oct 2018).

ما در حال ورود به «عصر پتابایت» هستیم و رویکردهای سنتی تحلیل‌ها و داده‌ها به پایان عمر خود رسیده‌اند. تحلیل ابری یک راه‌حل جایگزین نوظهور برای تحلیل داده‌ها در مقیاس کلان است. سیستم‌های ابری داده‌محور شامل ذخیره و رایانش در یک محیط پراکنده (توزیع‌شده) و مجازی‌سازی شده هستند. یک مزیت اصلی این محصولات، انتشار سریع ابزارهای تحلیلی پیشرفته بین کاربران است بدون اینکه سرمایه‌گذاری هنگفتی در بخش خرید فناوری انجام شود. این راه‌حل‌ها البته با چالش‌های بسیاری نیز همراه هستند مثل امنیت، سطح سرویس و حکمرانی داده‌ها. نگرانی‌هایی در خصوص رایانش ابری بوجود آمده است که از آن جمله می‌توان به از دست رفتن کنترل و حریم خصوصی، تعهدات قانونی، مسائل سیاسی برون مرزی و غیره اشاره کرد. طبق نظر «ائتلاف امنیت فضای ابری»، سه تهدید امنیتی مهم در فناوری ابری عبارتند از «از دست رفتن و نشر داده‌ها»، «خرابی تجهیزات سخت‌افزاری» و «رابط کاربری نامطمئن». تمامی داده‌های موجود در ابر در دسترس عرضه‌کننده‌ی سرویس هستند و بنابراین عرضه‌کننده‌ی سرویس می‌تواند ناخواسته یا عمدانه داده‌ها را تغییر داده یا آنها را بدون اخذ مجوز از شرکت‌ها در اختیار شخص ثالث قرار دهد. پژوهش‌های محدودی در این حوزه انجام شده است. در نتیجه، فرصت بسیار خوبی برای وارد کردن مدلسازی تحلیلی، رایانشی و مفهومی به مقوله‌ی علوم خدمات، جهت‌گیری خدمات و اطلاعات ابری وجود دارد. با وجود این، رایانش ابری یک ابتکار عمل مهم برای یک کارشناس تحلیل محسوب می‌شود.

¹ Chime

² FDIC-insured

³ JavaScript Object Notification

⁴ Elastic Data Warehouse

سوالاتی برای مرور بخش ۹.۹

۱۷. رایانش ابری را تعریف کند. ارتباط آن با «پلتفرم به عنوان یک سرویس»، «نرم افزار به عنوان یک سرویس» و «زیرساخت به عنوان یک سرویس» چیست؟
۱۸. مثال‌هایی از شرکت‌هایی که به عرضه خدمات ابری می‌پردازند ارائه کنید.
۱۹. رایانش ابری چگونه بر هوش تجاری تأثیر می‌گذارد؟
۲۰. «داده به عنوان یک سرویس» چگونه نحوه مدیریت داده‌ها را دستخوش تغییر و تحول کرده است؟
۲۱. انواع مختلف پلتفرم‌های ابری کدام‌ها هستند؟
۲۲. چرا «تحلیل به عنوان یک سرویس» مقرون به صرفه است؟
۲۳. حداقل سه عرضه‌کننده اصلی خدمات ابری را نام ببرید.
۲۴. حداقل سه مثال از عرضه‌کنندگان «تحلیل به عنوان یک سرویس» ارائه کنید.

تا به اینجا، مثال‌های بسیاری از بکارگیری تکنیک‌های تحلیلی توسط سازمان‌ها برای دستیابی به بینش در خصوص فرایندهای موجود از طریق گزارش‌گیری آموزنده، تحلیل‌های پیشگویانه، پیش‌بینی و تکنیک‌های بهینه‌سازی ارائه کردیم. در این بخش در مورد یک روند نوظهور بسیار مهم یعنی بکارگیری داده‌های مکانی در تحلیل‌ها صحبت خواهیم کرد. شکل ۹.۱۵ دسته‌بندی ما از کاربردهای تحلیل‌های مبتنی بر مکان را نشان می‌دهد. ما ابتدا به مرور برنامه‌های کاربردی که از داده‌های مکانی ایستا که معمولاً *داده‌های مکانی-جغرافیایی*^۱ نامیده می‌شوند استفاده می‌کنند می‌پردازیم. سپس به بررسی رشد انفجاری برنامه‌های کاربردی که در آنها از تمامی داده‌های مکانی تولید شده توسط دستگاه‌های امروزی استفاده می‌شود خواهیم پرداخت. تمرکز ما در این بخش ابتدا بر کاربردهای تحلیلی است که توسط سازمان‌ها برای تصمیم‌گیری بهتر در مدیریت عملیات، هدف‌گیری مشتریان، تبلیغات و غیره توسعه می‌یابند. سپس به جستجو و تحقیق برنامه‌های کاربردی تحلیلی که برای استفاده‌ی مستقیم کاربران توسعه یافته‌اند می‌پردازیم که برخی از آنها از داده‌های مکانی نیز بهره می‌برند.

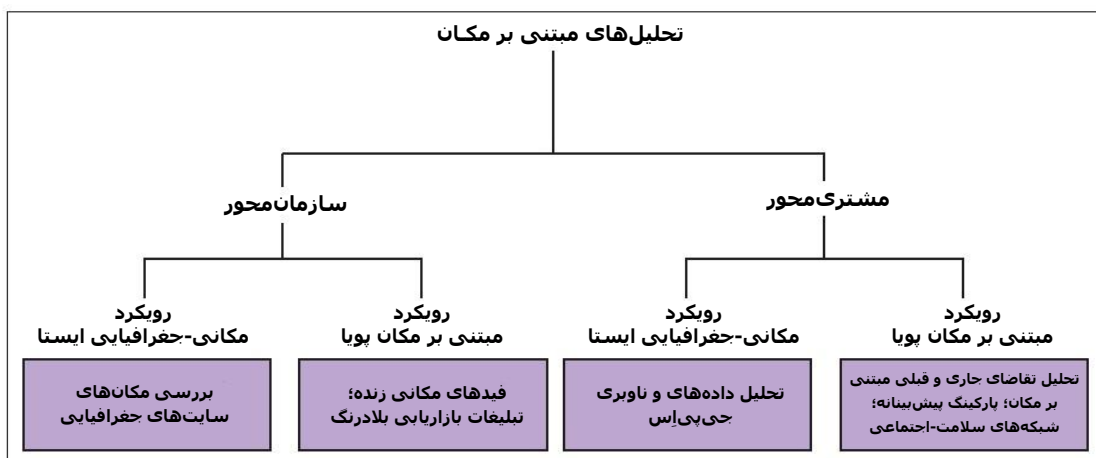
ابزارهای مصورسازی یک دید منسجم و کامل از عملکرد کلی یک سازمان را نشان می‌دهد که اطلاعات اقدام‌پذیری را در اختیار ما قرار می‌دهد. این اطلاعات می‌تواند شامل مقادیر جاری و پیش‌بینی شده از عوامل مختلف کسب‌وکار و شاخص‌های کلیدی عملکرد باشد. نگاه به شاخص‌های کلیدی عملکرد به صورت اعداد کلی از طریق نمودارها و گراف‌های گوناگون می‌تواند طاقت‌فرسا باشد. ریسک از دست دادن فرصت‌های رشد بالقوه یا عدم شناسایی حوزه‌های مشکل‌دار در اینجا بسیار بالا است. سازمان‌ها به عنوان روشی جایگزین برای مشاهده‌ی گزارشات معمولاً از نقشه‌های دیداری که مبتنی بر داده‌های نگاشت شده بر حسب منقطه‌ی جغرافیایی و مکانی هستند و معمولاً بر اساس کدپستی گروه‌بندی می‌شوند استفاده می‌کنند. این مصورسازی‌های مبتنی بر نقشه توسط سازمان‌ها برای مشاهده‌ی داده‌های تجمیع شده و دستیابی به بینش‌های معنادارتر مبتنی بر مکان‌ها بکار گرفته شده‌اند. این تکنیک‌های تحلیلی مبتنی بر مکان که از کدگذاری جغرافیایی مکان سازمان و مشتری استفاده می‌کنند مانعی بر سر راه

¹ Geospatial Data

درک تأثیرات «حقیقی مبتنی بر مکان» هستند. مکان‌های مبتنی بر کدپستی می‌توانند دید یکپارچه و تجمیعی از یک منطقه‌ی جغرافیایی بزرگ را به ما بدهند. این سطح ضعیف از دانه‌بندی ممکن است کمکی به شناسایی فرصت‌های رشد در یک منطقه نکند زیرا مکان مشتریان هدف می‌تواند به سرعت تغییر کند. بنابراین، کمپین‌های تبلیغاتی یک سازمان در صورتیکه از کدهای پستی استفاده می‌کنند ممکن است مشتریان درست را هدف قرار ندهد. برای رفع این مشکل، سازمان‌ها از اطلاعات مکانی و جغرافیایی در تحلیل‌ها استفاده می‌کنند. افزودن مؤلفه‌های مکانی بر اساس طول و عرض جغرافیایی به تکنیک‌های تحلیلی رایج می‌تواند سازمان‌ها را قادر کند تا یک بُعد جدید از «چه جایی» را به تحلیل‌های کسب‌وکارشان بیفزایند که در حال حاضر فقط به سوالات «چه کسی»، «چه چیزی»، «چه زمانی» و «چه میزان» پاسخ می‌دهند.

داده‌های مبتنی بر مکان در حال حاضر توسط سیستم‌های اطلاعات جغرافیایی (جی‌آی‌اس)¹ در دسترس قرار می‌گیرند. از این سیستم‌ها برای ضبط، ذخیره‌سازی، تحلیل و مدیریت داده‌های مربوط به یک مکان با استفاده از فناوری‌های حسگر یکپارچه، سیستم‌های مکان‌یاب جهانی که بر روی گوشی‌های هوشمند نصب شده‌اند یا از طریق استقرار آرف‌آی دی در صنایع خرده‌فروشی یا مراقبت سلامت استفاده می‌شود.

با ادغام و یکپارچه‌سازی اطلاعات مربوط به یک مکان با سایر داده‌های حساس کسب‌وکار، سازمان‌ها اکنون می‌توانند اطلاعات مکانی را تولید کنند. اطلاعات مکانی سازمان‌ها را قادر می‌سازد تا به بینش‌های حیاتی دست یافته و تصمیمات بهتری را از طریق بهینه‌سازی فرایندها و برنامه‌های کاربردی اتخاذ کنند. سازمان‌ها اکنون نقشه‌های تعاملی را ایجاد می‌کنند که می‌توان جزئیات مربوط به هر مکان را مشاهده کرد و تحلیلگران قادر به تحقیق و تفحص در روندهای جدید و همبستگی بین عوامل خاص مکانی بین شاخص‌های کلیدی عملکرد متعدد هستند. تحلیلگران اکنون می‌توانند روندها و الگوهای درآمد، فروش و سودآوری را در مناطق جغرافیایی مشخص و تعیین کنند.



شکل ۹.۱۵ دسته‌بندی کاربردهای تحلیل مبتنی بر مکان

¹ Geographic Information Systems (GIS)

esri.com

grindgis.com

(<http://grindgis.com/blog/gis-applications-uses>)

sensefly.com sstsoftware.com

TheGuardian.com, 2013

ST_GEOMETRY

¹ ESRI

² ArcGIS

³ SQL/MM

گریت کلیپس^۱ یکی از بزرگترین و سریع‌الرشدترین سالن‌های آرایش مو با بیش از ۳,۰۰۰ سالن در سرتاسر ایالات متحده و کانادا می‌باشد. موفقیت فرآیند گریت کلیپس به استراتژی رشد از طریق افتتاح سریع فروشگاه‌های جدید در مکان‌ها و بازارهای درست بستگی دارد. شرکت نیازمند این بود که مکان‌ها را بر اساس نیازمندی‌ها و الزامات برای یک پایگاه مشتریان بالقوه، روندهای جمعیت‌شناختی و تأثیر فروش بر فانشیزهای موجود در مکان هدف تجزیه و تحلیل نماید. انتخاب یک محل خوب از بالاترین اهمیت برخوردار است. طبق فرایندهای جاری، تجزیه و تحلیل یک سایت نیازمند زمان بسیار زیادی است و ارزیابی دستی داده‌های بدست آمده از منابع متعدد نیز انرژی و زمان بسیار زیادی را طلب می‌کند.

با وجود سالانه هزاران مکان مختلف که امکان تجزیه و تحلیل آنها وجود دارد، تأخیر در این تحلیل‌ها این ریسک را داشت که سایت‌ها و محل‌های خوب در اختیار سایر رقبا قرار بگیرد در عین اینکه اینکار بسیار پرهزینه بود. گریت کلیپس از پیمانکاران بیرونی برای رفع این مشکل استفاده کرد و یک برنامه کاربردی جریان کار انتخاب سایت را برای ارزشیابی مکان‌های مورد نظر جهت افتتاح سالن جدید با استفاده از قابلیت‌های تحلیلی مکانی-جغرافیایی آلتریکس^۲ ایجاد نمود. یک مکان جدید بر اساس مجاورت و راحتی در ارائه خدمت به کلیه‌ی شبکه مشتریان گریت کلیپس در آن منطقه مورد ارزشیابی قرار می‌گیرد. این راه‌حل مبتنی بر آلتریکس همچنین امکان ارزشیابی هر مکان جدید را بر اساس داده‌های رفتار مشتری و جمعیت‌شناختی فراهم می‌کرد که کاملاً با پروفایل مشتریان فعلی گریت کلیپس و تأثیر بالقوه‌ی درآمد حاصل از افتتاح سایت جدید بر سایت‌های موجود همسو و همراستا بود. در نتیجه‌ی بکارگیری تکنیک‌های تحلیلی مبتنی بر مکان، گریت کلیپس توانست زمان ارزیابی مکان‌های جدید را تا نزدیک به ۹۵٪ کاهش دهد. تحلیل‌های سنگین و زمان‌بر به صورت خودکار درآمده و در قالب یک برنامه‌ی کاربردی تحلیل گردآوری داده، نگاشت و گزارش‌گیری توسعه داده شدند که مدیران غیرفنی املاک و مستغلات به راحتی می‌توانستند از آن استفاده کنند. همچنین، شرکت توانست تحلیل‌های پیش‌گویانه و کنشگرانه را برای چند مکان فرانشیز جدید پیاده‌سازی کند که کل این فرایند تنها چند دقیقه به طول می‌انجامد.

سوالاتی برای بحث و گفتگو

۱. از تحلیل مکانی-جغرافیایی چگونه در گریت کلیپس استفاده می‌شود؟

¹ NAVTEQ

² Tele Atlas

³ Informatica

⁴ Trillium

⁵ Great Clipse

⁶ Alteryx

۲. یک شرکت در ارزشیابی مکان‌های سایت‌های جدید خود چه معیارهایی را باید در نظر بگیرد؟
۳. آیا می‌توانید به کاربردهای دیگری از داده‌های مکانی-جغرافیایی اشاره کنید؟

Source: Based on Alteryx.com. Great Clips. alteryx.com/sites/default/files/resources/files/case-study-great-chips.pdf (accessed Sept 2018).

استارباکس از سیستم اطلاعات جغرافیایی و تحلیل‌ها برای رشد جهانی استفاده می‌کند

مورد کاربردی ۹.۱۱

یکی از چالش‌های اصلی هر سازمانی که به دنبال رشد و توسعه‌ی حضور خود در بازار است، تصمیم‌گیری در مورد مکان شعبه‌ی بعدی که می‌خواهد افتتاح کند می‌باشد. استارباکس نیز با همین مسئله روبرو است. برای شناسایی مکان‌ها و موقعیت‌های شعب جدید، بیش از ۷۰۰ نفر از کارمندان استارباکس (که به آنها شرکا گفته می‌شود) در ۱۵ کشور از یک راه‌حل هوش تجاری و برنامه‌ریزی بازار مبتنی بر آرک‌جی‌آی‌اس^۱ به نام اطلس^۲ استفاده می‌کنند. اطلس امکاناتی از قبیل جریان‌کارها، تحلیل و اطلاعات عملکرد فروشگاه را به گونه‌ای در اختیار شرکا قرار می‌دهد که شرکای منطقه‌ای در زمان شناسایی فرصت‌های جدید کسب‌وکار بتوانند تصمیم‌گیری کنند.

طبق گزارشات منتشر شده در منابع متعدد، تصمیم‌گیران منطقه‌ای برای درک روندهای جمعیتی و تقاضا از اطلس استفاده می‌کنند. به عنوان مثال، در چین، بیش از ۱،۲۰۰ شعبه‌ی استارباکس وجود دارد و شرکت تقریباً هر روز یک شعبه‌ی جدید افتتاح می‌کند. اطلاعاتی از قبیل مناطق تجاری، خوشه‌های خرده‌فروشی، ترافیک و جمعیت‌شناختی در تصمیم‌گیری در خصوص محل شعبه‌ی جدید بسیار اهمیت دارند.

پس از تحلیل یک بازار جدید و بررسی همسایگی، یک مدیر می‌تواند با بزرگنمایی بر روی هر منطقه‌ای از شهر می‌تواند پی ببرد که به عنوان مثال طی ۲ ماه گذشته در کدام بخش از شهر سه برج اداری جدید افتتاح شده است. پس از مشاهده‌ی این منطقه بر روی نقشه، یک پنجره‌ی جریان‌کار را می‌توان ایجاد کرد که به مدیر کمک می‌کند تا کار اخذ تأییدیه، اخذ مجوز، ساخت و نهایتاً افتتاح را برای سایت جدید انجام دهد.

با ادغام و یکپارچه‌سازی داده‌های آب و هوایی و سایر داده‌های محلی عملیات زنجیره تأمین و تقاضا را بهتر می‌توان مدیریت نمود. استارباکس در حال یکپارچه‌سازی سیستم‌های کسب‌وکار سازمانی خود با راه‌حل‌های سیستم‌های اطلاعات جغرافیایی در قالب سرویس‌های وب است تا بتواند نگاه جدیدی به کسب‌وکارش و جهان داشته باشد. به عنوان مثال، استارباکس داده‌های مربوط به دمای پیش‌بینی شده‌ی هوا توسط اکیوویدر^۳ را در سیستم‌های خود ادغام کرده است. این داده‌های دمای پیش‌بینی شده‌ی هوا می‌توانند به متمرکزسازی و محلی‌سازی فعالیت‌های بازاریابی کمک کنند. اگر هفته‌ی بسیار گرمی را در ممفیس پیش رو داشته باشیم، تحلیلگران استارباکس می‌توانند گروهی از قهوه‌خانه‌ها را انتخاب کرده و اطلاعات بسیار کاملی از الگوهای گذشته و آینده‌ی آب و هوا و همچنین ویژگی‌های شعب آن منطقه را بدست آورند. از این دانش می‌توان برای طراحی یک برنامه تبلیغاتی محلی‌سازی شده برای مثلاً کاپوچینو استفاده کرد و به استارباکس کمک می‌کند تا خواسته‌ها و نیازهای مشتریان را از یک هفته قبل پیش‌بینی نمایند.

رویدادهای مهم نیز می‌توانند بر کسب‌وکار قهوه‌خانه‌ها تأثیر بگذارند. زمانیکه ۱۵۰،۰۰۰ نفر برای رژه‌ی افتخار^۴ در سان‌دیگو وارد خیابان‌ها شدند، کافه‌های محلی توانستند به مشتریان بسیاری خدمت‌رسانی کنند. استارباکس برای حصول اطمینان از اینکه مشتریان بهترین تجربه‌ی ممکن را خواهند داشت از این دانش رویدادهای محلی برای برنامه‌ریزی تخصیص نیروی انسانی و موجودی در مکان‌های نزدیک به رژه استفاده کرد.

سوالاتی برای بحث و گفتگو

¹ ArcGIS-based

² Atlas

³ AccuWeather

⁴ Pride Parade

۱. برای تصمیم‌گیری در خصوص مکان جدید یک شعبه چه اطلاعات جمعیت‌شناختی و سیستم اطلاعات جغرافیایی مرتبط است؟
۲. گفته شد که استارباکس مشتریان خود را تشویق به استفاده از اپلیکیشن موبایل خود می‌کند. این شرکت چه نوع اطلاعاتی را می‌تواند برای کمک به برنامه‌ریزی بهتر عملیات از این اپلیکیشن جمع‌آوری کند؟
۳. آیا دسترسی رایگان به اینترنت وای‌فای در شعب استارباکس می‌تواند اطلاعاتی را برای تحلیل‌های بهتر در اختیار استارباکس قرار دهد؟

Sources: Digit.HBS.org. (2015). "Starbucks: Brewing up a Data Storm!" <https://digit.hbs.org/submission/starbucks-brewingup-a-data-storm/> (accessed October 2018); Wheeler, C. (2014). "Going Big with GIS." www.esri.com/esri-news/arcwatch/0814/going-big-with-gis (accessed October 2018); Blogs.ESRI.com. "From Customers to CxOs, Starbucks Delivers World-Class Service." (2014). <https://blogs.esri.com/esri/ucinsider/2014/07/29/starbucks/> (accessed October 2018).

علاوه بر کاربردهای تحلیل تراکنش‌های خرده‌فروشی که در اینجا به آنها اشاره شد، کاربردهای بسیار زیاد دیگری از ترکیب اطلاعات جمعیت‌شناختی با سایر داده‌های تولید شده توسط سازمان‌ها وجود دارد. به عنوان مثال، شرکت‌های ارتباطات و عملیات شبکه اغلب روزانه حجم انبوهی از داده‌ها را تولید می‌کنند. توانایی تحلیل سریع داده‌ها با سطح تفکیک‌پذیری خاص مکانی می‌تواند به شناسایی بهتر علل ریزش مشتریان و کمک به تدوین استراتژی‌های خاص برای مکان‌های مختلف به منظور افزایش کارایی عملیاتی، کیفیت خدمات و درآمد کمک کند.

شرکت‌های ارتباطی با استفاده از تحلیل مکانی-جغرافیایی می‌توانند تراکنش‌های روزانه را از یک شبکه دریافت کرده و مناطق جغرافیایی خاصی که دچار قطعی زیاد ارتباطات صوتی، داده‌ها، متنی یا اینترنتی شده‌اند را شناسایی نماید. تحلیل‌ها می‌توانند به تعیین علل دقیق بر اساس منطقه کمک کرده و برای ارائه خدمات بهتر به مشتریان تا سطح فردی برای هر مشتری پیش بروند. با انجام تمرین چندرسانه‌ای زیر می‌توانید این موضوع را در عمل ببینید.

در شبکه‌ی دانشگاه ترادیتا یک ویدئوی بی‌اس‌آی در مورد تماس‌های قطع شده‌ی موبایل موجود است. این ویدئو را در سایت یوتیوب از آدرس www.teradatauniversitynetwork.com/Library/Samples/BSI-The-Case-of-the-Dropped-Mobile-Calls (accessed October 2018). تماشا کنید. یک شرکت مخابراتی در حال راه‌اندازی خط جدیدی از گوشی‌های همراه است. با مشکلات قطعی تماس‌ها مواجه است. عرضه‌ی محصول را مشکل مواجه شده و منطقه‌ی شمال شرق با توجه به مقایسه‌ی تأثیرات قطعی تماس‌ها بر سود در آن منطقه‌ی جغرافیایی بیش از سایر مناطق دچار مشکل شده است. این شرکت برای تحلیل مشکلات ناشی از عیوب هندسیت گوشی‌های هوشمند، پوشش‌دهی دکل‌ها و ایرادات نرم‌افزاری، بی‌اس‌آی را استخدام نمود. داده‌های کل منطقه‌ی شمال شرق به خوشه‌های جغرافیایی تقسیم شده و شرکت با شناسایی داده‌های مربوط به تک تک مشتریان این مشکل را برطرف نمود. تیم بی‌اس‌آی از تحلیل‌های مکانی-جغرافیایی برای شناسایی مکان‌هایی که پوشش‌دهی شبکه منجر به قطعی تماس می‌شده استفاده کرده و پیشنهاد داد تا چند دکل دیگر در مکان‌هایی که مشتریان ناراضی داشته‌اند نصب شود.

پس از تماشای ویدئو، خواهید دید که این تحلیل چگونه آماده شده است: slideshare.net/teradata/bsi-teradata-the-case-of-the-dropped-mobile-calls (accessed October 2018).

این تمرین یک مثال از ترکیب تحلیل‌های مکانی-جغرافیایی با تحلیل‌های کلان‌داده‌ها که به تصمیم‌گیری بهتر کمک می‌کنند را نشان داد.

بسیاری از دستگاه‌هایی که مشتریان و متخصصان استفاده می‌کنند دائماً در حال ارسال اطلاعات مکانی خود هستند. خودروها، اتوبوس‌ها، تاکسی‌ها، تلفن‌های سیار، دوربین‌ها و دستگاه‌های ناوبری شخصی همگی به لطف فناوری‌های موقعیت‌یاب متصل به شبکه از قبیل جی‌پی‌اس، وای‌فای و مثلث‌بندی دکل سلولی، موقعیت و مکان خود را ارسال می‌کنند. میلیون‌ها مصرف‌کننده و شرکت از دستگاه‌های مجهز به قابلیت‌های مکانی برای یافتن خدمات نزدیک، دوستان و اعضای خانواده، ناوبری، ردیابی اموال و حیوانات خانگی، ارسال و مشارکت در ورزش‌ها، بازی‌ها و سرگرمی‌ها استفاده می‌کنند. این افزایش خدمات مبتنی بر مکان منجر به شکل‌گیری یک پایگاه داده‌های عظیم از اطلاعات مکانی تاریخی و در حال جریان شده است. البته که با توجه به عدم تمرکز این پایگاه داده، به خودی خود چندان مفید نیست. گردآوری خودکار داده‌ها از طریق تلفن‌های سلولی انجام شده و نقاط دسترسی هات‌اسپات وای‌فای بعد جدید و جالبی برای پژوهش و تحقیقات بازار، گردآوری داده‌ها و میکروآنالیز اینگونه مجموعه داده‌های انبوه فراهم آورده‌اند.

با تحلیل و یادگیری از این الگوهای جابجایی در مقیاس بزرگ، می‌توان گروه‌های متمایزی از رفتارها را در مقوله‌های خاص شناسایی کرد. این رویکرد به کسب‌وکارها اجازه می‌دهد تا الگوهای مشتریان را بهتر درک کرده و تصمیمات آگاهانه‌تری را در خصوص تبلیغات، قیمت‌گذاری و غیره بگیرند. با بکارگیری الگوریتم‌هایی که چندبعدی بودن داده‌های مکانی را کاهش می‌دهند می‌توان مکان‌ها را بر اساس فعالیت و جابجایی بین آنها مشخص کرد. این الگوریتم‌ها روندها، مفاهیم و روابط را از میان حجم انبوهی از داده‌های مکانی چندبعدی کشف کرده و بازنمودهای قابل درک برای انسان را تولید می‌کنند. سپس می‌توان از اینگونه داده‌ها به صورت خودکار برای پیش‌بینی‌های هوشمند و یافتن مهم‌ترین تشابهات و تطابق‌ها بین افراد و مکان‌ها استفاده کرد. تحلیل‌های مبتنی بر مکان در بازاریابی مشتری‌محور نیز کاربرد دارند. بسیاری از شرکت‌ها در حال حاضر پلتفرم‌هایی را به منظور تحلیل رد مکانی کاربران موبایل بر اساس داده‌های مکانی-جغرافیایی بدست آمده از جی‌پی‌اس و هدف قرار دادن مشتریان هدف با کوپن‌ها بر روی گوشی‌های هوشمندی به محض عبور از کنار یک خرده‌فروشی عرضه می‌کنند. این بیانگر روندی نوظهور در فضای خرده‌فروشی است که شرکت‌ها به دنبال بهبود کارایی کمپین‌های بازاریابی هستند - نه فقط با هدف قرار دادن هر مشتری بر اساس مکان و موقعیت لحظه‌ای و آنی بلکه با بکارگیری تحلیل‌های پیشگویانه‌ی پیچیده‌تر به صورت بلادرنگ بر روی پروفایل رفتاری مشتری به منظور یافتن مجموعه مناسب و درستی از مشتریان برای کمپین‌های تبلیغاتی.

اما یک انشعاب دیگر از تحلیل‌های مبتنی بر مکان، استفاده از واقعیت افزوده است. در سال ۲۰۱۶، پوکمون گو تبدیل به یک حس‌گری بازار شد. پوکمون گو یک بازی مبتنی بر واقعیت افزوده مکان‌محور است که کاربران را تشویق می‌کند تا اقلام مجازی را از مکان‌های جغرافیایی منتخب مطالبه کنند. کاربر می‌تواند بازی از هر جایی در شهر شروع کرده و علائم نشانه‌گذاری شده بر روی اپلیکیشن را برای رسیدن به یک کالای خاص دنبال کند. اقلام مجازی زمانیکه کاربر دوربین گوشی خود را به سمت آن اقلام مجازی می‌گیرد بر روی اپلیکیشن قابل مشاهده می‌شوند. سپس او می‌تواند آن کالا را مطالبه کند. کاربردهای کسب‌وکار اینگونه فناوری‌ها نیز در حال ظهور هستند. به عنوان مثال، یک اپلیکیشن به نام کندی‌بار به کسب‌وکارها اجازه می‌دهد تا این اقلام مجازی را با استفاده گوگل مپس بر روی نقشه قرار دهند. جایگیری این کالا را با استفاده از گوگل استریت ویو می‌توان به خوبی تنظیم کرد. پس از پیکربندی و تنظیم تمامی کالاهای مجازی به کمک اطلاعات و مکان، کسب‌وکارها می‌توانند اقلامی را که برای کاربران قابل مشاهده هستند در همان لحظه ارسال کنند. شرکت‌ها همچنین از کندی‌بار می‌توانند به عنوان تحلیل‌ها برای هدف‌گیری بهتر اقلام مجازی استفاده کنند. جنبه‌ی واقعیت مجازی این اپلیکیشن موجب بهبود تجربه‌ی کاربر شده و یک

محیط بازی را در زندگی واقعی فراهم می‌کند. در عین حال، یک پلتفرم بازاریابی قدرتمند را برای کسب‌وکارها برای دستیابی به مشتریان‌شان ارائه می‌کند. همانگونه که از مطالب این بخش متوجه شدیم، تحلیل‌های مبتنی بر مکان و برنامه‌های کاربردی مرتبط با آنها شاید مهمترین موضوعات برای سازمان‌ها در آینده‌ای نزدیک باشند. یک مضمون مشترک در این بخش استفاده از داده‌های عملیاتی یا بازاریابی توسط سازمان‌ها بود. ما در ادامه به بررسی برنامه‌های کاربردی تحلیلی که مستقیماً کاربران را هدف قرار داده و برخی اوقات از اطلاعات مکانی بهره می‌برند خواهیم پرداخت.

رشد انفجاری صنعت اپلیکیشن‌سازی برای پلتفرم‌های گوشی‌های هوشمند (آی‌اَس، اندورید، ویندوز و غیره) و بکارگیری تحلیل‌ها، فرصت‌های بسیاری را برای توسعه‌ی اپلیکیشن‌ها بوجود آورده است در جائیکه مصرف‌کنندگان از تحلیل‌ها بدون آنکه مطلع باشند استفاده می‌کنند. این اپلیکیشن‌ها با گروه قبلی که ذکر شد از این جهت تفاوت دارند که برای استفاده‌ی مستقیم مصرف‌کننده در نظر گرفته شده‌اند در حالیکه آنها برای سازمان‌هایی طراحی شده بودند که سعی در کاوش داده‌های بدست آمده از خرید/استفاده‌ی کاربران برای ایجاد یک پروفایل برای بازاریابی محصولات یا خدمات خاصی بودند. قابل پیش‌بینی است که این اپلیکیشن‌ها به مصرف‌کنندگان این توان را می‌دهد تا با بکارگیری تحلیل‌های خاص تصمیمات بهتری را اتخاذ کنند. ما در اینجا به ذکر دو مثال اکتفا می‌کنیم:

- ویزا¹ یک اپلیکیشن وب اجتماعی است که به کاربران در شناسایی مسیر کمک کننده در در مورد مشکلات بالقوه از قبیل تصادفات، ایستگاه‌های پلیس؛ تله‌های سرعت و کارگران در حال کار بر اساس ورودی‌های سایر کاربران هشدار می‌دهد که به یک اپلیکیشن مسیریابی بسیار محبوب تبدیل شده است. گوگل این اپلیکیشن را چند سال پیش خریداری کرد و از آن زمان به ارتقاء آن پرداخته است. این اپلیکیشن یک مثال از تجمیع اطلاعات تولید شده توسط کاربران و در دسترس دادن مشتریان است.
- اپلیکیشن‌های بسیاری به کاربران اجازه می‌دهند تا نظرات و امتیازات خود در مورد شرکت‌ها، محصولات و غیره را ارسال کرده و آنها را به یک فرم تجمیع شده برای سایر کاربران ارسال کنند تا آنها نیز بتوانند تصمیم‌گیری و انتخاب کنند. این اپلیکیشن‌ها را همچنین می‌توان به عنوان اپلیکیشن‌های مبتنی داده‌های اجتماعی دانست که مستقیماً مصرف‌کنندگان را هدف قرار می‌دهند که داده‌ها توسط مصرف‌کنندگان تولید شده است. یکی از اپلیکیشن‌های محبوب‌تر در این گروه، پلپ نام دارد. اپلیکیشن‌های مشابه دیگری در سرتاسر جهان وجود دارند.
- یک اپلیکیشن دیگر مرتبط با بحث حمل و نقل که از تحلیل‌های پیشگویانه استفاده می‌کند پارک‌پی‌جی است که از حدود سال ۲۰۱۰ در پیتسبورگ ایالت پنسیلوانیا استقرار یافت. اینکه اپلیکیشن که با همکاری دانشگاه کارنجی ملون توسعه یافته است شامل قابلیت‌های پیشگویانه برای تخمین فضای موجود برای پارک خودرو است. پارک‌پی‌جی‌ای رانندگان را به سمت فضاهای پارکینگ خالی راهنمایی می‌کند. این اپلیکیشن تعداد فضاهای پارکینگ موجود در چندین پارکینگ در ناحیه‌ی هنرهای فرهنگی پیتسبورگ را محاسبه می‌کند. فضاهای خالی هر ۳۰ ثانیه به روزآوری می‌شوند و راننده را به نزدیک‌ترین فضای پارکینگ موجود راهنمایی می‌کند. بسته به رویدادهای جاری و تاریخچه‌ی تقاضا، این اپلیکیشن قادر به پیش‌بینی فضای پارکینگ موجود و ارائه اطلاعات در مورد پارکینگ‌های خالی در زمان رسیدن راننده به مقصد است. الگوریتم زیربنایی اپلیکیشن از داده‌های مربوط به رویدادهای جاری در آن ناحیه - مثلاً یک بازی

¹ Wayze

بسکتبال - برای پیش‌بینی میزان افزایش تقاضای فضای پارکینگ در آن روز استفاده کرده و بدین ترتیب در زمان ارزشمند مردم برای جستجوی فضای پارکینگ در آن شهر شلوغ صرفه‌جویی می‌کند. موفقیت این اپلیکیشن منجر به تکثیر اپلیکیشن‌های پارکینگ در بسیاری از شهرهای مهم شده و به کاربر اجازه می‌دهد تا فضای پارکینگ را از قبل رزرو کرده، دستگاه پارکبان را شارژ کرده و حتی درخواست فضای پارکینگ کنند. بسیاری از اینگونه اپلیکیشن‌ها را می‌توان هم از آپ‌استور آی‌فون و هم گوگل پلی بارگذاری کرد.

اپلیکیشن‌های مبتنی بر تحلیل‌ها نه فقط برای سرگرمی یا سلامت بلکه برای افزایش میزان بهره‌وری در حال ظهور هستند. اپلیکیشن‌های الکترونیکی گوگل به نام جیمیل به تحلیل میلیاردها تبادل نامه‌های الکترونیکی پرداخته و پاسخ‌های خودکار را برای نامه‌ها تولید می‌کند. زمانیکه یک کاربر یک نامه‌ی الکترونیکی را دریافت کرده و آن را در اپلیکیشن جیمیل خود باز می‌کند، اپلیکیشن نیز پاسخ‌های کوتاهی را برای آن نامه الکترونیکی به کاربر پیشنهاد می‌دهد که وی می‌تواند از بین آنها انتخاب کرده و برای فرستنده‌ای اصلی ارسال کند. همانطور که از این مثال‌های اپلیکیشن‌های متمرکز بر مصرف‌کنندگان مشخص است، تحلیل‌های پیشگویانه امکان توسعه‌ی نرم‌افزارهایی را فراهم آورده‌اند که مستقیماً توسط مصرف‌کنندگان بکار گرفته می‌شوند. ما معتقدیم که رشد اپلیکیشن‌های تحلیلی مشتری‌محور همچنان ادامه خواهد داشت و فرصت‌های بسیاری را برای خوانندگان این کتاب بوجود خواهد آورد.

یکی از نگرانی‌های اصلی در بکارگیری این فناوری‌ها، نقض حریم خصوصی است. اگر شخصی بتواند جابجایی یک تلفن سلولی را ردگیری کند، حریم خصوصی آن مشتری در معرض خطر خواهد بود. برخی از توسعه‌دهندگان اپلیکیشن‌ها معتقدند که آنها فقط به دنبال گردآوری اطلاعات جریان جمعیتی هستند نه اطلاعات قابل شناسایی تک تک افراد. اما داستان‌های بسیاری در رسانه‌ها به نقض این حریم و اصل عمومی پرداخته‌اند. هم کاربران و هم توسعه‌دهندگان اینگونه اپلیکیشن‌ها باید از تأثیر مخرب انتشار اطلاعات خصوصی و همچنین گردآوری اینگونه اطلاعات مطلع و آگاه باشند که در فصل ۱۴ بیشتر در این خصوص صحبت خواهیم کرد.

سوالاتی برای مرور بخش ۹.۱۰

۱. تحلیل‌های رایج و قدیمی چگونه از داده‌های مکان‌محور استفاده می‌کنند؟
۲. مکان‌های کدگذاری شده‌ی جغرافیایی چگونه به تصمیم‌گیری بهتر کمک می‌کنند؟
۳. تحلیل‌های مکانی-جغرافیایی چه ارزشی ارائه می‌کنند؟
۴. کاربردهای دیگر تحلیل‌های مکانی-جغرافیایی در سایر بخش‌ها از جمله سرشماری دولت، بازاریابی مصرف‌کننده و غیره را نام ببرید.
۵. برای کاربردهای دیگر اپلیکیشن‌های تحلیلی مشتری‌محور در اینترنت جستجو کنید.
۶. تحلیل‌های مکان‌محور چگونه می‌توانند به مشتریان و مصرف‌کنندگان کمک کنند؟
۷. اپلیکیشن‌های حمل و نقل دیگری که ممکن است از تحلیل‌های مبتنی بر مکان استفاده کنند را نام ببرید.
۸. اگر می‌توانستید به داده‌های مکانی تلفن‌های سلولی دسترسی داشته باشید چه کاربردهای دیگری را تصور می‌کردید؟

የገቢዎች ስም	የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት
	የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት
	የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት
	የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት
	የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት

የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	127.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	128.
የገቢው አይነት			
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	129.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	130.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	131.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	132.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	133.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	134.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	135.
የገቢው አይነት			
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	136.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	137.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	138.
የገቢው አይነት			
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	139.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	140.
የገቢው አይነት			
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	141.
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	142.
የገቢው አይነት			
የገቢው አይነት	የገቢው መጠን	የገቢው ዓይነት	143.

	teradatauniversitynetwork.com	.۲۶
	teradatauniversitynetwork.com	.۲۷
Sports Analytics	teradatauniversitynetwork.com	.۲۸
	teradatauniversitynetwork.com	.۲۹
	asterdata.com teradata.com	.۳۰
	IBM.com	.۳۱
	claudera.com	.۳۲
	mapr.com	.۳۳
	hortonworks.com	.۳۴
	marklogic.com	.۳۵
	youtube.com	.۳۶
	google.com/scholar	.۳۷
	google.com/scholar	.۳۸
	google.com/scholar	.۳۹
		.۴۰

Youtube.com	.۴۱
pandora.com	.۴۲
Humanyze.com	.۴۳
	.۴۴
archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones	

- کاربرد رباتیک در صنایع مختلف
- رباتیک در پزشکی
- رباتیک در کشاورزی
- رباتیک در خدمات
- رباتیک در حمل و نقل
- رباتیک در صنایع دفاعی
- رباتیک در صنایع معدنی
- رباتیک در صنایع نساجی
- رباتیک در صنایع غذایی
- رباتیک در صنایع پتروشیمی
- رباتیک در صنایع فولاد
- رباتیک در صنایع کاشی و سرامیک
- رباتیک در صنایع نساجی
- رباتیک در صنایع غذایی
- رباتیک در صنایع پتروشیمی
- رباتیک در صنایع فولاد
- رباتیک در صنایع کاشی و سرامیک

۱۰

رُباتیک: کاربردهای صنعتی و مصرفی

در فصل ۲ به اختصار به معرفی رباتیک که یکی از کاربردهای اولیه و عملی مفاهیم توسعه یافته در هوش مصنوعی بود پرداختیم. در این فصل، ما به معرفی چندین کاربرد ربات‌ها در محیط‌های صنعتی و همچنین شخصی خواهیم پرداخت. علاوه بر یادگیری در مورد کاربردهای اجرا شده و نوظهور ما به شناسایی مؤلفه‌های عمومی یک ربات می‌پردازیم. از منظر ملاحظات مدیریتی نیز ما در مورد تأثیرات رباتیک بر مشاغل و همچنین مسائل قانونی آنها بحث خواهیم کرد. مباحثی که در این فصل مطرح می‌شود گسترده بوده و تمامی حوزه‌های دیگر هوش مصنوعی را تحت تأثیر قرار می‌دهد و به همین دلیل ممکن است با مباحث فصل ۱۴ کمی همپوشانی داشته باشد. اما تمرکز ما در این فصل بر ربات‌های فیزیکی و نه فقط کاربردهای نرم‌افزاری هوش مصنوعی است.

این فصل از بخش‌های زیر تشکیل شده است:

۱۰.۱. پرونده‌ی آغازین: حمایت عاطفی از بیماران و کودکان توسط ربات‌ها

۱۰.۲. مروری بر رباتیک

۱۰.۳. تاریخچه‌ی رباتیک

۱۰.۴. کاربردهای بارز رباتیک

۱۰.۵. مؤلفه‌های رباتیک

۱۰.۶. دسته‌بندی مختلف ربات‌ها

۱۰.۷. خودروهای مستقل: ربات‌ها در حرکت

۱۰.۸. تأثیر ربات‌ها بر مشاغل فعلی و آینده

۱۰.۹. ملاحظات قانونی ربات‌ها و هوش مصنوعی

همانطور که در این فصل شرح داده شد، ربات‌ها بر ساخت و تولید صنعتی و سایر فعالیت‌های فیزیکی تأثیر گذاشته‌اند. امروزه، با پژوهش و تکامل هوش مصنوعی، رباتیک می‌تواند بر دنیای اجتماعی نیز سوار شود. به عنوان مثال، بیمارستان‌ها امروزه تلاش می‌کنند تا حمایت‌های عاطفی و اجتماعی از بیماران و خانواده‌هایشان به عمل آورند. این حمایت به ویژه در زمان درمان کودکان حساسیت پیدا می‌کند. کودکان در بیمارستان در یک محیط ناآشنا هستند که دستگاه‌های پزشکی به آنها متصل است و در بسیاری از موارد پزشکان ممکن است محدودیت‌های حرکتی برایشان توصیه کنند. این محدودیت منجر به استرس، اضطراب و افسردگی کودکان و در نتیجه اعضای خانواده آنها می‌شود. بیمارستان‌ها سعی دارند تا برای کاهش آسیب‌ها، متخصصان حمایت از کودکان یا تراپی‌های حیوانات خانگی همراه را ارائه دهند. این تراپی‌ها کودکان و والدین آنها را برای کاهش آسیب‌ها، متخصصان حمایت کرده و یک حمایت عاطفی موقت را برایشان به همراه دارد. با توجه به کم بودن تعداد این متخصصان، بین عرضه و تقاضای متخصصان درمان کودکان یک شکاف وجود دارد. همچنین در بسیاری از مراکز به دلیل ترس از حساسیت، گرد و غبار و گاز گرفتگی که ممکن است موجب وخیم‌تر شدن شرایط بیماران شود امکان ارائه‌ی تراپی برای حیوانات خانگی وجود ندارد. برای پر کردن این شکاف، استفاده از ربات‌های اجتماعی برای حل مشکل افسردگی و اضطراب بین کودکان در حال بررسی است. یک مطالعه نشان داد که (جئونگ و همکاران ۲۰۱۵) حضور فیزیکی یک ربات برای واکنش‌های عاطفی به مراتب مؤثرتر از تعامل با یک ماشین مجازی در یک مرکز بیمارستانی امراض کودکان بوده است. پژوهشگران برای مدت زمان طولانی می‌دانستند (مثلاً گوریس و همکاران ۲۰۱۰) که بیش از ۶۰ درصد از ارتباطات انسانی به صورت شفاهی نیست بلکه از طریق حالات چهره اتفاق می‌افتد. بنابراین، یک ربات اجتماعی باید بتواند ارتباطات عاطفی را همانند یک متخصصان کودکان از خود بروز دهد. یک ربات معروف در این حوزه، «هاگیل^۱» نام دارد. به کمک هوش مصنوعی، هاگیل قادر به درک ابراز حالات چهره، خلق و خوی، ژست و هوش انسانی است. این ربات به صورت عضوی از پرسنل متخصصان بیمارستان است که می‌تواند کمک‌های عمومی سلامت عاطفی را به کودکان ارائه دهد.

¹ Huggable

هاگیل شبیه یک عروسک خرسی است و بدن نرم پشمالوی آن ظاهری کودکانه و حس دوستانه‌ای به کودکان می‌دهد. هاگیل به کمک بازوهای مکانیکی خود می‌تواند کارهای خاصی را به سرعت انجام دهد. یک ربات هاگیل به جای دستگاه‌های با فناوری پیشرفته از یک دستگاه اندروئید تشکیل شده است که میکروفون، بلندگو و دوربین آن در حسگرهای داخلی آن تعبیه شده‌اند و یک تلفن سیار به عنوان یک سیستم عصبی مرکزی عمل می‌کند. این دستگاه اندروئید امکان برقراری ارتباط بین حسگرهای داخلی و رابط تله‌اپراتوری را فراهم می‌کند. مؤلفه‌های بازوی بخش‌بندی شده‌ی آن امکان جابجایی آسان حسگرها را فراهم کرده و امکان استفاده‌ی مجدد از آن را افزایش می‌دهد. این حسگرهای لامسه‌ای به همراه هوش مصنوعی ربات را قادر می‌سازند تا تماس‌های فیزیکی را پردازش کرده و آنها را با ابراز حالت‌ها بکار بگیرد. حسگرهای بکار گرفته شده در هاگیل، داده‌های فشار و تماس فیزیکی را از طریق یک بُرد یو یو^۱ به دستگاه تله‌اپراسیون یا دستگاه خارجی منتقل می‌کنند. دستگاه اندروئید داده‌ها را از حسگرهای خارجی دریافت کرده و آنها را به موتورهایی که به بدنه‌ی ربات متصل هستند ارسال می‌کند. این موتورها به حرکت و جابجایی ربات کمک می‌کنند. خازن‌ها در بخش‌های مختلف ربات تعبیه شده‌اند که آنها را نقاط فشار می‌نامند. این نقاط فشار به ربات کمک می‌کنند تا کودکی که قادر نیست تا درد خود را شفاهاً ابراز کند ولی می‌تواند با لمس کردن ربات درد خود را ابراز کند را درک کند. دستگاه اندروئید این داده‌های حسگر فشار و تماس فیزیکی را به نحوی قابل درک تفسیر کرده و به آن پاسخ می‌دهد. تلفن اندروئید امکان برقراری ارتباط بین سایر دستگاه‌ها را با حفظ یک طراحی مینیمالیست فراهم می‌کند. قدرت پردازش ربات و دستگاه اندروئید به اندازه‌ای خوب است که بتواند به صورت بلادرنگ با کودک ارتباط برقرار کند. شکل ۱۰.۱ یک نمای شماتیک از ربات هاگیل را نشان می‌دهد. از هاگیل برای کودکان تحت درمان در بیمارستان کودکان بوستون استفاده شده است. طبق گزارشات، اورورا کودک ۱۰ ساله‌ای که مبتلا به سرطان خون بوده در مرکز اختلالات خونی و سرطان کودکان بوستون/دانا فاربر تحت درمان بوده است. طبق اظهارات والدین اورورا، «کارهای بسیاری باید در بیمارستان انجام می‌شد ولی حضور هاگیل برای کودکان خیلی خوب بود. بئاتریس یکی دیگر از کودکانی که به خاطر شرایط مزمن مکرراً به بیمارستان مراجعه می‌کرد و از کلاس‌ها و دوستانش دور افتاده بود و نمی‌توانست هیچکدام از کارهای معمولی که هم سن و سال‌هایش انجام می‌دادند را انجام دهد. او بسیار عصبی بود و از پروسه‌ی درمان بیزار بود ولی در تعاملاتی که با هاگیل داشت تمایل بیشتری به مصرف داروهایش داشت گویی طبیعی‌ترین کاری بود که می‌توانست انجام دهد. او به ربات توصیه کرده بود که کمی سریع‌تر کار کند تا دفعه‌ی بعد بتوانند بازی دالی موشه را درست اجرا کنند.

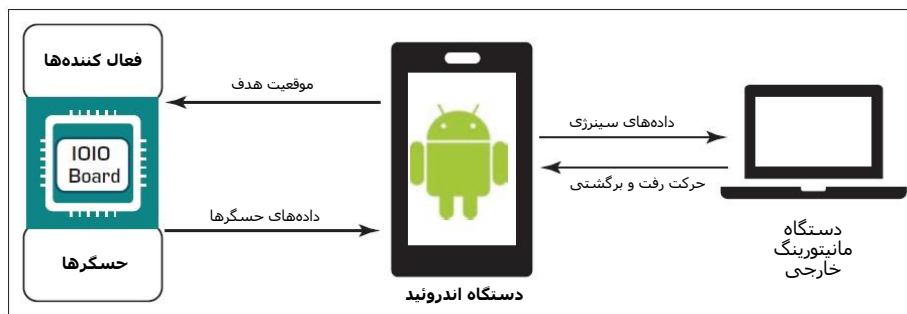
در طی این تعاملات با هاگیل، کودکان او را در آغوش می‌گرفتند، دست او را می‌گرفتند و مثل کسی که به حمایتش نیاز دارند با او رفتار می‌کردند. کودکان رفتار مؤدبانه‌ای با او داشتند و اصطلاحاتی همچون «نه، متشکرم» و «یه لحظه لطفاً» را بکار می‌بردند. در پایان، زمانیکه وقت خداحافظی می‌رسید یکی از کودکان هاگیل را در آغوش کشید و دیگر آرزو کرد که کاش بیشتر با هم بازی می‌کردند.

یکی دیگر از مزایای اینگونه ربات‌های حامی عاطفی، پیشگیری از عفونت‌ها است. بیماران ممکن است به بیماری‌های مسری مبتلا باشند ولی ربات‌ها پس از بار استفاده برای پیشگیری از گسترش عفونت، استریلیزه می‌شوند. بنابراین، هاگیل فقط برای ارائه‌ی حمایت‌های عاطفی به کودکان استفاده نمی‌شود بلکه می‌تواند ابزار مفیدی برای کاهش شیوع و گسترش بیماری‌های عفونی نیز باشد.

بر اساس مطالعه‌ی جدیدی که توسط پژوهشگران آزمایشگاه رسانه‌ی ام‌آی‌تی گزارش شد، تفاوت‌های بین ربات‌های اجتماعی مثل هاگیل و سایر فناوری‌های تعامل مجازی مورد بررسی قرار گرفت. یک گروه متشکل از ۵۴ کودک که در بیمارستان بودند

¹ IOIO Board

با سه تعامل اجتماعی متمایز آزموده شدند: یک خرس عروسکی معمولی، یک شخصیت مجازی از هاگیل بر روی تبلت و یک ربات اجتماعی. این خرس عروسکی فقط حالت فیزیکی داشت ولی فاقد هرگونه دیالوگ و گفتار اجتماعی بود. یک نسخه‌ی مجازی از هاگیل بر روی تبلت که تعاملات زبانشناسی و صحبت با انسان‌ها را دارا بود و همان ویژگی‌های ربات‌ها را داشت ولی یک نسخه‌ی مجازی سه‌بعدی از ربات هاگیل بود. هم شخصیت مجازی و هم ربات توسط یک تله‌پراتور عمل می‌کردند و به همین دلیل نحوه‌ی تعامل و پاسخگویی آنها یکسان بود. کودکان بر اساس سن و جنسیت به صورت گروه‌بندی شده می‌توانستند با یکی از این سه شیوه‌ی تعاملی کار کنند. متخصصان مراقبت از کودکان اطلاعات لازم را در اختیار کودکان قرار می‌دادند و شخصیت مجازی و ربات از خارج از اتاق به صورت جداگانه این ربات و شخصیت مجازی را کنترل می‌کردند. تحلیلگر لحن آی‌بی‌ام واتسون برای شناسایی پنج حالت عاطفی انسانی و پنج ویژگی شخصیتی بکار گرفته شد. تعاملات بین هر یک از این سه نوع عامل مجازی و کودکان توسط پژوهشگران فیلمبرداری و تحلیل شد. این نتایج نشان داد که کودکان بیشتر تحت تأثیر شخصیت مجازی و ربات قرار گرفته‌اند تا عروسک خرسی. بیشترین تماس‌های فیزیکی بین کودکان و عوامل مجازی ابتدا با ربات هاگیل و پس از آن با هاگیل شخصیت مجازی و نهایتاً خرس عروسکی بوده است. همچنین، کودکان از ربات هاگیل مراقبت می‌کردند و آن را هل نمی‌دادند یا به طرف خود نمی‌کشیدند. جالب اینکه، حتی زمانیکه شخصیت مجازی بر روی تبلت صداهای حاکی از درد و ناراحتی بروز می‌داد تعدادی کمی از کودکان بودند که با خشونت به آن پاسخ می‌دادند. اما عروسک خرسی بیچاره را به این طرف و آن طرف پرتاب می‌کردند. این نتایج نشان داد که کودکان بیش از دو گزینه‌ی دیگر با ربات هاگیل ارتباط برقرار می‌کردند.



شکل ۱۰.۱ یک نمای شماتیک عمومی از ربات هاگیل

تا سال ۲۰۵۰ نرخ جمعیت ۶۵ سال و بالاتر کشورهای مهم جهان به زودی بیش از جمعیت جوان خواهد شد. نیاز حمایت عاطفی از افراد مسن را نمی‌توان نادیده گرفت مخصوصاً در جاییکه پراکندگی جغرافیایی و شکاف تکنولوژیکی برقراری ارتباط با اعضای خانواده‌هایشان را دشوار کرده است. پارو یک ربات اجتماعی است که برای تعامل با انسان‌ها طراحی شده و به عنوان رباتی برای افراد مسن در خانه‌ی سالمندان استفاده می‌شود. پارو در اصل به عنوان جایگزین حیوان درمانی می‌تواند استفاده شود مخصوصاً در بیمارستان‌ها که ریسک عفونت بسیار بالاست.

پارو قادر به تفسیر حس لامسه‌ی انسانی بوده و می‌تواند گفتار را تا حد محدودی ضبط کرده و مجموعه‌ی محدودی از سخنان شفاهی را بیان کند و سرش را تکان بدهد. پارو یک ربات سیار نبوده و شبیه فک دریایی است. پارو در دو خانه سالمندان با ۲۳ بیمار تست شد. نتایج نشان داد که ربات‌های اجتماعی همانند پارو تعامل اجتماعی را افزایش می‌دهند. پارو نه تنها خنده را به

لب‌های بیماران آورد بلکه تجربیات بسیار پرنشاط و شادابی را برای بیماران به همراه داشت. اگرچه پارو همانند انسان‌ها قادر به پاسخگویی کامل نبود ولی پاسخ‌هایی که می‌داد برای بسیاری از بیماران معنادار بوده و با او ارتباط عاطفی برقرار می‌کردند. این ربات‌ها می‌توانند روتین‌های یکنواخت و خسته‌کننده‌ی سالمندان را شکسته و شادی و نشاط را به زندگی‌شان بازگردانند. این ربات‌ها به آنها احساس خواسته شدن و عزت نفس داده و میزان استرس و اضطراب را کاهش می‌دهند.

سوالاتی برای مرور پرونده‌ی آغازین

۴۲. از یک ربات که بتواند حمایت عاطفی به بیماران بدهد چه ویژگی‌هایی را انتظار دارید؟

۴۳. آیا می‌توانید کاربردهای دیگری از ربات‌هایی مثل هاگبل را نام ببرید؟

۴۴. از وب سایت <https://www.universal-robots.com/case-stories/aurolab> بازدید کرده و در مورد ربات‌های

مشارکت‌جو تحقیق کنید. این ربات‌ها چگونه می‌توانند در محیط‌ها و شرایط دیگر مفید باشند؟

همانطور که در فصل‌های مختلف این کتاب دیدیم، هوش مصنوعی کاربردهای بسیار جالب و منحصر‌بفردی را بوجود آورده است. داستان‌هایی درباره‌ی هاگبل و پارو ایده‌ی استفاده از ربات‌ها برای یکی از دشوارترین جنبه‌های کاری را به ما معرفی کرده یعنی حمایت عاطفی هم از کودکان و هم بزرگسالان. ترکیب فناوری‌هایی از قبیل یادگیری ماشین، تلفیق صدا، تشخیص صدا، پردازش زبان طبیعی، بینایی ماشین، اتوماسیون و غیره ترکیب این فناوری‌ها برای برآورده کردن بسیاری از نیازها را امکان‌پذیر ساخته است. این کاربردها می‌توانند به صورت کاملاً مجازی مثل آی‌بی‌ام واتسون که برنده‌ی مسابقه‌ی جئوپاردی! شد باشند یا پیاده‌سازی اتوماسیون صنعتی، تولید خودروهای خودران و حتی حمایت‌های عاطفی باشند. ما در این فصل مثال‌هایی مشابه بیشتری را خواهیم دید.

Sources: J. Broekens, M. Heerink, & H. Rosendal. (2009). "Assistive Social Robots in Elderly Care: A Review." *Gerontechnology*, 8, pp. 94–103. doi: 10.4017/gt.2009.08.02.002.00; S. Fallon. (2015). "A Blue Robotic Bear to Make Sick Kids Feel Less Blue." <https://www.wired.com/2015/03/blue-robotic-bear-make-sick-kids-feel-less-blue/> (accessed August 2018). Also see the YouTube video at <https://youtu.be/UaRCCA2rRR0> (accessed August 2018); K. Goris et al. (2010, September). "Mechanical Design of the Huggable Robot Probo." Robotics & Multibody Mechanics Research Group. Brussels, Belgium: Vrije Universiteit Brussel; S. Jeong et al. (2015). "A Social Robot to Mitigate Stress, Anxiety, and Pain in Hospital Pediatric Care." *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*; S. Jeong & D. Logan. (2018, April 21–26). "Huggable: The Impact of Embodiment on Promoting Socio-emotional Interactions for Young Pediatric Surgeons." MIT Media Lab, Cambridge, MA, CHI 2018, Montréal, Quebec, Canada.

هر دانشمند علم رباتیک دیدگاه خودش را از تعریف ربات دارد. ولی یک مفهوم مشترک از ربات یک ماشین یک دستگاه فیزیکی یا نرم‌افزار است که به کمک هوش مصنوعی قادر است یک مسئولیت را به طور مستقل انجام دهد. یک ربات می‌تواند محیط را حس کرده و بر آن تأثیر بگذارد. کاربردهای رباتیک در زندگی روزمره ما در حال افزایش هستند. این تکامل و بکارگیری فناوری‌ها را انقلاب صنعتی چهارم می‌نامند.

کاربردهای رباتیک در حوزه‌های ساخت و تولید، سلامت و فناوری اطلاعات در دهه‌ی گذشته منجر به رشد سریع تغییرات در آینده‌ی صنایع شده است. ربات‌ها صرفاً از انجام وظایف تکراری از پیش تعریف شده (اتوماسیون) فراتر رفته و قادر به واکنش نسبت به شرایط پیش‌بینی نشده برای انجام وظایف تخصصی در مراقبت سلامت، تولید، ورزش، خدمات مالی و تقریباً هر صنعتی هستند. این قابلیت‌سازی با موقعیت‌های جدید منجر به یک استقلال شده که در مقایسه با ربات‌های نسل قبل یک تغییر

عظیم محسوب می‌شود. در فصل ۲ به معرفی تعریفی از ربات‌ها پرداختیم و کاربردهایی از آنها را در برخی صنایع ذکر کردیم. در این فصل، با اشاره به کاربردهای مختلف و بررسی عمیق‌تر موضوع به تکمیل این معرفی خواهیم پرداخت. اگرچه تصور ما از یک ربات ممکن است آرتودی‌تو یا سی‌تری‌پیو از فیلم‌های جنگ ستارگان باشد ولی ربات‌ها را به روش‌های دیگری نیز تجربه کرده‌ایم. کارخانجات برای مدت زمان طولانی از ربات‌ها برای تولید و ساخت استفاده کرده‌اند. در بخش مصرف، رومبا یکی از کاربردهای اولیه‌ی ربات‌ها بود که می‌توانست کف زمین را تمیز کند. شاید بهترین مثال از ربات‌هایی که به زودی تجربه خواهیم کرد خودروهای خودران باشد. تک‌ریپابلیک خودروهای خودران را اولین ربات‌هایی نامیده که به زودی به آنها اعتماد خواهیم کرد. ما در بخش ۱۰.۷ نگاه عمیق‌تری به وسایل نقلیه‌ی خودران خواهیم داشت. با رشد یادگیری ماشین و به ویژه سیستم‌های تشخیص تصویر، کاربردهای ربات‌ها تقریباً در تمامی صنایع در حال افزایش است. ربات‌ها می‌توانند سوسیس‌ها را به اندازه‌های مناسب برای پیتزا برش داده و تعیین کنند که تعداد و نوع مناسبی از تکه‌های پپرونی پیش از پخت بر روی پیتزا قرار گرفته باشد. انجام جراحی‌ها به کمک ربات‌ها به سرعت در حال افزایش است. بخش ۱۰.۴ کاربردهای بسیاری از ربات‌ها را ارائه می‌کند.

سوالاتی برای مرور بخش ۱۰.۲

۱. ربات را تعریف کنید.
۲. تفاوت بین اتوماسیون و مستقل بودن چیست؟
۳. مثال‌هایی از ربات‌ها در حال استفاده بزنید. آخرین کاربردهای آنلاین را یافته و آنها را با کلاس به اشتراک بگذارید.

ویکیپدیا شامل تاریخچه‌ی جالبی از رباتیک است. انسان‌ها برای مدت زمان مدیدی مجذوب ایده‌ی ماشین‌هایی بوده‌اند که به آنها خدمت بدهند. اولین ایده‌ی رباتیک در ۳۲۰ قبل از میلاد و زمانی مطرح شد که ارسطو فیلسوف یونانی گفت «اگر هر ابزاری طبق دستور یا حتی طبق میل خودش بتواند کاری را انجام دهد آنگاه استادکارها نیازی به کارآموز و اربابان نیازی به برده‌ها نخواهند داشت.» در سال ۱۴۹۵ لئوناردو داوینچی تصاویر و استراتژی‌هایی از یک ربات را که شبیه انسان بود ارائه کرد. بین سال‌های ۱۷۰۰ و ۱۹۰۰ اتوماسیون‌های مختلفی خلق شدند از جمله یک ساختار اتوماسیون عالی که توسط ژاک دو ووکانسون ساخته شد که یک اردک کوچکی بود که می‌توانست بال بزند و به نظر می‌رسید که می‌تواند غذا بخورد و آن را هضم کند. در سرتاسر انقلاب صنعتی، رباتیک با محرک‌هایی از قبیل پیشرفت‌های برق و نیروی بخار همراه بود. با افزایش تقاضای مصرف، مهندسان سعی کردند تا روش‌های جدیدی برای افزایش تولید از طریق اتوماسیون و خلق ماشین‌هایی که بتوانند وظایفی که انجامشان برای انسان‌ها خطرناک بودند را عرضه کنند. در سال ۱۸۹۳، «مرد بخاری» که یک نمونه از ربات‌های انسان‌نما بود توسط پروفیسور کانادایی جورج مور پیشنهاد شد. این ربات از جنس فولاد بود و قدرت خود را از یک موتور بخار می‌گرفت و می‌توانست با سرعتی نزدیک به ۹ مایل در ساعت به طور مستقل حرکت کند و حتی بارهای سنگین را با خود بکشد. در سال ۱۸۹۸ نیکولا تسلا یک نمونه‌ی زیردریایی را ارائه کرد. این رویدادها منجر به یکپارچگی رباتیک با صنایع تولید، فضا، دفاع، هوافضا، پزشکی، آموزش و سرگرمی شده است.

در سال ۱۹۱۳، اولین خط مونتاژ تسمه نقاله‌ای متحرک جهان توسط هنری فورد شروع به کار کرد. به کمک این تسمه نقاله، یک خودرو در عرض ۹۳ دقیقه مونتاژ میشد. بعدها در ۱۹۲۰، عبارت ربات توسط کارل گِپک در نمایشنامه‌ی کارخانه‌ی ربات‌سازی روسوم برای اولین بار بکار گرفته شد. سپس یک ربات عروسکی در ژاپن ساخته شد.

در دهه ۵۰، نوآوران ماشین‌هایی را تولید کردند که قادر به انجام وظایف خطرناک و تکراری برای ساخت و تولید صنعتی و صنایع دفاع بودند. از آنجائیکه ربات‌ها اصولاً برای صنایع سنگین طراحی شده بودند، نیاز بود تا وظایفی مثل کشیدن، بلند کردن، جابجا کردن و هل دادن را همانند انسان‌ها انجام دهند. بنابراین، ربات‌های بسیاری همانند بازوی انسان طراحی شدند. مثال‌هایی از آنها شامل یک گجت رنگ‌پاش برای یک آپاراتوس کنترل موقعیت توسط دابلویو. ال. وی پولارد در سال ۱۹۳۸ بود. شرکت دو ویلیس این ربات را خریداری کرده و سپس به تأمین‌کننده‌ی اصلی بازوهای رباتیک در ایالات متحده تبدیل شد.

در اواسط دهه ۵۰، اولین بازوی رباتیک تجاری به نام پلنِت‌بات توسعه یافته و جنرال موتورز بعدها از آن در یک کارخانه تولیدی برای تولید رادیاتورها استفاده کرد. مجموعاً ۸ پلنِت‌بات فروخته شد. طبق اظهارات این شرکت، این ربات می‌توانست تقریباً ۲۵ حرکت را انجام داده و در عرض چند دقیقه برای انجام مجموعه دیگری از عملیات تنظیم شود. با وجود این، پلنِت‌بات به دلیل رفتار نامعمول سوخت هیدرولیک درونش نتوانست نتایج مطلوبی ارائه کند.

جورج دوول و جو اینگل برگر ربات یونی‌میت را برای اتوماسیون ساخت و تولید لامپ‌های تصویر تلویزیون‌ها طراحی کردند. این ربات حدود ۴،۰۰۰ پوند وزن داشت و توسط مجموعه دستورات برنامه‌ریزی شده بر روی یک استوانه‌ی مغناطیسی کنترل میشد. بعدها از این ربات در شرکت جنرال موتورز برای توالی و انباشت تولید قطعات فلزی قالب‌ریزی شده استفاده شد. این بازو به همراه ارتقاءهای خاص به یکی از مهمترین ویژگی‌های خطوط مونتاژ تبدیل شد. مجموعاً ۸،۵۰۰ ماشین فروخته شد و نیمی از آنها در صنایع خودروسازی بکار گرفته شدند. بعدها یونی‌میت به گونه‌ای ارتقاء یافت تا قادر به انجام جوشکاری نقطه‌ای، قالب‌ریزی و انباشت ابزارهای ماشین‌آلات باشد.

در دهه ۶۰، رالف موشر و تیمش دو بازوی رباتیک که جدا از یکدیگر کار می‌کردند را به نام‌های هندی من و من‌میت طراحی کردند. هندی‌من یک ربات الکتروهیدروییکی دو بازویی بود و طراحی بازوی من‌میت بر اساس ستون فقرات انسان انجام شده بود. بازوها انعطاف‌پذیری برای انجام رویه‌های بازرسی قطعات هواپیماها را به ربات‌ها می‌دادند. انگشتان به گونه‌ای طراحی شده بودند که بتوانند اشیاء را تنها با یک فرمان در دست بگیرند.

ربات‌های سیار جدید وارد صحنه شدند. اولین ربات از این نوع شیکی نام داشت که در سال ۱۹۶۳ طراحی شد. این ربات می‌توانست آزادانه حرکت کرده و از برخورد با موانع سر راهش اجتناب کند. یک آنتن رادیویی به دو سر آن متصل بود و یک حسگر دیداری بر روی واحد پردازش مرکزی آن قرار داشت. شیکی به دو چرخ متصل بود و دو حسگرش می‌توانستند موانع را حس کنند. با استفاده از حل مسئله‌ی منطقی این ربات می‌توانست شکل اشیاء را تشخیص داده، آنها را جابجا کرده و یا در اطراف آنها حرکت کند.

عصر رقابت‌های فضایی با ارسال اسپوتنیک روسیه آغاز شده و ایالات متحده پیشرفت‌های فناوری بسیاری را که منجر به رشد رباتیک شد بوجود آورد. در ۱۹۷۶ طی مأموریت ناسا به مریخ یک مریخ‌نورد به نام وایکینگ برای شرایط جوّی مریخ طراحی شد. بازوی وایکینگ باز شده و یک لوله را برای گردآوری نمونه‌ها از سطح مریخ تولید می‌کرد. یک سری مشکلات فنی در طی مأموریت پیش آمد ولی دانشمندان نتوانستند آنها را از راه دور رفع کنند.

در سال ۱۹۸۶ اولین محصولات آموزشی مبتنی بر لگو توسط شرکت هوندا وارد بازار شدند. در سال ۱۹۹۴، دانته ۲ که یک ربات متحرک هشت پا بود توسط دانشگاه کارنجی ملون ساخته شد که نمونه گازهای آتشفشانی را از قله‌ی اسپور جمع‌آوری کرد.

رباتیک با افزایش پژوهش و سرمایه‌گذاری‌ها به طور نمایی گسترش یافت. پژوهش و کاربردهای رباتیک به ژاپن، کره و کشورهای اروپایی تسری یافت. برآورد شده است که تا سال ۲۰۱۹ نزدیک به ۲.۶ میلیون ربات مهم در جهان وجود داشته باشد. ربات‌ها دارای کاربردهای گسترده‌ای از رشته‌های حمایت اجتماعی، صنایع دفاع، سرگرمی و اسباب‌بازی سازی، مراقبت سلامت، صنایع غذایی و ... دارند. ربات‌های بسیاری در حال حاضر وارد مراحل بعدی شده‌اند و به پژوهش در عمق دریاها تا بین ستاره‌ای و فراخورشیدی می‌پردازند. و همانطور که اشاره شد، خودروهای خودران به انبوه‌سازی ربات‌ها کمک می‌کنند.

سوالاتی برای مرور بخش ۱۰.۳

۱. برخی از نقاط عطف مهم در تاریخچه‌ی ساخت و تولید که منجر به توجه و علاقه‌ی جاری به رباتیک شده است را نام ببرید.
۲. قابلیت‌های شبکی را چگونه می‌توانید با ربات‌های امروزی مقایسه کنید؟
۳. ربات‌ها چگونه به مأموریت‌های فضایی کمک کردند؟

در این بخش به بررسی مثال‌هایی از کاربردهای ربات‌ها در صنایع مختلف می‌پردازیم. هر یک از این کاربردها در قالب یک مورد کاربردی کوچک به همراه سؤالات بحث و گفتگو در انتهای بخش ارائه شده است.

یک شرکت تولید موبایل در چین به نام چینجینگ پرسیژن تکنولوژی روی به استفاده از بازوهای رباتیک برای تولید قطعات تلفن‌های سیار نمود. این شرکت قبلاً از ۶۵۰ کارگر برای انجام فعالیت‌های کارخانه استفاده می‌کرد. در حال حاضر ربات‌ها اکثر عملیات آن را انجام داده و شرکت نیروی کار خود را به ۶۰ نفر یعنی به میزان ۹۰٪ کاهش داده است. در آینده، شرکت قصد دارد تا تعداد کارکنان خود را به ۲۰ نفر برساند. به کمک ربات‌ها، شرکت نه تنها به افزایش تولید ۲۵۰ درصدی دست یافت بلکه توانست میزان ضایعات خود را نیز از ۲۵ درصد به ۵ درصد کاهش دهد.

Compiled from C. Forrest. (2015). "Chinese Factory Replaces 90% of Humans with Robots, Production Soars." TechRepublic. <https://www.techrepublic.com/article/chinese-factory-replaces-90-of-humans-with-robotsproduction-soars/> (accessed September 2018); J. Javelosa & K. Houser. (2017). "Production Soars for Chinese Factory Who Replaced 90% of Employees with Robots." Future Society. <https://futurism.com/2-productionssoars-for-chinese-factory-who-replaced-90-of-employees-with-robots/> (accessed September 2018).

آدیداس یک تولید کننده سرشناس پوشاک ورزشی در سرتاسر جهان است. با در نظر گرفتن روندها، نوآوری و سفارشی‌سازی شرکت آدیداس شروع به اتوماسیون کارخانجاتی از قبیل اسپیدفکتوری در آنسباخ آلمان و آتلانتا گرجستان نمود. یک زنجیره‌ی تأمین سنتی از تأمین مواد اولیه تا محصول نهایی حدود ۲ ماه زمان نیاز داشت ولی به کمک این اتوماسیون این مدت زمان به چند روز یا هفته کاهش یافت. پیاده‌سازی رباتیک در آنجا کمی با سایر صنایع تولیدی متفاوت بود زیرا مواد اولیه‌ی مورد استفاده در ساخت و تولید کفش آدیداس از مواد نساجی نرم بودند. آدیداس با شرکت آکسلر برای پیاده‌سازی رباتیک در زنجیره تأمین

¹ Changing Precision Technology

² Adidas

خود کار می‌کند. آدیداس از فناوری‌هایی از قبیل بازوهای رباتیک و بافندگی رایانه‌ای بهره می‌برد. در اسپیدفکتوری رباتی که بخشی از یک کتانی را تولید می‌کند یک کد کیوآر قابل اسکن را بر روی آن بخش الصاق می‌کند. در طی فرایند کنترل کیفیت، اگر هر بخشی از کتانی معیوب تشخیص داده شود، رباتی که آن را تولید کرده قابل ردیابی و تعمیر خواهد بود. آدیداس این فرایند را بهینه‌سازی کرده که امکان عرضه‌ی چندین هزار کفش سفارشی به بازار و مشاهده‌ی عملکرد و بهینه‌سازی فرایند بر آن اساس را میسر می‌کند. در چند سال آینده، شرکت قصد دارد تا حدود ۱ میلیون جفت کفش با سبک‌های سفارشی را به صورت سالانه روانه‌ی بازار کند. در بلند مدت، این استراتژی از تولید محصولات بر حسب تقاضا به جای تولید حجم زیادی از موجودی محصولات پشتیبانی خواهد کرد.

Compiled from "Adidas's High-Tech Factory Brings Production Back to Germany." (2017, January 14). *The Economist*.

<https://www.economist.com/business/2017/01/14/adidas-high-tech-factory-brings-production-back-to-germany> (accessed September 2018); D.

Green. (2018). "Adidas Just Opened a Futuristic New Factory – and It Will Dramatically Change How Shoes Are Sold." *Business Insider*.

<http://www.businessinsider.com/adidas-hightech-speedfactory-begins-production-2018-4> (accessed September 2018).

کاربرد روزافزون هوش مصنوعی و اتوماسیون در صنایع منجر به توسعه‌ی ربات‌ها شده است. اما توانایی‌های شناختی انسان‌ها غیرقابل جایگزینی است. ترکیب ربات‌ها و انسان‌ها منجر به استفاده از ربات‌های مشارکتی در یک واحد تولیدی بی‌ام‌دبلیو شده است. با اینکار، شرکت توانسته است کارایی واحد تولیدی خود را به حداکثر رسانده و محیط کاری مدرن و پیشرفته‌ای را بوجود آورد.

کارخانه‌ی اسپارتان‌بورگ بی‌ام‌دبلیو در کارولینای جنوبی از ۶۰ ربات مشارکتی که در کنار انسان‌ها کار می‌کردند بهره می‌برد. این ربات‌ها به پرداخت کاری داخل درب‌های خودروی بی‌ام‌دبلیو و عایق کاری رطوبتی می‌پردازند. این درزبندی‌ها از تجهیزات الکترونیکی تعبیه شده در داخل درب‌ها و کل خودرو محافظت می‌کند. تا قبل از این، کارگرهای انسانی این وظیفه‌ی سنگین چسباندن فویل با چسب با استفاده از رولرهای دستی را بر عهده داشتند. با استفاده از ربات‌ها، یک بازوی رباتی این وظیفه را با دقت انجام می‌دهد. ربات‌ها با سرعت پایین کار کرده و به محض اینکه حسگرها هرگونه مانعی را بر سر راهشان شناسایی کنند برای ایمنی کارگران خط مونتاژ بلافاصله متوقف می‌شوند.

در کارخانه‌ی دینگوفلینگ بی‌ام‌دبلیو در آلمان، یک ربات سنگین در بخش مونتاژ سیستم انتقال اکسل برای بلند کردن چرخ دنده‌های مخروطی از سقف آویزان شده است. این چرخ دنده‌ها ممکن است تا ۵.۵ کیلوگرم وزن داشته باشند. ربات این چرخ دنده‌ها را به دقت و بدون اینکه آسیبی به آنها برسد جایگذاری می‌کند.

Compiled from M. Allinson. (2017, March 4). "BMW Shows Off Its Smart Factory Technologies at Its Plants Worldwide."

BMW Press Release. Robotics and Automation. <https://roboticsandautomationnews.com/2017/03/04/bmw-shows-off-its-smart-factory-technologies-at-its-plants-worldwide/11696/> (accessed September 2018);

"Innovative Human-Robot Cooperation in BMW Group Production." (2013, October 9).

[https://www.press.bmwgroup.com/global/article/detail/T0209722EN/innovative-human-robot-cooperation-in-bmw-groupproduction?](https://www.press.bmwgroup.com/global/article/detail/T0209722EN/innovative-human-robot-cooperation-in-bmw-groupproduction?language=en) language=en (accessed September 2018).

تگا یک ربات اجتماعی است که برای ارائه پشتیبانی گسترده از کودکان پیش دبستانی از طریق مشارکت دادن آنها را از طریق داستان سرایی و ارائه کمک در مورد واژگان طراحی شده است. مانند هاگبل، تگا نیز یک ربات اندرویدی و شبیه یک شخصیت انیمیشنی است. دارای دوربین خارجی و بلندگوهای داخلی بوده و برای کارکرد تا شش ساعت قبل از نیاز به شارژ طراحی شده است. تگا از قابلیت‌های

اندروید برای چشم‌های بیانی، توانایی‌های محاسباتی و فیزیکی استفاده می‌کند حرکات پاسخ کودکان به عنوان یک سیگنال پاداش به یک تقویت کننده به تگا داده می‌شود. الگوریتم یادگیری تگا از یک کنترل کننده اجتماعی، پردازش حسگر و کنترل موتوری برای حرکت دادن بدن و خم شدن و چرخش به چپ یا راست استفاده می‌کند. تگا نه تنها برای گفتن داستان، بلکه برای برگزاری گفتگو در مورد آن داستان‌ها طراحی شده است. با کمک یک اپلیکیشن روی تبلت، تگا به عنوان یک همسال و هم تیمی، نه به عنوان یک مربی با کودک ارتباط برقرار می‌کند. کودکان با تبلت و تگا ارتباط برقرار می‌کنند با تماشای حالات عاطفی کودکان بازخورد و واکنش‌ها را ارائه می‌کند. تگا همچنین به یادگیری واژگان کمک می‌کند و پاسخ‌های فیزیکی و عاطفی کودک را درک می‌کند که او را قادر می‌سازد تا با کودک رابطه برقرار کند. آزمایشات نشان داده است تگا می‌تواند تأثیر مثبتی بر علاقه کودک به آموزش، تفکر آزاد و رشد فکری داشته باشد. برای اطلاعات بیشتر، ویدیو را در <https://www.youtube.com/watch?v=16in922JTsw> تماشا کنید.

Compiled from E. Ackerman. (2016). *IEEE Spectrum*. <http://spectrum.ieee.org/automaton/robotics/homerobots/tega-mit-latest-friendly-squishable-social-robot> (March 5, 2017); J. K. Westlund et al. (2016). "Tega: A Social Robot." Video Presentation. *Proceedings of the Eleventh ACM/IEEE International Conference on Human Robot Interaction*; H. W. Park et al. (2017). "Growing Growth Mindset with a Social Robot Peer." *Proceedings of the Twelfth ACM/IEEE International Conference on Human Robot Interaction*; Personal Robots Group. (2016). <https://www.youtube.com/watch?v=sF0tRCqvyT0> (accessed September 2018); Personal Robots Group, MIT Media Lab. (2016). AAAS. https://www.eurekalert.org/pub_releases/2016-03/nsf-rlc031116.php (accessed September 2018).

پختن همبرگر یک شغل کم درآمد و معمولی است که افراد بسیاری با یک حقوق پایین می‌توانند مشغول به انجام آن شوند. اینگونه مشاغل به دلیل حضور ربات‌ها در طی زمان از بین خواهند رفت.

یکی از این پیاده‌سازی‌های رباتیک در صنایع غذایی در یک رستوران برگر در سانفرانسیسکو است. ماشین برگر ساز یک ربات سنتی بازو و پاهای رباتی نیست که می‌تواند حرکت کند و به عنوان یک انسان کار کنند. در عوض، این یک دستگاه کامل برگر ساز است که می‌تواند از آماده کردن همبرگر برای پخت و پز و تهیه یک وعده غذایی کامل را انجام دهد. این دستگاه با ترکیب یک طعم مناسب به کمک دستور العمل‌های سرآشپز میشلین‌استار و مقرون به صرفه بودن توانسته قدرت واقعی رباتیک را نمایش دهد. رستوران دو دستگاه ۱۴ فوتی را بکار گرفته که می‌توانند حدود ۱۲۰ برگر را در ساعت تهیه کنند. هر دستگاه از ۳۵۰ حسگر، ۲۰ رایانه و نزدیک به ۷۰۰۰ قطعه تشکیل شده است.

نان، پیاز، گوجه فرنگی، ترشیجات، ادویه‌ها و سس‌ها در لوله‌های شفاف بر روی یک تسمه نقاله پر می‌شوند. هنگامی که سفارش از طریق دستگاه تلفن همراه ثبت می‌شود سفارش در عرض پنج دقیقه آماده می‌شود. ابتدا، یک رول برگر با فشار هوا بر روی تسمه نقاله قرار می‌گیرد. بخش‌های مختلف ربات یکی پس از دیگری از بریدن رول به دو نیم برای آماده‌سازی سفارش کار می‌کنند را فشار می‌دهد لوله شفاف روی تسمه نقاله اجزای مختلف ربات یکی کار می‌کنند بعد از دیگری برای آماده کردن سفارش، از برش رول به دو نیم، مالیدن کره بر روی نان، خرد کردن سبزیجات و ریختن سس کار می‌کنند. به کمک حسگرهای حرارتی و یک الگوریتم، زمان و دمای پخت و پز تعیین شده و پس از پخت، همبرگر توسط یک بازوی رباتیک بر روی نان قرار می‌گیرد. زمانیکه دستگاه از لحاظ سفارش یا تأمین مواد اولیه دچار مشکل شود از طریق یک ساعت اپل به کارگران اطلاع داده می‌شود.

Compiled from "A Robot Cooks Burgers at Startup Restaurant Creator." (2018). TechCrunch. <https://techcrunch.com/video/a-robot-cooks-burgers-at-startup-restaurant-creator/> (accessed September 2018); L. Zimmeroff. (2018, June 21). "A Burger Joint Where Robots Make Your Food." <https://www.wsj.com/articles/a-burgerjoint-where-robots-make-your-food-1529599213> (accessed September 2018).

استفاده از ربات‌ها برای تهیه غذاهای مقرون به صرفه توسط یک رستوران فست فود در بوستون که غذاهای غلات و بشقاب سالاد سرو می‌کند نشان داده شده است. اسپایس یک رستوران مقرون به صرفه است که توسط فارغ التحصیلان مهندسی دانشگاه ام‌آی‌تی تاسیس شده است. مایکل فرید ربات‌هایی را ساخت که می‌توانند آشپزی کنند. این رستوران تعداد کمی از افراد با را خوب استخدام می‌کند و از روبات‌ها برای انجام کارهای زیادی استفاده می‌کند. سفارشات در یک کیوسک با صفحه نمایش لمسی گرفته می‌شود. پس از تایید سفارش، سیستم‌های مکانیزه شروع به تهیه غذا می‌کنند. مواد در یخچال در درون سط‌هایی قرار دارند و از طریق لوله‌های شفاف عبور داده می‌شوند و با استفاده از دستگاه سیار این مواد به مخزن مورد نظر انتقال داده می‌شوند. غذا توسط یک صفحه فلزی که به کناره‌ی مخزن رباتیک متصل شده است گرم می‌شود. دمای حدود ۴۵۰ درجه فارنهایت حفظ می‌شود، و غذا نزدیک به دو دقیقه پخته می‌شود. این فرایند همانند شستن لباس در ماشین لباسشویی است. پس از آماده شدن غذا، مخزن رباتیک کج شده و غذا را به کاسه منتقل می‌کند. بعد از هر دور پخت، قابلمه رباتیک خودش را با یک جریان پرفشار آب گرم شسته و سپس به موقعیت اولیه خود باز گشته و آماده‌ی طبخ غذای بعدی است. نام مشتری نیز بر روی کاسه قرار می‌گیرد. سپس یک پیشخدمت غذا را سرو می‌کند. اسپایس همچنین در تلاش است تا یک ربات را برای پخت پنکیک بکار بگیرد.

Compiled from B. Coxworth. (2018, May 29). "Restaurant Keeps Its Prices Down – With a Robotic Kitchen." New Atlas. <https://newatlas.com/spyce-restaurant-robotic-kitchen/54818/> (accessed September 2018); J. Engel. (2018, May 3). "Spyce, MIT-Born Robotic Kitchen Startup, Launches Restaurant: Video." Xconomy. <https://www.xconomy.com/boston/2018/05/03/spyce-mit-born-robotic-kitchen-startup-launchesrestaurant-video/> (accessed September 2018).

با افزایش جمعیت، صنعت کشاورزی نیز برای پاسخگویی ب تقاضا در حال گسترش است. برای حفظ قیمت‌ها و همچنین کیفیت با افزایش عرضه‌ی محصولات غذایی، شرکت چندملیتی هندی ماهیندرا آند ماهیندرا به دنبال بهبود فرایند برداشت انگوره‌های رومیزی است. این شرکت یک مرکز پژوهش و توسعه را در دانشگاه و مؤسسه‌ی پلی‌تکنیک ویرجینیا تأسیس کرد که با سایر مراکز ماهیندرا که در فنلاند، هند و ژاپن قرار دارند همکاری می‌کند. انگورها را برای آبگیری، شراب‌گیری و مصرف پذیرایی بکار می‌برند. میزان رسیده بودن و ظاهر انگوره‌های رومیزی با دو کاربرد دیگر بسیار متفاوت است از این رو کنترل کیفیت در اینجا بسیار حیاتی است. تصمیم‌گیری در این خصوص که کدام انگورها آماده‌ی برداشت هستند کار بسیار پردردسری است و باید اطمینان حاصل کرد که انگورها از کیفیت و یک‌دستی خوبی برخوردارند. این تصمیم‌گیری معمولاً نیازمند آموزش افراد خبره است که به آسانی امکان‌پذیر نیست. استفاده از برداشت رباتیک به جای افراد خبره مورد تحقیق و بررسی قرار گرفت و ربات‌ها می‌توانند با استفاده از حسگرهایی که کیفیت و در عین سرعت فرایند را مد نظر می‌گیرند به این اهداف دست پیدا کنند.

Compiled from L. Rosencrance. (2018, May 31). "Tabletop Grapes to Get Picked by Robots in India, with Help from Virginia Tech." RoboticsBusinessReview. <https://www.roboticsbusinessreview.com/agriculture/tabletop-grapes-picked-robots-india-virginia-tech/> (accessed September 2018); "Tabletop Grapes to Get Picked by Robots in India." Agtechnews.com. <http://agtechnews.com/Ag-Robotics-Technology/Tabletop-Grapes-to-Get-Picked-by-Robots-in-India.html> (accessed September 2018).

¹ Spyce

² Mahindra & Mahindra

به دلایل واضح و روشن، ارتش برای مدت زمان مدیدی است که در کاربردهای رباتیک سرمایه‌گذاری کرده است. ربات‌ها می‌توانند در جاهایی که ریسک از بین رفتن زندگی انسان‌ها بسیار بالاست جایگزین انسان شوند. ربات‌ها همچنین می‌توانند به مناطقی که به دلایل شرایط حاد - گرما، آب و غیره دسترسی به آنها برای انسان‌ها ممکن نیست دسترسی پیدا کنند. علاوه بر رشد اخیر پهپادها در کاربردهای نظامی، چندین ربات خاص نیز در طول زمان ساخته شده‌اند که به برخی از آنها در بخش‌های بعد اشاره می‌کنیم.

مارس^۱ (سیستم رباتیک مسلح پیشرفته مدولار) یک نسخه‌ی به‌روز آوری شده از ربات‌های سیستم تشخیص و شناسایی رصد سلاح‌های ویژه (سوردز)^۲ است که توسط ارتش ایالات متحده در جریان جنگ عراق استفاده شدند. این ربات برای شناسایی، رصد و هدف‌گیری طراحی شده و دارای دید ۳۶۰ درجه است. بسته به شرایط، مارس می‌تواند قدرت شلیک بسیار بیشتری را در زاویه‌ی دید کوچک خود ایجاد کند. انواع تسلیحات از قبیل گاز اشک‌آور، لیزرهای غیرکشنده و نارنجک‌انداز را با خود حمل کند. مارس یک ربات نظامی است که به طور مستقل عمل کرده و بنابراین ریسک کشته شدن سربازان در طول روز و شب را کاهش می‌دهد. این ربات چندین نوع حسگر برای ردیابی اثر حرارتی دشمن در طی شب و روز دارد و از دوربین‌های دید در شب برای پایش فعالیت‌های دشمن در طی شب استفاده می‌کند. مارس با دریافت فرمان به سمت دشمن شلیک می‌کند. کاربردهای دیگر آن شامل جابجایی بارهای سنگین از یک مکان به مکان دیگر است. این ربات گزینه‌های مختلفی را از اعمال نیروی غیرکشنده مثل هشداردهی یک حمله فراهم می‌کند. همچنین دارای یک سیستم ارتباطی دو طرفه است. این ربات می‌تواند همچنین از سلاح‌های غیرکشنده مثل خنده‌زا، اسپری فلفل و دود برای متفرق کردن جمعیت استفاده کند. این ربات از فاصله‌ی یک کیلومتری قابل کنترل بوده و برای به‌گونه‌ای طراحی شده تا سرعت خود را کاهش یا افزایش دهد، از پله‌ها بالا برود و بر روی مسیرهای ناهموار حرکت کند.

Compiled from T. Dupont. (2015, October 15). "The MAARS Military Robot." Prezi. <https://prezi.com/fsrlswo0qklp/the-maars-military-robot/> (accessed September 2018); Modular Advanced Armed Robotic System. (n.d.). Wikipedia. https://en.wikipedia.org/wiki/Modular_Advanced_Armed_Robotic_System (accessed September 2018); "Shipboard Autonomous Firefighting Robot - SAFFIR." (2015, February 4). YouTube. https://www.youtube.com/watch?time_continue=252&v=K4OtS534oYU (accessed September 2018).

سفیر (ربات آتش‌نشان خودکار مستقر در کشتی)^۳ آتش‌سوزی در کشتی یکی از بزرگ‌ترین ریسک‌ها را برای جان افراد حاضر در کشتی به همراه دارد. آتش‌سوزی در کشتی مشکلات متفاوت و حیاتی را بوجود می‌آورد. به دلیل فضای محدود، چالش‌هایی از قبیل دود، گاز و امکان محدود فرار از محل وجود دارد. اگرچه رویه‌هایی از قبیل مانورهای آزمایشی آتش‌سوزی، هشدارها، سیستم‌های اطفای حریق و اقدامات دیگری برای مقابله با آتش‌سوزی در دریا در نظر گرفته شده است ولی فناوری امروزی سعی کرده تا با این تهدید به شکل بهتری مقابله کند. یک تیم نیروی دریایی ایالات متحده در مرکز پژوهش‌های دریایی ربات سفیر را طراحی کردند. این ربات ۵ فوت و ۱۰ اینچ قد دارد به‌گونه‌ای طراحی نشده که کاملاً مستقل و خودگردان عمل کند. دارای یک ساختار رباتیک انسان‌نما است و به همین خاطر می‌تواند از راهروهای تنگ و گوشه و کناره‌های کشتی عبور کرده و از پله‌ها بالا برود. این ربات به‌گونه‌ای طراحی شده است که با

¹ MAARS (Modular Advanced Armes Robotic System)

² SWORDS

³ SAFFIR (Shipboard Autonomous Firefighting Robot)

موانع موجود در گذرگاه‌های یک کشتی کار کند. سفیر می‌تواند از تجهیزات محافظتی در برابر آتش مانند پوشش‌های ضد حریق، سرکوب کننده ها، و حسگرهایی که برای انسان طراحی شده‌اند استفاده کند. محرک‌های خطی سبک و کم اصطکاک کارایی و کنترل آن را بهبود بخشد. مجهز به چندین حسگر است: دوربین معمولی، دوربین گازی و مادون قرمز برای دید در شب و در دود سیاه. بدنه آن به گونه‌ای طراحی شده است که نه تنها در برابر آتش بلکه در پرتاب نارنجک‌های خاموش کننده نیز مقاوم باشد. این ربات می‌تواند حدود نیم ساعت بدون نیاز به شارژ کار کند. سفیر همچنین می‌تواند تعادل خود را روی یک سطح ناهموار حفظ کند.

Compiled from K. Drummond. (2012, March 8). "Navy's Newest Robot Is a Mechanized Firefighter." wired. om. <https://www.wired.com/2012/03/firefight-robot/> (accessed September 2018); P. Shadbolt. (2015, February 15). "U.S. Navy Unveils Robotic Firefighter." CNN. <https://www.cnn.com/2015/02/12/tech/mcisaffir-robot/index.html> (accessed September 2018); T. White. (2015, February 4). "Making Sailors 'SAFFiR' – Navy Unveils Firefighting Robot Prototype at Naval Tech EXPO." America's Navy. https://www.navy.mil/submit/display.asp?story_id=85459 (accessed September 2018).

پپر^۱ یک ربات نیمه انسان نما است که توسط سافت‌بانک رباتیکز^۲ ساخته شده و می‌تواند احساسات انسان را درک کند. یک صفحه نمایش روی سینه آن قرار دارد. می‌تواند اخم کردن، لحن صدا، لبخند زدن و اقدامات کاربر مانند زاویه سر و در هم قفل شدن انگشتان شخص را تشخیص دهد. به این ترتیب پپر می‌تواند تشخیص دهد که خلق و خوی یک فرد خوب است یا بد. پپر می‌تواند به طور مستقل راه رفته، افراد را شناسایی کرده و حتی می‌تواند خلق و خوی آنها را با صحبت کردن بهبود بخشد.

قدر پپر ۱۲۰ سانتی متر (حدود ۴ فوت) است. دارای سه چرخ جهت‌دار است که به آن متصل شده و امکان حرکت در کل محیط را می‌دهد. پپر می‌تواند سر خود را خم کند و بازوها و انگشتان خود را حرکت دهد و مجهز به دو دوربین با کیفیت بالا برای درک محیط است. به دلیل عملکردهای ضد برخورد، برخوردهای غیرمنتظره پپر کاهش یافته و می‌تواند انسان‌ها و همچنین موانع اطراف را تشخیص دهد. همچنین می‌تواند چهره انسان را به خاطر بسپارد و پرداخت از طریق گوشی هوشمند و کارت بانکی را قبول می‌کند. از دستورات به زبان‌های ژاپنی، انگلیسی و چینی پشتیبانی می‌کند.

پپر در صنایع خدماتی و همچنین منازل مستقر است. اگرچه پپر چندین مزیت برای برقراری ارتباط موثر با مشتریان دارد، اما به دلیل مسائل امنیتی و ناکفایتی در مکان‌ها مورد انتقاد قرار گرفته است. مثال‌های زیر اطلاعاتی در مورد کاربردها و معایب آن ارائه می‌دهد:

- تعامل با روبات‌ها در هنگام خرید، چهره هوش مصنوعی در محیط‌ها و فضای تجاری را تغییر می‌دهد. نستله ژاپن، یک تولید کننده پیشرو قهوه، از پپر برای فروش دستگاه‌های نسکافه برای ارتقاء تجربه مشتری استفاده کرده است. پپر می‌تواند طیف محصولاتی را که نستله ارائه می‌کند توضیح دهد و پاسخ‌های انسان را با استفاده از تشخیص چهره و صدا متوجه شود. با استفاده از یک سری سوالات و پاسخ به آنها، ربات نیاز مصرف کننده را شناسایی کرده و می‌تواند محصول مناسب را توصیه کند.
- برخی از هتل‌ها مانند کورتیارد ماریوت و ماندارین اورینتال از پپر برای افزایش رضایت مشتریان و کارایی استفاده می‌کنند. هتل‌ها از پپر برای افزایش تعامل مشتری، راهنمایی مهمانان به سمت فعالیت‌هایی که در حال اجرا هستند و تبلیغ برنامه‌های پاداش استفاده می‌کنند. هدف دیگر جمع‌آوری داده‌های مشتری و تنظیم دقیق ارتباطات بر اساس ترجیح و اولویت‌های مشتریان است. پپر در چند قدمی ورودی هتل‌های پارک موضوعی دیزنی لند مستقر شد و بلافاصله موجب افزایش تعامل با مشتریان شد. هتل‌ها از پپر برای گفتگو با مهمانان هنگام ورود یا خروج یا برای راهنمایی به اسپا، باشگاه ورزشی و سایر امکانات استفاده می‌کنند. همچنین می‌تواند مهمانان را

¹ Pepper

² SoftBank Robotics

در مورد کمپین‌ها و تبلیغات مطلع کند و به کارکنان کمک کند تا از انجام کارهای پیش پا افتاده ثبت نام مهمان در طرح‌های وفاداری مشتریان اجتناب کنند. واکنش‌های مشتریان نسبت به این موضوع تا حد زیادی مثبت بوده است

- سنترال الکترونیک کوآپریتو (سی‌ای‌سی)^۱ یک تعاونی توزیع برق واقع در سیلک‌واتر ایالت اوکلاهما است که از پپر برای نظارت بر قطعی برق استفاده کرده است. سی‌ای‌سی به بیش از ۲۰۰۰۰ مشتری در هفت شهرستان در اوکلاهما خدمات ارائه می‌دهد. پپر برای خواندن اطلاعات مربوط به قطع برق به مرکز عملیات متصل شده است و با متصل کردن آنها به نقشه‌های سیستم اطلاعات جغرافیایی (جی‌آی‌اس) همچنین می‌تواند در مورد مکان‌های زنده خودروهای خدماتی به مرکز عملیات اطلاع دهد. در سی‌ای‌سی، از پپر برای برگزاری کنفرانس‌هایی استفاده می‌شود که شرکت کنندگان می‌توانند اطلاعات بیشتری در مورد شرکت و خدمات آن داشته باشند. پپر به طیف وسیعی از سوالات در مورد مصرف انرژی پاسخ می‌دهد. در آینده، این شرکت برنامه دارد تا برای برآوردن کردن نیازهای سرمایه‌گذاری بیشتری بر روی روبات‌ها انجام دهد. شکل ۱۰.۲ پپر را نشان می‌دهد که به عنوان یک عضو تیم در طول مصاحبه با یک کارمند کاندید برای ارائه اطلاعات در مورد برنامه‌های سی‌ای‌سی و غیره شرکت می‌کند

- فابیو، یک ربات پپر است که به عنوان دستیار خرده فروشی در یک فروشگاه لوکس مواد غذایی و شراب در انگلستان و اسکاتلند نصب شده است. یک هفته پس از راه‌اندازی آن، فروشگاه خدمات آن را متوقف کرد زیرا مشتریان را گیج می‌کرد و آنها ترجیح می‌دادند تا خدمات را از کارکنان انسانی به جای فابیو دریافت کنند. فابیو پاسخ‌های عمومی در مورد پرسش‌هایی مانند محل قفسه اقلام ارائه می‌کرد. با این حال، به دلیل سر و صدای پس زمینه، نمی‌توانست به طور کامل بفهمد که مشتری چه چیزی را درخواست می‌کند. با قرار دادن فابیو در بخشی خاص که فقط تعداد کمی از مشتریان را به خود جلب می‌کرد، فرصت دیگری را به او دادند. سپس مشتریان از ناتوانی فابیو در حرکت در سوپرمارکت و راهنمایی آنها به سمت یک بخش خاص شکایت کردند. با کمال تعجب، کارکنان بازار به جای اینکه آن را به عنوان یک رقیب ببینند، به فابیو عادت کردند.



شکل ۱۰.۲ ربات پپر به عنوان یک شرکت‌کننده در یک میتینگ گروهی

Source: Central Electric Cooperative.

¹ Central Electronic Cooperative (CEC)

- پپر چندین مشکل امنیتی نیز داشت که پژوهشگران اسکاندیناویایی به آنها اشاره کردند. طبق یافته‌های آنها، دسترسی غیرمجاز در سطح روت به ربات آسانی وجود داشت. آنها همچنین متوجه شدند که ربات مستعد جمله‌ی جستجوی فراگیر بود. کارکردهای پپر را میشد با استفاده از رابط‌های برنامه‌نویسی کاربردی مختلف از طریق زبان‌هایی مثل پایتون، جاوا و سی++ برنامه‌ریزی کرد. این ویژگی موجب میشد تا به تمامی حسگرهای آنها دسترسی یافته و موجب ناامنی آن میشد. یک حمله‌کننده می‌توانست با ربات ارتباط برقرار کرده و از میکروفون، دوربین و سایر امکانات آن برای جاسوسی افراد و مکالمات آنها استفاده کند. این یکی از مشکلات عمومی بسیاری از ربات‌ها و اسپیکرهای هوشمند است.

Compiled from "Pepper Humanoid robot helps out at hotels in two of the nation's most-visited destinations (2017)". SoftBank Robotics.

<https://usblog.softbankrobotics.com/pepper-heads-to-hospitality-humanoidrobot-helps-out-at-hotels-in-two-of-the-nations-most-visited-destinations> (accessed November 2018); R. Chirgwin. (2018, May 29). "Softbank's 'Pepper' Robot Is a Security Joke." *The Register*.

https://www.theregister.co.uk/2018/05/29/softbank_pepper_robot_multiple_basic_security_flaws/ (accessed September 2018); A. France. (2014, December 1). "Nestlé Employs Fleet of Robots to Sell Coffee Machines in Japan." *The Guardian*.

<https://www.theguardian.com/technology/2014/dec/01/nestle-robots-coffee-machines-japan-georgeclooney-pepper-android-softbank> (accessed September 2018); Jiji. (2017, November 21). "SoftBank Upgrades Humanoid Robot Pepper." *The Japan Times*.

<https://www.japantimes.co.jp/news/2017/11/21/business/tech/softbank-upgrades-humanoid-robot-pepper/#.W6B3qPZFzIV> (accessed September 2018); C. Prasad. (2018, January 22). "Fabio, the Pepper Robot, Fired for 'Incompetence' at Edinburgh Store." *IBN Times*.

<https://www.ibntimes.com/fabio-pepper-robot-fired-incompetence-edinburgh-store-2643653> (accessed September 2018).

طی دهه گذشته، استفاده از رباتیک در جراحی‌ها شکل گرفته است. یکی از معروف‌ترین سیستم‌های رباتیک بکار رفته در جراحی، سیستم داوینچی^۱ است که هزاران عمل جراحی انجام داده است. طبق اظهارات جراحان، داوینچی همه‌جا حاضرترین ربات بکار رفته در یونیت‌های جراحی بیش از هر ربات دیگر است که برای انجام عمل‌های جراحی به اصطلاح تهاجمی متعدد طراحی شده و می‌تواند جراحی‌های ساده و در عین حال پیچیده را انجام دهد. اجزای مهم داوینچی عبارتند از کنسول جراح، سبد سمت بیمار، ابزار استاپلر و سیستم بینایی. کنسول جراح جایی است که جراح ماشین را به کار می‌گیرد و تصاویر سه‌بعدی با تفکیک‌نمایی بسیار بالا از داخل بدن بیمار تهیه می‌کند. کنسول دارای کنترل‌های اصلی است که جراح می‌تواند با انگشتان رباتیک آنها را در دست گرفته و بیمار را عمل کند. حرکت‌ها بسیار دقیق بوده جراح کنترل را به طور کامل در اختیار دارد و می‌تواند از حرکت ناخواسته‌ی انگشتان رباتیک جلوگیری کند. سبد سمت بیمار محلی است که بیمار در طی فرایند عمل جراحی در آنجا قرار می‌گیرد. دارای سه یا چهار بازو متصل است که جراح با استفاده از کنترل‌های اصلی آنها را کنترل می‌کند و هر بازو دارای نقاط گردش ثابتی است که توسط بازوها حرکت می‌کنند. مؤلفه‌ی سوم، ابزار استاپلر هستند که در حین عمل جراحی در دسترس قرار می‌گیرند. آنها در مجموع هفت درجه‌ی آزادی حرکت هستند و هر ابزار برای هدف خاصی طراحی شده است. اهرم‌ها برای تغییر ابزار سریعاً رها می‌شوند. مؤلفه‌ی آخر یک سیستم بینایی است که دارای یک دستگاه پردازش تصویر و اندوسکوپ سه‌بعدی است که تصاویر واقعی از آناتومی بیمار تهیه می‌کنند. یک مانیتور نیز با ارائه‌ی یک پرسپکتیو گسترده در طی فرایند جراحی به جراح کمک می‌کند. بیمارانی که توسط سیستم داوینچی جراحی شده‌اند سریع‌تر از سایر بیماران بهبود یافته‌اند زیرا برش‌هایی که بازوهای رباتیک انجام می‌دهند بسیار کوچک و دقیق هستند. جراح باید آموزش‌های آنلاین و عملی دیده و حداقل پنج عمل جراحی در حضور جراحان معتبر انجام دهد تا بتواند از سیستم داوینچی استفاده کند. این فناوری اگرچه هزینه‌ی جراحی را بالا می‌برد ولی توانایی آن برای کاهش درد و افزایش دقت موجب فراگیر شدن آن شده است.

Compiled from "Da Vinci Robotic Prostatectomy – A Modern Surgery Choice!" (2018). Robotic Oncology.

¹ Da Vinci

<https://www.roboticoncology.com/da-vinci-robotic-prostatectomy/> (accessed September 2018); "The da Vinci® Surgical System." (2015, September). Da Vinci Surgery. <http://www.davincisurgery.com/da-vincisurgery/da-vinci-surgical-system/> (accessed September 2018).

۱- اسنو (Snoo)

اسنو^۱ یک تخت نوزاد رباتیک مجهز به وای‌فای است که توسط ایو بهار، متخصص اطفال دکتر هاروی کارپ و مهندسان ام‌آی‌تی توسعه یافت. طبق اظهارات طراحانش، اسنو از استراتژی خواب معروف دکتر کارپ به نام پنج‌اس استفاده می‌کند که شامل قندان کردن، به پهلو خواباندن، ساکت کردن، تاب خوردن و مکیدن است. اسنو یک تخت نوزاد الکتریکی است که کودکان را به طور خودکار می‌خواباند و دقیقاً حس‌هایی که نوزاد در طی سه ماه آخر بارداری در شکم مادر تجربه می‌کند را خلق می‌کند. نوزادان زمانیکه صدای نويز سفید را می‌شنوند، تکان خوردن‌ها را حس می‌کنند و فنداق پیچ می‌شوند احساس راحتی می‌کنند. زمانیکه نوزاد به گهواره وصل می‌شود اسنو حس می‌کند که آیا کودک راحت است یا نه و حرکات او را دنبال می‌کند و تخت را به شکل حرکت رحم مادر حرکت می‌دهد تا زمانیکه کودک آرام شود. برای کنترل سرعت و نويز سفید می‌توان یک اپلیکیشن را بر روی گوشی هوشمند اسنو نصب کرد. همچنین اسنو را می‌توان پس از هشت دقیقه خاموش کرده یا اینکه سرتاسر شب کار کند. شرکت آن را به عنوان ایمن‌ترین رختخوابی که ساخته شده تبلیغ می‌کند که یک کمر بند برای جلوگیری از غلط خوردن یا افتادن کودک از پشت دارد. اسنو به والدین کمک می‌کند تا مجبور نباشند برای انجام اینکار چندین مرتبه در طی شب از خواب بیدار شوند.

Compiled from S. M. Kelly. (2017, August 10). "A Robotic Crib Rocked My Baby to Sleep for Months." CNN Tech.

<https://money.cnn.com/2017/08/10/technology/gadgets/snoo-review/index.html> (accessed September 2018); L. Ro. (2016, October 18). "World's First Smart Crib SNOO Will Help Put Babies to Sleep." Curbed. <https://www.curbed.com/2016/10/18/13322582/snoo-smart-crib-vves-behar-dr-harvey-karp-happiest-baby> (accessed September 2018).

۲- مدی (MEDI)

مدی^۲ که مخفف هوش طراحی ماشین و مهندسی است در شش بیمارستان در کانادا و یک بیمارستان در ایالات متحده وجود دارد. مدی به کاهش استرس ناشی از جراحی‌های سخت، تست‌ها و تزریقات کودکان کمک می‌کند. قد آن دو فوت و وزن آن حدود ۱۱ پوند است و شبیه یک اسباب‌بازی است. دکتر تانیا پران پس از شنیدن صدای حاکی از شغف و شادی کودکان از دیدن ربات پیشنهاد کرد که از این ربات در بیمارستان‌ها استفاده کنند. او بیان می‌کند که از آنجائیکه تخصص کافی در خصوص مدیریت درد در اینگونه موقعیت‌ها وجود ندارد این فناوری می‌تواند کمک بزرگی در این زمینه نماید. ربات قادر است به ۱۹ زبان صحبت کرده و خودش را به راحتی با فرهنگ‌های مختلف وفق بدهد. الدبران این ربات را ساخته و آن را نانو نامید. هزینه‌ی آن حدود ۸,۰۰۰ دلار یا بیشتر است. پران با افزودن نرم‌افزاری که بتواند در محیط‌های بیمارستانی با کودکان کار کند مدی را خلق کرد. مدی در طول درمان‌های مختلف با کودکان گفتگو می‌کند. در ابتدا برای واکسن آنفولانزا برنامه‌ریزی شده بود و از آن زمان تا کنون در تست‌های دیگر بکار گرفته شده است. مدی می‌تواند حتی برای کودکان داستان بگوید. این ربات با کاهش استرس کودک و آرام کردن آنها نه تنها به کودکان بلکه به پرستاران کمک می‌کند. والدین گفته‌اند زمانیکه کودکان بیمارستان را ترک می‌کنند، آنها حرفی از درد و سوزن نمی‌زنند بلکه با خاطرات خوش آنجا را ترک می‌کنند.

Compiled from A. Bereznak. (2015, January 7). "This Robot Can Comfort Children Through Chemotherapy." Yahoo Finance.

<https://finance.yahoo.com/news/this-robot-can-comfort-children-through-10736533404.html> (accessed September 2018); R. McHugh & J.

¹ Snoo

² MEDI

Rascon. (2015, May 23). "Meet MEDi, the Robot Taking Pain Out of Kids' Hospital Visits." NBC News. <https://www.nbcnews.com/news/us-news/meet-medi-robot-taking-pain-out-kids-hospital-visits-n363191> (accessed September 2018).

Compiled from M. Kelly. (2018, July 16). "This Adorable Robot Wants to Make Air Travel Less Stressful." The Verge. <https://www.theverge.com/2018/7/16/17576334/klm-royal-dutch-airlines-robot-travel-airport> (accessed September 2018); S. O'Kane. (2018, May 17). "Raden is the Second Startup to Bite the Dust After Airlines Ban Some Smart Luggage." Circuit Breaker. <https://www.theverge.com/circuitbreaker/2018/5/17/17364922/radensmart-luggage-airline-ban-bluesmart> (accessed September 2018).

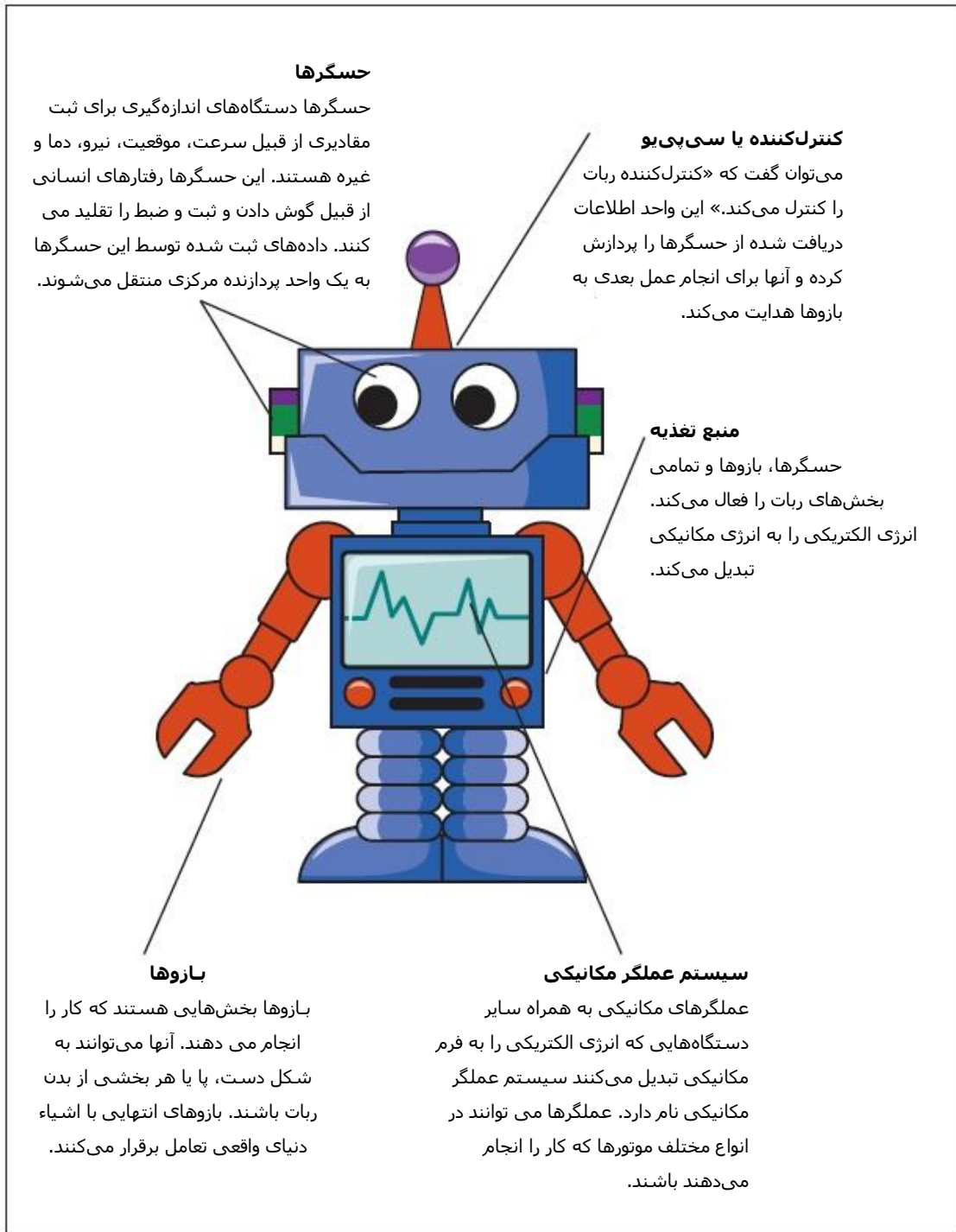
¹ Care-E Robot

² AGROBOT

Compiled from "Berry Picking at Its Best with Sensor Technology." Pepperl+Fuchs. <https://www.pepperlfuchs.com/usa/en/27566.htm> (accessed September 2018); R. Bogue. (2016). "Robots Poised to Revolutionise Agriculture." *Industrial Robot: An International Journal*, 43(5), pp. 45–456; "Robots in Agriculture." (2015, July 6). Intorobotics. <https://www.intorobotics.com/35-robots-in-agriculture/> (accessed September 2018).

سوالاتی برای مرور بخش ۱۰.۴

۲۷. کاربردهای رباتیک در کشاورزی را شناسایی کنید.
۲۸. چگونه یک ربات پشتیبانی اجتماعی همچون پپر یا میدی می‌تواند در مراقبت سلامت مفید باشد؟
۲۹. بر اساس کاربردهای مهمی که در این بخش ذکر شدند یک ماتریس تهیه کنید که سطرهای آن قابلیت‌های ربات و ستون‌های آن صنعت را نشان دهد. چه تشابهات و تفاوت‌هایی بین این ربات‌ها می‌بینید؟



شکل ۱۰.۳ اجزای اصلی یک ربات

کنترل کننده‌ی توان کنترل کننده‌ی توان، عامل محرک یک ربات است. اکثر ربات‌ها با نیروی باتری کار می‌کنند اما تعداد اندکی نیز هستند که با منبع تغذیه‌ی برق جاری مستقیم کار می‌کنند. در طراحی ربات‌ها باید عوامل دیگری از قبیل مصرف، توان کافی برای حرکت دادن تمامی اجزا را در نظر گرفت. حسگرهای نیرو، حسگرهای فراصوتی، حسگرهای مسافت، اسکنرهای لیزری و غیره به ربات‌ها کمک می‌کنند تا بر اساس محیط اطرافشان تصمیم بگیرند. از حسگرها برای شناسایی گفتار، بینایی، دما، موقعیت، فاصله، لامسه، نیرو، صدا و زمان استفاده می‌شود حسگرهای دیداری یا دوربین‌ها برای ساخت یک تصویر از محیط استفاده می‌شود تا ربات بتواند از محیط پیرامون خود آگاه شده و بین آنچه باید انتخاب کند و آنچه باید رها کند تمایز قائل شود. در ربات‌های مشارکتی، از حسگرها برای پیشگیری از تصادف با انسان‌ها و سایر ربات‌ها نیز استفاده می‌شود. بدین ترتیب، انسان‌ها و ربات‌ها می‌توانند بدون ترس از اینکه ربات‌ها ناخواسته به انسان آسیب برسانند در کنار یکدیگر کار کنند. حسگرها اطلاعات را گردآوری کرده و آن را به صورت الکترونیکی به واحد پردازنده مرکزی ارسال می‌کنند.

بازوها یا مانیپولاتورها یک بازو چیزی به جز بدن یک ربات نیست. همچنین می‌تواند بیانگر دستگاه‌ها و ادواتی باشد که بر محیط تأثیر می‌گذارند مثل دست‌ها، پاها، بازوها، بدنه و انگشتان. سی پی یو کنترل اقداماتی که بازو انجام می‌دهد را بر عهده دارد. یک کارکرد ضروری آنها جابجا کردن ربات و سایر اشیاء از یک محل به محل دیگری است و مشخصه‌های آنها به نقشی که برای ربات در نظر گرفته شده است بستگی دارد. ربات‌های صنعتی دارای بازوهای انتهایی هستند که به کارهای ربات کمک می‌کنند. بسته به نوع ربات، بازوهای انتهایی می‌توانند مغناطیسی، مشعل جوشکاری یا مکنده باشند.

سیستم ناوبری یا عملگر^۱ بازوها دستگاه‌هایی هستند که تعیین می‌کنند یک ربات چگونه حرکت کند. به کمک یک بازو، انرژی الکتریکی به انرژی مکانیکی تبدیل شده و ربات قادر خواهد بود تا به عقب، جلو، چپ، راست حرکت کرده و اشیاء را برداشته، انداخته و کارش را انجام دهد. بازو می‌تواند یک سیلندر هیدرولیکی یا یک موتور الکتریکی باشد. سیستم عملگر روشی است که تمام اجزاء یک ربات در قالب یک واحد گردهم می‌آیند.

کنترل کننده/سی پی یو این واحد مغز ربات بوده و هوش مصنوعی در آن تعبیه شده است. سی پی یو همچنین به یک ربات اجازه می‌دهد تا با اتصال تمامی سیستم‌ها با یکدیگر کارش را انجام دهد. همچنین دستوراتی را به ربات می‌دهد تا از جابجایی محیطی بدن و یا هر اقدام دیگری یادگیری داشته باشد.

سوالاتی برای مرور بخش ۱۰.۵

۱. اجزای معمول یک ربات کدام‌ها هستند؟
۲. کارکرد حسگرها در یک ربات چیست؟
۳. انواع مختلف حسگرها که ممکن است در یک ربات وجود داشته باشند کدام‌ها هستند؟
۴. کارکرد یک بازو یا مانیپولاتور چیست؟

¹ Effectors

ربات‌های از پیش تنظیم شده^۱ ربات‌های از پیش تنظیم شده از قبل برنامه‌ریزی شده‌اند. این ربات‌ها برای انجام کار یکسان در طول زمان طراحی شده و ۷ روز هفته و ۲۴ ساعته بدون وقفه کار می‌کنند. ربات‌های از پیش تنظیم شده از رفتار خود را تغییر نمی‌دهند. از این ربات‌ها اغلب در بخش‌های ساخت و تولید مثل صنعت جابجایی و حمل و نقل، ساخت و تولید وسیله‌ی نقلیه، جابجایی مواد و جوشکاری برای صرفه‌جویی در زمان و پول استفاده می‌شود. این ربات‌ها کارها را در محیط‌هایی که برای انسان‌ها خطرناک است انجام می‌دهند. ربات‌ها می‌توانند اشیاء سنگین را جابجا کرده، کارهای مونتاژ را انجام داده، قطعات را بازرسی کرده و مواد شیمیایی را جابجا کنند. یک ربات از پیش تنظیم شده از کارهایشان را بر اساس آنچه برایش برنامه‌ریزی شده انجام می‌دهد. این ربات‌ها می‌توانند نقش مهمی در رشته‌ی پزشکی داشته باشند زیرا وظایفی که انجام می‌دهند باید از کارایی بسیار بالایی در حد قابل قیاس با انسان‌ها باشد.

ربات‌های همکار یا کوبات‌ها^۲ کوبات‌ها ربات‌هایی هستند که می‌توانند با کارگران انسانی همکاری کرده و به آنها در رسیدن به اهدافشان کمک کنند. استفاده از کوبات‌ها در بازار رشد داشته و چشم‌انداز بسیار خوبی برای ربات‌های همکار وجود دارد. طبق یک نظرسنجی به عمل آمده توسط مارکتز آند مارکتز، ارزش بازار کوبات‌ها در سال ۲۰۲۰ چیزی حدود ۳.۳ میلیارد دلار خواهد بود. ربات‌های همکار کارکردهای مختلفی دارند. بسته به مورد کاربرد، از ربات‌های مشارکتی یا همکار استفاده می‌شود. ربات‌های همکار کاربردهای مختلفی در تولید و همچنین صنعت پزشکی دارند.

ربات‌های مستقل ربات‌های مستقل ربات‌هایی هستند که دارای یک سیستم هوش مصنوعی توکار بوده و به طور مستقل و بدون دخالت انسان‌ها کار می‌کنند. این ربات‌ها وظایفی را بسته به محیط انجام داده و خود را با تغییرات آن وفق می‌دهند. با استفاده از هوش مصنوعی، یک ربات مستقل یاد می‌گیرد تا رفتارش را تغییر داده و در انجام وظایفش پیشرفت کند. آنها می‌توانند همانند انسان حرکت کنند. ربات‌های مستقل دارای کاربردهای خانگی، نظامی، آموزشی و مراقب سلامت هستند. آنها می‌توانند همانند انسان‌ها حرکت کرده، از موانع اجتناب کرده و حمایت‌های اجتماعی-عاطفی بدهند. از برخی از این ربات‌ها برای مقاصد خانگی مثل جاروبرقی‌های مستقل مثل آی‌ربات رومبا استفاده می‌شود. از ربات‌های مستقل همچنین در بیمارستان‌ها برای دادن دارو و ارسال دقیق اطلاعات به پرستارانی که در آن شیفت و سایر شیفت‌ها کار می‌کنند استفاده می‌شود.

ربات‌های کنترل از راه دور حتی اگر ربات‌ها بتوانند وظایف مستقلی را انجام دهند، آنها فاقد مغز انسان هستند. از این رو، بسیاری از وظایف نیاز به نظارت انسانی دارند. این ربات‌ها را می‌توان از طریق وای فای، اینترنت یا ماهواره کنترل کرد. انسان‌ها ربات‌های کنترل از راه دور را برای انجام کارهای پیچیده یا خطرناک کنترل و هدایت می‌کنند. ارتش از این ربات‌ها برای منفجر

¹ Preset Robot

² Cobots

کردن بمب‌ها یا به عنوان سربازان شبانه روزی در میدان جنگ استفاده می‌کند. در زمینه تحقیقات برنامه های فضایی، دامنه استفاده از آنها گسترده است. از کوبات های کنترل از راه دور نیز برای انجام جراحی‌های نسبتاً تهاجمی استفاده می‌شود. **ربات‌های مکمل** ربات‌های مکمل قابلیت‌های انسانی موجود را تقویت کرده یا جایگزین قابلیت‌هایی که انسان از دست داده یا ندارد می‌شوند. این نوع ربات می‌توانند مستقیماً به بدن انسان متصل شوند. به بدن کاربر متصل شده و با آنها ارتباط برقرار کنند. ربات را می‌توان توسط بدن انسان و در برخی موارد حتی با فکر کردن به یک عمل خاص کنترل کرد. کاربردهای آن شامل خدمت به عنوان یک بازوی مصنوعی روباتیک یا افزایش دقت در جراحی است. تحقیقات گسترده ای در مورد ساخت اندام های مصنوعی در حال انجام است.

سوالاتی برای مرور بخش ۱۰.۶

۱. دسته‌بندی‌های اصلی ربات‌ها را شناسایی کنید.
۲. قابلیت‌های یک کوبات را تعریف کنید.
۳. فرق بین یک ربات از پیش تنظیم شده و یک ربات مستقل در چیست؟ برای هر کدام مثال بزنید.

یک ربات که در نهایت ممکن است زندگی اکثر مردم را تحت تأثیر قرار دهد، یک خودروی خودران است. مانند بسیاری از فناوری‌های دیگر، خودروهای خودران اخیراً در اوج تبلیغات بوده‌اند، اما مردم چالش‌های فنی، رفتاری و نظارتی آنها را نیز می‌شناسند. با این وجود، فناوری و فرایندهایی در حال تکامل هستند تا خودروی خودران را در آینده حداقل در یک شرایط خاص اگر نگوئیم در سرتاسر جهان به واقعیت تبدیل کنند. نسخه های اولیه خودروهای خودران توسط آنتن رادیویی توسعه یافته در سال ۱۹۲۵ فعال شدند. در سال ۱۹۸۹، محققان در کارنجی ملون از شبکه‌های عصبی برای کنترل یک وسیله نقلیه خودران استفاده کردند. از آن زمان، بسیاری از فناوری‌ها برای تسریع توسعه خودروهای خودران گرد هم آمده‌اند. اینها شامل موارد زیر هستند:

• **تلفن‌های همراه:** با کمک پردازنده های کامپیوتری کم مصرف و لوازم جانبی دیگر مانند دوربین، تلفن های همراه در همه جا حاضر شده‌اند. بسیاری از فناوری‌های توسعه‌یافته برای تلفن‌ها، مانند آگاهی از موقعیت مکانی و دید رایانه‌ای، در خودروها کاربرد پیدا می‌کنند

• **اینترنت بی سیم:** اتصال با ظهور شبکه های وای‌فای و جی ۴ بسیار امکان پذیرتر شده است. در آینده، رشد جی ۵ شاید برای خودروهای خودران مهم باشد تا به پردازنده‌های آنها اجازه دهند در زمان واقعی با یکدیگر ارتباط برقرار کنند

• **مراکز رایانه‌ای در خودروها:** تعدادی از فناوری‌های جدید در خودروهای امروزی موجود است، مانند دوربین‌های دید عقب و سنسورهای جلو و عقب که به وسایل نقلیه کمک می‌کنند اشیاء موجود در محیط را شناسایی کرده و راننده را در مورد آنها آگاه کنند یا حتی اقدامات لازم را به طور خودکار انجام دهند. به عنوان مثال، کروز کنترل تطبیقی به طور خودکار سرعت ماشین را بر اساس سرعت وسیله نقلیه جلویی تنظیم می‌کند.

• نقشه ها: نقشه های ناوبری در تلفن های همراه یا سیستم های ناوبری در اتومبیل ها کار راننده را در جهت یابی آسان کرده است. این نقشه ها یک وسیله نقلیه خودران را قادر می سازد تا مسیر خاصی را دنبال کند

• یادگیری عمیق: با پیشرفت در یادگیری عمیق، توانایی تشخیص یک شی یک عامل بسیار مهم برای ماشین های خودران است. به عنوان مثال، توانایی تشخیص یک فرد از یک شی مانند درخت، یا اینکه آیا جسم در حال حرکت است یا ثابت، در انجام اقدامات در یک وسیله نقلیه در حال حرکت بسیار مهم است.

قلب یک سیستم خودروی خودران یک مسافت یاب لیزری (یا تشخیص نور و برد - دستگاه لیدار) است که روی سقف خودرو قرار دارد. لیدار یک تصویر سه بعدی از محیط اطراف خودرو تولید می کند و سپس آن را با نقشه های جهانی با وضوح بالا ترکیب می کند تا مدل های مختلف داده را برای اقدام برای اجتناب از موانع و پیروی از قوانین راهنمایی و رانندگی تولید کند. علاوه بر این، بسیاری از دوربین های دیگر نیز نصب شده اند. برای مثال، دوربینی که در نزدیکی آینه دید عقب قرار دارد، چراغ های راهنمایی را تشخیص می دهد و فیلم می گیرد. قبل از تصمیم گیری برای ناوبری، خودرو تمام داده های جمع آوری شده از سنسور و دوربین را فیلتر کرده و از محیط اطراف خود یک نقشه می سازد و سپس با استفاده از جی پی اس دقیقاً خود را در آن نقشه قرار می دهد. این فرایند نگاشت و محلی سازی نامیده می شود.

این خودرو همچنین از سنسورهای دیگری مانند چهار دستگاه رادار که در سپر جلو و عقب قرار دارند، تشکیل شده است. این دستگاه ها به وسیله نقلیه اجازه می دهد تا فواصل دور را ببیند تا بتوانند از قبل تصمیم بگیرند و با ترافیک سریع مقابله کنند. رمزگذار چرخ مکان وسیله نقلیه را تعیین می کند و سوابق حرکت آن را حفظ می کند. الگوریتم هایی مانند شبکه های عصبی، تصمیم گیری مبتنی بر قوانین و رویکرد ترکیبی برای تعیین سرعت، جهت و موقعیت خودرو استفاده می شوند و از داده های جمع آوری شده برای تعیین سرعت خودرو استفاده می شود. برای جلوگیری از موانع، وسیله نقلیه را در جاده هدایت کنید. وسایل نقلیه خودران باید بر روی نقشه های دقیق جاده ها تکیه کنند. بنابراین، مهندسان قبل از ارسال خودروهای بدون راننده به جاده ها، چندین بار مسیری را رانندگی و اطلاعات پیرامون آن را جمع آوری می کنند. هنگامی که وسایل نقلیه بدون راننده در حال کار هستند، مهندسان داده های به دست آمده توسط آنها را با داده های تاریخی مقایسه می کنند. یک شهر کامل وجود دارد که تنها با هدف آزمایش وسایل نقلیه خودران ساخته شده است. در میشیگان واقع شده است. این شهر ساکن واحدی ندارد و وسایل نقلیه خودران بدون خطر در دنیای واقعی در خیابان ها پرسه می زنند. این شهر که ام سیتی نام دارد، واقعاً شهری برای است وسایل نقلیه رباتیک است. ام سیتی شامل تقاطع ها، علائم راهنمایی و رانندگی، ساختمان ها، کارهای ساختمانی و موانع متحرک مانند انسان و دوچرخه شبیه به موانع موجود در شهرهای واقعی است. خودروهای خودران نه تنها در این محیط بسته آزمایش می شوند، بلکه در دنیای واقعی نیز مورد استفاده قرار می گیرند .

واحد وی^۱ مو^۱ گوگل یکی از پیشگامان اولیه وسایل نقلیه خودران است. آنها در جاده های کالیفرنیا آزمایش شده اند، اما قبل از شروع رانندگی در کنار خودروهای انسان گردان، شرکت ها باید آنها را به طور کامل آزمایش کنند زیرا یک حادثه منفی می تواند مانع از پذیرش آنها شود. به عنوان مثال، در بهار ۲۰۱۸، یک وسیله نقلیه خودران که توسط اوبر آزمایش می شد،

¹ Waymo

یک عابر پیاده را در تمپ، آریزونا، کشت. این امر منجر به تعلیق کلیه آزمایش‌های عمومی وسایل نقلیه خودران توسط اوبر شد. این فناوری هنوز در حال توسعه است، اما این توسعه به اندازه‌ای پیش رفته که آزمایش محدود در جاده‌های عمومی ایمن باشد. ممکن است در آینده نزدیک از این واقعیت متعجب شویم که شخصی که در صندلی راننده کنار شما در ترافیک قرار دارد، ممکن است اصلاً آن را رانندگی نکند.

در سال ۲۰۱۶، وزارت حمل و نقل ایالات متحده به استقبال از وسایل نقلیه بدون راننده برای سرعت بخشیدن به توسعه آنها کرد. در سپتامبر ۲۰۱۶، این وزارت خانه اولین دستورالعمل‌های رانندگی خودران را اعلام کرد. یک اعلامیه پیشگامانه توسط اداره ملی ایمنی ترافیک بزرگراه یک ماه بعد به سیستم هوش مصنوعی کنترل کننده وسیله نقلیه خودران گوگل اجازه داد تا در پاسخ به پیشنهاد این شرکت به اداره ملی ایمنی ترافیک بزرگراه در نوامبر ۲۰۱۵ به عنوان راننده در نظر گرفته شود. برخی از ایالت‌ها در حال حاضر قوانین خاصی دارند که رانندگی خودران را ممنوع می‌کند. به عنوان مثال، تا لحظه نگارش این مقاله، ایالت نیویورک اجازه رانندگی با هندزفری را نمی‌دهد. بدون مقررات روشن، آزمایش وسایل نقلیه خودران یک چالش است. اگرچه چند ایالت مانند آریزونا، کالیفرنیا، نوادا، فلوریدا و میشیگان در حال حاضر به وسایل نقلیه خودران در جاده‌ها اجازه می‌دهند، کالیفرنیا تنها ایالتی است که در این مرحله مقررات صدور مجوز دارد. گوگل ممکن است برای خودروهای خودران شناخته شده‌ترین باشد، اما تنها شرکت نیست. تعداد انگشت شماری از قدرتمندترین شرکت‌ها مانند اوبر و تسلا نیز در یک رقابت هستند. هر شرکت بزرگ خودروسازی یا با شرکت‌های فناوری کار می‌کند یا فناوری خود را برای توسعه وسایل نقلیه خودران یا حداقل برای شرکت در این انقلاب.

خودروهای خودران مسائل و مشکلاتی نیز به همراه دارند:

• **چالش‌های تکنولوژی:** چندین چالش تکنولوژیکی در مورد استفاده از خودروهای خودران وجود داشته است. چندین مانع نرم افزاری و مکانیکی هنوز باید برطرف شود تا بتوان یک خودروی کاملاً مستقل را تولید کرد. به عنوان مثال، گوگل همچنان در تلاش است تا نرم افزار خود را تقریباً روزانه برای خودروی خودران خود به روز کند. چندین شرکت دیگر هنوز در تلاشند تا میزان اختیارات را در زمانی که یک راننده انسانی کنترل یک وسیله نقلیه اتوماتیک را به دست می‌گیرد، تعیین کنند.

• **چالش‌های زیست محیطی:** فناوری و قابلیت‌های مکانیکی هنوز نمی‌توانند بسیاری از عوامل محیطی تحت تاثیر خودروهای خودران را برطرف کنند. به عنوان مثال، هنوز نگرانی‌هایی در مورد عملکرد آنها در آب و هوای بد وجود دارد. به همین ترتیب، چندین سیستم در شرایط شدید مانند برف و تگرگ آزمایش نشده‌اند. آنجا چندین موقعیت ناوبری دشوار در جاده وجود دارد، مانند زمانی که یک حیوان روی آن می‌پرد.

• **چالش‌های نظارتی:** همه شرکت‌هایی که قصد دارند با خودروهای خودران درگیر شوند، باید به موانع نظارتی رسیدگی کنند. هنوز سوالات بی‌پاسخ زیادی در مورد مقررات رانندگی خودران وجود دارد. چندین سؤال در مورد مسئولیت شامل موارد زیر است: مجوز شامل چه مواردی می‌شود؟ آیا رانندگان جدید حتی اگر راننده نباشند ملزم به دریافت گواهینامه سنتی خواهند بود؟ در مورد جوانان یا افراد مسن دارای معلولیت چطور؟ برای راه اندازی این خودروهای جدید چه چیزی لازم است؟ دولت‌ها باید به سرعت کار کنند تا به فناوری‌های رو به رشد برسند.

با توجه به اینکه ایمنی عمومی در خط است، مقررات خودرو باید یکی از سخت‌گیرانه‌ترین مقررات در دنیای مدرن باشد.

• **مسائل مربوط به اعتماد عمومی:** اکثر مردم هنوز باور ندارند که یک خودروی خودران بتواند آنها را ایمن نگه دارد. اعتماد و پذیرش مصرف‌کننده از عوامل بسیار مهم هستند. به عنوان مثال، اگر شرایطی وجود داشته باشد که یک خودروی خودران مجبور به انتخاب بین زندگی یک مسافر در مقابل زندگی یک عابر پیاده شود، چه باید کرد؟ مصرف‌کنندگان ممکن است ایده کامل خودروهای بدون راننده را رد کنند. هیچ فناوری نمی‌تواند بی‌نقص باشد، اما سوال اینجاست که کدام شرکت می‌تواند مشتریان خود را به بهترین نحو متقاعد کند که زندگی خود را به آنها بسپارند.

پیشرفت‌هایی مشابه پیشرفت‌های خودروهای خودران در دیگر وسایل نقلیه خودمختار در حال بررسی است. به عنوان مثال، چندین شرکت قبلاً آزمایش کامیون‌های خودران را راه‌اندازی کرده‌اند. کامیون‌های خودران، اگر به طور کامل مستقر شوند، تأثیر تحول‌آفرین گسترده‌ای بر مشاغل در صنعت حمل و نقل خواهند داشت. به طور مشابه، تراکتورهای خودران در حال آزمایش هستند. در نهایت، پهپادها و هواپیماهای خودران نیز در حال توسعه هستند. این تحولات تأثیر زیادی بر مشاغل آینده خواهد داشت و در عین حال مشاغل جدید دیگری نیز در این فرایند ایجاد خواهد کرد. وسایل نقلیه خودران علیرغم موانع فنی و نظارتی مرتبط، بخشی از این دنیای فناوری شده‌اند. وسایل نقلیه خودران هنوز به توانایی‌های دانش‌راندگان انسانی دست پیدا نکرده‌اند، اما با پیشرفت فناوری، وسایل نقلیه رانندگی قابل اطمینان‌تر به واقعیت تبدیل خواهد شد. مانند بسیاری از فناوری‌ها، تأثیر کوتاه‌مدت ممکن است مبهم باشد، اما تأثیر بلندمدت هنوز مشخص نشده است.

سوالاتی برای مرور بخش ۱۰.۷

۱. برخی از پیشرفت‌های کلیدی فناوری که موجب رشد خودروهای خودران شده‌اند کدامها هستند؟
۲. نمونه‌هایی از مسائل نظارتی در خودروهای خودران را بیان کنید.
۳. تحقیقات آنلاین برای شناسایی آخرین پیشرفت‌های خودروهای خودران در حال انجام است. نمونه‌هایی از تحولات مثبت و منفی را ذکر کنید.
۴. کدام نوع از وسایل نقلیه خودران احتمالاً بیشترین تأثیر تحول‌آفرین را بر مشاغل خواهند داشت و چرا؟

رباتیک یک موهبت برای صنعت تولید بوده است. علاوه بر اتوماسیونی که با روباتیک امکان پذیر است، فناوری‌های جدید مانند سیستم‌های تشخیص تصویر، مشاغلی را خودکار می‌کنند که قبلاً به انسان برای بازرسی و کنترل کیفیت نیاز داشتند. کارشناسان مختلف صنعت گزارش می‌دهند که تا سال ۲۰۲۵، تا ۲۵ درصد از مشاغل فعلی با ربات‌ها یا هوش مصنوعی جایگزین خواهند شد. کتاب فقط انسان‌ها نیاز به درخواست دارند: برندگان و بازندگان عصر ماشین‌های هوشمند اثر دُونپورت و کیربی (۲۰۱۶) بر این موضوع تمرکز دارد. البته بسیاری از محققان دیگر، روزنامه نگاران، مشاوران و آینده پژوهان پیش‌بینی‌های خود را ارائه کرده‌اند. در این بخش به بررسی برخی از موضوعات مرتبط می‌پردازیم. این مسائل مربوط به هوش مصنوعی به طور کلی و

روباتیک به طور خاص است. بنابراین، فصل ۱۴ نیز به این موضوعات خواهد پرداخت، اما ما می خواهیم در این فصل این موارد را در زمینه رباتیک بررسی کنیم. به عنوان یک فعالیت گروهی، ویدیوی زیر را تماشا کنید:

<https://www.youtube.com/watch?v=GHC63Xgc0-8>.

همچنین <https://www.youtube.com/watch?v=ggN8wCWSIx4> را تماشا کنید. دریافت‌های شما از این ویدیوها چیست؟ محتمل ترین سناریو از نظر شما چیست؟ چگونه می توانید برای روزی آماده شوید که در واقع انسان ها ممکن است نیازی به درخواست برای بسیاری از مشاغل نداشته باشند؟

توانایی واتسون آی‌بی‌ام در هضم حجم وسیعی از داده ها در ادبیات تحقیقات پزشکی و ارائه آخرین اطلاعات به پزشک در مقالات نوشته شده است. فرصت های مشابهی برای افزایش شغل در بسیاری از زمینه های دیگر دیده شده است. مورد زیر را در نظر بگیرید: فناوری های مبتنی بر هوش مصنوعی مانند علم روایت و بینش های خودکار که می توانند داده های ساختاریافته را دریافت کنند، شامل تجسم های تولید شده توسط نرم افزارهایی مانند تابلو و یک پیش نویس اولیه از یک داستان را برای روایت آنچه که نتایج می رسانند تهیه می کنند. البته، به نظر می رسد که شغل یک روزنامه نگار یا حتی یک دانشمند داده را تهدید کند. در واقع، این کار می تواند با ارائه پیش نویس اولیه یک داستان، آن شغل را بهبود بخشد. سپس داستان نویسی می تواند روی موضوعات پیشرفته تر و استراتژیک تر مربوط به آن داده ها و تجسم تمرکز کند. قدرت قوام و جامعیت نیز می تواند در تکمیل مشاغل کمک کننده باشد. به عنوان مثال، همانطور که مایستر (۲۰۱۷) اشاره می کند، چت بات ها احتمالاً می توانند بسیاری از اطلاعات اولیه منابع انسانی را در اختیار کارمندان جدید قرار دهند. چت بات ها در ارائه چنین اطلاعاتی به کارکنان از راه دور نیز می توانند مفید باشند. یک چت بات بیشتر از یک انسان اطلاعات کامل و ثابتی را در هر بار ارائه می دهد. البته این بدان معناست که کارگرانی که وظیفه اصلی آنها بازگو کردن چنین اطلاعاتی برای هر کارمند جدید است یا به عنوان اولین منبع اطلاعات ممکن است مورد نیاز نباشد. هرناندز (۲۰۱۸) هفت دسته شغلی را شناسایی کرد که رباتیک به طور خاص و هوش مصنوعی به طور کلی در آنها گسترش می یابد. او همچنین چندین مطالعه دیگر را نقل کرد. بر اساس مطالعه ای مک کنزی، هوش مصنوعی می تواند منجر به ایجاد ۲۰۹۵۰ میلیون شغل جدید در ۱۰۹۱۵ سال آینده شود. مک کنزی همچنین پیش بینی می کند که ۷۵۹۳۷۵ میلیون نفر ممکن است به دلیل رباتیک و هوش مصنوعی نیاز به تغییر شغل و شغل در همان دوره زمانی داشته باشند. به گفته هرناندز، هفت مورد شغلی زیر احتمال افزایش دارد:

۱. **توسعه هوش مصنوعی:** این یک حوزه ی رو به رشد آشکار است. همانطور که شرکت های بیشتری محصولات و خدمات مبتنی بر هوش مصنوعی را توسعه می دهند، نیاز به چنین توسعه دهندگانی همچنان افزایش می یابد. به عنوان مثال، شرکت آی ربات که جاروبرقی های رباتیک تولید می کند، استخدام خود را از مهندسان سخت افزار به مهندسان نرم افزار تغییر می دهد تا در حال توسعه بعدی خود باشد. تولید محصولاتی که سازگارتر و مبتنی بر هوش مصنوعی هستند. جاروبرقی های ربات جدیدتر قادر به "دیدن" دیوار خواهند بود. آنها همچنین می توانند به مالک هشدار دهند که پاکسازی چقدر طول کشیده و منطقه را جارو کرده است
۲. **تعاملات مشتری و ربات:** از آنجایی که شرکت های بیشتری ربات ها را در این سازمان ها مستقر می کنند، پذیرش چنین روبات هایی هم از سوی کارمندان و هم از سوی مشتریان نامشخص است. یک دسته شغلی جدید برای مطالعه تعاملات بین یک ربات و همکاران و مشتریان و آموزش مجدد ربات یا در نظر گرفتن این اطلاعات در طراحی نسل

بعدی پدید آمده است. واضح است که مطالعه چنین تعاملاتی ممکن است استفاده از علم تجزیه و تحلیل/داده را نیز امکان پذیر کند.

۳. مدیران ربات: اگرچه ربات‌ها ممکن است بخش عمده‌ای از کار خود را در یک موقعیت خاص انجام دهند، انسان‌ها همچنان باید بر آنها نظارت کرده و اطمینان حاصل کنند که کار مطابق انتظار پیش می‌رود. علاوه بر این، در صورت بروز هر گونه شرایط غیرعادی، یک کارگر انسانی باید هشدار داده شود و به وضعیت واکنش نشان دهد. این در بسیاری از محیط‌ها که ربات‌ها در حال انجام بخش عمده‌ای از وظایف در زمینه‌هایی مانند تولید هستند، صادق است. هرناندز (۲۰۱۸) نمونه‌ای از روبات‌های کُبات را ارائه می‌کند که به عنوان نگهبان کار می‌کنند. این ربات‌ها هر زمان که یک فرد متجاوز شناسایی شود یا متوجه چیز غیرعادی شود، به انسان هشدار می‌دهند. البته، یک مدیر ربات انسانی معمولاً می‌تواند بر ربات‌های بیشتری نسبت به کارگران انسانی نظارت کند، زیرا نقش اصلی مدیر نظارت بر آنها و واکنش به موقعیت‌های غیر معمول است.

۴. برچسب‌زن‌های داده: ربات‌ها یا الگوریتم‌های هوش مصنوعی از مثال‌ها یاد می‌گیرند. و هرچه مثال‌های بیشتری به آنها داده شود، یادگیری آنها بهتر می‌شود (برای توضیح بیشتر در مورد این موضوع به فصل ۵ مراجعه کنید). برای مثال، سیستم‌های تشخیص تصویر تقریباً در هر محیطی (برای مثال به فصل ۵ در مورد یادگیری عمیق مراجعه کنید) به تعداد مثال‌هایی نیاز دارند. امکان بهبود قابلیت تشخیص آن سیستم‌ها وجود دارد. این نه تنها برای تشخیص چهره، بلکه برای برنامه‌های تصویری برای تشخیص سرطان از تصاویر اشعه ایکس، ویژگی‌های آب و هوا از تصاویر رادار و غیره بسیار مهم است. این امر مستلزم آن است که انسان‌ها تصاویر نمونه را ببینند و آنها را به عنوان نماینده یک شخص، ویژگی یا طبقه خاص برچسب گذاری کنند. این کار خسته کننده است و نیاز به انسان دارد. بسیاری از شرکت‌ها صدها نفر برچسب‌زن را استخدام کرده‌اند تا تصاویر را مشاهده کرده و آنها را تگ گذاری کنند. با رشد چنین کاربردهایی، نیاز به برچسب‌زن‌ها نیز افزایش می‌یابد. این کارگران برای بهبود مستمر ربات‌ها و الگوریتم‌های هوش مصنوعی از طریق ضبط مثال‌های مثبت کاذب یا جدیدتر مورد نیاز هستند.

۵. خلبانان و هنرمندان ربات: ربات‌ها به طور کلی و هواپیماهای بدون سرنشین به طور خاص برای ارائه عکس‌های اکشن با استفاده از دوربین‌های بالای سر یا زوایایی استفاده می‌شوند که انجام آن برای انسان اگر غیرممکن نباشد، دشوار است. پهپادها همچنین می‌توانند مانند پرندگان یا گل‌ها لباس بپوشند و نمای کلی منحصر به فرد و همچنین محیط را بهبود بخشند. به همین ترتیب، ربات‌های دیگر ممکن است لباس‌های منحصر به فردی بپوشند تا فضایی فرهنگی ایجاد کنند. چنین طراحان/آرایشگرانی توسط بسیاری از شرکت‌ها استخدام می‌شوند که خدماتی را برای رویدادهایی مانند کنسرت، عروسی و غیره ارائه بدهند. علاوه بر این، خلبانی هواپیماهای بدون سرنشین در حال تبدیل شدن به یک مهارت بسیار تخصصی برای برنامه‌های سرگرمی، تجاری و نظامی است. این مشاغل با تکامل برنامه‌ها افزایش خواهند یافت.

۶. رانندگان تست و بازرسان کیفیت. خودروهایی خردران به یک واقعیت تبدیل شده‌اند. با هر اتوماسیون وسایل نقلیه، حداقل برای آینده قابل پیش‌بینی، نیاز فزاینده‌ای به رانندگان ایمنی خواهد بود تا عملکرد هر وسیله نقلیه را نظارت کرده و در موقعیت‌های غیرعادی اقدامات مناسب را انجام دهند. شغل آنها مستلزم استفاده از کنترل از راه دور همانند خلبانان پهپاد نیست بلکه شامل نظارت مستمر از عملکرد خودرو و واکنش به موقعیت‌های اضطراری می‌باشد. مشاغل مشابه در سایر کاربردهای رباتیک نیز وجود دارد زیرا ربات‌ها برای کار در تنظیمات خاص آموزش دیده و آزمایش می‌شوند

۷. دانشمندان آزمایشگاه هوش مصنوعی. این ما را به اولین دسته از مشاغل جدیدی که شناسایی کردیم - کدنویسان هوش مصنوعی - می‌رساند. در حالی که کار آنها توسعه الگوریتم‌های ربات‌ها یا برنامه‌های هوش مصنوعی است، دسته مشابهی از کاربران بسیار متخصص نیز در حال ظهور هستند - افرادی که در استفاده از این سیستم‌های سخت‌افزاری و نرم‌افزاری برای کاربردهای خاص آموزش دیده و به کار گرفته می‌شوند. به عنوان مثال، پزشکان برای دریافت گواهی استفاده از ربات‌ها در جراحی‌ها، کار دیپ‌لوژی و اورولوژی و غیره باید آموزش‌های بیشتری را طی کنند. دسته دیگری از چنین متخصصانی شامل دانشمندانی است که این ربات‌ها و الگوریتم‌های هوش مصنوعی را برای دامنه آنها سفارشی‌سازی می‌کنند. به عنوان مثال، تعداد کمی از شرکت‌ها از ابزارهای هوش مصنوعی برای شناسایی مولکول‌های دارویی جدید برای توسعه و آزمایش گزینه‌های درمانی جدید برای بیماری‌ها استفاده می‌کنند. هوش مصنوعی می‌تواند چنین توسعه‌ای را سرعت بخشد. این دانشمندان نه تنها تخصص حوزه خود را توسعه می‌دهند، بلکه دانش یا حداقل توانایی دانشمندان داده را نیز توسعه می‌دهند تا برای ایجاد کاربردهای جدید با دانشمندان داده کار کند.

اگرچه فهرست قبلی چندین دسته از مشاغل را مشخص می‌کند که احتمالاً توسعه یا افزایش می‌یابند، میلیون‌ها شغل نیز احتمالاً از بین خواهند رفت. به عنوان مثال، اتوماسیون در حال حاضر بر تعداد مشاغل در لجستیک تأثیر گذاشته است. هنگامی که کامیون‌های خودران تبدیل به یک واقعیت شوند، حداقل برخی از مشاغل با درآمد خوب در حمل و نقل احتمالاً از بین خواهند رفت. ممکن است در مورد زمان وقوع این تغییر عظیم اختلاف نظر وجود داشته باشد، اما تأثیر بلندمدت آن بر مشاغل مسلم است. مسئله اصلی این بار این است که بسیاری از مشاغل اقتصاد دانشی "یقه سفید" آنهایی هستند که احتمال بیشتری دارد خودکار شوند. و این تغییر در تاریخ بی‌سابقه است. بسیاری از دانشمندان علوم اجتماعی، اقتصاددانان و متفکران برجسته نگران تحولاتی هستند که این موج بعدی اتوماسیون رباتیک ایجاد خواهد کرد. و در حال بررسی راه‌های مختلف هستند. به عنوان مثال، مفهوم درآمد پایه جهانی^۱ پیشنهاد شده است. طرفداران درآمد پایه جهانی استدلال می‌کنند که دادن حداقل درآمد پایه به هر شهروند تضمین می‌کند که با وجود از دست دادن گسترده مشاغل، هیچ‌کس گرسنه نمی‌ماند. که احتمال وقوع دارد. برخی دیگر، به عنوان مثال، لی (۲۰۱۸)، استدلال کرده‌اند که ارائه درآمد پایه جهانی ممکن است نیاز انسان به دستاوردها و مشارکت‌های معنادار در زندگی را برآورده نکند. لی یک کمک هزینه سرمایه‌گذاری اجتماعی (SIS) را پیشنهاد می‌کند که افراد را به رسمیت می‌شناسد. کمک به جامعه برای ارائه حمایت و مراقبت، خدمات اجتماعی یا آموزش. این کمک هزینه برای به رسمیت شناختن خدمات یک فرد در یکی از این دسته‌ها پرداخت می‌شود. کتاب لی بر این موضوع تمرکز می‌کند و یکی از ایده‌های متعددی است که در مورد چگونگی برنامه‌ریزی و رسیدگی به اختلالات آینده اتوماسیون پیشنهاد شده است. هدف ما در این بخش این است که به سادگی این مسائل را به شما گوشزد کنیم.

سوالاتی برای مرور بخش ۱۰.۸

۲۳. کدام مشاغل در معرض نابودی به خاطر انقلاب جدید رباتیک هستند؟
۲۴. حداقل سه دسته‌ی جدید از مشاغل را احتمال دارد منجر به افزایش چشمگیر کارمندان جدید شود شناسایی کنید.
۲۵. آیا وظایفی که برچسب‌زن‌های داده بر عهده دارند فقط یکبار است یا مستمر؟
۲۶. در خصوص مفاهیم درآمد پایه جهانی و کمک هزینه سرمایه‌گذاری اجتماعی تحقیق کنید.

¹ Universal Basic Income

همانگونه که در بخش قبل گفته شد، تأثیر هوش مصنوعی به طور عام و رباتیک به طور خاص بسیار گسترده بوده و می‌توان هر دو موضوع را در مقوله‌ی ربات‌ها و به شکل گسترده‌تر برای هوش مصنوعی مورد مطالعه قرار داد. ملاحظات قانونی رباتیک و هوش مصنوعی در این فصل و در فصل ۱۴ مورد بحث و بررسی قرار می‌گیرند. بسیاری از مسائل قانونی اگرچه ما از فناوری‌های هوش مصنوعی، ربات‌ها و خودروهای خودران استفاده می‌کنیم هنوز لاینحل باقی مانده‌اند. در این بخش به بررسی برخی از ابعاد کلیدی تأثیرات قانونی مرتبط با هوش مصنوعی می‌پردازیم. مطالب این بخش توسط پروفیسور مایکل شوستر دستیار مطالعات قانونی در دانشکده‌ی کسب‌وکار اسپیرز در دانشگاه ایالتی آکلاهما ارائه شده است. وی در امور قانونی مرتبط با هوش مصنوعی یک خبره محسوب می‌شود که مطالب بسیاری را در این حوزه منتشر کرده است.

خودروهای خودران و سایر سیستم‌هایی که توسط هوش مصنوعی کنترل می‌شوند یک جعبه‌ی پاندورا برای مسئولیت شبه‌جرم بالقوه محسوب می‌شوند (در جائیکه اقدامات اشتباه شما را ملزم به پرداخت خسارت به شخص دیگر می‌کند). تصور کنید که یک موتورسوار زمانیکه یک خودروی خودران وارد خط حرکت آن شده و با آن تصادم می‌کند مصدوم شود. این اتفاقی بود که منجر به پرونده‌ی نیلسون علیه جنرال موتورز شد و می‌بایست به سوالات دشواری در خصوص مسئولیت شبه‌جرم ناشی از هوش مصنوعی پاسخ می‌داد. پرونده مختومه شد و مشخص نشد که چه کسی می‌بایست در صورت مصدومیت ناشی از یک سیستم تحت کنترل هوش مصنوعی خسارت پرداخت کند. کاندیدهای بالقوه برای مسئولیت شامل برنامه‌نویسان سیستم هوش مصنوعی، سازندگان محصولی که شامل هوش مصنوعی است و مالکان محصولی هستند که در آن زمان به کسی آسیب رسانده است. دعاوی قصور پزشکی نیز به همین منوال توسط فناوری‌های نوین تغییر می‌کنند. با واگذار کردن برخی تصمیم‌گیری توسط پزشکان به هوش مصنوعی، پرونده‌های حقوقی برای اقدامات پزشکی آسیب‌زننده از پرونده‌های مسئولیت حرفه‌ای (علیه پزشکان) به مسئولیت محصول (علیه سازندگان سیستم‌های هوش مصنوعی) منتقل می‌شود. یک مثال اولیه از این پدیده، دعاوی حقوقی علیه جراحی‌هایی بود که با استفاده از ربات سیستم‌های جراحی داوینچی انجام شده بود.

بکارگیری سیستم‌های هوش مصنوعی که قادر به ابداعات مستقل یا به کمک انسان بودند سوالات بسیاری را در مورد حق ثبت این اختراعات بوجود آورد. حق ثبت اختراع تا پیش از این به ابداعات و اختراعات نوینی داده میشد که از دید یک طرف معمولی که در رشته‌ی مربوطه کار می‌کند یک پیشرفت آشکار محسوب نمی‌شد. بر همین اساس، این استاندارد بررسی می‌کرد که آیا یک فناوری جدید برای انسان واضح و بدیهی است یا خیر ولی با همه‌گیر شدن هوش مصنوعی، دامنه‌ی آنچه بدیهی و واضح است نیز گسترش

یافت. اگر یک شخص معمولی در یک صنعت به یک سیستم هوش مصنوعی که قادر به ابداع چیزهای جدید است دسترسی داشته باشد، بسیاری از بهبودها در فناوری‌های شناخته شده می‌تواند بدیهی و آشکار باشد. از آنجائیکه این بهبودها بدیهی می‌شوند، امکان ثبت اختراع وجود نخواهد داشت. ابداعات هوش مصنوعی دستیابی به حق ثبت اختراع را برای انسان دشوارتر می‌کند زیرا این فناوری بسیار معمول‌تر و متداول‌تر شده است. فراتر رفتن از سطح ابداعات انسانی مسائل و مشکلاتی از لحاظ اینکه هوش مصنوعی بتواند مستقلاً ابداع کند را به همراه دارد. اگر شخصی نتواند در ابداع و اختراع نقشی ایفا کند (بلکه صرفاً هدفی که باید به آن برسد را تعیین کند یا داده‌های پشتیبان را تهیه کند)، یک مخترع محسوب نخواهد شد. اگر هوش مصنوعی یا ابداع را بدون نقش یک مخترع انسانی خلق کند آیا او مالک حق ثبت اختراع است یا اصلاً باید ثبت اختراعی صورت بپذیرد؟ برخی با استناد به قانون اساسی ایالات متحده معتقدند که حقوق ثبت اختراع را فقط می‌توان به «مخترعانی» داد که لزوماً یک عامل انسانی در آنها نقش داشته باشد و بنابراین کنگره نمی‌تواند از لحاظ قانون اساسی اجازه‌ی ثبت اختراع را به فناوری‌هایی که توسط هوش مصنوعی خلق شده‌اند بدهد. برخی دیگر معتقدند که چرا باید به مالک رایانه، خالق هوش مصنوعی یا دیگران به خاطر ابداعات رایانه‌ای حق ثبت اختراع داد. اینها مسائلی است که باید حل شود.

رباتیک و هوش مصنوعی جایگزین تعداد زیادی از مشاغلی خواهند شد که در حال حاضر توسط انسان‌ها انجام می‌شوند. مفسران در مورد اینکه آیا فناوری‌های جدید تعدادی شغل جدید در قبال مشاغل جایگزین شده توسط اتوماسیون ایجاد خواهد کرد یا خیر، اختلاف نظر دارند. اگر دامنه مشاغل جدید کمتر از مشاغل از دست رفته باشد، ممکن است مسائل مالیاتی ایجاد شود. یک نگرانی خاص به مالیات‌های حقوق و دستمزد فدرال می‌پردازد که به موجب آن کارگران و کارفرمایان مالیات را از پیش می‌پردازند در مورد دستمزدهای پرداخت شده توسط کارمند اگر تعداد کل کارگران و دستمزد خالص با اتوماسیون مشاغل کاهش یابد، پایه مالیات بر حقوق کاهش می‌یابد. با توجه به اینکه این مالیات‌ها برای پایداری برنامه‌های شبکه ایمنی مختلف که توسط دولت اداره می‌شود مهم هستند (به عنوان مثال، تامین اجتماعی)، کسری مالیات حقوق و دستمزد می‌تواند پیامدهای اجتماعی قابل توجهی داشته باشد. در سال ۲۰۱۷، بیل گیتس (بنیان‌گذار مایکروسافت) پیشنهادی را برای مالیات بر روبات‌هایی که برای خودکارسازی مشاغل موجود انسان استفاده می‌شوند، ارائه کرد. این مالیات جدید از نظر تئوری مکمل مالیات حقوق و دستمزد موجود برای تضمین تداوم بودجه برای برنامه‌های دولتی است. مفسران بر اساس توصیه (یا نیاز) برای چنین مالیاتی تقسیم می‌شوند. انتقاد رایج این است که اخذ مالیات از روبات‌ها مانع از پیشرفت فناوری می‌شود، که برخلاف سیاست پذیرفته شده تشویق چنین تلاش‌هایی است. هنوز راه حل رضایت بخشی برای این بحث وجود ندارد برسد.

فراتر از آن چیزی که قانون است یا باید باشد، هوش مصنوعی تأثیرات قابل توجهی در بخش‌های مجزا از قانون خواهد داشت. نمونه بارز این تأثیر در حوزه بررسی اسناد است - بخشی از دعوی که در آن اصحاب دعوا اسناد ارائه شده توسط مخالفان خود را ارزیابی می‌کنند. برای ارتباط با پرونده هزینه‌های مرتبط با این فرایند می‌تواند قابل توجه باشد، زیرا برخی موارد مستلزم بررسی میلیون‌ها صفحه توسط وکلایی است که صدها دلار در ساعت صورت حساب می‌دهند. مشتریان شرکتی که به دنبال کاهش هزینه‌ها هستند - و شرکت‌های حقوقی که به دنبال مزیت رقابتی هستند - سیستم‌های بررسی اسناد مبتنی بر هوش مصنوعی را برای به حداقل رساندن تعداد ساعات صورت حساب اتخاذ کرده‌اند (یا قصد دارند آن را اتخاذ کنند). به طور مشابه، برخی از شرکت‌ها فناوری‌های

خاص صنعت را برای ایجاد مزیت رقابتی اتخاذ کرده اند. حداقل یک شرکت حقوقی بزرگ استفاده از یک سیستم مبتنی بر هوش مصنوعی را برای تجزیه و تحلیل نقاط قوت و ضعف خود راه اندازی کرد.

با پیشرفت وضعیت هوش مصنوعی، این روند به سمت هوش «سطح انسانی» ادامه می دهد. اما با «شخصی تر شدن» آن، سؤالاتی در مورد اینکه آیا باید حقوقی را که معمولاً به انسانها اعطا می شود، به هوش مصنوعی داد مطرح می شود. اولین متمم قانون اساسی ایالات متحده مربوط به آزادی بیان، تجمع و مذهب است، اما آیا این حقوق باید شامل هوش مصنوعی نیز شود؟ به عنوان مثال، ممکن است کسی استدلال کند که این حقوق، دولت را از دیکته کردن آنچه که یک ربات می تواند بگوید (نقض حق آن برای آزادی بیان) منع می کند. در ابتدا، این پیشنهاد به نظر دور از ذهن است، اما شاید اینطور نباشد. در مورد این موضوع، قابل توجه است که دادگاه عالی ایالات متحده اخیراً برخی از حقوق آزادی بیان و آزادی های مذهبی را به نهادهای شرکتی تعمیم داده است. بر این اساس، پیشینه داخلی برای استطاعت قانون اساسی در حمایت از عوامل غیر انسانی وجود دارد. علاوه بر این، در سال ۲۰۱۷، عربستان سعودی به "سوفیا"، یک ربات انسان نما که توسط شرکت هانسون رباتیک هنگ کنگ در سال ۲۰۱۵ ساخته شد، تابعیت اعطا کرد. اینکه چگونه این مشکل خود به خود (داخلی و جهانی) حل خواهد شد، نیازمند گذر زمان است.

فعالیت های بسیاری وجود دارند که انسانها پیش از انجام آنها باید گواهینامه صادر شده توسط دولت یا سازمان حرفه ای را دریافت کنند (به عنوان مثال، وکالت یا پزشکی). با پیشرفت وضعیت هوش مصنوعی، هوش مصنوعی به طور فزاینده ای قادر به انجام این کارها به شکل مستقل از مشارکت انسانی خواهد بود. با در نظر گرفتن این موضوع، استانداردهایی برای تعیین اینکه آیا یک فناوری هوش مصنوعی قادر به ارائه خدمات رضایت بخش در زمینه های حرفه ای تعریف شده است یا خیر باید ایجاد شود. اگر یک ربات مستقل قادر به قبولی در آزمون وکالت یک ایالت باشد، آیا می تواند بدون نظارت انسان مشاوره حقوقی بدهد؟ تا جایی که بسیاری از گروه های حرفه ای برای حفظ شایستگی به آموزش سالانه نیاز دارند، این سیاستها چگونه در مورد فناوری های هوش مصنوعی اعمال می شود؟ آیا ملزم کردن یک رایانه که وظایف قانونی را انجام می دهد به اینکه در کلاس های آموزش حقوقی مستمر شرکت کند کار ارزشمندی است؟ این مسائل زمانی حل خواهد شد که هوش مصنوعی شروع به انجام کارهایی می کند که در حال حاضر منحصراً توسط متخصصان انسانی مانند پزشکان و وکلای انجام می شود.

علاوه بر انتخاب های خط مشی که قانون چیست یا باید باشد، ممکن است هوش مصنوعی بر اجرای قانون تأثیر بگذارد. رشد سریع فناوری به زودی به نیروهای پلیس امکان دسترسی به مقادیر زیادی از داده های تقریباً واقعی و ظرفیت محاسباتی برای تعیین محل وقوع جرایم می دهد. تخلفات شناسایی شده ممکن است طیف وسیعی از تخلفات عمومی (مثلاً عبور از چراغ قرمز) تا اعمال خصوصی بیشتر مانند گزارش درآمد کمتر در اظهارنامه مالیاتی را شامل شود. ظرفیت تشخیص این گونه اعمال مجرمانه در مقیاس بزرگ، سؤالات اجرایی مختلفی را ایجاد می کند. احتیاط در تعقیب تخلفات از دیرباز بخشی از اجرای قانون بوده است. آیا باید این قدرت انتخاب در مورد مسائل تصمیم گیری کلیشه ای را به سیستم های هوش مصنوعی واگذار کرد؟ علاوه بر این، برنامه های اجرایی مبتنی بر ماشین به طور مداوم با سؤالاتی در مورد قانون اساسی مواجه شده اند (به عنوان مثال، استفاده از دوربین هایی برای شناسایی

رانندگانی که از چراغ قرمز عبور می‌کنند). در حالی که این استدلال‌ها تاکنون ناموفق بوده‌اند، احتمالاً با گسترش این رویه، دوباره مورد اعتراض قرار خواهند گرفت. فراتر از اجرا برخی از آنها امکان پیاده سازی هوش مصنوعی در قوه قضاییه را مطرح کرده اند. به عنوان مثال، پیشنهاد شده است که مجازات مبتنی بر داده‌ها ممکن است با موفقیت بیشتری به اهداف تعیین شده دست یابد (مثلاً آموزش موفق در حبس یا اجتناب از تکرار جرم). چنین مکانیزمی، البته، مسائل شفافیت بالقوه و استدلال‌های مربوط به اعطای قدرت بیش از حد به سیستم‌های هوش مصنوعی را مطرح خواهد کرد .

صرف نظر از مسائلی که در دو بخش اخیر مطرح شده است، فناوری‌ها و کاربردهای رباتیک به سرعت در حال پیشرفت هستند. به عنوان مدیران، باید همچنان به نحوه مدیریت این فناوری‌ها فکر کنید و در عین حال از رفتارهای آنی و مسائل حقوقی در پیاده سازی فناوری‌ها آگاه باشید.

سوالاتی برای مرور بخش ۱۰.۹

۱. برخی از مسائل قانونی مهم برای رباتیک و هوش مصنوعی را شناسایی کنید؟
۲. مسئولیت آسیب‌رسانی یکی از مسائل بدیهی اولیه برای هر فناوری بوده است. برخی از چالش‌های اساسی در شناسایی چنین مسئولیتی را نام ببرید.
۳. اخبار جدید در مورد مداخلات غیرقانونی در انتخابات منجر به بحث‌هایی در مورد اینکه چه کسی مسئول کنترل آسیب است بوجود آورده است. زمانیکه سیستم‌های رسانه‌های اجتماعی خودکار و چت‌بات‌ها توانایی انتشار «اخبار کذب» را دارند چه کسانی باید بر آنها نظارت کرده و از چنین اقداماتی پیشگیری کنند؟
۴. برخی از مسائل مرتبط با اعمال قانون در بکارگیری هوش مصنوعی کدامها هستند؟

- اتوماسیون یا خودکارسازی صنعتی اولین موج ربات‌ها را به همراه آورده است ولی اکنون ربات‌ها کاملاً مستقل شده و در بسیاری از حوزه‌ها کاربرد دارند
- کاربردهای رباتیک در صنایع مختلف از قبیل کشاورزی، مراقب سلامت و خدمات مشتریان گسترش یافته است
-
-
-
-
-
-
-
-

... .۱۴۴

<https://www.youtube.com/watch?v=GHC63Xgc0-8> .۱۴۵

<https://www.youtube.com/watch?v=ggN8wCWSIx4>

... .۱۴۶

... .۱۴۷

... .۱۴۸

... .۱۴۹

... .۱۵۰

... ..

... ..

... .. .1

https://www.qinetiq-na.com/wp-content/uploads/brochure_maars.pdf2

... ..

... .. .3

... ..

... .. .4

... ..

... .. .5

... ..

... .. .6

... ..

... .. .7

... .. .8

... .. .9

... .. .10

... ..

... .. .11

... .. .12

تصمیم‌گیری گروهی، سیستم‌های مشارکتی و پشتیبانی هوش مصنوعی

- تصمیم‌گیری گروهی: تصمیم‌گیری که توسط یک گروه از افراد انجام می‌شود.
- سیستم‌های مشارکتی: سیستم‌هایی که به افراد امکان می‌دهد تا با یکدیگر همکاری کنند و با همکاران خود در تصمیم‌گیری مشارکت کنند.
- پشتیبانی هوش مصنوعی: استفاده از سیستم‌های هوش مصنوعی برای کمک به تصمیم‌گیری.
- مزایای تصمیم‌گیری گروهی: افزایش دقت تصمیم‌گیری، کاهش خطا، افزایش رضایت و تعهد اعضا، افزایش خلاقیت و نوآوری.
- چالش‌های تصمیم‌گیری گروهی: طولانی شدن فرآیند تصمیم‌گیری، احتمال بروز تنش و تعارض، احتمال بروز تصمیم‌گیری‌های غیرعقلانی.
- مزایای سیستم‌های مشارکتی: افزایش شفافیت و پاسخگویی، افزایش انگیزه و تعهد اعضا، افزایش دقت تصمیم‌گیری.
- چالش‌های سیستم‌های مشارکتی: نیاز به فرهنگ سازمانی مناسب، نیاز به آموزش و توانمندسازی اعضا.
- مزایای پشتیبانی هوش مصنوعی: افزایش سرعت و دقت تصمیم‌گیری، کاهش خطا، امکان دسترسی به داده‌های بیشتر.
- چالش‌های پشتیبانی هوش مصنوعی: نیاز به داده‌های باکیفیت، نیاز به تخصص در زمینه هوش مصنوعی.

در این فصل، ما به معرفی چندین موضوع مرتبط با پشتیبانی تصمیم گروهی و همکاری می‌پردازیم. افراد با هم کار می‌کنند و گروه‌ها (یا تیم‌ها) بسیاری از تصمیمات پیچیده را در سازمان‌ها می‌گیرند. افزایش پیچیدگی تصمیم‌گیری سازمانی نیاز به جلسات و کار گروهی را به دنبال دارد. حمایت از کار گروهی که در آن اعضای تیم ممکن است در مکان‌های مختلف باشند و در زمان‌های مختلف کار کنند، بر جنبه‌های مهم ارتباطات، همکاری رایانه‌ای و متدولوژی‌های محل کار تأکید دارد. پشتیبانی گروهی جنبه حیاتی سیستم‌های پشتیبانی تصمیم (DSS) است. سیستم‌های پشتیبانی گروهی مؤثر با پشتیبانی از رایانه برای افزایش سود و کاهش ضرر در عملکرد کار و فرایندهای اساسی تکامل یافته‌اند. ابزارها و روش‌های جدیدی برای حمایت از کار گروهی بکار گرفته می‌شود. اینها شامل هوش جمعی، جمع‌سپاری و انواع مختلف هوش مصنوعی است. در نهایت، همکاری انسان-ماشین و ماشین-ماشین موجب افزایش توان همکاری و حل مسئله شده است. تمامی این موضوعات در بخش‌های زیر ارائه شده‌اند:

- ۱۱.۱. پرونده‌ی آغازین: هندریک موتور اسپورت با تیم‌های مشارکتی تعالی می‌یابد
- ۱۱.۲. تصمیم‌گیری در گروه‌ها: مشخصه‌ها، فرایندها، منافع و اختلال عملکردها
- ۱۱.۳. پشتیبانی از کار گروهی و مشارکت تیمی با سیستم‌های رایانه‌ای
- ۱۱.۴. پشتیبانی الکترونیکی از همکاری و ارتباطات گروهی
- ۱۱.۵. پشتیبانی رایانه‌ای مستقیم برای تصمیم‌گیری گروهی
- ۱۱.۶. هوش جمعی و هوش مشارکتی
- ۱۱.۷. جمع‌سپاری به عنوان روشی برای پشتیبانی تصمیم

۱۱.۸. هوش مصنوعی و پشتیبانی هوش ازدحامی برای تصمیم‌گیری گروهی و همکاری تیمی

۱۱.۹. همکاری انسان-ماشین و تیم ربات‌ها

هندریک موتور اسپورت (اچ‌ام‌اس)^۱ یک شرکت پیشرو در مسابقات اتومبیل‌رانی است (با بیش از ۵۰۰ کارمند) که در سری رقابت‌های مانسٹر انرجی جام ناسکار^۲ فعالیت می‌کند. هدف اصلی اچ‌ام‌اس این است که هر سال تا حد امکان در مسابقات برنده شود. این شرکت با چهار اتومبیل مسابقه‌ای و تیم‌های آنها وارد مسابقات می‌شود. این شرکت همچنین اتومبیل‌های مسابقه‌ای خودش را تولید می‌کند که شامل ساخت یا بازسازی ۵۵۰ موتور خودرو در هر سال است. در این نوع کسب‌وکار، کار تیمی بسیار مهم است، زیرا افراد مختلف با مهارت‌ها و دانش‌های مختلف و چندین تیم حرفه‌ای به موفقیت شرکت کمک می‌کنند.

اچ‌ام‌اس در طول فصل مسابقه (۳۸ هفته در سال) در تمامی مسابقات اتومبیل‌رانی در سراسر ایالات متحده شرکت می‌کند. این شرکت هر هفته به یک پیست اتومبیل‌رانی متفاوت می‌رود. در طول زمان خارج از فصل (۱۴ هفته)، شرکت داده‌های به‌دست‌آمده و درس‌های آموخته شده در طول آخرین فصول مسابقات را تجزیه و تحلیل می‌کند و برای فصل بعد آماده می‌شود. دفتر مرکزی این شرکت شامل ۱۹ ساختمان پراکنده در ۱۰۰ هکتار است.

این شرکت باید در طول مسابقات باید سریع تصمیم بگیرد - برخی از این تصمیمات به صورت لحظه‌ای، گاهی اوقات در کسری از ثانیه باید گرفته شوند. اعضای تیم‌های مختلف باید مشارکت داشته باشند و آنها در مکان‌های مختلف هستند. ارتباطات و همکاری بسیار مهم است. مسابقات اتومبیل‌رانی بر اساس کار تیمی، رانندگان، مهندسان، برنامه ریزان، مکانیک‌ها و سایر افرادی است که شرکت می‌کنند. اعضا باید با یکدیگر ارتباط برقرار کرده و برای تصمیم‌گیری همکاری کنند. محیط به اندازه‌ای پر سر و صدا است که نمی‌توان در طول مسابقه صحبت کرد. با این حال، اعضای تیم باید داده‌ها، نمودارها و تصاویر را به اشتراک بگذارند و به سرعت چت کنند. چندین تصمیم باید در لحظه گرفته شود که به پیروزی در مسابقات کمک کند (مثلاً چه میزان سوخت در چند ثانیه آینده در طی مسابقه به یک ماشین افزوده شود). اعضای تیم باید با یکدیگر ارتباط برقرار کرده و داده‌ها از جمله داده‌های دیداری را به اشتراک بگذارند. حدود ۴۵ تا ۵۰ ثانیه طول می‌کشد تا یک ماشین یک دور ۲.۵ مایلی را در دیتونا ۵۰۰ طی کند. در طول مسابقه، مهندسان ارشد باید دائماً با سوخت‌رسان‌ها در ارتباط باشند. داده‌های لحظه آخری در طول جلسه مسابقه رایج است.

هرگونه دانش کسب شده در هر دور مسابقه را می‌توان برای بهبود دور بعدی بکار برد. در مسابقات، تصمیمات مربوط به سوخت‌رسانی اهمیت حیاتی دارند. تصمیمات بسیار دیگری نیز باید در طی فصل مسابقه گرفته شود. به عنوان مثال، پس از هر

¹ Hendrick Motorsport (HMS)

² Monster Energy NASCAR Cup

مسابقه، شرکت باید تعداد زیادی از افراد را به همراه تجهیزات و ملزومات از یک مکان به مکان دیگر منتقل کند. جابجایی‌ها باید سریع، کارآمد و مقرون به صرفه باشند. بازهم در اینجا به کار تیمی و هماهنگی نیاز است.

۱۴ هفته برای آماده شدن برای فصل بعد فرصت است. علاوه بر این، حجم قابل توجهی از داده‌ها برای تجزیه و تحلیل، شبیه‌سازی، بحث، و دستکاری وجود دارد. برای این کار، افراد نه تنها به ابزارهای ارتباطی و همکاری، بلکه به انواع مختلف تجزیه و تحلیل‌ها نیز نیاز دارند.

اِچ‌اِم‌اِس تصمیم گرفت از مایکروسافت تیمز^۱ که یک پلتفرم مبتنی بر چت برای فضای کاری تیمی در مایکروسافت آفیس ۳۶۵ است استفاده کند. این پلتفرم به عنوان یک مرکز ارتباطی برای اعضای تیم در مسیرهای مسابقه و در هر مکان دیگری در سازمان استفاده می‌شود. مایکروسافت تیمز داده‌ها را در قالب‌های مختلف در فضای کاری تیمز خود ذخیره می‌کند. بنابراین، خدمه خودرو، مهندسان و مکانیک‌ها می‌توانند در چند ثانیه تصمیماتی بگیرند که ممکن است به پیروزی در مسابقات کمک کند. این همچنین تجزیه و تحلیل محاسباتی را در یک مکان مرکزی امکان پذیر می‌کند.

مایکروسافت تیمز شامل چندین برنامه فرعی است و به راحتی به نرم افزارهای دیگر در آفیس ۳۶۵ متصل می‌شود. آفیس ۳۶۵ چندین ابزار دیگر را برای افزایش همکاری ارائه می‌دهد (به عنوان مثال، شیرپوینت). به عنوان مثال، در راه حل اِچ‌اِم‌اِس، یک لینک کار به اکسل و همچنین شیرپوینت وجود دارد. همچنین از وان‌نوت مایکروسافت برای اشتراک‌گذاری یادداشت‌های جلسه استفاده می‌شود. قبل از مایکروسافت تیمز، این شرکت از اسلک^۳ استفاده می‌کرد (بخش ۱۱.۴)، اما فاقد اسلک امنیت و عملکرد کافی بود.

اعضا باید حجم عظیمی از داده‌های جمع‌آوری شده در طول فصل مسابقه را به اشتراک بگذارند و در مورد آنها بحث کنند. توجه داشته باشید که چندین کارمند با مهارت‌ها و وظایف متعدد در اینجا فعالیت می‌کنند. راه حل شامل ایجاد یک مرکز همکاری^۴ برای پروژه‌های همزمان بوده است. توجه داشته باشید که هر پروژه متفاوت بسته به نوع پروژه ممکن است به استعدادها و داده‌های متفاوتی نیاز داشته باشد. همچنین، راه حل شامل سایر ابزارهای فناوری اطلاعات بوده است. به عنوان مثال، اِچ‌اِم‌اِس از داشبورد پاور بی‌آی^۵ برای ارائه‌ی دیداری داده‌ها استفاده می‌کند. برخی از داده‌ها را می‌توان به صورت صفحات گسترده مبتنی بر اکسل پردازش کرد.

مایکروسافت تیمز همچنین به صورت یک اپلیکیشن موبایل در دسترس است. فایل دیتای هر تیم در خانه و حتی زیر خودرو در پیست موجود است. بنابراین، بسته نرم افزاری قادر است به موقعیت‌های مهم بلافاصله پاسخ دهد.

¹ Microsoft Teams

² Microsoft Office 365

³ Slack

⁴ Collaboration hub

⁵ Power BI

نتایج اصلی اینکار بهبود بهره وری، ارتباطات روان‌تر، همکاری آسان‌تر و کاهش نیاز به زمان مصرف شده در جلسات حضوری بود. افراد می‌توانند به صورت آنلاین گفتگو کرده، شرکای خود را بدون ترک محل کار فیزیکی مشاهده کنند. این شرکت اعتراف می‌کند که بدون تیمز، نمی‌توانست به موفقیت دست یابد. امروزه تیمز هر آنچه که شرکت نیاز دارد را در سر انگشتان خود دارد.

سوالاتی برای مرور پرونده‌ی آغازین

۴۵. محرک‌های اصلی برای استفاده از مایکروسافت تیمز چه بودند؟
۴۶. برخی از بحث‌هایی که در طی فصل مسابقات مطرح می‌شوند و اینکه چگونه از فناوری برای پشتیبانی از آنها استفاده می‌شود را نام ببرید.
۴۷. برخی تصمیمات اتخاذ شده در زمان خارج از فصل و اینکه چگونه توسط فناوری پشتیبانی می‌شوند را نام ببرید.
۴۸. بحث کنید که چرا مایکروسافت تیمز انتخاب شد و چگونه از تصمیم‌گیری گروهی کار تیمی پشتیبانی می‌کند؟
۴۹. ارتباطات و همکاری بین این گروه‌ها را ردیابی کنید.
۵۰. کارکرد فضای کاری مایکروسافت تیمز را مشخص کنید.
۵۱. ویدیوی [youtube.com/watch?time_continue=108&v=xnFd9IOaTE](https://www.youtube.com/watch?time_continue=108&v=xnFd9IOaTE) را تماشا کرده و چکیده‌ای از محتوای آن تهیه کنید.

اولین درس این است که امروزه بسیاری از وظایف باید توسط تیم‌های همکاری انجام شود تا به موفقیت برسند. دوم، زمان موضوعی حیاتی است. بنابراین، شرکت‌ها باید از فناوری برای سرعت بخشیدن به عملیات و تسهیل ارتباطات و همکاری در کار تیمی استفاده کنند. سوم، امکان استفاده از نرم افزارهای موجود برای پشتیبانی وجود دارد، اما بهتر است از یک فروشنده اصلی استفاده کنید که محصولات تکمیلی دارد تا بتواند نرم افزار همکاری/ارتباطات را تکمیل کند. چهارم، چت کردن می‌تواند ارتباطات را تسریع کند و پشتیبانی از فناوری دیداری می‌تواند مفید باشد. پنجم، اعضای تیم به واحدهای مختلفی تعلق دارند و مهارت‌های متنوعی دارند. نرم افزار آنها را گرد هم می‌آورد. اعضای تیم باید اهداف روشنی داشته باشند و نحوه دستیابی به آنها را بدانند. در نهایت، همکاری می‌تواند هم در درون و هم بین گروه‌ها باشد.

Sources: Compiled from Ruiz-Hopper (2016) and Microsoft (2017).

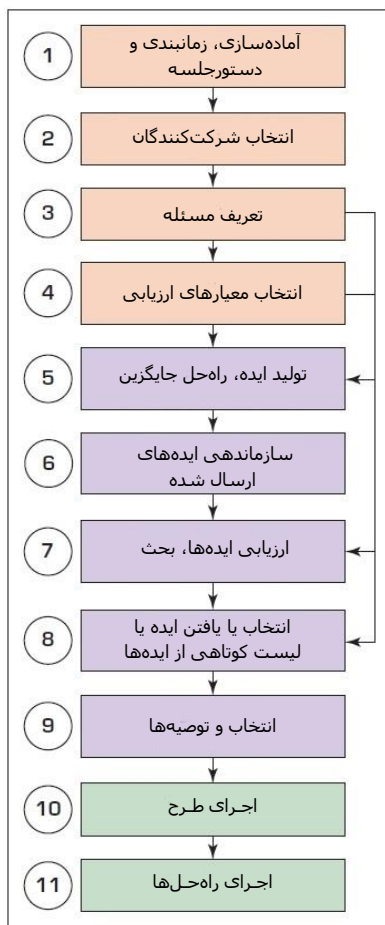
مدیران و سایر کارکنان دانشی به طور مداوم در حال تصمیم‌گیری، طراحی محصولات، توسعه‌ی خط مشی‌ها و استراتژی‌ها، تولید سیستم‌های نرم افزاری و غیره هستند. اغلب آنها این کار را به صورت گروهی انجام می‌دهند. وقتی افراد به صورت گروهی (یعنی تیمی) کار می‌کنند، کار گروهی یا کار تیمی انجام می‌دهند. **کار گروهی**^۱ به کاری گفته می‌شود که توسط دو یا چند نفر با هم انجام می‌شود. یکی از جنبه‌های کار گروهی، تصمیم‌گیری گروهی است. **تصمیم‌گیری گروهی**^۲ به موقعیتی اطلاق می‌شود که در آن افراد با هم تصمیم می‌گیرند. ابتدا به ویژگی‌های کار گروهی نگاه می‌کنیم.

¹ Group Work

² Group Decision Making

برخی از کارکردها و مشخصه‌های کارگروهی به شرح زیر هستند:

- اعضای گروه ممکن است در مکان‌های مختلفی باشند.
- اعضای گروه ممکن است در زمان‌های متفاوتی کار کنند.
- اعضای گروه ممکن است برای سازمان یکسان یا سازمان‌های متفاوتی کار کنند.
- یک گروه ممکن است موقت یا دائمی باشد.
- یک گروه ممکن است در یک سطح مدیریتی یا چندین سطح مدیریتی باشند.
- یک گروه ممکن است خلق هم‌افزایی کرده یا منجر به تعارض شود.
- یک گروه می‌تواند تولید بهره‌وری یا زیان کند.
- یک گروه ممکن است موظف باشد که مسئولیتش را بسیار سریع انجام دهد.
- ممکن است ملاقات تمامی اعضای تیم در یک مکان و در یک زمان بسیار پرهزینه یا ناممکن باشد به ویژه زمانی که جلسه‌ای برای موارد اضطراری برگزار شود.
- برخی از داده‌ها، اطلاعات یا دانش مورد نیاز گروه‌ها ممکن است در چندین منبع باشد که برخی از آنها خارج از سازمان باشند.
- ممکن است به تخصص اعضای تیم گروه نیاز باشد.
- گروه‌ها وظایف بسیاری را انجام می‌دهند ولی گروه مدیران و تحلیلگران اغلب بر تصمیم‌گیری یا حل مسئله تمرکز دارند.
- تصمیمات اتخاذ شده توسط یک گروه را راحت‌تر می‌توان اجرا کرد در صورتیکه توسط تمامی اعضای حمایت شود.
- کارگروهی دارای مزایای بسیاری است ولی متأسفانه اختلال عملکردهایی نیز دارد.
- رفتار گروه تحت تأثیر چندین عامل بوده و ممکن است بر تصمیمات گروه تأثیر بگذارد.



شکل ۱۱.۱ فرایند تصمیم‌گیری گروهی

گروه‌ها معمولاً درگیر دو نوع تصمیم‌گیری اصلی هستند:

۱. تصمیم‌گیری با یکدیگر
۲. فعالیت‌ها یا وظایف پشتیبان مرتبط با فرایند تصمیم‌گیری. به عنوان مثال، گروه ممکن است معیارهایی ارزیابی راه‌حل‌های جایگزین، اولویت‌بندی راه‌حل‌های ممکن و کمک به طراحی استراتژی برای اجرای آنها را انتخاب کنند.

فرایند تصمیم‌گیری گروهی مشابه فرایند تصمیم‌گیری عمومی است که در فصل ۱ توضیح داده شد ولی چندین گام بیشتر دارد. گام‌های فرایند تصمیم‌گیری گروهی در شکل ۱۱.۱ به تصویر کشیده شده است.

- گام ۱. آماده‌سازی برای جلسات در رابطه با دستور جلسه، زمان، مکان، شرکت‌کنندگان و زمانبندی.
- گام ۲. تعیین موضوع جلسه (مثلاً تعریف مسئله)
- گام ۳. انتخاب شرکت‌کنندگان در جلسه
- گام ۴. انتخاب معیارهای ارزیابی آلترناتیوها و راه‌حل انتخابی.
- گام ۵. تولید ایده‌های جایگزین (طوفان ذهنی)
- گام ۶. سازماندهی ایده‌های تولید شده در گروه‌های مشابه
- گام ۷. ارزیابی ایده‌ها، بحث و گفتگو و طوفان ذهنی
- گام ۸. انتخاب یک فهرست کوتاه
- گام ۹. انتخاب یک راه‌حل توصیه شده
- گام ۱۰. اجرای طرح راه‌حل
- گام ۱۱. اجرای راه‌حل

این فرایند به صورت متوالی نشان داده شده است ولی همانطور که در شکل ۱۱.۱ می‌بینید، احتمال دارد در قسمت‌هایی فرایند به صورت حلقه‌ای باشد. همچنین در صورتیکه هیچ راه‌حلی یافت نشود، فرایند را می‌توان مجدداً شروع کرد.

۱۱.۱.۱.۱. فرایند انتخاب تیم

زمانیکه یک گروه درگیر گام‌های شکل ۱۱.۱ می‌شود موارد زیر معمولاً صادق است:

- تصمیمات اتخاذ شده می‌بایست اجرا شوند.
- اعضای گروه معمولاً در وضعیت یا سطح یکسانی هستند.
- نتیجه‌ی یک جلسه تا حدودی به دانش، عقاید و قضاوت‌های شرکت‌کنندگان و حمایت آنها از پیامد جلسه بستگی دارد.
- پیامد یک جلسه به ترکیب گروه و فرایند تصمیم‌گیری مورد استفاده بستگی دارد.
- اعضای گروه با رتبه‌بندی شخص حاضر یا از طریق مذاکره اختلاف نظرات را حل و فصل می‌کنند.
- اعضای گروه ممکن است در یک مکان و رو در رو بوده یا یک تیم مجازی باشند که در آن صورت در مکان‌های متفاوتی بوده و به صورت الکترونیکی تیم تشکیل می‌شود. همچنین ممکن است در زمان‌های متفاوتی ملاقات کنند.

۱۱.۱.۱.۲. فرایند تصمیم‌گیری

برخی افراد برگزاری جلسات را یک ضرورت می‌بینند و برخی دیگر معتقدند که جلسات اتلاف وقت است. خیلی چیزها ممکن است در جلسات خوب پیش نرود. شرکت‌کنندگان ممکن است هدف جلسه را به خوبی درک نکنند یا تمرکزشان را از دست بدهند یا اینکه موضوعاتی که در دستور جلسه نیست را مطرح کنند. بسیاری از شرکت‌کنندگان ممکن است از صحبت کردن هراس داشته باشند یا تعدادی اندکی از اعضا بر بحث تسلط داشته باشند. سوء تفاهمات ممکن است به دلیل تفسیرهای متفاوت از زبان، ژست و حالات بوجود آید. در بینش فناوری ۱۱.۱ فهرستی از عواملی که می‌توانند اثربخشی چنین جلساتی را خدشه دار

کنند آورده شده است. کار تیمی علاوه بر چالش برانگیز بودن می تواند پرهزینه باشد. یک جلسه متشکل از چندین مدیر ارشد یا اجرایی می تواند چندین هزار دلار هزینه در بر داشته باشد.

کار گروهی دارای مزایا و معیاب بالقوه ای است. مزایای آن حاصل از کار کردن در یک گروه است. اختلال عملکردهای ناخوشایندی که در خلال یک جلسه کار گروهی ممکن است رخ دهد را زیان های فرایند می نامند. مثال هایی از هر کدام از آنها در بینش فناوری ۱۱.۱ آورده شده است.

بینش فناوری ۱۱.۱ مزایا و اختلال عملکردهای کار گروهی

برخی از مزایا و اختلال عملکردهای کار گروهی در جدول زیر آورده شده است:

اختلال عملکردهای فرایند گروهی رو در رو	مزایای کار گروهی
<p>فشارهای اجتماعی برای انطباق بر پیروی از جمع ممکن است منجر به گروه اندیشی یا تمایل به هم عقیده شدن با سایر افراد گروه شده و ایده های جدید را نپذیرند که منجر به فشار هم نویی می شود.</p> <p>فرایندی کند و زمان بر است.</p> <p>ممکن است برخی اطلاعات مرتبط وجود نداشته باشند</p> <p>یک جلسه ممکن است از فاقد هماهنگی بوده یا دستور جلسه ی ضعیفی داشته یا به خوبی برنامه ریزی نشده باشد</p> <p>یک جلسه ممکن است با محدودیت زمان، موضوع، عقیده یا نظر یک یا چند نفر یا ترس از مشارکت به خاطر بروز تعارض مواجه باشد.</p> <p>برخی از اعضای گروه تمایل دارند تا بر جلسه اعمال نفوذ کنند در حالیکه برخی دیگر سعی دارند تا برای انجام اکثر کارها به دیگران اتکا کنند. گروه ممکن است راه حل های خوب را نادیده بگیرد، اهداف ضعیفی را تعریف کند یا اینکه شرکت کنندگان گروه اشتباه انتخاب شده باشند.</p> <p>برخی اعضا ممکن است از صحبت کردن هراس داشته باشند.</p> <p>گروه ممکن است نتواند به یک اتفاق نظر برسد.</p> <p>کار کردن در یک گروه می تواند خلاقیت شرکت کنندگان و فرایند را بر انگیزد.</p> <p>ممکن است مصالحه و سازش های با کیفیت پایین انجام شود.</p>	<ul style="list-style-type: none"> • یادگیری به همراه دارد. گروه ها در درک مسائل بهتر از فرد عمل می کنند. آنها می توانند به یکدیگر بیاموزند. • افراد به آسانی می توانند مالکیت مشکلات و راه حل های آنها را بر عهده بگیرند. • اعضای گروه عزت نفس خود را در تصمیم نهایی دخیل می کنند و به همین دلیل خودشان را به آن تصمیم متعهد می کنند. • گروه ها در شناسایی خطاها بهتر از افراد عمل می کنند • اطلاعات و دانشی که در اختیار یک گروه است بیش از تک تک اعضای آن گروه است. اعضا می توانند دانش خود را برای خلق دانش جدید با یکدیگر ترکیب کنند. بدین ترتیب آلترناتیوهای خلاقانه تر و بیشتری را می توان برای حل مسئله تولید کرد و به راه حل های بهتری دست یافت (مثلاً از طریق طوفان ذهنی) • یک گروه ممکن است در طی فرایند حل مسئله هم افزایی ایجاد کند و بدین ترتیب اثربخشی و/یا کیفیت کار گروهی می تواند بیش از مجموع اثربخشی تک تک اعضای گروه باشد.

- همکاری در یک گروه می تواند به ارتباطات بهتر و دقیق تر منجر شود.
- گروه ممکن است تمرکز خود را از دست بدهد
- اغلب زمان های غیربهره ور وجود دارد (مثلاً معاشرت ها، آماده سازی، انتظار برای افرادی که با تأخیر حاضر می شوند)
- گرایش به ریسک متعادل می شود. گروه ها اعضای محافظه کار را تشویق کرده و اعضای با ریسک پذیری بالا را تعدیل می کنند.
- ممکن است گرایش برای تکرار مطالبی که قبلاً گفته شده است وجود داشته باشد.
- هزینه های جلسات ممکن است بالا باشد (مثلاً سفر، زمان صرف شده در جلسه).

سوالاتی برای مرور بخش ۱۱.۲

۴. کار گروهی را تعریف کنید.
۵. پنج ویژگی کار گروهی را نام ببرید.
۶. گام های تصمیم گیری گروهی را توضیح دهید.
۷. فعالیت های اصلی که در کار گروهی اتفاق می افتند را نام ببرید.
۸. پنج مزیت کار گروهی را نام برده و توضیح دهید.
۹. پنج اختلال عملکرد تصمیمات گروهی را نام برده و توضیح دهید.

زمانیکه افراد در قالب تیم کار می کنند به ویژه زمانیکه اعضای تیم در مکان های متفاوتی بوده و ممکن است اختلاف زمانی داشته باشند باید به مجموعه مختلفی از منابع اطلاعاتی با قالب های متعدد دسترسی داشته و از این طریق با یکدیگر ارتباط برقرار کرده و همکاری کنند. اینکار موجب پیچیدگی میتینگ ها به ویژه میتینگ های مجازی و افزایش احتمال ضرر و زیان های فرایندی می شود. از این رو، بسیار مهم است که برای برگزاری میتینگ ها از فرایندها و روال های مشخصی پیروی شود.

کار گروهی ممکن است به سطوح متفاوتی از هماهنگی نیاز داشته باشد. برخی اوقات یک گروه در سطح فردی عمل کرده و تک تک اعضا فعالیت هایی را انجام می دهند که نیازی به هماهنگی ندارد. مانند یک تیم دو میدانی که نماینده یک کشور شرکت کننده در یک دوی ۱۰۰ متر، بهره وری گروه در اینجا معادل بهترین نتایج فردی است. در مواقع دیگر، اعضای گروه ممکن است با هماهنگی تعامل داشته باشند. در این سطح، مانند یک تیم در مسابقه رله، کار نیاز به هماهنگی دقیق بین تلاش های فردی مستقل دارد. گاهی ممکن است یک تیم در سطح کاری هماهنگ عمل کند. همانند مسابقات قایقرانی، تیم هایی که در این سطح کار می کنند باید برای موفقیت تلاش کنند. مکانیسم های مختلفی از کار گروهی در سطوح مختلف هماهنگی پشتیبانی می

کند. اکثر سازمان های کوچک و بزرگ، از برخی روش ها و ابزارهای ارتباطی مبتنی بر رایانه و همکاری برای حمایت از افرادی که در تیم ها یا گروه ها کار می کنند، استفاده می کنند. از ایمیل ها گرفته تا تلفن های همراه و سرویس پیام کوتاه (SMS) و همچنین فناوری های کنفرانس، چنین ابزارهایی بخشی ضروری از زندگی کاری امروزی هستند. در ادامه به برخی از فناوری ها و کاربردهای مرتبط اشاره می کنیم.

برای همکاری مؤثر گروه ها، روش ها و فناوری های ارتباطی مناسبی مورد نیاز است. ما از این فناوری ها به عنوان سیستم های پشتیبانی گروهی¹ یاد می کنیم. اینترنت و مشتقات آن به عنوان مثال، اینترنت، اینترنت اشیا و اکسترانت زیرساخت هایی هستند که بسیاری از ارتباطات و همکاری ها بر روی آن ها انجام می شود. وب از تصمیم گیری مشارکتی درون سازمانی و بین سازمانی پشتیبانی می کند. چندین دهه است که از رایانه برای تسهیل کار گروهی و تصمیم گیری مشارکتی استفاده می شود. اخیراً، ابزارهای مشارکتی به دلیل افزایش قابلیت ها و توانایی آنها در صرفه جویی در زمان و هزینه (مثلاً در هزینه سفر) و تسریع در تصمیم گیری بیشتر مورد توجه قرار گرفته اند. ابزارهای رایانه ای را می توان بر اساس دسته بندی های زمانی و مکانی طبقه بندی کرد.

ابزارهای مورد استفاده برای پشتیبانی از همکاری، گروه ها و فناوری رایانش مشارکتی به مکان اعضای گروه و زمان ارسال و دریافت اطلاعات مشترک بستگی دارد. دی سانکتیس و گالوپ (۱۹۸۷) چارچوبی را برای طبقه بندی فناوری های پشتیبانی ارتباطات فناوری اطلاعات پیشنهاد کردند. در این چارچوب، ارتباطات به چهار سلول تقسیم می شود که با فناوری های پشتیبانی رایانه ای در شکل ۱۱.۲ نشان داده شده اند. چهار سلول در دو بعد زمان و مکان سازماندهی شده اند. هنگامی که اطلاعات تقریباً به طور همزمان ارسال و دریافت می شود، ارتباط در حالت همزمان (زمان واقعی) است. تلفن، پیام رسانی فوری (IM) و ملاقات های حضوری نمونه هایی از ارتباطات همزمان هستند. ارتباط ناهمزمان رخ می دهد هنگامی که گیرنده اطلاعاتی مانند یک ایمیل را در زمانی متفاوت از ارسال دریافت می کند (یا مشاهده می کند). فرستنده و گیرنده می توانند در یک مکان یا در مکان های مختلف باشند. همانطور که در شکل ۱۱.۲ نشان داده شده است، ترکیب زمان و مکان را می توان به عنوان یک ماتریس یا چارچوب چهار سلولی مشاهده کرد. چهار سلول چارچوب به شرح زیر است:

- **مکان/زمان یکسان.** حاضرین در جلسه همانند جلسات معمول به طور رو در رو ملاقات می کنند یا تصمیمات در یک اتاق تصمیم گیری مجهز و ویژه اتخاذ می شود. این روش بسیار مهمی حتی در شرایط مبتنی بر وب محسوب می شود زیرا برخی اوقات حاضرین باید محل کار معمول خود را برای دوری از هرگونه حواس پرتی ترک کنند.
- **زمان یکسان/مکان متفاوت.** شرکت کنندگان در مکان های متفاوتی هستند ولی در زمان یکسانی با یکدیگر ارتباط می گیرند (مثلاً ویدئو کنفرانس یا IM).
- **زمان متفاوت/مکان یکسان.** افراد به صورت شیفتی کار می کنند. یک شیفت اطلاعات را برای شیفت بعدی می گذارد.

¹ Group Supporting Systems (GSS)

- **زمان متفاوت/مکان متفاوت.** شرکت‌کنندگان در مکان‌های متفاوتی هستند و اطلاعات را در زمان‌های متفاوتی ارسال و دریافت می‌کنند. این حالت زمانی رخ می‌دهد که اعضای تیم سفر می‌کنند یا زمانبندی‌های متضادی دارند یا در مناطقی با اختلاف زمانی کار می‌کنند.

گروه‌ها و کار گروهی در سازمان‌ها در حال افزایش است. در نتیجه، گروه‌افزارها نیز برای پشتیبانی مؤثر در کار گروهی و عمدتاً برای مشارکت و برقراری ارتباطات رشد و تکامل می‌یابند.

	زمان یکسان	زمان متفاوت
مکان یکسان	<ul style="list-style-type: none"> • پیام‌رسانی فوری • چت، اتاق تصمیم • جی‌اس‌اس مبتنی بر وب • سیستم پرزنت چندرسانه‌ای • وایت‌برد • به اشتراک‌گذاری اسناد • فضای کاری 	<ul style="list-style-type: none"> • جی‌اس‌اس در اتاق تصمیم • جی‌اس‌اس مبتنی بر وب • سیستم مدیریت جریان کار • به اشتراک‌گذاری اسناد • ایمیل، وی‌میل • پخش ویدئوکنفرانس
مکان متفاوت	<ul style="list-style-type: none"> • جی‌اس‌اس مبتنی بر وب • وایت‌برد مجازی • به اشتراک‌گذاری اسناد • ایمیل، وی‌میل • کنفرانس صوتی • کنفرانس رایانه‌ای • ویدئوکنفرانس • محل کار مجازی 	<ul style="list-style-type: none"> • جی‌اس‌اس مبتنی بر وب • وایت‌برد مجازی • به اشتراک‌گذاری اسناد • ایمیل، وی‌میل • سیستم مدیریت جریان کار • کنفرانس رایانه‌ای با حافظه • پخش ویدئوکنفرانس • ویدئو مِمو

شکل ۱۱.۲ چارچوب زمان/مکان

گروه‌ها علاوه بر تصمیم‌گیری از زیرفرایندهای تصمیم‌گیری از قبیل طوفان ذهنی نیز پشتیبانی می‌کنند. فناوری همکاری به عنوان عامل محرک برای افزایش بهره‌وری و ارتقاء عملکرد کارکنان و سازمان شناخته شده است. گروه‌ها برای تصمیم‌گیری به چند روش همکاری می‌کنند. به عنوان مثال، گروه‌ها برای گام‌های ذکر شده در شکل ۱۱.۱ پشتیبانی ارائه می‌کنند. گروه‌ها می‌توانند به شناسایی مسائل، تعیین معیارهای انتخاب راه‌حل‌ها، ایجاد راه‌حل‌ها (مثلاً طوفان ذهنی)، ارزشیابی راه‌حل‌ها و انتخاب بهترین راه‌حل و اجرای آن کمک کنند. گروه‌ها ممکن است در یک یا چند گام درگیر باشند. علاوه بر این، می‌توانند داده‌های ضروری را گردآوری کنند.

از فناوری‌های بسیاری می‌توان برای همکاری استفاده کرد که برخی از آنها رایانه‌ای بوده و در چندین بخش در این فصل توضیح داده شده‌اند.

مطالعات نشان می‌دهد که بکارگیری فناوری‌های همکاری بهره‌وری را افزایش می‌دهد به عنوان مثال، راه‌حل‌های همکاری‌های دیداری موجب افزایش رضایت و بهره‌وری کارکنان می‌شود.

ما پشتیبانی رایانه‌ای را به دو بخش تقسیم می‌کنیم. در بخش ۱۱.۴ به معرفی پشتیبانی اصلی از فعالیت‌های عمومی در ارتباطات و همکاری‌ها می‌پردازیم. توجه داشته باشید که صدها و شاید هزاران محصول تجاری برای پشتیبانی از ارتباطات و همکاری‌ها وجود دارد که ما فقط به ذکر یک نمونه از آنها می‌پردازیم. بخش ۱۱.۵ پشتیبانی مستقیم از تصمیم‌گیری هم برای کل فرایند و هم گام‌های اصلی فرایند را پوشش می‌دهد. توجه داشته باشید که برخی محصولات از قبیل مایکروسافت تیمز که در پرونده‌ی آغازین معرفی شد هم از فعالیت‌های عمومی و هم فعالیت‌های فرایند تصمیم‌گیری پشتیبانی می‌کنند.

سوالاتی برای مرور بخش ۱۱.۳

۴. چرا برخی شرکت‌ها از رایانه‌ها برای پشتیبانی از کار گروهی استفاده می‌کنند؟
۵. جی‌اس‌اس چیست؟
۶. اجزای چارچوب مکان/زمان را توضیح دهید.
۷. اهمیت همکاری در تصمیم‌گیری را توضیح دهید.

ابزارها و روش‌های بسیاری برای تسهیل کار گروهی، همکاری الکترونیک و ارتباطات وجود دارند. در بخش‌های بعدی فقط به معرفی برخی ابزارها که از این فرایند پشتیبانی می‌کنند خواهیم پرداخت. تمرکز ما در اینجا بر پشتیبانی غیرمستقیم برای تصمیم‌گیری است. در بخش ۱۱.۵ ما به بررسی پشتیبانی مستقیم می‌پردازیم.

ابزارهای رایانه‌ای بسیاری برای پشتیبانی از گروه‌ها توسعه یافته‌اند. این ابزارها را گروه‌افزار می‌نامند زیرا هدف اصلی آنها پشتیبانی غیرمستقیم از کار گروهی است که در این بخش توضیح داده شد. برخی برنامه‌های ایمیل، اتاق‌های گپ و گفتگو، آی‌ام و تله‌کنفرانس‌ها این نوع پشتیبانی غیرمستقیم را ارائه می‌کنند. گروه‌افزارها مکانیزمی را برای اعضای تیم‌ها فراهم می‌کنند تا عقاید، داده‌ها، اطلاعات، دانش و سایر منابع را به اشتراک بگذارند. فناوری‌های رایانشی مختلف به شکل‌های متفاوتی از کار گروهی پشتیبانی می‌کنند که به وظیفه و اندازه‌ی گروه، سطح امنیت مورد نیاز و عوامل دیگر بستگی دارد.

گروه‌افزارهای بسیاری برای ارتقاء سطح همکاری تعداد اندک و زیاد افراد در اینترنت و اینترنت وجود دارند. یک مثال بارز آنها مایکروسافت تیمز است. ویژگی‌های محصولات گروه‌افزارها که از ارتباطات، همکاری و هماهنگی پشتیبانی می‌کنند در جدول ۱۱.۱ آورده شده است. در ادامه به تعریف برخی از آنها می‌پردازیم.

محصولات و ویژگی‌های توضیح داده شده در جدول ۱۱.۱ ممکن است همزمان یا غیرهمزمان باشند. کنفرانس وبی و آی‌ام و همچنین صدا روی پروتکل اینترنت جزو ویژگی‌های همزمان هستند. روش‌های مرتبط با حالت غیرهمزمان شامل ایمیل و فضای کاری آنلاین^۱ هستند که شرکت‌کنندگان می‌توانند در زمان‌های متفاوت با یکدیگر همکاری کنند. گوگل درایو (drive.google.com) و مایکروسافت شیرپوینت (<http://office.microsoft.com/en-us/SharePoint/collaboration-software-SharePoint-FX103479517.aspx>) به کاربران اجازه می‌دهند تا فضاهای کاری آنلاین را برای انبارش، به اشتراک‌گذاری و کار گروهی بر روی انواع مختلف اسناد راه‌اندازی کنند. محصولات مشابه شامل پلتفرم ابری گوگل^۲ و ابر فضای کاری سیتریکس^۳ هستند. شرکت‌هایی از جمله **Dropbox.com** روش ساده‌ای را برای به اشتراک‌گذاری اسناد ارائه می‌کنند. سیستم‌های مشابهی از قبیل به اشتراک‌گذاری تصاویر (مثلاً اینستاگرام، واتساپ، فیسبوک) برای مصارف خانگی در حال تکامل هستند.

جدول ۱۱.۱ محصولات و ویژگی‌های گروه‌افزارها

عمومی (همزمان یا غیرهمزمان)

- مجهز به سیستم پیام‌رسانی، ایمیل
- رابط مرورگر
- ایجاد صفحه وب مشترک
- به اشتراک‌گذاری هایپرلینک فعال
- مجهز به کارکردهای جستجو (بر اساس موضوع یا کلیدواژه)
- به اشتراک‌گذاری فایل (گرافیک، ویدئو، صوتی و غیره)
- ابزارهای جریان کار
- پورتال‌های سازمانی برای ارتباطات، همکاری و جستجو
- صفحات نمایش مشترک
- اتاق‌های تصمیم‌الکترونیک
- شبکه‌های هم‌تا به هم‌تا

همزمان

- پیام‌رسانی فوری
- ویدئو کنفرانس، کنفرانس‌های چندرسانه‌ای
- کنفرانس‌های صوتی
- وایت‌برد مشترک، وایت‌برد هوشمند
- ویدئوهای فوری

¹ Online Workspace

² Google Cloud Platform

³ Citrix Workspace Cloud

- طوفان ذهنی
- رأی‌گیری و سایر پشتیبانی‌های تصمیم (فعالیت‌هایی از قبیل زمانبندی، موافقت عمومی)
- چَت با انسان
- چَت با بات‌ها

غیرهمزمان

- فضاهای کاری مجازی
- توئیت‌ها
- توانایی ارسال/دریافت ایمیل، اس‌ام‌اس
- توانایی دریافت هشدارهای اطلاع‌رسانی از طریق ایمیل یا اس‌ام‌اس
- توانایی تجمیع/گسترش موضوعات^۱ در فروم‌های گفتگو
- مرتب‌سازی پیام‌ها (بر اساس تاریخ، نویسنده یا خوانده شده/خوانده نشده)
- پاسخ‌دهندگان خودکار^۲
- لاگ‌های چَت‌ها
- بولتن‌های خبری و گروه‌های بحث و گفتگوی الکترونیکی
- وبلاگ‌ها و ویکی‌ها
- ابزارهای طراحی و/یا برنامه‌ریزی مشارکتی

محصولات گروه‌افزارها یا به صورت مستقل که از یک وظیفه پشتیبانی می‌کنند (مثلاً ویدئوکفرانس‌ها) یا به صورت یکپارچه که شامل چندین ابزار هستند عرضه می‌شوند. به طور کلی، محصولات فناوری گروه‌افزارها نسبتاً ارزان بوده و به سادگی می‌توان آنها را در سیستم‌های اطلاعاتی موجود بکار گرفت.

پیشرفت سیستم‌های مبتنی بر وب فضا را برای بهبود میتینگ‌های مجازی^۳ که به صورت الکترونیکی و با حضور اعضای تیم در مکان‌های مختلف حتی در کشورهای مختلف برگزار می‌شوند فراهم آورده است. میتینگ‌ها و ابزارهای پرزنت آنلاین توسط ابزارهایی از قبیل ویکس^۴، **GoToMeeting.com**، اسکایپ. کام^۵ و بسیاری دیگر ارائه می‌شوند. این سیستم‌ها شامل سمینارهای وبی (وبینارها)، به اشتراک‌گذاری صفحه نمایش، کنفرانس صوتی، کنفرانس ویدئویی، رأی‌گیری، جلسات پرسش و پاسخ و غیره هستند. مایکروسافت آفیس ۳۶۰ شامل یک قابلیت برگزاری میتینگ یا جلسات مجازی است. حتی گوشی‌های هوشمند نیز امروزه دارای قابلیت‌های تعاملی کافی برای برگزاری جلسات زنده از طریق اپلیکیشن‌هایی از قبیل فیس‌تایم^۶ هستند.

¹ Discussion threads

² Auto Responders

³ Virtual Meeting

⁴ Webex

⁵ Skype.com

⁶ FaceTime

گردش کار مشترک به محصولات نرم‌افزاری اشاره دارد که با فرایندهای مشارکتی و پروژه‌محور سر و کار دارند. این برنامه‌ها به صورت مرکزی مدیریت شده ولی این امکان وجود دارد که کارگرانی از دپارتمانی و بخش‌های مختلف به آنها دسترسی یافته و با آنها کار کنند. هدف ابزارهای گردش کار مشارکتی این است که کارگران دانشی را تربیت و تقویت کند. تمرکز یک راه‌حل سازمانی برای گردش کار مشارکتی این است که به کارگران اجازه دهد تا در یک محیط یکپارچه با یکدیگر ارتباط برقرار کرده، مذاکره کرده و مشارکت و همکاری کنند. برخی از عرضه‌کنندگان پیشگام این نرم‌افزارهای گردش کار عبارتند از فایل‌نت^۲ و اکشن تکنالوجیز^۳. گردش کار مشارکتی با فضای کار مشارکتی مرتبط ولی متفاوت از آن است.

یک فضای کار مشترک فضایی است که افراد می‌توانند از هر مکان در یک زمان یا در زمان‌های متفاوت با یکدیگر کار کنند. این در اصل یک اتاق کنفرانس فیزیکی است که تیم‌ها برای برگزاری جلسات استفاده می‌کردند. سپس به یک فضای کاری مشترک توسعه یافت که به نام «فضای همکاری» شناخته می‌شد. برخی از اینها در شرکت‌ها بوده و برخی دیگر اجاره داده می‌شوند. فناوری‌های رایانه‌های متفاوتی برای پشتیبانی از کار گروهی در یک ساختار فیزیکی وجود دارند. برای ۱۲ مزیت فضاهای کاری مشترک نگاه کنید به پنا (۲۰۱۷).

یک فضای کاری مشترک مجازی محیطی است که مجهز به پشتیبانی دیجیتالی توسط اعضای گروهی است که در مکان‌های متفاوتی قرار دارند و می‌توانند اطلاعات را به اشتراک گذشته و همکاری کنند. یک مثال ساده، گوگل درایو است که امکان به اشتراک‌گذاری صفحات گسترده را فراهم می‌کند.

فضای کاری مشترک به کارکنان خبره در فناوری اجازه دسترسی به سیستم‌ها و ابزارها از طریق هر دستگاهی که بخواهند را می‌دهد. افراد می‌توانند به شکلی امن از هر کجا با یکدیگر همکاری کنند. فضای کاری دیجیتال موجب افزایش بهره‌وری و نوآوری تیم می‌شود. فضای کاری دیجیتال به کارکنان اختیار و توان داده و نوآوری آنها را آزاد می‌کند. به کارگران اجازه می‌دهد تا برای کارهای مشترک به سایر کارگران دسترسی پیدا کنند. برای جزئیات بیشتر و سایر فناوری‌های همکاری و مشارکت به لارس نوریس (۲۰۱۸) نگاه کنید.

پرایس‌واتر‌هاوس‌کوپرز^۴ یک اتاق جنگ تولید ایده را در دفتر خود در پاریس به صورت یک مرکز بزرگ مشارکت و همکاری برای پشتیبانی از جلسات مشتریان راه‌اندازی کرد.

پنج عرضه‌کننده‌ی مهم به شرح زیر می‌باشند:

¹ Collaborative Workflow

² FileNet

³ Action Technologies

⁴ PriceWaterhouseCoopers (PwC)

⁵ Flexible Workspace

GoToMeeting

-
-
-
-

فضای کاری اسلک یک فضای دیجیتال است که آن هم تیمی‌ها به مشارکت و همکاری در خصوص یک کار می‌پردازند. این فضای کاری می‌تواند در یک سازمان بوده یا سازمان‌های بزرگ چندین فضای اسلک بهم متصل داشته باشند.

هر فضای کاری شامل چندین کانال موضوعی است. این کانال‌ها می‌توانند به صورت عمومی، خصوصی یا مشترک تعریف شوند. سایر اجزای اسلک شامل پیام‌ها، جستجوها و نوتیفیکیشن‌ها هستند. چهار گروه از افراد با اسلک سر و کار دارند: مالکان فضای کاری، مدیران فضای کاری، اعضا و مهمان‌ها. برای راهنمای اسلک نگاه کنید به get.slack.help/hc/en-us/articles/115004071768-What-is-Slack-

اسلک ویژگی‌های بسیاری داشته و می‌تواند اپلیکیشن‌های مجازی امن را برای تقریباً هر دستگاهی ارائه کند.

به لحاظ سنتی، مشارکت و همکاری بین اعضای زنجیره تأمین و عمدتاً آنهایی که نزدیک به یکدیگر بودند صورت می‌گرفت (مثلاً سازنده و توزیع کننده آن یا یک توزیع کننده و یک خرده‌فروش). حتی اگر شرکای بیشتری نیز حضور داشته باشند، تمرکز بر بهینه‌سازی جریان اطلاعات و محصول بین گره‌های موجودی در زنجیره تأمین معمول است. روش‌های پیشرفته از قبیل برنامه‌ریزی، پیش‌بینی و تأمین مشترک این ساختار اصلی را تغییر نمی‌دهند.

مشارکت و همکاری سنتی منجر به یک زنجیره تأمین یکپارچه عمودی می‌شود. با وجود این، فناوری‌های وب می‌توانند شکل زنجیره تأمین، فعالان آن و سایر نقش‌های فردی را شدیداً دستخوش تغییر کنند. در یک شبکه‌های مشارکتی، شرکا در هر نقطه از شبکه می‌توانند با یکدیگر تعامل داشته و شرکای سنتی را کنار بگذارند. تعامل ممکن است بین چندین سازنده یا توزیع کننده و همچنین با فعالان جدید مثل عوامل نرم‌افزار صورت بگیرد.

هدف از یک هاب مشترک این است که یک نقطه مرکزی برای مشارکت و همکاری گروهی باشد. پلتفرم‌های هاب مشترک باید شرکت کنندگان را قادر کنند تا بتوانند به چندین شکل آنلاین با یکدیگر تعامل داشته باشند.

¹ Slack Workspace

² Surface Hub For Business

این محصول افراد را در هر جایی که باشند و در هر زمان که بخواهند از یک وایت بورد دیجیتال استفاده کنند و اپلیکیشن‌ها و نرم افزارها را یکپارچه کنند به یکدیگر متصل می‌کند. این محصول به ایجاد یک فضای کاری مشترک در جائیکه چندین دستگاه به صورت بی سیم به یکدیگر متصل هستند کمک می‌کند تا محیط کاری قدرتمندی را ایجاد کند.

همکاری در شبکه های اجتماعی همکاری و مشارکت های مرتبط با کسب و کارها را عمدتاً می‌توان در فیسبوک و لینکدین مشاهده کرد.

facebook.com/workspace

نرم افزار همکاری اجتماعی برای تیمها علاوه بر نرم افزارهای همکاری عمومی که دو نفر یا دو تیم می‌توانند استفاده کنند، پلتفرم‌های نرم افزاری ویژه ای برای تشکیل تیمها و پشتیبانی از فعالیت‌های آنها وجود دارد. چند مثال بارز طبق **collaboration- software.financesonline.com/c/social-collaborationsoftware/** عبارتند از رایک^۳، رایبور^۴، آزندو^۵، پلتفرم اجتماعی

¹ Social Collaboration

² LinkedIn Lookup

³ Wrike

⁴ Ryver

⁵ Azendoo

زیمبرا^۱، سیم پیج^۲، زوهو^۳، آسانا^۴، جایو^۵، چتر^۶ و سوشیال تیبلز^۷. برای مشاهده ی بهترین نرم افزارهای مشارکت و همکاری اجتماعی به تفکیک گروه به technologyadvice.com/social-collaboration-software نگاه کنید.

همانطور که قبلاً گفته شد، صدها یا شاید هزاران محصول نرم افزاری مشارکت و همکاری وجود دارد. علاوه بر این، قابلیت‌های آنها دائماً در حال تغییر است. با این فرض که موضوع اصلی مورد توجه ما، پشتیبانی تصمیم گیری است ما تنها به معرفی نمونه‌ی کوچکی از این ابزارها می‌پردازیم. ما از دسته بندی و مثال دکتر تایم و فهرست سال ۲۰۱۸ استفاده می‌کنیم (دیگنثو ۲۰۱۸).

- ابزارهای ارتباطی: یامر^۸ (همکاری اجتماعی)، اسلک، اسکایپ، گوگل هنگ آوتز، گو تو میتینگ.
- ابزارهای طراحی: اینویژن^۹، مورال^{۱۰}، رد پن^{۱۱}، لوگو میکر^{۱۲}
- ابزارهای مستندسازی: آفیس آنلاین، گوگل داکز^{۱۳}، زوهو
- ابزارهای به اشتراک گذاری فایل: گوگل درایو، دراپ باکس، باکس
- ابزارهای مدیریت پروژه: آسانا، پودیو^{۱۴}، ترلو^{۱۵}، وُرک فلو مَکس^{۱۶}، ابزار کانبان^{۱۷}
- ابزارهای نرم افزاری: جیت هاب^{۱۸}، یوزراسنپ^{۱۹}
- ابزارهای گردش کار: اینتگریتی^{۲۰}، بی پی لاجیکس^{۲۱}

سایر ابزارهایی که از ارتباطات و/یا همکاری پشتیبانی می کنند

نوت جوی^{۲۲} (ایجاد یادداشت های مشترک برای تیم)

1 Zimbra
2 SamePage
3 Zoho
4 Asana
5 Jive
6 Chatter
7 Social Tables
8 Yammer
9 InVision
10 Mural
11 Red Pen
12 Logo Maker
13 Google Docs
14 Podio
15 Trello
16 WorkFlowMax
17 Kanban Tool
18 GibHub
19 Usersnap
20 Integrity
21 BP Logix
22 Notejoy

کاهوتز^۱ (ذینفعان را گردهم آورده و جوامع مورد علاقه را تشکیل می دهد)
ناوبریج^۲ (ارتباطات تیمی، توانایی مشاهده‌ی شرکت کنندگان)
فضای کاری واک ابوت^۳ (یک دفتر مجازی سه بعدی برای تیم های راه دور)
ریل تایم بورد^۴ (یک مشارکت و همکاری دیداری سازمانی)
کوئورا^۵ (مکانی محبوب برای ارسال پرسش ها برای جمع مردم)
پینترست (یک فضای کاری تجارت الکترونیک را فراهم می کند که می توان متن و تصاویر با موضوعات منتخب را گردآوری کرد)
آی بی ام کانکشن کلوز^۶ (مجموعه ابزار جامع برای برقراری ارتباطات و همکاری)
اسکدا^۷ (فضای کاری را برای همکاری‌ها زمانبندی می کند)
زینک^۸ (یک ابزار مشارکت و همکاری اجتماعی)
اسکریب لر^۹ (یک اتاق همکاری آنلاین برای طوفان ذهنی مجازی)
کولوکیا^{۱۰} (یک پلتفرم یادگیری ماشین برای گردش کار)
برای جزئیات بیشتر به استوارد^{۱۱} (۲۰۱۷) نگاه کنید.

سوالاتی برای مرور بخش ۱۱.۴

۱۹. گروه‌افزار را تعریف کنید.
۲۰. ابزارهای گروه‌افزار مهم را نام برده و آنها را به انواع همزمان و غیرهمزمان تقسیم کنید.
۲۱. ابزارهای خاص برای کنفرانس وبی و قابلیت‌های آنها را شناسایی کنید.
۲۲. گردش کار مشترک را توضیح دهید.
۲۳. فضای کار مشارکتی چیست؟ چه مزایایی دارد؟
۲۴. مشارکت و همکاری اجتماعی را توضیح دهید.

تصمیمات عمدتاً در جلسات گرفته می‌شوند که برخی از آنها برای اتخاذ یک تصمیم خاص و یکباره برگزار می‌شوند. به عنوان مثال، مدیران در جلسات سهامداران برگزیده می‌شوند، سازمان‌ها در جلسات بودجه تخصیص می‌دهند، شهرداری‌ها برای تصمیم‌گیری

¹ Kahootz

² Nowbridge

³ Walkabout Workapce

⁴ RealtimeBoard

⁵ Quora

⁶ IBM Connection Closed

⁷ Skedda

⁸ Zinc

⁹ Scribblar

¹⁰ Collokia

¹¹ Steward

جهت استخدام افراد برای سمت‌های ارشد در جلسات تصمیم می‌گیرند و دولت فدرال ایالات متحده به صورت دوره‌ای برای تعیین نرخ بهره کوتاه مدت جلسه برگزار می‌کند. برخی از این تصمیمات پیچیده هستند؛ برخی دیگر می‌توانند بحث برانگیز باشند مثل تخصیص منابع توسط یک دولت شهری. اختلال عملکردهای فرایندی در چنین موقعیت‌هایی می‌توانند بزرگ باشند. از این رو، برای پیشگیری از این مشکلات پیشنهاد شده است که از پشتیبانی رایانه‌ای استفاده شود. این سیستم‌های پشتیبانی رایانه‌ای در ادبیات موضوع و تحت نام‌های متفاوتی از جمله سیستم‌های پشتیبان تصمیم گروهی (جی دی اس اس)، سیستم‌های پشتیبان گروه (جی اس اس)، کار مشترک با پشتیبانی رایانه‌ای (سی اس سی دلیو) و سیستم‌های میتینگ الکترونیک (ای‌ام‌اس) ظاهر شده‌اند. این سیستم‌ها موضوع مورد بحث ما در این بخش هستند. علاوه بر پشتیبانی از کل فرایندها، ابزارهایی وجود دارند که از یک یا چند فعالیت در فرایند تصمیم‌گیری گروهی پشتیبانی می‌کنند (مثلاً طوفان ذهنی).

طی دهه ۸۰، پژوهشگران متوجه شدند که پشتیبانی رایانه‌ای برای تصمیم‌گیری مدیریتی را باید به گروه‌ها گسترش دهند زیرا تصمیمات مهم سازمانی توسط گروه‌هایی از قبیل کمیته‌های اجرایی و نیروهای کارگروه ویژه اتخاذ می‌شوند. نتیجه‌ی اینکار شکل‌گیری متدولوژی سیستم‌های پشتیبان تصمیم گروهی^۱ بود. یک سیستم پشتیبان تصمیم گروهی (جی دی اس اس) یک سیستم رایانه‌ای تعاملی است که برای تسهیل حل مشکلات نیمه ساخت‌یافته و ساخت‌نیافته توسط گروهی از تصمیم‌گیرندگان بکار می‌رود. اهداف جی دی اس اس شامل بهبود بهره‌وری جلسات تصمیم‌گیری از طریق سرعت بخشیدن به فرایند تصمیم‌گیری و/یا افزایش کیفیت تصمیمات گرفته شده است.

ویژگی‌ها و قابلیت‌های مهم یک جی دی اس اس و ویژگی‌های جی دی اس اس به شرح زیر هستند:

- از فرایند تصمیم‌گیری گروهی و عمدتاً از طریق اتوماسیون زیرفرایندها و بکارگیری ابزارهای فناوری اطلاعات پشتیبانی می‌کند.
- یک سیستم اطلاعاتی است با هدف خاصی طراحی شده است نه صرفاً یک پیکربندی از اجزای سیستم‌های موجود و می‌توان آن را به گونه‌ای طراحی کرد که به یک نوع مسئله یا مجموعه‌ای از تصمیمات سازمانی در سطح گروهی رسیدگی کند.
- از تولید ایده‌ها، رفع تعارضات و آزادی بیان حمایت می‌کند. شامل مکانیزم‌های توکار برای نفی رفتارهای گروهی منفی از قبیل تعارض و تضادهای مخرب، تفکر گروهی و ... است.

نسل اول جی دی اس اس برای پشتیبانی از جلسات رو در رو در یک اتاق تصمیم^۲ طراحی شد. امروزه، پشتیبانی عمدتاً از طریق وب برای تیم‌های مجازی ارائه می‌شود. اعضای یک گروه می‌توانند در زمان یکسان یا در زمان‌های متفاوت با یکدیگر ملاقات کنند. جی دی اس اس به ویژه زمانی مفید است که تصمیمات بحث برانگیزی می‌بایست اتخاذ شوند (مثلاً تخصیص منابع). برنامه‌های جی دی اس اس به یک تسهیلگر برای یک مکان فیزیکی یا یک هماهنگ‌کننده یا رهبر برای جلسات مجازی آنلاین نیاز دارند.

¹ Group Decision Support Systems (GDSS)

² Decision Room

جی‌دی‌اس‌اس می‌تواند فرایند تصمیم‌گیری را به روش‌های گوناگونی بهبود بخشد. برای مثال، اگرچه برخی برنامه‌های کاربردی به گروه اجازه می‌دهند تا از روش‌ها و تکنیک‌های ساخت‌نیافته برای تولید ایده و فکر استفاده کنند ولی جی‌دی‌اس‌اس به طور کلی به فرایند برنامه‌ریزی جلسات ساختار می‌دهد. علاوه بر این، جی‌دی‌اس‌اس دسترسی سریع و آسان به اطلاعات برون‌سازمانی و ذخیره شده مورد نیاز برای تصمیم‌گیری را فراهم می‌کند. همچنین از پردازش موازی اطلاعات و تولید ایده توسط شرکت‌کنندگان پشتیبانی کرده و امکان بحث و گفتگوی رایانه‌ای غیرهمزمان را فراهم می‌کند. جی‌دی‌اس‌اس برگزاری جلسات بزرگ‌تر را امکان‌پذیر می‌کند که بدون آن مدیریت اینگونه جلسات ناممکن است و یک گروه بزرگ‌تر به معنای این است که اطلاعات، دانش و مهارت‌های کامل تری در جلسه ارائه خواهد شد. نهایتاً، رأی‌گیری گمنام با نتایج آنی انجام شده و تمامی اطلاعاتی که از این سیستم عبور می‌کنند را می‌توان برای تحلیل‌های آتی ثبت و ضبط نمود (حافظه‌ی سازمانی).

در طول زمان، روشن شد که پشتیبانی از تیم‌ها باید گسترده‌تر از چیزی باشد که جی‌دی‌اس‌اس در یک اتاق تصمیم‌پشتیبان نموده است. همچنین، روشن شد که آنچه واقعاً مورد نیاز بود، پشتیبانی از تیم‌های مجازی هم در وضعیت‌های مکان متفاوت/زمان یکسان و هم مکان متفاوت/زمان متفاوت بوده است. همچنین، روشن شد که تیم‌ها در اکثر موارد تصمیم‌گیری به پشتیبانی غیرمستقیم نیاز دارند (مثلاً کمک به جستجوی اطلاعات یا مشارکت و همکاری‌ها) تا اینکه برای فرایند تصمیم‌گیری به پشتیبانی مستقیم نیاز داشته باشند. اگرچه جی‌دی‌اس‌اس به پشتیبانی از تیم‌های مجازی گسترش پیدا کرد ولی نتوانست پاسخگوی سایر نیازها باشد. علاوه بر این، جی‌دی‌اس‌اس اولیه برای رسیدگی به تصمیمات بحث‌برانگیز و دارای تضاد طراحی شده بود. بنابراین، به نسل جدیدی از جی‌دی‌اس‌اس نیاز بود که از کارهای مشترک پشتیبانی کند. همانطور که در ادامه خواهیم دید، محصولاتی از قبیل استورم‌بورد این نیاز را برآورده می‌کنند.

دو گزینه برای استقرار فناوری جی‌دی‌اس‌اس وجود دارد: (۱) در یک اتاق تصمیم با هدف خاص و (۲) به صورت یک گروه افزار اینترنتی با برنامه‌های کلاینت که در هر مکان که اعضای گروه حضور دارند اجرا می‌شوند.

اتاق‌های تصمیم اولین جی‌دی‌اس‌اس در تاسیسات بسیار پرهزینه، سفارشی و تک‌منظوره به نام اتاق‌های تصمیم (یا اتاق‌های جلسات الکترونیک) نصب شد که رایانه‌های شخصی و یک صفحه نمایش بزرگ در جلوی هر اتاق وجود داشت. ایده‌ی اصلی این بود که فقط مدیران اجرایی و مدیران ارشد از این امکانات و تجهیزات گران‌قیمت استفاده کنند. یک نرم‌افزار در یک اتاق جلسه‌ی الکترونیکی معمولاً بر روی یک شبکه محلی اجرا می‌شود و این اتاق‌ها نسبتاً مبلمان و چیدمان لوکسی داشتند. اتاق‌های جلسات الکترونیک به شکل‌ها و اندازه‌های مختلفی طراحی و چیده می‌شدند. یک طراحی متداول به صورت یک اتاق مجهز به ۱۲ الی ۳۰ رایانه شخصی بهم متصل بود که معمولاً بر روی میز کار قرار می‌گرفتند. یک رایانه شخصی سرور به یک سیستم پروژکتور صفحه نمایش بزرگ متصل بوده و برای نمایش کار در ایستگاه‌های کاری به یک شبکه متصل بود و اطلاعات را از ایستگاه کاری تسهیلگرها گردآوری و تجمیع می‌کرد. در اتاق‌های جداگانه که مجهز به رایانه‌های شخصی متصل به سرور بودند و گروه‌های کوچکی می‌توانستند در آنجا مشاوره بگیرند و برخی اوقات در مجاورت اتاق تصمیم بودند. خروجی حاصل از این گروه‌های مشاوره را میشد بر روی یک صفحه نمایش بزرگ نمایش داد. تعداد اندکی از شرکت‌ها چنین اتاق‌هایی را به صورت روزانه اجاره می‌دادند. امروزه چند اتاق ارتقاء یافته هنوز وجود دارد که معمولاً به قیمت بالایی اجاره داده می‌شوند. :

گروه‌افزارهای مبتنی بر اینترنت از اواخر دهه ۹۰ میلادی، متداول‌ترین روش برای جی‌اس‌اس و جی‌دی‌اس‌اس به عنوان یک گروه‌افزار مبتنی بر اینترنت استفاده میشد که به اعضای گروه اجازه می‌داد تا از هر مکان و در هر زمانی کار کنند (مثلاً وبکس، گو تو میتینگ، ادوبی کانکت، آی‌بی‌ام کانکشنز، مایکروسافت تیمز). این گروه‌افزار اغلب شامل کنفرانس صوتی و ویدئویی بود. دسترسی نسبتاً ارزان‌قیمت به گروه‌افزارها که در بخش ۱۱.۴ توضیح داده شد به همراه قدرت بالا و هزینه‌ی پایین رایانه‌ها و در دسترس بودن دستگاه‌های سیار بر جذابیت بسیار زیاد این نوع سیستم‌ها افزود.

فرایندی که در شکل ۱۱.۱ به تصویر کشیده شد را توسط انواع مختلف محصولات نرم‌افزاری می‌توان پشتیبانی کرد. در این بخش، ما یک مثال از این محصولات یعنی استورم‌بوردها^۱ را توضیح می‌دهیم که از بخش‌های مختلف این فرایند پشتیبانی می‌کند.

استورم‌بوردها **stormbard.com** از شکل‌های مختلف تصمیم‌گیری گروهی و طوفان ذهنی پشتیبانی می‌کند. توالی فعالیت‌های استورم‌بوردها در ادامه آورده شده است:

۱. تعریف مسئله و اهداف کاربران (می‌خواهند به چه چیزی دست پیدا کنند).
۲. تولید ایده‌ها از طریق طوفان ذهنی
۳. سازماندهی ایده‌ها در گروه‌هایی با سلائق مشابه، جستجوی الگوها و انتخاب ایده‌های قابل اجرا.
۴. همکاری و مشارکت، پالایش مفاهیم و ارزشیابی اهداف جلسه (با استفاده از معیارها).
۵. این نرم‌افزار کاربران را قادر می‌سازد تا ایده‌های پیشنهادی را با تمرکز بر معیارهای انتخاب اولویت‌بندی کنند. به تمامی شرکت‌کنندگان اجازه می‌دهد تا افکار خود را بیان کرده و تیم را به سمت انسجام و یکپارچگی هدایت می‌کند.
۶. فهرست کوتاهی از ایده‌های برتر را تهیه می‌کند.
۷. این نرم‌افزار بهترین ایده‌ها را برای پیاده‌سازی توصیه می‌کند.
۸. پیاده‌سازی و اجرای پروژه را طرح‌ریزی می‌کند.
۹. پروژه را مدیریت می‌کند.
۱۰. پیشرفت پروژه را به صورت دوره‌ای پایش و مرور می‌کند.

برای تماشای یک ویدئو نگاه کنید به youtube.com/watch?v=0buRzu4rhJs

ابزارهای گروه‌افزار جامع از جمله تینک‌تنک اگرچه ابزارهای نرم‌افزاری متداول دارای قابلیت‌های بسیاری برای پشتیبانی از تصمیم‌گیری و افزایش بهره‌وری در محل کار هستند مثل مایکروسافت آفیس ۳۶۵ بهتر است در مورد نرم‌افزارهای تخصصی که برخی از قابلیت‌های منحصر بفرد گروه‌افزارها را به تصویر می‌کشند نیز بیاموزیم. میتینگ‌روم^۲ یکی از اولین بسته‌های نرم‌افزاری

¹ Stormboard

² MeetingRoom

جامع برای جلسات الکترونیک زمان یکسان/مکان یکسان بود. محصول بعد از آن به نام گروه سیستمز آنلاین^۱ قابلیت‌های مشابهی را ارائه می‌کرد و در حالت غیرهمزمان بر روی وب اجرا می‌شد (میتینگ‌روم فقط بر روی شبکه LAN اجرا میشد). آخرین محصول گروه سیستمز، تینک‌تنک^۲ نام دارد که مجموعه‌ای از ابزارها است که فعالیت‌های گوناگون تصمیم‌گیری گروهی را تسهیل می‌کنند. به عنوان مثال، زمان سیکل طوفان ذهنی را کاهش می‌دهد. تینک‌تنک همکاری تیم‌های مجازی یا رو در رو را از طریق فرایندهای قابل سفارشی‌سازی در جهت دستیابی سریع‌تر و مؤثرتر به اهداف گروه‌ها بهبود می‌بخشد. تینک‌تنک قابلیت‌های زیر را ارائه می‌کند:

- می‌تواند همکاری، گردش کار، اولویت‌بندی و تجزیه و تحلیل تصمیم مؤثر را فراهم کند.
- طوفان فکری ناشناس آن برای تولید ایده‌ها و نظرات، راه ایده‌آلی برای جلب خلاقیت و تجربه شرکت‌کنندگان است.
- رابط کاربری پیشرفته وب ۲.۰ محصول تضمین می‌کند که شرکت‌کنندگان برای پیوستن به تیم به آموزش خاصی نیاز ندارند، بنابراین می‌توانند ۱۰۰ درصد روی حل مشکلات و تصمیم‌گیری تمرکز کنند.
- با تینک‌تنک، تمام دانش به اشتراک گذاشته شده توسط شرکت‌کنندگان در اسناد و صفحات گسترده ضبط و ذخیره می‌شود، به صورت خودکار به صورت جلسه جلسه تبدیل می‌شود و در پایان جلسه در دسترس همه شرکت‌کنندگان قرار می‌گیرد.

(thinktank.net/case-study)

در زیر دو نمونه از استفاده تینک‌تنک آورده شده است

- همکاری تحول آفرین بین شرکای زنجیره تامین را ممکن می‌سازد. از ابزارها و روش‌های اطلاعاتی جمعی برای پشتیبانی جلسات آنها استفاده می‌شد. شرکا در مورد چگونگی کاهش هزینه‌ها، سرعت بخشیدن به فرایندها و بهبود کارایی توافق کردند. در گذشته هیچ پیشرفتی در این زمینه‌ها وجود نداشت.
- دانشگاه نبراسکا و کالج قلب و عروق آمریکا با استفاده از ابزارها و روش‌های تینک‌تنک برای سازماندهی مجدد پرونده‌های سلامت الکترونیکی برای کمک به مشاوران پزشکی در صرفه‌جویی زمان با یکدیگر همکاری کردند. زمان ویزیت بیماران ۵ تا ۸ دقیقه کوتاه شد. پیشرفت‌های دیگری نیز به دست آمد. هم در زمینه‌ی مراقبت از بیمار و هم صرفه‌جویی مالی پیشرفت‌های بسیاری به دست آمد.

سایر پشتیبانی‌های تصمیم‌گیری لیستی از انواع دیگر پشتیبانی‌های ارائه شده توسط سیستم‌های هوشمند در ادامه آمده است

- استفاده از سیستم‌های دانش و محصولی به نام اکسپرت چویس سافت‌ور^۳ برای تصمیم‌گیری گروهی چند معیاره.
- یک روش تصمیم‌گیری گروهی میانجی برای مدیریت دارایی زیرساخت توسط یون و همکاران (۲۰۱۷) پیشنهاد شد.
- برای یک سیستم پشتیبانی تصمیم‌گیری گروهی در مدیریت لجستیک و زنجیره تامین، به یزدانی و همکاران (۲۰۱۷) مراجعه کنید

¹ GroupSystems OnLine

² ThinkTank

³ Expert Choice Software

یک فعالیت اصلی در تصمیم‌گیری گروهی، ایده‌پردازی است. طوفان ذهنی فرایندی برای تولید ایده‌های خلاقانه است. این شامل بحث‌های گروهی آزادانه برای حل مشکلات و تدوین استراتژی و تخصیص منابع است. ایده‌های مشارکت‌کنندگان توسط اعضا مورد بحث قرار می‌گیرد. تلاش برای تولید هر چه بیشتر ایده‌ها، صرف نظر از اینکه چقدر عجیب به نظر می‌رسند انجام می‌شود. ایده‌های تولید شده توسط گروه مورد بحث و ارزیابی قرار می‌گیرد. شواهدی وجود دارد که نشان می‌دهد گروه‌ها نه تنها ایده‌های بیشتر بلکه بهترین نیز تولید می‌کنند (مک ماهون و همکاران، ۲۰۱۶)^۱. طوفان ذهنی با مدیریت دستی دارای برخی از محدودیت‌های کار گروهی است که در بخش ۱۱.۲ توضیح داده شده است. بنابراین، اغلب استفاده از پشتیبانی رایانه‌ای توصیه می‌شود.

طوفان ذهنی با پشتیبانی رایانه‌ای برنامه‌های رایانه‌ای می‌توانند از فعالیت‌های مختلف طوفان ذهنی پشتیبانی کنند. پشتیبانی معمولاً برای طوفان ذهنی آنلاین، به صورت همزمان یا غیرهمزمان است. امیدواریم طوفان ذهنی الکترونیکی بسیاری از اختلال عملکردهای ذکر شده در فرایندها در بخش ۱۱.۲ را حذف کرده و به تولید بسیاری از ایده‌های جدید کمک کند. نرم افزار طوفان ذهنی می‌تواند به طور مستقل یا بخشی از یک بسته پشتیبانی عمومی باشد. ویژگی‌های اصلی بسته‌های نرم‌افزاری به شرح زیر است:

- خلق تعداد زیادی از ایده‌ها
- مشارکت گروه‌های بزرگ
- به‌روزآوری‌های بلادرنگ
- کدگذاری رنگی اطلاعات
- ویرایش مشارکتی
- طراحی جلسات طوفان ذهنی
- به اشتراک گذاری ایده‌ها
- مشارکت مردمی
- نگاشت ایده‌ها (مثلاً ایجاد نقشه‌های ذهنی)
- انتشار متن، ویدئو، اسناد و غیره
- طوفان ذهنی از راه دور
- ایجاد یک بایگانی الکترونیکی
- کاهش طفره‌های اجتماعی

محدودیت‌های اصلی پشتیبانی الکترونیکی از نرم‌افزارها موجب افزایش بار شناختی و ترس از بکارگیری فناوری جدید و نیاز به کمک‌های فنی شده است.

شرکت‌هایی که به ارائه خدمات طوفان ذهنی آنلاین و پشتیبانی از کار گروهی می‌پردازند

برخی شرکت‌ها و خدمات و پشتیبانی‌هایی که ارائه می‌دهند به شرح زیر می‌باشند:

¹ McMahan et al.

- ایزی تاکس میتینگز^۱: یک ابزار مبتنی بر ابر برای طوفان ذهنی و به اشتراک گذاری ایده‌ها.
- بابل.آس^۲: یک ماشین تفکر مجازی که نمایش گرافیکی از ایده‌ها و مفاهیم را ارائه کرده به تولید ایده‌ها کمک کرده و همپوشانی بین ایده‌ها و افکار را نشان می‌دهد (به صورت دیداری و رنگی).
- مایندومو^۳: ابزاری برای همکاری آنی که دارای قابلیت چت می‌باشد.
- مورال^۴: ابزاری که امکان گردآوری و مرتب‌سازی ایده‌ها را در قالب فایل‌های رسانه‌ای غنی فراهم می‌آورد. به صورت یک پین‌برد طراحی شده است که شرکت‌کنندگان را دعوت به حضور در جلسه می‌کند.
- آی‌مایندکیو^۵: سرویس مبتنی بر ابر که امکان ایجاد نقشه‌های ذهنی و دیاگرام‌های پایه را فراهم می‌آورد.

برای یک ارزشیابی از ۲۸ ابزار طوفان ذهنی آنلاین نگاه کنید به blog.lucidmeetings.com/blog/25-tools-for-online-brainstorming-and-decision-making-in-meetings/.

هوش مصنوعی از طوفان ذهنی پشتیبانی می‌کند

در فصل ۱۲ ما به معرفی بکارگیری بات‌ها می‌پردازیم. برخی از نرم‌افزارها به کاربران اجازه می‌دهند تا یک بات را ایجاد کرده و منتشر کنند که این بات نمایانگر فردی است که به طور بی‌نام می‌تواند با آن ارتباط برقرار کرد. از هوش مصنوعی نیز می‌توان برای تشخیص الگو و شناسایی ایده‌هایی که مشابه یکدیگر هستند استفاده کرد. هوش مصنوعی همچنین در جمع‌سپاری کاربرد دارد که کاربرد گسترده‌ای برای تولید ایده‌ها و رأی‌گیری دارد.

یک سیستم پشتیبان گروهی (جی‌اس‌اس) که قبلاً توضیح داده شد هر ترکیبی از سخت‌افزار و نرم‌افزار است که موجب ارتقاء کار گروهی شود. جی‌اس‌اس یک عبارت عمومی است که در بر گیرنده‌ی تمامی شکل‌های رایانش ارتباطی و مشارکتی است. جی‌اس‌اس پس از اینکه پژوهشگران فناوری پی بردند که فناوری را می‌توان برای پشتیبانی از بسیاری از فعالیت‌هایی که در جلسات رو در رو رخ می‌دهد توسعه داد (مثلاً تولید ایده، رتبه‌بندی بی‌نام). همچنین، تمرکز به جای به حداقل رساندن تعارض‌ها بر مشارکت و همکاری بود.

یک جی‌اس‌اس کامل یک نرم‌افزار سیستم اطلاعاتی با طراحی ویژه است ولی امروزه قابلیت‌های ویژه‌ی آن در ابزارهای استاندارد بهره‌وری فناوری اطلاعات گنجانده شده‌اند. برای مثال، مایکروسافت آفیس ۳۶۵ شامل مایکروسافت تیمز است. همچنین ابزارهایی برای کنفرانس‌های وبی دارد. همچنین محصولات تجاری بسیاری پشتیبانی از یک یا دو جنبه‌های کار تیمی (مثلاً کنفرانس ویدئویی، تولید ایده، به اشتراک گذاری صفحه نمایش، ویکی‌ها) توسعه یافته‌اند.

¹ eZ Talks Meetings

² Bubble.us.

³ Mindomo

⁴ Mural

⁵ iMindQ

جی‌اس‌اس چگونه کار گروهی را بهبود می‌بخشد هدف جی‌اس‌اس پشتیبانی از شرکت‌کنندگان در کار گروهی برای بهبود بهره‌وری و اثربخشی جلسات از طریق روان‌سازی و سرعت‌دهی به فرایند تصمیم‌گیری و/یا بهبود کیفیت نتایج است. جی‌اس‌اس تلاش می‌کند تا بهره‌ی حاصل از فرایند و کارها را افزایش داده و اتلاف آنها را کاهش دهد. در مجموع، جی‌اس‌اس در انجام اینکار موفق بوده است. بهبود از طریق ارائه‌ی پشتیبانی به اعضای گروه برای تولید و تبادل ایده‌ها، نظرات و اولویت‌ها بدست می‌آید. ویژگی‌های خاص از قبیل توانایی شرکت‌کنندگان در یک گروه برای همکاری همزمان بر روی یک وظیفه یا فعالیت (مثلاً تولید ایده یا رأی‌گیری) و بی‌نام بودن موجب بهبود می‌شود. در ادامه برخی از فعالیت‌های خاص پشتیبانی جی‌اس‌اس آورده شده است:

- پشتیبانی موازی از پردازش اطلاعات و تولید ایده‌ها (ایده‌پردازی طوفان ذهنی)
- ارائه‌ی اطلاعات، دانش و مهارت‌های کامل‌تر به شرکت‌کنندگان گروه‌های بزرگ
- اجازه دادن به گروه برای استفاده از تکنیک‌ها و روش‌های ساخت‌یافته و ساخت‌نیافته
- ارائه‌ی دسترسی سریع و آسان به اطلاعات برون سازمانی
- اجازه‌ی برگزاری جلسات بحث و گفتگو به صورت موازی
- کمک به شرکت‌کنندگان در ترسیم تصویر بزرگ و کلان
- بی‌نام بودن که اجازه‌ی مشارکت افراد خجالتی را در جلسات می‌دهد
- فراهم کردن اقداماتی که از کنترل جلسات توسط افراد تهاجمی پیشگیری می‌کند
- ارائه روش‌های متعدد برای مشارکت در یک رأی‌گیری آنی و بی‌نام
- ارائه‌ی ساختاری برای فرایند برنامه‌ریزی برای در مسیر نگهداشتن گروه
- امکان تعامل و همکاری همزمان چند کاربر (یعنی برگزاری کنفرانس).
- ضبط کلیه اطلاعات ارائه شده در یک جلسه (یعنی حافظه‌ی سازمانی).

برای مطالعه‌ی داستان‌های موفقیت جی‌اس‌اس نگاه کنید به نمونه پرونده‌های وب سایت‌های عرضه‌کنندگان. همانطور که در بسیاری از این موارد خواهید دید، رایانش مشارکتی منجر به بهبودهای چشمگیر فرایندها و صرفه‌جویی در هزینه‌ها شده است. توجه داشته باشید که فقط برخی از این قابلیت‌ها در یک بسته‌ی نرم‌افزاری توسط شرکت‌های عرضه‌کننده ارائه شده است.

سوالاتی برای مرور بخش ۱۱.۵

۱. جی‌دی‌اس‌اس را تعریف کرده و محدودیت‌های نرم‌افزارهای اولیه‌ی جی‌اس‌اس را نام ببرید.
۲. مزایای جی‌دی‌اس‌اس را نام ببرید
۳. مزایای فرایندی حاصل از جی‌دی‌اس‌اس را نام ببرید.
۴. اتاق تصمیم را تعریف کنید.
۵. جی‌اس‌اس مبتنی بر وب را توضیح دهید.
۶. توضیح دهید که جی‌دی‌اس‌اس چگونه از طوفان ذهنی و تولید ایده پشتیبانی می‌کند.

هوش جمعی (سی‌آی)^۱ به کل هوش یک گروه اشاره دارد. همچنین به خرد یا دانش جمع اشاره می‌کند. افراد حاضر در یک گروه از دانش و مهارت‌هایشان برای حل مسائل و ارائه‌ی ایده‌ها و بینش‌های جدید استفاده می‌کنند. مزایای اصلی آن توانایی حل مسائل پیچیده و/یا طراحی محصولات و خدمات جدیدی است که از این نوآوری‌ها نشأت می‌گیرند. یکی از مراکز مهم پژوهشی در خصوص هوش جمعی (سی‌آی)، مرکز هوش جمعی ام‌آی‌تی^۲ (cci.mit.edu) است. یکی از جنبه‌های مهم مطالعاتی سی‌سی‌آی این است که افراد و رایانه‌ها چگونه می‌توانند با یکدیگر همکاری کنند تا تیم‌ها بتوانند از نوآوری بیشتری نسبت به تک افراد، گروه‌ها یا رایانه‌ها برخوردار شوند. سی‌آی در چندین رشته از جامعه‌شناسی گرفته تا علوم سیاسی کاربرد دارد. توجه ما در اینجا بر ارتباط سی‌آی با تصمیم‌گیری رایانه‌ای است. ما در این بخش و بخش ۱۱.۷ (موضوع جمع‌سپاری) به این موضوع می‌پردازیم. در بخش ۱۱.۸ به معرفی هوش ازدحامی که یکی از کاربردهای هوش جمعی است خواهیم پرداخت. برای اطلاع از مزایای سی‌آی به 50Minutes.com (۲۰۱۷) مراجعه کنید.

انواع هوش جمعی یک روش برای دسته‌بندی هوش جمعی این است که آن را به سه حوزه‌ی مهم کاربردی تقسیم کنیم: شناخت، همکاری و هماهنگی. هر یک از اینها را می‌توان مجدداً به زیرگروه‌ها تقسیم کرد. برای یک مرور کلی به هوش جمعی در ویکیپدیا نگاه کنید. توجه ما در اینجا بر کاربردهایی است که به واسطه‌ی آنها هم‌افزایی گروهی به حل مسائل و تصمیم‌گیری کمک می‌کند. افراد دانش و تجربه‌ی خود را به اشتراک می‌گذارند و تعاملات گروهی و پشتیبانی رایانه‌ای به اتخاذ تصمیمات بهتر کمک می‌کند.

توماس دابلو مالون، بنیان‌گذار و مدیر سی‌سی‌سی‌آی در ام‌آی‌تی، سی‌آی را یک چتر گسترده می‌داند. هوش جمعی از نظر وی «گروه‌هایی از افراد هستند که به طور جمعی به روش‌هایی که هوشمند به نظر می‌رسد عمل می‌کنند.» کار سی‌سی‌آی که به نام edge.org/conversation/Thomas_w__malone ^۳ شناخته می‌شود را می‌توانید در ویدئوی ۳۱:۴۵ دقیقه‌ای در آدرس edge.org/conversation/Thomas_w__malone مشاهده کنید.

هوش جمعی را به کمک ابزارها و پلتفرم‌های بسیاری که در بخش‌های ۱۱.۰۴ و ۱۱.۵ توضیح داده شده‌اند می‌توان پشتیبانی کرد. علاوه بر این، اینترنت، اینترنت و اینترنت اشیا می‌تواند نقش مهمی در تسهیل سی‌آی از طریق توانمندسازی افراد برای به اشتراک گذاری دانش و ایده‌ها ایفا کند.

¹ Collective Intelligence (CI)

² MIT Center for Collective Intelligence (CCI)

³ Edge

بنیاد کارنجی^۱ به دنبال روش‌هایی برای این بود تا افراد بتوانند برای سرعت بخشیدن به بهبودها و به اشتراک‌گذاری داده‌ها و آموخته‌ها در سرتاسر شبکه‌های خود با یکدیگر همکاری کنند. راه‌حل این کار یک فضای کاری آنلاین به نام کارنجی هاب^۲ بود که به عنوان یک نقطه‌ی دسترسی به منابع و درگیرسازی افراد در کار گروهی و مشارکت و همکاری شناخته شد. این هاب از چندین محصول نرم‌افزاری که برخی از آنها در بخش ۱۱.۴ توضیح داده شدند مثل گوگل درایو برای ایجاد فضای کاری مشارکتی استفاده می‌کند. جنبه‌های مهم پروژه‌ی کارنجی کالکشن اینتلیجنس^۳ به شرح زیر هستند:

۱.

۲.

۳.

۴.

طبق بریج‌واتر (۲۰۱۸)، دولت‌ها از اینترنت اشیاء برای پشتیبانی از تصمیم‌گیری و سیاست‌گذاری استفاده می‌کنند. دولت‌ها سعی می‌کنند تا دانش و اطلاعات را از مردم گردآوری کرده و اینکار را به طور فزاینده‌ای از طریق اینترنت اشیاء انجام می‌دهند. بریج‌واتر اشاره می‌کند که دولت‌های امارات متحده عربی از اینترنت اشیاء برای ارتقاء تصمیم‌گیری‌های عمومی استفاده می‌کنند. سیستم‌های اینترنت اشیاء ایده‌ها و خواسته‌های شهروندان را گردآوری می‌کند. پلتفرم هوش جمعی اجازه می‌دهد تا گروه‌های ذینفع تعریف شده را تا سطوح بسیار پایین هدف قرار داد. طرح‌های املاک و مستغلات تحت تأثیر نظرات ساکنانی است که در مجاورت‌های طرح‌های توسعه و عمرانی مذکور قرار دارند. پروژه‌ی شهرهای هوشمند این کشور با سی‌آی تلفیق شده است (فصل ۱۳). علاوه بر اینترنت اشیاء، فعالیت‌هایی در شبکه‌ها و سی‌آی وجود دارند که در مورد کاربردی ۱۱.۱ نشان داده شده‌اند.

¹ Carengie Foundation

² Carengie Hub

³ Carengie Collection Intelligence

⁴ Thorn and Huang

مقدمه

مدیریت آب یکی از مهم‌ترین چالش‌های هر جامعه محسوب می‌شود. به طور کلی، تقاضا برای آب به سرعت در حال افزایش است در حالیکه منابع عرضه‌ی آب (مثلاً به دلیل آلودگی) می‌تواند کاهش یابد. مدیریت آب نیازمند مشارکت ذینفعان بسیاری از مصرف‌کنندگان و عرضه‌کنندگان به دولت‌های محلی گرفته تا کارشناسان بهداشت آب است. ذینفعان باید با یکدیگر همکاری کنند. هدف در اینجا استفاده و حفظ مسئولانه‌ی آب است. اداره حسابداری پس‌دابیوسی گزارش 150CO47 به نام «همکاری: حفظ آب طریق مشارکت مؤثر» را در

pwc.com/hu/hu/kiadvanyok/assets/pdf/pwc_water_collaboration.pdf منتشر کرد. این گزارش به شرح مشکل و مزایا و ریسک‌های آن می‌پردازد. در این گزارش دیدگاه‌های ذینفعان مختلف به اشتراک گذاشته شده، به شناسایی عوامل موفقیت همکاری‌ها پرداخته شده و عواملی را برای ارزشیابی راه‌حل‌های جایگزین برای مسئله مدیریت آب ارائه می‌کند. یک چارچوب جالب، مدلسازی مشارکتی که در دانشگاه ایالتی اورگان با همکاری دانشگاه پوردو-دانشگاه ایندیانا توسعه یافته است.

چالش

برنامه‌ریزی و مدیریت فعالیت‌های حفظ آب کار راحتی نیستند. ایده‌ی اصلی در اینجا توسعه‌ی یک ابزار کاربرپسند است که تمامی ذینفعان بتوانند در این فعالیت‌ها مشارکت داشته باشند. مشارکت کلیه ذینفعان در بکارگیری دستورالعمل‌های علمی تدوین شده برای طراحی روش‌های حفظ آب امری ضروری است. در اینجا به معرفی برخی از الزامات این ابزار می‌پردازیم:

- ابزار باید تعاملی بوده و توسط انسان هدایت و عملیاتی شود.
- ابزار باید مبتنی بر وب و کاربر پسند باشد.
- هم افراد و هم گروه‌ها باید قادر به استفاده از آن باشند
- کاربران باید قادر به مشاهده و ارزشیابی راه‌حل‌های طراحی شده بر اساس معیارهای کمی و کیفی باشند.

راه‌حل: ری‌ستور

احیاء منابع آب‌خیز با استفاده از بهینه‌سازی زمانی-مکانی (ری‌ستور)^۱ یک ابزار مبتنی بر وب است که الزامات فوق را برآورده می‌کند. این یک ابزار مبتنی بر هوش مصنوعی و الگوریتم‌های بهینه‌سازی تحلیلی است. این الگوریتم به پردازش مدل‌های شبیه‌سازی پویا پرداخته و به کاربران اجازه می‌دهد تا مکان منابع آبی جدید را به صورت مکانی بهینه کنند. علاوه بر استفاده از مدل‌های شبیه‌سازی پویا، کاربران می‌توانند دیدگاه‌های شخصی و معیارهای کیفی خود را نیز اضافه کنند. ری‌ستور روش‌های جایگزین را ارائه می‌کند که کاربران می‌توانند در مورد آنها بحث کرده و آنها را ارزشیابی کنند. لحاظ نمودن اولویت‌های انسانی در راه‌حل‌های رایانه‌ای موجب شده تا راه‌حل‌های نهایی قابل پذیرش‌تر باشند. بخش هوش مصنوعی این پروژه شامل یادگیری ماشینی و جمع‌سپاری است تا اطلاعات را جمع افراد استخراج نماید. دلیل این همکاری

¹ Watershed Restoration Using Spatio-Temporal Optimization (WRESTORE)

مشترک این است که آب یک منبع ضروری است و نباید به طور متمرکز کنترل شود. فناوری‌های هوش مصنوعی به دموکراتیزه کردن مدیریت آب کمک کرده و در عین حال از قدرت و توان انسان و رایانه‌ها برای حل مشکلات عدیده‌ی مدیریت آب بهره می‌گیرد.

الگوریتم‌های یادگیری ماشینی از آنچه انسان‌ها انجام می‌دهند یاد می‌گیرند. بازخوردهای انسانی به هوش مصنوعی در شناسایی بهترین راه‌حل‌ها و راهبردها کمک می‌کند. بنابراین، انسان‌ها و ماشین‌ها برای حل مشکلات با یکدیگر تلفیق می‌شوند.

نتایج

توسعه‌دهندگان ری‌ستور در حال آزمایش و تجربه‌ی این فناوری در چندین مکان هستند و تا به امروز توانسته‌اند مشارکت کامل کلیه‌ی ذینفعان را بدست آورند. نتایج اولیه نشان می‌دهند که خلق ایده‌های نوآورانه برای توسعه‌ی منابع آبی و روش‌های توزیع آب به کمک ری‌ستور صرفه‌جویی چشمگیری در منابع آب به همراه داشته است.

سوالاتی برای مورد کاربردی ۱۱.۱

۱. از جمع‌سپاری برای یافتن اطلاعات از یک جمع مردم استفاده می‌شود. در این مورد چرا به چنین چیزی نیاز بود؟
۲. ری‌ستور چگونه به عنوان یک ابزار سی‌آی عمل می‌کند؟
۳. در مورد کنترل متمرکز و همکاری مشارکتی بحث کنید. مزایا و معایب هر کدام را ذکر کنید.
۴. چرا مدیریت منابع آبی کار دشواری است؟
۵. یک مدل بهینه‌سازی/شبیه‌سازی/هوش مصنوعی چگونه می‌تواند از کار گروهی در این مورد پشتیبانی کند؟

Sources: Compiled from Basco-Carrera et al. (2017), KTVZ.com (Channel 21, Oregon, March 21, 2018), and Babbar-Sebens et al. (2015).

برای چندین دهه، پژوهشگران به مطالعه‌ی رابطه‌ی بین سی‌آی و کار پرداخته‌اند. به عنوان مثال، داگ اینگنپرت یکی از پیشگامان سی‌آی توضیح می‌دهد که مردم چگونه برای پاسخ به یک چالش مشترک با یکدیگر همکاری کرده و چگونه می‌توانند از حافظه‌ی جمعی، ادراک، برنامه‌ریزی، استدلال و ... خود برای دستیابی به یک دانش قدرتمند استفاده کنند. از زمان کار اینگنپرت، تأثیر فناوری موجب افزایش سی‌آی سازمان‌ها و ایجاد جوامع دانشی مشارکتی شده است. به طور خلاصه، سی‌آی سعی دارد تا هوش انسانی را برای حل مشکلات اجتماعی و کاری افزایش دهد. این اساساً به این معنی است که سی‌آی اجازه‌ی مشارکت و مداخله‌ی بیشتری در تصمیم‌گیری سازمانی به افراد می‌دهد. در سی‌آی‌ام‌آی‌تی، تحقیقاتی در مورد اینکه چگونه افراد و رایانه‌ها می‌توانند برای بهبود کارها همکاری کنند انجام می‌شود (همچنین به بخش ۱۱.۹ مراجعه کنید) سی‌آی‌ام‌آی‌تی بر نقش شبکه‌ها از جمله اینترنت، اینترنت و اینترنت اشیا تمرکز دارد. محققان در آنجا دریافته‌اند که ساختار سازمان‌ها تمایل به تخت بودن و تفویض بیشتر تصمیمات به تیم‌ها دارند. تمامی اینها منجر به تمرکززدایی محل کار می‌شود. برای اطلاعات بیشتر به وبلاگ ام‌آی‌تی ۳ آوریل ۲۰۱۶ در executive.mit.edu/blog/willcollective-intelligence-change-the-way-we-work نگاه کنید. برای یک بحث جامع در مورد اینکه چگونه سی‌آی می‌تواند کل دنیا را تغییر دهد نگاه کنید به مولگان (۲۰۱۷).

قرار دادن افراد در گروه‌ها و اینکه از آنها انتظار داشته باشیم تا به کمک فناوری با یکدیگر همکاری کنند ممکن است آرزو به نظر برسد. محققان مدیریت و علوم رفتاری این موضوع را بررسی می‌کنند که چگونه افراد را به همکاری در گروه‌ها وادار کنند. آنچه که برخی آن را هوش مشارکتی می‌نامند، کولمن (۲۰۱۱) تصریح می‌کند هوش مشارکتی از ۱۰ جزء تشکیل شده است: (۱) تمایل به اشتراک گذاری، (۲) دانستن نحوه اشتراک گذاری، (۳) تمایل به همکاری، (۴) دانستن اینکه چه چیزی را به اشتراک بگذارید (۵) دانستن چگونگی ایجاد اعتماد، (۶) درک پویایی تیم، (۷) استفاده از هاب‌های صحیح برای شبکه، (۸) راهنمایی و مربیگری درست، (۹) پذیرش ایده‌های جدید، و (۱۰) استفاده از ابزار رایانه و فناوری. لیست مشابهی در thebalancecareers.com/collaboration-skills-with-examples-2059686 ارائه شده است. تکنیک‌ها و ابزارهای رایانه‌ای عوامل توانمندسازی برای برقراری ارتباطات، همکاری و درک افراد از یکدیگر هستند.

اعضای تیم‌ها و گروه‌ها به ارائه ایده‌ها و بینش‌ها می‌پردازند. سازمان‌ها برای پیشرفت باید از دانش افراد که بخشی از این دانش توسط هوش جمعی ایجاد می‌شوند بهره‌برداری کنند. یک روش برای اینکار توسط مطالعه‌ی که مؤسسه‌ی آی‌بی‌ام برای ارزش کسب‌وکار توسط هوش مشارکتی انجام داده است معرفی شده است. این مطالعه در www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-collective-intelligence.html در دسترس است. همچنین یک چکیده اجرایی در دسترس است. این مطالعه سه نکته‌ی مهم را نشان داده است:

۱. سی‌آی می‌تواند نتایج سازمانی را از طریق نگاشت درست دانش و تجربه‌ی اعضای گروه‌ها (مشتریان، شرکا و کارکنان) ارتقاء بخشد.
۲. هدف قرار دادن و مشارکت اعضای مناسب بسیار اهمیت دارد.
۳. سی‌آی باید به مسئله‌ی مقاومت شرکت‌کنندگان در برابر تغییر رسیدگی کند.

از همه اینها گذشته، آی‌بی‌ام نتیجه گرفت که «هوش مشارکتی یک منبع قدرتمند برای خلق ارزش با استفاده از تجربه و بینش تعداد بسیار زیادی از افراد در سرتاسر جهان است.»

سوالاتی برای مرور بخش ۱۱.۶

۵. هوش جمعی (سی‌آی) چیست؟
۶. مزایای اصلی هوش جمعی را نام ببرید.
۷. رایانه‌ها چگونه از هوش جمعی پشتیبانی می‌کنند؟
۸. هوش جمعی چگونه می‌تواند کار و زندگی ما را دستخوش تغییر کند؟
۹. سی‌آی چگونه می‌تواند بر ساختار و تصمیم‌گیری سازمانی تأثیر بگذارد؟

۱۰. مورد کارنجی توضیح داد که چگونه ابزارهای همکاری و مشارکت استاندارد می‌توانند یک زیرساخت هوش جمعی ایجاد کنند. مورد ری‌ستور یک چارچوب تحلیلی دلسازی را توضیح داد که ذینفعان می‌توانستند به کمک آن همکاری کنند. تشابه و تفاوت‌های بین این دو مورد چیست؟

۱۱. هوش مشارکتی را توضیح دهید.

۱۲. چگونه می‌توان از هوش جمعی برای کسب‌وکار خلق ارزش کرد؟

جمع سپاری به برون سپاری وظایف به گروه بزرگی از مردم (جمعیت) اشاره دارد. یکی از دلایل اصلی انجام این کار، پتانسیل خرد جمعی برای بهبود تصمیم‌گیری و کمک به حل مشکلات دشوار است (پاور (۲۰۱۴)). بنابراین، جمع سپاری را می‌توان به عنوان یک روش هوش جمعی در نظر گرفت. این بخش به سه بخش تقسیم می‌شود: الزامات جمع سپاری، جمع سپاری به عنوان مکانیزم پشتیبانی تصمیم، و اجرای جمع سپاری برای حل مشکلات.

جمع سپاری تعاریف متعددی دارد زیرا برای اهداف مختلفی در رشته‌های متعدد استفاده می‌شود. برای آموزش در مورد جمع سپاری و مثال‌ها، ویدیو را (۱۴:۵۱ دقیقه) در آدرس زیر مشاهده کنید [youtube.com/watch?v=IXhydxSSNOY](https://www.youtube.com/watch?v=IXhydxSSNOY). جمع سپاری به این معنی است که یک سازمان به چند دلیل در حال برون سپاری کار است: مهارت‌های لازم ممکن است در داخل در دسترس نباشد، سرعت اجرا مورد نیاز است، مشکلات بیش از حد پیچیده برای حل باشند، یا نیاز به نوآوری خاصی است.

چند مثال

- از سال ۲۰۰۵، دوریتوس^۱ یک مسابقه‌ی «Crash the Super Bowl» را برای ایجاد یک ویدئوی ۳۰ ثانیه‌ای برای مسابقه‌ی سوپر بول راه‌اندازی کرده است. این شرکت طی ۱۰ سال گذشته برای تیزرهای تبلیغاتی که توسط مردم تهیه شده است ۷ میلیون دلار جایزه پرداخت کرده است.
- ایر بی‌ان بی^۲ از ویدئوهای ارسالی توسط کاربران (۱۵ ثانیه‌ای) که محل سفرهایشان را توضیح می‌دهند استفاده می‌کند.
- آیدیا استورم^۳ شرکت دل (ideastorm.com) مشتریان را قادر می‌سازد تا در مورد ویژگی‌های آیدیا استورم که ترجیح می‌دهند رأی‌دهی کنند. دل از جوامع فنی مثل لینوکس (linux.org) بهره می‌برد. جمع ایده‌های خود را ارسال کرده و برخی اوقات اعضا به آنها رأی می‌دهند.
- پژوهشگران پروکتر آند گمبلر^۴ مسائل و مشکلات خود را در innocentive.com و ninesigma.com منتشر کرده و برای حل آنها جایزه نقدی در نظر می‌گیرند. آنها از سایر ارائه‌دهندگان خدمات جمع سپاری مثل yourencore.com استفاده می‌کنند.

¹ Doritos Inc.

² Airbnb

³ Ideas Storm

⁴ Procter & Gambler

- شرکت لگو^۱ پلتفرمی به نام لگو آیدیا^۲ دارد که از طریق آن کاربران می‌توانند ایده‌های خود را برای مجموعه لگوهای جدید ارائه کرده و در مورد ایده‌های ارسالی توسط جمع رأی‌گیری می‌شود. ایده‌های پذیرفته شده در صورتیکه به مرحله‌ی تجاری‌سازی برسند به نام ارائه‌دهندگان ثبت می‌شوند.
- پیسی کولا ایده‌های جدید در مورد طعم چپیس سیب‌زمینی را برای برند لی^۳ این شرکت از جمع مردم درخواست می‌کند. طی سال‌های گذشته این شرکت بیش از ۱۴ میلیون پیشنهاد دریافت کرده است. سهم این مشارکت‌ها در افزایش فروش حدود ۸ درصد برآورد شده است.
- شهرداری‌های کانادا نقشه‌های الکترونیکی بلادرنگ از شهرها را برای اطلاع‌رسانی در مورد نواحی پرریسک به دوچرخه‌سواران ایجاد کرده‌اند. کاربران می‌توانند از نقشه‌ها در زمان تصادف، سرقت دوچرخه، مخاطرات جاده‌ای و غیره استفاده کنند. برای جزئیات بیشتر به کیت (۲۰۱۸) نگاه کنید.
- سازمان‌های اطلاعاتی ایالات متحده از اطلاعات افراد عادی (جمع) برای پیش‌بینی رویدادهای آتی جهان از نتایج انتخابات گرفته تا جهت‌دهی به قیمت‌ها استفاده می‌کنند.
- شرکت هرسی^۴ برخی از راه‌حل‌های خود برای ارسال شکلات در آب و هوای گرم را از طریق جمع‌سپاری بدست آورده است. برای جزئیات بیشتر نگاه کنید به دیگنان^۵ (۲۰۱۶). جایزه در نظر گرفته شده ۲۵,۰۰۰ دلار بوده است.

این مثال‌ها برخی از مزایای جمع‌سپاری از قبیل افزایش عملکرد و سرعت و بهبود قابلیت‌های حل مسئله و نوآوری را نشان می‌دهند.

انواع اصلی جمع‌سپاری

هاو^۶ (۲۰۰۸) یکی از پیشگامان جمع‌سپاری کاربردهای جمع‌سپاری را به انواع (یا مدل‌های) زیر تقسیم کرده است:

۱. هوش (یا خرد) جمعی. مردم در جمع‌ها به حل مسئله پرداخته و بینش‌ها و ایده‌های جدیدی را ارائه می‌دهند که منجر به نوآوری در محصول، فرایند یا خدمت می‌شود.
۲. خلق جمعی. افراد به خلق انواع مختلف محتوا و به اشتراک‌گذاری آن با دیگران می‌پردازند (به رایگان یا با هزینه). محتوای خلق شده ممکن است برای حل مسئله، تبلیغات یا تجمیع دانش بکار برده شود. خلق محتوا را می‌توان با تفکیک وظایف بزرگ به بخش‌های کوچکتر انجام داد.
۳. رأی‌گیری جمعی. افراد عقاید، نظرات و امتیازات خود را در مورد ایده‌ها، محصولات یا خدمات داده و اطلاعات حاصل از ارزشیابی یا فیلترینگ را ارائه می‌کنند. یک مثال از رأی‌گیری در رقابت‌های امریکن آیدل است.
۴. پشتیبانی و تأمین مالی جمعی. مردم در فعالیت‌های پشتیبانی به دلایل اجتماعی و کاری مشارکت می‌کنند. مثل امور خیریه و تأمین مالی خرد در فعالیت‌های جدید.

¹ LEGO

² LEGO Idea

³ Lay

⁴ Hershey

⁵ Dignan

⁶ Howe

روش دیگر برای طبقه بندی جمع سپاری بر اساس نوع کاری است که انجام می دهد. چند نمونه با یک فروشنده جمع سپاری برای هر یک در زیر:

- طراحی لوگو – طراحی بیل
- حل مسئله - InnoCentive ، - NineSigma.
- نوآوری کسب و کار - چاردیکس
- نامهای تجاری – Name This
- طراحی محصول و ساخت – Pronto ERP
- پاکسازی داده ها – Amazom Mechanical Turk
- تست نرم افزار uTest
- رصد روند - TrendWatching
- تصاویر - Flickr Creative Commons

برای فهرست جامعی از جمع سپاری، هوش جمعی و شرکت های مرتبط، boardofinnovation.com را ببینید.

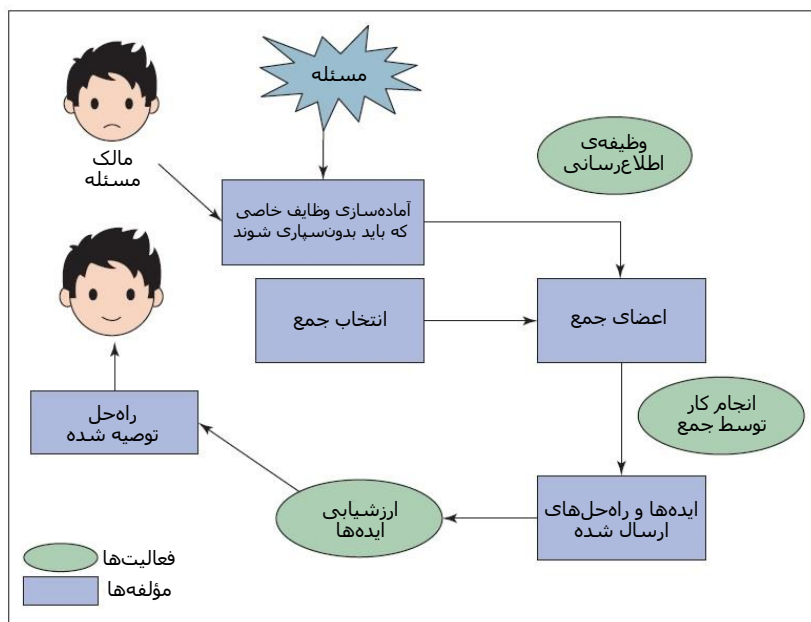
فرایند جمع سپاری

فرایند جمع سپاری در کاربردهای مختلف می تواند متفاوت باشد که به ماهیت مسئله ای که باید حل شود و روش بکار رفته بستگی دارد. با وجود این، گام های زیر در اکثر کاربردها جمع سپاری سازمانی مشترک هستند اگرچه جزئیات اجرا ممکن است متفاوت باشد. فرایند در شکل ۱۱.۳ به تصویر کشیده شده است.

۱. شناسایی مسئله و وظایفی که باید برون سپاری شوند.
۲. انتخاب جامعه ی هدف
۳. اطلاع رسانی وظیفه به جمع (یا یک جمع ناشناس از طریق فراخوان عمومی)
۴. درگیرسازی جمع در تحقق وظیفه (مثلاً ایده پردازی، حل مسئله)
۵. گردآوری محتوای تولید شده توسط کاربر
۶. ارزشیابی مطالب ارسالی توسط مدیریت
۷. انتخاب بهترین راه حل
۸. جبران خدمات جمع
۹. اجرای راه حل

توجه داشته باشید که ما فرایند را به صورت متوالی نشان داده ایم در صورتیکه ممکن است حلقه های برگشتی به گام های قبلی وجود داشته باشد.

اگرچه فعالیت‌های بالقوه‌ی بسیاری در جمع‌سپاری وجود دارد ولی فعالیت‌های اصلی شامل پشتیبانی از فرایند تصمیم‌گیری مدیریتی و/یا ارائه‌ی راه‌حل برای یک مسئله هستند. یک مسئله‌ی پیچیده که تصمیم‌گیری را برای یک تصمیم‌گیر یا یک گروه کوچک دشوار می‌کند را ممکن است بتوان به کمک یک جمع حل کرد که می‌تواند تعداد ایده‌های بیشتری را برای حل مسئله ارائه کنند. با وجود این، استفاده‌ی نامناسب از جمع‌سپاری می‌تواند نتایج منفی را به همراه داشته باشد (گران ۲۰۱۵). در خصوص چگونگی اجتناب از عواقب نامناسب جمع‌سپاری به بهانداری و همکاران ۲۰۱۸ نگاه کنید.



شکل ۱۱.۳ فرایند جمع‌سپاری

نقش جمع‌سپاری در تصمیم‌گیری

جمع‌ها می‌توانند ایده‌ها و نظرات را به نحوی مشارکتی یا رقابتی ارائه کنند. با وجود این، نقش جمع ممکن است در مراحل مختلف فرایند تصمیم‌گیری متفاوت باشد. ما ممکن است از یک جمع برای تصمیم‌گیری در خصوص نحوه‌ی پاسخگویی به یک رقیب یا کمک به تصمیم‌گیری در خصوص اینکه آیا طرح پیشنهادی مفید و مؤثر بوده است یا خیر استفاده کنیم. چپو و همکاران (۲۰۱۴) از مدل فرایند تصمیم‌گیری هربرت سایمون برای تشریح نقش‌های بالقوه‌ی یک جمع استفاده کردند. مدل سایمون از سه فاز اصلی پیش از اجرا تشکیل شده است: *اطلاعات* (گردآوری اطلاعات و به اشتراک‌گذاری با هدف حل مسئله یا بهره‌برداری از یک فرصت، شناسایی مسئله و تعیین میزان اهمیت مسئله)، *طراحی* (گردآوری ایده‌ها و راه‌حل‌های جایگزین) و *انتخاب* (ارزشیابی آلترناتیوهای تولید شده و توصیه یا انتخاب بهترین اقدام). جمع‌سپاری می‌تواند انواع مختلفی از پشتیبانی‌ها را برای این فرایند تصمیم‌گیری مدیریتی ارائه کند. اکثر کاربردهای جمع‌سپاری در فاز طراحی (مثلاً ایده‌پردازی) و در فاز انتخاب (رأی‌گیری) هستند. در برخی موارد، می‌توان در تمامی فازهای فرایند پشتیبانی ارائه کرد.

اگرچه با استفاده از فراخوان عمومی می‌توان اینکار را نسبتاً به آسانی انجام داد ولی آنهایی نیازمند حل مسائل دشوار هستند معمولاً تمایل دارند تا از افراد خبره برای حل مسائل کمک بگیرند. برای یک شرکت که به دنبال مساعدت در یافتن چنین افراد خبره‌ای به ویژه از بیرون از سازمان می‌باشد، می‌تواند از یک عرضه‌کننده‌ی شخص ثالث کمک بگیرد. این عرضه‌کنندگان دارای صدها، هزاران یا حتی میلیون‌ها حلال از پیش ثبت شده هستند که می‌توانند این وظیفه را مطابق مورد کاربردی ۱۱.۲ انجام دهند.

اینوستیو چگونه به جی‌اس‌کی در حل یک مسئله‌ی دشوار کمک کرد

مورد کاربردی ۱۱.۲

گلاکسو اسمیت کی لاین (جی‌اس‌کی)^۱ یک شرکت داروسازی/مراقب سلامت بین‌المللی در بریتانیا با بیش از ۱۰۰,۰۰۰ کارمند است. این شرکت به دنبال نوآوری است. با وجود این، علیرغم حضور بین‌المللی و بزرگی آن، مشکلات بسیاری دارد که برای حل آنها به تخصص برون‌سازمانی نیاز دارد.

مسئله

شرکت در خصوص یک فناوری تحول‌آفرین که نویدبخش درمان بسیاری از بیماری‌های دشوار بود به تحقیق پرداخت. شرکت به دنبال کشف این موضوع بود که از کدام بیماری به عنوان بستر آزمایشی برای درمان‌های نوآورانه استفاده کند. شرکت باید مطمئن میشد که انتخاب به گونه‌ای باشد که تمامی جنبه‌های روش درمان جدید را مورد بررسی قرار دهد. علیرغم بزرگی شرکت، جی‌اس‌کی برای پشتیبانی و چک کردن فعالیت‌های پژوهشی درون‌سازمانی خود به تخصص برون‌سازمانی نیاز داشت.

راه‌حل

جی‌اس‌کی تصمیم گرفت تا راه‌حل مسئله را به کمک شرکت اینوستیو (Innocentive.com) به افراد خبره و کارشناسان جمع‌سپاری نماید. اینوستیو یک شرکت جمع‌سپاری بین‌المللی در ایالات متحده است که چالش‌ها و مشکلات مشتریان خود مثل جی‌اس‌کی را دریافت می‌کند. این چالش‌ها به همراه جوایزه بالقوه در مرکز چالش اینوستیو برای حل‌کنندگان ارسال می‌شوند. حل‌کنندگانی که مایل هستند تا در این چالش شرکت کنند طبق دستورالعمل‌ها عمل کرده و ممکن است توافقاتی را امضا کنند. راه‌حل‌های ارسالی مورد ارزشیابی قرار گرفته و جوایز در اختیار برندگان قرار می‌گیرد.

موقعیت جی‌اس‌کی

در مجموع ۳۹۷ نفر در این چالش مشارکت کردند در حالیکه میزان پاداش بسیار ناچیز بود (۵,۰۰۰ دلار). حل‌کنندگان در چندین کشور مستقر بودند و ۶۶ راه‌حل پیشنهادی را ارسال کردند. کل فرایند به مدت ۷۵ روز به طور انجامید.

¹ GlaxoSmithKline (GSK)

نتایج

راه‌حل برنده به معرفی حوزه‌ای پرداخت که تیم‌های جی‌اس‌کی به آن توجه نکرده بودند. پیشنهاد دهنده یک شخص بلغاری بود که ایده‌ی وی مبتنی بر یک نشریه‌ی مکزیک‌ی بود. چندین پیشنهاد برنده نیز ایده‌های مفیدی را ارائه کردند. همچنین، این فرایند موجب همکاری بین تیم جی‌ای‌کی و پژوهشگران برنده گردید.

سوالاتی برای مورد کاربردی ۱۱.۲

۱. چرا جی‌اس‌کی تصمیم به جمع‌سپاری گرفت؟
۲. چرا شرکت از اینوسنتیو استفاده کرد؟
۳. در مورد ماهیت جهانی این پرونده اظهار نظر کنید.
۴. از این پرونده چه درس‌هایی آموخته‌اید؟
۵. به نظر شما چرا جایزه‌ی ناچیزی در حد ۵,۰۰۰ دلار کافی بود؟

Sources: Compiled from InnoCentive Inc. Case Study GlaxoSmithKline. Waltham, MA., GSK Corporate Information (gsk.com) and InnoCentive.com/our-solvers

جمع‌سپاری برای بازاریابی

بیش از ۱ میلیون مشتری در شرکت کراود تپ^۱ که پلتفرمی به نام سوزی^۲ را ارائه کرده ثبت شده‌اند که بازاریابان را قادر می‌سازد تا به انجام مطالعات جمع‌سپاری بپردازند.

سوالاتی برای مرور بخش ۱۱.۷

۴. جمع‌سپاری را تعریف کنید.
۵. فرایند جمع‌سپاری را توضیح دهید.
۶. مزایای اصلی این فناوری را نام ببرید.
۷. برخی از حوزه‌هایی که مناسب جمع‌سپاری هستند را نام ببرید.
۸. چرا برای جمع‌سپاری فرایند حل مسئله امکان دارد به یک عرضه‌کننده نیاز داشته باشید؟

هوش مصنوعی همانطور که در فصل ۲ دیدیم یک رشته‌ی متنوع است. فناوری‌های آن را می‌توان برای پشتیبانی از تصمیم‌گیری گروهی و مشارکت و همکاری تیمی بکار برد.

¹ Crowd Tap

² Suzy

هدف اصلی هوش مصنوعی، خودکارسازی تصمیم‌گیری و/یا پشتیبانی از فرایند تصمیم‌گیری است. این هدف برای تصمیماتی که توسط گروه‌ها گرفته می‌شوند نیز صادق است. با وجود این، ما نمی‌توانیم تصمیمی که توسط یک گروه گرفته شده است را خودکارسازی کنیم. کل کاری که می‌توانیم انجام دهیم پشتیبانی از برخی گام‌های فرایند تصمیم‌گیری گروهی است. یک محل منطقی برای شروع، شکل ۱۱.۱ است. ما به ارزیابی گام‌های مختلف فرایند پرداخته و خواهیم دید که کجا می‌توان از هوش مصنوعی استفاده کرد.

۱. آماده‌سازی جلسه. هوش مصنوعی برای یافتن زمان مناسب برگزاری جلسات بکار می‌رود. هوش مصنوعی می‌تواند به زمانبندی جلسه به نحوی که تمامی شرکت‌کنندگان بتوانند حضور داشته باشند کمک کند.
۲. شناسایی مسئله. از فناوری‌های هوش مصنوعی برای تشخیص الگوهایی که بتوانند حوزه‌های نیازمند توجه را شناسایی کنند استفاده می‌شود. از هوش مصنوعی می‌توان در انواع دیگر تحلیل‌ها برای شناسایی مسائل بالقوه استفاده کرد.
۳. ایده‌پردازی. هوش مصنوعی به خاطر عطشی که برای خلاقیت دارد معروف است. اعضای تیم می‌توانند خلاقیت خود را با بکارگیری پشتیبانی هوش مصنوعی افزایش دهند.
۴. سازماندهی ایده‌ها. از پردازش زبان طبیعی (ان‌ال‌پی) می‌توان برای مرتب‌سازی ایده‌ها و سازماندهی آنها برای بهبود ارزشیابی‌ها استفاده کرد.
۵. تعامل و همکاری گروهی. هوش مصنوعی می‌تواند ارتباطات و مشارکت و همکاری بین اعضای گروه را تسهیل نماید. این فعالیت نقش حیاتی در فرایند در دستیابی به اتفاق نظر همگان دارد. همچنین هوش مصنوعی ازدحامی برای افزایش تعاملات بین اعضای گروه طراحی شده تا بتوان خرد و دانش جمعی آنها را ارتقاء بخشید.
۶. پیش‌بینی‌ها. هوش مصنوعی از پیش‌بینی‌هایی که برای ارزیابی تأثیر ایده‌های تولید شده بر عملکرد و/یا تأثیرات در آینده مورد نیاز هستند پشتیبانی می‌کند. یادگیری ماشینی، یادگیری عمیق و هوش مصنوعی ازدحامی ابزارهای مناسبی در این حوزه هستند.
۷. گروه‌های چندملیتی. مشارکت و همکاری بین افرادی که در کشورهای مختلف قرار دارند در حال افزایش است. هوش مصنوعی تعامل آنی و لحظه‌ای بین گروه‌هایی که به زبان‌های متفاوت سخن می‌گویند را امکان‌پذیر می‌سازد.
۸. بات‌ها در پشتیبانی از جلسات مفید هستند. اعضای گروه‌ها ممکن است از الکسا و سایر بات‌ها مشاوره بگیرند. چت‌بات‌ها می‌توانند در لحظه به استعلام‌ها پاسخ دهند.
۹. سایر مشاورین. آی‌بی‌ام واتسون می‌تواند مشاوره‌های مفیدی را در حین برگزاری جلسات ارائه کرده و اطلاعاتی که توسط شرکت‌کنندگان و الکسا ارائه می‌شوند را تکمیل نماید.

در سال ۲۰۱۸، آمازون کام به دنبال مکانی برای دفتر مرکزی دوم خود بود. یک ربات به نام آیرا از شرکت ولز فارگو سکیوریتیز با استفاده از یادگیری عمیق مکان مورد نظر را بوستون پیش‌بینی کرد (یوریف ۲۰۱۸). (در زمان نگارش این فصل هنوز این تصمیم گرفته نشده بود).

برای یک رویکرد آکادمیک در خصوص چگونگی بهبود تصمیم‌گیری گروهی توسط هوش مصنوعی نگاه کنید به ژیا (۲۰۱۷).

سازمان‌ها امروزه به دنبال روش‌هایی برای افزایش و بهبود همکاری و مشارکت با کارکنان، شرکای تجاری و مشتریان خود هستند. برای دستیابی به یک بینش در خصوص اینکه هوش مصنوعی چگونه ممکن است مشارکت و همکاری‌ها تأثیر بگذارد، سیسکو سیستمز یک نظرسنجی جهانی را به نام AI Meets Collaboration را در خصوص تأثیر هوش مصنوعی از جمله بکارگیری دستیارهای مجازی در فضای کاری انجام داد. یافته‌های مهم این نظرسنجی به شرح زیر هستند:

۱. دستیارهای مجازی موجب افزایش بهره‌وری، خلاقیت و رضایت شغلی می‌شوند. بات‌ها نیز به کارکنان کمک می‌کنند تا بر وظایف با ارزش بالا تمرکز کنند.
 ۲. بات‌ها به عنوان بخشی از تیم‌های کارکنان پذیرفته شده‌اند.
 ۳. بات‌ها موجب بهبود کنفرانس‌ها می‌شوند و همچنین می‌توانند از جلسات یادداشت برداری کرده و آنها را زمانبندی کنند.
 ۴. هوش مصنوعی می‌تواند از تشخیص چهره برای ورود افراد به جلسات استفاده کند.
 ۵. ویژگی‌های شخصی بر احساسی که کارکنان نسبت به هوش مصنوعی در محل کار دارند تأثیر می‌گذارد.
 ۶. کارکنان به طور کلی دوست دارند تا از هوش مصنوعی در تیم‌هایشان استفاده کنند.
 ۷. امنیت یکی از دغدغه‌های اصلی هوش مصنوعی از جمله دستیارهای مجازی در تیم‌ها می‌باشد.
 ۸. ابزارهای اصلی هوش مصنوعی که بسیار مفید هستند شامل پردازش زبان طبیعی و پاسخ صوتی هستند؛ هوش مصنوعی همچنین می‌تواند عناوین اصلی جلسات را خلاصه‌برداری کرده و نیازهای شرکت‌کنندگان را درک کند. هوش مصنوعی می‌تواند از اهداف سازمانی و مهارت‌های کارگران آگاه شده و توصیه‌های متناسب با آنها را ارائه دهد.
- در خصوص چگونگی پشتیبانی جلسات مجازی با هوش مصنوعی توسط سیسکو سیستمز در محصولات پیشرفته‌شان به بینش فناوری ۱۱.۲ مراجعه کنید.

بینش فناوری ۱۱.۱ سیسکو چگونه همکاری‌ها را به هوش مصنوعی بهبود می‌بخشد

سیسکو سیستمز^۱ برای محصولات همکاری خود مانند اسپارک^۲ و وبکس معروف است. اولین قدم در معرفی هوش مصنوعی، به دست آوردن پلتفرم هوش مصنوعی مایند مِلد برای استفاده در محصولات همکاری سیسکو بود. هدف این پروژه بهبود تداخلات مکالمه‌ای برای هر برنامه یا دستگاهی بود تا کاربران بتوانند زمینه مکالمات را بهتر درک کنند. مایند مِلد از یادگیری ماشینی برای بهبود دقت ارتباطات صوتی و متنی استفاده می‌کند. برای انجام این کار، از پردازش زبان طبیعی و پنج نوع یادگیری ماشینی استفاده می‌کند. سیسکو همچنین در حال ادغام آی‌بی‌ام واتسون با محصولات همکاری سازمانی خود است. همانطور که ممکن است از فصل ۶ به خاطر بیاورید، واتسون یک مشاور قدرتمند است. ابزارهای همکاری هوش مصنوعی می‌توانند کارایی، سرعت تولید ایده و بهبود کیفیت تصمیمات اتخاذ شده توسط گروه‌ها را افزایش دهند. فناوری بهبود یافته سیسکو در اتاق‌های کنفرانس و هر جا دیگری استفاده خواهد شد. یکی از پروژه‌های بزرگ هوش مصنوعی، دستیار اسپارک است.

مونیکا^۳ برای پاسخ دادن به سوالات کاربران با استفاده از یادگیری ماشینی آموزش دیده است. علاوه بر این، کاربران می‌توانند از مونیکا^۴ برای تعامل با پلتفرم همکاری اسپارک با استفاده از دستورات زبان طبیعی استفاده کنند. این یک دستیار سازمانی مشابه الکسا و دستیار گوگل است (فصل ۱۲). مونیکا اولین دستیار صوتی آماده سازمانی در جهان است که به طور خاص برای پشتیبانی از جلسات طراحی شده است. این ربات دارای هوش مصنوعی مکالمه‌ای با دامنه عمیق است که قابلیت‌های شناختی را به آن اضافه می‌کند. مونیکا می‌تواند در چندین مرحله شکل ۱۱.۱ به کاربران کمک کند، مانند:

- سازماندهی جلسات
- ارائه اطلاعات به شرکت کنندگان قبل و در طول جلسات
- کنترل و راهبری دستگاه‌های اسپارک
- کمک به سازماندهی کنندگان در یافتن اتاق جلسه و رزرو آن
- کمک در اشتراک گذاری صفحات نمایش و راه‌اندازی یک تخته سفید
- یادداشت برداری از جلسه و سازماندهی آنها

در آینده‌ای نزدیک، مونیکا از فعالیت‌های داخلی و خارجی شرکت کنندگان مطلع خواهد شد و جلسات را با استفاده از این اطلاعات برنامه‌ریزی خواهد کرد. کارکردهای بیشتر برای پشتیبانی از مراحل فرایند در شکل ۱۱.۱ در آینده اضافه خواهند شد. برای اطلاعات بیشتر درباره دستیار، به [youtube.com/watch?v=80cFSEbR_6k](https://www.youtube.com/watch?v=80cFSEbR_6k) مراجعه کنید.

¹ Cisco Systems

² Spark

³ Monica

⁴ Monicait

نکته: سیسکو اسپارک با افزودن قابلیت‌های بیشتر هوش مصنوعی به وبکس تیمز تبدیل خواهد شد. علاوه بر این، جلسات وبکس شامل کنفرانس ویدئویی برای همکاری و سایر پشتیبانی‌ها برای جلسات خواهد بود

منابع: گوئکه (۲۰۱۷)، فینگان (۲۰۱۸)، گلدشتاین (۲۰۱۷).

عبارت هوش ازدحامی به رفتار جمعی سیستم‌های غیرمتمرکز چه طبیعی و چه مصنوعی اشاره دارد. اینگونه سیستم‌ها از اشیائی تشکیل شده‌اند که با یکدیگر و با محیط اطراف تعامل دارند. اقدامات ازدحامی به صورت مرکزی کنترل نمی‌شوند ولی منجر به یک رفتار هوشمند می‌شوند. در واقع، مثال‌های بسیاری از اینگونه رفتارها وجود دارد. مشاهده شده است که گروه‌های طبیعی هوش گروهی خود را با تشکیل ازدحامها تقویت می‌کنند. موجودات اجتماعی، از جمله انسان‌ها می‌توانند عملکرد تک تک اعضای خود را هنگام کار با یکدیگر به عنوان یک سیستم واحد بهبود بخشند. برخلاف حیوانات و سایر گونه‌هایی که تعامل بین اعضای گروه در آنها طبیعی است، مردم برای نمایش هوش جمعی نیاز به فناوری دارند. این مفهوم در مطالعات و پیاده‌سازی هوش مصنوعی و رباتیک استفاده می‌شود. عمده کاربردها در حوزه پیش‌بینی‌ها است. یک مثال از مطالعه‌ای در دانشگاه آکسفورد (بریتانیا) شامل پیش‌بینی نتایج تمام ۵۰ بازی فوتبال لیگ برتر انگلیس در طول پنج هفته بود. گروهی از داوران مستقل هنگام کار به تنهایی ۵۵ درصد دقت را به دست آوردند. با این حال، هنگام پیش‌بینی با استفاده از هوش مصنوعی ازدحامی، موفقیت پیش‌بینی آنها به ۷۲ درصد افزایش یافت (۳۱ درصد بهبود). بهبود مشابهی در چندین مطالعه دیگر ثبت شده است. علاوه بر بهبود دقت پیش‌بینی، مطالعات نشان می‌دهد که استفاده از هوش مصنوعی ازدحامی منجر به تصمیمات اخلاقی بیشتری نسبت به حالت انفرادی می‌شود.

فناوری هوش مصنوعی ازدحامی

هوش مصنوعی ازدحامی الگوریتم‌هایی را برای ارتباطات متقابل بین افرادی که ازدحام انسانی را ایجاد می‌کنند، ارائه می‌دهد. این ارتباطات موجب ادغام دانش، شهود، تجربه و خرد افراد در یک واحد هوش ازدحامی بهبودیافته می‌شود. نتایج هوش ازدحامی را می‌توانید در ارائه‌ی TED مشاهده کنید [Youtube.com/watch?v=Eu-RyZt_Uas](https://www.youtube.com/watch?v=Eu-RyZt_Uas). هوش مصنوعی ازدحامی توسط چندین شرکت شخص ثالث مورد استفاده قرار گرفته است (به عنوان مثال، یونانی‌موس ای‌آی^۱، همانطور که در مورد کاربردی ۱۱.۳ نشان داده شده است).

ایکس پرایز چگونه ویژنرینگ را بهینه‌سازی می‌کند

مورد کاربردی ۱۱.۳

ایکس پرایز یک سازمان غیرانتفاعی است که جوایز را از طریق مسابقاتی که برای ترویج نوآوری‌هایی که توان بالقوه‌ای برای تغییر جهان دارند در نظر می‌گیرد. کانال اصلی برای طراحی جوایزی که بزرگ‌ترین چالش‌های بشری را حل کند ویژنرینگ نام دارد که سعی دارد تا از قدرت و توان جمع جهانی برای توسعه‌ی راه‌حل‌هایی برای چالش‌های مهم استفاده کند. رویداد مهم این سازمان یک نشست سالانه است که جوایز در آن طراحی شده و پیشنهادات مورد ارزشیابی قرار می‌گیرند. افراد خبره در ایکس پرایز مفاهیم را توسعه داده و آنها را به مسابقات جایزه‌ای تبدیل می‌کنند. جوایز توسط سازمان‌های پیشگام اهدا می‌شود.

¹ Unanimous AI

به عنوان مثال، در سال ۲۰۱۸ آی‌بی‌ام واتسون یک جایزه‌ی ۵ میلیون دلاری را به نام « رویکردها و همکاری هوش مصنوعی » اهدا نمود. در این مسابقه ۱۲۴ تیم ثبت نام نمودند و ۶۲ تیم در دور دوم مسابقه در سال ۲۰۱۸ باقی ماندند. تیم‌ها برای خلق اهداف و راه‌حل‌های خودشان برای یک چالش بزرگ دعوت شده بودند.

مسئله

هر ساله، یک جلسه متشکل از ۲۵۰ عضو «ویژنرز سامیت آیدپیشن» از افراد خبره (کارآفرینان، سیاستمداران، دانشمندان و غیره) برای کشف و اولویت‌بندی موضوعات برای مسابقه ایکس‌پرایز دعوت شدند. یافتن مسائل تراز اول و مهم جهانی به دلیل تعداد متغیرهای زیاد می‌تواند کار دشواری باشد. تنها در طی چند روز، کارشناسان خبره باید از خرد جمعی خود برای توافق در خصوص چالش‌های مهم سال آینده‌ی ایکس‌پرایز استفاده کنند. روشی که برای پشتیبانی از تصمیم‌گروهی بکار می‌رود یک عامل حیاتی برای موفقیت محسوب می‌شود.

راه‌حل

در سال ۲۰۱۷ نشست سالانه برای تعیین اینکه چه چالشی برای سال ۲۰۱۸ در نظر گرفته شود، سازمان از پلتفرم هوش مصنوعی ازدحامی (متعلق به شرکت یونانیموس‌ای‌آی) استفاده کرد. چندین گروه کوچک که توسط الگوریتم‌های هوش مصنوعی مدیریت می‌شدند برای کشف موضوعات چالش برانگیز ایجاد شدند. مأموریت آنها کشف ایده‌ها و توافق بر روی راه‌حل ارجح بود. هدف استفاده از استعدادها و توان ذهنی شرکت کنندگان بود.

به عبارت دیگر، هدف این بود که از ویژگی تفکر جمعی هوش مصنوعی ازدحامی برای تولید هم‌افزایی هر گروه با الگوریتم‌های هوش مصنوعی که به عنوان میانجی عمل می‌کردند استفاده شود. بدین ترتیب، تصمیمات هوشمندانه‌تری توسط گروه‌ها در مقایسه با تک تک افراد گرفته می‌شد. گروه‌های مختلف به ارزیابی شش موضوع از پیش تعیین شده پرداختند: انرژی و زیرساخت، پتانسیل یادگیری انسانی، فضا و مرزهای جدید، محیط و گیاهان، جامعه‌ی مدنی و رفاه و سلامت. گروه‌ها از طریق طوفان ذهنی به بررسی مسائل پرداختند. سپس، هر شرکت کننده یک جدول ارزشیابی سفارشی ایجاد کرد. جداول توسط الگوریتم‌هایی ترکیب و تحلیل شدند.

هوش مصنوعی ازدحامی با بهینه‌سازی نقش و سهم هر یک از شرکت کنندگان جایگزین روش‌های رأی‌گیری متداول شد.

نتایج

استفاده از هوش مصنوعی ازدحامی منجر به نتایج زیر شد:

- پشتیبانی از تولید پاسخ‌های بهینه و پذیرش سریع پاسخ‌ها توسط شرکت کنندگان
- مشارکت کلیه شرکت کنندگان
- فراهم کردن یک سیستم رأی‌گیری بهتر نسبت به سال‌های گذشته

سوالاتی برای مورد کاربردی ۱۱.۳

۱. چرا بحث و گفتگوی گروهی در این مورد کاربردی موضوع پیچیده‌ای است؟

۲. چرا دستیابی به یک اتفاق آراء زمانیکه افراد خبره‌ی تراز اول درگیر هستند دشوارتر از زمانی است که افراد غیرخبره درگیر هستند؟
۳. سهم و نقش هوش مصنوعی ازدحامی در این مورد چه بوده است؟
۴. سیستم رأی‌گیری ساده را با رأی‌گیری هوش مصنوعی ازدحامی مقایسه کنید.

Sources: Compiled from Unanimous AI (2018), xprize.org, and xprize.org/about.

هوش مصنوعی ازدحامی برای پیش‌بینی‌ها

یونانی‌موس ای‌آی از هوش مصنوعی ازدحامی برای پیش‌بینی‌ها در موقعیت‌هایی که ارزیابی در آنها دشوار است استفاده کرد. مثال‌هایی از آن به شرح زیر می‌باشند:

- پیش‌بینی امتیازات کسب شده در سوپر بول شماره ۵۲
- پیش‌بینی برندگان بازی‌های فصل این‌فیل.
- پیش‌بینی چهار مهاجم برتر دربی کنتاکی در سال ۲۰۱۷
- پیش‌بینی دریافت‌کنندگان برتر جایزه اسکار سال ۲۰۱۸

سوالاتی برای مرور بخش ۱۱.۸

۵. ارتباط بین بکارگیری هوش مصنوعی در فعالیت‌های شکل ۱۱.۱ را توضیح دهید.
۶. در مورد روش‌های مختلفی که هوش مصنوعی می‌تواند مشارکت و همکاری گروهی را تسهیل کند توضیح دهید.
۷. هوش مصنوعی چگونه می‌تواند از ارزشیابی گروهی ایده‌ها پشتیبانی کند؟
۸. هوش مصنوعی چگونه می‌تواند از ایده‌پردازی پشتیبانی کند؟
۹. تشابه بین هوش مصنوعی ازدحامی با گونه‌های زنده در چیست؟
۱۰. چگونه از هوش مصنوعی ازدحامی می‌توان برای بهبود کار گروهی و پیش‌بینی‌های گروهی استفاده کرد؟

از زمان شروع انقلاب صنعتی، انسان‌ها و ماشین‌ها با یکدیگر همکاری کرده‌اند. تا اواخر سال‌های ۱۹۰۰، مشارکت و همکاری تنها در حوزه‌های ساخت و تولید بود. ولی از آن زمان به بعد، به دلیل پیشرفت‌های تکنولوژیکی و تغییرات در ماهیت کارها، همکاری انسان-ماشین به بسیاری از حوزه‌های دیگر از جمله کارهای ذهنی و شناختی و همکاری در کارهای مدیریتی و اجرایی تسری یافت. طبق نیلری (۲۰۱۷)، همکاری انسان و هوش مصنوعی آینده‌ها کارها را دستخوش تغییر خواهد کرد. انسان‌ها و ماشین‌ها می‌توانند به طرق مختلف همکاری کنند که به وظایفی که انجام می‌دهند بستگی دارد. همکاری با ربات‌ها در سناریوی ساخت و تولید یک انشعاب از مدل قدیمی‌تری است که در آن انسان‌ها و ربات‌ها برای کنترل و پایش تولید با یکدیگر همکاری می‌کردند و ربات‌ها کارهای فیزیکی را انجام می‌دادند که نیازمند سرعت، قدرت، دقت و توجه مداوم بود. ربات‌ها همچنین در محیط‌های

خطرناک به انجام کار می‌پردازند. به طور کلی، ربات‌ها مکمل قابلیت‌های انسان‌ها هستند. یک مثال از آن مراکز توزیع آمازون است که بیش از ۵۰,۰۰۰ ربات سیار به انجام کارهای گوناگون که عمدتاً شامل جابجایی مواد و کمک به تحقق سفارشات مشتریان است می‌پردازند. فناوری رباتیک امکان ارائه‌ی راه‌حل‌های کاملاً مشارکتی را فراهم می‌کند. برای جزئیات ویدئوی kuka.com/en-us/technologies/humanrobot-collaboration را تماشا کنید. سیستم کوکا امکان اجرای مشاغل پیچیده را به نحو مقرون به صرفه فراهم می‌کند.

یک سیستم همکاری انسان-ربات یومی نام دارد. برای مشاهده‌ی کارکرد این سیستم ویدئوی ۴:۳۸ دقیقه‌ای را در youtube.com/watch?v=2KfXY2SvImQ تماشا کنید. توجه داشته باشید که این ربات دو بازو دارد/

پیشرفت‌ها در هوش مصنوعی امکان خودکارسازی فعالیت‌های غیردستی را فراهم کرد. اگرچه برخی سیستم‌های هوشمند به طور کاملاً خودکار هستند ولی مثال‌های بسیار زیادی از همکاری انسان-ماشین در مشاغل شناختی وجود دارد (یعنی در بازاریابی و امور مالی). یک مثال دیگر در تصمیمات سرمایه‌گذاری است. فردی از رایانه برای سرمایه‌گذاری خود مشورت می‌خواهد و پس از دریافت مشاوره می‌تواند سوالات بیشتری را پرسیده و برخی از ورودی‌ها را تغییر دهد. تفاوت آن با گذشته در این است که امروزه رایانه‌ها (ماشین‌ها) می‌توانند توصیه‌های بسیار دقیق‌تری را با استفاده یادگیری ماشینی و یادگیری عمیق ارائه کنند. یک مثال دیگر از این نوع همکاری شامل تشخیص‌های پزشکی موقعیت‌های پیچیده است. به عنوان مثال، آی‌بی‌ام واتسون مشاوره‌های پزشکی را با مجوز پزشک و پرستاران برای بهبود مشاغل‌شان ارائه می‌کند. در واقع، رشته‌ی ارائه مشاوره‌ی ماشینی به انسان‌ها به سطوح بالایی دست یافته است. برای اطلاعات بیشتر در مورد افزایش قدرت مشارکت و همکاری هوش مصنوعی به کارتر (۲۰۱۷) نگاه کنید.

مشاغل مدیریت ارشد

یک وظیفه‌ی مهم و اصلی مدیران، تصمیم‌گیری است که به یکی از حوزه‌های کاربرد همکاری انسان-ماشین بدل شده است. استفاده از هوش مصنوعی و ابزارهای تحلیلی موجب بهبود چشمگیر تصمیم‌گیری شده است که در سرتاسر کتاب این موضوع به تصویر کشیده شد. برای یک مرور کلی به ولادوفسکی-برگر (۲۰۱۷) مراجعه کنید.

مک‌کنزی آند کامپانی و ام‌آی‌تی دو فعال اصلی در پژوهش همکاری بین مدیران و ماشین‌ها هستند. به عنوان مثال، دیوهرست آند ویلموت (۲۰۱۴) در خصوص افزایش استفاده از همکاری انسان-ماشین با استفاده از یادگیری عمیق گزارش داده‌اند. یک شرکت هنگ‌کنگی حتی از یک الگوریتم تصمیم‌گیری برای اعضای هیئت مدیره خود استفاده کرده است. شرکت‌ها از مشاوره‌های جمع‌سپاری برای پشتیبانی حل مسائل پیچیده استفاده می‌کنند.

زمانی در آینده، راه‌رفتن و صحبت کردن ربات‌های انسان‌نما که در زمان استراحت انسان‌ها با آنها معاشرت کنند امری عادی خواهد شد. زمانی، ربات‌ها به همکاران شناختی تبدیل شده و به افراد کمک می‌کنند تا بهره‌وری خود را افزایش دهند.

طبق تویی (۲۰۱۵)، مطالعه‌ای در یکی از کارخانجات بی‌ام‌دلیو نشان داد که همکاری انسان-ربات می‌تواند بهره‌ورتر از فعالیت‌ها انسان‌ها یا ربات‌ها به تنهایی باشد. همچنین این مطالعه نشان داد که همکاری بین این دو زمان بیکاری را تا ۸۵ درصد کاهش داده است. این بدان دلیل است که انسان‌ها و ماشین‌ها نقاط قوت یکدیگر را تقویت می‌کنند (مار ۲۰۱۷)^۱.

چالش‌های زیر را باید در نظر گرفت:

- طراحی یک تیم انسان-ماشین که بر نقاط قوت هر شریک تأکید و تمرکز کند.
- تبادل اطلاعات بین انسان‌ها و ربات‌ها
- آماده‌سازی کارکنان شرکت در تمامی دپارتمان‌ها برای همکاری (مار ۲۰۱۷)
- تغییر فرایندهای کسب‌وکار برای امکان همکاری انسان-ماشین (موران ۲۰۱۸)^۲.
- حصول اطمینان از ایمنی ربات‌ها و کارکنانی که با یکدیگر همکاری می‌کنند.

فناوری‌هایی که از ربات‌ها در نقش همکاران پشتیبانی می‌کنند.

یوریف (۲۰۱۸b) مثال‌های زیر را از تسهیل یا در نظر گرفتن ربات‌ها در نقش همکاران ارائه کرده است.

۱. از واقعیت مجازی می‌توان به عنوان یک ابزار آموزشی قدرتمند استفاده کرد (مثلاً برای ایمنی).
۲. در ژاپن یک ربات با یک آژانس تبلیغاتی برای ایده‌پردازی کار می‌کند.
۳. یک ربات می‌تواند رئیس شما باشد.
۴. ربات‌ها در خطوط مونتاژ بخشی از کارها را انجام داده و می‌توانند کیفیت کار را به همراه انسان‌ها چک کنند.
۵. ابزارهای هوش مصنوعی به اندازه‌گیری فشار خون و حجم عضلات قلب در چند ثانیه می‌پردازند (به جای چند دقیقه که توسط یک رادیولوژیست انجام می‌شود). این اطلاعات موجب تسهیل تصمیمات توسط رادیولوژیست‌ها می‌شود.

تلفیق انسان‌ها و هوش مصنوعی برای ارائه بهترین خدمات به مشتریان

شرکت جنسیس از شرکت پژوهشی فارستر برای انجام یک نظرسنجی جهانی در سال ۲۰۱۷ برای پی بردن به اینکه شرکت‌ها چگونه از هوش مصنوعی برای بهبود خدمات مشتریان استفاده می‌کنند استفاده کرد. این مطالعه با عنوان «هوش مصنوعی با رنگ و بوی انسانی» به صورت رایگان در آدرس [genesis.com/resources/artificialintelligence-with-the-human-touch](https://www.genesis.com/resources/artificialintelligence-with-the-human-touch) در دسترس است. یک ویدئوی مرتبط در [youtube.com/watch?v=NP2qqwGTNPk](https://www.youtube.com/watch?v=NP2qqwGTNPk) موجود است. نتایج حاصل از این مطالعه به شرح زیر هستند:

۱. هوش مصنوعی با افزایش کارایی و بهره‌وری کارگران، ارائه‌ی تجربه‌ی بهتر به مشتریان و کشف جریان‌های درآمدی جدید موجب دگرگونی و تغییر شکل سازمان‌ها شده است.
۲. یکی از اهداف اصلی همکاری انسان-ماشین افزایش رضایت مشتریان و عوامل سازمان‌ها به جای کاهش هزینه است.

¹ Marr

² Moran

۳. توانایی عوامل انسانی برای برقراری ارتباط عاطفی با مشتریان به منظور افزایش رضایت خود و مشتریان بر خدمات ارائه شده توسط هوش مصنوعی ارجحیت دارد.

۴. با تلفیق نقاط قوت انسان‌ها و هوش مصنوعی، شرکت‌ها می‌توانند به رضایت بیشتر مشتریان از خدمات (۷۱٪) و عوامل خود (۶۹٪) دست پیدا کنند.

توجه داشته باشید که هوش مصنوعی در پشتیبانی از بازاریابی و تبلیغات پیشگام است. برای کاربرد هوش مصنوعی در پشتیبانی از مدیریت ارتباط با مشتریان و جمع‌سپاری و هوش جمعی برای پشتیبانی از بازاریابی نگاه کنید به لُتن (۲۰۱۸).

ربات‌های مشارکتی (کوبات‌ها) ربات‌های مشارکتی (کوبات‌ها) برای کار با انسان‌ها، کمک به انجام وظایف مختلف اجرایی طراحی شده‌اند. این ربات‌ها خیلی هوشمند نیستند ولی هزینه‌ی پایین آنها و استفاده‌پذیری بالای آنها موجب محبوبیت‌شان شده است. برای جزئیات نگاه کنید به توبی^۱ (۲۰۱۵).

یکی از جهت‌گیری‌های رباتیک در آینده، خلق تیم‌هایی از ربات‌ها است که برای انجام کارهای پیچیده طراحی شده‌اند. تیم ربات‌ها در حوزه ساخت و تولید متداول هستند و به یکدیگر سرویس داده یا به گروهی از ربات‌ها در انجام کارهای مونتاژ ملحق می‌شوند. یک مثال جالب از بکارگیری یک تیم از ربات‌ها در آماده‌سازی زمین در مریخ است.



شکل ۱۱.۴ یک تیم از ربات‌ها برای رفتن به مریخ آماده می‌شوند

پیش از آنکه بشر بر روی مریخ فرود آید، دانشمندان باید اطلاعات بیشتری در مورد این «سیاره سرخ» کسب کنند. ایده‌ی اصلی استفاده از تیم ربات‌ها بود. مراکز پژوهشی هوش مصنوعی آلمان^۲ آزمایشات شبیه‌سازی را در صحرای یوتا انجام دادند. جزئیات

¹ Tobe

² German Research Centers for Artificial Intelligence (DFKI)

این شبیه‌سازی توسط استاف رایترز^۱ (۲۰۱۶) توضیح داده شده است. این فرایند در یک ویدئوی ۴:۵۴ دقیقه‌ای در [youtube.com/watch?v=pvKIzldni68/showing robots' collaboration](https://www.youtube.com/watch?v=pvKIzldni68/showing_robots_collaboration) به تصویر کشیده شده است. برای اطلاعات بیشتر به robotic.dfki-bermen.de/en/research/projects/ft-utah.html نگاه کنید.

دی‌اف‌کی‌آی تنها نهادی نیست که برای اکتشاف در سطح مریخ برنامه‌ریزی می‌کند. ناسا قصد دارد تا ازدحامی از زنبورهای ربات با بال‌های متحرک به نام مارس‌بیز^۲ را که به صورت گروهی برای اکتشاف سطح زمین و هوای سیاره سرخ فعالیت می‌کنند ارسال کند. دلیل استفاده از ساختار بال‌های متحرک این است که پروازهایی که سطح انرژی پایینی نیاز دارند را امکان‌پذیر سازد. هر ربات به اندازه‌ی یک زنبور است. مارس‌بیز به عنوان بخشی از یک شبکه ارتباطی بی‌سیم در کنار یکدیگر شبکه‌ای از حسگرها را ایجاد می‌کنند. اطلاعات به یک پایگاه سیار ارسال می‌شود که مرکز اصلی ارتباطات و یک ایستگاه شارژ مجدد برای مارس‌بیز محسوب می‌شود. برای اطلاعات بیشتر به کانگ (۲۰۱۸) نگاه کنید.

همکاری ربات‌ها با یکدیگر در ام‌آی‌تی مورد مطالعه و پژوهش قرار گرفته است. آنها از سیستم پرسپشن برای حس کردن محیط استفاده کرده و سپس یافته‌هایشان را برای یکدیگر ارسال نموده و کارهایشان را هماهنگ می‌کنند. به عنوان مثال، یک ربات می‌تواند یک در را برای ربات دیگری باز کند. برای مطالعه بیشتر به ویدئوی [ft.com/video/ea2d4877-f3fb-403d-84a8-a4d2d4018c5e](https://www.ft.com/video/ea2d4877-f3fb-403d-84a8-a4d2d4018c5e) نگاه کنید.

[youtube.com/watch?v=FB14Y55V2Z4](https://www.youtube.com/watch?v=FB14Y55V2Z4)

[youtube.com/watch?v=ULKyXnQ9xWA](https://www.youtube.com/watch?v=ULKyXnQ9xWA)

پیش

سوالاتی برای مرور بخش ۱۱.۹

۵. چرا همکاری انسان-ماشین در حال افزایش است؟
۶. برخی از مزایای چنین همکاری را نام ببرید.
۷. توضیح دهید که چگونه از رباتیک مشارکتی می‌توان در ساخت و تولید استفاده کرد؟
۸. استفاده از ربات‌ها را توضیح دهید.
۹. ربات‌ها در مریخ چه خواهند کرد؟

¹ Staff Writers

² Marsbees

³ Hughes & Hughes

- گروه/افزار به محصولات نرم افزاری اطلاق می‌شود که از گروه‌ها پشتیبانی مشارکتی به عمل می‌آورند (از جمله برگزاری جلسات).
- گروه‌افزار می‌تواند از تصمیم‌گیری و حل مسئله به طور مستقیم یا غیر مستقیم با بهبود ارتباط بین اعضای تیم پشتیبانی به عمل آورد.
- افراد در کارهایشان با هم همکاری می‌کنند (به نام کارگروهی). گروه افزار (یعنی نرم‌افزار رایانش مشترک) از کار گروهی پشتیبانی می‌کند.
- اعضای گروه ممکن است در یک سازمان یا در سازمان‌های مختلف در همان یا در مکان‌های مختلف بوده و ممکن است در یک زمان یا زمان‌های متفاوتی کار کنند.
- چارچوب زمان/مکان راه مناسبی برای تشریح الگوهای ارتباطی و همکاری و پشتیبانی از کار گروهی باشد. فناوری‌های مختلف می‌توانند شرایط و حالت‌های زمان/مکان مختلف را پشتیبانی کنند.
- کار گروهی می‌تواند مزایای زیادی را به همراه داشته باشد، از جمله بهبود در تصمیم‌گیری، افزایش بهره‌وری و سرعت و کاهش هزینه‌ها.
- ارتباط می‌تواند همزمان (یعنی زمان یکسان) یا غیرهمزمان (یعنی ارسال و دریافت در زمان‌های متفاوت) باشد.
- اینترنت، اینترانت و اینترنت اشیا از جلسات مجازی و تصمیم‌گیری از طریق مشارکت ابزار و دسترسی به تجزیه و تحلیل داده‌ها، اطلاعات، و دانش پشتیبانی می‌کنند.
- گروه‌افزارها برای پشتیبانی مستقیم معمولاً شامل قابلیت طوفان ذهنی، کنفرانس، برنامه‌ریزی جلسات گروهی؛ برنامه‌ریزی؛ حل و فصل تعارض‌ها؛ کنفرانس ویدیویی؛ به اشتراک گذاری الکترونیکی اسناد؛ رای دادن، سیاست‌گذاری؛ و تجزیه و تحلیل داده‌های سازمانی است.
- جی‌دی‌اس‌اس به هر ترکیبی از سخت افزار و نرم افزاری که موجب تسهیل جلسات تصمیم‌گیری شود اطلاق می‌شود. همچنین پشتیبانی مستقیم در جلسات حضوری و مجازی به عمل آورده و سعی در افزایش مزایای فرایند و کاهش ضرر و زیان‌های کار گروهی دارد.
- هوش جمعی مبتنی بر این فرض است که خرد ترکیبی چندین نفر که با یکدیگر همکاری می‌کنند بیشتر از خرد تک تک افرادی است که کار می‌کنند.
- هر یک از چندین پیکربندی هوش جمعی را می‌توان به شکل متفاوتی به کمک فناوری پشتیبانی کرد.
- چندین پلتفرم همکاری، مانند مایکروسافت تیمز و اسلک می‌توانند هوش جمعی را تسهیل کنند.
- ایده پردازی و طوفان ذهنی فعالیت‌های کلیدی در کار گروهی برای تصمیم‌گیری محسوب می‌شوند. چندین نرم افزار همکاری و برنامه‌های هوش مصنوعی این فعالیت‌ها را پشتیبانی می‌کنند.
- جمع‌سپاری یک فرایند برون سپاری کار به یک جمع است. اینکار می‌تواند به حل مسئله، ایده‌پردازی و سایر فعالیت‌های نوآورانه کمک کند.

- از جمع‌سپاری می‌توان برای پیش‌بینی توسط گروه‌هایی از افراد از جمله جمع‌ها استفاده کرد. نتایج، پیش‌بینی‌های بهتری را نشان داده‌اند، به خصوص زمانی که بین پیش‌بینی‌کننده‌ها ارتباط برقرار باشد تا زمانیکه هیچ ارتباط فعالی وجود نداشته است.
- یکی از روش‌های برقراری ارتباط در جمع‌سپاری مبتنی بر هوش ازدحامی است. یک فناوری معروف به هوش مصنوعی ازدحامی موفقیت بسیار زیادی کسب کرده است.
- هوش مصنوعی می‌تواند از بسیاری از فعالیت‌های تصمیم‌گیری گروهی پشتیبانی کند.
- همکاری انسان و ماشین می‌تواند یکی از روش‌های اصلی انجام کارها در آینده باشد.
- ماشین‌هایی که زمانی از ساخت و تولید پشتیبانی می‌کردند در حال حاضر نیز برای پشتیبانی از کارهای شناختی استفاده می‌شود، از جمله کارهای مدیریتی.
- برای همکاری و مشارکت انسان و ماشین‌ها به صورت کار گروهی اتخاذ تمهیدات ویژه امری ضروری است.
- ربات‌ها ممکن است در تیم‌های انحصاری کار کنند. با افزایش هوش ربات‌ها این تیم‌های انحصاری در تولید و احتمالاً در سایر فعالیت‌ها (به عنوان مثال، اکتشاف مریخ) بیش از پیش انجام می‌شوند.

- ۱.
- ۲.
- ۳.
- ۴.
- ۵.
- ۶.
- ۷.
- ۸.

.....	.۹
.....	.۱۰
.....	.۱۱
.....	.۱
..... realtimeboard.com۱
.....	.۲
.....	.۳
.....	.۴
.....	.۵
..... socialtext.com۶
.....	.۷
.....	.۸
..... youtube.com/watch?v=JJHAHQmiI3c۹
.....	.۱۰
..... Dropbox.com۱۱
.....	.۱۲
.....	.۱۳
.....	.۱۴
.....	.۱۵
..... crowdtap.com۱۶

¹ Kuzer

۱۲

سیستم‌های دانشی: سیستم‌های خبره، توصیه‌گرها،
چَت‌بات‌ها، دستیارهای شخصی مجازی و مشاورات روبرو

پیشرفت در فناوری‌های هوش مصنوعی و به‌ویژه پردازش زبان طبیعی، سیستم‌های یادگیری و دانش ماشینی و عمیق، همراه با افزایش کیفیت و عملکرد سایر سیستم‌های هوشمند و دستگاه‌های تلفن همراه و برنامه‌های آن‌ها، باعث توسعه چت‌بات‌ها برای اجرای ارزان و سریع بسیاری از وظایف مربوط به ارتباطات، همکاری و بازیابی اطلاعات شده است. استفاده از چت‌بات‌ها در تجارت به سرعت در حال افزایش است که تا حدودی به این دلیل تناسب آنها با سیستم‌ها و دستگاه‌های سیار می‌باشد. در واقع، ارسال

- چت‌بات‌ها می‌توانند به مشتریان کمک کنند تا مشکلات خود را حل کنند.
- چت‌بات‌ها می‌توانند به مشتریان کمک کنند تا محصولات و خدمات را بفهمند.
- چت‌بات‌ها می‌توانند به مشتریان کمک کنند تا سفارشات خود را ثبت کنند.

- چت‌بات‌ها می‌توانند به مشتریان کمک کنند تا مشکلات خود را حل کنند.
- چت‌بات‌ها می‌توانند به مشتریان کمک کنند تا محصولات و خدمات را بفهمند.
- چت‌بات‌ها می‌توانند به مشتریان کمک کنند تا سفارشات خود را ثبت کنند.

پیام احتمالاً فعالیت اصلی در دنیای موبایل یا سیار است. در دو تا سه سال گذشته، هزاران ربات در سراسر جهان توسط سازمان‌ها (خصوصی و عمومی) و افراد مورد استفاده قرار گرفته‌اند. بسیاری از مردم از این پدیده به عنوان *انقلاب چت‌بات* یاد می‌کنند. چت‌بات‌های امروزی بسیار پیچیده‌تر از همتایان قبلی خود هستند. از آنها به طور گسترده‌ای به عنوان مثال، در بازاریابی، خدمات مالی، دولتی و مشتری؛ مراقبت‌های بهداشتی؛ و در تولید استفاده می‌شود. چت‌بات‌ها در مقایسه با رایانه‌های معمولی ارتباط شخصی‌تری را برقرار می‌کنند و در جمع‌آوری داده‌ها برتری دارند. چت‌بات‌ها می‌توانند به صورت مستقل یا بخشی از سایر سیستم‌های دانشی باشند:

موضوعات این فصل را به چهار دسته تقسیم می‌کنیم: سیستم‌های خبره، چت‌بات‌ها برای ارتباط و همکاری، دستیاران شخصی مجازی (محصولات بومی مانند الکسا)، و چت‌بات‌هایی که به عنوان مشاوران حرفه‌ای استفاده می‌شوند. برخی از موضوعات مربوط به پیاده‌سازی و اجرای سیستم‌های هوشمند در آخر توضیح داده شده است. این فصل دارای بخش‌های زیر است:

۱۲.۱. پرونده‌ی آغازین: تعالی سِفورا به کمک چت‌بات‌ها

۱۲.۲. سیستم‌های خبره و توصیه‌گرها

۱۲.۳. مفاهیم، محرک‌ها و مزایای چت‌بات‌ها

۱۲.۴. چت‌بات‌های سازمانی

۱۲.۵. دستیارهای شخصی مجازی

۱۲.۶. چت‌بات‌ها به عنوان مشاوران شخصی (مشاوران روبو)

۱۲.۷. مسائل پیاده‌سازی و اجرا

سِفورا^۱ یک شرکت لوازم آرایشی و زیبایی مستقر در فرانسه است که در سطح جهانی تجارت می‌کند. اجناس خود را از طریق شبکه فروشگاه‌های بزرگ لوازم آرایشی و بهداشتی خودش به فروش می‌رساند. علاوه بر این، سِفورا به فروش آنلاین در آمازون و در فروشگاه آنلاین خود می‌پردازد. این شرکت صدها برند معتبر و همچنین بسیاری از برندهای خود را می‌فروشد. این شرکت در یک بازار بسیار رقابتی که مراقبت از مشتری و تبلیغات نقش بسیار حیاتی را ایفا می‌کنند فعالیت می‌کند. سِفورا برخی از محصولات مردانه را نیز به فروش می‌رساند اما اکثر محصولات زیبایی آن بازار زنان را هدف قرار داده است.

اولین استفاده‌ی سِفورا از چت‌بات‌ها از طریق سرویس‌های پیام‌رسانی رخ داد. هدف اولین بات این بود که به جستجوی اطلاعات از قبیل ویدئوها، تصاویر، نکات و غیره برای منابع شرکت باشد. این بات در حالت پرسش و پاسخ (Q&A) عمل می‌کند. محتوای مربوطه را بر اساس علایق مشتریان توصیه می‌کند. هدف این شرکت جلب توجه مشتریان جوانی است که از پیام‌رسان کیک^۲ استفاده می‌کنند. محققان سِفورا دریافتند که مشتریانی که با کیک‌بات^۳ کار می‌کنند عمیقاً درگیر گفتگو می‌شوند. سپس بات آنها را تشویق می‌کند تا محصولات جدید را شناسایی کنند. بات جدیدتر سِفورا به نام دستیار رزرواسیون^۴ در فیسبوک مسنجر^۵ قرار گرفت که مشتریان را قادر می‌سازد تا برای رزرو و یا برنامه‌ریزی مجدد قرار ملاقات اقدام کنند. یکی دیگر از بات‌هایی که سِفورا در اپلیکیشن کیک ارائه کرد شید-مچینگ^۶ نام دارد. رنگ لب‌ها را با عکس‌های آپلود شده توسط کاربران (صورت و لب) مطابقت داده و بهترین گزینه را به آنها توصیه می‌کند. این بات همچنین به کاربران اجازه می‌دهد تا عکس رنگ‌های پیشنهادی را با استفاده از هنرمند مجازی سِفورا که در فیسبوک مسنجر اجرا می‌شود امتحان کنند. بات‌ها به صورت اپلیکیشن‌های موبایل نیز اجرا می‌شوند. اگر کاربران تمایل به دریافت توصیه داشته باشند، برای خرید محصولات به فروشگاه اینترنتی شرکت هدایت می‌شوند. کاربران می‌توانند عکس‌های سلفی را آپلود کنند تا برنامه بتواند خواسته آنها را با کالای موجود مطابقت دهد. بیش از ۴ میلیون بازدیدکننده در اولین سال فعالیت هنرمند مجازی ۹۰ میلیون طیف رنگ را امتحان کردند.

¹ Sephora

² Kik

³ Kikbot

⁴ Reservation Assistant

⁵ Facebook Messenger

⁶ Shade-Matching

مجموعه پرسش و پاسخ پایگاه دانش از طریق ایجاد ارتباط بین این پایگاه و کارشناسان فروشگاه ایجاد شد. برای این منظور از تکنیک‌های کسب دانش (فصل ۲) استفاده شد. ربات‌های این شرکت از پردازش زبان طبیعی که برای درک واژگان معمولی کاربران آموزش دیده‌اند استفاده می‌کنند.

مشتریان این شرکت عاشق ربات‌ها شدند. علاوه بر این، سفورا اهمیت ارائه کمک و راهنمایی برای کاربرانی که انگیزه بازگشت (با هزینه معقول!)، داشتند را آموخت.

بات سفورا برای پی بردن به سلیقه و اولویت‌های کاربران از آنها سوالاتی را می‌پرسد. سپس همانند یک سیستم توصیه‌گر عمل می‌کند (بخش ۱۲.۲) و محصولات را پیشنهاد می‌دهد. کاربران کیک و مسنجر می‌توانند اقلام را بدون اینکه از سرویس پیام‌رسان خارج شوند خریداری کنند.

نهایتاً، شرکت در طول زمان دانش بات سفورا را بهبود بخشیده و بات‌های جدیدی را برای انجام وظایف بیشتر طرح‌ریزی کرده است.

Note: Sephora was selected by *Fast Company Magazine*, March/April 2018, as one of the "World's Most Innovative Companies." Sephora is known for its digital transformation and innovation (Rayome, 2018). Also, Sephora's bots are considered among the top marketing chatbots (Quoc, 2017).
Sources: Compiled from Arthur (2016), Rayome (2018), and Taylor (2016), theverge.com/2017/3/16/14946086/sephora-virtual-assistant-ios-app-update-ar-makeup/, and sephora.com/.

سوالاتی برای مرور پرونده‌ی آغازین

۵۲. مزایای بات‌ها در این شرکت را نام برده و در مورد آنها بحث کنید.
۵۳. مزایای بات‌ها برای مشتریان را نام برده و در مورد آنها بحث کنید.
۵۴. چرا بات‌ها از طریق منسجر و کیک راه‌اندازی شدند؟
۵۵. اگر رقبا از روش مشابهی استفاده کنند چه اتفاقی برای سفورا خواهد افتاد؟

در دنیای بسیار رقابتی خرده‌فروشی محصولات زیبایی، مراقبت از مشتری و بازاریابی بسیار مهم است. صرفاً بکارگیری کارکنان انسانی می‌تواند بسیار پرهزینه باشد. علاوه بر این، مشتریان به صورت ۲۴ ساعته در حال خرید هستند و فروشگاه‌های فیزیکی در ساعات و روزهای محدودی باز هستند. علاوه بر این، ترکیبات زیادی از محصولات زیبایی خاص (به عنوان مثال، بسیاری از سایه‌ها/رنگ‌ها) موجود است. سفورا تصمیم گرفت از چت‌بات‌ها در فیسبوک مسنجر و کیک برای تعامل با مشتریان خود استفاده کند. چت‌بات‌ها، موضوع این فصل، ۲۴ ساعته و با هزینه کمتر از طریق دستگاه‌های تلفن همراه در دسترس هستند. بات‌ها اطلاعات را به طور مداوم به مشتریان ارائه می‌دهند و به سرعت مشتریان را به خرید آنلاین آسان هدایت می‌کنند. سفورا چت‌بات‌های خود را روی پیام‌رسانی خدمات قرار داده است. منطق این کار این بود که مردم دوست دارند در سرویس‌های پیام‌رسانی با دوستانشان چت کنند، و همچنین ممکن است دوست داشته باشند با مشاغل چت کنند.

علاوه بر چندین سرویس به مشتریان، استفاده از چت‌بات‌ها به سفورا کمک می‌کند تا در مورد مشتریان اطلاعات کسب کند. این نوع چت‌بات‌ها رایج‌ترین نوع برای مراقبت از مشتری و بازاریابی هستند. در این فصل، ما چندین نوع دیگر از سیستم‌های دانشی،

از جمله سیستم‌های خبره پیشگام، توصیه‌گرها، دستیاران شخصی مجازی ارائه شده توسط چندین شرکت بزرگ فناوری و مشاوران روبرو را پوشش می‌دهیم.

در فصل ۲ ما به معرفی مفهوم سیستم‌های تصمیم‌گیری مستقل پرداختیم. یک سیستم خبره یک گروه از سیستم‌های تصمیم‌گیری مستقل بوده و جزو اولین کاربردهای هوش مصنوعی محسوب می‌شود. بکارگیری سیستم‌های خبره از اوایل و اواسط دهه ۶۰ در مؤسسات پژوهشی آغاز شد (مثلاً دانشگاه استنفورد، آی‌بی‌ام) و در طی دهه ۸۰ به کاربرد تجاری رسید.

در ادامه به معرفی مفاهیم اصلی مرتبط با فناوری سیستم‌های خبره می‌پردازیم.

در این بخش، ما به بررسی سیستم‌های خبره می‌پردازیم. سیستم‌های خبره، سیستم‌هایی هستند که دانش تخصصی را در مورد یک حوزه خاص، مانند تشخیص پزشکی، طراحی مهندسی یا بازی شطرنج، در اختیار می‌گذارند. این سیستم‌ها قادرند با استفاده از قواعد و قوانین مشخص، تصمیمات منطقی و مبتنی بر دانش بگیرند. سیستم‌های خبره معمولاً شامل یک پایگاه دانش (که حاوی اطلاعات تخصصی است) و یک موتور استنتاج (که برای حل مسئله از این دانش استفاده می‌کند) می‌باشند. این سیستم‌ها می‌توانند به کاربران کمک کنند تا با مسائل پیچیده مواجه شوند و به تصمیمات بهتری دست یابند.

سیستم‌های خبره در دهه ۱۹۸۰ به‌طور گسترده‌ای مورد استفاده قرار گرفتند. این سیستم‌ها در صنایع مختلفی مانند پزشکی، مهندسی، مالی و آموزش به کار آمدند. یکی از دلایل موفقیت این سیستم‌ها، توانایی آن‌ها در تقلید از فرآیند تفکر انسانی در حل مسائل تخصصی بود. با پیشرفت فناوری‌های هوش مصنوعی، سیستم‌های خبره به‌طور فزاینده‌ای پیچیده‌تر شده‌اند و قادرند وظایف بسیار متنوعی را انجام دهند. امروزه، سیستم‌های خبره در کنار سایر فناوری‌های هوش مصنوعی مانند یادگیری عمیق، به‌کاربرد می‌آیند و نقش مهمی در توسعه سیستم‌های هوشمند ایفا می‌کنند.

در ادامه:



¹ Expert System

- در صورتی که دانشجو در این زمینه تخصص داشته باشد
- در صورتی که دانشجو در این زمینه تجربه داشته باشد
- در صورتی که دانشجو در این زمینه مهارت داشته باشد
- در صورتی که دانشجو در این زمینه توانایی داشته باشد
- در صورتی که دانشجو در این زمینه علاقه داشته باشد
- در صورتی که دانشجو در این زمینه انگیزه داشته باشد

این موارد می‌تواند به دانشجو کمک کند تا در انتخاب رشته تحصیلی خود تصمیم‌گیری بهتری داشته باشد و به او کمک کند تا در انتخاب رشته تحصیلی خود تصمیم‌گیری بهتری داشته باشد.

تخصص^۱ منظور از تخصص، یک دانش گسترده و خاص است که در اختیار شخص متخصص یا کارشناس می‌باشد. سطح تخصص تعیین کننده موفقیت تصمیمی است که توسط یک شخص خبره گرفته شده است. تخصص اغلب از طریق آموزش، یادگیری و تجربه در عمل بدست می‌آید که شامل دانش صریح مثل نظریه‌های آموخته شده از کتب یا کلاس درس و دانش پنهان کسب شده از تجربه است. در ادامه فهرستی از انواع دانش‌های ممکن که در برنامه‌های کاربردی سیستم خبره بکار می‌روند آورده شده است:

- در صورتی که دانشجو در این زمینه تخصص داشته باشد
- در صورتی که دانشجو در این زمینه تجربه داشته باشد
- در صورتی که دانشجو در این زمینه مهارت داشته باشد
- در صورتی که دانشجو در این زمینه توانایی داشته باشد
- در صورتی که دانشجو در این زمینه علاقه داشته باشد
- در صورتی که دانشجو در این زمینه انگیزه داشته باشد

این موارد می‌تواند به دانشجو کمک کند تا در انتخاب رشته تحصیلی خود تصمیم‌گیری بهتری داشته باشد و به او کمک کند تا در انتخاب رشته تحصیلی خود تصمیم‌گیری بهتری داشته باشد.

- در صورتی که دانشجو در این زمینه تخصص داشته باشد
- در صورتی که دانشجو در این زمینه تجربه داشته باشد
- در صورتی که دانشجو در این زمینه مهارت داشته باشد
- در صورتی که دانشجو در این زمینه توانایی داشته باشد
- در صورتی که دانشجو در این زمینه علاقه داشته باشد
- در صورتی که دانشجو در این زمینه انگیزه داشته باشد

¹ Expertise

صدها شرکت در سرتاسر جهان بین سال‌های ۱۹۸۰ تا ۲۰۱۰ از سیستم خبره استفاده کرده‌اند. با این حال، از سال ۲۰۱۱، استفاده از آنها بیشتر به دلیل ظهور سیستم‌های دانشی بهتر که سه نوع از آنها در این فصل توضیح داده شده است به سرعت کاهش یافته است. با این حال، درک ویژگی‌ها و مزایای اصلی سیستم‌های خبره بسیار مهم است زیرا بسیاری از آنها منجر به تکامل سیستم‌های دانشی جدیدتری شدند. هدف اصلی سیستم خبره انتقال تخصص به یک ماشین است تا افراد غیرمتخصص بتوانند از این تخصص استفاده کنند. یک مثال معمولی برای سیستم خبره، تشخیص یا عیب‌یابی است. برای مثال، بسیاری از ما می‌توانیم خودمان مشکلات رایانه‌هایمان را تشخیص داده و رفع کنیم. حتی فراتر از این، رایانه‌ها می‌توانند به تنهایی مشکلات را یافته و رفع کنند. یکی از زمینه‌هایی که چنین توانایی در آن بکار گرفته می‌شود حوزه‌ی پزشکی است که در مثال زیر توضیح داده شده است:

یک سیستم خبره‌ی مبتنی بر وب در کشور کره توسعه یافت تا مردم بتوانند وضعیت سلامت روانی و ذهنی خود را بررسی کنند. هر کسی در دنیا می‌تواند به این سیستم دسترسی یافته و به صورت رایگان از آن استفاده کند. دانش مورد نیاز برای این سیستم از یک نظرسنجی انجام شده از ۳،۲۳۲ مهاجر گره‌ای گردآوری شده است. نتایج نظرسنجی توسط کارشناسان و افراد خبره و از طریق بحث و گفتگو در گروه‌های تمرکز مورد تجزیه و تحلیل و بررسی قرار گرفت. برای اطلاعات بیشتر به بانه (۲۰۱۳) نگاه کنید.

مزایای سیستم خبره بسته به مأموریت و ساختار سیستم خبره، قابلیت‌ها و مزایای بالقوه‌ی آنها به شرح زیر می‌باشند:

- انجام وظایف روتین (مثلاً عیب‌یابی، غربال‌کننده‌ها، تحلیل اعتبار) که نیازمند تخصص سریع تر از انسان‌ها می‌باشد.
- کاهش هزینه‌های عملیات
- بهبود سازگاری و کیفیت کار (مثلاً کاهش خطاهای انسانی)
- سرعت‌دهی به تصمیم‌گیری و اتخاذ تصمیمات سازگار
- افزایش انگیزه‌ی کارکنان برای افزایش بهره‌وری
- حفظ تخصص‌های کمیاب کارکنانی که بازنشسته می‌شوند.
- کمک به انتقال و استفاده مجدد از دانش
- کاهش هزینه آموزش کارکنان با استفاده از خودآموزی
- حل مسائل پیچیده بدون نیاز به کارشناسان و حل سریع تر آنها
- توجه به مسائلی که حتی از چشم کارشناسان و خبرگان دور می‌ماند
- متمرکزسازی تصمیم‌گیری (مثلاً با استفاده از «ابر»)
- تسهیل به اشتراک گذاری دانش

این مزایا می توانند مزیت رقابتی بسیار خوبی را برای شرکت‌هایی که از سیستم خبره استفاده می کنند به همراه داشته باشند. علاوه بر این، برخی شرکت‌ها با استفاده از سیستم‌های خبره صرفه‌جویی قابل توجهی را انجام داده‌اند. علیرغم این مزایا، بکارگیری سیستم خبره در حال افول است. علل این امر و محدودیت‌های مربوطه در ادامه این بخش توضیح داده شده‌اند.

از سیستم خبره به صورت تجاری در حوزه‌های متعددی از جمله موارد زیر استفاده شده است:

- **امور مالی.** سیستم خبره مالی شامل تجزیه و تحلیل سرمایه‌گذاری، اعتبارات و گزارشات مالی؛ ارزشیابی عملکرد و بیمه؛ برنامه‌ریزی مالیات؛ پیشگیری از کلاهبرداری‌ها و برنامه‌ریزی مالی می باشد.
- **پردازش داده‌ها.** سیستم خبره پردازش داده شامل برنامه‌ریزی، انتخاب تجهیزات، نگهداری تجهیزات، ارزشیابی تأمین کننده و مدیریت شبکه می باشد.
- **بازاریابی.** سیستم خبره بازاریابی شامل مدیریت ارتباط با مشتری، پژوهش و تحلیل بازار، برنامه‌ریزی محصول و برنامه‌ریزی بازار می‌باشد. همچنین ارائه مشاوره پیش از فروش نیز توسط این سیستم‌ها ارائه می‌شود.
- **منابع انسانی.** مثال‌هایی از سیستم خبره منابع انسانی شامل برنامه‌ریزی، ارزیابی عملکرد، زمانبندی پرسنل، مدیریت حقوق بازنشستگی، مشاوره‌های قانونی و طراحی پرسننامه.
- **ساخت و تولید.** سیستم خبره ساخت و تولید شامل برنامه‌ریزی تولید پیکربندی محصولات پیچیده مدیریت کیفیت طراحی محصول انتخاب سایت تولید و نگهداری و تعمیرات تجهیزات (از جمله عیب‌یابی) می‌باشد.
- **اتوماسیون فرایند کسب‌وکار.** سیستم خبره برای اتوماسیون میز کار، مدیریت مرکز تماس و اعمال قوانین و مقررات بکار گرفته شده است.
- **امنیت ملی.** این سیستم‌های خبره شامل ارزیابی تهدیدهای تروریستی و تشخیص تأمین مالی‌ها برای اهداف تروریستی می باشد.
- **مدیریت مراقب سلامت.** سیستم‌های خبره برای بیوانفورماتیکز و سایر مسائل مدیریت مراقب سلامت توسعه یافته‌اند.
- **الزامات قانونی و انطباق با آنها.** قوانین و مقررات می‌توانند پیچیده باشند. سیستم خبره از یک فرایند گام به گام برای حصول اطمینان از انطباق با قوانین و مقررات استفاده می‌کند.
- **طراحی وب سایت.** یک طراحی وب سایت خوب نیازمند این است که به متغیرهای زیادی توجه شده و اطمینان حاصل کنیم که عملکرد سایت مطابق با استانداردها باشد. سیستم خبره می‌تواند منجر به یک فرایند طراحی مناسب و خوب شود.

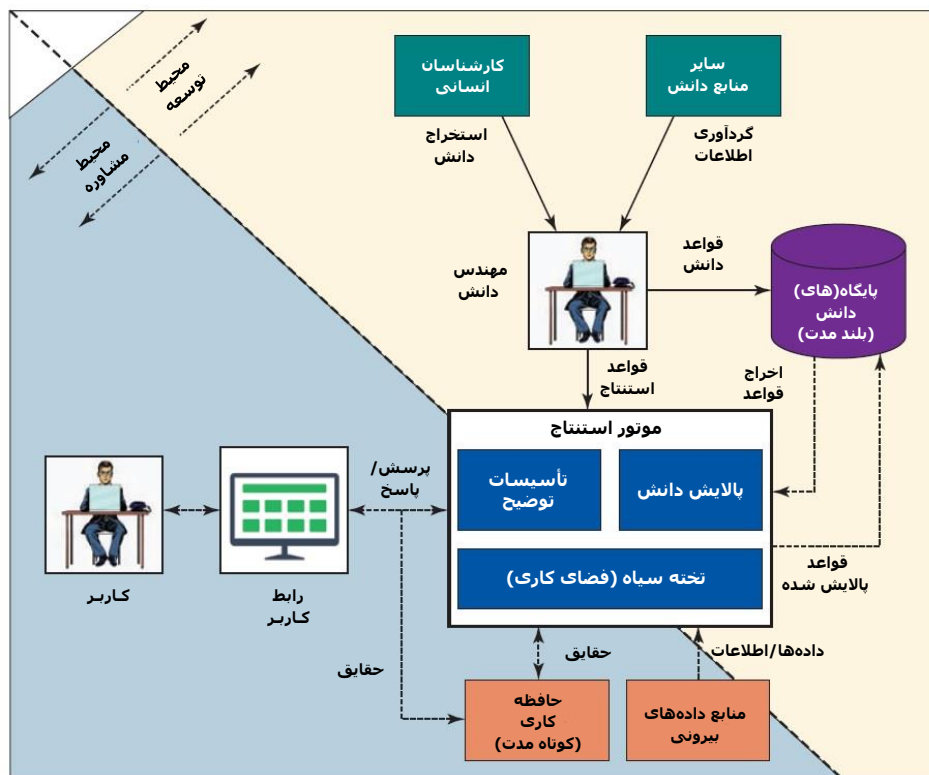
اکنون که به مفاهیم پایه و اصلی سیستم خبره آشنا شدیم زمان آن رسیده تا نگاهی به ساختار داخلی سیستم خبره و چگونگی دستیابی به اهداف آنها بیندازیم.

همانطور که از بخش ۲.۵ و شکل ۲.۵ به خاطر داریم فرایند استخراج دانش و کاربرد آن به دو بخش متمایز تفکیک شده است. در سیستم خبره ما این دو بخش را محیط توسعه و محیط مشاوره می‌نامیم (شکل ۱۲.۱). یک سازنده‌ی سیستم خبره، اجزا و مؤلفه‌های ضروری سیستم خبره را ساخته و پایگاه دانش را با پرزنت‌های مناسب دانش خبره در محیط توسعه^۱ بارگذاری می‌کند. یک فرد غیرکارشناس از محیط مشاوره برای دستیابی به مشاوره و حل مسائل با استفاده از دانش خبره‌ای که در سیستم تعبیه شده است استفاده می‌کند. این دو محیط معمولاً جدا از یکدیگر هستند.

کامپوننت‌های اصلی سیستم خبره کامپوننت‌های اصلی سیستم‌های خبره معمولاً شامل موارد زیر است:

- **اکتساب دانش.** این دانش که عمدتاً توسط مهندسان دانش از کارشناسان و خبرگان انسانی کسب می‌شود. این دانش ممکن است در منابع متعددی اخذ، یکپارچه سازی، صحت‌گذاری و اعتبارسنجی شود.
- **پایگاه دانش.** این یک مخزن دانش است. دانش به دامش مرتبط با دامنه و دانش مرتبط با حل مسئله و رویه‌های راه حل تقسیم می‌شود. همچنین، داده‌های ورودی که توسط کاربران ارائه شده اند ممکن است در یک پایگاه دانش ذخیره شوند.
- **نمایش دانش.** دانش اغلب به صورت قواعد و قوانین کسب و کار سازماندهی می‌شود (که به نام قواعد تولید نیز معروفند).
- **موتور/استنتاج.** که به نام ساختار کنترل یا مفسر قوانین نیز شناخته می‌شود «مغز» سیستم خبره محسوب می‌شود و قابلیت‌های استدلال و استنتاج یعنی توانایی پاسخگویی به سوالات کاربران، ارائه‌ی توصیه‌ها برای راه حل‌ها، انجام پیش‌بینی‌ها و انجام سایر وظایف مربوطه را امکان‌پذیر می‌کند.
- **رابط کاربری.** این کامپوننت به کاربر اجازه می‌دهد تا با موتور استنتاج به تعامل بپردازد. در سیستم‌های خبره کلاسیک، اینکار از طریق نوشتن دستورها یا با استفاده از منوها انجام می‌شود. در سیستم‌های دانشی امروزی، اینکار با زبان‌های طبیعی و صوت انجام می‌شود.

¹ Development Environment



شکل ۱۲.۱ معماری عمومی سیستم‌های خبره

Exsys.com

استفاده از سیستم خبره در شناسایی عوامل شیمیایی، بیولوژیکی و رادیولوژیکی

مورد کاربردی ۱۲.۱

حملات تروریستی با استفاده از عوامل شیمیایی، بیولوژیکی یا رادیولوژیکی (سی بی آر) به دلیل توان بالقوه‌ای که منجر به از بین رفتن جان بسیاری می‌شود دغدغه و نگرانی بسیاری را موجب شده است. در ایالات متحده و سایر کشورها میلیاردها دلار هزینه صرف طرح‌ها و پروتکل‌های دفاع در برابر اقدامات تروریستی که می‌تواند توسط سی بی آر انجام شود شده است. با این حال، سی بی آر طیف وسیعی از عوامل ورودی را با بسیاری از ارگانیزم‌های خاصی که می‌توانند به چندین شکل استفاده شوند پوشش می‌دهد. پاسخ به موقع به چنین حملاتی نیاز به شناسایی سریع عوامل ورودی درگیر دارد. این می‌تواند فریند پیچیده‌ای باشد که شامل روش‌ها و ابزارهای مختلفی است.

آژانس حفاظت از محیط زیست ایالات متحده (ای پی ای) با همکاری دکتر لارنس اچ کیث، رئیس اینستنس ریفرنس سورس و مشاوران دیگر، دانش، تجربه و تخصص و همچنین اطلاعات خود را برای توسعه‌ی مشاور سی بی آر با استفاده از نرم‌افزار اکسیس کورویدر اسناد ای پی ای در دسترس عموم قرار دادند.

یکی از مهم‌ترین بخش‌های سی بی آر ارائه‌ی مشاوره به شکل رویه‌های گام به گام منطقی تعیین هویت الف عامل سمی زمانی که اطلاعات کمی در دسترس باشد یا هیچ اطلاعاتی در دسترس نباشد، که در آغاز یک تروریست معمولی است حمله کنند. این سیستم به کارکنان پاسخگویی کمک می‌کند تا حتی در چنین محیط پر تنش بر اساس یک برنامه عملیاتی تثبیت شده به کار خود ادامه دهند. صفحه نمایش دوتایی این سیستم اطلاعات را در سه سطح ارائه می‌دهد: (۱) سطح ارشد/اجرایی با پاسخ‌های مختصر، (۲) سطح آموزشی با اطلاعات عمیق، و (۳) سطح تحقیق با پیوندهایی به اسناد دیگر، نمایش اسلاید، فرم‌ها، و سایت‌های اینترنتی سی بی آر. محتوای مشاور سی بی آر شامل موارد زیر است:

- چگونه هشدارهای تهدیدآمیز را دسته‌بندی کنند.
- چگونه یک ارزشیابی اولیه از تهدید به عمل آورند.
- چه اقدامات واکنشی فوری را اتخاذ کنند.
- چگونه مکان اولیه و ورودی امن به آن را ارزشیابی کنند.
- بهترین نمونه‌ها را از کجا و چگونه گردآوری کنند.
- چگونه نمونه‌ها را برای تحلیل بسته‌بندی و ارسال کنند.

محتوای محدود شامل عوامل سی بی آر و روش‌های تحلیل آنهاست. مشاور سی بی آر را می‌توان برای پاسخگویی به حوادث و/یا آموزش بکار برد. این سیستم دارای دو منوی متفاوت یکی برای پاسخ‌های اضطراری و دیگری برای آموزش است. این یک برنامه‌ی نرم‌افزاری محدود است و در دسترس عموم قرار ندارد.

سوالاتی برای مورد کاربردی ۱۲.۱

۶. مشاور سی بی آر چگونه می‌تواند به تصمیم‌گیری سریع کمک کند؟
 ۷. کدامیک از ویژگی‌های مشاور سی بی آر آن را به یک سیستم خبره تبدیل کرده است؟
 ۸. در چه موقعیت‌های دیگری می‌توان از سیستم‌های خبره مشابه استفاده کرد؟
- از سیستم‌های خبره همچنین در موقعیت‌های پرفشار که تصمیم‌گیرندگان انسانی اغلب نیاز دارند تا اقدامات را در کسری از ثانیه با در نظر گرفتن دانش عینی و ذهنی در پاسخ به موقعیت‌های اضطراری اتخاذ کنند بکار برده می‌شوند.

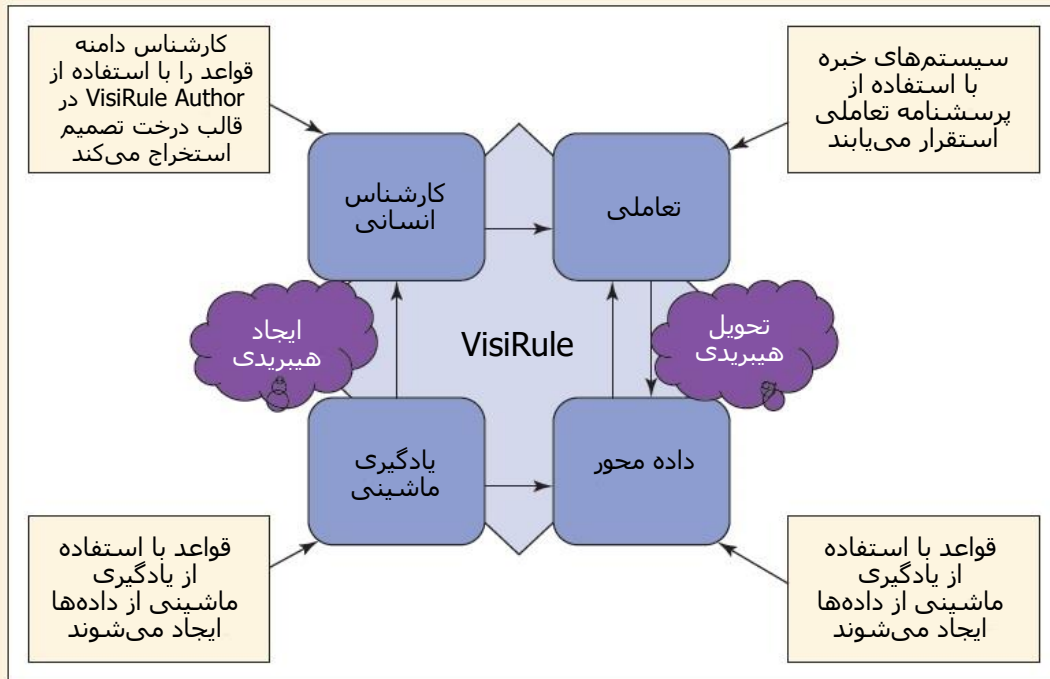
مزایای زیادی که تا پیش از این مطرح شد در پیاده‌سازی و اجرای بسیاری از سیستم‌های خبره در سرتاسر جهان تأثیر داشته‌اند. با وجود این، همانند بسیاری از فناوری‌های دیگر، سیستم‌های خبره‌ی کلاسیک با سیستم‌های بهتری جایگزین شده‌اند. در اینجا به برخی از محدودیت‌های سیستم‌های خبره که منجر به کاهش بکارگیری آنها شده‌اند می‌پردازیم.

۳. کسب دانش از کارشناسان انسانی به دلیل کمبود مهندسان دانشی خوب و نیاز به مصاحبه با چندین کارشناس خبره برای یک برنامه کاربردی می‌تواند بسیار پرهزینه باشد.
۴. هر دانش کسب شده‌ای باید با هزینه‌ی بالا مکرراً به‌روزرآوری شود.
۵. مبنای قاعده‌مند این سیستم‌ها چندان محکم و قابل اتکا یا انعطاف‌پذیر نبوده و ممکن است موارد استثنای بسیار زیادی برای قواعد وجود داشته باشد. سیستم‌های دانشی بهبود یافته از رویکردهای آماری و داده‌محور برای استدلال استفاده کرده و موفقیت‌های بهتری را کسب می‌کنند. علاوه بر این، روش استنتاج مبتنی بر مورد فقط در صورتی بهتر عمل می‌کند که تعداد موارد مشابه کافی وجود داشته باشد. بنابراین، معمولاً نمی‌تواند از سیستم خبره پشتیبانی کند.
۶. رابط کاربری قاعده‌محور می‌بایست توسط فناوری‌های دیگری از قبیل ارتباطات صوتی، نقشه‌های تصویری پشتیبانی شود. این موضوع می‌تواند استفاده از سیستم‌های خبره را بیش از اندازه دشوار و طاقت‌فرسا کند.
۷. قابلیت استنتاج فناوری قاعده‌محور در مقایسه با بکارگیری مکانیزم‌های جدیدتر از قبیل یادگیری ماشین بسیار محدود است.

نسل جدید سیستم‌های خبره به جای استفاده از سیستم قدیمی کسب و نمایش دانش، سیستم‌های خبره جدیدتر که مبتنی بر الگوریتم‌های یادگیری ماشینی و سایر فناوری‌های هوش مصنوعی هستند برای ایجاد سیستم‌های بهتر استقرار یافته‌اند. یک مثال در مورد کاربردی ۱۲.۲ ارائه شده است.

ویزی‌رول^۱ یک شرکت قدیمی سیستم خبره است که مدل کسب‌وکار خود را در طول زمان تغییر داد. ویزی‌رول (در بریتانیا) ابزارهای ترسیم نمودار را برای تسهیل ایجاد و ساخت سیستم‌های خبره عرضه می‌کند. با استفاده از ترسیم نمودار کاربران می‌توانند دانش موجود در سیستم‌های خبره را راحت‌تر استخراج کرده و بکار بگیرند. فرایند ساخت پایگاه دانش را در قسمت چپ شکل ۱۲.۲ مشاهده می‌کنید. در سمت چپ، ایجاد و ساخت هیبریدی را مشاهده می‌کنند. با استفاده از یک درخت تصمیم، کارشناسان دامنه می‌توانند قواعد بیشتری را مستقیماً از داده‌های مربوطه ایجاد کنند (مثلاً داده‌های تاریخی). علاوه بر این قواعد را می‌توان با استفاده از یادگیری ماشینی ایجاد کرد.

سمت راست تصویر (گوشه بالایی سمت راست) تحویل هیبریدی (مشاوره) را به تصویر می‌کشد. این سیستم با استفاده از پرسش و پاسخ‌ها می‌تواند مشاوره دهد. علاوه بر این، قواعد را می‌توان برای پردازش از راه‌دور داده‌ها و به‌روزرسانی مخزن داده‌ها بکار برد. توجه داشته باشید که گزینه‌ی تحویل دوگانه مبتنی بر توانایی یادگیری ماشینی برای کشف الگوهای پنهان در داده‌هایی است که بتوان از مدل‌های تصمیم‌گیری پیشگویانه استفاده کنند.



شکل ۱۲.۲ فرایند سیستم‌های توصیه‌گر

ویزی‌رول همچنین چت‌بات‌هایی را برای بهبود بخش تعاملی فرایند ارائه کرده و یک نقشه تعاملی را تأمین می‌کند. بر اساس وب سایت شرکت visirule.co.uk، مزایای اصلی این محصول به شرح زیر هستند:

- نیازی به برنامه‌نویسی ندارد.

¹ VisiRule

- دیاگرام‌ها توسط کارشناسان انسانی ترسیم شده یا به صورت خودکار از داده‌ها استخراج می‌شوند.
- حاوی ابزارهای خودارزیابی با امکان تولید گزارش و اسناد است.
- دانش تولید شده می‌توان به آسانی به صورت کد ایکس‌ام‌ال اجرا کرد.
- توضیح و توجیه مورد نیاز را ارائه می‌کند.
- مشاوره‌های تعاملی کارشناسان می‌تواند مشتریان جدیدی را جذب کند.
- برای آموزش و مشاوره‌های کارکنان می‌تواند مورد استفاده قرار گیرد.
- شرکت‌ها به آسانی می‌توانند به مخزن دانش سازمانی دست پیدا کنند.
- نمودارها باید از ابزارهای ویزی‌رول که با درخت‌های تصمیم و ابزارهای ترسیم فلوجارت کشیده شده‌اند استفاده کنند.
- نمودارها امکان ایجاد مدل‌هایی را می‌دهند که بلافاصله می‌توان آنها را اجرا و صحت‌گذاری کرد.

از همه اینها گذشته، ویزی‌رول یک سیستم خبره‌ی جامع مبتنی بر هوش مصنوعی را ارائه می‌کند.

Source: Courtesy of VisiRule Corp. UK. Used with permission.

سوالاتی برای مورد کاربردی ۱۲.۲

۱. کدامیک از محدودیت‌های سیستم‌های خبره‌ی اولیه توسط سیستم ویزی‌رول رفع شده‌اند؟
۲. شکل ۱۲.۱ و ۱۲.۲ را مقایسه کنید. چه تفاوت‌هایی بین زیرسیستم‌های ایجاد (شکل ۱۲.۲) و توسعه (شکل ۱۲.۱) وجود دارد؟
۳. شکل ۱۲.۱ و ۱۲.۲ را مقایسه کنید. چه تفاوت‌هایی بین زیرسیستم‌های تحویل (شکل ۱۲.۲) و مشاوره (شکل ۱۲.۱) وجود دارد؟
۴. تمامی فناوری‌های هوش مصنوعی را شناسایی کرده و نقش آنها در سیستم ویزی‌رول را نام ببرید.
۵. برخی مزایای این سیستم خبره برای کاربران را نام ببرید.

از سیستم‌های خبره همچنین در موقعیت‌های پرفشار که تصمیم‌گیرندگان انسانی اغلب نیاز دارند تا اقدامات را در کسری از ثانیه با در نظر گرفتن دانش عینی و ذهنی در پاسخ به موقعیت‌های اضطراری اتخاذ کنند بکار برده می‌شوند.

سه نوع از کاربردهای اصلی هوش مصنوعی که بر محدودیت‌های قبلی که به آنها اشاره کردیم غلبه کردند چت‌بات‌ها، دستیارهای شخصی مجازی و مشاوران ربات هستند که در ادامه این فصل به آنها خواهیم پرداخت. از دیگر فناوری‌های هوش مصنوعی که فعالیت‌های مشابهی را انجام می‌دهند در فصل‌های ۴ تا ۹ معرفی شده‌اند. معروف‌ترین آنها آی‌بی‌ام واتسون است. برخی از قابلیت‌های مشاوره آن مشابه سیستم‌های خبره بوده ولی بسیار برتر از آن هستند.

فناوری هوش مصنوعی مشابه دیگر یعنی سیستم توصیه‌گر درخش بعدی معرفی شده است. گونه‌های جدیدتر آن از یادگیری ماشین و تحلیل‌های آی‌بی‌ام واتسون استفاده می‌کنند.

یکی از سیستم‌های دانشی که شدیداً برای توصیه‌ی محصولات یا خدمات یک به یک استفاده می‌شود، سیستم توصیه‌گر یا سیستم توصیه یا موتور توصیه نام دارد. چنین سیستمی سعی دارد تا اهمیت (اولویت یا امتیاز) اینکه یک کاربر از یک محصول یا خدمت خوشش بیاید را پیش‌بینی کند. پس از مشخص شدن این اهمیت، سلیق و ترجیحات کاربر به اطلاع یک فروشنده رسانده شده و می‌تواند محصول یا خدمت خودش را با آنچه کاربر خواسته است تطبیق دهد. برای اطلاعات جامع نگاه کنید به آگاروال (۲۰۱۶). برای یک مطلب آموزشی جامع و مطالعه موردی نگاه کنید به

analyticsvidhya.com/blood/2015/10/recommendation-engine

سیستم‌های توصیه بسیار متداول هستند و در بسیاری از حوزه‌ها کاربرد دارند. کاربردهای معمول آنها شامل فیلم، موسیقی و کتاب است. با وجود این، سیستم‌هایی نیز برای سفر، رستوران‌ها، بیمه و دوست‌یابی آنلاین وجود دارد. توصیه‌ها معمولاً به ترتیب اولویت ارائه می‌شوند. بسیاری از مردم توصیه‌های آنلاین را به جستجوهای معمول ترجیح می‌دهند زیرا شخصی‌تر، کندتر و برخی اوقات از دقت کمتری برخوردارند.

مزایای سیستم‌های توصیه استفاده از این سیستم‌ها می‌تواند منجر به مزایای فوق‌العاده‌ای برای خریداران و فروشندگان شود (ماکادیا ۲۰۱۸).

مزایای این سیستم‌ها برای مشتریان:

- شخصی‌سازی. آنها توصیه‌هایی که بسیار نزدیک به آن چیزی است که نیاز دارند یا دوست دارند را دریافت می‌کنند. این البته به کیفیت روش مورد استفاده نیز بستگی دارد.
- کشف. آنها ممکن است توصیه‌هایی را برای محصولاتی دریافت کنند حتی از وجود آنها اطلاع نداشته‌اند ولی واقعاً به آنها نیاز داشته‌اند.
- رضایت مشتری. رضایت مشتری با تکرار توصیه‌ها افزایش می‌یابد.
- گزارشات. برخی توصیه‌گرها گزارشات و دیگر توضیحاتی را در مورد محصولات انتخابی ارائه می‌کنند.
- افزایش گفتگو با فروشنده. از آنجائیکه توصیه‌ها ممکن است همراه با توضیحات باشند، خریداران ممکن است خواهان تعامل بیشتری با فروشندگان باشند.

مزایای آن برای فروشندگان:

- نرخ بالای تبدیل مشتری. با ارائه‌ی توصیه‌های شخصی‌سازی شده محصول، خریداران تمایل به خرید بیشتری دارند.
- افزایش فروش جانبی. سیستم‌های توصیه می‌توانند محصولات دیگری را نیز توصیه کنند. به عنوان مثال، آمازون کام سایر محصولاتی که افراد دیگری به همراه محصولی که شما سفارش داده‌اید را نمایش می‌دهند.
- افزایش وفاداری مشتری. با افزایش مزایای این سیستم‌ها برای مشتریان، وفاداری آنها به فروشنده نیز افزایش می‌یابد.
- امکان سفارشی‌سازی انبوه. اطلاعات بیشتری را در مورد سفارشات شخصی‌سازی شده بالقوه ارائه می‌کند.

از روش‌های دیگری نیز برای ایجاد سیستم‌های توصیه استفاده می‌شود. دو روش کلاسیک عبارتند از فیلترینگ مشارکتی^۱ و فیلترینگ مبتنی بر محتوا^۲.

فیلترینگ مشارکتی این روش مدلی را می‌سازد که خلاصه‌ای از رفتار گذشته‌ی خریداران، نحوه‌ی جستجوی آنها در اینترنت، آنچه به دنبالش بوده‌اند، آنچه خریداری کرده‌اند و تا چه اندازه از محصولات خوششان آمده را تهیه می‌کند. علاوه بر این، در فیلترینگ مشارکتی آنچه خریدارانی با پروفایل‌های مشابه خریداری کرده‌اند و چگونه به خریدهایشان امتیاز داده‌اند در نظر گرفته می‌شود. این روش از الگوریتم‌های هوش مصنوعی برای پیش‌بینی اولویت‌ها بر اساس صفات محصولاتی که مشتریان خریداری کرده یا قصد خرید آنها را دارند می‌پردازد. فروشنده با اطلاع از این اولویت‌ها محصولاتی را با صفات مشابه به مشتریان توصیه می‌کند. به عنوان مثال، سیستم ممکن است یک کتاب متن کاوی را به مشتری که به داده کاوی علاقمند بوده توصیه کند. هر کدام از اینها مزایا و محدودیت‌هایی دارند (نگاه کنید به مثال Wikipedia.org/wiki/Recommender_system). برخی اوقات این دو روش در قالب یک روش با یکدیگر ترکیب می‌شوند. روش‌های فیلترینگ دیگری نیز وجود دارند. مثال‌هایی از آنها شامل فیلترینگ مبتنی بر قاعده و فیلترینگ مبتنی بر فعالیت است. روش‌های جدیدتر شامل یادگیری ماشین و فناوری‌های هوش مصنوعی هستند که در مورد کاربردی ۱۲.۳ نشان داده شده است.

توصیه گر نتفلیکس: یک عامل کلیدی موفقیت

مورد کاربردی ۱۲.۳

طبق ir.netflix.com نتفلیکس بزرگ‌ترین شبکه تلویزیونی اینترنتی جهان با بیش از ۱۱۸ میلیون عضو در بیش از ۱۹۰ کشور است که از بیش از ۱۵۰ میلیون ساعت فیلم و نمایش تلویزیونی از جمله سریال‌ها، مستندها و ... در روز لذت می‌برند. اعضا می‌توانند تعداد نامحدودی از نمایش‌ها را بدون آگهی‌های تبلیغاتی به ازای پرداخت ماهانه تماشا کنند.

چالش‌ها

نتفلیکس چندین میلیون عنوان نمایش داشته و در حال حاضر نمایش‌های خاص خودش را تولید می‌کند. این مجموعه‌ی بزرگ از عناوین اغلب برای مشتریانی که در برای انتخاب کردن مشکل دارند دردسر ساز می‌شود. یک چالش دیگر این است که نتفلیکس کسب‌وکار خود را از ایالات متحده و کانادا به ۱۹۰ کشور دیگر گسترش داده است. نتفلیکس در یک محیط بسیار رقابتی فعالیت می‌کند که در آن فعالان بزرگی همچون اپل، آمازون، کام و گوگل فعالیت می‌کنند. نتفلیکس به دنبال راهی برای متمایز کردن خودش در محیط رقابتی با ارائه‌ی توصیه‌های مفید به مشتریانش بود.

موتور جستجوی اولیه

نتفلیکس در ابتدا صرفاً یک شرکت کسب‌وکار سفارش پستی دی‌وی‌دی بود. در آن زمان به دلیل اینکه مشتریانش برای اینکه انتخاب کنند کدام دی‌وی‌دی را اجاره کنند در موجودی محصولاتش به مشکلاتی برخورد. راه‌حل این مشکل توسعه‌ی یک موتور توصیه بود (به نام سینه‌مچ^۳) که به اعضا می‌گفت که ممکن است از کدام عناوین نمایش‌ها و فیلم‌ها خوششان بیاید. سینه‌مچ از ابزارهای داده‌کاوی برای جستجو در پایگاه داده‌های از میلیاردها تاریخچه‌ی اجاره‌ی دی‌وی‌دی و امتیازدهی فیلم‌ها استفاده کرد. نتفلیکس با استفاده از الگوریتم‌های انحصاری توانست عناوین‌هایی را به مشتریان توصیه کند. توصیه از طریق مقایسه‌ی اولویت‌ها، ترجیحات و علایق مشتریان با آنهایی که سلیقه مشابهی داشته‌اند

¹ Collaborative Filtering

² Content-based Filtering

³ Cinematch

و با استفاده از گونه‌ای از فیلترینگ مشارکتی محقق میشد. سینه‌مچ مثل یک کارمند در یک فروشگاه کوچک فیلم و دی‌وی‌دی بود که عناوینی که می‌دانست شما خوشتان خواهد آمد را کنار گذشته و زمانیکه به فروشگاه می‌رفتید به شما توصیه می‌کرد. برای افزایش دقت سینه‌مچ، نتفلیکس مسابقه‌ای را در اکتبر ۲۰۱۶ آغاز کرد که به اولین شخص یا تیمی که برنامه‌ای را می‌نوشت که بتواند دقت پیش‌بینی سینه‌مچ را حداقل ۱۰ درصد افزایش دهد ۱ میلیون دلار پرداخت می‌کرد. شرکت متوجه شد که اینکار زمان‌بر است و به همین دلیل هر ساله یک جایزه‌ی پیشرفت ۵۰,۰۰۰ دلاری را برای مسابقه در نظر می‌گیرد. پس از گذشت بیش از ۲ سال از این مسابقه، جایزه بزرگ به تیم پراگماتیک کی‌آز از بالتیمور رسید که ترکیبی از دو تیم راه یافته به مرحله پایانی بودند. برای آشنایی با نحوه کارکرد الگوریتم‌های توصیه‌ی فیلم‌ها به [quora.com/How-does-the-Netflix-movie-recommendation-algorithm-work/](https://www.quora.com/How-does-the-Netflix-movie-recommendation-algorithm-work/) مراجعه کنید.

دوران جدید

با گذشت زمان، نتفلیکس به سمت کسب‌وکار پخش فیلم و سپس تلویزیون اینترنتی رفت. همچنین، گسترش فناوری/ابر موجب شد تا بهبودهایی در سیستم توصیه به وجود آید. این سیستم جدید ارائه توصیه بر اساس آنچه مردم در گذشته دیده‌اند را متوقف کرده و در عوض، از ابر آمازون برای تقلید از مغز انسان برای یافتن آنچه مردم واقعاً در فیلم‌ها و نمایش‌های مورد علاقه‌شان دوست دارند استفاده می‌کند. این سیستم مبتنی بر هوش مصنوعی بوده و از فناوری یادگیری عمیق استفاده می‌کند. شرکت اکنون از کلان‌داده‌ها استفاده کرده و قادر به استخراج بینش از این توصیه‌ها می‌باشد. برای تولیدات این شرکت از تجزیه و تحلیل نیز استفاده می‌شود. یک تغییر بزرگ دیگر در تبدیل این شرکت به یک حوزه‌های جهانی بود. در گذشته، توصیه‌ها بر اساس اطلاعات گردآوری شده در داخل کشور (یا منطقه) که کاربران زندگی می‌کردند ارائه می‌شدند. توصیه‌ها مبتنی بر چیزهایی بودند که سایر افراد در همان کشور از آن لذت برده بودند. این روش به دلیل تفاوت‌های فرهنگی، سیاسی و اجتماعی در محیط بین‌المللی به خوبی جواب نداد. در سیستم اصلاح‌شده، دیدگاه‌های مردم در بسیاری از کشورهای دیگر و عادات و سلیقه‌شان در نظر گرفته می‌شود. پیاده‌سازی و اجرای سیستم جدید دشوار بود به ویژه زمانیکه یک کشور یا منطقه جدید افزوده میشد. توصیه‌ها در ابتدا بدون داشتن اطلاعات کافی و زیاد در مورد مشتریان جدید ارائه می‌شدند. اصلاح سیستم توصیه نزدیک به ۱ سال کار با ۷۰ مهندس به طول انجامید. برای جزئیات نگاه کنید به پاپر^۱ (۲۰۱۶).

نتایج

در نتیجه‌ی پیاده‌سازی سیستم توصیه، فروش و عضویت‌های نتفلیکس رشد بسیار سریعی پیدا کرد. مزایای این سیستم عبارتند از:

- **توصیه‌های مؤثر.** بسیاری از اعضای نتفلیکس فیلم‌هایشان را بر اساس توصیه‌هایی که متناسب با ذائقه و سلیقه فردی‌شان توصیه می‌شود انتخاب می‌کنند.
- **رضایت مشتری.** بیش از ۹۰ درصد از اعضای نتفلیکس می‌گویند که از خدمات نتفلیکس بسیار راضی هستند و آن را به سایر اعضای خانواده و دوستانشان نیز توصیه می‌کنند.
- **مالی.** تعداد اعضای نتفلیکس از ۱۰ میلیون نفر در سال ۲۰۰۸ به ۱۱۸ میلیون نفر در سال ۲۰۱۸ افزایش یافته است. فروش و سود این شرکت دائماً در حال رشد است. در بهار ۲۰۱۸، سهام نتفلیکس به قیمت بیش از ۴۰۰ دلار به ازای هر سهم فروخته شد که این عدد در سال گذشته ۱۴۰ دلار بوده است.

Sources: Based on Popper (2016), Arora (2016), and StartUp (2016).

سوالاتی برای مورد کاربردی ۱۲.۳

۱. چرا سیستم توصیه‌گر مفید است؟ (ارتباط آن با بازاریابی هدفمند یک به یک را توضیح دهید).

¹ Popper

۲. توضیح دهید که توصیه‌ها چگونه تولید می‌شوند؟
۳. آمازون الگوریتم‌های توصیه‌ی خود را برای عموم افشا کرد ولی نتلفیکس اینکار را نکرد. چرا؟
۴. در مورد فعالیت‌های پژوهشی که سعی دارند تا از مغز انسان تقلید کنند تحقیق کنید.
۵. تغییرات ناشی از جهانی شدن شرکت نتلفیکس را توضیح دهید.

سوالاتی برای مرور بخش ۱۲.۲

۱۰. سیستم‌های خبره را تعریف کنید.
۱۱. هدف اصلی یک سیستم خبره چیست؟
۱۲. افراد خبره را توضیح دهید.
۱۳. منظور از تخصص چیست؟
۱۴. برخی حوزه‌های قابل بررسی مختص سیستم خبره را نام ببرید.
۱۵. کامپوننت‌های اصلی سیستم خبره را نام برده و هریک را به اختصار توضیح دهید.
۱۶. چرا کاربرد سیستم‌های خبره در حال کاهش است؟
۱۷. سیستم‌های توصیه را تعریف کرده و عملیات و مزایای آنها را توضیح دهید.
۱۸. ارتباط بین سیستم‌های توصیه و هوش مصنوعی را توضیح دهید.

از آنهایی که استفاده نکرده‌اند علاقمند به اینکار هستند. نسل هزاره‌ها تنها نسلی نیستند که از چت‌بات‌ها استفاده می‌کنند اگرچه آنها بیش از سایرین از این فناوری استفاده می‌کنند. موضوع بحث ما در این بخش این است که چت‌بات‌ها چه هستند و چکار می‌کنند.

چت‌بات^۱ که شکل اختصاری چت ربات است و با نام‌های «بات» یا «روبو» نیز شناخته می‌شود یک سرویس رایانه‌ای است که امکان برقراری آسان مکالمات بین انسان‌ها و ربات‌های انسان‌نمای رایانه‌ای یا کاراکترهای تصویری را برخی اوقات از طریق اینترنت فراهم می‌کند. مکالمه می‌تواند در قالب نوشتار و بیش از همه صدا و تصویر باشد. مکالمات اغلب شامل پرسش و پاسخ‌های کوتاه بوده و به زبان طبیعی انجام می‌شوند. چت‌بات‌های هوشمندتر منجر به پردازش زبان طبیعی هستند که رایانه بتواند دیالوگ‌های ساخت‌نیافته را درک کند. تعاملات نیز ممکن است با گرفتن عکس و آپلود کردن آنها رخ دهد (مثلاً همانند سامسونگ بیکس‌بای در سامسونگ S8 و 8). برخی شرکت‌ها در حال آزمایش چت‌بات‌های *آموزنده*^۲ هستند که از تجربیات تجمیع شده دانش بیشتری را کسب می‌کنند. توانایی رایانه برای گفتگو و مکالمه با انسان از طریق یک سیستم دانشی (مثلاً مبتنی بر قاعده) و یک قابلیت درک زبان طبیعی میسر می‌شود. این سرویس اغلب در خدمات پیام‌رسانی مثل فیسبوک مسنجر یا وی‌چت و توئیتر در دسترس است.

¹ Chatbot

² Learning Chatbots

چت‌بات‌ها چندین دهه پیش به وجود آمدند. آنها سیستم‌های خبره‌ی ساده‌ای بودند که ماشین‌ها را قادر به پاسخگویی به سوالات کاربران می‌کردند. اولین نوع شناخته شده این نوع ماشین‌ها الیزا^۱ نام داشت (en.wikipedia.org/wiki/ELIZA). الیزا و ماشین‌های مشابه برای کار در حالت پرسش و پاسخ توسعه یافتند.

search.cpan.org/dist/Chatboat-Eliza/Chatbot/Eliza.pm

سوفیا یک چت‌بات است که در هنگ کنگ ایجاد شده و جایزه‌ی شهروندی را در اکتبر ۲۰۱۷ از عربستان سعودی دریافت کرد. از آنجائیکه مسلمان نیست حجاب ندارد. می‌تواند به سوالات بسیاری پاسخ دهد. برای جزئیات نگاه کنید به newsweek.com/Saudi-arabia-robot-sophia-muslim-694152/

انواع بات‌ها را می‌توان بر اساس قابلیت‌هایشان طبقه‌بندی کرد. سه طبقه‌بندی بات‌ها به شرح زیر هستند:

۱. **بات‌های معمولی.** این بات‌ها در اصل عوامل هوشمند مکالمه‌ای هستند (فصل ۲). آنها می‌توانند وظایف ساده و تکرارپذیر را برای صاحبان‌شان انجام دهند مثل نشان دادن مانده حساب بانکی، کمک به خرید آنلاین کالاها و خرید یا فروش آنلاین سهام.
۲. **چت‌بات‌ها.** در این گروه، بات‌های توانمندتری را معرفی می‌کنیم به عنوان مثال آنهایی که می‌توانند مکالمات با انسان را شبیه‌سازی کنند. این فصل عمدتاً با این چت‌بات‌ها مرتبط است.
۳. **بات‌های هوشمند.** این بات‌ها پایگاه دانشی دارند که با کسب تجربه بهبود پیدا می‌کند. یعنی این بات‌ها می‌توانند یاد بگیرند مثلاً اولویت‌ها و ترجیحات مشتریان (مثلاً مثل الکسا و برخی مشاوران روبو).

یک محدودیت اصلی انواع قدیمی‌تر بات‌ها به‌روزرآوری پایگاه دانش آنها بود که هم کند و هم پرهزینه بود. این بات‌ها برای دامنه‌های تخصصی محدود و/یا کاربران خاصی طراحی شده بودند. بهبود پشتیبانی از این فناوری سالیان زیادی طول کشید. پردازش زبان طبیعی روز به روز بهتر شد. پایگاه‌های دانش امروزه در فضای «ابر» در یک مکان متمرکز به‌روزرآوری می‌شوند. دانش توسط تعداد زیادی از کاربران به اشتراک گذاشته می‌شود و بدین ترتیب هزینه به ازای هر کاربر کاهش می‌یابد. دانش ذخیره شده با سوالاتی که توسط کاربران پرسیده می‌شود انطباق داده می‌شود. پاسخ‌ها به طور قابل توجهی توسط ماشین‌ها بهبود یافته است. از سال ۲۰۰۰، ماشین‌های هوش مصنوعی توانمندتری را برای دیالوگ‌های پرسش و پاسخ شاهد بوده‌ایم. حدود

¹ Eliza

سال‌های ۲۰۱۰، ماشین‌های هوش مصنوعی محاوره‌ای به نام چت‌بات‌ها نام‌گذاری شده و بعدها به شکل دستیارهای شخصی مجازی توسعه یافتند که پیشگام آن الکسا آمازون است.

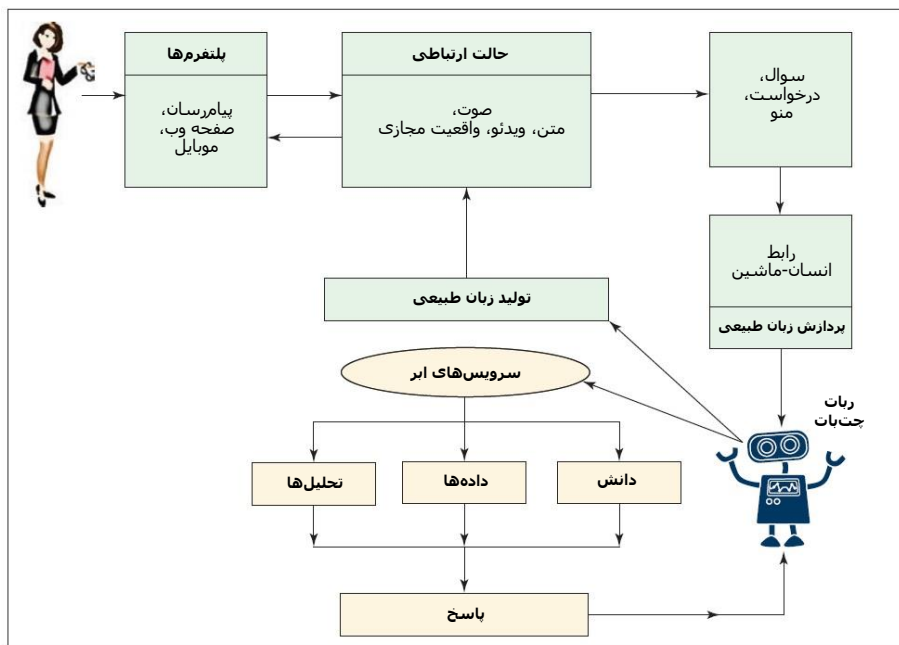
محرك‌های چت‌بات‌ها محرك‌های اصلی عبارتند از:

- توسعه‌دهندگان در حال ایجاد ابزارهای قدرتمندی برای ساخت سریع و ارزان چت‌بات‌ها با کارکردهای مفید هستند.
- کیفیت چت‌بات‌ها در حال بهبود است و به همین دلیل مکالمات برای کاربران مفیدتر شده است.
- تقاضا برای چت‌بات‌ها به دلیل توان بالقوه‌ای که در کاهش هزینه و بهبود خدمات مشتری و خدمات بازاریابی دارند در حال رشد هستند.
- استفاده از چت‌بات‌ها رشد سریع بدون نیاز به استخدام و آموزش کارکنان خدمات مشتریان را فراهم می‌کند.
- شرکت‌ها با استفاده از چت‌بات‌ها می‌توانند از سیستم‌های پیام‌رسان و اپلیکیشن‌های مربوطه که برای مشتریان جذاب هستند بهره‌برداری کنند.

کامپوننت‌های اصلی چت‌بات‌ها عبارتند از:

- شخص
 - یک رایانه، آواتار یا ربات (ماشین هوش مصنوعی).
 - پایگاه دانش که بتوان آن را در یک ماشین تعبیه کرد یا به یک « ابر » متصل نمود.
 - یک رابط انسان-رایانه که برقراری دیالوگ در حالات نوشتاری و گفتاری را فراهم می‌کند.
 - یک زبان پردازش طبیعی که ماشین را قادر می‌سازد تا زبان طبیعی را درک کند.
- چت‌بات‌های پیشرفته می‌توانند ژست، سرنخ و تغییرات صوتی انسان را درک کنند.
- فرایند تعامل شخص-ماشین** کامپوننت‌هایی که در بالا ذکر شدند چارچوبی را برای محاوره بین انسان-بات فراهم می‌کنند. شکل ۱۲.۳ فرایند محاوره و گفتگو را نشان می‌دهد.

- شخص (سمت چپ تصویر) به دنبال یافتن اطلاعات یا کمک است.
- شخص یک سوال مرتبط را به صورت صوتی، متن و غیره از بات می‌پرسد.
- پردازش زبان طبیعی سوال را به زبان ماشین ترجمه می‌کند.
- چت‌بات‌ها سوال را به خدمات ابر منتقل می‌کنند.
- ابر حاوی یک پایگاه دانش، منطق کسب‌وکار و ابزارهای تحلیل برای ارائه پاسخ به سوال است.
- پاسخ به برنامه‌ی تولید زبان طبیعی و سپس به شخصی که سوال را پرسیده است منتقل می‌شود.



شکل ۱۲.۲ فرایند گفتگو با چت‌بات‌ها

عوامل و مزایای زیر موجب افزایش کاربرد چت‌بات‌ها شده است:

- نیاز به کاهش هزینه‌ها
- قابلیت‌های روزافزون هوش مصنوعی به ویژه فناوری‌های پردازش زبان طبیعی و فناوری‌های صوتی.
- توانایی تبدیل به زبان‌های مختلف (از طریق ترجمه ماشینی).
- افزایش کیفیت و توانایی ثبت دانش.
- عرضه‌ی دستگاه‌ها توسط فروشندگان (مثلاً دستیارهای شخصی مجازی مثل الکسا آمازون و دستیار گوگل شرکت آلفابت).
- استفاده از آنها برای ارائه‌ی خدمات اقتصادی و فوق‌العاده و انجام پژوهش بازار.
- استفاده برای تشخیص متن و تصویر.
- استفاده برای تسهیل خرید.
- پشتیبانی از تصمیم‌گیری.

چت‌بات‌ها و ماشین‌های هوش مصنوعی مشابه در طول زمان بهبود یافته‌اند. چت‌بات‌ها هم برای کاربران و هم سازمان‌ها دارای مزیت هستند. به عنوان مثال، چندین بیمارستان از پذیرشگرهای ربات برای هدایت بیماران به محل درمان استفاده می‌کنند. زورا رباتیکز یک ربات به نام نائو را ساخته است که به عنوان یک همراه افراد سالمند و بیمار عمل می‌کند. بات‌ها به صورت یک تراپی برای آنهایی که دچار زوال عقل هستند عمل می‌کنند.

برای اطلاع از دایرکتوری چت‌بات‌ها که شامل بیش از ۱,۲۵۰ بات در ۵۳ کشور تا آوریل ۲۰۱۸ است نگاه کنید به chatbots.org/ و botlist.co/bots. مثال‌هایی از چت‌بات‌ها و کارهایی که می‌توانند انجام بدهند در chatbot.org آورده شده است:

- **روبوکوک^۱**: این یک بات برای توصیه‌ی موسیقی و مهمانی است که برای کوکاکولا در مجارستان ایجاد شده است.
- **کیپ^۲**: این بات برای کمک به خرید در پلتفرم پیام‌رسانی اسلک طراحی شده است. به کیپ بگویید چه چیزی می‌خواهید بخرید و کیپ آن را یافته و حتی برای شما خریداری می‌کند.
- **وال‌نات^۳**: این چت‌بات می‌تواند مهارت‌های مربوط به شما را کشف کرده و در یادگیری آنها به شما کمک کند. مجموعه بزرگی از داده‌ها را برای کشف مهارت‌ها می‌تواند تجزیه و تحلیل کند.
- به اشتراک‌گذاری خودرو توسط تاکسی‌بات^۴: اگر مطمئن نیستید که کدامیک از اوپر، لیفت، گرب یا کامفورت دل‌گرو ارزان‌ترین سرویس را می‌دهند می‌توانید از این بات بپرسید. علاوه بر این، می‌توانید کدهای تخفیف جاری را دریافت کنید.
- **شاپی‌بات^۵**: زمانیکه تصویری از یک محصول را برای این بات ارسال کنید، محصولات مشابه را در عرض چند ثانیه برایتان ارسال می‌کند. به شاپی‌بات بگویید به دنبال چه محصولی و به چه قیمتی هستید و بهترین گزینه را برایتان پیدا می‌کند.
- **کانسرنینگ دیزا پُرد تریپ^۶**: می‌تواند به سوالات شما در مورد رویدادها، رستوران‌ها و جاذبه‌های مقاصد گردشگری مهم کمک کند.
- **بی‌اُتی^۷**: اولین چت‌بات بولیویایی که به زبان اسپانیایی با شما صحبت کرده و به سوالات‌تان در مورد کشور بولیوی، فرهنگ، جغرافیا، جامعه و ... پاسخ می‌دهد.
- **هیزی^۸**: دستیار دیجیتالی شما که هدف آن کمک به از بین بردن فاصله‌ی بین شما و اقدام بعدی شما برای شغل‌تان است. متقاضیان کار می‌توانند مستقیماً با هیزی همانند موسسات شغلیابی و جابجایی شغل یا دوستان‌شان ارتباط برقرار کنند.
- **گرین کارت^۹**: این ویزابات^{۱۰} به کاربران کمک می‌کند تا درخواست‌هایشان برای گرین کارت ایالات متحده را به شکلی مناسب ارسال کنند.

¹ RoboCoke

² Kip

³ WalNut

⁴ Taxi Bot

⁵ ShopiiBot

⁶ Concerning Desired Trips

⁷ BO.T

⁸ Hazie

⁹ Green Card

¹⁰ Visabot

- **زوم^۱ (Zoom.ai (botlist.co/bots/369-zoomai)** یک دستیار مجازی مستقل و خودکار برای هر کس در محل کار است.
- **آکیتا^۲**. این چت بات (botlist.co/bots/1314-akita) می تواند شما را به کسب و کارهای منطقه اطرافتان متصل کند.

همانگونه که می بینید، از چت بات ها می توان برای انجام وظایف متفاوتی استفاده کرد. مورگان (۲۰۱۷) بات ها را به گروه های زیر دسته بندی می کند: آموزش، بانک داری، بیمه، خرده فروشی، سفر، مراقب سلامت و تجربه مشتری.

گروه های اصلی کاربردهای چت بات ها امروزه برای مقاصد زیادی و در صنایع و کشورهای بسیاری مورد استفاده قرار می گیرند. ما کاربردهای آنها را به گروه های زیر تقسیم می کنیم:

- چت بات ها برای فعالیتهای سازمانی از جمله ارتباطات، همکاری، خدمات مشتری و فروش بکار می روند که در بخش ۱۲.۴ توضیح داده شده است.
 - چت بات هایی که به عنوان دستیارهای شخصی عمل می کنند که در بخش ۱۲.۵ معرفی شده اند.
 - چت بات هایی که به عنوان مشاور و عمدتاً در خصوص موضوعات مرتبط با امور مالی عمل می کنند (بخش ۱۲.۶).
- برای بحث و گفتگو در خصوص این گروه ها به فِرون^۳ (۲۰۱۷) نگاه کنید.

سوالاتی برای مرور بخش ۱۲.۳

۱. چت بات ها را تعریف کرده و کاربرد آنها را توضیح دهید.
۲. کامپوننت های اصلی چت بات ها را نام ببرید.
۳. محرک های اصلی فناوری چت بات ها چه هستند؟
۴. چت بات ها چگونه کار می کنند؟
۵. چرا چت بات ها را ماشین های هوش مصنوعی می دانند؟

¹ Zoom

² Akita

³ Ferron

مزایای چتبات‌ها برای سازمان‌ها به سرعت در حال افزایش بوده و موجب کاهش هزینه‌های دیالوگ‌ها و سازگاری بیشتر آنها می‌شود. چتبات‌ها می‌توانند با مشتریان و شرکای تجاری به شکل مؤثرتری تعامل برقرار کرده، در هر زمان در دسترس بوده و از هر مکان می‌توان به آنها دسترسی پیدا کرد. کسب‌وکارها مشخصاً به رشد و تکامل چتبات‌ها توجه دارند. طبق بیور^۱ (۲۰۱۶)، کسب‌وکارها باید به دلایل زیر به دنبال چتبات‌ها باشند:

- هوش مصنوعی به مرحله‌ای رسیده است که در آن چتبات‌ها می‌توانند به طور فزاینده‌ای درگیر مکالمات با انسان‌ها شوند و به کسب‌وکارها اجازه دهند تا از این فناوری ارزان قیمت و قابل دسترس برای همگان برای درگیرسازی مشتریان بیشتر استفاده کنند.
- چتبات‌ها مخصوصاً برای کاربردهای موبایل و اپلیکیشن‌ها مناسب هستند. پیام‌رسانی قلب تجربه‌ی موبایل است که استفاده سریع از اپلیکیشن‌های چت گوپای این مسئله است.
- اکوسیستم چتبات از استحکام و قوت بالایی برخوردار است که بسیاری از چتبات‌های شخص ثالث، بات‌های بومی، کانال‌های توزیع و شرکت‌های فناوری را شامل می‌شود.
- چتبات‌ها می‌توانند برای اپلیکیشن‌های پیام‌رسانی و برنامه‌نویسانی که بات‌ها را برای این پلتفرم‌های ارائه می‌کنند بسیار پرمفعت باشند مشابه آنچه فروشگاه‌های اپلیکیشن‌ها را به اکوسیستم‌های پول‌ساز تبدیل کرده است.

مطالعه‌ای که در سال ۲۰۱۶ انجام شده نشان می‌دهد که ۸۰ درصد از کسب‌وکار تا سال ۲۰۲۰ خواهان چتبات‌ها هستند. businessinsider.com/80-of-businesses-want-chatbots-by-2020-2016-12. برای اطلاعات بیشتر نگاه کنید به نایت^۲ (۲۰۱۷).

همانطور که در قسمت ابتدایی این فصل دیدیم و در چندین مثال بعدی در این فصل خواهیم دید، چتبات‌ها در ارائه خدمات بازاریابی و مشتریان (به عنوان مثال، ماه، ۲۰۱۶)^۳، به دست آوردن مشتریان بالقوه و متقاعد کردن مشتریان برای خرید محصولات و خدمات، ارائه اطلاعات حیاتی به خریداران بالقوه، بهینه سازی کمپین‌های تبلیغاتی (به عنوان مثال، یک بات به نام باروج؛ رجوع کنید به (رادو ۲۰۱۶)^۴ و غیره بسیار مفید هستند. مشتریان می‌خواهند تا با همان اپلیکیشن‌هایی که قبلاً کار می‌کردند تجارت کنند. به همین دلیل، بسیاری از بات‌ها در فیسبوک مسنجر، اسنپ‌چت، واتساپ، کیک و وی‌چت ارائه می‌شوند. با استفاده از صدا و پیامک، می‌توان شخصی سازی و همچنین تجربه عالی مشتری را ارائه داد. چتبات‌ها فروشندگان را قادر می‌سازند تا روابط شخصی خود را با مشتریان بهبود دهند. علاوه بر حوزه‌های بازاریابی، چتبات‌های زیادی در زمینه‌هایی مانند خدمات مالی (به عنوان مثال، بانک‌ها) و خدمات مدیریت منابع انسانی و همچنین مدیریت تولید و عملیات برای ارتباطات، همکاری و سایر

¹ Beaver

² Knight

³ Mah

⁴ Radu

مشاغل داخلی و خارجی وجود دارند. به طور کلی، شرکت‌ها از چت‌بات‌ها در پلتفرم‌های پیام‌رسان برای اجرای کمپین‌های بازاریابی (به عنوان مثال، پرونده‌ی آغازین را ببینید) و ارائه تجربه عالی برای مشتری استفاده می‌کنند.

بهبود تجربه مشتری

چت‌بات‌های سازمانی با ارائه یک پلتفرم مکالمه برای تماس سریع و ۲۴ ساعته با شرکت‌ها، موجب بهبود تجربه مشتری می‌شوند. وقتی مشتریان از این سیستم سود ببرند، تمایل بیشتری به خرید و تبلیغ یک برند خاص دارند. چت‌بات‌ها همچنین می‌توانند در ارائه تجربه بهتر مشتری مکمل انسان‌ها باشند.

مثال‌هایی از چت‌بات‌های سازمانی

اشلیشت^۱ (۲۰۱۶) یک راهنمای مبتدی برای چت‌بات‌ها ارائه می‌دهد. او مثال فرضی زیر را در مورد خرید امروز در نورداستروم (یک فروشگاه بزرگ) در مقایسه با بکارگیری چت‌بات‌ها ارائه می‌دهد. اگر می‌خواستید از نورداستروم به صورت آنلاین کفش بخرید، می‌توانید به وب سایت آنها مراجعه کرده، به اطراف نگاه کنید تا کفش مورد نظر خود را پیدا کنید و سپس آنها را خریداری کنید. اگر نورداستروم یک بات بسازد، که مطمئن هستیم که اینکار را انجام خواهند داد، شما به سادگی می‌توانید به نورداستروم در فیسبوک پیام دهید. از شما می‌پرسد که به دنبال چه هستید و شما به سادگی آنچه به دنبالش هستید را می‌گویید. به جای مرور در یک وب سایت، با بات نورداستروم مکالمه می‌کنید که کاملاً همانند تجربه‌ای است که در صورت مراجعه به فروشگاه فیزیکی خواهید داشت.

در ادامه سه مثال دیگر را ارائه می‌کنیم.

لینکدین چت‌بات‌هایی را معرفی کرده است که می‌توانند وظایفی از قبیل مقایسه‌ی تقویم‌های افراد شرکت‌کنندگان در جلسات را مقایسه کرده و زمان و مکان برگزاری جلسه را توصیه کنند. برای جزئیات بیشتر به سی‌بی‌اس نیوز^۲ (۲۰۱۶) نگاه کنید.

مسترکارت دو بات مبتنی بر پلتفرم‌های پیام‌رسان دارد که یکی از بانک‌ها و دیگری برای تجار و بازرگانان است.

مشتریان در سرتاسر جهان می‌توانند از طریق پیام‌رسان فیسبوک با کوکاکولا چت کنند. بات‌ها به مکالماتی که هر روزه شکل شخصی‌تری به خود می‌گیرند حس خوبی را به مشتریان القا می‌کنند. بات‌ها می‌توانند داده‌های مشتریان از قبیل سلاقی، مشکلات، گویش و لهجه محلی و اخلاق و رفتار آنها را گردآوری کرده و سپس تبلیغات سفارشی را متناسب با هر کاربر ارائه کنند. یک ویدئوی ۵ دقیقه‌ای در مورد فیسبوک در آدرس cnbc.com/2016/04/13/why-facebook-is-going-all-in-on-chatbots.html موجود است که جلسات پرسش و پاسخ را با دیوید مارکوس^۳ ارائه می‌کند که وی علاقه فزاینده فیسبوک به چت‌بات‌ها را توضیح می‌دهد.

چرا از خدمات پیام‌رسانی استفاده کنیم؟ تا به اینجا، اشاره کردیم که شرکت‌ها از خدمات پیام‌رسان مانند فیسبوک مسنجر، وی‌چت، کیک، اسکایپ و واتساپ استفاده می‌کنند. دلیل آن این است که در سال ۲۰۱۷، بیش از ۲.۶ میلیارد نفر در سرویس‌های پیام‌رسانی چت می‌کردند. پیام‌رسانی به گسترده‌ترین رفتار دیجیتالی تبدیل شده است. وی‌چت محصول چین اولین شرکتی بود

¹ Schlicht

² CBS News

³ David Marcus

که خدمات خود را با قابلیت‌های «چت با کسب‌وکار» را به صورت تجاری ارائه کرد که در مورد کاربردی ۱۲.۴ نشان داده شده است.

چت‌بات‌های فیسبوک پس از وی‌چت، فیسبوک مکالمات کاربران را با چت‌بات‌های کسب‌وکار در مقیاس وسیع در مسنجر خود راه اندازی کرد و به کاربران پیشنهاد کرد که می‌توانند به یک کسب‌وکار دقیقاً همانطور که به دوستانشان پیام می‌دهند پیام دهند. این سرویس به کسب‌وکارها اجازه می‌دهد تا پیام‌های متنی را با کاربران تبادل کنند. علاوه بر این، بات‌ها توانایی یادگیری دارند که آنها را قادر می‌سازد تا ورودی افراد را به دقت تجزیه و تحلیل کرده و پاسخ‌های صحیح را ارائه دهند. به طور کلی، تا اوایل سال ۲۰۱۸، بیش از ۳۰,۰۰۰ بات شرکتی در فیس‌بوک مسنجر وجود داشت. برخی از شرکت‌ها از بات‌های پیام‌رسان برای تشخیص چهره‌ها در تصاویر استفاده می‌کنند و به گیرندگان تبلیغات هدفمند پیشنهاد می‌دهند. به گفته گوئین^۱ (۲۰۱۶)، فیسبوک به توسعه‌دهندگان نرم افزار اجازه دسترسی به ابزارهای ساخت دستیار شخصی خود به نام «ام»^۲ را می‌دهد که ترکیبی از هوش مصنوعی با مهارت انسانی را برای کارهایی مانند سفارش غذا یا ارسال گل فراهم می‌کند. با استفاده از ابزار «ام»، توسعه‌دهندگان می‌توانند برنامه‌هایی برای مسنجر بسازند که درک بیشتری از درخواست‌های ارائه‌شده به زبان‌های طبیعی داشته باشند. مزیت اصلی این بات‌ها برای فیسبوک گردآوری داده‌ها و ایجاد پروفایل‌های کاربران است. در زیر مثال دیگری از نحوه استفاده از چت‌بات‌ها برای تسهیل خدمات مشتری و بازاریابی ارائه شده است (مورد کاربردی ۱۲.۵).

ابَر چت‌بات وی‌چَت

مورد کاربردی ۱۲.۴

وی‌چت یک سرویس پیام‌رسان جامع و بزرگ در چین و سایر کشورها با حدود ۱ میلیارد عضو در اوایل سال ۲۰۱۸ است. وی‌چت از پیشگامان بازرگاری بات‌ها در سال ۲۰۱۳ بود (mp.weixin.qq.com). کاربران می‌توانند از چت‌بات‌ها برای فعالیت‌های زیر استفاده کنند:

- گرفتن تاکسی
- سفارش غذا برای ارسال
- خرید بلیت سینما و سایر اقلام
- خرید سفارشی یک جفت کفش نایکی
- ارسال سفارش به نزدیک‌ترین شعبه استارباکس
- پیگیری پیشرفت روزانه تمرین‌های ورزشی
- خرید جدیدترین کالکشن بربری
- تعیین وقت ملاقات پزشک
- پرداخت صورتحساب‌ها
- میزبانی یک تماس کنفرانس کاری
- ارسال پیام‌های صوتی، اسنپ‌شات‌ها برای دوستان
- ارسال پیام‌های صوتی برای ارتباط با شرکت‌ها

¹ Guynn

² M

- برقراری ارتباط و درگیرسازی مشتریان
- ارائه‌ی چارچوبی برای کار تیمی و همکاری.
- انجام تحقیقات بازار
- ارائه‌ی اطلاعات و توصیه در خصوص محصولات و خدمات
- راه‌اندازی یک شرکت نوپا در وی‌چت

گریفیث^۱ (۲۰۱۶) اطلاعاتی را در خصوص یک شرکت چینی فروش آنلاین مُد و فشن به نام میسی ارائه کرده است. این شرکت از حساب وی‌چت خود برای گردآوری اطلاعات فروش استفاده کرده است. هر زمان که کاربران جدید حساب میسی را دنبال می‌کردند، یک پیام خوشامد برایشان ارسال می‌شد. وی‌چت به دلیل کاربردی و مفید بودنش به زبان انگلیسی و سایر زبان‌های دنیا موجود است. فیسبوک قابلیت‌های مشابهی را در سال ۲۰۱۵ افزوده است.

سوالاتی برای مورد کاربردی ۱۲.۴

۱. برخی از فعالیت‌های جدیدی که وی‌چت انجام می‌دهد را پیدا کنید.
۲. چه چیزی این چت‌بات را منحصر بفرد کرده است؟
۳. بات وی‌چت را با بات‌های عرضه شده توسط فیسبوک مقایسه کنید.

چگونه ورا گولد مارک از چت‌بات‌ها برای افزایش فروش استفاده می‌کند

مورد کاربردی ۱۲.۵

ورا گولد مارک یک شرکت ساختمانی لوکس در پنجاب هند است.

مسئله

ورا گولد مارک (وی‌جی‌ام)^۲ در یک بازار بسیار رقابتی فعالیت می‌کند. این شرکت به عنوان یک سازنده‌ی آپارتمان‌ای لوکس که معمولاً گرانقیمت هستند باید سعی کند تا خریداران بالقوه‌ی زیادی را جذب کند و به همین خاطر به بیشترین تعداد مشتریان بالقوه با قیمت منطقی نیاز دارد. گپ و گفت زنده با مشتریان می‌تواند بسیار پرهزینه باشد زیرا به عوامل با سطح اطلاعات بالا به صورت ۲۴ ساعته و در ۷ روز هفته نیاز دارد. وی‌جی‌ام واحد آپارتمان‌های زیادی دارد که در زودترین زمان ممکن باید فروخته شوند.

راه‌حل

وی‌جی‌ام تصمیم گرفت تا از چت‌بات‌ها در کنار یا به جای گفتگوی زنده یا حضوری پرهزینه استفاده کند. خریداران می‌توانند بر روی دکمه‌ی « با یک ربات صحبت کنید» در صفحه فیسبوک شرکت کلیک کنند و هرگونه اطلاعاتی که نیاز دارند را دریافت کنند. این روش چت به وی‌جی‌ام کمک می‌کند تا محصولات موجودش را تبلیغ کند. زمانیکه کاربران بر روی دکمه کلیک می‌کنند می‌توانند چت کرده و اطلاعات مربوط به قیمت، زمان تحویل، محل ساخت و ساز و سایر پروژه‌های وی‌جی‌ام را دریافت کنند. کاربران می‌توانند همچنین از توثیقات استفاده کنند. چت‌بات‌ها در مورد پروژه‌های شرکت می‌توانند پاسخگو باشند. فیسبوک امکان دسترسی به پروفایل‌های کاربران را برای وی‌جی‌ام فراهم کرده است (با اجازه‌ی کاربران)

¹ Griffith

² Vera Gold Mark (VGM)

که تیم‌های فروش وی‌جی‌ام می‌توانند از آنها برای تغییر و اصلاح استراتژی‌های فروش استفاده کنند. این سیستم به صورت ۲۴ ساعته و ۷ روز هفته در دسترس است. ارتباط صوتی نیز به زودی به آن افزوده خواهد شد (۲۰۱۸).

نتایج

وی‌جی‌ام در حال حاضر جایگاه مثبت و خوبی پیدا کرده و به عنوان یک شرکت حرفه‌ای شناخته شده است. وی‌جی‌ام نقدهای مثبتی به واسطه‌ی خدمات مشتریانش دریافت کرده است. سازنده از صداقت بیشتری برخوردار است زیرا پاسخ‌ها و وعده‌های مکتوب به مشتریان می‌دهد. کارکنان فروش در وی‌جی‌ام می‌توانند تعداد مشتریان بالقوه‌ی بیشتری را جذب کنند و از آنجائیکه اطلاعات بیشتری در مورد مشتریان آینده دارند بهتر می‌توانند آپارتمان‌های موجود را با نیازهای آنها تطبیق دهند. این سیستم قادر است تا خریداران بین‌المللی را بدون افزایش هزینه جذب کند. با توجه به دسترسی ۲۴/۷ به سیستم، خریداران بین‌المللی براحتی می‌توانند واحدهای موجود وی‌جی‌ام را ارزشیابی کنند. از چت‌بات‌ها همچنین به عنوان یک ابزار آموزشی برای کارکنان جدید استفاده می‌شود. در زمانیکه این مطلب نوشته شد، هیچگونه داده‌های مالی موجود نبود. این فناوری توسط کنیت تکنولوژی^۱ هندوستان **kenyt.com** در اختیار سایر سازندگان نیز قرار گرفته است که چت‌بات‌های هوشمند برای املاک و مستغلات عرضه می‌کنند.

Sources: Based on Garg (2017) and facebook.com/veragoldmark/ (accessed April 2018).

سوالاتی برای مورد کاربردی ۱۲.۵

۱. مزایای وی‌جی‌ام را نام ببرید.
۲. مزایای آن را برای خریداران نام ببرید.
۳. کنیت تکنولوژی چه نقشی دارد؟

مجله‌ی چت‌بات‌ها یک مرور سه بخشی بر کاربرد چت‌بات‌ها در خرده‌فروشی و تجارت الکترونیک دارد. برای جزئیات بیشتر به chatbotsmagazine.com/chatbots-for-retail-and-commerce-part-three-c112a89c0b48 نگاه کنید.

دومین حوزه‌ای که چت‌بات‌های سازمانی در آن فعال هستند بخش خدمات مالی است. ما در اینجا به اختصار در مورد کاربرد آنها در بانک‌داری بحث و گفتگو می‌کنیم. در بخش ۱۲.۶ به معرفی ربات‌های مشاور مالی برای سرمایه‌گذاری می‌پردازیم.

بانک‌داری یک نظرسنجی در سال ۲۰۱۷ (مورگان) نشان داد که اکثر مردم ایالات متحده تا سال ۲۰۱۹ برای بانک‌داری از چت‌بات‌ها استفاده خواهند کرد. چت‌بات‌ها می‌توانند از تحلیل‌های پیشگانه و پیام‌رسانی شناختی برای انجام وظایفی از قبیل پرداخت‌ها استفاده کنند. آنها می‌توانند در مورد خدمات و تسهیلات شخصی‌سازی شده به مشتریان اطلاع‌رسانی کنند. کارت‌های اعتباری بانک‌ها را می‌توان از طریق چت‌بات‌های فیسبوک مسنجر تبلیغ کرد. اینگونه به نظر می‌رسد که مشتریان بیشتر تمایل دارند تا برای انجام کارهایشان با چت‌بات‌ها صحبت کنند تا اینکه با کارکنان فروش بانک‌ها سر و کار داشته باشند.

¹ Kenyt Technologies

در این بخش، ما به بررسی چگونگی استفاده از فناوری‌های نوین در مراقبت‌های بهداشتی می‌پردازیم. این فناوری‌ها شامل دستگاه‌های تشخیصی، ابزارهای جراحی، و سیستم‌های نظارتی می‌باشند. این فناوری‌ها به پزشکان کمک می‌کند تا تشخیص‌های دقیق‌تری داشته باشند و عملیات جراحی را با دقت و ایمنی بیشتری انجام دهند. همچنین، این فناوری‌ها به بیمارستان‌ها کمک می‌کند تا هزینه‌های مراقبت‌های بهداشتی را کاهش دهند و کیفیت خدمات را بهبود بخشند.

در ادامه، ما به بررسی چگونگی استفاده از این فناوری‌ها در مراقبت‌های بهداشتی می‌پردازیم. این فناوری‌ها به پزشکان کمک می‌کند تا تشخیص‌های دقیق‌تری داشته باشند و عملیات جراحی را با دقت و ایمنی بیشتری انجام دهند. همچنین، این فناوری‌ها به بیمارستان‌ها کمک می‌کند تا هزینه‌های مراقبت‌های بهداشتی را کاهش دهند و کیفیت خدمات را بهبود بخشند.

مراقبت سلامت چت‌بات‌ها در حوزه‌ی مراقبت سلامت شدیداً فعال هستند و به میلیون‌ها نفر در سرتاسر جهان کمک می‌کنند (لارسون ۲۰۱۶)؟ مثال‌هایی از این چت‌بات‌ها به شرح زیر هستند:

- **Verbal Access**: یک چت‌بات مبتنی بر هوش مصنوعی که به بیماران کمک می‌کند تا با پزشکان خود در مورد مشکلات سلامتی خود مشورت کنند.
- **North Side Inc.**: یک چت‌بات که به بیماران کمک می‌کند تا با پزشکان خود در مورد مشکلات سلامتی خود مشورت کنند.
- **Hunt**: یک چت‌بات که به بیماران کمک می‌کند تا با پزشکان خود در مورد مشکلات سلامتی خود مشورت کنند.
- **Larson**: یک چت‌بات که به بیماران کمک می‌کند تا با پزشکان خود در مورد مشکلات سلامتی خود مشورت کنند.
- **Zora Robotics**: یک چت‌بات که به بیماران کمک می‌کند تا با پزشکان خود در مورد مشکلات سلامتی خود مشورت کنند.

¹ POSB
² Citi ank
³ Verbal Access
⁴ North Side Inc.
⁵ Hunt
⁶ Larson
⁷ Zora Robotics

آموزش از معلمان چت‌بات در چندین کشور برای آموزش موضوعاتی از زبان انگلیسی (در کشور کره) گرفته تا ریاضیات (در روسیه) استفاده می‌شود. یک چیز قطعی است و آن اینکه چت‌بات‌ها با همه‌ی دانش‌آموزان یکسان رفتار می‌کنند. دانش‌آموزان نیز چت‌بات‌های آموزش آنلاین را دوست دارند. ترجمه‌ی زبان ماشینی دانش‌آموزان را قادر می‌سازد تا در کلاس‌های آنلاین به هر زبانی غیر از زبان مادری خودشان شرکت کنند. نهایتاً، از چت‌بات‌ها می‌توان به عنوان معلم خصوصی استفاده کرد.

دولت طبق لایحه‌ی ۲۰۱۷، چت‌بات‌ها به عنوان یک ابزار گفتگوی جدید برای استفاده‌ی عموم در دولت گسترش یافته‌اند. متداول‌ترین کاربرد آن در ارائه دسترسی به اطلاعات دولتی و پاسخگویی به سوالات مرتبط با دولت است.

سفر و گردشگری چت‌بات‌ها به عنوان راهنمای گردشگری در چندین کشور (مثلاً نروژ) فعالیت دارند. نه تنها هزینه آنها کمتر است (یا حتی رایگان)، بلکه ممکن است نسبت به راهنماهای انسانی اطلاعات بیشتری داشته باشند. چت‌بات‌ها در چندین هتل در ژاپن به عنوان راهنما فعالیت می‌کنند. در هتل‌ها آنها به عنوان دربان عمل کرده و اطلاعات و توصیه‌های شخصی به مشتریان می‌دهند (مثلاً در مورد رستوران). چت‌بات‌ها می‌توانند رزرواسیون اتاق در هتل، وعده‌های غذایی و رویدادها را انجام دهند. در هتل‌های شلوغ، چت‌بات‌ها به طور تمام وقت بر روی تلفن‌های هوشمند در دسترس هستند. همانند سایر خدمات رایانه‌ای، چت‌بات‌ها سریع، ارزان، سهل الوصول و همیشه خوب هستند. آنها تجربه‌ای عالی برای مشتری رقم می‌زنند. یک مثال از خدمات سفرهای خارجی در مورد کاربردی ۱۲.۶ آورده شده است.

چالش‌ها

چت‌بات‌ها در داخل سازمان‌ها تا به اینجا ما با چت‌بات‌هایی آشنا شده‌ایم که در خارج از محیط سازمان‌ها و عمدتاً در حوزه‌های بازاریابی و مراقبت از مشتری کار می‌کنند. با وجود این، شرکت‌ها اخیراً شروع به استفاده از چت‌بات‌ها برای خودکارسازی وظایف پشتیبانی ارتباطات، همکاری‌ها و فرایندهای کسب‌وکار درون سازمانی کرده‌اند. طبق نظر هانت (۲۰۱۷)، «چت‌بات‌های درون سازمانی و سازمان‌ها انقلابی در نحوه فعالیت شرکت‌ها به وجود آورده‌اند.» چت‌بات‌ها در سازمان‌ها می‌توانند وظایف زیاد و فعالیت‌های پشتیبانی از تصمیم‌گیری را انجام دهند. به عنوان مثال، نگاه کنید به نیولندز (۲۰۱۷). چت‌بات‌ها می‌توانند موجب کاهش هزینه، افزایش بهره‌وری، کمک به گروه‌های کاری و پیشبرد روابط با شرکای تجاری شوند. مثال‌های بارز از وظایفی که چت‌بات‌ها می‌توانند انجام دهند به شرح زیر می‌باشند:

- کمک به مدیریت پروژه
- مدیریت ورود داده‌ها
- تنظیم جداول زمانبندی
- روان‌سازی پرداخت‌ها با شرکا
- مشاوره در خصوص اعطای مجوز تأمین مالی
- پایش کار و کارگران
- تحلیل کلان‌داده‌های درون سازمانی
- یافتن محصولات با تخفیف یا ارزان قیمت‌تر

- ساده‌سازی تعاملات
- تسهیل استراتژی‌های مبتنی بر داده
- استفاده از یادگیری ماشینی
- تسهیل و مدیریت امور مالی شخصی

شرکت هواپیمایی ترانسوایا از بات‌ها برای ارتباط و مراقبت از مشتری استفاده می‌کند.

مورد کاربردی ۱۲.۶

تاریخچه

صنعت مسافرت‌های هوایی صنعتی بسیار رقابتی به ویژه در اروپا است. مشتریان جوان بیش از پیش از دستگاه‌های بی‌سیم و سایت‌های رسانه‌های اجتماعی و چت استفاده می‌کنند. مشتریان تمایل دارند تا با استفاده از فناوری‌های مورد علاقه‌شان و از طریق پلتفرم‌های محبوب‌شان با شرکت‌ها ارتباط برقرار کنند. محبوب‌ترین این پلتفرم‌ها فیسبوک مسنجر می‌باشد که بیش از ۱.۲ میلیارد نفر و عمدتاً از طریق تلفن‌های هوشمند در آن چت می‌کنند. این کاربران امروزه نه تنها بین خودشان بلکه با دنیای کسب‌وکار نیز ارتباط برقرار می‌کنند. پلتفرم‌های پیام‌رسان از قبیل مسنجر، واتساپ و وی‌چت برای این گروه از مشتریان به یک عادت و نُرم تبدیل شده‌اند. فروشندگان در حال ساخت اپلیکیشن‌های هوشمند برای پلتفرم‌های پیام‌رسان از جمله بات‌ها هستند.

بات ترانسوایا

ترانسوایا^۱ با یادگیری از سایر شرکت‌ها تصمیم گرفت تا یک بات بر روی فیسبوک مسنجر ایجاد کند. برای اینکار، یک مشاور فناوری اطلاعات واحد کسب‌وکار دیجیتال کائیزانت^۲ به نام میرابین^۳ را استخدام کرد که در حوزه‌ی رابطه‌های مکالمه به ویژه از طریق بات‌ها تخصص داشتند. فعالیت‌ها، فرایندهای کسب‌وکار، بازاریابی و مراقبت از مشتری ترانسوایا با تجربه‌ی تکنولوژیکی میرابین درهم آمیخت و این بات در عرض چند هفته راه‌اندازی شد. اکنون امکان گفتگوی آنی با مشتریان وجود دارد. کاربرد اولیه آن جستجوی پروازهای ترانسوایا است که اطلاعات پرواز و همچنین امکان خرید بلیت را فراهم می‌آورد. این سیستم در حال حاضر با فرایندهای کسب‌وکار ترکیب شده که موجب تسهیل سایر تراکنش‌ها از طریق بات می‌شود. ارائه‌ی یک ابزار دیجیتالی به انتخاب مشتری موجب شد تا ترانسوایا بتواند سهم بازار خود را افزایش داده و رشد و ترقی یابد. توجه داشته باشید که کی‌ال‌ام مالک شرکت ترانسوایا اولین شرکت هواپیمایی اروپایی بود که در سال ۲۰۱۶ چت‌بات مشابهی را بر روی فیسبوک مسنجر پیاده‌سازی کرد.

Sources: Compiled from Cognizant (2017) and transavia.com.

سوالاتی برای مورد کاربردی ۱۲.۶

۱. عوامل محرک مشتریان برای استفاده از دستگاه‌های سیار و چت چیست؟
۲. چرا این بات بر روی فیسبوک مسنجر قرار گرفت؟
۳. مزایای استفاده از کائیزانت چه بود؟

¹ Transavia

² Cognizant Digital Business

³ Mirabeau

با توجه به تعداد بسیار زیاد بات‌ها، تعجب برانگیز نیست که بسیاری از برنامه‌نویسان شروع به ارائه‌ی ابزارها و پلتفرم‌هایی برای کمک به ساخت چت‌بات‌ها کرده‌اند که در بینش فناوری ۱۲.۱ به آنها می‌پردازیم.

بینش فناوری ۱۲.۱ عرضه‌کنندگان پلتفرم‌های چت‌بات

چندین شرکت به عرضه‌ی پلتفرم‌هایی برای ساخت چت‌بات‌های سازمانی می‌پردازند. شرکت‌ها می‌توانند چت‌بات‌ها را به آسانی با استفاده از این ابزارها برای ورود به دنیای پلتفرم‌های پیام رسان یا وبسایت‌هایشان ایجاد کنند. برخی از این ابزارها دارای قابلیت‌های یادگیری ماشینی هستند تا اطمینان حاصل کنند که بات‌ها از طریق تعامل قادر به یادگیری هستند. طبق هانت (۲۰۱۷)، برخی از عرضه‌کنندگان بزرگ عبارتند از:

۱. **چتی‌پیپل**^۱: این ابزار ساخت چت‌بات به ساخت چت‌بات‌هایی که به کمترین مهارت در برنامه‌نویسی نیاز دارند کمک می‌کند.

شرکت‌ها به آسانی می‌توانند صفحات رسانه‌های اجتماعی خود را به حساب چتی‌پیپل وصل کنند. این بات قادر است تا:

- برنامه‌ریزی پرداخت‌ها بر اساس اطلاعات تماس رسانه‌های اجتماعی
- استفاده از ارائه‌دهندگان بزرگ خدمات پرداخت مثل اپل‌پی و پی‌پال.
- تشخیص تغییرات در عبارات کلیدی
- پشتیبانی از پیام‌رسانی

۲. **کوندی**^۲: این بات مالی به مشتریان کمک می‌کند تا پرداخت‌هایشان را مستقیماً از اپلیکیشن‌های پیام‌رسان مخصوصاً مسنجر،

اسکایپ و تلگرام و از طریق یک مرورگر وب به فروشندگان انجام دهند. کاربران با استفاده از این بات می‌توانند:

- پرداخت صورتحساب
- تنظیم یادآوری کننده برای پرداخت صورتحساب
- انتقال پول از طریق ارسال پیام متنی.

این بات امن بوده و از حریم خصوصی کاربران محافظت می‌کند. فروشندگان به آسانی می‌توانند آن را نصب کرده و استفاده کنند.

۳. **توایلا**^۳: این پلتفرم ساخت چت‌بات برای بهبود مراقبت از مشتری و انجام چت‌های زنده ارائه شده است. این پلتفرم

یک پلتفرم پیام‌رسانی برای مشتریانی که ترجیح می‌دهند از چت استفاده کنند عمل می‌کند. هدف اصلی آن تسهیل کردن کارکنان دپارتمان‌های منابع انسانی از وظایف روتین است.

محبوب‌ترین پلتفرم‌ها عبارتند از:

¹ ChattyPeople

² Kundi

³ Twyla

- **آی‌بی‌ام واتسون.** این بسته از یک شبکه عصبی متشکل از ۱ میلیارد کلمه برای درک بسیار خوب زبان طبیعی (مثلاً انگلیسی، ژاپنی) استفاده می‌کند. واتسون ابزارهای توسعه‌ی رایگان از قبیل جاوا اسکریپت^۱، نود اسکریپت^۲، پایتون اسکریپت^۳ و آی‌اواس اسکریپت^۴ را ارائه می‌کند.
- **مایکروسافت بات فریم‌ورک^۵.** مایکروسافت همانند آی‌بی‌ام ابزارهای متنوعی که قابلیت ترجمه به ۳۰ زبان را دارند عرضه می‌کند. این یک پلتفرم متن باز است. این سیستم از سه بخش تشکیل شده، بات کانکتور، پورتال برنامه‌نویس و بات دایرکتوری که توسط سرویس هوشمند درک زبان مایکروسافت^۶ که قصد و نیت کاربر را درک می‌کند به یکدیگر متصل شده‌اند. این سیستم همچنین شامل یک فناوری یادگیری فعال است. یک ابزار ساده‌سازی شده‌ی آن آژور^۷ نام دارد؛ نگاه کنید به بخش ۱۲.۷ و عافق^۸ (۲۰۱۷). برای یک جدول مقایسه‌ای از ۲۵ پلتفرم چت‌بات نگاه کنید به دیویدووا^۹ (۲۰۱۷). برای فهرست سایر پلتفرم‌ها نگاه کنید به اسماعیل^{۱۰} (۲۰۱۷).

Sources: Compiled from Hunt (2017) and Davydova (2017).

سوالاتی برای بحث

۱. تفاوت بین یک بات سازمانی معمولی و یک پلتفرم در چیست؟
۲. در مورد مزایای چتی‌پیپل بحث کنید.
۳. در مورد نیاز به استفاده از کوندی بحث کنید.
۴. در مورد اینکه به چه دلیل مشتریان پلتفرم‌های پیام‌رسان را ترجیح می‌دهند بحث کنید.

برای اطلاعات بیشتر در مورد پلتفرم‌های چت‌بات برای ساخت چت‌بات‌های سازمانی به entrepreneur.com/article/289788 نگاه کنید.

بات‌ها در صنایع خاص همانگونه که دیدیم، بات‌ها می‌توانند متخصص (مثلاً برای مشاوره سرمایه‌گذاری، خدمات مشتری) یا خبره در یک صنعت خاص (مثلاً بانک‌داری، خطوط هوایی) باشند. یک بات جالب و مهم در صنعت ضایعات، آلتو نام دارد (از شرکت بیو‌های تک گلوبال) که کاربران را قادر می‌سازد تا با تجهیزات صنعتی ارتباط هوشمند برقرار کنند. اینکار به مالکان

¹ Java SKD

² Node SKD

³ Python SKD

⁴ iOS SKD

⁵ Microsoft's Bot Framework

⁶ Microsoft Language Understanding Intelligent Service (LUIS)

⁷ AZURE

⁸ Afaq

⁹ Davydova

¹⁰ Ismail

تجهیزات کمک می‌کند تا تصمیماتی را بگیرند که سطح عملکرد را بهبود بخشیده و موجب تسهیل و روان‌سازی روتین‌های نگهداری و تعمیرات و برقراری ارتباطات شود.

دانش مورد نیاز چت‌بات‌ها به وظایف‌شان بستگی دارد. اکثر بات‌های بازاریابی و مراقبت از مشتری نیازمند دانش انحصاری هستند که معمولاً در داخل سازمان تولید و نگهداری می‌شود. این دانش مشابه دانش سیستم‌های خبره است؛ در بسیاری از موارد، چت‌بات‌های سازمانی بسیار شبیه سیستم‌های خبره عمل می‌کنند به استثنای اینکه رابط کاربر در اینجا به زبان طبیعی بوده و اغلب بر اساس صوت کار می‌کند. به عنوان مثال، دانش بات سفورا مختص آن شرکت و محصولات آن شرکت است و در قالب پرسش و پاسخ سازماندهی شده است.

از طرف دیگر، چت‌بات‌هایی که در داخل سازمان مورد استفاده قرار می‌گیرند (مثلاً برای آموزش کارکنان یا ارائه مشاوره در خصوص امنیت و انطباق با قوانین و مقررات دولتی) ممکن است مختص یک سازمان نباشند. یک شرکت می‌تواند این دانش را خریداری کرده و آن را متناسب با موقعیت و نیازهای خاص خودش تغییر دهد. چت‌بات‌های جدیدتر از یادگیری ماشینی برای استخراج دانش از داده‌ها استفاده می‌کنند.

دستیارهای شخصی در سازمان‌ها چت‌بات‌های سازمانی همچنین می‌توانند نقش دستیارهای شخصی مجازی را ایفا کنند که در بخش ۱۲.۵ توضیح داده خواهد شد. به عنوان مثال، این بات‌ها می‌توانند به استعلام‌های کاری پاسخ داده و به افزایش قابلیت‌های تصمیم‌گیری و بهره‌وری کارکنان کمک کنند.

سوالاتی برای مرور بخش ۱۲.۴

۲۵. برخی از بات‌های بازاریابی را توضیح دهید.
۲۶. بات‌ها در خدمات مالی چه کاری می‌توانند انجام دهند؟
۲۷. بات‌ها چگونه می‌توانند به خریداران کمک کنند؟
۲۸. برخی از مزایای چت‌بات‌های سازمانی را نام ببرید.
۲۹. منابع دانش برای چت‌بات‌های سازمانی را توضیح دهید.

در قسمت قبل چت‌بات‌های سازمانی را معرفی کردیم که می‌توان از آنها برای انجام مکالمات و گفتگوها استفاده کرد. در بازاریابی و فروش می‌توانند مدیریت ارتباط با مشتری را تسهیل کنند (مدیریت ارتباط با مشتریان، انجام جستجو برای مشتریان، ارائه اطلاعات و انجام بسیاری از وظایف خاص در سازمان‌ها برای مشتریان و کارکنان خود). برای یک پوشش جامع، از جمله مسائل پژوهشی، نگاه کنید به کاستا و همکاران (۲۰۱۸).

یک نوع در حال ظهور از چت‌بات‌ها به عنوان یک دستیار شخصی مجازی هم برای افراد و هم سازمان‌ها طراحی شده است که به عنوان **دستیار شخصی مجازی**^۱ شناخته می‌شود. این عامل نرم افزاری به افراد در بهبود کارها، تصمیم‌گیری‌ها و تسهیل سبک زندگی‌شان کمک می‌کند. دستیارهای شخصی مجازی اساساً انشعابی از عامل‌های نرم افزاری هوشمند هستند که با مردم تعامل می‌کنند. دستیارهای شخصی مجازی چت‌بات‌هایی هستند که هدف اصلی آنها کمک به افراد در انجام بهتر وظایف خاص است. در حال حاضر میلیون‌ها نفر از سیری با محصولات اپل، دستیار گوگل و الکسا از آمازون استفاده می‌کنند. پایگاه‌های دانشی دستیارها معمولاً جهان‌شمول بوده و به شکل متمرکز در فضای «ابر» نگهداری می‌شوند که موجب شده تا برای تعداد بسیار زیادی از کاربران اقتصادی باشند. کاربران هر زمان که بخواهند می‌توانند کمک و مشاوره از آنها دریافت کنند. در این بخش ما برخی از کاربردهای جالب آنها را توضیح می‌دهیم. مجموعه اول از کاربردها شامل دستیارهای شخصی مجازی به ویژه الکسا آمازون و سیری اپل و دستیار گوگل هستند. اُ برایان^۲ (۲۰۱۶) در مورد کارهایی که چت‌بات‌های دستیار شخصی برای شرکت‌ها می‌توانند انجام دهند بحث کرده است. مجموعه دوم، در مورد برنامه‌های رایانه‌ای است که عمدتاً به عنوان مشاور در موضوعات خاص (اکثراً سرمایه‌گذاری) استفاده می‌شوند.

وظیفه‌ی اصلی دستیارهای شخصی مجازی کمک به کاربران در انجام جستجوی اطلاعات با بکارگیری صوت است. بدون وجود دستیار، کاربران برای یافتن اطلاعات مجبور هستند تا در اینترنت به گشت و گذار بپردازند و در بسیاری از موارد جستجو را رها می‌کنند. در خصوص کسب‌وکارها، کاربران می‌توانند برای کمک و راهنمایی با یک نماینده خدمات مشتریان تماس بگیرند. اینکار می‌تواند برای فروشندگان کار پرهزینه‌ای باشد. تفویض انجام جستجو به یک ماشین می‌تواند صرفه‌جویی پولی قابل توجهی برای فروشندگان به همراه داشته و مشتریان را از اینکه در صف انتظار خدمات نمی‌مانند خشنود کند. به عنوان مثال، لنوو^۳ از دستیار آن‌هولد در سرویس نقطه‌ی تکی جستجو^۴ برای کمک به مشتریان در یافتن پاسخ‌های سؤالاتشان استفاده می‌کند.

زمانیکه سیری و الکسا در حال توسعه بودند زوکربرگ تصمیم گرفت تا دستیار شخصی خودش را برای کمک به انجام کارهای منزل و کارهای اداری‌اش به عنوان مدیرعامل فیسبوک توسعه دهد. دیدگاه او نسبت به این دستیار شخصی همانند شخصیت جارویس در داستان مرد آهنین بود. زوکربرگ این بات را به گونه‌ای آموزش داد تا صدای او را تشخیص دهد و دستورات اولیه مربوط به لوازم خانگی را درک کند. این دستیار می‌تواند چهره مراجعین را تشخیص داده و جابجایی دختر کوچک زوکربرگ را پایش و رصد کند. برای جزئیات بیشتر نگاه کنید به اولانوف^۵ (۲۰۱۶).

بسیاری از ویژگی‌های این دستیار را می‌توانید در ویدئویی ۲:۱۳ دقیقه‌ای در آدرس [youtube.com/watch?v=vvimBPJ3XGQ](https://www.youtube.com/watch?v=vvimBPJ3XGQ) و [youtube.com/watch?v=vPoT2vdVvKc](https://www.youtube.com/watch?v=vPoT2vdVvKc) به همراه روایت‌گری مورگان فریمن تماشا کنید. امروزه دستیارهای مشابهی با قیمت بسیار اندک یا حتی رایگان وجود دارند. معروف‌ترین این دستیارها، الکسا آمازون است.

¹ Virtual Personal Assistant (VPA)

² O'Brien

³ Lenovo

⁴ Single Point of Search

⁵ Ulanoff

بهترین دستیار شخصی مجازی در سال ۲۰۱۸ از بین چندین دستیار، الکسا آمازون بود که توسط شرکت آمازون برای رقابت با سیری اپل طراحی شد و محصول برتری نسبت به آن به شمار می‌آید (شکل ۱۲.۴). الکسا با یک بلندگوی هوشمند مثل اکو آمازون کار می‌کند.



شکل ۱۲.۴ الکسا و اکو آمازون

الکسا آمازون یک دستیار صوتی شخصی مجازی مبتنی بر ابر است که می‌تواند کارهای بسیاری انجام دهد:

- پاسخگویی به سوالات در حوزه‌های مختلف
- کنترل عملیات گوشی‌های هوشمند با دستورات صوتی
- ارائه اطلاعات به روز آب و هوا و ترافیک
- کنترل لوازم خانگی هوشمند و سایر دستگاه‌ها با استفاده از یک هاب خودکارسازی خانگی
- تهیه لیست کارهایی که باید انجام شود
- مرتب‌سازی موزیک‌ها در پلی‌باکس
- کنترل دستگاه‌های خودرو
- خرید برای کاربر
- برقراری تماس‌های تلفنی و پیام‌های متنی

الکسا توانایی تشخیص صداهای مختلف را برای ارائه‌ی پاسخ‌های شخصی‌سازی شده دارد. همچنین، از ترکیبی از گفتار و تماس برای ارائه اخبار، گرفتن تاکسی اوبر و بازی‌ها استفاده می‌کند. با گذشت زمان، قابلیت‌ها و مهارت‌هایش رشد می‌کنند. برای قابلیت‌های روزافزون الکسا به جانسون (۲۰۱۷) نگاه کنید. برای اطلاع از چیزهایی که قادر به شنیدن و به خاطر سپردن آنهاست به اورموس^۱ (۲۰۱۸) نگاه کنید.

در خصوص نحوه کارکرد الکسا ویدیوی ۳:۵۵ دقیقه‌ای [youtube.com/watch?v=jCtfRdqPlbw](https://www.youtube.com/watch?v=jCtfRdqPlbw) را تماشا کنید. برای اطلاعات بیشتر cnet.com/pictures/what-canamazon-echo-and-alexa-do-pictures و مانگالیندان^۲ (۲۰۱۷) و tomsguide.com/us/pictures-story/1012-alexa-tricks-and-easter-eggs.html را ببینید.

مهارت‌های الکسا علاوه بر قابلیت‌های استاندارد (ذاتی) ذکر شده، مردم می‌توانند از اپلیکیشن‌های الکسا برای دانلود قابلیت‌های سفارشی الکسا استفاده کنند. هدف از این مهارت‌ها، آموزش چیزهای جدید به الکسا است. مثال‌هایی از مهارت‌های الکسا در ادامه آورده شده است:

- گرفتن تاکسی اوبر و یافتن هزینه‌ی سفر
- سفارش دادن پیتزا
- سفارش غذاهای بیرون‌بر
- ارائه مشاوره‌ی مالی
- روشن کردن خودروی هیوندای جنسیس از داخل منزل (کوروسیک ۲۰۱۶).

این مهارت‌ها توسط فروشندگان شخص ثالث ارائه می‌شود و باید دستورات فراخوانی را فعال کنند. ده‌ها هزار از این فروشندگان وجود دارند. به عنوان مثال، شخصی می‌تواند بگوید: «الکسا، یک تاکسی برای من بگیر تا سال ۴:۳۰ بعد از ظهر من را از محل کار سوار کند.» برای اطلاعات بیشتر در مورد الکسا آمازون به کلی^۳ (۲۰۱۸) نگاه کنید و برای مزیت‌های آن به رایزینگر^۴ (۲۰۱۶) مراجعه کنید. الکسا مجهز به رابط کاربری آن‌لین پی است که با دستور صوتی می‌تواند فعال شود. اینکار با ترکیب نرم‌افزار الکسا با اسپیکر هوشمند آمازون یعنی اکو انجام شده است.

اسپیکرها و رابط‌های صوتی الکسا آمازون یک خانواده‌ی سه‌تایی از بلندگوها را به نام‌های اکو، دات و تگ دارد که دستگاه‌های ارتباط صوتی هستند. از طریق یک خط تلویزیون فایر و برخی دستگاه‌های غیرآمازونی می‌توان به الکسا دسترسی پیدا کرد. برای ارتباط بین الکسا و اکو به جیکاس^۵ (۲۰۱۶) نگاه کنید.

¹ Oremus

² Mangalindan

³ Kelly

⁴ Reisinger

⁵ Jikas

اِکو آمازون اِکو یک اسپیکر هوشمند بی سیم هندزفری است که با صدا کنترل می شود. در واقع همراه سخت افزار اِکسا است که این دو در کنار یکدیگر کار می کنند. اِکو همیشه روشن و در حال گوش دادن است. زمانیکه اِکو سوال، دستور یا درخواستی را می شنود، صدا را برای اِکسا و از آنجا برای ابر ارسال می کند. سرورهای آمازون پاسخها را با سوالات تطبیق داده و آنها را در کسری از ثانیه برای اِکسا ارسال می کنند. آمازون اِکسا/اِکو در برای از خودروهای فورد موجود است.

آمازون اِکو دات آمازون اِکو دات برادر کوچکتر اِکو است که تمام قابلیت های اِکسا را ارائه می کند ولی فقط یک اسپیکر خیلی کوچک دارد. آن را با هر گونه سیستم اسپیکر می توان وصل کرد.

آمازون اِکو تپ آمازون اِکو تپ برادر کوچکتر اِکو است که می توان از آن در حال Go استفاده کرد. کاملاً بی سیم و قابل حمل بوده و با استفاده از یک شارژر می توان آن را شارژ کرد.

دات و تپ هر دو از اِکو گران تر هستند ولی کیفیت و کارکردهای پایین تری نسبت به اِکو دارند. با وجود این، افرادی که اسپیکرهای خانگی خوبی دارند می توانند از دات استفاده کنند. برای بحث در مورد این سه اسپیکر به [trustedreviews.com/news/amazon-echo-show-vs-echo-2948302](https://www.trustedreviews.com/news/amazon-echo-show-vs-echo-2948302) نگاه کنید.

نکته: اسپیکرهای غیر آمازونی در حال حاضر برای اِکسا موجود هستند (مثلاً یوفی^۱، جینی^۲) که برخی از آنها ارزان قیمت هستند. نکته: اِکسا به اندازه ای هوشمند است که می تواند اعتراف کند که پاسخ سوالی را نمی داند ولی امروزه برای جوابهایی که نمی داند به منابع شخص ثالث رجوع می کند. برای جزئیات و مثالها به uk.finance.yahoo.com/news/alexa-recommendthird-party-skills-192700876.html مراجعه کنید.

اِکسا برای سازمانها اِکسا در ابتدا برای مصارف فردی طراحی شد ولی استفاده از آن در سازمانها افزایش یافته است. وی وورک^۳ پلتفرمی را برای کمک به شرکتها توسعه داده است تا مهارت های اِکسا را در اتاق های جلسه بکار بگیرند. برای جزئیات نگاه کنید به کروک^۴ (۲۰۱۷) و [yahoo.com/news/destiny-2-alexa-skills-let-140946575.html](https://www.yahoo.com/news/destiny-2-alexa-skills-let-140946575.html).

سیری (مخفف رابط تشخیص و تفسیر گفتار) یک دستیار شخصی مجازی هوشمند است که بخشی از چنیدن سیستم عامل اپل می باشد. سیری می تواند به سوالات پاسخ داده، توصیه کرده و با تفویض درخواستها به مجموعه ای از خدمات وب در فضای ابر اقداماتی را انجام دهد. این نرم افزار می تواند خودش را با زبان کاربر تطبیق داده، با توجه به ترجیحات به جستجو پرداخته و نتایج شخصی سازی شده را باز گرداند. سیری برای آیفون و کاربران آی پد به صورت رایگان در دسترس است. سیری قابلیت یکپارچه سازی در ریموت سیری^۵ اپل را دارد. سیری با استفاده از کارپلی در برخی از خودروها بکار گرفته شده که توسط آیفون (۵ و بالاتر) کنترل می شود. سیری ۲ مدل سال ۲۰۱۸-۲۰۱۷ آن است.

¹ Eufy

² Genie

³ WeWork Corp

⁴ Crook

⁵ Siri Remote

ویو^۱ در سال ۲۰۱۶، داگ کیتلاوس^۲ خالق سیری نرم‌افزار ویو^۱ را معرفی کرد که یک رابط هوشمند برای همه چیز است. ویو^۱ نسل بعدی تعاملات مجازی هوشمند محسوب می‌شود (نگاه کنید به ماتنی^۳ ۲۰۱۶). برخلاف سایر دستیارها، منبع ویو^۱ در اختیار تمام برنامه‌نویسان قرار دارد (محصولات اکوسیستم شخص ثالث). ویو^۱ در حال حاضر جزو شرکت سامسونگ است. در سال ۲۰۱۷، سامسونگ دستیار شخصی خودش را برای گلکسی اس ۸ راه‌اندازی کرد.

رقابت در حوزه‌های دستیارهای شخصی مجازی با بهبود قابلیت‌های دستیار گوگل که به عنوان رقیب سیری برای گوشی‌های هوشمند اندرویدی توسعه یافت در حال افزایش است. یک نمایش جالب از آن را می‌توانید در [youtube.com/watch?v=WTMbf0qYWW](https://www.youtube.com/watch?v=WTMbf0qYWW) مشاهده کنید. برخی از قابلیت‌های پیشرفته‌ی آن در ویدئویی در [youtube.com/watch?v=17rY2ogJQQ](https://www.youtube.com/watch?v=17rY2ogJQQ) موجود است. برای جزئیات به کلی (۲۰۱۶) نگاه کنید.

چندین شرکت دیگر نیز دستیارهای شخصی مجازی تولید می‌کنند. به عنوان مثال، مایکروسافت کورتانا^۴ بسیار معروف است. در سپتامبر ۲۰۱۶، مایکروسافت دو برنامه‌ی کورتانا و بینگ^۵ را با یکدیگر ادغام کرد (هچمن ۲۰۱۶).^۶ الکسا و کورتانا اکنون با یکدیگر همکاری می‌کنند. توجه داشته باشید که برآورد شده است که تا سال ۲۰۲۲ دستیارهای شخصی صوتی ۵۵ درصد از کل خانوارهای ایالات متحده را پوشش خواهند داد. برای این موضوع و آینده‌ی دستیارهای شخصی به پترز^۷ (۲۰۱۷) مراجعه کنید.

اپل و گوگل دستیارهای شخصی خودشان را برای صدها میلیون کاربر دستگاه‌های موبایل عرضه کرده‌اند. مایکروسافت بیش از ۲۵۰ میلیون رایانه شخصی را به دستیار شخصی خود مجهز کرده است. آمازون الکسا^۸ کو تعداد دستیارهای بسیار بیشتری را فروخته است. رقابت بر سر چت‌بات‌های صوتی است.

همانطور که قبلاً نیز گفته شد، دانش برای دستیارهای شخصی در فضای «ابر» نگهداری می‌شود. دلیل اینکار این است که دستیارهای کالاهایی هستند که در دسترس میلیون‌ها کاربر قرار دارند و باید اطلاعات پویا و به‌روز ارائه کنند (مثلاً وضعیت آب و هوا، اخبار، قیمت سهام). زمانیکه پایگاه دانش به صورت متمرکز باشد، نگهداری آن در یک مکان انجام می‌شود. این برعکس دانش در بسیاری از بات‌های سازمانی است که به‌روزآوری آنها به صورت غیرمتمرکز انجام می‌شود. بنابراین، دانش سیری بر روی یک آیفون همیشه توسط ای‌ای‌پی‌ال^۸

¹ Viv

² Dog Kittlaus

³ Matney

⁴ Microsoft Cortana

⁵ Bing

⁶ Hachman

⁷ Perez

⁸ AAPL

به‌روزآوری می‌شود. دانش برای مهارت‌های الکسا باید به صورت محلی یا توسط فروشندگان شخص ثالثی که آنها را تولید می‌کنند نگهداری می‌شود.

سوالاتی برای مرور بخش ۱۲.۵

۷. یک دستیار شخصی مجازی هوشمند را توضیح دهید.
۸. ارتباط بین الکسا آمازون و اِکو چیست؟
۹. قابلیت‌های الکسا آمازون را توضیح دهید.
۱۰. اِکو دات و تپ را توضیح دهید.
۱۱. سیری اپل و دستیار گوگل را توضیح دهید.
۱۲. دانش دستیارهای شخصی چگونه نگهداری می‌شود؟
۱۳. ارتباط بین دستیارهای شخصی مجازی و چت‌بات‌ها را توضیح دهید.

می‌دانیم که اکثریت تصمیمات خرید و فروش سهام در بورس به ویژه مؤسسات مالی توسط رایانه‌ها اتخاذ می‌شود. با وجود این، رایانه‌ها می‌توانند حساب‌های شخصی را نیز به شکلی شخصی‌سازی شده مدیریت کنند. طبق نظرسنجی ای. تی. کرنی، مشاوران روبا^۱ به عنوان ارائه‌دهندگان آنلاین شناخته می‌شوند که خدمات مشاوره سرمایه‌گذاری خودکار، کم هزینه و شخصی‌سازی شده را معمولاً از طریق پلتفرم‌های سیار ارائه می‌کنند. این مشاوران روبا از الگوریتم‌هایی استفاده می‌کنند که محصولات سرمایه‌گذاری را تخصیص، گسترش، متوازن سازی مجدد و مبادله می‌کنند. کاربران زمانیکه برای خدمات روبا نام‌نویسی می‌کنند وارد بخش اهداف سرمایه‌گذاری^۲ و اولویت‌های آنها می‌شوند. سپس روبا با استفاده از الگوریتم‌های هوش مصنوعی پیشرفته سرمایه‌گذاری‌های آلترناتیو شخصی‌سازی شده را برای انتخاب از بین صندوق‌ها یا صندوق‌های قابل معامله پیشنهاد می‌دهند. با گفتگو با یک مشاور روبا، یک برنامه‌های هوش مصنوعی به پالایش پورتفولیوی سرمایه‌گذاری می‌پردازد. تمامی اینها به صورت دیجیتالی و بدون نیاز به صحبت با انسان انجام می‌شود. برای جزئیات نگاه کنید به کیپل (۲۰۱۶).

¹ Robo Advisors

² Investment Objectives

شرکت‌های بسیاری پس از ظهور شرکت پیشگام بترمنت^۱ در سال ۲۰۱۰ شکل گرفتند (فیوچر ادوایزر آند هجیل^۲ در ۲۰۱۰ و پرسونال کپیتال^۳، ولتفرانت آند سیگ‌فیگ^۴ در ۲۰۱۱ و ۲۰۱۲). شرکت‌های معروف دیگری (شاب اینتلیجنت پورتفولیو^۵، آکورنز^۶، ونگارد راس^۷ و آلی^۸) در سال‌های ۲۰۱۴ و ۲۰۱۵ به خیل ارائه دهندگان مشاوران رобо پیوستند. در ۲۰۱۶ و ۲۰۱۷، موسسه‌های کارگزاری ایبی*ترید^۹ و تی‌دی آم‌ری‌ترید^{۱۰} به همراه فیدلیتی^{۱۱} و مریل اج^{۱۲} به این جمع پیوستند. شکی نیست که مشاوران رобо معادلات بازی کسب‌وکار مدیریت سرمایه را برهم زده‌اند اگرچه عملکرد آنها تا به اینجا چندان تفاوتی با خدمات مالی و روش‌های قدیمی نداشته است.

صندوق‌های قابل معامله که حق کمیسیون آنها بسیار کمتر از صندوق‌های مشترک است هزینه‌ها را کاهش دهند. حق کمیسیون سالانه با توجه به حداقل دارایی مورد نیاز متغیر است. خدمات ویژه بسیار گرانتر هستند زیرا فرصت مشاوره با کارشناسان انسانی را فراهم می‌کنند.

با رشد و بلوغ مشاوران، مشخص شد که برخی اوقات در انجام کارشان به تنهایی اثربخش نیستند. از این رو، در اواخر ۲۰۱۶ چندین مشاوران کاملاً خودکار شروع به افزودن حواس انسانی می‌نامیم کردند (مثلاً نگاه کنید به یول^{۱۳} ۲۰۱۷ و هوانگ^{۱۴} ۲۰۱۷). شرکت‌ها در حال افزودن گزینه‌های انسانی یا شراکت با یک شرکت دیگر هستند. به عنوان مثال، یوبی‌اس ولث منیجمنت آمریکا^{۱۵} با یک مشاور رобо محض به نام سیگ‌فیگ شراکت کرده است. مشاوران رобо که از حواس انسانی برخوردارند از لحاظ تخصص بسیار متنوع هستند. به عنوان مثال، بترمنت، شاب اینتلیجنت ادوایزری و ونگارد پرسونال ادوایزر سرویس از برنامه‌ریزان مالی معتبر استفاده می‌کنند، سایر شرکت‌ها از تخصص کمتری برخوردارند. برای جزئیات بیشتر به هوانگ (۲۰۱۷) مراجعه کنید. مورد کاربردی ۱۲.۷ توضیح می‌دهد که بترمنت چگونه حواس انسانی را به کار گرفته است.

¹ Betterment Inc.

² Future Advisor & Hedgeable

³ Personal Capital

⁴ Wealthfront and SigFig

⁵ Schwab Intelligent Portfolios

⁶ Acorns

⁷ Vanguard RAS

⁸ Alley

⁹ E*Trade

¹⁰ TD Ameritrade

¹¹ Fidelity

¹² Merrill Edge

¹³ Eule

¹⁴ Huang

¹⁵ UBS Wealth Management Americas

کیفیت مشاوره‌های ارائه شده توسط مشاوران رобо ممکن است تعجب کنید که مشاوره‌هایی که مشاوران رобо ارائه می‌کنند تا چه اندازه خوب است. پاسخ این است که این به دانش آنها، نوع سرمایه‌گذاری، موتور جستجوی ماشین هوش مصنوعی و مسائل دیگری بستگی دارد. با وجود این، به خاطر داشته باشید که این مشاوره‌ها بی‌طرفانه و سازگار هستند. حتی ممکن است در یکی از مهمترین جنبه‌های مشاوره‌ی سرمایه‌گذاری بهتر از انسان‌ها عمل کنند و بدانند که چگونه مالیات را به شکل قانونی به حداقل برسانند. این بیانگر این مطلب است که تمامی سرمایه‌گذاران از فرصت مالیات زیان در سطح سازمانی برخوردارند. برعکس، برخی معتقدند که جایگزین کردن کارگزاران سرمایه‌گذاری با ربوها کار دشواری است. دی آئیل^۱ (۲۰۱۸) معتقد است که انسان‌ها همچنان در ارائه خدمات مشاوره سلطه دارند (مثال نوردیا بانک^۲ توسط پوجانپالو^۳ ۲۰۱۷). برای فهرستی از بهترین مشاوران ربو نگاه کنید به یول (۲۰۱۷)، اُشیا^۴ (۲۰۱۶) و investorjunkie.com/35919/roboadvisors برای پوشش جامع مشاوران ربو در امور مالی و سرمایه‌گذاری از جمله شرکت‌های مهم در صنعت مشاوره به مک‌کلیان^۵ (۲۰۱۶) نگاه کنید. یک مشاور ربویی تجاری نوظهور در دانشگاه کورنل^۶ تحت نام جی‌اس‌فیر^۷ در حال توسعه است. علاوه بر این، مشاوران ربو در کشورهای دیگری غیر از ایالات متحده نیز وجود دارند (مثلاً مارول استون کپیتال^۸ در سنگاپور).

مؤسسات مالس و رقابت آنها چندین مؤسسه بزرگ مالی و بانک با ایجاد مشاوران ربو خودشان یا شراکت با آنها به این پدیده واکنش نشان داده‌اند. ارزیابی برندگان و بازندگان این رقابت کار دشواری است زیرا داده‌های کافی و بلندمدتی از آنها وجود ندارد. تا به اینجا به نظر می‌رسد مشاوران ربو مورد توجه مشتریان قرار گرفته‌اند که اصولاً به خاطر این است که هزینه آنها ۱۰ درصد هزینه‌ی مشاوران انسانی می‌باشد. برای بحث و داده‌ها به مارینو (۲۰۱۶) نگاه کنید. توجه داشته باشید که برخی به خطر استفاده از مشاوران ربو در کاهش بازار بورس به دلیل استفاده آنها از صندوق‌های قابل معامله اشاره کرده‌اند.

بسیاری از مؤسسات و برخی سرمایه‌گذاران حقیقی سهام‌هایشان را با استفاده از الگوریتم‌های هوش مصنوعی خریداری می‌کنند. برخی ترجیح می‌دهند تا صندوق‌های مشترکی را که هوش مصنوعی دارایی آنها را انتخاب می‌کند خریداری کنند. ای‌کی‌وبات یک چنین صندوقی است. عملکرد سال ۲۰۱۷ آن بالاتر از حد متوسط بوده است. الگوریتم‌های هوش مصنوعی که ای‌کی‌وبات از آنها استفاده می‌کند می‌توانند ۱ میلیون داده را روزانه پردازش کنند. برای جزئیات به ال^۹ (۲۰۱۸) نگاه کنید.

¹ De Aenlle

² Nordea Bank

³

⁴ O'Shea

⁵ McClellan

⁶ Cornell University

⁷ Gsphere

⁸ Marvelstone Capital

⁹ Eli

علاوه بر مشاوران سرمایه‌گذار، چندین نوع مشاور روبرو در حوزه‌های سفر گرفته تا پزشکی و قانونی فعالیت دارند. در ادامه مثال‌هایی از این مشاوران غیرمالی آورده شده است:

- [Utrip](#) یک مشاور سفر است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای سفر خود پیدا کنید. این مشاور می‌تواند به شما در انتخاب بهترین هتل، پرواز و رستوران‌ها کمک کند. همچنین می‌تواند به شما در برنامه‌ریزی سفر خود کمک کند.
 - [Utrip.com](#) یک مشاور سفر است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای سفر خود پیدا کنید. این مشاور می‌تواند به شما در انتخاب بهترین هتل، پرواز و رستوران‌ها کمک کند. همچنین می‌تواند به شما در برنامه‌ریزی سفر خود کمک کند.
 - [Utrip.com](#) یک مشاور سفر است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای سفر خود پیدا کنید. این مشاور می‌تواند به شما در انتخاب بهترین هتل، پرواز و رستوران‌ها کمک کند. همچنین می‌تواند به شما در برنامه‌ریزی سفر خود کمک کند.
- علاوه بر این، مشاوران غیرمالی دیگری نیز وجود دارند که در زمینه‌های مختلف فعالیت می‌کنند. در ادامه به برخی از این مشاوران اشاره می‌کنیم:
- [noHold AI](#) یک مشاور است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای خرید بلیت هواپیما پیدا کنید.
 - [Utrip](#) یک مشاور سفر است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای سفر خود پیدا کنید.
 - [Ad a Health](#) یک مشاور است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای خرید بیمه‌های درمانی پیدا کنید.
 - [Health Tap](#) یک مشاور است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای خرید بیمه‌های درمانی پیدا کنید.
 - [YourMD](#) یک مشاور است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای خرید بیمه‌های درمانی پیدا کنید.
 - [Florence](#) یک مشاور است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای خرید بیمه‌های درمانی پیدا کنید.
 - [OneStopHealth](#) یک مشاور است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای خرید بیمه‌های درمانی پیدا کنید.
 - [HealthBot](#) یک مشاور است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای خرید بیمه‌های درمانی پیدا کنید.
 - [GYANT](#) یک مشاور است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای خرید بیمه‌های درمانی پیدا کنید.
 - [Buoy](#) یک مشاور است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای خرید بیمه‌های درمانی پیدا کنید.
 - [Mewhat](#) یک مشاور است که به شما کمک می‌کند تا بهترین گزینه‌ها را برای خرید بیمه‌های درمانی پیدا کنید.

¹ noHold AI

² Utrip

³ Ad a Health

⁴ Health Tap

⁵ YourMD

⁶ Florence

⁷ OneStopHealth

⁸ HealthBot

⁹ GYANT

¹⁰ Buoy

¹¹ Mewhat

•

• shopadvisor.com/our-platform

۱.

۲.

۳.

بات‌های خرید برای درک نیازها و ترجیحات مشتری سوالاتی اندکی می‌پرسند. سپس بهترین مورد تطبیق یافته را به مشتریان توصیه می‌کنند. بدین ترتیب مشتری احساس می‌کند که خدمات شخصی‌سازی شده‌ای را دریافت کرده است. این دستیارها فرایند تصمیم‌گیری را برای مشتری سهل و آسان می‌کنند. دستیارهای هوشمند همچنین در خصوص مسائل مورد توجه مشتریان از طریق مکالمات پرسش و پاسخ مشاوره می‌دهند. برای یک آزمون هدایت شده به دموی smartassistant.com/advicebots نگاه کنید. توجه داشته باشید که این بات‌ها اساساً سیستم‌های توصیه‌گر هستند و

¹ Enduarncce

² Lovett

³ ELIZA

کاربران باید از آنها مشورت بخواهند در حالیکه سایر سیستم‌های توصیه‌گر (مثلاً آمازون.کام) حتی در صورتیکه مشتری درخواست نکرده باشد به ارائه مشاوره می‌پردازند.

یک دستیار خرید بسیار معروف در حوزه‌ی مُد، علی‌بابا فِشِنِ ای‌آی^۱ است که به مشتریانی که در فروشگاه‌ها خرید می‌کنند کمک می‌کند. زمانیکه خریدار وارد اتاق پرو می‌شود، مشاور فِشِنِ ای‌آی وارد عمل می‌شود. برای جزئیات بیشتر در این خصوص به سان (۲۰۱۷) نگاه کنید.

نوع دیگری از مشاورهای خرید به عنوان مشاور شخصی مجازی برای خریداران عمل می‌کنند. این نوع برای عوامل هوشمند تجارت الکترونیک توسعه یافت مثل bizrate.com و pricegrabber.com.

شاید بتوان آی‌بی‌ام واتسون را با معلومات‌ترین مشاور مجازی دانست. مثال‌هایی از کاربردهای آن به شرح زیر هستند:

-
-
-
-

سوالاتی برای مرور بخش ۱۲.۶

۱۳. مشاور روبو را توضیح دهید.

۱۴. توضیح دهید مشاوران روبو چگونه برای سرمایه‌گذاری کار می‌کنند؟

۱۵. در مورد برخی از نواقص مشاوران روبو در سرمایه‌گذاری بحث کنید.

¹ Alibaba Fashion AI

² Macy On Call

³ Watson Medical Advisor

⁴ Deep Thunder

⁵ Connie Robot

⁶ Noyes

- ۱۶. همکاری بین انسان و ماشین را در مشاوران روبو توضیح دهید.
- ۱۷. نقش آی‌بی‌ام واتسون به عنوان یک مشاور را توضیح دهید.

چندین مسئله‌ی مرتبط با پیاده‌سازی و اجرا مختص چت‌بات‌ها و دستیارهای شخصی وجود دارد. مثال‌هایی از سیستم‌های مهم در ادامه توضیح داده شده است.

بسیاری از چت‌بات‌ها از جمله دستیارهای شخصی مجازی در تشخیص صدا ضعیف عمل می‌کنند. هیچ سیستم بازخورد مناسبی هنوز برای سیستم‌های تشخیص صدای کاربران و اینکه تا چه اندازه قادر به تشخیص کاربران هستند وجود ندارد. علاوه بر این، سیستم‌های تشخیص صدا ممکن است ندانند که چه زمان باید کاری را انجام دهند و نیاز به مداخله‌ی انسان باشد. چت‌بات‌هایی که در داخل سازمان‌ها استفاده می‌شوند باید به یک سیستم ان‌ال‌پی متصل باشند. این می‌تواند مشکل‌ساز باشد ولی مشکل بزرگتر زمانی بوجود می‌آید که چت‌بات‌ها به اینترنت متصل شوند که با مسائل امنیتی و مشکلات اتصال مواجه می‌شویم. برخی چت‌بات‌ها باید چندزبانه باشند که به همین دلیل باید به یک مترجم زبان ماشینی متصل شوند.

در ادامه نکاتی در رابطه با معایب و محدودیت‌های بات‌ها (که در زمان نگارش این کتاب در سال‌های ۲۰۱۷ و ۲۰۱۸ مشاهده شده‌اند) ارائه شده که برخی از آنها در طول زمان رفع خواهند شد:

- برخی بات‌ها حداقل در ابتدای کار خود عملکرد ضعیفی دارند که موجب دلسردی کاربران می‌شوند.
- برخی بات‌ها نتوانسته‌اند برند خود را به خوبی مطرح کنند. طراحی ضعیف ممکن است منجر به تصویر ذهنی ضعیف از برند شود.
- کیفیت بات‌های مبتنی بر هوش مصنوعی به استفاده از الگوریتم‌های پیچیده بستگی دارد که ساخت و بکارگیری آنها پرهزینه است.
- استفاده از برخی بات‌ها راحت و آسان نیست.
- برخی بات‌ها به شکلی نامنسجم کار می‌کنند.
- چت‌بات‌های سازمانی چالش‌های امنیتی و یکپارچگی زیادی را بوجود می‌آورند.

برای روش‌های حذف برخی از این معایب و محدودیت‌ها به کایا ۲۰۱۷ نگاه کنید.

دستیارهای مجازی در معرض حمله کورتانا، سیری، الکس و دستیار گوگل از طرف کسانی که به طور کلی با ماشین‌ها موافق نیستند یا اینکه فقط دوست دارند آنها را مورد تمسخر قرار دهند مورد حمله قرار گرفته‌اند. در برخی موارد، مدیران این بات‌ها سعی دارند تا پاسخی در قبال این حملات بدهند در سایر موارد برخی ماشین‌ها پاسخ‌های بی‌معنایی به این حملات بی‌معنا می‌دهند.

اگرچه کیفیت اکثر سیستم‌ها کامل نیست ولی در طول زمان بهبود می‌یابد. با وجود این، کیفیت بات‌هایی که اطلاعات کاربران را گرفته و به خوبی برنامه‌ریزی شده‌اند می‌تواند بسیار بالا باشد. به طور کلی، هرچه شرکت سرمایه‌گذاری بیشتری برای تملک یا اجاره‌ی یک چت‌بات انجام دهد، دقت آن بیشتر خواهد بود. علاوه بر این، بات‌هایی مثل الکسا و دستیار گوگل که به تعداد زیادی از کاربران خدمت‌رسانی می‌کنند از دقت بالاتری برخوردارند.

کیفیت مشاوران روبو با توجه به زمان کوتاهی که از پیدایش مشاوران روبو برای خدمات مالی می‌گذرد، ارزیابی کیفیت مشاوره‌های آنها کار دشواری است. بک‌اند بنچمارکینگ^۱ (theroboreport.com)، گزارشات فصلی را در خصوص شرکت‌های مشاوران روبو منتشر می‌کند. برخی از این گزارشات رایگان هستند. طبق گزارشات این سرویس، ربات شاب اینتلیجنت پورتفولیو بهترین بات در سال ۲۰۱۷ بوده است. با وجود این، توجه داشته باشید که عملکرد پورتفولیو باید در یک بازه زمانی طولانی ارزیابی شود (مثلاً بین ۵ تا ۱۰ سال). یک مسئله اساسی در خصوص بات‌ها، از دست رفتن حواس انسانی است. برای اینکه مشتریان بتوانند پاسخ‌های بات‌ها را درک کنند به اعتماد و پاسخگویی به سوالات پیچیده نیاز است. همچنین بات‌ها نمی‌توانند حس همدلی یا دوستی را انتقال دهند. طبق نایت (۲۰۱۷)، این مشکل راه‌حل دارد. اول، بات‌ها باید فقط وظایفی که برای آنها مناسب هستند را انجام دهند. دوم، آنها باید مزیت مشهود و ملموسی برای مشتری داشته باشند. نهایتاً، از آنجائیکه بات‌ها با مشتریان در تماس مستقیم هستند تعاملات باید کاملاً برنامه‌ریزی شده باشد تا مشتری خشنود و راضی باشد. علاوه بر این، توجه داشته باشید که مشاوران روبو مشاوره‌های شخصی‌سازی شده ارائه می‌دهند. برای اطلاعات در مورد اینکه کدام روبو می‌تواند پاسخگوی نیازهای شما باشد به یول (۲۰۱۷) نگاه کنید که یک کارت امتیازی برای شرکت‌های پیشگام در این حوزه ارائه کرده است. نهایتاً، جیلانی^۲ (۲۰۱۶) یک راهنما برای مشاوران روبو و همچنین خطرات احتمالی آنها ارائه کرده است.

مایکروسافت تی تی^۳ یکی از چت‌بات مبتنی بر توئیتر ناموفق از شرکت مایکروسافت بود که فعالیت آن متوقف شد. این بات اطلاعات را از اینترنت گردآوری می‌کرد ولی مایکروسافت دانش کافی و مناسب را برای مواجهه با اطلاعات نامناسب در اینترنت در نظر نگرفته بود (مثلاً ترول‌ها، اخبار جعلی). از این رو، خروجی تی تی بی‌فایده بوده و کاربران را دلخور می‌کرد. در نتیجه، مایکروسافت این سرویس را متوقف کرد.

الکسا ابزار مفیدی برای کنترل منازل هوشمند است. کریست^۴ (۲۰۱۷) یک فرایند شش مرحله‌ای برای استفاده از الکسا در منازل هوشمند پیشنهاد داده است:

۱. یک اسپیکر تهیه کنید (مثلاً اکو).
۲. مکانی برای آن اسپیکر در نظر بگیرید.

¹ Backend Benchmarking

² Gilani

³ Microsoft Tay

⁴ Crist

۳. دستگاه‌های خانگی هوشمند را راه‌اندازی کنید.

۴. برنامه‌های مربوطه را با الکسا تنظیم کنید.

۵. گروه و صحنه را راه‌اندازی کنید.

۶. فرایند را در حین عملیات تنظیم کنید.

این گام‌ها در cnet.com/uk/how-to-get-started-withan-alexa-smart-home/ نشان داده شده‌اند.

چگونه می‌توانیم

قبل تر ما به معرفی برخی از شرکت‌ها که به ارائه پلتفرم‌های چت‌بات‌ها می‌پردازند پرداختیم. علاوه بر این، چندین شرکت می‌توانند بات‌ها را برای کاربران تولید کنند یا اینکه یک بات ساده را خودشان تولید کنند. یک راهنمای گام به گام به همراه ابزارهای بکار رفته توسط ایگنات (۲۰۱۷) معرفی شده است. این بات بر روی فیسبوک مسنجر ساخته شد. یک راهنمای دیگر برای ساخت بات بر روی فیسبوک مسنجر توسط نیولندز (۲۰۱۷) ارائه شده که گام‌های زیر را توصیه می‌کند:

۱. یک نام منحصر بفرد برای بات انتخاب کنید.

۲. یک راهنما در خصوص نحوه ساخت بات و برقراری ارتباط با آن به مشتریان بدهید.

۳. برای اینکه یک جریان مکالمه‌ی طبیعی داشته باشید آزمایشاتی را انجام دهید.

۴. بات را هوشمند بسازید ولی از اصطلاحات و واژگان ساده استفاده کنید.

۵. تمامی ویژگی‌های مد نظرتان را به یکباره پیاده‌سازی نکنید.

۶. بات را برای بهبود مستمر عملکردش بهینه‌سازی و نگهداری کنید.

چندین منبع رایگان برای ساخت چت‌بات‌ها وجود دارد. اکثر آنها شامل دستورالعمل‌های «How-to» هستند. چندین سرویس پیام‌رسان (مثلاً فیسبوک مسنجر، تلگراف) هم پلتفرم‌های چت‌بات و هم چت‌بات‌های خاص خودشان را ارائه می‌کنند. برای فهرست سال ۲۰۱۷ پلتفرم‌های چت‌بات‌های سازمانی و قابلیت‌هایشان به entrepreneur.com/article/296504 نگاه کنید.

استفاده از سرویس بات مایکروسافت آژور آژور یک سازنده‌ی بات جامع ولی بسیار پیچیده است. سرویس بات آن پنج الگو برای

ساخت سریع و آسان بات‌ها ارائه کرده است. طبق [docs.microsoft.com/en-us/bot-framework/azure-bot-](http://docs.microsoft.com/en-us/bot-framework/azure-bot-service/overview/)

[serviceoverview/](http://docs.microsoft.com/en-us/bot-framework/azure-bot-service/overview/) هر کدام از الگوها که در جدول ۱۲.۱ نشان داده شده‌اند را می‌توان بکار برد.

جدول ۱۲.۱ الگوهای آژور	
الگو	شرح
پایه	ساخت یک بات که از روش گفتگو برای پاسخگویی به ورودی کاربر استفاده می‌کند.
فرم	ساخت یک بات که ورودی کاربران که از طریق مکالمات هدایت شده که با استفاده از فرم فلو ایجاد می‌شوند را گردآوری می‌کند.

درک زبان	ساخت یک بات که از مدل‌های زبان طبیعی برای درک نیت کاربر استفاده می‌کند.
کنشگر	ساخت یک بات که از کارکردهای آژور برای اطلاع‌رسانی در مورد رویدادها به کاربر استفاده می‌کند.
پرسش و پاسخ	ساخت یک بات که از یک پایگاه دانش برای پاسخگویی به سوالات کاربران استفاده می‌کند.

نکته: مایکروسافت همچنین یک چارچوب برای ساخت بات‌ها ارائه کرده است (مشابه آنچه بر روی فیسبوک مسنجر انجام می‌شود). برای اطلاعات بیشتر در مورد مایکروسافت بات و یک دستورالعمل آموزشی به عافاق (۲۰۱۷) نگاه کنید.

- چت‌بات‌ها می‌توانند صرفه‌جویی اقتصادی برای سازمان‌ها داشته باشند، یک ارتباط ۲۴ ساعته و ۷ روز هفته با مشتریان و یا شرکای تجاری برقرار کرده و گفته‌هایشان منسجم است.
- یک سیستم خبره اولین محصول هوش مصنوعی با کاربرد تجاری بود.
- انتقال دانش سیستم خبره از کارشناسان به ماشین‌ها موجب شد تا ماشین‌ها از تخصص مورد نیاز برای حل مسئله برخوردار شوند.
- سیستم‌های خبره‌ی کلاسیک از قواعد کسب‌وکار برای نمایش دانش و تولید پاسخ به سوالات کاربران استفاده می‌کنند.
- مؤلفه‌های اصلی سیستم‌های خبره شامل اکتساب دانش، نمایش دانش، پایگاه دانش، رابط کاربر و موتور رابط هستند.
- مؤلفه‌های بیشتر می‌توانند شامل یک زیرسیستم توضیح و یک سیستم پالایش دانش باشند.
- سیستم خبره به حفظ دانش کمیاب در سازمان کمک می‌کند.
- انواع جدید سیستم‌های دانشی نسبت به سیستم‌های خبره کلاسیک برتری بالایی دارند.
- ما سه نوع چت‌بات اصلی را شناسایی کردیم: دستیارهای شخصی مجازی، چت‌بات‌های سازمانی و مشاوران روبو.
- یکی از کاربردهای نسبتاً جدید سیستم‌های دانشی، دستیارهای شخصی مجازی هستند. مثال‌های بارز اینگونه دستیارها شامل آمازون الکسا، اپل سیری، دستیار گوگل هستند.
- دانش دستیارهای شخصی مجازی به صورت متمرکز در فضای « ابر » نگهداری شده و معمولاً از طریق گفتگوهای پرسش و پاسخ منتشر می‌شوند.
- دستیارهای شخصی می‌توانند دستورات صوتی را دریافت و اجرا کنند.
- دستیارهای شخصی می‌توانند مشاوره‌های شخصی‌سازی شده را به انتخاب خودشان ارائه کنند.
- نسل ویژه‌ای از دستیارها به نام مشاوران شخصی هستند مثل مشاوران روبو که مشاوره‌های شخصی‌سازی شده به سرمایه‌گذاران می‌دهند.
- سیستم‌های توصیه‌گر امروزی از چندین فناوری هوش مصنوعی برای ارائه توصیه‌های شخصی‌سازی شده در مورد محصولات و خدمات استفاده می‌کنند.
- کاربران می‌توانند از طریق پیام‌های نوشتاری، صوت و تصویر با چت‌بات‌ها ارتباط برقرار کنند.
- چت‌بات‌ها حاوی یک پایگاه دانش و یک رابط زبان طبیعی هستند.

- از چت‌بات‌ها عمدتاً برای جستجوی اطلاعات، برقراری ارتباط و همکاری و ارائه مشاوره در حوزه‌های خاص و محدود استفاده می‌شود.
- چت‌بات‌ها می‌توانند با ارائه اطلاعات و خدمات به مشتریان موجب تسهیل خرید آنلاین شوند.
- چت‌بات‌ها با سیستم‌های پیام‌رسان به خوبی کار می‌کنند (مثلاً فیسبوک مسنجر، وی‌چت).
- چت‌بات‌های سازمانی به انواع مشتریان خدمت رسانی کرده و می‌توانند با شرکای تجاری همکاری کنند. آنها همچنین می‌توانند به کارکنان سازمان خدمت دهند.
- دستیارهای شخصی مجازی برای کار با کاربران طراحی شده‌اند و امکان سفارشی‌سازی آنها برای کاربران وجود دارد.
- دستیارهای شخصی مجازی به عنوان محصولات « بومی و ذاتی » برای انبوه کاربران طراحی شده‌اند.
- یک دستیار شخصی مجازی معروف آمازون الکسا می‌باشد که از طریق یک اسپیکر کوچک به نام اکو (یا دیگر اسپیکرهای هوشمند) می‌توان به آن دسترسی یافت.
- فروشندگان متعددی دستیارهای شخصی مجازی را عرضه می‌کنند. معروف‌ترین آنها آمازون الکسا، اپل سیری و دستیار گوگل هستند.
- دستیارهای شخصی مجازی در حوزه‌های خاص دارای تخصص بوده و می‌توانند به عنوان مشاوران سرمایه‌گذاری فعالیت کنند.
- مشاوران رباتی به ارائه مشاوره‌های سرمایه‌گذاری آنلاین و شخصی‌سازی شده با هزینه‌ای بسیار کمتر از مشاوران انسانی می‌پردازند. تا به امروز کیفیت کار آنها در حد مشاوران انسانی بوده است.
- مشاوران رباتی می‌توانند برای موارد خاص و تخصصی با مشاوران انسانی ترکیب نمود.

۱۵۱.

۱۵۲.

۱۵۳.

۱۵۴.

۱۵۵.

۱۵۶.

۱۵۷.

۱۵۸.

	.159
	.160
	.161
	.162
	.163
	.164
	.165
	.13
nuance.com	.14
shopadvisor.com/our-platform	.15
chatbot.org	.16
	.17
	.18
howtogeek.com/253719/do-i-need-anamazon-echo-to-use-alexa	.19
	.20
	.21
gravityinvestment.com/digital-advice-platformdemo	.22
visirule.co.uk	.23
	.24
	.25

¹ Simon Property Group

² Baidu's Melody

quora.com	.۲۶
	.۲۷
	.۲۸
	.۲۹
personalityforge.com	.۳۰
	.۳۱
http://international-adviser.com/news/1035281/alibababacked-retail-robo-adviser-youyu-launches-hongkong	
exsys.com	.۳۲
	.۳۳

¹ Nina
² Nuance Communication
³ Tommy Hilfiger
⁴ Botsify
⁵ Personality Forge
⁶ Youyu

اینترنت اشیاء، پلتفرمی برای کاربردهای هوشمند

اینترنت اشیاء از سال ۲۰۱۴ موضوع داغ فناوری بوده است. کاربردهای آنها به سرعت در حوزه‌ها و بخش‌های مختلف صنعت،

- خدمات دولتی و ارتش در حال افزایش است (مانیکا و همکاران ۲۰۱۵)^۱. برآورد می‌شود که بین سال‌های ۲۰۲۵-۲۰۲۰ بین ۲۰ تا ۵۰ میلیارد «شیء» به اینترنت متصل خواهند شد. اینترنت اشیاء تعداد زیادی از اشیاء هوشمند را به اینترنت متصل کرده و داده‌هایی را گردآوری می‌کند که توسط ابزارهای تحلیلی و سایر سیستم‌های هوشمند پردازش می‌شوند. این فناوری با ابزارهای هوش مصنوعی ترکیب شده و کاربردهای هوشمندی را بوجو می‌آورد که بارزترین آنها خودروهای خودران؛ خانه‌های هوشمند و شهرهای هوشمند هستند.

خدمات، دولت و ارتش در حال افزایش است (مانیکا و همکاران ۲۰۱۵)^۱. برآورد می‌شود که بین سال‌های ۲۰۲۵-۲۰۲۰ بین ۲۰ تا ۵۰ میلیارد «شیء» به اینترنت متصل خواهند شد. اینترنت اشیاء تعداد زیادی از اشیاء هوشمند را به اینترنت متصل کرده و داده‌هایی را گردآوری می‌کند که توسط ابزارهای تحلیلی و سایر سیستم‌های هوشمند پردازش می‌شوند. این فناوری با ابزارهای هوش مصنوعی ترکیب شده و کاربردهای هوشمندی را بوجو می‌آورد که بارزترین آنها خودروهای خودران؛ خانه‌های هوشمند و شهرهای هوشمند هستند.

۱۳.۱. پرونده‌ی آغازین: سی‌ان‌اچ اینداستریال از اینترنت اشیاء برای تعالی خود استفاده می‌کند

۱۳.۲. اصول و مبانی اینترنت اشیاء

۱۳.۳. مزایا و محرک‌های اصلی اینترنت اشیاء

۱۳.۴. اینترنت اشیاء چگونه کار می‌کند

۱۳.۵. حسگرها و نقش آنها در اینترنت اشیاء

۱۳.۶. کاربردهای منتخب اینترنت اشیاء

۱۳.۷. خانه‌ها و لوازم خانگی هوشمند

۱۳.۸. شهرها و کارخانجات هوشمند

۱۳.۹. وسایل نقلیه‌ی خودران (مستقل)

¹ Manyka et al.

سی‌ان‌اچ اینداستریال ان.وی.^۱ یک شرکت هلندی تولیدکننده‌ی وسایل نقلیه‌ی کشاورزی، راه‌سازی و بازارهای تجاری است. این شرکت به تولید و خدمت‌رسانی به بیش از ۳۰۰ نوع وسیله نقلیه پرداخته و در ۱۹۰ کشور با بیش از ۶۵،۰۰۰ کارمند فعالیت دارد. کسب‌وکار این شرکت علیرغم اینکه در محیطی بسیار رقابتی فعالیت می‌کند دائماً در حال رشد است.

برای مدیریت و هماهنگی چنین کسب‌وکار پیچیده‌ای از دفتر مرکزی شرکت در لندن، شرکت به یک سیستم ارتباطی فوق‌العاده بزرگ و همچنین قابلیت‌های تحلیلی مؤثر و یک شبکه خدمات مشتریان نیاز داشت. به عنوان مثال، در دسترس بودن قطعات یدکی امری حیاتی بود. تجهیزات و ادوات مشتریان در صورت خرابی نیازمند این بود که قطعات تعویض شوند. فشارهای بازار رقابتی بسیار شدید بودند به ویژه در بخش کشاورزی که شرایط آب و هوایی، تغییر فصل‌ها و فشارهای مربوط به برداشت محصول می‌توانند بر پیچیدگی عملیات بیفزایند. پایش و کنترل مناسب تجهیزات یک عامل رقابتی بسیار مهم محسوب می‌شود. پیش‌بینی خرابی دستگاه‌ها و تجهیزات امری ضروری است. برقراری سریع ارتباط با مشتریان و تجهیزاتی که از سی‌ان‌اچ خریداری کرده‌اند به اندازه‌ی پایش و گردآوری مؤثر و کارآمد داده‌ها اهمیت دارد. هم سی‌ان‌اچ و هم مشتریانانش باید تصمیمات مستمری را اتخاذ می‌کنند که نیازمند جریان بلادرنگ اطلاعات و ارتباطات بود.

سی‌ان‌اچ با بهره‌گیری از شرکت پی‌تی‌سی ترنسفورمیشنال^۲ به عنوان عرضه‌کننده‌ی اینترنت اشیا یک سیستم مبتنی بر اینترنت اشیا را با تغییر ساختار درون سازمانی برای حل این مشکلات پیاده‌سازی کرده و شکل جدیدی به وسایل نقلیه‌ی صنعتی متصل خود داد. پیاده‌سازی اولیه در بخش کشاورزی اتفاق افتاد. جزئیات پیاده‌سازی این سیستم توسط شرکت پی‌تی‌سی (۲۰۱۵) ارائه شده است. نکات مهم این سیستم اینترنت اشیا در ادامه ارائه شده است:

- تمامی وسایل نقلیه (آنهايي که مجهز به حسگرها بوده و به این سیستم متصل هستند) را در صدها مکان در سرتاسر جهان به مرکز کنترل و فرماندهی سی‌ان‌اچ متصل می‌کند. این اتصال پایش عملکرد را امکان‌پذیر می‌کند.
- شرایط و عملیات محصولات و همچنین محیط پیرامون آنها را از طریق حسگرها پایش می‌کند. همچنین داده‌های محیط بیرون از قبیل شرایط آب و هوایی را گردآوری می‌کند.
- امکان سفارشی‌سازی عملکرد محصولات را در محل مشتری فراهم می‌کند.

¹ CNH Industrial N.V.

² PTC Transformational

- داده‌های لازم برای بهینه‌سازی عملیات تجهیزات را فراهم می‌کند.
- عملکرد رانندگان وسائل نقلیه‌ی تولیدی شرکت را تحلیل کرده و تغییراتی که برای بهبود کارایی وسیله‌ی نقلیه لازم هستند را توصیه می‌کند.
- دامنه‌ی تأمین سوخت وسایل نقلیه را پیش‌بینی می‌کند.
- به رانندگان در خصوص ضرورت و زمان انجام نگهداری‌های پیشگیرانه (مثلاً با پایش میزان کارکرد و یا پیش‌بینی خرابی) و سفارش‌گذاری قطعات برای این خدمات اطلاع‌رسانی می‌کند. بدین ترتیب انجام نگهداری‌های پیشگیرانه امکان‌پذیر می‌شود.
- از اضافه‌بار کامیون‌ها که از گارانتی سی‌ان‌اچ تخطی می‌کنند آگاه می‌شود.
- تشخیص زودهنگام خرابی محصولات را امکان‌پذیر می‌کند.
- با اتصال کامیون‌ها به پرسنل برنامه‌ریزی و مبدأ و مقصد تحویل امکان تحویل به موقع محصولات را فراهم می‌کند.
- به کشاورزان در برنامه‌ریزی بهینه‌ی چرخه‌ی کشت و زرع از زمان آماده‌سازی زمین تا برداشت کمک می‌کند (با تحلیل شرایط آب و هوایی).
- داده‌های گردآوری شده را تحلیل کرده و آنها را با استانداردها مقایسه می‌کند.

تمامی این کارها اکثراً به صورت بی‌سیم انجام می‌شوند.

طبق مارکوس^۱ (۲۰۱۵)، سی‌ان‌اچ زمانی خرابی تجهیزات خود در محل مشتری را با استفاده از اینترنت اشیاء به نصف رساند. قطعات سفارش‌گذاری شده بسیار سریع ارسال می‌شدند. اینترنت اشیاء همچنین به کشاورزان در پایش مزارع و تجهیزات برای بهبود کارایی کمک کرد. علاوه بر این، در توسعه‌ی محصول از تحلیل داده‌های گردآوری شده استفاده شد.

Sources: Compiled from PTC, Inc. (2015), Marcus (2015), and cnhindustrial.com/en-us/pages/homepage.aspx.

سوالاتی برای مرور پرونده‌ی آغازین

۱. چرا اینترنت اشیاء تنها راه‌حل مناسب برای مشکلات سی‌ان‌اچ بود؟
۲. مزایای اصلی اینترنت اشیاء را نام برده و توضیح دهید.
۳. سی‌ان‌اچ چگونه می‌تواند در توسعه‌ی محصول از داده‌های گردآوری شده مربوط به کارکرد تجهیزات بهره‌بردار؟
۴. گفته می‌شود که اینترنت اشیاء تلماتیک و وسائل نقلیه‌ی متصل را امکان‌پذیر می‌کند. توضیح دهید.
۵. چرا اینترنت اشیاء « رکن استراتژی آینده‌ی کسب‌وکار » به حساب می‌آید؟
۶. گفته می‌شود که اینترنت اشیاء خدمات جدیدی را برای سی‌ان‌اچ امکان‌پذیر می‌کند (مثلاً برای فروش و همکاری با شرکا). توضیح دهید.
۷. شکل ۱۳.۱ را نگاه کنید و ارتباط آن را با کاربرد اینترنت اشیاء در سی‌ان‌اچ توضیح دهید.
۸. امکانات پشتیبانی تصمیم را شناسایی کنید.

¹ Marcus

۹. اینترنت اشیا از کدام تصمیماتی که شرکت گرفته و کدام تصمیماتی که مشتریان گرفته‌اند پشتیبانی می‌کند؟

ابتدا، آموختیم که اینترنت اشیا چگونه زیرساختی را برای انواع جدید کاربردهایی که هزاران شیء را به یک مرکز تصمیم‌گیری متصل می‌کنند فراهم می‌کند.

دوم، آموختیم که داده‌های گردآوری شده توسط حسگرها چگونه از وسیله‌ی نقلیه و محیط پیرامون آنها و ارسال برای پردازش‌های تحلیلی جریان می‌یابد.

سوم، سازنده‌ی وسایل نقلیه و مالکان و کاربران آنها می‌توانند از مزایای بسیار زیاد بکارگیری این سیستم بهره‌مند شوند. نهایتاً، اینترنت اشیا یک چارچوب ارتباط و همکاری را برای تصمیم‌گیرندگان، تولیدکننده و کاربران محصولات خریداری شده فراهم می‌کند.

در این فصل ما به بررسی بیشتر فناوری‌های درگیر و فرایند عملیاتی اینترنت اشیا می‌پردازیم. همچنین کاربرد مهم آن در سازمان‌ها منازل، شهرهای هوشمند و وسایل نقلیه‌ی خودران را توضیح خواهیم داد.

اینترنت اشیا^۱ یک عبارت نوظهور با تعاریف متعدد است. به طور کلی، اینترنت اشیا به یک شبکه‌ی رایانه‌ای اشاره دارد که اشیا زیادی (افراد، حیوانات، دستگاه‌ها، حسگرها، ساختمان‌ها) را که هر یک مجهز به یک ریزپردازنده هستند به یکدیگر متصل می‌کند. اشیا عمدتاً به صورت بی‌سیم به اینترنت متصل شده و اینترنت اشیا را تشکیل می‌دهند. اینترنت اشیا می‌تواند داده‌ها را تبادل کرده و بین اشیا و محیط پیرامون آنها ارتباط برقرار می‌کند. بدین ترتیب اینترنت اشیا به افراد و اشیا اجازه می‌دهد تا در هر زمان و مکان به یکدیگر متصل شوند. حسگرهای تعبیه شده که داده‌ها را گردآوری و مبادله می‌کنند بخش مهمی از اشیا و اینترنت اشیا را تشکیل می‌دهند. تحلیلگران پیش‌بینی می‌کنند که تا سال ۲۰۲۵ بیش از ۵۰ میلیارد دستگاه (شیء) به اینترنت متصل خواهند شد و ستون فقرات کاربردهای اینترنت اشیا را شکل خواهند داد. چالش‌ها و فرصت‌های این فناوری تحول‌آفرین (مثلاً کاهش هزینه‌ها، ایجاد مدل‌های کسب‌وکار جدید، بهبود کیفیت) در مصاحبه‌ای با پیتر اوتزاشنایدر^۲ معاون مدیریت محصول جاوا در اوراکل مورد بحث قرار گرفته است (کویتکا ۲۰۱۴)^۳. علاوه بر این، می‌توانید به نشست‌های گفتگوی iotcommunity.com بپیوندید. برای اطلاع از چشم‌انداز شرکت اینتل در خصوص یک جهان کاملاً متصل به موری^۴ (۲۰۱۶) مراجعه کنید.

تعبیه کردن رایانه‌ها و سایر دستگاه‌هایی که بتوان آنها را از هر کجا فعال و غیرفعال کرده و متصل کردن تمامی دستگاه‌ها به اینترنت (و/یا به یکدیگر)، ارتباطات و همکاری‌های گسترده‌ای را بین کاربران و اشیا امکان‌پذیر می‌سازد. با اتصال دستگاه‌های بسیاری که بتوانند با یکدیگر صحبت کنند می‌توان کاربردهایی با کارکردهای جدید را ایجاد کرده، بهره‌وری سیستم‌های موجود را افزایش داد. این نوع تعامل، فضا برای کاربردهای بسیار زیادی را فراهم می‌کند. در خصوص کاربردهای کسب‌وکار اینترنت اشیا

¹ Internet of Things (IoT)

² Peter Utzschneider

³ Kvitka

⁴ Murray

نگاه کنید به جامته^۱ (۲۰۱۶). علاوه بر این، از « کنسرسیوم اینترنت اشیا » (iofthings.org) و کنفرانس‌های سالانه آن بازدید کنید. برای یک اینفوگرافی و یک راهنما به intel.com/content/www/us/en/internet-of-things/infographics.guide-to-ito.html مراجعه کنید.

چندین تعریف برای اینترنت اشیا وجود دارد. کوین آشتون^۲ که عبارت « اینترنت اشیا » را برای اولین بار به کار برد این تعریف را ارائه کرده است: « اینترنت اشیا به معنی حسگرهای متصل به اینترنت و رفتار اینترنت‌گونه داشتن با برقراری ارتباطات باز و آد هوک، به اشتراک‌گذاری آزاد داده‌ها و مجاز نمودن کاربردهای غیرمنتظره است تا رایانه‌ها بتوانند دنیای اطرافشان را درک کرده و به یک سیستم عصبی برای بشریت تبدیل شوند ».

تعریف کاربردی ما این است:

اینترنت اشیا شبکه‌ای از دستگاه‌های رایانشی متصل بجهت از جمله انواع مختلف اشیا (مثلاً ماشین‌های دیجیتالی) است. هر شیء در شبکه یک شناسار منحصر بفرد بوده و قادر به گردآوری و انتقال خودکار داده‌ها در سرتاسر شبکه می‌باشد. داده‌های گردآوری شده تا زمانیکه تحلیل نشوند هیچ ارزشی ندارند.

توجه داشته باشید که اینترنت اشیا به افراد و اشیا اجازه می‌دهد تا هر در زمان و مکان صرف‌نظر از سرویس یا موضوع کسب‌وکار با یکدیگر ارتباط و تعامل داشته باشند.

طبق میلر^۳ (۲۰۱۵)، اینترنت اشیا یک شبکه متصل است که در آن:

- تعداد زیادی از اشیا می‌توانند به یکدیگر متصل شوند.
- هر شیء دارای یک تعریف منحصر بفرد است (آدرس آی‌پی).
- هر شیء توانایی دریافت، ارسال و ذخیره‌سازی خودکار داده‌ها را دارد.
- هر شیء عمدتاً به صورت بی‌سیم در شبکه حضور دارد.
- هر شیء بر اساس ارتباط ماشین به ماشین ساخته شده است.

توجه داشته باشید که برخلاف اینترنت معمول که افراد را با استفاده از فناوری رایانشی به یکدیگر متصل می‌کند، اینترنت اشیا، «اشیا» را به یکدیگر و به حسگرهایی که داده‌ها را گردآوری می‌کنند متصل می‌کند. در بخش ۱۳.۴ ما فرایند اینترنت اشیا را توضیح خواهیم داد.

یک مثال متداول از اینترنت اشیا، وسایل نقلیه‌ی خودران هستند (بخش ۱۳.۹). برای اینکه یک وسیله‌ی نقلیه بتواند به صورت خودران کار کند باید حسگرهای کافی برای پایش خودکار محیط و موقعیت پیرامون خود را داشته و برای تغییر هرگونه تنظیمانی از جمله سرعت، جهت و ... خودرو در زمان لازم اقدام مناسب را اتخاذ کند. یک مثال دیگر، شرکت اسمارت‌بین است. این شرکت سطل‌های زباله‌ای را

¹ Jamthe

² Kevin Ashton

³ Miller

طراحی کرده و دارای حسگرهایی برای تشخیص میزان محتویات آن هستند. زمانیکه حسگر تشخیص دهد که محتویات سطل زباله به سطح مورد نظر رسیده به صورت خودکار شرکت را جهت تخلیه زباله‌ها مطلع می‌کند. یکی از تصوراتی که مردم از اینترنت اشیاء دارند این است که یک یخچال بتواند زمانیکه تشخیص می‌دهد که مواد غذایی تمام شده است به صورت خودکار غذا سفارش دهد. کلوروکس^۱ یک فیلتر بریتای^۲ جدید را معرفی کرده که یک مکانیزم متصل به وای‌فای بتواند زمانیکه تشخیص می‌دهد که زمان تغییر فیلتر آب یخچال فرا رسیده به صورت خودکار سفارش‌گذاری کند. در این مثال‌ها، نیازی به ارتباط بین انسان با انسان یا حتی انسان با ماشین نیست.

اینترنت اشیاء می‌تواند به ما در بسیاری از زمینه‌ها کمک کند. در ادامه به بررسی برخی از این کاربردها می‌پردازیم.

- **امنیت و سلامت:** اینترنت اشیاء می‌تواند به ما در نظارت بر سلامت و امنیت کمک کند. به عنوان مثال، می‌تواند به ما در نظارت بر سلامت کودکان کمک کند. می‌تواند به ما در نظارت بر سلامت سالمندان کمک کند. می‌تواند به ما در نظارت بر سلامت حیوانات خانگی کمک کند.
- **مدیریت انرژی:** اینترنت اشیاء می‌تواند به ما در مدیریت انرژی کمک کند. می‌تواند به ما در نظارت بر مصرف انرژی کمک کند. می‌تواند به ما در کاهش مصرف انرژی کمک کند.
- **مدیریت ترافیک:** اینترنت اشیاء می‌تواند به ما در مدیریت ترافیک کمک کند. می‌تواند به ما در نظارت بر ترافیک کمک کند. می‌تواند به ما در کاهش ترافیک کمک کند.
- **مدیریت زباله:** اینترنت اشیاء می‌تواند به ما در مدیریت زباله کمک کند. می‌تواند به ما در نظارت بر زباله کمک کند. می‌تواند به ما در کاهش زباله کمک کند.

زمانیکه میلیاردها شیء با تمامی خدمات پشتیبانی و زیرساخت‌های فناوری اطلاعات متصل به اینترنت متصل می‌شوند، می‌توانیم مجموعه‌ای غول پیکر را به شکل یک اکوسیستم عظیم ببینیم. اکوسیستم اینترنت اشیاء به تمامی اجزایی اشاره دارد که کاربران را قادر می‌سازد تا کاربردهای اینترنت اشیاء را خلق کنند. این اجزا شامل درگاه‌ها، تحلیل‌ها، الگوریتم‌های هوش مصنوعی، سرورها، انبارهای داده، امنیت و دستگاه‌های اتصال می‌باشند. یک شمای تصویری در شکل ۱۳.۱ نشان داده شده است که کاربردهای اینترنت اشیاء در سمت چپ و اجزا و پلتفرم‌های سازنده در سمت راست نشان داده شده‌اند. یک مثال از کاربردهای اینترنت اشیاء در پرونده‌ی آغازین معرفی شد. این تصویر شبکه‌ای از حسگرها را نمایش می‌دهد که اطلاعات را گردآوری کرده و به یک مکان مرکزی برای پردازش و نهایتاً پشتیبانی تصمیم ارسال می‌کنند. بنابراین، کاربردهای اینترنت اشیاء یک زیرسیستم از اکوسیستم اینترنت اشیاء است.

¹ Clorox

² Brita

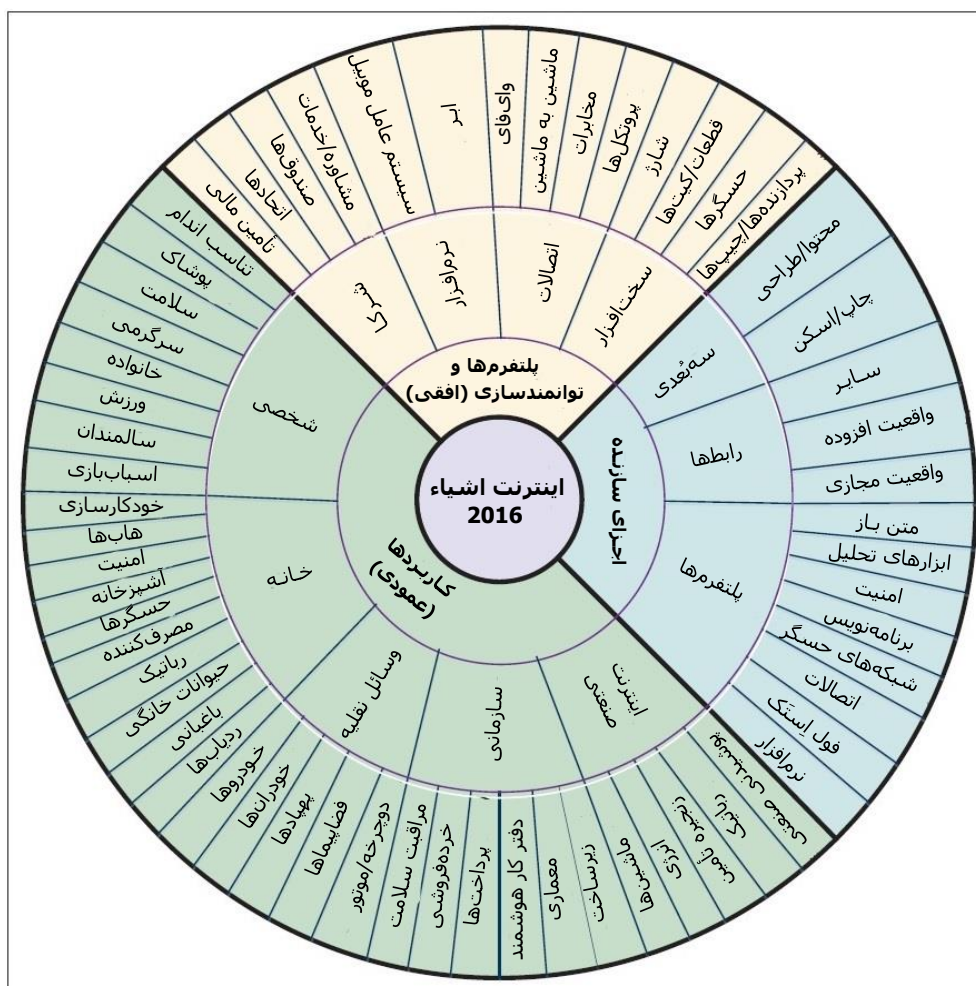
³ McKaffetry

⁴ Bert

⁵ Greengard

یک بحث پایه، اصطلاحات، شرکت‌های مهم و پلتفرم‌ها توسط مئولا^۱ (۲۰۱۸) ارائه شده است.

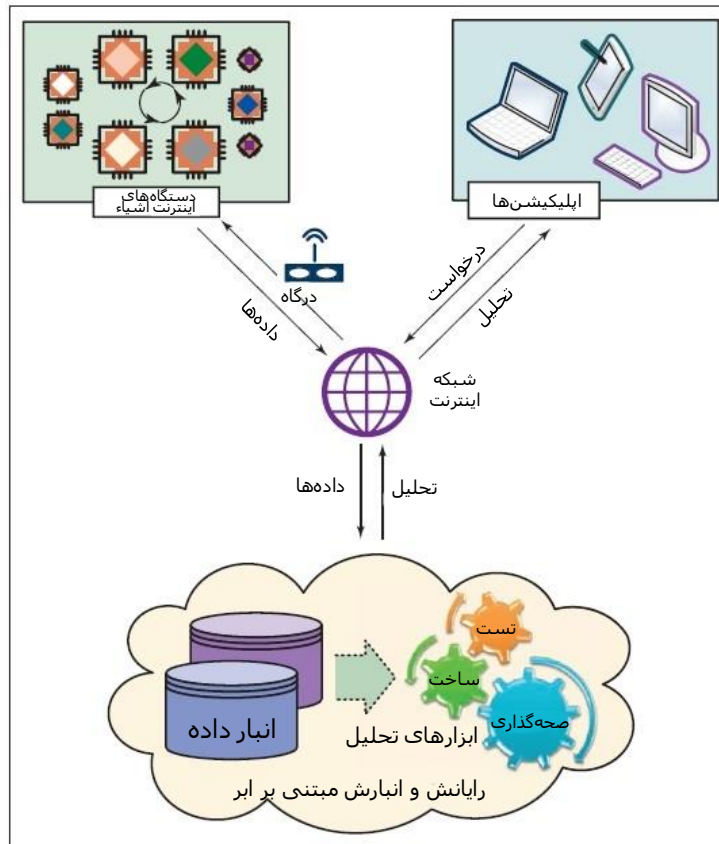
اینترنت اشیا به انواع مختلف اشیاء و دستگاه‌ها از خودرو و لوازم خانگی گفته تا لوازم پزشکی، رایانه‌ها، سخت‌افزار، نرم‌افزار، حسگرهای داده و غیره اشاره دارد. اتصال اشیاء و برقراری ارتباط بین آنها یکی از قابلیت‌های ضروری یک کاربرد اینترنت اشیا می‌باشد ولی برای کاربردهای پیچیده‌تر ما به اجزای بیشتری نیاز داریم: یک سیستم کنترلی و یک مدل کسب‌وکار. اینترنت اشیا، اشیاء را قادر می‌سازد تا به صورت بی‌سیم در سطح شبکه حس کرده و حس شوند. یک مثال غیر اینترنت، یک سیستم کنترل دما در اتاق است. یک مثال غیراینترنتی دیگر، چراغ راهنمایی در تقاطع خیابان‌ها است که حسگر دوربین‌ها قادر به تشخیص خودروهای در حال تردد در طرفین خیابان بوده و یک سیستم کنترلی زمان تغییر چراغ را بر اساس قواعد برنامه‌ریزی شده تنظیم می‌کند. بعدها، ما به معرفی کاربردهای اینترنتی خواهیم پرداخت.



شکل ۱۳.۱ اکوسیستم اینترنت اشیا ۲۰۱۶

¹ Meola

۱. **سخت‌افزار**. شامل دستگاه‌ها و لوازم فیزیکی، حسگرها و عملگرهای مکانیکی است که داده‌ها توسط آنها تولید و ضبط می‌شود. این دستگاه‌ها تجهیزاتی هستند که باید کنترل، پایش یا ردیابی شوند. دستگاه‌های حسگر اینترنت اشیا می‌توانند حاوی یک پردازنده یا هرگونه دستگاه رایانشی دیگر که داده‌های ورودی را تجزیه می‌کند باشند.



شکل ۱۳.۲ اجزای سازنده اینترنت اشیا

۲. **اتصال**. برای گردآوری داده‌ها از حسگرها و اشیاء و ارسال آن داده‌ها به فضای ابر جهت تحلیل باید یک ایستگاه یا هاب مرکزی وجود داشته باشد. دستگاه‌ها برای برقراری ارتباط با سایر شبکه‌ها یا اپلیکیشن‌ها به یک شبکه متصل می‌شوند. همچنین ممکن است مستقیماً به اینترنت وصل شوند. دستگاه‌هایی که مستقیماً به اینترنت متصل نیستند می‌توانند از طریق درگاه‌ها به پلتفرم ابر دسترسی پیدا کنند.

۳. **بک‌اند نرم‌افزاری**. در این لایه، داده‌های گردآوری شده مدیریت می‌شوند. بک‌اند نرم‌افزاری، شبکه‌ها و دستگاه‌های متصل را مدیریت کرده و داده‌ها را یکپارچه می‌کند. اینکار ممکن است در فضای ابر انجام شود.

۴. **اپلیکیشن‌ها.** در این بخش از اینترنت اشیا، داده‌ها به اطلاعات معنادار تبدیل می‌شوند. بسیاری از اپلیکیشن‌ها می‌توانند بر روی گوشی‌های هوشمند، تبلت‌ها و رایانه‌های شخصی اجرا شده و کار مفیدی را با داده‌ها انجام دهند. اپلیکیشن‌های دیگر می‌توانند بر روی سرور اجرا شده و نتایج یا هشدارها را از طریق داشبوردها یا پیام‌ها برای ذینفعان ارسال کنند. برای کمک به ساخت سیستم‌های اینترنت اشیا، می‌توان از پلتفرم‌های اینترنت اشیا استفاده کرد. برای اطلاعات بیشتر به مئولا (۲۰۱۸) نگاه کنید.

ibm.com/us-en/marketplace/internet-of-things-cloud

سوالاتی برای مرور بخش ۱۳.۲

۱. اینترنت اشیا چیست؟
۲. ویژگی‌های مهم اینترنت اشیا را نام ببرید.
۳. چرا اینترنت اشیا اهمیت دارد؟
۴. تغییرات مهمی که به واسطه‌ی اینترنت اشیا به وجود آمده‌اند را نام ببرید.
۵. منظور از اکوسیستم اینترنت اشیا چیست؟
۶. اجزای اصلی فناوری اینترنت اشیا چه هستند؟

هدف اصلی سیستم‌های اینترنت اشیا، بهبود بهره‌وری، کیفیت، سرعت و کیفیت زندگی است. اینترنت اشیا دارای چندین مزیت بالقوه‌ی مهم است به ویژه زمانیکه با هوش مصنوعی ترکیب شود. برای بحث و مثال‌ها به جامته (۲۰۱۵) نگاه کنید.

مزیت‌های اصلی اینترنت اشیا در ادامه آورده شده است:

- کاهش هزینه از طریق خودکارسازی فرایندها.

¹ Amazon AWS IoT

² Microsoft Azure IoT Suite

³ Predix IoT

⁴ IBM Watson IoT Platform

- بهبود بهره‌وری کارکنان
- ایجاد جریان‌های درآمدی جدید
- بهینه‌سازی بهره‌برداری از دارایی‌ها
- افزایش پایداری
- تغییر و بهبود همه چیز
- پیش‌بینی احتمال نیازهای ما
- امکان دستیابی به بینش در محیط‌های وسیع (حسگرهایی که داده‌ها را گردآوری می‌کنند)
- امکان اتخاذ تصمیمات/خریدهای هوشمندانه‌تر
- افزایش دقت پیش‌بینی‌ها
- شناسایی سریع مسائل (حتی پیش از آنکه رخ دهند).
- امکان تولید و انتشار بلادرنگ اطلاعات
- امکان ردیابی سریع و ارزان قیمت فعالیت‌ها.
- کارآمدتر کردن فرایندهای کسب‌وکار.
- امکان برقراری ارتباط بین مصرف‌کنندگان و مؤسسات مالی
- تسهیل استراتژی رشد.
- بهبود بنیادی استفاده از تحلیل‌ها.
- امکان تصمیم‌گیری بهتر بر اساس اطلاعات بلادرنگ.
- ارائه‌ی راه‌حل مسئله و بازیابی در شرایط خرابی.
- پشتیبانی از یکپارچه‌سازی تأسیسات.
- ارائه دانش بهتر در مورد مشتریان برای بازاریابی و خدمات شخصی‌سازی شده.

در ادامه محرک‌های اصلی اینترنت اشیاء ارائه شده است:

- بین ۲۰ تا ۵۰ میلیارد شیء ممکن است بین سال‌های ۲۰۲۵-۲۰۳۰ به اینترنت متصل شوند.
- سیستم‌ها/اشیاء متصل مستقل (مثل ربات‌ها، خودروها) کاربردهای جدیدی را برای اینترنت اشیاء ایجاد می‌کنند.
- اینترنت باند پهن به طور گسترده‌تری در دسترس است.
- هزینه‌ها دستگاه‌ها و حسگرها دائماً در حال کاهش است.
- هزینه‌ی اتصال دستگاه‌ها در حال کاهش است.
- دستگاه‌های بیشتری ساخته شده و به راحتی به یکدیگر متصل می‌شوند (مثلاً نگاه کنید به فنویک^۱ ۲۰۱۶).
- از حسگرهای بیشتری در دستگاه‌ها استفاده می‌شود.
- نفوذ گوشی‌های هوشمند به شدت در حال افزایش است.

¹ Fenwick

- در دسترس بودن دستگاه‌های پوشیدنی در حال افزایش است.
- سرعت جابجایی داده‌ها تا ۶۰ تراهرتز افزایش می‌یابد.
- پروتکل‌هایی برای اینترنت اشیاء تدوین شده است (مثلاً وی‌گیگ^۱).
- انتظارات مشتریان در حال افزایش است، خدمات نوآورانه برای مشتریان تبدیل به یک ضرورت شده است.
- در دسترس بودن ابزارها و پلتفرم‌های اینترنت اشیاء در حال افزایش است.
- در دسترس بودن تحلیل‌های قدرتمند که با اینترنت اشیاء بکار گرفته می‌شوند در حال افزایش است.

مزایا و محرک‌هایی که در بالا ذکر شد فرصت‌های زیادی را برای تعالی سازمان‌ها در مسائل اقتصادی (مثلاً سینکِر^۲ ۲۰۱۷)، در بسیاری از صنایع و محیط‌های مختلف ایجاد می‌کنند. مؤسسه بین‌المللی مک‌کنزی^۳ (مانیکا و همکاران ۲۰۱۵) فهرست جامعی از محیط‌هایی که از اینترنت اشیاء می‌توان استفاده کرد به همراه مثال‌هایی برای هر محیط ارائه کرده‌اند. مطالعه‌ای در سال ۲۰۱۷ (استاف^۴ ۲۰۱۷)، یک افزایش چشمگیر در قابلیت‌ها و مزایای اینترنت اشیاء را نشان داد.

این شبکه توسط شرکت ارتباطات تاتا^۵ هند و شرکت اچ‌پی ایالات متحده بر روی پلتفرم اچ‌پی‌ای یونیورسال آی‌آئی^۶ ساخته می‌شود. اشیائی که قرار است متصل شوند در ۲,۰۰۰ مجموعه متشکل از دستگاه‌های رایانش، اپلیکیشن‌ها و راه‌حل‌های اینترنت اشیاء قرار دارند که بر روی شبکه «لو را»^۷ که یک پروتکل ارتباطی بی‌سیم برای شبکه‌های گسترده است متصل می‌شوند. این اشیاء در ساختمان‌ها، تأسیسات، پردیس‌های دانشگاهی، سیستم‌های امنیتی، وسائل نقلیه و ناوگان‌ها و تأسیسات مراقب سلامت هوشمند قرار دارند.

این پروژه‌ها قرار است طی چند فاز با اپلیکیشن‌هایی که قرار است ابتدا تست شوند اجرا شود. این شبکه به ۴۰۰ میلیون نفر خدمت‌رسانی خواهد کرد. برای جزئیات نگاه کنید به شاه (۲۰۱۷).

سوالاتی برای مرور بخش ۱۳.۳

۱. مزایای اینترنت اشیاء برای سازمان‌ها را نام ببرید.

¹ WiGig

² Sinclair

³ McKinsey Global Institute

⁴ Staff

⁵ Tata Communications

⁶ HPE Universal IoT Platform

⁷ Lo Ra

۲. مزایای اینترنت اشیا برای مشتریان را نام ببرید.
۳. مزایای اینترنت اشیا برای تصمیم‌گیری را نام ببرید.
۴. محرک‌های اصلی اینترنت اشیا را نام ببرید.

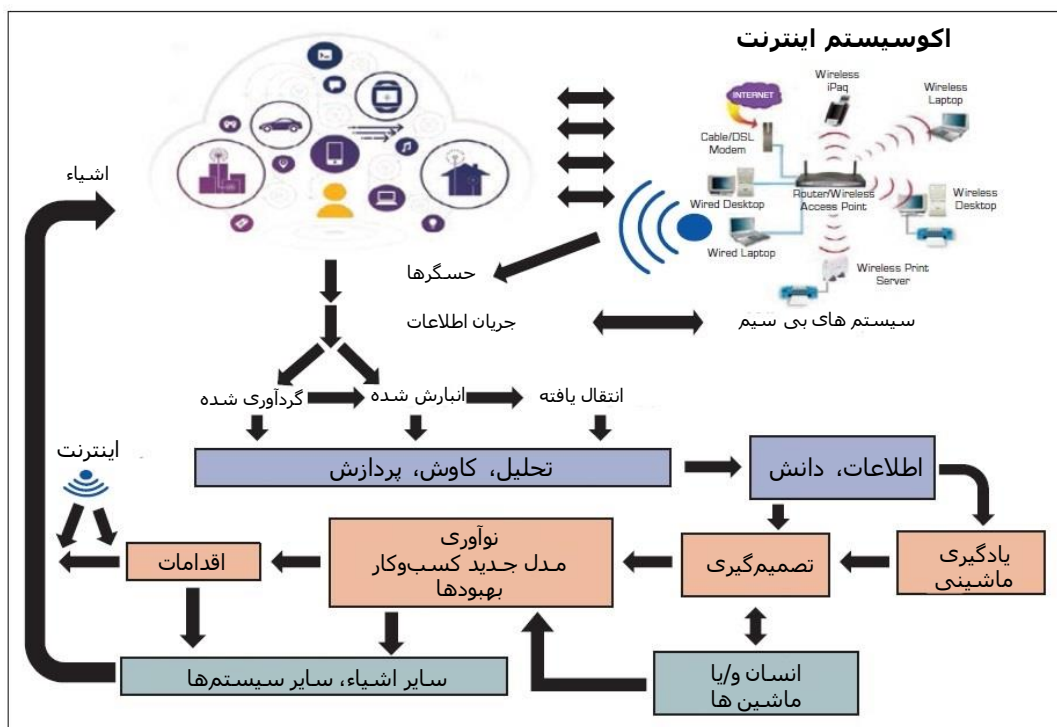
اینترنت اشیا یک اپلیکیشن نیست بلکه یک زیرساخت، پلتفرم یا چارچوب است که برای پشتیبانی از اپلیکیشن‌ها بکار می‌رود. در ادامه فرایند جامع کاربردهای اینترنت اشیا توضیح داده شده است. در بسیاری از موارد، اینترنت اشیا تنها از بخش‌هایی از این فرایند پیروی می‌کند.

فرایند در شکل ۱۳.۳ توضیح داده شده است. اکوسیستم اینترنت (بالای تصویر) شامل تعداد زیادی از اشیا است. حسگرها و سایر دستگاه‌ها به گردآوری اطلاعات از این اکوسیستم می‌پردازند. اطلاعات گردآوری شده را می‌توان نمایش داده، ذخیره کرده و تحلیل و پردازش نمود (مثلاً از طریق داده‌کاوی). بر اساس این تحلیل‌ها، اطلاعات به دانش و/یا هوش تبدیل می‌شوند. سیستم‌های خبره یا یادگیری ماشینی ممکن است در تبدیل این دانش به پشتیبانی تصمیم کمک کنند که بهبود نتایج و اقدامات مؤید این مسئله است.

تصمیمات تولید شده می‌توانند به ایجاد اپلیکیشن‌های نوآورانه، مدل‌های کسب‌وکار جدید و بهبود فرایندهای کسب‌وکار کمک کنند. اینها منجر به «اقدامات» می‌شوند که ممکن است بر سناریوی اصلی یا سایر اشیا تأثیر بگذارند. پرونده‌ی آغازین این فرایند به تصویر کشیده است. توجه داشته باشید که اکثر اپلیکیشن‌های موجود در بخش بالای تصویر قرار دارند که آنها را حسگر به بینش^۱ می‌نامند بدین معنی که تا خلق دانش یا ارائه‌ی اطلاعات جدید پیش می‌روند. با وجود این، امروزه تمرکز بر حرکت به کل چرخه است (یعنی حسگر به اقدام).

اینترنت اشیا ممکن است حجم انبوهی از داده‌ها (کلان‌داده‌ها) را تولید کند که نیازمند تحلیل توسط روش‌های مختلف هوش تجاری از جمله یادگیری عمیق یا روش‌های هوش مصنوعی پیشرفته باشد.

همانطور که قبلاً گفته شد، اینترنت اشیا دانش و/یا هوش را به وجود می‌آورد که به عنوان پشتیبانی برای تصمیم‌گیرندگان ارسال شده یا ورودی سیستم‌های پشتیبان تصمیم خودکار می‌شود. انتقال از گردآوری داده‌ها به پشتیبانی تصمیم به دلیل حجم بالای داده‌ها که برخی از آنها نامربوط هستند ممکن است کار ساده‌ای نباشد. در اینترنت اشیا در مقیاس بزرگ داده‌های گردآوری شده پیش از آنکه برای پشتیبانی تصمیم بکار گرفته شوند معمولاً باید فیلتر و پاکسازی شوند به ویژه اگر به عنوان یک مبنا برای تصمیم‌گیری خودکار بکار می‌روند.



شکل ۱۳.۳ فرایند اینترنت اشياء

سوالاتی برای مرور بخش ۱۳.۴

۱. اجزای اصلی اینترنت اشياء را تویح دهید.
۲. طبق فرایند شکل ۱۳.۳ توضیح دهید اینترنت اشياء چگونه کار می‌کند.
۳. اینترنت اشياء چگونه از تصمیم‌گیری پشتیبانی می‌کند؟

همانگونه که در پرونده‌ی آغازین این فصل دیدیم، حسگرها با گردآوری داده‌های عملکرد اشیائی که به اینترنت متصل هستند و پایش محیط پیرامون نقش اساسی و مهمی را اینترنت اشياء ایفا می‌کنند. حسگرها می‌توانند داده‌ها را ارسال کرده و برخی اوقات حتی آنها را پیش از ارسال پردازش کنند.

یک حسگر یک دستگاه الکترونیکی است که به صورت خودکار داده‌های مربوط به رویدادها یا تغییرات محیط را گردآوری می‌کند. در بسیاری از اپلیکیشن‌های اینترنت اشياء از حسگرها استفاده شده است. داده‌های گردآوری شده برای پردازش به سایر دستگاه‌های الکترونیکی ارسال می‌شوند. چندین نوع حسگر و روش‌های مختلفی برای گردآوری داده‌ها وجود دارد. حسگرها اغلب سیگنال‌هایی را تولید می‌کنند که به شکل نمایشگرهای قابل خواندن توسط انسان تبدیل می‌شوند. علاوه بر استفاده‌ی آنها در اینترنت اشياء،

حسگرها اجزای ضروری در بحث رباتیک و خودروهایی خودران نیز هستند. هر حسگر معمولاً از لحاظ حداکثر فاصله برای تشخیص دارای محدودیت است. حسگرهای با دامنه بسیار کوتاه را حسگرهای مجاورتی^۱ می‌نامند که نسبت به حسگرهای با دامنه‌ی وسیع قابل اطمینان‌تر هستند. هر شبکه‌ی اینترنت اشیا می‌تواند میلیون‌ها حسگر داشته باشد. نحوه کارکرد حسگرها در اینترنت اشیا در مورد کاربردی ۱۳.۱ توضیح داده شده است.

استفاده از حسگرها، اینترنت اشیا و هوش مصنوعی برای کنترل محیطی در فرودگاه بین‌المللی آتن، یونان

مورد کاربردی ۱۳.۱

مسئله

سالانه بیش از ۲۰ میلیون مسافر از این فرودگاه استفاده می‌کنند که تعداد آنها هر سالانه به میزان ۱۰ درصد افزایش می‌یابد. مشخصاً، تعداد پروازها بسیار زیاد بوده و هر ساله در حال افزایش است. این رشد مورد افزایش آلودگی هوا نیز می‌شود. این فرودگاه تعهد محکمی نسبت به حفظ محیط زیست دارد و به همین دلیل مدیریت به دنبال یک راه‌حل کنترل محیط زیست بوده است. هدف از اینکار حذف کربن از فرودگاه بوده است. تعداد زیاد هواپیماها در آسمان و بر روی زمین و این واقعیت که هواپیماها بسیار جایجا می‌شوند نیازمند فناوری‌های پیشرفته‌ای برای راه‌حل مشکل بود.

راه‌حل

یک روش منطقی برای حل مشکل هواپیماها استفاده از اینترنت اشیا به همراه حسگرهای مبتنی بر هوش مصنوعی بود که امکان پایش، تحلیل و گزارش‌دهی محیط زیست را فراهم می‌کرد که تمامی اینها اطلاعات پس زمینه‌ای را برای تصمیمات مربوط به حداقل رساندن آلودگی هوا ارائه می‌کردند. دو شرکت ای‌اکس‌ام^۲ از یونان که متخصص در تحلیل‌های پیش‌بینی اینترنت اشیا و راه‌حل‌های اینترنت اشیا نوآورانه است و لیلیوم^۳ از ایالات متحده که متخصص در حسگرهای مبتنی بر هوش مصنوعی از جمله حسگرهای محیطی است تخصص‌شان را در هم آمیختند. هدف از این پروژه، پایش مناسب کیفیت هوا در داخل و خارج از فرودگاه و شناسایی به موقع و آنی مکان هواپیما بر روی زمین و اتخاذ اقدامات اصلاحی در صورت نیاز بود.

تحلیل و پایش آد هاک کیفیت هوا

فرودگاه در حال حاضر یک شبکه‌ی پایش کیفیت هوا دارد. این راه‌حل شامل پلتفرم حسگرهای بهم متصل لیلیوم به شکلی مقرون به صرفه است. حسگرها مختلف به سنجش دما، رطوبت، فشار اتمسفر، سطح اوزون و مواد ریز معلق در هوا می‌پردازند. خواننده‌های حسگرها برای گزارش‌گیری و تحلیل به اینترنت اشیا ارسال می‌شود. حسگرها با استفاده از ویژگی‌های هوش مصنوعی بهبود می‌یابند. از این رو، دقت آنها نیز افزایش می‌یابد. علاوه بر این، امنیت و مصرف انرژی نیز می‌تواند کنترل شود.

¹ Proximity sensors

² EXM

³ Libelium

موقعیت هواپیما در فرودگاه

برای شناسایی موقعیت دقیق هواپیما در زمان نشست و برخاست، در این پروژه از مکانیزم‌های اندازه‌گیری آکوستیک استفاده شد. اینکار با استفاده از حسگرهای نويز که در مکان‌های متفاوتی قرار گرفته‌اند انجام شد. حسگرها به اندازه‌گیری آبی سطح نويز می‌پردازند که توسط ابزارهای تحلیلی ارزشیابی می‌شوند. در مجموع، این سیستم یک راه‌حل اینترنت اشیاء غیرتہاجمی را ارائه می‌دهد.

جایگیری حسگرها به دلیل مسائل ایمنی، امنیتی و قوانین و مقررات کار دشواری بود. از این رو، زیرسیستم پایش صدا می‌بایست مستقل (خودکار) بوده و حامل پانل‌های خورشیدی و باتری‌هایی برای تأمین الکتریسیته باشد. علاوه بر این، این سیستم از یک سیستم ارتباطی بی‌سیم دوگانه به نام جی‌پی‌اس استفاده می‌کند. داده‌های نويز گردآوری شده با انواع هواپیما و پروازها در قسمت بک‌اند اینترنت اشیاء مرتبط می‌شوند. تمامی داده‌ها توسط دپارتمان محیط زیست فرودگاه تحلیل شده و برای تصمیمات بهبود و کنترل آلودگی بکار گرفته می‌شوند.

پشتیبانی فناوری

این راه‌حل ترکیبی از یک سیستم اینترنت اشیاء با تحلیل‌های مبتنی بر هوش مصنوعی، مصورسازی و گزارش‌گیری بوده و در فضای « ابر » اجرا می‌شود. علاوه بر این، این سیستم دارای حسگرهای در محل و زیرساخت‌های ارتباطی است. حسگرهای بی‌سیم کم توان به پایش میان مصرف آب و گاز در داخل فرودگاه و همچنین کیفیت هوا در محل‌های پارکینگ هواپیما می‌پردازند محصولات عرضه‌کنندگانی از قبیل میکروسافت آژور و آی‌بی‌ام بلومیکس^۱ از این پروژه پشتیبانی کرده و انعطاف‌پذیری لازم را فراهم می‌آورند.

Sources: Compiled from Hedge (2017) and Twentyman (2017).

سوالاتی برای مورد کاربردی ۱۳.۱

۱. نقش اینترنت اشیاء در این پروژه چیست؟
۲. نقش حسگرها چیست؟
۳. مزایای این پروژه چه هستند؟

در کاربردهای مقیاس بزرگ، حسگرها داده‌ها را گردآوری کرده و برای پردازش در فضای « ابر » ارسال می‌کنند. چنین پلتفرم برای این فرایند مورد استفاده قرار می‌گیرد که در مورد کاربردی ۱۳.۲ توضیح داده شده‌اند.

¹ IBM Bluemix

راکول اتومیشن^۱ یکی از بزرگ‌ترین ارائه‌دهندگان راه‌حل‌های اطلاعاتی و اتوماسیون صنعتی در جهان است که در بیش از ۸۰ کشور در سراسر جهان مشتری و حدود ۲۲۵۰۰ کارمند دارد. یکی از حوزه‌های تمرکز کسب‌وکار آن کمک به شرکت‌های نفت و گاز در امر اکتشاف است. یک مثال از آنها هیلکورپ انرژی^۲ است که در آلاسکا به حفاری نفت می‌پردازد. تجهیزات مورد استفاده در حفاری، استخراج و پالایش نفت بسیار گرانتقیمت هستند. وقوع یک نقص کوچک در تجهیزات می‌تواند حدود ۱۰۰،۰۰۰ تا ۳۰۰،۰۰۰ دلار در روز برای شرکت هزینه داشته باشد. برای مقابله با این مشکل، شرکت به یک فناوری برای پایش از راه دور قطعات این تجهیزات و پیش‌بینی خرابی‌های احتمالی در آینده نیاز داشت. راکول اتومیشن با گردآوری داده‌ها از محل‌های اکتشاف و تحلیل آنها برای بهبود تصمیم‌گیری‌های نگهداری پیشگیرانه در موارد بحرانی از این فرصت برای گسترش فعالیت‌های خود در صنایع نفت و گاز بهره گرفت که بدین ترتیب زمان خرابی به حداقل رسیده و عملکرد بهبود می‌یابد. شرکت از چشم‌انداز «کانکتد اینترپرایز»^۳ با نرم‌افزار میکروسافت برای پایش و پشتیبانی از تجهیزات نفت و گاز که در مناطق دوردست قرار گرفته‌اند استفاده می‌کند. راکول در حال حاضر راه‌حلی را برای پیش‌بینی خرابی تجهیزات در کل زنجیره تأمین نفت، پایش بلادرنگ سلامت و عملکرد و پیشگیری از خرابی‌ها در آینده بکار می‌گیرد. راه‌حل‌ها در حوزه‌های زیر ارائه شده‌اند:

- **حفاری.** هیلکورپ انرژی تجهیزات حفاری خود را در آلاسکا مستقر کرده است که ۲۴ ساعته به حفاری نفت مشغولند. یک خرابی در تجهیزات می‌تواند هزینه‌های هنگفتی برای هیلکورپ به همراه داشته باشد. هیلکورپ درایوهای متغیر الکتریکی تجهیزات حفاری را که قرار است در «ابر» پردازش شوند برای کنترل هزاران ماشین که مایل‌ها از اتاق کنترل از آنها فاصله دارند به یکدیگر متصل کرده است. حسگرها داده‌ها را ضبط کرده و از طریق درگاه کنترل راکول این داده‌ها به میکروسافت آژور کلاود ارسال می‌شوند. راه‌حل‌های دریافتی از طریق داشبوردهای دیجیتالی که اطلاعاتی در مورد دما، فشار، نرخ جریان و ده‌ها پارامتر دیگر که به مهندسان در رصد عملکرد و سلامت تجهیزات کمک می‌کنند برای مهندسان هیلکورپ ارسال می‌شوند. این داشبوردها همچنین هشدارهایی را در مورد هرگونه مسئله‌ای احتمالی نمایش می‌دهند. زمانیکه یکی از قطعات تجهیزات حفاری هیلکورپ خراب شود سریعاً شناسایی و در کمتر از یکساعت تعویض می‌شود و شش ساعت در زمان ردیابی خرابی و هزینه‌های تولید از دست رفته صرفه‌جویی می‌شود.
- **ساخت پمپ گازهای هوشمندتر:** امروزه برخی از کامیون‌ها از گاز طبیعی مایع (ال‌ان‌جی) به عنوان سوخت استفاده می‌کنند. شرکت‌های نفتی در حال ارتقاء جایگاه‌های سوخت‌گیری جهت پمپ‌های ال‌ان‌جی هستند. راکول اتومیشن حسگرها و درایوهای فرکانس متغیر را در این پمپ‌ها قرار داده تا بتواند داده‌های آنی در مورد عملیات تجهیزات، موجودی سوخت و نرخ مصرف را گردآوری کند. این داده‌ها برای پردازش به پلتفرم ابر راکول ارسال می‌شوند. سپس راکول با استفاده از میکروسافت آژور گزارشات و داشبوردهای تعاملی را تولید می‌کند. نتایج برای ذینفعان مربوطه ارسال شده و ایده‌ی خوبی از وضعیت سلامت دارایی‌های سرمایه‌ای آنها ارائه می‌کند.

¹ Rockwell Automation

² Hilcorp Energy

³ Connected Enterprise

راه‌حل کانکتد اینترپرایز شرکت راکول با ارائه‌ی داده‌های عملیات به پلتفرم ابر و کمک به آنها در کاهش خرابی‌ها و نگهداری و تعمیرات پرهزینه به رشد بسیاری از شرکت‌های نفت و گاز مثل هیلکورپ انرژی سرعت بخشید و منجر به یک فرصت کسب‌وکار جدید برای پیشکسوتان دوران صنعتی مثل راکول شد.

Sources: customers.microsoft.com (2015); Rockwell Automation: *Fueling the Oil and Gas Industry with IoT*; <https://customers.microsoft.com/Pages/CustomStory.aspx?recid=19922>; Microsoft.com. (n.d.). "Customer Stories | Rockwell Automation," <https://www.microsoft.com/en-us/cloud-platform/customer-stories-rockwell-automation> (accessed April 2018).

سوالاتی برای مورد کاربردی ۱۳.۲

۱. پلتفرم حفاری نفت و گاز چه نوع اطلاعاتی را گردآوری می‌کند؟
۲. آیا این کاربرد با سه V کلان‌داده‌ها (حجم، تنوع، سرعت) تطابق دارد؟
۳. چه صنایع دیگری می‌توانند از داشبوردها و اندازه‌گیری‌های عملیاتی مشابهی استفاده کنند؟

حسگرهای بسیار مختلفی وجود دارند. برخی از آنها دما و برخی دیگر رطوبت را اندازه‌گیری می‌کنند. حسگرهای بسیاری اطلاعات را گردآوری و ارسال می‌کنند. برای فهرست ۵۰ کاربرد حسگرها به همراه تعداد زیادی از مقالات مرتبط به libelium.com/resources/top_50-iot_sensor_applications_ranking مراجعه کنید.

یک نوع حسگر معروف که نقش بسیار مهمی را در اینترنت اشیا ایفا می‌کند شناسایی فرکانس رادیویی است.

بینش فناوری ۱۳.۱ حسگرهای شناسایی فرکانس رادیویی (آراف‌آی‌دی)

The first part of the paper discusses the importance of network security in the context of cloud computing. It highlights the challenges faced by organizations in securing their data and applications in a multi-tenant environment. The authors propose a comprehensive security framework that addresses various threats and vulnerabilities. This framework includes a combination of access control, encryption, and intrusion detection mechanisms. The second part of the paper focuses on the implementation of this security framework. It details the architectural components and the integration of various security technologies. The authors also discuss the performance and scalability of the proposed solution. The paper concludes with a summary of the findings and a call for further research in the field of cloud security.

The second part of the paper discusses the importance of network security in the context of cloud computing. It highlights the challenges faced by organizations in securing their data and applications in a multi-tenant environment. The authors propose a comprehensive security framework that addresses various threats and vulnerabilities. This framework includes a combination of access control, encryption, and intrusion detection mechanisms. The second part of the paper focuses on the implementation of this security framework. It details the architectural components and the integration of various security technologies. The authors also discuss the performance and scalability of the proposed solution. The paper concludes with a summary of the findings and a call for further research in the field of cloud security.

The authors also discuss the performance and scalability of the proposed solution. The paper concludes with a summary of the findings and a call for further research in the field of cloud security.

علاوه بر این، حسگرهای هوشمند برای اینترنت اشیا شامل نرم‌افزارهای ویژه برای تبدیل داده‌ها، پردازش دیجیتالی و قابلیت‌های ارتباط با دستگاه‌های بیرونی هستند.

طبق یک مطالعه مهم (بورکاکي و همکاران ۲۰۱۸)^۱، حسگرها در حال هوشمندتر شدن هستند. به عنوان آنهایی که بر روی وسایل نقلیه نصب می‌شوند. وسایل نقلیه می‌توانند از یک ماشین کاملاً سخت‌افزاری به یک دستگاه الکترونیکی نرم‌افزارمحور تبدیل شوند. نرم‌افزار می‌تواند ۳۵ درصد از هزینه‌های تولید وسیله نقلیه را به خود اختصاص دهد. برای اطلاعات بیشتر به اسکانل^۲ (۲۰۱۷)، جمیلی^۳ (۲۰۱۷) و تکناوویو^۴ (۲۰۱۷) نگاه کنید.

سوالاتی برای مرور بخش ۱۳.۵

۱. حسگر را تعریف کنید.
۲. نقش حسگرها در اینترنت اشیا را توضیح دهید.
۳. آراف‌آی‌دی چیست؟ یک حسگر آراف‌آی‌دی چیست؟
۴. آراف‌آی‌دی چه نقشی در اینترنت اشیا ایفا می‌کند؟
۵. حسگرهای هوشمند را تعریف کرده و نقش آنها را در اینترنت اشیا توضیح دهید.

sncf.com

¹ Burkacky et al.

² Scannell

³ Gemelli

⁴ Technavio

یک خانه‌ی هوشمند، خانه‌ای مجهز به اجزا و کامپوننت‌های خودکار است که به یکدیگر متصل هستند (عمدتاً به صورت بی‌سیم) مثل لوازم خانگی، چراغ‌ها، سیستم امنیتی و سرگرمی که به صورت مرکز کنترل شده و می‌توانند با یکدیگر ارتباط برقرار کنند. برای یک توضیح نگاه کنید به techterms.com/definition/smart_home.

خانه‌های هوشمند به گونه‌ای طراحی شده‌اند تا آسایش، امنیت، هزینه‌ی انرژی پایین و راحتی را برای ساکنان‌شان به ارمغان بیاورند. آنها می‌توانند از طریق گوشی‌های هوشمند یا اینترنت ارتباط برقرار کنند. آنها را می‌توان به صورت لحظه‌ای یا در بازه‌های زمانی مطلوب کنترل کرد. اکثر خانه‌های موجود هنوز هوشمند نشده‌اند ولی به آسانی و با هزینه بسیار پایین می‌توان آنها را تجهیز نمود تا از یک سطح حداقل از هوشمندی برخوردار شوند. برقراری اتصالات از طریق چندین پروتکل میسر می‌شود که معروف‌ترین آنها اکس‌آی‌ا، یوپی‌بی‌آ، زی-ویو^۳ و این‌اُشین^۴ هستند. این محصولات از مقیاس‌پذیری بالایی برخوردارند که بدین ترتیب می‌توان دستگاه‌های بیشتری را در طول زمان به خانه هوشمند متصل کرد. برای یک مرور نگاه کنید به smarthomeenergy.co.uk/what-smart-home و techterms.com/definition/smart_home و پیتسکر^۵ (۲۰۱۷).

در ایالات متحده که سایر کشورها نیز پیرو آن بوده‌اند، هزاران خانه به اینگونه سیستم‌های مجهز شده‌اند.

اجزای معمول خانه‌های هوشمند در ادامه آورده شده است:

Nestnest.com/workswith-nest

watercop.com

¹ XIO
² UPB
³ Z-Wave
⁴ EnOcean
⁵ Pitsker

•
•
tomsguide.com/us/huble-hugo-smarthome-camera.news-24240.html
•
•

اقدامات امنیتی در بسیاری از تسهیلات زندگی در جوامع سطح بالا و برای افراد سالمندی که به تنهایی زندگی می‌کنند بکار می‌روند. به عنوان مثال، سیستم نمایشگر آی‌هلت‌هوم تاج^۳ داده‌ها را گردآوری کرده و با استفاده از نرم‌افزار این شرکت با افراد مرتبط در شرکت ارتباط برقرار می‌کند. این سیستم دسترسی از راه‌دور به داده‌های سلامت بیماران را به پزشکان و مراقبان می‌دهد. برنامه آی‌هلت‌هوم با استفاده از این فناوری همچنین قرار ملاقات‌های روزانه و زمان مصرف داروها را به سالمندان یادآوری می‌کند. این سیستم همچنین زمان گرفتن فشار خون و چگونگی برقراری تماس با مراقبان را به این افراد یادآوری می‌کند.

یک لوازم خانگی هوشمند شامل ویژگی‌هایی است که می‌توانند عملیات لوازم خانگی را بر اساس اولویت‌های کاربر از راه دور کنترل کنند. یک لوازم خانگی هوشمند^۴ ممکن است از یک شبکه خانگی^۵ یا اینترنت برای برقراری ارتباط با سایر دستگاه‌ها در خانه هوشمند استفاده کند.

¹ Nest Cam Indoor

² Instaview

³ iHealthHome Touch

⁴ Smart Appliance

⁵ Home Network



شکل ۱۳.۴ اجزا و مؤلفه‌های یک خانه هوشمند

nest.com/thermostats/nest-

[learning-thermostat/overview/?alt=3](https://nest.com/thermostats/nest-learning-thermostat/overview/?alt=3)

nest.com/smoke-co-

[alarm/overview](https://nest.com/smoke-co-alarm/overview)

¹ McGrath

² Haier

[.nest.com/cameras/nest-cam-indoor/overview](https://www.nest.com/cameras/nest-cam-indoor/overview)

[.en.wikipedia.org/wiki/Nest_Labs](https://en.wikipedia.org/wiki/Nest_Labs)

دو کیت معروف برای خانه‌های هوشمند به شرح زیر هستند (پیتسکر ۲۰۱۷):

۱. **آمازون اکو**. این کیت شامل آمازون اکو، بلکین ومو مینی^۲، فیلیپس هیو وایت استارتر کیت^۳، اکوبی لایت^۴ و آمازون فایر تی وی استیک^۵ با ریموت صوتی الکسا می‌باشد که قیمت کل این مجموعه در اکتبر ۲۰۱۷ معادل ۴۹۵ دلار بوده است.
۲. **گوگل هوم**. این کیت شامل گوگل هوم، اسمارت اسپیکر، بلکین ومو مینی، فیلیپس هیو وایت استارتر کیت، ترموستات یادگیرنده نست و گوگل کروم کست^۶ (برای سرگرمی) می‌باشد. قیمت کل این مجموعه در اکتبر ۲۰۱۷ معادل ۵۲۰ دلار بوده است.

- **یخچال هوشمند سامسونگ**. دوربین‌ها محتوای یخچال را بررسی کرده، حسگرها نیز دما و رطوبت را اندازه‌گیری می‌کنند.
- **آشپز رباتی گورمت^۷**. آشپزی‌های جذابی را انجام می‌دهد.
- **دستگاه ۱۰ کاره آشپزخانه**. این دستگاه غذایی را درست کرده و دارای ۱۰ سبک آشپزی است (مثلاً پختن نان، تهیه‌ی سُس).

¹ Kastrenakes

² Belkin Wemo Mini

³ Philips Hue white starter kit

⁴ Ecobee Lite

⁵ Amazon Fire TV stick

⁶ Google Chromecast

⁷ Gourmet Robotic Cooker

- **ال جی هام بات توربو پلاس^۱**. این دستگاه می تواند بر یک بخش از خانه که نیازمند توجه ویژه است تمرکز کند. یک دوربین خانه را زمانیکه مالک در منزل نیست از راه دور رصد می کند (مشابه گوگل نِست).
- **یخچال آرتری دی تو هایر^۲**. طبق موریس (۲۰۱۶)، این یخچال ممکن است کاربردی ترین یخچال نباشد ولی ارزش سرگرمی زیادی دارد. این یخچال شبیه ربات آرتری دی تو در داستان جنگ ستارگان است که می تواند برایتان نوشیدنی سرو کرده و سیستم روشنایی و صوتی را راه اندازی کند.
- **یخچال اینستاویو ال جی^۳**. این یخچال که مجهز به الکسا می باشد شامل یک نمایشگر لمسی ال سی دی ۲۹ اینچی است که کارکردهایی از قبیل تعیین تاریخ انقضاء غذاها و اطلاع دادن به کاربر را ارائه می کند. برای جزئیات نگاه کنید به دیاز^۴ (۲۰۱۷).
- **شوینده ی حجم بالای هوشمند ویرپول^۵**. این ماشین کاملاً خودکار دارای کنترل های کوچکی است. این ماشین در مصرف انرژی صرفه جویی کرده و با ارسال مبلغی ناچیز به مؤسسه ی « زیستگاه بشریت » در هر نوبت شستشو، فعالیت های بشردوستانه نیز انجام می دهد.
- **ماشین ظرفشویی ال دی تی ۸۷۸۶ اس تی ال جی^۶**. این ماشینی مجهز به یک دوربین است که برای صرفه جویی در مصرف آب آنچه که قبلاً شسته شده است را ردیابی می کند. علاوه بر این، انعطاف پذیری بالایی در عملیات فراهم می کند.

برخی از روندهای خانه های هوشمند ار ادامه آورده شده است:

- تلویزیون ها می توانند به عنوان یک هاب یا شبکه هوشمند برای لوازم خانگی هوشمند سامسونگ عمل کنند.
- محصولات دالبی آتموس شامل اسپیکرها، گیرنده ها و سایر محصولات سرگرمی هستند.
- دوربین های امنیت خانگی هوشمند « خودت انجام بده » از ورود اشخاص غیرمجاز پیش از هشدار به پلیس اطمینان حاصل می کنند.
- کنترل آب برای شیرآلات، آب پاش ها و تشخیص دهنده سیلاب ها موجود هستند. علاوه بر این، یک ربات می تواند به کاربران آموزش دهد تا در مصرف آب خانه صرفه جویی کنند (hydro.com/us/en).

برای اطلاعات بیشتر در مورد اتوماسیون خانگی به smarthome.com/sh-learning-center-what-cani-control.html نگاه کنید. اپلیکیشن های متنوعی برای کنترل خانگی در smarthome.com/android_apps.html موجود است. کامپوننت های هوشمند برای خانه ها در فروشگاه های بهبود خانه ها (مثلاً لوز^۷) موجود بوده و می توان آنها را مستقیماً از سازندگان خریداری کرد (مثلاً نِست). آمازون و اینتل برای تسهیل در ساخت کامپوننت های هوشمند در سال

¹ LG HUM-BOT Turbo+

² Haier R3D2 Refrigerator

³ LG Instaview Refrigerator

⁴ Diaz

⁵ Whirlpool

⁶ LG LDT8786ST

⁷ Lowes

۲۰۱۷ با یکدیگر شراکت کرده و پلتفرم‌هایی را برای پیشبرد اکوسیستم خانه‌های هوشمند برای برنامه‌نویسان ارائه کردند. برای جزئیات نگاه کنید به pcmag.com/news/350055/amazon-intel-partner-to-advancesmart-home-tech/.

برای آگاهی از لوازم خانگی هوشمند در نمایشگاه سی‌ای‌اس ۲۰۱۸ ویدیوی [youtube.com/watch?v=NX-9LivJh0/](https://www.youtube.com/watch?v=NX-9LivJh0/) را تماشا کنید.

دستیار شخصی مجازی که در فصل ۱۲ معرفی کردیم کاربران را قادر می‌سازد تا از طریق صدا با چت‌بات‌هایی از قبیل الکسا/کو و دستیار گوگل ارتباط برقرار کنند. اینگونه دستیارها را می‌توان برای مدیریت لوازم خانگی در خانه‌های هوشمند بکار برد. در یک خانه‌ی هوشمند تمام عیار، دستگاه‌ها نه تنها نیازهای اولیه را برآورده می‌کنند بلکه قادر به پیش‌بینی آنها نیز هستند. پیش‌بینی می‌شود که در آینده‌ی نزدیک، یک خانه‌ی هوشمند مبتنی بر هوش مصنوعی از یک اکوسیستم هوشمند و هماهنگ با بات‌ها تشکیل شود که قادر به مدیریت و انجام وظایف اولیه و حتی برقراری رابطه‌ی عاطفی با افراد باشد. برای یک پیش‌بینی از بات‌های آینده نگاه کنید به کومو و همکاران^۱ (۲۰۱۷). آمازون و اینتل توان خود را برای توسعه‌ی چنین اکوسیستم‌های خانه‌های هوشمندی که شامل قابلیت‌های پردازش زبان طبیعی باشد ترکیب کرده‌اند. خانه‌های هوشمند همچنین دارای ربات‌های هوشمندی هستند که می‌توانند برای ساکنان خانه اسنک سرو کرده، از افراد معلول مراقبت کرده و حتی مهارت‌های دشوار را به کودکان آموزش دهند.

پتانسیل خانه‌های هوشمند بسیار جذاب است ولی زمان زیادی می‌خواهد تا تعداد بسیاری از این خانه‌های هوشمند بوجد آیند. برخی از موانع و محدودیت‌های این موضوع که توسط وانکاتاکریشنان^۲ (۲۰۱۷) مطرح شده در ادامه آورده شده است.

۱. **سازگاری.** محصولات و فروشندگان بسیار زیادی وجود دارند که انتخاب از بین آنها می‌تواند خریداران را دچار سردرگمی کند. بسیاری از این محصولات قادر نیستند تا با یکدیگر « صحبت » کنند و به همین دلیل به استانداردهای صنعتی بیشتری نیاز است. علاوه بر این، تطابق دادن محصولات با نیازهای مصرف‌کنندگان کار دشواری است.
۲. **ارتباطات.** مصرف‌کنندگان مختلف دارای ایده‌ها و نظرات مختلفی در مورد تعریف خانه‌های هوشمند هستند. از این رو، قابلیت‌ها و مزایای یک خانه هوشمند باید به طور واضح و شفاف برای کاربران توضیح داده شود.
۳. **تمرکز.** برندها باید بر بخش‌های مشتریان که علاقه‌مندتر به خانه‌های هوشمند هستند تمرکز کنند (مثلاً چن وای).

علاوه بر این مشکلات توجیه هزینه، تجاوز به حریم خصوصی، امنیت و سهولت استفاده نیز وجود دارند. برای اطلاع از آینده‌ی خانه‌های هوشمند از جمله نقش آمازون و وال‌مارت نگاه کنید به واینرایش^۳ (۲۰۱۸). خانه‌ها، لوازم خانگی و ساختمان‌های هوشمند را می‌توان در شهرهای هوشمند در نظر گرفت که موضوع بخش بعدی ما هستند.

¹ Coumau

² Vankatakrisnan

³ Weinreich

سوالاتی برای مرور بخش ۱۳.۷

۱. یک خانه‌ی هوشمند را توضیح دهید.
۲. مزایای یک خانه‌ی هوشمند چه هستند؟
۳. لوازم خانگی هوشمند مهم را نام ببرید.
۴. نحوه‌ی کارکرد نِسْت را توضیح دهید.
۵. نقش بات‌ها در خانه‌های هوشمند را توضیح دهید.

ایده‌ی شهرهای هوشمند حدود سال ۲۰۰۷ و زمانیکه آی‌بی‌ام پروژه‌ی سیاره‌ی هوشمند خود را راه‌اندازی کرد و سپسکو برنامه‌ی شهرها و جوامع هوشمند خود را آغاز کرد شکل گرفت. ایده این است که در شهرهای هوشمند، فناوری‌های دیجیتالی (عمدتاً موبایل) موجب تسهیل خدمات عمومی برای شهروندان، بهره‌برداری بهتر از منابع و تأثیرات زیست‌محیطی منفی کمتر شوند. تاونزاند^۱ (۲۰۱۳) یک نگاه تاریخی گسترده به این فناوری‌ها داشته است. وی در مروری در کتابش مثال‌های زیر را ارائه کرده است: «در زاراگوزا اسپانیا، شما با استفاده از یک کارت شهروندی می‌توانید دسترسی رایگان به شبکه وای‌فای در سطح شهر داشته باشید، از کتابخانه کتاب قرض بگیرید، دوچرخه سوار شوید و بلیت اتوبوس تهیه کنید. در نیویورک، یک گروه از دانشمندان شهروند حسگرهایی را در فاضلاب‌های محلی قرار نصب کرده‌اند تا از وقوع گرفتگی معابر و آبراه‌ها مطلع شده و هشدار دهند.» طبق پیش‌بینی ادیتورز (۲۰۱۵)، شهرهای هوشمند در سال ۲۰۱۶ از ۱.۶ میلیارد شیء متصل استفاده خواهند کرد. نهایتاً، شهرهای هوشمند می‌توانند به چندین موجودیت هوشمند از قبیل دانشگاه‌ها و کارخانجات متصل شوند (نگاه کنید به لیسی ۲۰۱۶). برای اطلاعات بیشتر در مورد شهرهای هوشمند نگاه کنید به شوارتز (۲۰۱۵). علاوه بر این، ویدئوی «Cisco Bets Big on 'Smart Cities'» را در آدرس money.cnn.com/video/technology/2016/03/21/cisco-ceo-smart-cities.cnnmoney تماشا کنید. یک ویدئوی دیگر به نام «Smart Cities of the Future» در youtube.com/watch?v=mQR8hxMP6SY موجود است. یک ویدئوی کامل‌تر در youtube.com/watch?v=LJznAJe5uQ قابل مشاهده است.

شهرها یک شبه هوشمند نمی‌شوند که این موضوع در مورد کاربردی ۱۳.۳ به تصویر کشیده شده است که شهر آمستردام و رشد و تکامل آن به یک شهر هوشمند را نشان می‌دهد. در بسیاری از کشورها، دولت‌ها در حال توسعه‌ی کاربردهای شهر هوشمند هستند. به عنوان مثال، هند شروع به توسعه‌ی ۱۰۰ شهر هوشمند کرده است (-enterpriseinnovation.net/article/india-eyes-development-100-smart-cities-) (1301232910).

آمستردام در مسیر تبدیل شدن به یک شهر هوشمند

مورد کاربردی ۱۳.۳

¹ Townsend

در عرض هفت سال، شهر آمستردام (هلند) با استفاده از فناوری‌های اطلاعات به یک شهر هوشمند تبدیل شد. این مورد کاربردی طبق گزارش دانشکده‌ی مدیریت اسلوان ام‌آی‌تی به توضیح گام‌هایی که بین سال‌های ۲۰۰۹ تا ۲۰۱۶ برای تبدیل شدن به یک شهر هوشمند برداشته شد می‌پردازد. ابتکار عمل این شهر هوشمند شامل پروژه‌های زیر بوده است: حمل و نقل، کیفیت زندگی، جابجایی، امنیت، سلامت و اقتصاد و همچنین زیرساخت‌ها، منبع داده‌های کلان و باز و آزمایشگاه‌های زندگی آزمایشی. یافته‌های مهم تیم ام‌آی‌تی در رابطه با تغییر شکل آمستردام به شرح زیر هستند:

- **داده‌های بخش خصوصی برای تغییر سیاست‌ها بسیار حیاتی هستند.** در این پروژه گروه‌های اصلی متشکل از موجودیت‌های غیردولتی (مثلاً با استفاده از یک جی‌پی‌اس برای مدیریت ترافیک) بوده‌اند. به عنوان مثال، بخش خصوصی در این پروژه برای تغییر وضعیت ترافیک حضور داشته است (کاهش ۲۵ درصدی تعداد خودروها و افزایش ۱۰۰ درصدی تعداد اسکوترها در عرض پنج سال).
 - **داشتن افسران ارشد فناوری در شهرهای هوشمند ضروری است.** شهرهای هوشمند نیازمند گردآوری حجم انبوهی از داده‌ها با استفاده از ابزارها و الگوریتم‌های متعدد هستند. مسائلی از قبیل هزینه و امنیت در اینجا حیاتی هستند.
 - **انتظارات از نقش اینترنت اشیا، کلان‌داده‌ها و هوش مصنوعی برای مدیریت شود.** شهروندان توقع تغییرات و بهبود سریع را در حوزه‌هایی از قبیل پارکینگ تا ترافیک دارند. گردآوری داده‌ها بسیار کند و اجرای تغییرات بسیار دشوار است.
 - **ابتکار عمل‌های شهر هوشمند باید با موجودی داده آغاز شوند.** مشکل اصلی در آمستردام این بود که داده‌ها در ۱۲،۰۰۰ پایگاه داده در ۳۲ دپارتمان ذخیره شده بودند. اینها به روش‌های متفاوتی بر روی سخت‌افزارهای متفاوتی سازماندهی شده بودند و به همین دلیل به موجودی داده‌ها نیاز بود. این فعالیت اولیه بسیار خسته‌کننده و دشوار بود و خروجی مشهودی در ابتدای کار نداشت.
 - **پروژه‌های پایلوت یک استراتژی عالی هستند.** پروژه‌های پایلوت درس‌هایی را برای پروژه‌های آینده فراهم می‌کنند. این شهر بیش از ۸۰ پروژه‌ی پایلوت داشت. به عنوان مثال، جمع‌آوری انواع زباله‌ها و قرار دادن آنها در کیسه‌های رنگی. پروژه‌های موفق در مقیاس بزرگ‌تر ادامه یافتند.
 - **ورودی شهروندان یک عامل موفقیت کلیدی است.** روش‌های متعددی برای ترغیب شهروندان به ارائه‌ی ورودی وجود دارد. مشارکت دانشگاه‌ها و مؤسسات پژوهشی نیز حیاتی است. علاوه بر این، از شبکه‌های اجتماعی می‌توان برای تسهیل مشارکت و درگیرسازی شهروندان استفاده کرد.
- ابتکار عمل شهر هوشمند ممکن است در مراحل اولیه خود باشد ولی تا به همین جا کیفیت زندگی شهروندان را بهبود بخشیده و موجب افزایش رشد اقتصادی شهر شده است. یک عامل موفقیت بحرانی تمایل مقامات شهر برای به اشتراک‌گذاری داده‌ها با شرکت‌های فناوری بوده است.
- اینترنت اشیا یک مؤلفه‌ی مهم در پروژه‌ها بوده است. ابتدا، جریان داده‌ها از حسگرها و پایگاه‌های داده را برای پردازش تحلیلی میسر می‌کرد. دوم، اینترنت اشیا و وسائل نقلیه‌ی خودران را که منجر به کاهش آلودگی، تصافات و ازدحام ترافیک میشد امکان‌پذیر ساخت. نهایتاً، اینترنت اشیا داده‌های بلادرنگ را برای کمک به تصمیم‌گیرندگان برای توسعه و بهبود سیاست‌ها فراهم می‌کرد. در آوریل ۲۰۱۶، این شهر برنده‌ی جایزه‌ی « سرمایه‌ی نوآوری » اروپا شد (جایزه‌ای به مبلغ ۹۵۰،۰۰۰ یورو).

سوالاتی برای مورد کاربردی ۱۳.۳

۵. ویدیوی [youtube.com/watch?v=FinLi65Xtik](https://www.youtube.com/watch?v=FinLi65Xtik) را تماشا کرده و در مورد فناوری‌های بکار رفته اظهار نظر کنید.
۶. یک نسخه از مطالعه‌ی موردی ام‌آی‌تی را از sloanreview.mit.edu/case-study/data-driven-citymanagement دریافت کرده و گام‌های فرایند و کاربردهای اینترنت اشیا را نام ببرید.
۷. اجزای هوشمند بکار رفته در این پروژه را شناسایی کنید.

در این بخش، شما به بررسی و تحلیل موارد کاربردی در زمینه اینترنت اشیا (IoT) می‌پردازید. سوالات مطرح شده در بالا، شما را به درک عمیق‌تری از فناوری‌های بکار رفته در این پروژه‌ها و همچنین فرایندهای پیاده‌سازی و کاربردهای اینترنت اشیا در محیط‌های واقعی سوق می‌دهد. این مطالعه‌ها و پروژه‌ها، نمونه‌های برجسته‌ای از نحوه‌ی استفاده از داده‌ها و فناوری‌های نوین برای حل مشکلات واقعی و بهبود فرایندهای سازمانی و شهری هستند.

- در مورد چالش‌های امنیتی و حریم خصوصی در استفاده از داده‌های اینترنت اشیا، چه نظریاتی دارید؟
- چگونه می‌توانیم از داده‌های اینترنت اشیا برای بهبود فرایندهای عملیاتی و تصمیم‌گیری‌ها استفاده کنیم؟
- در مورد نقش داده‌های اینترنت اشیا در توسعه‌ی شهرهای هوشمند، چه دیدگاهی دارید؟
- چگونه می‌توانیم از داده‌های اینترنت اشیا برای بهبود فرایندهای عملیاتی و تصمیم‌گیری‌ها استفاده کنیم؟
- در مورد نقش داده‌های اینترنت اشیا در توسعه‌ی شهرهای هوشمند، چه دیدگاهی دارید؟
- چگونه می‌توانیم از داده‌های اینترنت اشیا برای بهبود فرایندهای عملیاتی و تصمیم‌گیری‌ها استفاده کنیم؟
- در مورد نقش داده‌های اینترنت اشیا در توسعه‌ی شهرهای هوشمند، چه دیدگاهی دارید؟

این سوالات، شما را به تفکر عمیق‌تری در مورد کاربردهای اینترنت اشیا و چالش‌های مرتبط با آن سوق می‌دهد. امید است که با بررسی و تحلیل موارد ذکر شده، به درک عمیق‌تری از این فناوری و نحوه‌ی استفاده از آن در محیط‌های واقعی دست یابید.

در این بخش، شما به بررسی و تحلیل موارد کاربردی در زمینه اینترنت اشیا (IoT) می‌پردازید. سوالات مطرح شده در بالا، شما را به درک عمیق‌تری از فناوری‌های بکار رفته در این پروژه‌ها و همچنین فرایندهای پیاده‌سازی و کاربردهای اینترنت اشیا در محیط‌های واقعی سوق می‌دهد. این مطالعه‌ها و پروژه‌ها، نمونه‌های برجسته‌ای از نحوه‌ی استفاده از داده‌ها و فناوری‌های نوین برای حل مشکلات واقعی و بهبود فرایندهای سازمانی و شهری هستند.

هنگ کنگ پروژه‌ای به نام جابجایی هوشمند را برای بهبود ایمنی راه‌ها و جاده‌ها آغاز کرد. یک کنسرسیوم از سازمان‌های دولتی و خصوصی خدمات حمل و نقل هوشمند را معرفی کردند که شامل یک مکانیسم هشدار برای تصادفات و دستیار کنترل برای یافتن پارکینگ بود. این سیستم همچنین تجاوز از محدودیت سرعت و خطوط اتوبان‌ها و ازدحام ترافیک را مدیریت می‌کرد. تمامی اینها موجب افزایش ایمنی و کارایی شدند. برای جزئیات نگاه کنید به استوپاس^۱ (۲۰۱۷).

حمل و نقل یکی از حوزه‌های مهم است که در آن، تحلیل‌ها و هوش مصنوعی می‌توانند موجب هوشمندتر شدن شهرها شوند. از حوزه‌های دیگر می‌توان به توسعه‌ی اقتصادی، مبارزه با جرائم و مراقبت سلامت اشاره کرد. برای جزئیات نگاه کنید به ساس^۲ (۲۰۱۷). مثال‌های دیگر از اجزای شهرهای هوشمند را می‌توانید در یک دانشگاه هوشمند، مراکز پزشکی هوشمند، فرودگاه‌ها، کارخانجات، بنادر، استادیوم‌های ورزشی و کارخانجات هوشمند پیدا کنید. هرکدام از این اجزا را می‌توان به صورت یک پروژه‌ی اینترنت اشیا مستقل و/یا بخشی از یک پروژه‌ی شهر هوشمند در نظر گرفت.



شکل ۱۳.۵ چارچوب بلوغ ساختمان شناختی آی‌بی‌ام

Source: IBM. "Embracing the Internet of Things in the new era of cognitive buildings." IBM Global Business Services, White Paper, 2016. Courtesy of International Business Machines Corporation, © International Business Machines Corporation. Used with permission.

¹ Estopace

² SAS

.DUP_The-smart-factory.pdf

<https://www2.deloitte.com/insights/us/en/focus/internet-of-things/technical-primer.html>

توماس (۲۰۱۶) چشم‌اندازی از تولید صنعتی در آینده‌ای نزدیک ارائه می‌کند. تولید صنعتی یک نظام کاملاً دیجیتالی و متصل، سریع و انعطاف‌پذیر خواهد بود. موضوع اصلی در تولید صنعتی وجود یک مرکز فرماندهی مجهز به فناوری‌های هوش مصنوعی در کارخانه است. هوش مصنوعی به همراه حسگرهای اینترنت اشیا و جریان اطلاعات امکان سازماندهی و توالی‌سازی بهینه‌ی فرایندهای کسب‌وکار را فراهم می‌آورد. کل زنجیره‌ی تولید از تأمین‌کنندگان مواد اولیه، لجستیک و ساخت و تولید گرفته تا فروش برای برنامه‌ریزی، هماهنگی و کنترل به سیستم‌های اینترنت اشیا متصل خواهند شد. برنامه‌ریزی بر اساس پیش‌بینی‌های تحلیلی تقاضا انجام می‌شود. فرایندهای تولید تا حد امکان به صورت خودکار انجام شده و به صورت بی‌سیم کنترل می‌شوند. لجستیک سریع‌اً طبق تقاضا ارائه شده و کنترل کیفیت نیز به صورت خودکار انجام می‌شود. اینترنت اشیا به همراه حسگرها برای نگهداری و تعمیرات پیش‌بینانه و پیشگیرانه بکار می‌رود. برخی از این اجزا در کارخانجات پیشرفته وجود دارند و کارخانجات بیشتری در آینده هوشمندتر خواهند شد.

برای اطلاعات بیشتر در مورد کارخانجات هوشمند نگاه کنید به لیبلوم (۲۰۱۵) و پوجاری^۱ (۲۰۱۷). برای کارخانجات هوشمند آینده مقاله‌ی belden.com/blog/industrial-ethernet/topic/smartfactory-of-the-future/page/0 را مطالعه کنید.

کاربرد اینترنت اشیا در کارخانجات در ویدئوی «Smart Factory Towards a Factory of Things» در youtube.com/watch?v=EUnnKAFcpuE نشان داده شده است. کارخانجات هوشمند دارای فرایندهای کسب‌وکار متفاوت، راه‌حل‌های فناوری جدید، تعاملات متفاوت بین انسان و ماشین و یک فرهنگ اصلاح شده هستند. برای آگاهی از فرایند تغییر و تکامل به سمت یک کارخانه‌ی هوشمند به باپکار و دیاس^۲ (۲۰۱۷) نگاه کنید. مؤسسه‌ی حسابداری دلویت (dupress.deloitte.com/smart-factory) یک دیاگرام از ویژگی‌های مهم یک کارخانه‌ی هوشمند را ارائه کرده است (شکل ۱۳.۶).

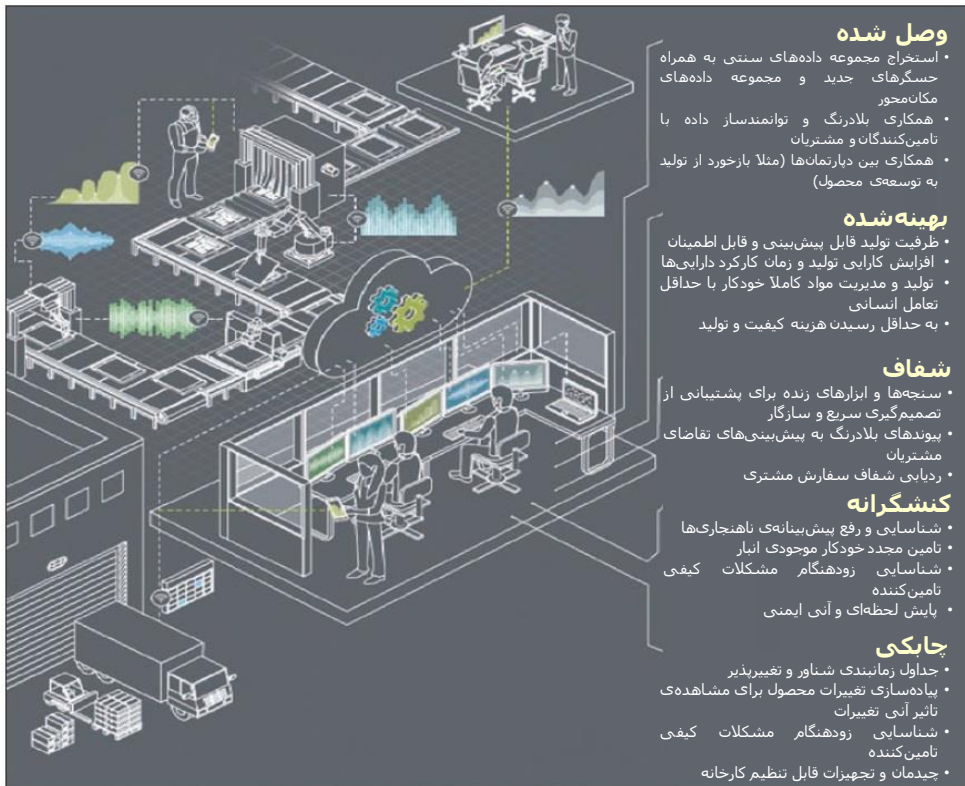
تقاضای جهانی برای دوچرخه‌های هوشمند به ویژه در شهرهای هوشمند به سرعت در حال افزایش است. موبایک اولین و بزرگ‌ترین شرکت به اشتراک‌گذاری دوچرخه است. این شرکت برای پاسخگویی به تقاضا با گروه فناوری فاکس‌کان^۳ برای ساخت هوشمندتر دوچرخه‌ها همکاری می‌کند. ساخت و تولید هوشمند شامل ایجاد یک زنجیره‌ی تأمین بین‌المللی از مواد اولیه تا تولید و فروش است. فاکس‌کان برای خاطر تخصص در فناوری‌های پیشرفته در ارائه‌ی فرایندهای ساخت و تولید کارآمد و مقرون به صرفه زبان‌زد است. این شرکت به بهینه‌سازی ساخت و تولید هوشمند مبتنی بر اینترنت می‌پردازد. خروجی تولید انتظار می‌رود که در آینده‌ی نزدیک دو برابر شود. برای جزئیات نگاه کنید به هامبلن (۲۰۱۶) و

enterpriseinnovation.net/article/foxconn-drives-mobike-smart-bike-production-1513651539

¹ Pujari

² Bhapkar & Dias

³ Foxconn Technology Group



شکل ۱۳.۶ پنج ویژگی کلیدی یک کارخانه‌ی هوشمند

Source: Burke, Hartigan, Laaper, Martin, Mussomeli, Sniderman, "The smart factory: Responsive, adaptive, connected manufacturing," Deloitte Insights (2017), <https://www.deloitte.com/insights/us/en/focus/industry-4-0/smart-factory-connected-manufacturing.html>. Used with permission.

آی‌بی‌ام چگونه شهرهای هوشمندتر را در سرتاسر جهان می‌سازد

مورد کاربردی ۱۳.۴

آی‌بی‌ام برای چندین سال از ابتکار عمل‌های شهرهای هوشمند حمایت و پشتیبانی کرده است. مثال‌های زیر از اسلاید شو تفت^۱ (eweek.com/cloud/how-ibm-is-making-cities-smarter-worldwide) استخراج شده‌اند.

- **مینیاپولیس (ایالات متحده)**. این ابتکار عمل از تصمیمات مؤثرتر برای تخصیص منابع شهری پشتیبانی می‌کند. علاوه بر این، عملیات چندین دپارتمان که بر روی یک پروژه کار می‌کنند را همراستا می‌کند. آی‌بی‌ام الگوریتم‌های تشخیص الگو برای مسئله و بهبود عملکرد را ارائه می‌دهد.

¹ Taft

- **مونپلیه (فرانسه).** نرم افزار آی بی ام به ابتکار عمل این شهر در مدیریت آب ها، حمل و نقل و مدیریت ریسک (تصمیم گیری) کمک می کند. رشد سریع این شهر باید پاسخگوی تقاضای روزافزون خدمات باشد. آی بی ام برای این منظور تحلیل و تفسیر فعالیت ها، مؤسسات پژوهشی و سایر شرکا را در منطقه در اختیار آنان قرار داده است.
 - **استکهلم (سوئد).** برای کاهش مشکلات ترافیک، فناوری های آی بی ام به تطبیق بهینه ی تقاضا و عرضه می پردازند. این ابتکار عمل از حسگرها و اینترنت اشیا برای کاهش مشکل ترافیک استفاده می کند.
 - **دوبوک (ایالات متحده).** چندین ابتکار عمل برای استفاده ی کارآمد از منابع و مدیریت مشکلات حمل و نقل اجرا شده است.
 - **کمبریج (کانادا).** این شهر از « برنامه ریزی زیرساخت هوشمند » آی بی ام برای اجرای فناوری های پشتیبان تصمیم و تحلیل کسب و کار استفاده می کند. این شهر با استفاده از الگوریتم های مبتنی بر هوش مصنوعی می تواند تصمیمات بهتری را اتخاذ کند (مثلاً تعمیر یا تعویض دارایی ها). علاوه بر این، فناوری های هوشمند آی بی ام به بهبود هماهنگی پروژه کمک می کنند.
 - **لیون (فرانسه).** مدیریت حمل و نقل یکی از پروژه های مهم در هر شهر بزرگ و یک هدف برای اکثر ابتکار عمل های شهرهای هوشمند محسوب می شود. فناوری های هوشمند می توانند ابزارهای پشتیبان تصمیم بلادرنگ و مؤثری را در اختیار پرسنل حمل و نقل قرار دهند. اینکار به کاهش ازدحام ترافیک کمک کرده است. با استفاده از تحلیل های پیش بینانه می توان مشکلات را پیش بینی کرده و در صورت وقوع آنها را سریعاً رفع نمود.
 - **ریودو ژانیرو (برزیل).** مدیریت و هماهنگی عملیات ۳۰ دپارتمان شهری کار بسیار پیچیده ای است. فناوری های آی بی ام از یک مرکز فرماندهی مرکزی برای برنامه ریزی عملیات و مدیریت شرایط اضطراری در تمامی حوزه ها پشتیبانی می کنند.
 - **مادرید (اسپانیا).** برای مدیریت کلیه ی موقعیت های اضطراری (آتش نشانی، حمل و نقل پلیس، بیمارستان ها) این شهر یک مرکز پاسخگویی مرکزی را ایجاد کرده است. داده ها توسط حسگرها، جی پی اس، دوربین های کنترل و ... گردآوری می شوند. این مرکز پس از حمله ی تروریستی ۲۰۰۴ مادرید ایجاد شد و با پشتیبانی فناوری های هوشمند آی بی ام مدیریت می شود.
 - **روچستر (ایالات متحده).** دپارتمان پلیس شهر روچستر از اینترنت اشیا و تحلیل پیش بینانه برای پیش بینی زمان و محل وقوع احتمالی جرائم استفاده می کند. این سیستم مبتنی بر هوش مصنوعی در چندین شهر دیگر نیز از دقت بالایی برخوردار بوده است.
- این مثال ها بهره برداری از چارچوب شهرهای هوشمند آی بی ام در حوزه های مختلف را با استفاده از ابتکار عمل های شهر هوشمند نشان می دهند. توجه داشته باشید که آی بی ام واتسون از اینترنت اشیا در بسیاری از پروژه های خود استفاده می کند.

سوالاتی برای مورد کاربردی ۱۳.۴

۱. خدماتی مختلفی که توسط اینترنت اشیا در یک شهر هوشمند بهبود یافته اند را نام ببرید.
۲. این فناوری ها چگونه از تصمیم گیری پشتیبانی می کنند؟
۳. در مورد ماهیت جهانی این مثال ها اظهار نظر کنید.

یک حوزه ی مهم بهبود در شهرهای هوشمند، حمل و نقل می باشد.

10. 10. 2019. 10. 10. 2019.

11. 11. 2019. 11. 11. 2019.

12. 12. 2019.

13. 13. 2019.

14. 14. 2019.

15. 15. 2019.

16. 16. 2019.

17. 17. 2019.

18. 18. 2019.

19. 19. 2019.

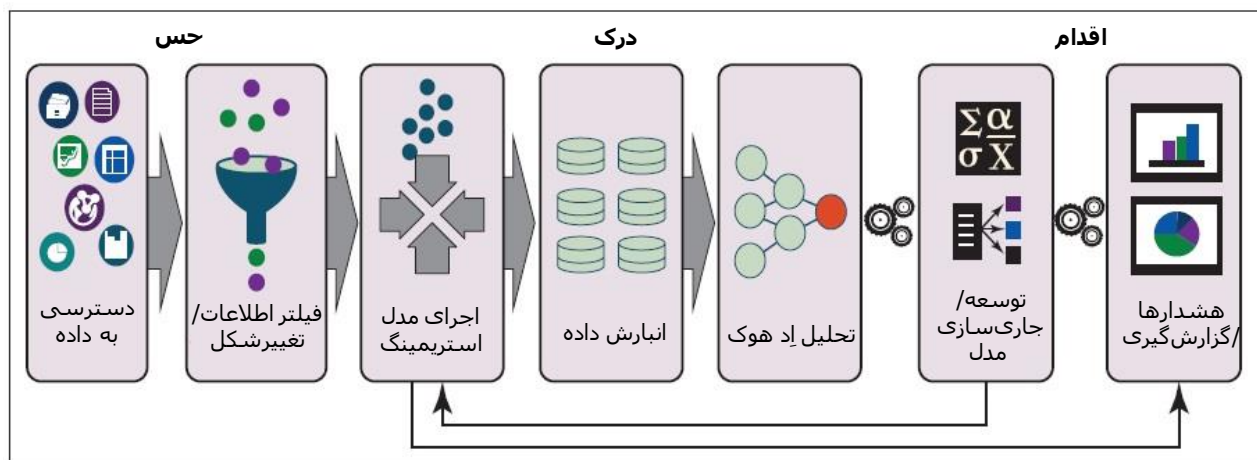
20. 20. 2019.

21. 21. 2019.

... ..

-
-
-

» https://www.sas.com/en_us/insights/big-data/internetof-things.html « SAS Analytics for IoT



شکل ۱۳.۷ اساس از چرخه‌ی عمر کامل ابزارهای تحلیلی اینترنت اشیا برای شهرهای هوشمند پشتیبانی می‌کند

Source: Courtesy of SAS Institute Inc. Used with permission.

در پایان، در شهرها هوشمند وسائل نقلیه‌ی خودران و متصل در همه جا تردد خواهند کرد (هامیلن، ۲۰۱۶ و بخش بعدی).

سوالاتی برای مرور بخش ۱۳.۸

۱. شهر هوشمند را توضیح دهید.
۲. برخی مزایای یک شهر هوشمند را برای شهروندان نام ببرید.
۳. نقش اینترنت اشیا در ابتکار عمل‌های شهر هوشمند چیست؟
۴. ابزارهای تحلیل چگونه با اینترنت اشیا ترکیب می‌شوند و چرا؟
۵. ساختمان‌های هوشمند و شناختی را توضیح دهید.
۶. یک کارخانه‌ی هوشمند چیست؟
۷. پشتیبانی فناوری در شهرهای هوشمند را توضیح دهید.

وسائل نقلیه‌ی مستقل که با نام خودروهای خودران، بدون راننده و خودروهای مستقل نیز شناخته شده‌اند در چندین مکان در جاده‌ها تردد می‌کنند. اولین پروژه‌ی تجاری خودروهای خودران توسط گوگل آغاز شد (به نام شوfer گوگل) و در چندین ایالت آمریکا که آماده‌ی استفاده از آنها هستند به واقعیت می‌پیوندد. فرانسه، سنگاپور، چین و چندین کشور دیگر نیز از این خودروها و اتوبوس‌ها در جاده‌هایشان استفاده می‌کنند. این خودروها برقی بوده و می‌توانند با کاهش گازهای منتشر شده، تصادفات، مرگ و میرها (تخمین سالانه ۳۰،۰۰۰ نفر تلفات در جهان) و ازدحام ترافیک انقلابی را بوجود آورند (توکونکا ۲۰۱۶). تا بدینجا، این خودروها در چندین شهر در سرتاسر جهان آزمایش شده و در برخی شهرها نیز در جاده‌ها تردد می‌کنند. کارشناسان برآورد کرده‌اند که تا سال ۲۰۲۰ حدود ۱۰ میلیون خودرو در جاده‌های ایالات متحده تردد خواهند کرد و چین نیز برنامه‌ریزی برای ۳۰ میلیون خودرو را تا سال ۲۰۲۱ در دستور کار خود دارد.

وی‌مو^۱ یک واحد از آلفابت^۲ است که به پروژه‌ی توسعه‌ی خودروهای خودران گوگل اختصاص دارد. گوگل تقریباً ۲۰ سال پیش به کمک دانشگاه استنفورد کار بر روی این پروژه را آغاز کرد. این ایده در سال ۲۰۰۵ و زمانیکه این پروژه جایزه چالش بزرگ دارپا^۳ را دریافت کرد سرعت گرفت. سپس وزارت دفاع ایالات متحده جایزه‌ی ۲ میلیون دلاری را به این پروژه اعطا کرد. گوگل در سال ۲۰۰۹ و پس از انجام شبیه‌سازی رایانه‌ای به مدت چندین سال و طی مسافت مجازی ۲.۵ میلیارد مایل برای خودروهای خودران در آزمایشات واقعی این خودروها پیشگام شد. گام بعدی این بود که قوانین و مقررات برای استفاده از خودروهای خودران در جاده‌ها را اخذ کند. تا سال ۲۰۱۸، این قوانین در ۱۰ ایالت به تصویب رسید. برخی از این ایالت اجازه‌ی تردد این خودروها را فقط در برخی مناطق خاص داده بودند. خودروهای خودران با رانندگان ربات در اوایل ۲۰۱۸ توسط وی‌مو در منطقه‌ی فینکس آریزونا تست شدند. ابتدا مهندسان سازمان در صندلی راننده می‌نشستند ولی در نوامبر ۲۰۱۸ خودروها به صورت کاملاً مستقل و بدون راننده تست شدند. شرکت آماده بود تا مینی‌ون‌های تجاری خود را در پنج ایالت در سال ۲۰۱۸ به بازار عرضه کند. تا پایان سال ۲۰۱۸، ون‌های وی‌مو مسافران داوطلبی را که خواهان دریافت خدمات (برنامه‌ی اِری‌رایدر^۴) بودند سوار می‌کرد اگرچه اکثر مسافران همچنان به دیده‌ی تردید به آن نگاه می‌کردند.

تکنیسین‌های شرکت همانند مسافران عادی خدمت مورد نظرشان را از طریق اپلیکیشن موبایل درخواست می‌دادند. یک مکانسیم هوش مصنوعی متوجه می‌شد که خودرو چگونه برای درخواست‌کننده تخصیص یافته و همچنین چگونه به صورت خودران به مقصد مورد نظر رهسپار می‌شود. وی‌مو که پیشگام خودروهای خودران است با شرکت کرایسلر^۵ همکاری نمود. توان رایانشی این خودروها توسط شرکت اینتل فراهم شد. هزینه‌ی بالای خودروها موجب محدودیت کاربردهای تجاری این خودروها در ابتدای امر شد. با وجود این، وی‌مو توافق کرد تا ناوگان مینی‌ون‌های خودران آویس^۶ را مدیریت کند. وی‌مو با لیفت^۷ بر روی وسائل نقلیه‌ی خودران جدید همکاری دارد. نهایتاً، وی‌مو برای ارائه خدمات نگهداری و تعمیرات و خدمات جاده‌ای خودروهای وی‌مو با اتونیشن^۸ شراکت کرده است. نکته: در خصوص مباحث قانونی از جمله اوبر به پرونده‌ی آغازین فصل ۱۴ نگاه کنید.

Sources: Compiled from Hawkins (2017), Ohnsman (2017), and Khoury (2018).

¹ Waymo

² Alphabet

³ DARPA

⁴ Early Rider Program

⁵ Chrysler

⁶ Avis

⁷ Lyft

⁸ AutoNation

سوالاتی برای مورد کاربردی ۱۳.۵

۱. چرا وی‌مو ابتدا از شبیه‌سازی استفاده کرد؟
۲. چرا به قوانین و مقررات نیاز است؟
۳. برنامه‌ریزی را چه بود؟
۴. چرا زمان زیادی طول میکشد تا مالکان خودروها بتوانند سوار خودروهای خودران شوند؟
۵. چرا لیفت، اوبر و آویس به خودروهای خودران علاقمند هستند؟



شکل ۱۳.۸ خودروی خودران وی‌مو (گوگل)

Source: SiliconValleyStock/Alamy Stock Photo.

یک مثال از نحوه همکاری انویدیا با ابتکار عمل تویوتا در بینش فناوری ۱۳.۲ ارائه شده است.

بینش فناوری ۱۳.۲ تویوتا و شرکت انویدیا برای تولید انبوه خودروهای خودران همکاری می‌کنند.

جای تعجب ندارد که تویوتا علاقمند به خودروهای هوشمند باشد. در واقع، انتظار می‌رود که خودروهای این شرکت تا سال ۲۰۲۰ وارد بازار شوند. تویوتا برای تولید چندین نوع وسیله نقلیه مستقل برنامه‌ریزی کرده است. یک نوع از آن برای سالمندان و افراد معلول است. نوع دیگر آن توانایی رانندگی کاملاً مستقل یا به عنوان دستیار (توسط مکانیسمی به نام «فرشته‌ی نگهبان»^۱) برای رانندگان را دارد. به عنوان مثال، توانایی این را خواهد داشت تا زمانیکه راننده به خواب میرود یا زمانیکه احساس می‌کند که تصادفی ممکن است رخ بدهد کنترل کامل خودرو را به دست بگیرد. راننده‌ی خسته می‌تواند از الکسا (یا دستگاهی مشابه آن) استفاده کرده و به فرشته‌ی نگهبان بگوید که ماشین را نگهدارد. وسائل نقلیه‌ی خودران به یک سیستم کنترلی هوشمند نیاز دارند و این جایی است که انویدیا^۲ وارد صحنه می‌شود. خودروهای خوران باید حجم انبوهی از داده‌های گردآوری شده توسط حسگرها و دوربین‌ها را به صورت بلادرنگ پردازش کنند. انویدیا از پیشگامان ابررایانه‌های مبتنی بر هوش مصنوعی (درایو پی‌اِکس^۳) برای این منظور است. این رایانه شامل یک پردازشگر ویژه (به نام زاویر^۴) است که توان حرکتی این خودروها را فراهم می‌کند انویدیا از طریق شراکت با تویوتا توانست از توان پردازشگر خود برای بکارگیری هوش مصنوعی در خودروهای خودران بهره‌برداری کند.

¹ Guardian angel

² Nvidia

³ Drive PX2

⁴ Xavier

ابرایانه‌ی انویدیا دارای یک سیستم عامل ویژه‌ی مبتنی بر الگوریتم هوش مصنوعی است که یک نقشه‌ی سه بعدی مبتنی بر ابر با تفکیک‌پذیری بالا را شامل می‌شود. به کمک این قابلیت‌ها «مغز» خودرو می‌تواند محیط اطراف خودرو را درک کند. از آنجائیکه خودرو می‌تواند موقعیت خودش را نیز دقیقاً شناسایی کند، از هرگونه خطرات احتمالی مطلع می‌شود (مثلاً جاده در دست تعمیرات یا وسیله‌ی نقلیه‌ای که از روبرو می‌آید). سیستم عامل این شرکت دائماً در حال ارتقاء بوده و بدین ترتیب خودرو را هوشمندتر می‌کند (قابلیت یادگیری هوش مصنوعی). سیستم زاویر «مغز» خودرو را بر روی یک تراشه‌ی مخصوص (ولتا¹) ارائه کرده است که قادر به انجام ۳۰ تریلیون عملیات یادگیری عمیق در ثانیه است. بدین ترتیب، می‌تواند الگوریتم‌های هوش مصنوعی پیچیده را که شامل یادگیری ماشینی هستند پردازش کند. انتظار می‌رود که انویدیا با استفاده از ولتا، عصر جدید و قدرتمندی در رایانش هوش مصنوعی به روی ما بگشاید.

Source: Compiled from Korosec (2017) and blogs.nvidia.com/blog/2016/09/28/Xavier/.

سوالاتی برای بحث و گفتگو

۱. یک خودرو برای اینکه مستقل یا خودران باشد به چه چیزهایی نیاز دارد؟
۲. نقش انویدیا در خودروهای خودران چیست؟
۳. نقش زاویر چیست؟
۴. چرا برای انجام یک فرایند از یک ابررایانه استفاده می‌شود؟

علیرغم فناوری پیچیده‌ی مورد نیاز، چندین شرکت خودروساز آماده‌ی فروش یا بکارگیری زود هنگام اینگونه خودروها هستند (مثلاً بی‌ام‌دبلیو، مرسدس، فورد، جی‌ام، تسلا و البته گوگل). دستاوردهای مرتبط با وسائل نقلیه‌ی خودران به شرح زیر می‌باشند:

- اوپر و سایر شرکت‌های اشتراک‌گذاری خودرو برای استفاده از خودروهای خودران برنامه‌ریزی کرده‌اند.
- بسته‌های پستی توسط خودروهای خودران به درب منازل ارسال می‌شود. نگاه کنید به uspsaig.gov/blog/no-driver-needed
- اتوبوس‌های خودران در فرانسه و فنلاند تست می‌شوند. در مورد اتوبوس‌های خودران در هلسینکی ویدئوی money.cnn.com/video/technology/2016/18/self-driving-buses-hit-the-road-in-helsinki.cnnmoney را تماشا کنید.
- تاکسی‌های خودران در سنگاپور فعالیت می‌کنند.

قانون خودران اولین قانون ملی در ایالات متحده است که به مسئله‌ی خودروهای خودران می‌پردازد. هدف از این قانون، تنظیم مقررات و قوانین ایمنی مسافران در وسائل نقلیه‌ی خودران است که درهای تولید سالانه‌ی ۱۰۰،۰۰۰ خودرو را تا سال ۲۰۲۱ گشوده است.

¹ Volta

خودروها در جاده‌ها باشیم لازم است تا به چندین مشکلات اجرایی آنها بپردازیم. در ادامه دلایل طولانی شدن زمان تجاری‌سازی کامل این وسائل نقلیه آورده

- هزینه‌ی فناوری‌های نقشه‌نگاری سه‌بعدی بلادرنگ می‌بایست کاهش و کیفیت آنها افزایش یابد.
- قابلیت‌های نرم‌افزار هوش مصنوعی باید افزایش یابد. به عنوان مثال، هوش مصنوعی باید با شرایط غیرمنتظره‌ی بسیاری از جمله رفتار رانندگان سایر خودروها سر و کله بزند.
- بری (۲۰۱۶) سوال جالبی را مطرح کرد: «آیا مشتریان، خودسازان و بیمه‌گذاران واقعاً برای خودروهای خودران آمادگی دارند؟» مشتریان به نظر می‌رسد که ظهور این خودروها را پذیرفته باشند ولی در برابر سوار شدن آنها مقاومت می‌کنند. با وجود این، برخی افراد جسورتر انتظار دارند تا این خودروها بهتر از انسان‌ها رانندگی کنند.
- فناوری در این حوزه نیازمند پژوهش بیشتری است که می‌تواند بسیار پرهزینه باشد. یک دلیل آن این است که تعداد حسگرهای زیادی که در خودرو و جاده‌ها بکار گرفته شده‌اند باید بهبود یافته و هزینه‌ی آنها نیز کاهش یابد.
- اینترنت اشیا تعداد زیادی از اشیاء را برای وسائل نقلیه خودران به یکدیگر متصل می‌کند از جمله اشیاء موجود در فضای ابر. سیستم‌های اینترنت اشیا نیز باید بهبود یابند. به عنوان مثال، تأخیر در ارسال داده‌ها باید رفع شود. برای اطلاعات بیشتر در خصوص مسائل و مشکلات عمومی اجرایی هوش مصنوعی به فصل ۱۴ مراجعه کنید.

سوالاتی برای مرور بخش ۱۳.۹

۱. وسائل نقلیه‌ی خودران چه هستند؟ چه ارتباطی با اینترنت اشیا دارند؟
۲. مزایای وسائل نقلیه‌ی خودران برای رانندگان، جامعه و شرکت‌ها چه هستند؟
۳. چرا اوبر و شرکت‌های مشابه به وسائل نقلیه‌ی خودران علاقمند هستند؟
۴. برای پشتیبانی از وسائل نقلیه‌ی خودران به چه فناوری‌های هوش مصنوعی نیاز است؟
۵. خودروهای پرنده چه هستند؟
۶. برخی از مشکلات اجرایی وسائل نقلیه‌ی خودران را نام ببرید.

¹ Volocopter

² Sorkin

مؤسسه‌ی جهانی مک‌کنزی (بوگین و همکاران^۱، ۲۰۱۵) یک راهنمای جامع از اینترنت اشیاء برای مدیران اجرایی ارائه کرده‌اند. در این راهنما به مسائل زیر پرداخته شده است:

- **همسویی سازمانی.** اگرچه در مورد چندین ابتکار عمل فناوری دیگر این موضوع صادق است، اما در خصوص اینترنت اشیاء، فرصت‌های بهبود عملیاتی و ایجاد فرصت‌های جدید کسب‌وکار به این معنی است که پرسنل فناوری اطلاعات و عملیاتی باید به جای کارکردهای جداگانه به عنوان یک تیم با یکدیگر همکاری کنند. همانطور که توسط نویسندگان راهنمای جامع گفته شده است، «اینترنت اشیاء مفاهیم مسئولیت‌های سازمانی را به چالش می‌کشد. مدیران ارشد مالی، بازاریابی و عملیات و همچنین رهبران واحدهای کسب‌وکار باید پذیرای این باشند که سیستم‌هایشان را به یکدیگر متصل کنند.
- **چالش‌های تعامل‌پذیری.** تعامل‌پذیری تا به حال یک عامل بازدارنده‌ی بزرگ بر سر راه رشد کاربردهای اینترنت اشیاء بوده است. تعداد کمی از دستگاه‌های اینترنت اشیاء هستند که به طور یکپارچه به یکدیگر متصل می‌شوند. دوم، مسائل فنی زیادی در مورد اتصال وجود دارد. بسیاری از مناطق دور افتاده هنوز اتصال وای‌فای مناسبی ندارند. مسائل مربوط به پردازش کلان‌داده نیز باعث کندی پیشرفت در پذیرش اینترنت اشیاء می‌شود. شرکت‌ها در تلاشند تا داده‌ها را در سطح حسگر کاهش دهند تا کمترین حجم از داده‌ها به فضای ابر انتقال یابد. زیرساخت فعلی به سختی از حجم عظیمی از داده‌های جمع‌آوری شده توسط اینترنت اشیاء پشتیبانی می‌کند. یک مشکل در اینجا، مقاوم‌سازی حسگرهای دستگاه‌ها برای جمع‌آوری و انتقال داده‌ها برای تجزیه و تحلیل است. علاوه بر این، مصرف‌کنندگان برای جایگزینی اشیاء آنالوگ خود با محصولات هوشمند دیجیتالی اینترنت اشیاء جدید زمان می‌برند. به عنوان مثال، جایگزینی تلفن همراه برای افراد آسان‌تر از خودرو، لوازم آشپزخانه و چیزهایی است که می‌توانند از داشتن حسگر و اتصال به اینترنت اشیاء بهره‌مند شوند.
- **امنیت.** امنیت داده‌ها به طور کلی یک مسئله است، اما در زمینه اینترنت اشیاء این مشکل بزرگتر است. هر دستگاهی که به اینترنت اشیاء متصل می‌شود، نقطه ورود دیگری برای هکرهای مخرب می‌شود تا وارد یک سیستم بزرگ شوند یا حداقل یک دستگاه خاص را از کار ببندازند. داستان‌هایی از هکرها وجود دارد که می‌توانند به صورت خودکار از

¹ Bughin et al.

سیستم خودرو یا کنترل از راه دور درب بازکن پارکینگ نفوذ کرده و آنها را کنترل کنند. چنین مسائلی مستلزم آن است که در بکارگیری اینترنت اشیا در مقیاس بزرگ از همان ابتدا ملاحظات امنیتی را در نظر بگیریم.

با توجه به اینکه اینترنت از امنیت خوبی برخوردار نیست، بکارگیری شبکه‌های اینترنت اشیا نیازمند تدابیر امنیتی ویژه به خصوص در بخش شبکه‌های بی سیم است. پرکینز (۲۰۱۶) وضعیت را به صورت زیر خلاصه می‌کند: «اینترنت اشیا یک حضور دیجیتال فراگیر را ایجاد می‌کند که سازمان‌ها و جامعه را به صورت یک کل به هم متصل می‌کند. بازیگران جدید در این عرصه شامل دانشمندان علم داده‌ها، یکپارچه‌سازهای برون سازمانی^۱؛ و نقاط انتهایی افشاشده^۲ هستند. تصمیم گیرندگان امنیتی باید اصول اساسی ریسک و انعطاف‌پذیری را برای ایجاد تغییر بپذیرند». برای یک کتاب الکترونیکی رایگان در مورد اینترنت اشیا به مک‌لینان (۲۰۱۷) نگاه کنید.

از دیگر مسائل و مشکلات به موارد زیر می‌توان اشاره کرد:

- **حریم خصوصی.** برای تضمین حریم خصوصی، ما به یک سیستم امنیتی خوب به اضافه‌ی یک خط‌مشی و یک سیستم محافظت از حریم خصوصی نیاز داریم (به فصل ۱۴ مراجعه کنید). به دلیل بزرگی شبکه‌ها و استفاده از اینترنت در شرایط با محافظت کمتر، ساخت این سیستم امنیتی و سیستم محافظت از حریم خصوصی در اینترنت اشیا می‌تواند دشوار باشد. برای مشاوره از کارشناسان امنیتی برتر، هو^۳ (۲۰۱۶) را ببینید.
- **اتصال سیلوه‌های داده.** میلیون‌های سیلو داده در اینترنت وجود دارد و بسیاری از آنها باید در اپلیکیشن‌های خاص اینترنت اشیا به هم متصل می‌شوند. این موضوع به عنوان نیاز به یک «تار و پود» و اتصال شناخته می‌شود. این می‌تواند یک موضوع پیچیده برای اپلیکیشن‌هایی باشد که سیلوه‌های مختلف از سازمان‌های مختلف را در بر می‌گیرند. در تعاملات ماشین به ماشین، انسان به انسان، انسان به ماشین، انسان به خدمات و حسگرها ما به اتصال نیاز داریم. برای یک بحث و گفتگو نگاه کنید به رینی و اندرسون^۴ (۲۰۱۷) و machineshop.io/blog/the-fabric-of-the-internet-of-things. برای اطلاع از نحوه انجام اتصال در آی‌بی‌ام واتسون به ibm.com/Internet-of-things/iot-solutions/ مراجعه کنید
- **آماده‌سازی معماری‌های فناوری اطلاعات و مدل‌های عملیاتی موجود برای اینترنت اشیا می‌تواند مسئله‌ی پیچیده‌ای در بسیاری از سازمان‌ها باشد.** برای یک تحلیل و راهنمای کامل در مورد این موضوع به دایچمن و همکاران^۵ (۲۰۱۵) رجوع کنید. ادغام اینترنت اشیا در فناوری اطلاعات برای جریان داده‌های مورد نیاز اینترنت اشیا و جریان داده‌های پردازش شده توسط اینترنت اشیا به سمت اقدامات امری ضروری است.
- **مدیریت.** در معرفی هر فناوری جدید، حمایت مدیریت ارشد ضروری است. بوئی^۶ (۲۰۱۶) با توجه به ضرورت مقابله با سیلوه‌های داده که قبلاً توضیح داده شد استخدام یک مدیر ارشد داده را برای موفقیت اینترنت اشیا توصیه می‌کند. استفاده از چنین مدیر ارشدی می‌تواند به اشتراک‌گذاری اطلاعات در کلیه کارکردها، نقش‌ها و سطوح کسب‌وکار را تسهیل کند. در نهایت، مشکلات بین دپارتمان‌ها برای مالکیت و کنترل اینترنت اشیا را رفع می‌کند.

¹ External Integrators

² Exposed Endpoint

³ Hu

⁴ Reinie and Anderson

⁵ Deichman et al.

⁶ Bui

• **مشتریان متصل.** شواهدی مبنی بر افزایش استفاده از اینترنت اشیا در بازاریابی و روابط با مشتری وجود دارد. علاوه بر این، اینترنت اشیا موجب افزایش تعامل مشتری می‌شود. طبق گفته پارک^۱ (۲۰۱۷)، استقرار موفقیت‌آمیز اینترنت اشیا برای مشتریان به «مشتریان متصل»^۲ نیاز دارد. اتصال باید برای داده‌ها، تصمیم‌ها، نتایج، و کارکنان مرتبط با هر تماس مرتبط با اینترنت اشیا و بازاریابی باشد. سازمان تحقیقاتی بلو هیل^۳ گزارشی رایگان در مورد این موضوع ارائه می‌دهد (به پارک مراجعه کنید). اینترنت اشیا ارتباط بهتری با مشتریان کلیدی را امکان‌پذیر می‌سازد و خدمات مشتری را بهبود می‌بخشد. از ملاحظات ویژه می‌توان به مهمان‌پذیری، مراقبت‌های بهداشتی و شرکت‌های حمل و نقل اشاره کرد.

فینالی، چویی و همکاران^۴ (۲۰۱۸) در مطالعه‌ای جدید در مورد نحوه‌ی پیاده‌سازی موفق اینترنت اشیا توصیه‌هایی را ارائه کرده‌اند. با توجه به کثرت اینگونه مسائل و مشکلات اجرایی، اتخاذ یک استراتژی اجرایی ضروری است.

اینترنت اشیا حجم انبوهی از داده‌ها را جمع‌آوری می‌کند که می‌تواند برای بهبود فعالیت‌های کسب‌وکار برون‌سازمانی (به عنوان مثال بازاریابی) و همچنین عملیات درون‌سازمانی مورد استفاده قرار گیرد. ساس (۲۰۱۷) یک چرخه استراتژی را پیشنهاد کرده است که شامل مراحل زیر است:

۱. **اهداف کسب‌وکار را مشخص کنید.** آنها باید با منافع و هزینه‌های درک شده هماهنگ شوند تا بتوان ابتکار عمل‌ها را توجیه کرد. این مرحله مستلزم سطح بالایی از برنامه‌ریزی و بررسی منابع است. تحلیل بازگشت سرمایه اولیه برای این منظور توصیه می‌شود.

۲. **یک استراتژی تحلیلی را بیان کنید.** برای حمایت از بازگشت سرمایه و تهیه یک مورد کسب‌وکار، برنامه‌ریزی نحوه تجزیه و تحلیل داده‌های بزرگ ضروری است. این شامل انتخاب یک پلتفرم تحلیلی است که یک عامل بحرانی موفقیت محسوب می‌شود. ممکن است بررسی فناوری‌های هوش مصنوعی نوظهور، مانند یادگیری عمیق در اینجا انجام شود. یک انتخاب مناسب تضمین‌کننده‌ی یک راه‌حل اینترنت اشیا قدرتمند است.

۳. **نیاز به ابزارهای تحلیلی پیشرفته را ارزیابی کنید.** ابزارهای تحلیل پیشرفته یک فناوری است که برای برخی، اما نه همه‌ی اپلیکیشن‌ها مورد نیاز است که برای معرفی قابلیت‌های بلادرنگ در اپلیکیشن‌ها طراحی شده است. همچنین داده‌ها را برای تصمیم‌گیری خودکار و عمدتاً بلادرنگ فیلتر می‌کند زیرا داده‌های مربوطه فقط از طریق فیلتر کردن بدست می‌آیند.

۴. **راه‌حل‌های تحلیلی مناسب را انتخاب کنید.** راه‌حل‌های تحلیلی متعددی در بازار وجود دارد که توسط فروشندگان زیادی عرضه می‌شود. در استفاده از یک یا چند مورد برای اینترنت اشیا، لازم است چندین معیار مانند تناسب برای اینترنت اشیا، سهولت استقرار، توانایی به حداقل رساندن خطرات پروژه، پیچیدگی ابزارها و اتصال به سیستم‌های فناوری اطلاعات موجود (به عنوان مثال، کیفیت درگاه‌های اینترنت اشیا) در نظر گرفته شود. گاهی اوقات ایده خوبی است که به گروهی از فروشندگان که محصولات ترکیبی را ارائه می‌دهند مراجعه کنید (مانند ساس و اینتل). در

¹ Park

² Connected Customers

³ Blue Hill

⁴ Finnaly, Chui et al.

نهایت، زیرساخت‌های مناسب مانند سرورهای ابری با کارایی بالا و سیستم‌های ذخیره‌سازی باید بررسی شوند. اینها باید به عنوان یک پلتفرم مقیاس پذیر، موثر و کارآمد با هم کار کنند.

۵. بهبود مستمر، نقطه پایانی حلقه است. مانند هر چرخه استراتژی، عملکرد باید نظارت شده و در مراحل مختلف فرایند به ویژه از آنجایی که اینترنت اشیا به سرعت در حال تغییر و تحول است بهبود صورت گیرد. گستردگی دستیابی به هدف معیار مهمی است و ارتقای اهداف باید مورد توجه قرار گیرد.

خلاصه‌ای از فرایند در شکل ۱۳.۹ ارائه شده است.

ولدون^۱ (۲۰۱۵) مراحل زیر را برای پیاده‌سازی موفق اینترنت اشیا توصیه می‌کند:

- ایجاد یک مورد کسب‌وکار برای توجیه پروژه اینترنت اشیا شامل تجزیه و تحلیل هزینه-فایده و مقایسه با پروژه‌های دیگر.
- توسعه‌ی یک نمونه اولیه‌ی اجرایی. با آن آزمایش کنید. یاد بگیرید و آن را بهبود ببخشید.
- اینترنت اشیا را در یک واحد سازمانی استقرار دهید. با آن آزمایش کنید و درس بیاموزید.
- در صورت موفقیت‌آمیز بودن آزمایش، استقرار اینترنت اشیا را در سطح سازمان برنامه‌ریزی کنید. توجه ویژه‌ای به پردازش و انتشار داده‌ها داشته باشید.

با گذشت زمان، شاهد افزایش تعداد کاربردهای اینترنت اشیا، چه در درون و چه بیرون سازمان‌ها و شرکت‌ها هستیم. از آنجایی که تمام شبکه‌های اینترنت اشیا به اینترنت متصل هستند، اتصال برخی از شبکه‌ها به یکدیگر امکان‌پذیر خواهد بود و اینترنت اشیا بزرگ‌تری را ایجاد می‌کنند. این برای بسیاری از سازمان‌ها فرصت‌های رشد و توسعه را ایجاد می‌کند

حوزه‌های بالقوه‌ی مختلفی برای توسعه و رشد وجود دارد. یکی از حوزه‌هایی که هوش مصنوعی اینترنت اشیا را تقویت می‌کند، اکوسیستم آن است. بسیاری از برنامه‌های کاربردی اینترنت اشیا پیچیده هستند و می‌توانند با یادگیری ماشینی که می‌تواند بینش‌هایی درباره داده‌ها ارائه دهد، بهبود یابد. علاوه بر این، هوش مصنوعی می‌تواند به ایجاد دستگاه‌هایی («اشیا») کمک کند که می‌توانند مشکلات را خودشان تشخیص داده و حتی آنها را رفع کنند. برای بحث بیشتر، مارتین^۲ (۲۰۱۷) را ببینید. یکی دیگر از مزایای آینده هوش مصنوعی هنگامی که با اینترنت اشیا ترکیب شود، «شکل دادن به یک جفت همزیستی» است (هافپر^۳ ۲۰۱۶). این جفت شدن می‌تواند سیستم‌های شناختی را ایجاد کند که قادر به مدیریت و درک داده‌هایی باشند که با ابزارهای تحلیلی معمول امکان‌پذیر نمی‌باشد. ترکیب هوش مصنوعی و اینترنت اشیا می‌تواند یک شناخت درونی^۴ را ایجاد کند که قابلیت‌های هوش مصنوعی را به درون اشیا (از قبیل ربات‌ها و ماشین‌های تولیدی) تزریق کند تا این اشیا بتوانند محیط پیرامون خود را درک کرده و سپس از آنها آموخته و عملکردشان

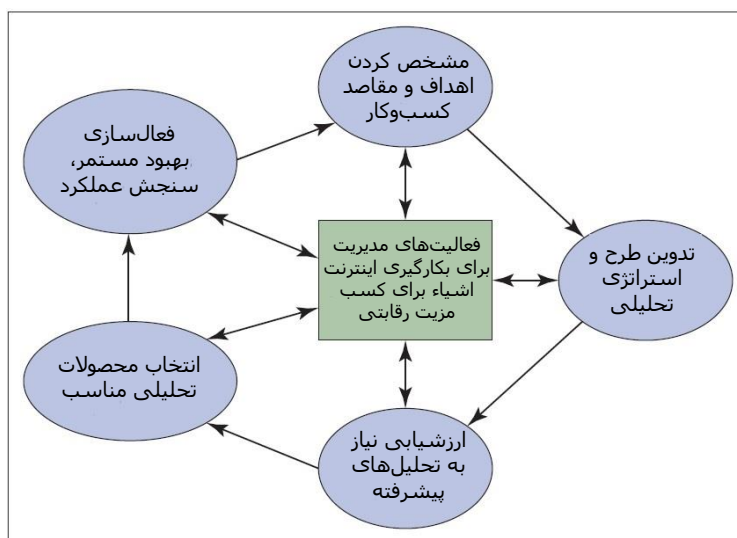
¹ Weldon

² Martin

³ Hupfer

⁴ Embodied Cognition

را بهبود دهند. برای جزئیات به هاپفر (۲۰۱۷) نگاه کنید. نهایتاً، هوش مصنوعی می‌تواند به یکپارچه‌سازی اینترنت اشیا با دیگر سیستم‌های فناوری اطلاعات کمک کند.



شکل ۱۳.۹ چرخه‌ی استراتژی اینترنت اشیا

- اینترنت اشیا یک فناوری انقلابی است که می‌تواند هم چیز را عوض کند.
- اینترنت اشیا به اکوسیستمی اطلاق می‌شود که در آن تعداد زیادی اشیا (مانند افراد، حسگرها و رایانه‌ها) از طریق اینترنت (اغلب به صورت بی‌سیم) به هم متصل هستند. بین سال‌های ۲۰۲۰ تا ۲۰۲۵، ممکن است ۵۰ میلیارد شیء متصل وجود داشته باشد. زیرسیستم‌های این اشیا متصل را می‌توان برای اهداف بسیاری استفاده کرد.
- استفاده از اینترنت اشیا می‌تواند فریندهای کسب‌وکار موجود را بهبود بخشد و کاربردهای کسب‌وکار جدیدی را ایجاد کند.
- میلیاردها شیء به اینترنت متصل خواهند شد و اکوسیستم اینترنت اشیا را تشکیل خواهند داد.
- اشیا موجود در اینترنت اشیا قادر به برقراری ارتباط خواهند بود و ساختار یک کنترل مرکزی را برای دستکاری اشیا و پشتیبانی از تصمیم‌گیری در برنامه‌های کاربردی اینترنت اشیا فراهم می‌سازد.
- اینترنت اشیا بسیاری از کاربردها صنعت، خدمات و دولت‌ها فعالیت می‌کند.

- کاربردهای اینترنت اشیا بر اساس تجزیه و تحلیل داده‌های جمع‌آوری شده توسط حسگرها یا سایر دستگاه‌هایی که برای پردازش در اینترنت جریان دارند وجود دارند.
- حسگرها می‌توانند داده‌ها را از تعداد زیادی از اشیاء جمع‌آوری کنند (مثلاً بیش از ۱ میلیون آسانسور در پرونده‌ی آغاین فصل ۱)
- اتصال اینترنت اشیا به سایر سیستم‌های فناوری اطلاعات نیازمند تلاش‌های مهم و اساسی است.
- برنامه‌های کاربردی اینترنت اشیا می‌توانند از تصمیمات اتخاذ شده توسط سازندگان تجهیزات و کاربران تجهیزات پشتیبانی کنند. (پرونده‌ی آغازین این فصل را ببینید.)
- آی‌بی‌ام واتسون یکی از عرضه‌کنندگان اصلی برنامه‌های کاربردی اینترنت اشیا در بسیاری از صنایع و خدمات (به عنوان مثال، تحقیقات پزشکی) است. پیش بینی می‌شود تا پایان سال ۲۰۱۸ تعداد کاربران آن به بیش از ۱ میلیارد نفر برسد.
- لوازم خانگی و خانه‌های هوشمند توسط اینترنت اشیا فعال می‌شوند.
- پروژه‌های شهر هوشمند در سراسر جهان توسط اینترنت اشیا پشتیبانی شده، کیفیت زندگی ساکنان شهرها را افزایش داده و از تصمیم‌های برنامه‌ریزان شهری و عرضه‌کنندگان فناوری‌ها پشتیبانی می‌کنند.
- اتومبیل‌های خودران ممکن است تصادفات، آلودگی، ترافیک و هزینه‌های حمل و نقل را کاهش دهند. خودروهای خودران هنوز به طور کامل اجرایی نشده‌اند، اما برخی از آنها در سال ۲۰۱۸ معرفی شدند.
- خانه‌ها و لوازم خانگی هوشمند محبوبیت زیادی دارند. مالکان آنها با هزینه‌ای اندک می‌توانند از چندین برنامه از امنیت خانه گرفته تا کنترل لوازم خانگی در خانه خود استفاده کنند.
- مفهوم شهرهای هوشمند در سطح جهانی با پروژه‌هایی در کشورهایی مانند هند، آلمان و ایالات متحده و سنگاپور در حال توسعه است. هدف شهرهای هوشمند فراهم کردن زندگی بهتر برای ساکنانشان است. حوزه‌های اصلی تحت پوشش عبارتند از حمل و نقل، مراقبت سلامت، صرفه جویی در انرژی، آموزش و خدمات دولتی.

- ۱.
- ۲.
- ۳.
- ۴.

.5
.6
.7
.8
.9

IoT Council

theinternetofthings.eu

.1
.2

<https://www.ptc.com/en/resource-center>

.3
.4

.5
.6
.7

smartcitiescouncil.com

.8
.9
.10

.11
.12

enterpriseinnovation.net/article/internet-things-next-big-wave-india-

1270947471/

.13

¹ AT & T

...	۱۴
youtube.com/watch?v=ZJr0X3XBmMA	۱۵
...	۱۶
youtube.com/watch?v=SfVUkGoCA7s “Smart Manufacturing”	۱۶
...	۱۷
...	۱۸
mckinsey.com/mgi/overview	۱۸

- ...
- ...
- ...
- ...

۱۴

مشکلات پیاده‌سازی: از اصول اخلاقی و حریم خصوصی گرفته تا تأثیرات سازمانی و اجتماعی

در این فصل پایانی، ما موضوعات مختلفی را در رابطه با پیاده‌سازی و آینده‌ی سیستم‌های هوشمند توضیح خواهیم داد و اینکار را با مسائل تکنولوژیکی مانند امنیت و اتصال آغاز می‌کنیم. سپس، به مسائل مدیریتی که شامل حقوقی، حریم خصوصی و اخلاق می‌باشد می‌پردازیم. در ادامه به بررسی تأثیرات آن بر سازمان‌ها، جامعه و مشاغل پرداخته و پس از آن روندهای فناوری که به آینده اشاره دارند را مرور می‌کنیم.

نکته: در این فصل ما تمام فناوری‌های اشاره شده در این کتاب را فناوری‌های هوشمند یا سیستم‌های هوشمند تلقی می‌کنیم. این فصل از بخش‌های زیر تشکیل شده است:

- فصل ۱۴.۱: پرونده‌ی آغازین: چرا اوپیر ۲۴۵ میلیون دلار به وی‌مو پرداخت کرد؟
- فصل ۱۴.۲: پیاده‌سازی سیستم‌های هوشمند: یک مرور
- فصل ۱۴.۳: مسائل حقوقی، حریم خصوصی و اخلاقی
- فصل ۱۴.۴: جاری‌سازی موفق سیستم‌های هوشمند
- فصل ۱۴.۵: تأثیرات سیستم‌های هوشمند بر سازمان‌ها
- فصل ۱۴.۶: تأثیرات بر مشاغل و کار
- فصل ۱۴.۷: خطرات بالقوه‌ی ربات‌ها، هوش مصنوعی و مدل‌های تحلیلی
- فصل ۱۴.۸: روندهای فناوری‌های مرتبط
- فصل ۱۴.۹: آینده‌ی سیستم‌های هوشمند

۱۴.۱. پرونده‌ی آغازین: چرا اوپیر ۲۴۵ میلیون دلار به وی‌مو پرداخت کرد؟

۱۴.۲. پیاده‌سازی سیستم‌های هوشمند: یک مرور

۱۴.۳. مسائل حقوقی، حریم خصوصی و اخلاقی

۱۴.۴. جاری‌سازی موفق سیستم‌های هوشمند

۱۴.۵. تأثیرات سیستم‌های هوشمند بر سازمان‌ها

۱۴.۶. تأثیرات بر مشاغل و کار

۱۴.۷. خطرات بالقوه‌ی ربات‌ها، هوش مصنوعی و مدل‌های تحلیلی

۱۴.۸. روندهای فناوری‌های مرتبط

۱۴.۹. آینده‌ی سیستم‌های هوشمند

در اوایل سال ۲۰۱۸، شرکت اوپیر تکنالاجیز ۲۴۵ میلیون دلار از سهام خود را به شرکت خودروهای خودران وی‌مو (یکی از شرکت‌های زیرمجموعه آلفابت) پرداخت کرد. این پرداخت برای حل و فصل دعاوی مطرح شده توسط وی‌مو مبنی بر استفاده‌ی غیرقانونی اوپیر از فناوری اختصاصی وی‌مو انجام شد.

این پرونده مربوط به حفاظت از مالکیت معنوی (اسرار تجاری) شرکت وی‌مو است. همانطور که از بخش ۱۳.۹ به خاطر دارید، وی‌مو پیشگام در عرضه‌ی خودروهای خودران بود. مهندس سابق وی‌مو (به نام لواندوفسکی)^۱ ۱۴,۰۰۰ فایل محرمانه‌ی خودروهای خودران وی‌مو را به صورت غیرقانونی دانلود کرده است. بدتر اینکه، لواندوفسکی ممکن است چندین مهندس برتر وی‌مو را متقاعد کرده باشد تا وی‌مو را ترک کرده و به او بپیوندند تا یک استارت‌آپ به نام اوتو کمپانی^۲ را برای توسعه وسایل نقلیه خودران تأسیس کنند. اوبر شرکت اتو را تصاحب کرد. برای اوبر، استفاده از چنین خودروهایی در سیستم حمل و نقل مسافر و دستیابی به سودآوری و رشد ضروری هستند. اوبر یک شرکت بزرگ حمل و نقل اشتراکی است که قصد دارد از ارائه خدمات خودروهای مشترک متعلق به افراد به سمت خدمات خودروهای مشترک برای کسب‌وکارها حرکت کند که در آن خودروهای خودران متعلق به اوبر و/یا خودروسازان هستند. به این ترتیب سود اوبر می‌تواند بسیار بیشتر باشد. علاوه بر این، اوبر قصد دارد ناوگان تاکسی‌های بدون راننده را راه اندازی کند.

بحث دعوی حقوقی بسیار پیچیده است و به مالکیت معنوی و توانایی کارکنان با فناوری پیشرفته برای کار کردن پس از ترک شغل خود به نفع رقبا می‌پردازد. وکلای می‌مو ادعا کردند که شرکت وی‌مو در صورت استفاده رقبا از اسرار تجاری آن شرکت متحمل خسارت بالقوه سنگینی خواهند شد. تیم حقوقی وی‌مو پرونده خود را بر اساس تحقیقات پزشکی قانونی دیجیتال استوار کرد که ثابت کرد لواندوفسکی عمداً پرونده‌های محرمانه را کپی کرده و سپس سعی کرده تا این کار خود را پنهان کند. توجه داشته باشید که اوبر اسرار تجاری را به سرقت نبرده است بلکه لواندوفسکی را که این اسرار را در اختیار داشته استخدام کرده است.

از نظر حقوقی، این پرونده منحصر به فرد و اولین مورد مربوط به خودروهای خودران بود، بنابراین هیچ پرونده‌ی قبلی برای استناد به آن وجود نداشت. این دو شرکت، شرکت‌های بزرگ فناوری در دره سیلیکون هستند. با کارمندانی که شرکت‌ها را ترک می‌کنند مصاحبه شده و به آنها یادآوری می‌شود که توافقنامه‌ای را در مورد اسرار تجاری که در ضمن کار برای شرکتی که ترک می‌کنند به دست آورده‌اند امضا کرده‌اند. لواندوفسکی در مصاحبه خروج از وی‌مو گفت که برنامه‌های آینده او شامل فعالیت‌های رقابتی که ممکن است با خودروهای خودران وی‌مو رقابت کند نخواهد شد. با این حال، او قبلاً با اوبر ملاقات کرده بود و این اسرار تجاری محرمانه را به شرکت جدید خود به نام اوتو تراکینگ^۳ فروخته بود. کاملاً مشخص شد که اوبر و لواندوفسکی هیچکدام حقیقت را نمی‌گویند.

رقبا پس از چهار روز در دادگاه به توافق رسیدند. پرونده در مقابل هیئت منصفه بود که این موضوع عنصر غیرقطعی و نامشخصی را وارد پرونده کرد. وی‌مو برای توافق رضایت داد زیرا، برای برنده شدن در پرونده باید اثبات می‌کرد که چه میزان خسارت واقعی

¹ Levandowski

² Otto Company

³ Otto Trucking

به شرکت وارد شده که قادر به چنین کاری نبود. محاسبه این خسارت بسیار دشوار است. علاوه بر این، هیچ مدرکی وجود نداشت که نشان دهد اوپر از اسرار تجاری وی مو استفاده کرده است و اوپر نیز لواندوفسکی را اخراج کرده بود. اوپر پذیرفت تا برای شرکت وی مو جبران خسارت کند زیرا پرونده‌ی حقوقی منجر به تأخیر در توسعه‌ی خودروهای خودران میشد که برای آینده‌ی اوپر بسیار حیاتی بود. همچنین، هزینه‌های رسیدگی به پرونده نیز در حال افزایش بودند (اوپر در پرونده‌های حقوقی دیگری که عمدتاً مربوط به رانندگان هستند نیز درگیر است). با توجه به جیب پر پول گوگل، جنگ با وی مو عاقبت خوشی نمی‌توانست داشته باشد. در واقع، وی مو پیام روشنی را فرستاده بود که از موقعیت خودروهای خودران خود به هر قیمتی محافظت می‌کند.

- اوپر حدود یک سوم از ۱ درصد سهام شرکت را به وی مو پرداخت کرد. ارزش سهام اوپر ۷۰ میلیارد دلار بود (ژانویه ۲۰۱۸) که بدین ترتیب مبلغی معادل ۲۴۵ میلیون دلار را به عنوان خسارت پرداخت کرد. اوپر به دنبال تبدیل شدن به سهامی عام است که می‌تواند ارزش آن را افزایش دهد.
- اوپر موافقت کرد تا از اطلاعات محرمانه‌ی وی مو در فناوری فعلی یا آینده‌ی خود استفاده کند. این یک شرط مهم وی مو بود.
- علت اینکه چرا این پرونده برای هر دو شرکت مهم بود این است که ارزش بازار وسائل نقلیه‌ی خودران تا سال ۲۰۵۰ به ۷ تریلیون دلار خواهد رسید (مارشال آند دیویس^۱ ۲۰۱۸). نکته: این معادل یک سوم کل بدهی جاری ایالات متحده است.
- ماهیت رقابت خودروهای خودران بین سال‌های ۲۰۱۶ و زمانیکه این پرونده در جولای ۲۰۱۸ به پایان رسید دستخوش تغییر اساسی شد. امروزه رقبای بسیار بیشتر و فناوری‌ها و فرایندهای شناخته‌شده‌تری وجود دارد (یعنی اسرار تجاری کمتر). نهایتاً، شرکت‌ها باید به کارکنان خود بگویند که چه چیزی اسرار تجاری محسوب نمی‌شود.

سوالاتی برای مرور پرونده‌ی آغازین

۵۶. مسائل حقوقی مربوط به این پرونده را شناسایی کنید.
۵۷. به نظر شما چرا وی مو موافقت کرد تا به جای پول، سهام شرکت اوپر را دریافت کند؟
۵۸. معنای مالکیت معنوی در این پرونده چیست؟
۵۹. قاضی فدرال در پایان گفت: « این پرونده اکنون به تاریخ پیوسته است ». منظور او چه بود؟
۶۰. خلاصه‌ای از آسیب‌های بالقوه‌ی این پرونده برای هر دو شرکت در صورت عدم دستیابی به یک توافق تهیه کنید.
۶۱. مزایای رسیدن به توافق در این پرونده را برای هر دو طرف بیان کنید.

¹ Marshall and Davies

خودروهای خودران محصول اصلی سیستم‌های هوشمند و هوش مصنوعی هستند که مزایای بالقوه زیادی برای طرفین آن دارند. همچنین رقابت شدید در صنعت و اهمیت اسرار تجاری به دست آمده در این مسیر اجتناب ناپذیر است. اختلافات حقوقی در شرایط رقابتی رایج است و حفاظت از مالکیت معنوی امری حیاتی است. حفاظت از مالکیت فکری موضوعی است که ما در فصل پایانی به آن پرداخته‌ایم. همچنین در این فصل به بررسی موضوعات دیگری که به پیاده‌سازی سیستم‌های هوشمند مربوط می‌شوند پرداخته‌ایم مثل: اخلاق، امنیت، حریم خصوصی، اتصال، یکپارچگی، استراتژی و نقش مدیریت ارشد. ما همچنین در این چکیده در مورد اهمیت آینده فناوری جدید وسایل نقلیه خودران آموختیم. این فناوری ممکن است تأثیر زیادی بر سازمان‌ها و ساختار و عملکرد آنها داشته باشد. علاوه بر این، ما در این فصل در مورد تأثیرات اجتماعی سیستم‌های هوشمند و به ویژه تأثیر آنها بر کار و مشاغل بحث می‌کنیم. ما همچنین برخی از پیامدهای بالقوه ناخواسته سیستم‌های هوشمند را بررسی می‌کنیم. در نهایت، ما به بررسی آینده‌ی احتمالی سیستم‌های هوشمند پرداخته و مناظرات بزرگ در خصوص خطرات و مزایای ممکن سیستم‌های هوشمند و به ویژه ربات‌ها و هوش مصنوعی را معرفی خواهیم کرد.

Sources: Compiled from A. Marshall & A. Davies. (2018, February 9). "The End of Waymo v. Uber Marks a New Era for Self-Driving Cars: Reality." *Wired*; A. Sage, et al. (2018, February 9). "Waymo Accepts \$245 Million and Uber's 'Regret' to Settle Self-Driving Car Dispute." *Reuters (Business News)*; K. Kokalitcheva. (2017, May 9). "The Full History of the Uber-Waymo Legal Fight." *Axios*.

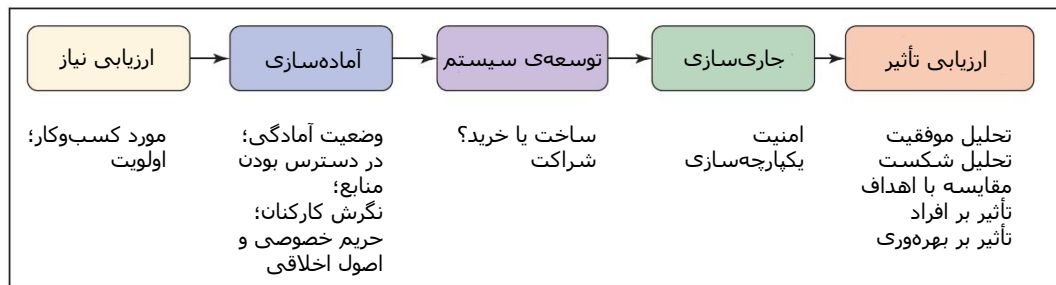
اکنون که اصول تجزیه و تحلیل، علم داده‌ها، هوش مصنوعی و فعالیت‌های پشتیبانی تصمیم را آموخته‌اید، ممکن است وسوسه شوید که بپرسید: با همه اینها در سازمانم چه کار می‌توانم انجام دهم؟ شما در مورد مزایای مهم آنها یاد گرفتید و در مورد چندین مطالعه‌ی موردی که در آن شرکت‌ها از سیستم‌های هوشمند استفاده می‌کنند مطلع شدید. بنابراین، چه کاری باید انجام دهید؟ ابتدا برخی از منابع توصیه شده در این کتاب را مطالعه کنید تا درک بهتری در مورد فناوری‌ها داشته باشید. در مرحله بعد، این فصل را که به مسائل اصلی مربوط به پیاده‌سازی سیستم‌های هوشمند در سازمان‌ها پرداخته است مطالعه کنید. پیاده‌سازی سیستم‌های تحلیلی/هوش مصنوعی می‌تواند کار پیچیده‌ای باشد. علاوه بر مسائل خاصی که در سیستم‌های هوشمند وجود دارد، مسائلی نیز وجود دارد که در بسیاری از دیگر سیستم‌های اطلاعاتی مبتنی بر رایانه مشترک هستند. در این بخش، انواع اصلی این مسائل را شرح می‌دهیم که برخی از آنها در این فصل مورد بحث قرار می‌گیرند. برای آشنایی با چندین عامل موفقیت پیاده‌سازی هوش مصنوعی که در نظرسنجی از ۳۰۰۰ مدیر اجرایی نشان داده شده است، به بوگین، مک‌کارتی و چویی^۱ (۲۰۱۷) رجوع کنید.

این فصل به سه بخش تقسیم می‌شود. در بخش اول، برخی از مسائل اجرایی مرتبط با مدیریت را شرح می‌دهیم. در بخش دوم، اثرات فناوری‌های هوشمند بر سازمان‌ها، مدیریت، کار و مشاغل را شرح می‌دهیم. بخش آخر به روندهای فناوری و آینده فناوری‌های هوشمند می‌پردازد. فرایند پیاده‌سازی سیستم‌های هوشمند مشابه فرایند عمومی در سایر سیستم‌های اطلاعاتی است. لذا به اختصار به ارائه آن می‌پردازیم. این فرایند در شکل ۱۴.۱ نشان داده شده است.

¹ Bughin, McCarthy and Chui

مرحله ۱ نیازسنجی در نیازسنجی نیاز باید یک مورد کسب‌وکار^۱ برای سیستم‌های هوشمند شامل بخش‌های اصلی آنها تهیه شود. این یک مرحله عمومی فناوری اطلاعات است و در اینجا مورد بحث قرار نخواهد گرفت.

مرحله ۲ آماده‌سازی در این مرحله لازم است آمادگی سازمان برای استفاده از ابزارهای تحلیل و هوش مصنوعی بررسی شود. بررسی منابع موجود، نگرش کارکنان نسبت به تغییر، اولویت‌های پروژه‌ها و ... ضروری است. این مرحله‌ی عمومی فناوری اطلاعات در اینجا مورد بحث قرار نخواهد گرفت با این حال، فکر کردن در مورد مسائل حقوقی، حریم خصوصی و اخلاقی مفید است زیرا آنها با فناوری‌های هوشمند مرتبط هستند که در بخش ۱۴.۳ توضیح داده شده است.



شکل ۱۴.۱ فرایند پیاده‌سازی

مرحله ۳ دستیابی به سیستم سازمان‌ها باید در مورد رویکرد داخلی یا برون‌سپاری (ساخت یا خرید) یا ترکیبی از این دو و احتمالاً با مشارکت با یک فروشنده یا شرکت دیگر تصمیم بگیرند. یک مشاور ممکن است در این مرحله کمک کند. این مرحله‌ی عمومی فناوری اطلاعات در اینجا مورد بحث قرار نخواهد گرفت.

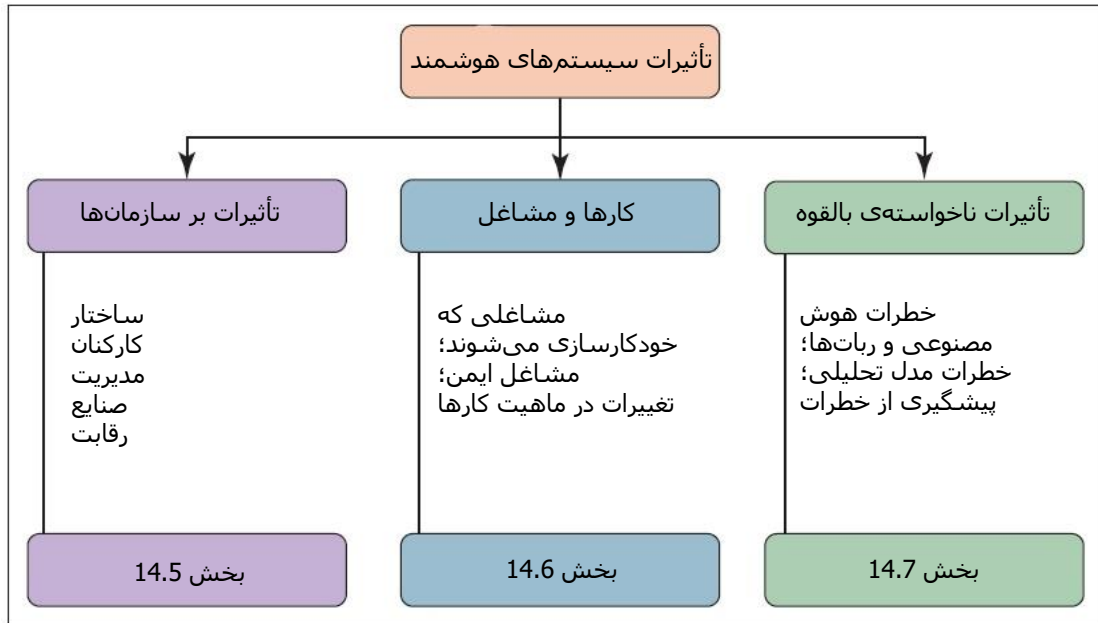
مرحله ۴ توسعه سیستم صرف‌نظر از اینکه چه کسی سیستم را توسعه خواهد داد، فعالیت‌های خاصی باید انجام شود. اینها شامل امنیت، ادغام با سیستم‌های دیگر، آماده‌سازی مدیریت پروژه و سایر فعالیت‌ها می‌شود. مجدداً، بسیاری از آنها عمومی هستند و در اینجا توضیح داده نمی‌شوند. فقط موارد انتخاب شده در بخش ۱۴.۴ توضیح داده شده است.

مرحله ۵ ارزیابی تأثیر بررسی عملکرد سیستم‌ها با توجه به برنامه‌ها ضروری است. مجدداً، بسیاری از آنها جنبه‌ی عمومی دارند و در اینجا توضیح داده نمی‌شوند.

سیستم‌های هوشمند بر تمام زندگی ما و بسیاری از مشاغل و سایر سازمان‌ها تأثیر می‌گذارند. پیدا کردن چیزی که تحت تأثیر قرار نگرفته است بسیار آسانتر از چیزی است که تحت تأثیر قرار گرفته است. در این بخش، این اثرات را به سه دسته تقسیم می‌کنیم که در شکل ۱۴.۲ به همراه شماره بخش‌هایی که در آنها توضیح داده شده‌اند نشان داده شده است. ما تأثیر بر افراد و کیفیت زندگی را که حوزه بسیار وسیعی هستند از این فهرست حذف کرده‌ایم (بهداشت، آموزش، سرگرمی، مبارزه با جرائم، خدمات اجتماعی و غیره).

¹ Business Case

در اینجا یک مثال در حوزه‌ی سرگرمی ارائه شده است. در آینده‌ی نزدیک، زمانیکه به دیزنی‌لند یا یکی از پارک‌های بین‌المللی دیزنی بروید شاهد ربات‌های آکروبات‌باز خواهید بود. این ربات‌ها را همه جا خواهید دید و بسیار شگفت‌انگیز هستند. برای یک مرور به ویدئوهای money.cnn.com/video/news/2018/07/04/disneyrobots-acrobatics-stuntronics-animatronics.cnnmoney/index.html و [youtube.com/watch?v=Z_QGsNpIOJ8](https://www.youtube.com/watch?v=Z_QGsNpIOJ8) نگاه کنید.



شکل ۱۴.۲ چشم‌انداز تأثیرات

سوالاتی برای مرور بخش ۱۴.۲

۱۹. مراحل اصلی فرایند پیاده‌سازی را نام ببرید.
۲۰. چرا پیاده‌سازی موضوع مهمی است؟
۲۱. تأثیرات مهم سیستم‌های هوشمند را توضیح دهید.

همانطور که علم داده، تجزیه و تحلیل، محاسبات شناختی و هوش مصنوعی در دسترس و فراگیر می‌شوند، همه ممکن است تحت تأثیر این برنامه‌ها قرار بگیرند. فقط به این دلیل که چیزی از طریق فناوری قابل انجام است، باعث نمی‌شود که آن را مناسب، قانونی یا اخلاقی بدانیم. متخصصان و مدیران علم داده‌ها و هوش مصنوعی باید از این دغدغه‌ها کاملاً آگاه باشند. مسائل متعدد مهم حقوقی، حریم خصوصی و اخلاقی مربوط به فناوری‌های هوشمند وجود دارد که به هم مرتبط هستند. به عنوان مثال، چندین موضوع حریم خصوصی بخشی از مسائل اخلاقی بوده یا جنبه حقوقی دارند. در اینجا ما فقط مثال‌ها و منابعی را ارائه می‌دهیم که در فصل ۱ به آن اشاره شد. هدف ما در

اینجا فقط این است که خواننده را با این مسائل آشنا کنیم. در خصوص دلیل اهمیت مسائل حقوقی، اخلاقی و حریم خصوصی هوش مصنوعی به کریگزمن^۱ (۲۰۱۷) مراجعه کنید.

معرفی فناوری‌های هوشمند ممکن است مجموعه‌ای از مسائل حقوقی را که قبلاً مربوط به سیستم‌های رایانه‌ای بوده‌اند را در بر بگیرد. به عنوان مثال، سؤالات مربوط به مسئولیت مشاوره‌های ارائه شده توسط ماشین‌های هوشمند در حال بررسی است. در این بخش، ما به ارائه‌ی مثال‌های بارز در این خصوص می‌پردازیم.

علاوه بر حل و فصل اختلافات در مورد نتایج غیرمنتظره و احتمالاً مخرب برخی از سیستم‌های هوشمند (نگاه کنید به پرونده‌ی آغازین و بخش ۱۴.۷)، مسائل پیچیده دیگری نیز ممکن است پدیدار شوند. به عنوان مثال، اگر شرکتی در حال ورشکستگی باشد، چه کسی مسئول استفاده از مشاوره‌ی یک برنامه مبتنی بر هوش مصنوعی خواهد بود؟ آیا خود شرکت در قبال عدم آزمایش کافی سیستم قبل از سپردن مسائل حساس و مهم به آن است؟ آیا موسسات حسابرسی و حسابداری در قبال عدم انجام آزمون‌های حسابرسی کافی و مناسب مسئول هستند؟ آیا توسعه دهندگان نرم افزار سیستم‌های هوشمند در این مسئولیت سهیم هستند؟ همانطور که اتومبیل‌های خودران رایج‌تر می‌شوند، زمانیکه حسگرها، شبکه یا سیستم هوش مصنوعی خودرو طبق برنامه عمل نمی‌کند، چه کسی مسئول آسیب‌ها خواهد بود؟ یک مورد اخیر که بحث را به صفحات اول جراید کشاند مربوط به تصادف خودروی تسلا است که در آن راننده در یک تصادف جان خود را از دست داد.

- زمانیکه نظر یک کارشناس در قالب یک برنامه رایانه‌ای کننویسی شده باشد ارزش نظر وی در دادگاه چیست؟
- چه کسی در قبال مشاوره (یا اطلاعات) اشتباه ارائه شده توسط یک برنامه هوشمند مسئول است؟ به عنوان مثال، اگر پزشک به واسطه‌ی رایانه تشخیص نادرستی داده و روشی را انجام دهد که منجر به مرگ بیمار شود، چه اتفاقی می‌افتد؟
- اگر یک مدیر قضاوت نادرستی را وارد یک برنامه هوشمند کند و منجر به یک آسیب یا فاجعه شود چه اتفاقی می‌افتد؟
- چه کسی صاحب دانش در یک پایگاه دانش است (به عنوان مثال، دانش یک چت‌بات)؟
- آیا مدیریت می‌تواند کارشناسان را وادار کند که تخصص خود را در یک سیستم هوشمند سهیم کنند؟ این موضوع چگونه جبران خواهند شد؟
- آیا رانندگی خودروهای خودران با رانندگان پشتیبان داخل خودرو در جاده‌های عمومی اشکالی ندارد؟ (بله، در چند ایالت، به ویژه در کالیفرنیا).
- چه کسی باید خودروهای خودران را تنظیم کند: شهرداری‌ها، ایالت‌ها یا دولت فدرال؟
- قانون‌گذاران دولت فدرال ایالات متحده در حال تدوین قوانین ملی برای اتومبیل‌های خودران (برای رانندگی ایمن) هستند.
- آیا استفاده از ربات‌های تحویل کالا یا غذا در پیاده‌روها مجاز است؟ (نه در سانفرانسیسکو بلکه در برخی شهرهای اروپایی).
- آیا رانندگان اوبر و شرکت‌های مشابه خویشتن فرما هستند؟ (نه در لندن).
- آیا ربات‌ها دارای حقوق انسانی هستند؟ (اگر مثل شهروندانی مثل سوفیا در عربستان سعودی باشند؟) اگر چنین حقوقی داشته باشند آیا مسئولیت‌های حقوقی نیز بر عهده آنها خواهد بود؟

¹ Krigsman

- آیا تاکسی‌های ربات وجهی قانونی دارند؟ آیا سفر با این تاکسی‌ها ارزان‌تر خواهد بود؟ (بله در سنگاپور و برخی جاهای دیگر می‌تواند ارزان‌تر باشد).

Source: Turban, Introduction to Information Technology, 2nd edition, John Wiley & Sons, 2006.

پرونده‌ی آغازین توجه ما را به یک موضوع حقوقی که برای شرکت‌های مرتبط با فناوری بسیار مهم است معطوف کرد: مالکیت و حفاظت از دارایی معنوی.

چندین مورد از مسائل اخلاقی که بعداً توضیح داده خواهد شد با مسائل حقوقی در هم آمیخته‌اند. به عنوان مثال، حقوق قانونی ربات‌ها را در نظر بگیرید. آیا ما به این حقوق نیاز داریم؟ برای چه (یک موضوع اخلاقی)؟ سپس تدوین حقوق قانونی ضروری خواهد بود. به عنوان مثال، فیسبوک در مورد تشخیص چهره با مشکلات حقوقی مواجه بود. قوانین ایمنی برای ربات‌ها مدت‌ها پیش تدوین شده است. در حال حاضر، قوانین بسیار کمی در مورد فناوری‌های هوشمند وجود دارد و بیشتر قوانین مربوط به ایمنی هستند.

علاوه بر قوانین مربوط به رباتیک و هوش مصنوعی، زیرشاخه‌ای از هوش مصنوعی وجود دارد که به کاربردهای هوش مصنوعی در حرفه حقوقی و حل برخی از مشکلات حقوقی مربوط می‌شود. طبق دوناهو^۱ (۲۰۱۸)، برخی از موضوعات مهم شامل موارد زیر هستند:

- تجزیه و تحلیل داده‌های حقوقی (به عنوان مثال، تعارضات قانونی) برای شناسایی الگو
- ارائه مشاوره حقوقی به مصرف‌کنندگان (به عنوان مثال DoNotPay.com را ببینید).
- بررسی اسناد
- تجزیه و تحلیل قراردادها
- حمایت از پژوهش‌های حقوقی
- پیش‌بینی نتایج (به عنوان مثال، احتمال برنده شدن)
- تأثیر هوش مصنوعی بر حرفه حقوقی

هوش مصنوعی می‌تواند وظایف معمول حقوقی مانند مدیریت اسناد و تهیه پیش‌نویس قراردادها را انجام دهد. برای جزئیات به کان (۲۰۱۷) مراجعه کنید. برای ۳۵ برنامه کاربردی در حوزه قانون و رویه‌های حقوقی، رایو^۲ (۲۰۱۸) ببینید. مسائل حقوقی ممکن است به شدت با موضوع بعدی ما یعنی حریم خصوصی مرتبط باشد.

حریم خصوصی برای افراد مختلف معانی متفاوتی دارد. به طور کلی، حریم خصوصی^۳، حق آزار ندیدن و حق رهایی از تجاوز و مداخله‌های غیرمنطقی است. حریم خصوصی از دیرباز با مسائل حقوقی، اخلاقی و اجتماعی در بسیاری از کشورها مرتبط بوده است. حق حفظ حریم خصوصی امروزه در همه ایالت‌های آمریکا و توسط دولت فدرال یا توسط قانون عمومی به رسمیت شناخته شده است. حریم خصوصی

¹ Donahue

² Rayo

³ Privacy

می‌تواند تعریف گسترده‌ای داشته باشد. با این حال، دو قانون زیر در تصمیمات دادگاه‌های گذشته به دقت رعایت شده است: (۱) حق حریم خصوصی مطلق نیست. حریم خصوصی باید با نیازهای جامعه متعادل شود. (۲) حق عموم مردم برای دانستن از حق افراد برای حفظ حریم خصوصی بالاتر است. این دو قانون نشان می‌دهد که چرا تعیین و اعمال آن در برخی موارد دشوار است. مسائل مربوط به حریم خصوصی آنلاین از ویژگی‌ها و سیاست‌های خاصی برخوردار هستند. یکی از زمینه‌هایی که ممکن است حریم خصوصی به خطر بیفتد در ادامه مورد بحث قرار می‌گیرد. با افزایش نمایی حجم داده‌های تولید شده در اینترنت، اهمیت حریم خصوصی نیز روز به روز در حال افزایش است. برای یک مرور کلی از حریم خصوصی با توجه به ارتباط آن با هوش مصنوعی به پرووازا^۱ (۲۰۱۷) نگاه کنید.

هدف فناوری‌های هوشمند ارائه خدمات و بازاریابی هدفمند به مصرف‌کنندگان است. آنها این کار را از طریق گردآوری اطلاعات در مورد این مشتریان انجام می‌دهند. در گذشته پیچیدگی فعالیت‌های گردآوری، مرتب سازی، بایگانی و دسترسی به اطلاعات به صورت دستی از چندین سازمان دولتی و سایر پایگاه‌های داده عمومی، در بسیاری از موارد، یک محافظت داخلی در برابر سوء استفاده از اطلاعات خصوصی محسوب می‌شد. اینترنت در ترکیب با پایگاه‌های داده در مقیاس بزرگ، بعداً کاملاً جدیدی از دسترسی و استفاده از داده‌ها را ایجاد کرده است. از قدرت ذاتی سیستم‌های هوشمند که می‌توانند به حجم وسیعی از داده‌ها دسترسی داشته و آنها را تفسیر کنند، می‌توان به نفع جامعه استفاده کرد. به عنوان مثال، با تجزیه و تحلیل سوابق به کمک تجزیه و تحلیل کسب‌وکار، می‌توان کلاهبرداری، جرم، سوء مدیریت دولت، فرار مالیاتی، کلاهبرداری از رفاه، خساست خانواده، استخدام کارگران غیرقانونی و غیره را از بین برده یا کاهش داد. با این حال، فرد به لحاظ از دست دادن حریم خصوصی باید چه بهایی را بپردازد تا دولت بتواند مجرمین را دستگیر کند؟ همین امر در سطح شرکت نیز صادق است. اطلاعات خصوصی در مورد کارمندان ممکن است به تصمیم‌گیری بهتر شرکت کمک کند، اما حفظ حریم خصوصی کارمندان ممکن است به خطر بیفتد.

استفاده از فناوری‌های هوش مصنوعی در مدیریت و اجرای قوانین و مقررات ممکن است نگرانی عمومی را در مورد حفظ حریم خصوصی اطلاعات افزایش دهد. این ترس‌ها که توسط توانایی‌های درک شده هوش مصنوعی ایجاد می‌شوند، باید در ابتدای هر اقدام برای توسعه هوش مصنوعی مورد توجه قرار گیرند.

آمازون اکو/الکسا و دستگاه‌های مشابه به هر آنچه در محیط اطراف رخ می‌دهد گوش می‌دهند. آنها همچنین ممکن است عکسبرداری کنند. به عبارت دیگر، دستیار صوتی شما جاسوسی شما را می‌کند. پیشرفته‌ترین این دستیارها، زوج اکو/الکسا است. امروزه می‌توانید از الکسا بخواهید محصولات آمازون را خریداری کند. آمازون و گوگل اختراعی را ثبت کرده‌اند که به دستیاران مجازی خانه شما این امکان را می‌دهد تا محصولات را برای شما تبلیغ کرده و به شما بفروشند. طرفداران حریم خصوصی از این موضوع خرسند نیستند اما مشتریان ممکن است راضی باشند. به عنوان مثال، الگن^۲ (۲۰۱۷) توضیح می‌دهد که چگونه الکسا به عنوان یک مشاور مد با استفاده از بررسی سبک پوشش لباس عمل می‌کند. این سیستم، دانش یک متخصص مد و دانش هوش مصنوعی را ترکیب می‌کند. یک سیستم توصیه به شما دو عکس را در یک زمان نشان داده و به شما می‌گوید کدام یک را بخرید (بر اساس رنگ، روند فعلی و غیره). آمازون برای مفید و کاربردی‌تر شدن این سیستم در حال بهبود حریم خصوصی است. این ممکن است کار آسانی نباشد زیرا سوابق شما در ابر آمازون ذخیره می‌شود.

هاف^۳ (۲۰۱۷) استدلال‌هایی در مورد خطرات دستیارها و حفاظت ارائه شده توسط آمازون ارائه می‌دهد.

¹ Provazza

² Elgen

³ Huff

بسیاری از کاربران از اطلاعات خصوصی که به واسطه‌ی استفاده از تلفن هوشمند آنها ردیابی می‌شود بی اطلاع هستند. بسیاری از اپلیکیشن‌ها داده‌های کاربر را با هر تماس تلفنی در گذر از یک دکل مخابراتی به دکل دیگر با استفاده از دستگاه‌های مجهز به جی‌پی‌اس که موقعیت مکانی کاربر و از طریق تلفن‌هایی که اطلاعات را در نقاط اتصال وای‌فای ارسال می‌کنند گردآوری می‌کنند. توسعه دهندگان اصلی اپلیکیشن‌ها مدعی هستند که به حریم خصوصی کاربران توجه ویژه‌ای دارند، اما جالب است بدانید که چه مقدار اطلاعات از طریق استفاده از یک دستگاه واحد در دسترس قرار می‌گیرد به خصوص زمانی که گوشی‌های هوشمند دارای اجزای بیشتر و بیشتر هوش مصنوعی هستند.

برای حفظ حریم خصوصی و امنیت اینترنت اشیاء به هو (۲۰۱۶) مراجعه کنید. داده‌های بیشتری از طریق شبکه‌های اینترنت اشیاء در جریان است. توجه داشته باشید که مسائل مربوط به حریم خصوصی داده‌های هوش مصنوعی در حال افزایش است، به خصوص زمانی که هوش مصنوعی با داده‌های مصرف‌کنندگان سروکار دارد. حجم داده‌های گردآوری شده نیز به شدت در حال افزایش است برای مثال، از طریق یادگیری ماشینی و چت‌بات‌ها. همچنین، در سازمان‌ها، کارفرمایان داده‌های بیشتری را در مورد کارکنان گردآوری و تجزیه و تحلیل می‌کنند. چگونه از داده‌ها و در برابر سوء استفاده از آنها محافظت کنیم؟

با رشد تعداد کاربران اینترنت به طور کلی و کاربران دستگاه‌های تلفن همراه به طور خاص، بسیاری از شرکت‌ها بر اساس میزان مصرف دستگاه، جستجو در اینترنت، شماره مخاطبین و ... شروع به استفاده از فناوری‌های هوشمند برای توسعه پروفایل‌های کاربران کرده‌اند. *وال استریت ژورنال*^۱ مجموعه‌ای عالی از مقالات با عنوان «آنچه می‌دانند» را منتشر می‌کند (WallStreetJournal.com 2016). این مقالات دائماً به روز می‌شوند تا بر آخرین فناوری و مسائل مربوط به حریم خصوصی/اخلاقی تأکید کنند. یکی از شرکت‌هایی که در این مجموعه مقالات به آنها اشاره شده است، رپلیف (در حال حاضر بخشی از تاوردیتا^۲) نام دارد. فناوری رپلیف^۳ مدعی است که می‌تواند مشخصات یک کاربر را فقط با دانستن آدرس ایمیل او ارائه دهد. واضح است که فناوری رپلیف آن را قادر می‌سازد تا اطلاعات مرتبط مهمی را جمع‌آوری کند. شرکت دیگری که هدف آن شناسایی دستگاه‌ها بر اساس استفاده از آنها بلوکاوا است که اخیراً با کوالیا (qualia.com) ادغام شده است. فناوری بلوکاوا^۴ کوالیا به یک پروفایل شخصی متصل می‌شود تا بتواند کاربر را حتی اگر با چندین دستگاه موبایل یا لپ‌تاپ کار می‌کند به عنوان یک فرد یا یک خانوار تشخیص دهد. همه این شرکت‌ها از ابزارهای تحلیلی مانند خوشه‌بندی و کاوش وابستگی برای توسعه پروفایل‌های کاربران استفاده می‌کنند. البته بسیاری از شرکت‌های نوپای تحلیلی در این فضا مدعی احترام به حریم خصوصی کاربران هستند، اما اغلب موارد تخلف گزارش می‌شود. به عنوان مثال، رپلیف اطلاعات غیرمجاز را از کاربران فیسبوک جمع‌آوری می‌کرده و متعاقباً فیسبوک جلوی آن را گرفت. یکی از کاربران گزارش داد که یک ساعت پس از آنکه آدرس ایمیل خود را به یک شرکت متخصص در نظارت بر اطلاعات کاربران داده است (reputation.com)، این شرکت توانسته شماره تأمین اجتماعی او را کشف کند. بنابراین، نقض حریم خصوصی باعث ایجاد ترس از رفتار مجرمانه در مورد اطلاعات می‌شود. این حوزه به طور کلی نگرانی بزرگی محسوب می‌شود و نیاز به مطالعه دقیق دارد. این مثال‌ها

¹ Wallstreet Journal

² Towerdata

³ Rappleaf

⁴ BlueCava

نه تنها نشان دهنده‌ی قدرت تجزیه و تحلیل در توانایی یادگیری بیشتر در مورد مشتریان هدف است بلکه به عنوان هشدار برای هوش مصنوعی و متخصصان تجزیه و تحلیل در مورد حساس بودن به حریم خصوصی و مسائل اخلاقی نیز عمل می‌کند. یکی دیگر از حوزه‌های کاربردی مرتبط با نگرانی‌های حفظ حریم خصوصی، تجزیه و تحلیل رفتارهای کارکنان بر اساس داده‌های جمع‌آوری شده از حسگرهایی است که در نشان سینه‌ی کارمندان تعبیه شده است. از شرکت‌های معروف می‌توان به هیومنایز^۱ اشاره کرد که کاربردهای متعددی را از طریق حسگر تعبیه شده در نشان سینه‌ی کارکنان خود گزارش داده است. این حسگرها تمام حرکات یک کارمند را ردیابی می‌کنند.

بارکلیز^۲ با استفاده از حسگرهای حرارتی و حرکتی، مدت زمانی که بانکدارانش پشت میزشان هستند را ردیابی می‌کند. این سیستم در شعبه‌های لندن، انگلستان نصب شده است. توضیح رسمی این بود که ادارات خالی در بانک برای تخصیص بهینه و احتمالاً کاهش فضای اداری شناسایی شود. شبکه اینترنت اشیا داشبوردهایی را ارائه کرد که نشان می‌داد کدام ادارات (مکعب‌ها) کمتر مورد استفاده قرار می‌گرفتند و روند استفاده از آنها چگونه بود. بانک به کارکنان و اتحادیه اعلام کرد که هدف این پروژه اندازه‌گیری بهره‌وری بلکه فقط میزان استفاده از فضای اداری را اندازه‌گیری می‌کند. از نتایج آن می‌توان برای مدیریت بهتر مصرف انرژی در ادارات و برنامه‌ریزی یک محیط کاری انعطاف پذیر استفاده کرد. در نتیجه، بارکلیز توانست در فضای اداری صرفه جویی کرده و آنها را به مبلغ ۴۵ میلیون دلار در سال اجاره دهد.

این بانک از یک سیستم ردیابی مشابه استفاده می‌کند تا متوجه شود کارمندان مختلف چه مقدار زمان را با مشتریان می‌گذرانند. اتحادیه این اپلیکیشن اینترنت اشیا را به دقت تحت نظر دارد تا مطمئن شود که برای جاسوسی از کارمندان استفاده نمی‌شود. سایر بانک‌ها در انگلستان از سیستم‌های مشابهی استفاده می‌کنند. برای جزئیات، به اخبار بلومبرگ^۳ (۲۰۱۷) مراجعه کنید. البته، موقعیت‌هایی مانند مواردی که شرح داده شد، نگرانی‌های عمده‌ای را در مورد حریم خصوصی ایجاد می‌کنند. آیا شرکت‌ها مجاز هستند تا کارمندان خود را به طور سرزده نظارت کنند؟ در نهایت، احتمال باج افزار یا حملات هکرها به ربات‌ها وجود دارد که می‌تواند علیه مشاغلی که کارمندان آنها از چنین ربات‌هایی استفاده می‌کنند استفاده شود. اسمیت (۲۰۱۸) در مورد تحقیقاتی گزارش داد که ۵۰ نقطه آسیب‌پذیر ربات‌ها را شناسایی کرد. حملات باج افزار ممکن است عملیات را مختل کرده و سازمان‌ها را مجبور به پرداخت باج‌های سنگینی کند.

موارد زیر چند نمونه دیگر از نقض احتمالی حریم خصوصی در دنیای فناوری هوشمند را نشان می‌دهند:

- پلیس دلاور^۴ از دس‌کام‌های^۵ هوش مصنوعی برای جستجوی افراد فراری در اتومبیل‌های رهگذر استفاده می‌کند. عکس‌ها و فیلم‌های گرفته شده به ابرها ارسال شده و توسط الگوریتم‌های هوش مصنوعی تجزیه و تحلیل می‌شوند.
- سیستم‌های تشخیص چهره فیسبوک نگرانی‌هایی را در مورد حفاظت از حریم خصوصی ایجاد می‌کنند.

¹ Humanize

² Barclay

³ Bloomberg News

⁴ Delaware

⁵ Dashcom

- اپیسنتر^۱ به کارکنان خود یک ریزتراشه ایمپلنت ارائه می‌دهد. این ریزتراشه مانند یک کارت مغناطیسی عمل کرده، درها را باز کرده، از فروشگاه شرکت برای شما غذا خریداری کرده و خیلی کارهای دیگری را انجام می‌دهد. اما مدیریت می‌تواند شما را نیز ردیابی کند. این ریزتراشه فقط در اختیار داوطلبان قرار داده می‌شود.

با رشد اخیر داده‌هایی که به واسطه‌ی استفاده‌ی ما از فناوری‌ها حاصل شده‌اند و توانایی شرکت‌ها برای دسترسی و استخراج آنها، بحث حفظ حریم خصوصی نیز به این موضوع منجر شده که داده‌های هر کاربر به چه کسی تعلق دارد؟ برای نکات برجسته در خصوص این مشکل به ستون بلومبرگ بیزینس‌ویک^۲ از ولش^۳ (۲۰۱۶) نگاه کنید. یک خودروهای نسبتاً جدید را به عنوان مثال در نظر بگیرید. این خودرو مجهز به حسگرهایی زیادی است که از حسگرهای فشار باد تایر تا ردیاب‌های جی‌پی‌اس را در بر می‌گیرند که می‌توانند مکان‌هایی که رفته‌اید، سرعت رانندگی‌تان، زمان تغییر مسیر و غیره را ردیابی کنند. خودرو حتی ممکن است وزن مسافر را که به صندلی جلو اضافه شده است بداند. همانطور که ولش خاطر نشان می‌کند، یک خودروی متصل به اینترنت (بیشتر خودروهای جدید هستند!) می‌تواند یک کابوس برای حفظ حریم خصوصی برای مالک یا یک «معدن طلا» از داده‌ها برای هر کسی که این داده‌ها را در اختیار داشته یا تجزیه و تحلیل می‌کند باشد. جنگ اساسی بین تولیدکنندگان خودرو و ارائه دهندگان فناوری مانند اپل (کارپلی^۴) و گوگل (اندروئید اوتو^۵) در حال شکل رفتن است که چه کسی مالک این داده‌ها است و چه کسی می‌تواند به آنها دسترسی داشته باشد. اهمیت این موضوع بیشتر نیز می‌شود زیرا با هوشمندتر شدن خودروها و در نهایت خودران شدن آنها، راننده/مسافر داخل خودرو می‌تواند یک چشم انداز بسیار هدفمند برای خدمات بازاربایان باشد. به عنوان مثال، برنامه ویز گوگل داده‌های جی‌پی‌اس را برای ردیابی ترافیک از میلیون‌ها کاربر جمع‌آوری می‌کند و به کاربران کمک کند تا بهترین مسیرها را پیدا کنند. اما همچنین تبلیغات پاپ آپ را بر روی صفحه نمایش کاربران نمایش می‌دهد. اسپاتیفای، پلپ و سایر برنامه‌هایی که معمولاً در خودروها استفاده می‌شوند رویکردهای مشابهی دارند.

نکته اصلی این است که متخصصان و کاربران سیستم‌های هوشمند باید از مسائل قانونی و اخلاقی مربوط به جمع‌آوری اطلاعاتی که ممکن است دارای امتیاز بوده یا تحت محافظت هستند آگاه باشند. مسائل مربوط به حریم خصوصی در بسیاری از موارد به عنوان مؤلفه‌های مهم اخلاقی در نظر گرفته می‌شود.

چندین موضوع اخلاقی به سیستم‌های هوشمند مربوط می‌شود. ارزش‌های شخصی عامل اصلی در تصمیم‌گیری‌های اخلاقی هستند. مطالعه مسائل اخلاقی به دلیل ماهیت چندبعدی آنها کار پیچیده‌ای است. یکی از داستان‌هایی که موجب ناراحتی بسیاری از کاربران شد (اگرچه غیرقانونی نبود) آزمایش فیسبوک برای ارائه اخبار مختلف به کاربران و نظارت بر واکنش‌های احساسی آنها بر اساس پاسخ‌ها، لایک‌ها و تحلیل احساسات بود. اکثر شرکت‌ها، از جمله شرکت‌های فناوری، برای شناسایی ویژگی‌هایی که بیشتر پسندیده یا ناپسند هستند آزمایش‌هایی را بر روی کاربران انجام می‌دهند و بر اساس آن، محصولات خود را به‌خوبی تنظیم می‌کنند. از آنجایی که فیسبوک

¹ Epicenter

² Bloomberg Businessweek

³ Welch

⁴ CarPlay

⁵ Android Auto

شرکت بسیار بزرگی است، انجام آزمایش بدون رضایت آگاهانه کاربران امری غیراخلاقی تلقی شد. در واقع، فیسبوک خطای خود را پذیرفت و بازبینی رسمی تری را از طریق هیئت‌های بازبینی داخلی و مکانیسم‌های انطباقی دیگری را برای آزمایش‌های آتی آغاز کرد. مورگان^۱ (۲۰۱۷) گفت برای حفظ اصول اخلاقی و شفافیت در هر موقعیتی لازم است تا این اصول پایه‌ی تمام کارهایی که هوش مصنوعی هم برای فروشندگان و هم برای مشتریان انجام می‌دهد قرار گیرد. بدین ترتیب افراد می‌توانند صادق بمانند و به اهداف هوش مصنوعی پایبند باشند، بنابراین می‌توانند نقش مهمی در زندگی و کار ما داشته باشد. برای اینکه چگونه مسائل اخلاقی با آلفابت تداخل دارد به ابتکار عمل‌های گوگل و کان^۲ (۲۰۱۷) مراجعه کنید.

بسیاری افراد در مورد مسائل اخلاقی در هوش مصنوعی، رباتیک و سایر سیستم‌های هوشمند سوالاتی را مطرح کرده‌اند. به عنوان مثال، بوسمان (۲۰۱۶) مسائل زیر را مطرح کرده است:

۱. تأثیر آنها بر مشاغل چیست؟ (به بخش ۱۴.۵ مراجعه کنید)
۲. ماشین‌ها (یعنی ربات‌ها) چگونه بر رفتار و تعاملات ما تأثیر می‌گذارند؟
۳. چگونه می‌توان ثروت ایجاد شده توسط ماشین‌های هوشمند را توزیع کرد؟ (به عنوان مثال، کاپلان، ۲۰۱۶)
۴. چگونه می‌توانیم در مقابل اشتباهات برنامه‌های هوشمند از خودمان محافظت کنیم؟ به عنوان مثال، برنامه‌های آموزشی در یادگیری ماشین چقدر باید باشد؟
۵. آیا سیستم‌های هوشمند می‌توانند منصفانه و بی طرف باشند؟ چگونه می‌توان جهت‌گیری و غرض‌ورزی در ایجاد و عملکرد سیستم‌های هوش مصنوعی را از بین برد؟
۶. چگونه می‌توان برنامه‌های کاربردی هوشمند را از دشمنان در امان نگه داشت؟
۷. چگونه می‌توان سیستم‌ها را در برابر پیامدهای ناخواسته محافظت کرد؟ (به عنوان مثال، حوادث در عملیات ربات). به عنوان مثال، محققان فیسبوک مجبور شدند یک سیستم هوش مصنوعی که زبان ضعیفی را ایجاد می‌کرد غیرفعال کنند.
۸. چگونه می‌توانیم کنترل یک سیستم هوشمند پیچیده را حفظ کنیم؟
۹. آیا باید حقوق قانونی ربات‌ها را توسعه دهیم؟ چگونه می‌توانیم رفتار انسانی با ماشین‌های هوشمند را تعریف و برنامه‌ریزی کنیم؟
۱۰. آیا باید اجازه دهیم یک جامعه ربات خودگردان به همراه جامعه ما وجود داشته باشد؟
۱۱. تا چه حد باید بر رفتارهای ناخواسته ربات‌ها تأثیر بگذاریم (یا حتی بتوانیم)؟
۱۲. چگونه می‌توانیم مسئله مالکیت ماشین هوشمند را حل کنیم؟

مسائل دیگر عبارتند از:

- نظارت الکترونیکی
- اخلاق در هوش تجاری و طراحی سیستم‌های هوش مصنوعی
- سرقت نرم افزار
- تجاوز به حریم خصوصی افراد

¹ Morgan

² Kahn

این سند معروف توسط سایبرسیتی‌زین شیب (cybercitizenship.org/ethics/commandments.html) منتشر شده است.

۱. شما نباید از رایانه برای آسیب رساندن به دیگران استفاده کنید.
۲. شما نباید در کار رایانه‌ی افراد دیگر دخالت کنید.
۳. شما نباید در فایل‌های دیگران جاسوسی نکنید
۴. شما نباید از رایانه برای سرقت استفاده کنید
۵. شما نباید از رایانه برای شهادت دروغ استفاده کنید
۶. شما نباید از نرم افزارهایی که برای آن پولی پرداخت نکرده‌اید استفاده کرده یا آنها را کپی کنید
۷. شما نباید بدون مجوز از منابع رایانه‌ای دیگران استفاده کنید.
۸. شما نباید خروجی فکری دیگران را تصاحب کنید.
۹. شما نباید به پیامدهای اجتماعی برنامه‌ای که می‌نویسید فکر کنید.
۱۰. شما نباید از رایانه به گونه‌ای استفاده کنید که نشان دهنده توجه و احترام باشد.

یکی از مسائل مهم آینده، اخلاق در خودروهای خودران است. به عنوان مثال، چه کسی آنها را توسعه خواهد داد، چگونه آنها را در وسایل نقلیه برنامه‌نویسی می‌کنند، و چگونه بکار گرفته می‌شوند؟ نگاه کنید به شارما^۱ (۲۰۱۷). برای بررسی ملاحظات اخلاقی در ادبیات تحقیق به nowpublishers.com/article/Details/ISY-012/ مراجعه کنید.

ام‌آی‌تی مدیا لب^۲ و مرکز اینترنت و جامعه دانشگاه هاروارد ابتکاری را برای تحقیق در مورد موضوعات اخلاقی و حکمرانی در هوش مصنوعی مدیریت می‌کنند. ساس یک فروشنده بزرگ ابزارهای تحلیل و هوش مصنوعی است که سه مرحله ضروری را برای اخلاق هوش مصنوعی پیشنهاد کرده است (sas.com/en_us/insights/articles/analytics/artificial-intelligence-ethics.html).

سوالاتی برای مرور بخش ۱۴.۳

۸. برخی از مسائل حقوقی سیستم‌های هوشمند را نام ببرید.
۹. نگرانی‌های مربوط به حریم خصوصی را در سیستم‌های هوشمند شرح دهید.
۱۰. از نظر شما، چه کسی باید اطلاعات استفاده شما از خودرو را در اختیار داشته باشد؟ چرا؟
۱۱. مسائل اخلاقی در سیستم‌های هوشمند را نام ببرید.
۱۲. فرمان سیستم‌های رایانه‌ای/اطلاعاتی را نام ببرید.

بسیاری از کارشناسان، مشاوران و پژوهشگران پیشنهاداتی را در مورد جاری‌سازی موفق سیستم‌های هوشمند ارائه می‌کنند. با توجه به اهمیت موضوع، واضح است که شرکت‌ها باید برای ظهور انبوه هوش مصنوعی و سایر فناوری‌های هوشمند آماده شوند. برخی از مسائل مربوط به استراتژی جاری‌سازی به شرح زیر هستند:

¹ Sharma

² MIT Media Lab

- چه زمانی باید پروژه‌های هوشمند را آغاز کرد و چگونه آنها را اولویت‌بندی کرد.
- چگونه تصمیم بگیرید که آیا این کار را خودتان انجام دهید یا از شرکا استفاده کنید، یا برون سپاری کنید.
- چگونه سرمایه‌گذاری در پروژه‌های هوشمند را توجیه کنید.
- چگونه بر مقاومت کارکنان غلبه کنیم (مثلاً ترس از دست دادن شغل).
- چگونه تیم‌های مناسب انسان-ربات را تشکیل دهیم.
- چگونه تعیین کنیم کدام تصمیمات باید به طور کامل توسط هوش مصنوعی خودکارسازی شوند.
- چگونه از سیستم‌های هوشمند (امنیت) و از حریم خصوصی محافظت کنیم.
- چگونه به از دست دادن شغل و بازآموزی کارکنان (بخش ۱۴.۵) رسیدگی کنیم.
- چگونه تعیین کنیم که آیا فناوری به روز لازم را ادر اختیار دارید یا خیر.
- چگونه تصمیم بگیریم که مدیریت ارشد چه حمایت‌هایی را باید ارائه دهد
- چگونه سیستم‌ها را با فرایندهای کسب‌وکار ادغام کنیم
- چگونه پرسنل پرسنل واجد شرایط را برای ساخت و استفاده از سیستم‌های هوشمند پیدا کنیم.

برای مسائل بیشتر در رابطه با استراتژی به کایرون^۱ (۲۰۱۷) مراجعه کنید. ما در این بخش فقط به چند موضوع می‌پردازیم و در خصوص سایر موارد شما را به منابع مختلف ارجاع می‌دهیم. اکثر موضوعات پیاده‌سازی ماهیت عمومی دارند و در اینجا به آنها پرداخته نخواهد شد.

چویی و همکاران (۲۰۱۷)، به نقل از مک‌کنزی آند کامپانی می‌گویند، «مدیران ارشد باید فرصت‌های تاکتیکی و استراتژیک (هوش مصنوعی) را درک کنند، سازمان‌های خود را دوباره طراحی کنند و متعهد شوند که به شکل‌دهی و بحث درباره آینده کار کمک کنند. به طور مشخص، مدیران باید برای ادغام سیستم‌های هوشمند در محل کار خود را برنامه‌ریزی کنند و متعهد شوند که یک محیط مشارکتی را برای تغییرات ایجاد کرده و منابع کافی را فراهم کنند. اسنایدر^۲ (۲۰۱۷) مدعی است که بسیاری از مدیران می‌دانند که سیستم‌های هوشمند کسب‌وکار آنها را تغییر خواهد داد، اما تلاش چندانی در مورد آن انجام نمی‌دهند. کی‌پی‌ام‌جی^۳، یک مشاور خدمات مدیریت ارشد، مراحل زیر را در مورد نیروی کار دیجیتال ارائه می‌دهد: «رویکرد جامع کی‌پی‌ام‌جی از استراتژی تا اجرا به شرکت‌ها در هر مرحله از پیاده‌سازی کمک خواهد کرد. مراحل عبارتند از:

- ایجاد زمینه‌های اولویت‌دار برای نوآوری‌های تکنولوژیکی
- تدوین استراتژی و برنامه برای کارکنان
- شناسایی عرضه‌کنندگان و شرکا برای اجرای طرح‌ها
- ایجاد یک استراتژی و برنامه‌ریزی برای تحقق منافع حاصل از ابتکار عمل‌های کار دیجیتال

Source: KPMG Internal Audit: Top 10 in 2018, Considerations for impactful internal audit

¹ Kiron

² Snyder

³ KPMG

یک راهنمای کامل کی‌پی‌ام‌جی توسط کایرون (۲۰۱۷) ارائه شده است. این راهنما شامل خودکارسازی فرایند رباتیک، خودکارسازی پیشرفته فرایند و خودکارسازی شناختی است. برای مسائل مربوط به رهبری در پیاده‌سازی به اینزورث^۱ (۲۰۱۷) مراجعه کنید.

از آنجایی که هوش مصنوعی و تحلیل کسب‌وکار اصطلاحات گسترده‌ای هستند که برای توصیف چندین فناوری که سطوح بلوغ آنها متفاوت است بکار می‌روند، مسائل پیاده‌سازی ممکن است کاملاً متفاوت باشند. شچوتسکایا^۲ (۲۰۱۷) سه مشکل اساسی زیر را بیان می‌کند:

۱. رویکرد توسعه. سیستم‌های تحلیلی کسب‌وکار و هوش مصنوعی نیازمند رویکردی متفاوت از سایر سیستم‌های رایانه‌ای/فناوری اطلاعات هستند. به طور خاص، شناسایی و برخورد با منابع داده‌های مختلف و اغلب بزرگ، امری ضروری است (به پرونده آغازین فصل ۱ و ۲ مراجعه کنید). پاکسازی و نگهداری این داده‌ها ضروری است. همچنین، اگر یادگیری درگیر باشد، باید از آموزش ماشینی استفاده کرد. بنابراین، روش‌های خاصی مورد نیاز است.
۲. یادگیری/از داده‌ها. بسیاری از ابزارهای تحلیل کسب‌وکار و هوش مصنوعی شامل یادگیری هستند. کیفیت داده‌های ورودی، کیفیت برنامه‌ها را تعیین می‌کند. همچنین مکانیسم یادگیری بسیار مهم است. بنابراین، دقت داده‌ها بسیار مهم است. در یادگیری، سیستم‌ها باید قادر به مقابله با شرایط متغیر محیطی باشند. داده‌ها باید در پایگاه‌های داده سازماندهی شوند نه در فایل‌ها.
۳. هیچ دیدگاه روشنی/از نحوه ایجاد بینش در دسترس نیست. هوش مصنوعی، اینترنت اشیاء و سیستم‌های تحلیل کسب‌وکار، بینش، نتیجه‌گیری و توصیه‌هایی را بر اساس تجزیه و تحلیل داده‌های جمع‌آوری شده تولید می‌کنند. با توجه به اینکه داده‌ها اغلب توسط انواع مختلف حسگرها جمع‌آوری می‌شوند، ممکن است دیدگاه روشنی از بینش‌های تولید شده نداشته باشیم. حوزه‌های مهم مرتبط شامل مشکلات مربوط به کلان‌داده‌ها، دسترسی ناکارآمد به اطلاعات، و قابلیت‌های محدود یکپارچه‌سازی (در ادامه بحث می‌شود) هستند.

به عنوان بخشی از فرایند توسعه، اتصال هوش مصنوعی و اپلیکیشن‌های تحلیلی به سیستم‌های فناوری اطلاعات موجود از جمله اینترنت و سایر سیستم‌های هوشمند امری ضروری است.

دولت استرالیا در آگوست ۲۰۱۷ مایکروسافت را مأمور ساخت مناطق ابر در مقیاس بزرگ کرد تا بتواند از قدرت فناوری‌های هوشمند بهره‌بردار. انتظار می‌رود این سیستم به طور چشمگیری نحوه پردازش داده‌ها و ارائه خدمات توسط دولت را مدرن کند. شهروندان می‌توانند داده‌های طبقه‌بندی نشده و محافظت‌نشده‌ی این سیستم را مدیریت کنند. زیرساخت در داخل یا نزدیک

¹ Ainsworth

² Shchutskaya

مراکز داده دولتی ساخته شده است. این سیستم دولت را قادر می‌سازد تا از برنامه‌های نوآورانه مبتنی بر یادگیری ماشین، ربات‌ها و ترجمه زبان استفاده کند و مراقبت‌های بهداشتی، آموزشی، خدمات اجتماعی و سایر عملکردهای دولت‌ها را بهبود بخشد. در نهایت، سیستم هم امنیت و هم حفاظت از حریم خصوصی را افزایش می‌دهد. یکپارچه‌سازی تقریباً لازمی کار با هر سیستمی است که تحت تأثیر ابزارهای تحلیل هوش مصنوعی یا کسب‌وکار است می‌باشد. به عنوان مثال، لازم است که برنامه‌های کاربردی هوشمند هم در استراتژی بازاریابی دیجیتال و هم برای اجرای بازاریابی یکپارچه شوند. برای بحث، نگاه کنید searchenginejournal.com/artificial-intelligence-marketing/200852/.

برای غلبه بر مشکل یکپارچه‌سازی، هوای چین (یک تولید کننده تلفن همراه) در حال نصب یک سیستم هوش مصنوعی با پایگاه دانش خود در داخل تراشه‌های محصولات خود است. سازندگان دیگر گوشی‌ها برای تعامل با دانش هوش مصنوعی به اتصال به «ابر» متکی هستند. برای پیامدهای اتصال به اینترنت اشیاء به رینی آند آندرسون (۲۰۱۷) نگاه کنید. برای ملاحظات در مورد ارائه دهندگان اتصال اینترنت اشیاء، به بارودی و همکاران ۲۰۱۸ مراجعه کنید.

بسیاری از برنامه‌های کاربردی هوشمند در «ابر» مدیریت و به روز شده و/یا به اینترنت معمولی متصل می‌شوند. متأسفانه با اتصال به اینترنت ممکن است آسیب‌پذیری‌های جدیدی را به وجود آورد. هکرها از فناوری‌های هوشمند برای شناسایی این آسیب‌پذیری‌ها استفاده می‌کنند. برای چگونگی استفاده مجرمان از هوش مصنوعی و مسائل مربوط به آن، به کراسمن^۱ (۲۰۱۷) مراجعه کنید. در بخش ۱۴.۷، خطرات بالقوه رباتیک را مورد بحث قرار می‌دهیم. ایمنی سرنشینان خودروهای خودران و سایرینی که ممکن است درگیر برخورد با خودروهای خودران باشند از اهمیت بالایی برخوردار است. مسئله ایمنی نیز همچنین، ایمنی افرادی که در نزدیکی ربات‌ها کار می‌کنند، برای چندین دهه مورد تحقیق قرار گرفته است. علاوه بر این، هک ربات‌ها، چت‌بات‌ها و سایر سیستم‌های هوشمند حوزه‌هایی هستند که نیاز به توجه دارند. در نهایت، ایمنی خود ربات‌ها هنگام کار در خیابان‌ها یک مسئله اساسی است. برخی از افراد به آنها حمله می‌کنند (مک فارلند، ۲۰۱۷).

بسته به ماهیت برنامه‌ها، روش‌های زیادی برای استفاده از سیستم‌های هوشمند وجود دارد. گتلیف^۲ (۲۰۱۷) راه‌های زیر را برای بهره‌گیری از قابلیت‌های فناوری هوشمند جهت افزایش کارایی و ارائه خدمات بیشتر به مشتری پیشنهاد کرد:

۱. تجربه مشتری را سفارشی کنید (مثلاً برای تعامل با مشتریان).
۲. تعامل با مشتری را افزایش دهید (به عنوان مثال، از طریق چت‌بات‌ها)
۳. از فناوری‌های هوشمند برای تشخیص مشکلات و ناهنجاری‌ها در داده‌ها استفاده کنید.

¹ Crosman

² Catliff

سینگ^۱ (۲۰۱۷) موارد زیر را به عنوان عوامل حیاتی موفقیت توصیه می‌کند: کشف، پیش‌بینی، توجیه و یادگیری از تجربه. راس^۲ (۲۰۱۷) موضوع نیاز به ارتقای مهارت‌های کارکنان و ایجاد نیروی کار توانمند با هوش مصنوعی را مطرح کرد. یکی از مهمترین موضوعات این است که چگونه می‌توان با ترس از دست دادن شغل کارکنان مقابله کرد. این در بخش ۱۴.۶ مورد بحث قرار گرفته است.

بیشتر مسائل مربوط به پذیرش سیستم‌های هوشمند مانند هر سیستم اطلاعاتی مشابه دیگر است. به عنوان مثال، کارکنان ممکن است در برابر تغییر مقاومت کنند، مدیریت ممکن است منابع کافی را فراهم نکند، ممکن است عدم برنامه ریزی و هماهنگی وجود داشته باشد و غیره. برای مقابله با چنین مسائلی، مورگان استنلی^۳ از صدها گفتگو و مصاحبه با متخصصان ایده‌هایی را استخراج کرد (به دی کامیلو^۴، ۲۰۱۸ مراجعه کنید). یکی از مسائل مهم داشتن یک استراتژی مناسب اتخاذ و جاری‌سازی استراتژی است که باید با فناوری‌های اجرا شده و افراد درگیر هماهنگ باشد. به طور کلی، رویکرد پذیرش عمومی برای سیستم‌های اطلاعاتی در اینجا نیز صادق است.

سوالاتی برای مرور بخش ۱۴.۴

۳۰. فرایند جاری‌سازی سیستم‌ها را شرح دهید.
۳۱. درباره‌ی نقش مدیریت ارشد در استقرار سیستم‌های هوشمند بحث کنید.
32. چرا اتصال موضوع بسیار مهمی است؟
33. مسائل و مشکلات توسعه‌ی سیستم را شرح دهید.
34. در مورد اهمیت امنیت و ایمنی و نحوه‌ی محافظت از آنها بحث کنید.
35. برخی از مسائل مربوط به پذیرش سیستم‌های هوشمند را شرح دهید.

سیستم‌های هوشمند اجزای مهمی در انقلاب اطلاعات و دانش هستند. برخلاف انقلاب‌های آهسته‌تر گذشته، مانند انقلاب صنعتی، این انقلاب به سرعت در حال وقوع است و بر هر جنبه‌ای از کار و زندگی ما تأثیر می‌گذارد. تأثیر بر سازمان‌ها، صنایع و مدیران که برخی از آنها در این بخش توضیح داده شده است جزو لاینفک و ذات این سیستم‌ها به شمار می‌روند. تفکیک تأثیر سیستم‌های هوشمند از سایر سیستم‌های رایانه‌ای کار دشواری است، به ویژه به دلیل گرایش به سمت یکپارچه‌سازی، یا حتی تعبیه کردن سیستم‌های هوشمند در سایر سیستم‌های اطلاعاتی مبتنی بر رایانه. سیستم‌های هوشمند می‌توانند پیامدهای خرد و کلان داشته باشند. چنین سیستم‌هایی می‌توانند بر افراد و مشاغل خاص و همچنین کار و ساختار بخش‌ها و واحدهای درون یک سازمان تأثیر

¹ Dingh

² Ross

³ Morgan Stanley

⁴ DiCamillo

بگذارند. آنها همچنین می‌توانند اثرات بلندمدت قابل توجهی بر کل ساختارهای سازمانی، کل صنایع، جوامع و کل یک جامعه داشته باشند (به عنوان مثال، در مورد تأثیر کلان، به بخش‌های ۱۴.۶ و ۱۴.۷ مراجعه کنید). رشد انفجاری ابزارهای تحلیل، هوش مصنوعی و رایانش شناختی تأثیر زیادی بر آینده سازمان‌ها خواهد داشت. تأثیر رایانه‌ها و سیستم‌های هوشمند را می‌توان به سه دسته کلی سازمانی، فردی و اجتماعی تقسیم کرد. در هر یک از اینها، رایانه‌ها ممکن است تأثیرات احتمالی زیادی داشته باشند. ما نمی‌توانیم همه آنها را در این کتاب مورد بررسی قرار دهیم. بنابراین در پاراگراف‌های بعدی به موضوعاتی که احساس می‌کنیم بیشتر مربوط به سیستم‌ها و سازمان‌های هوشمند هستند می‌پردازیم.

یک تغییر در ساختار سازمانی امکان ایجاد یک بخش تجزیه و تحلیل، یک بخش هوش تجاری، یک بخش علم داده‌ها و/یا یک بخش هوش مصنوعی است که در آن تجزیه و تحلیل نقش اصلی را ایفا می‌کند. چنین واحدهای ویژه‌ای (از هر نوع) می‌توانند با یک واحد تجزیه و تحلیل کمی ترکیب یا جایگزین شوند، یا می‌توانند یک موجودیت کاملاً جدید باشند. برخی از شرکت‌های بزرگ دارای واحدها یا بخش‌های پشتیبانی تصمیم‌گیری جداگانه هستند. به عنوان مثال، بسیاری از بانک‌های بزرگ دارای چنین بخش‌هایی در بخش خدمات مالی خود هستند. بسیاری از شرکت‌ها دارای بخش‌های کوچک علم داده‌ها یا واحدهای انبارش داده هستند. این نوع دپارتمان‌ها معمولاً علاوه بر فعالیت‌های مشاوره و توسعه اپلیکیشن، درگیر آموزش نیز می‌باشند. سایر شرکت‌ها به توانمندسازی یک مدیر ارشد فناوری در زمینه هوش تجاری، سیستم‌های هوشمند و برنامه‌های کاربردی تجارت الکترونیکی پرداخته‌اند. شرکت‌هایی مانند تارگت و وال‌مارت سرمایه‌گذاری‌های بزرگی در چنین واحدهایی کرده‌اند که به طور مداوم داده‌های خود را تجزیه و تحلیل می‌کنند تا با درک تعاملات مشتری و تامین‌کننده، کارایی بازاریابی و مدیریت زنجیره تامین را تعیین کنند. از سوی دیگر، بسیاری از شرکت‌ها تخصص‌های علم تحلیل/داده را در حوزه‌های عملکردی مانند بازاریابی، مالی و عملیات بکار می‌گیرند. به طور کلی، در حال حاضر در این حوزه فرصت‌های شغلی قابل توجهی وجود دارد. برای بحث در مورد نیاز به مدیر ارشد داده، به ولدون (۲۰۱۸) مراجعه کنید. همچنین، نیاز به یک مدیر ارشد هوش مصنوعی توسط لاوسون (۲۰۱۷) مورد بحث قرار گرفته است. رشد هوش تجاری و ابزارهای تحلیل منجر به تشکیل واحدهای جدید در شرکت‌های فناوری اطلاعات نیز شده است. به عنوان مثال، چند سال پیش، آی‌بی‌ام یک واحد کسب‌وکار جدید را با تمرکز بر تجزیه و تحلیل تشکیل داد. این گروه شامل واحدهایی در حوزه‌های هوش تجاری، مدل‌های بهینه‌سازی، داده‌کاوی و عملکرد کسب‌وکار است. مهمتر از آن، این گروه نه تنها بر روی نرم افزار، بلکه عمدتاً بر روی خدمات/مشاوره متمرکز است.

یکی از تأثیرات عمده سیستم‌های هوشمند این است که کسب‌وکارها را به شکل دیجیتالی تبدیل می‌کنند. در حالی که سایر فناوری‌های اطلاعات چنین تحولاتی را برای سالیان متمادی به همراه داشته‌اند با ظهور فناوری‌های هوشمند و عمدتاً هوش مصنوعی، این تغییرات و تحولات شتاب بیشتری گرفته است. در بسیاری از موارد، هوش مصنوعی تنها یک ابزار حمایتی برای انسان است. با این حال، با افزایش توانایی هوش مصنوعی، ماشین‌ها می‌توانند وظایف بیشتری را به تنهایی یا به کمک انسان انجام دهند. واقعیت این است که هوش مصنوعی در حال حاضر در حال تغییر برخی از مشاغل است. همانطور که در فصل دیده می‌شود 2، هوش مصنوعی در حال حاضر در حال تغییر تمامی حوزه‌های عملکردی کسب‌وکار، به ویژه بازاریابی و امور مالی است. دامنه تأثیر از اتوماسیون کامل بسیاری از وظایف، از جمله وظایف مدیریتی، تا افزایش همکاری انسان و ماشین (فصل ۱۱) را در بر می‌گیرد.

توضیحات جامعی از چگونگی ایجاد تحول دیجیتال توسط هوش مصنوعی توسط داگرتی و ویلسون^۱ (۲۰۱۸) ارائه شده است که به این نتیجه رسیده‌اند که کسب‌وکارهایی که موفق به ایجاد تحول مبتنی بر هوش مصنوعی نشوند در یک موضع ضعف رقابتی قرار خواهند گرفت. باترا و همکاران^۲ (۲۰۱۸) به پدیده مشابهی اشاره می‌کنند و از شرکت‌ها می‌خواهند از هوش مصنوعی برای ایجاد موجی از نوآوری‌ها استفاده کنند. برای اطلاعات بیشتر در مورد این موضوع، نگاه کنید به یوزیالکو^۳ (۲۰۱۷).

استفاده از فناوری‌های هوشمند و به ویژه هوش مصنوعی در بسیاری از موارد مشهود است. به عنوان مثال، آمازون. کام با استفاده از ربات‌ها، توانسته است تا هزینه را کاهش داده و تجارت آنلاین را کنترل کند. به طور کلی، با کاهش هزینه‌ها، افزایش تجارب مشتری، بهبود کیفیت و تسریع در تحویل، شرکت‌ها به مزیت رقابتی دست خواهند یافت. ریکرت^۴ (۲۰۱۷) گفتگوهایش با مدیران عامل در خصوص اینکه هوش مصنوعی و یادگیری ماشینی چگونه می‌توانند به شکست رقبا کمک کنند توضیح می‌دهد. آندرونیک^۴ (۲۰۱۷) به مزیت رقابتی اشاره می‌کند. مزایای آن عبارتند از ایجاد تقاضای بیشتر (به فصل ۲ مراجعه کنید)، فروش خودکار (فصل ۲) و شناسایی فرصت‌های فروش. یکی از عوامل مهم جدید این است که شرکت‌های جدید و محو شدن مرز بین حوزه‌ها بر تصویر رقابتی بسیاری از صنایع تأثیر می‌گذارد. به عنوان مثال، وسایل نقلیه خودران بر رقابت در صنعت خودرو تأثیر می‌گذارد. طبق گفته‌های ولدون (۲۰۱۷)، بکارگیری هوشمندانه‌ی ابزارهای تحلیلی منجر به مزیت رقابتی بسیار خوبی می‌شود. یک مثال از نحوه‌ی بکارگیری تحلیل‌ها، هوش مصنوعی و سایر فناوری‌های هوشمند برای دستیابی به یک مزیت رقابتی در **1-800-Flowers.com** در مورد کاربردی ۱۴.۱ شرح داده شده است.

1-800-Flowers.com چگونه از سیستم‌های هوشمند برای مزیت رقابتی

استفاده می‌کند.

مورد کاربردی ۱۴.۱

1-800-Flowers.com یک خرده‌فروش آنلاین پیشگام گل و گیاه و هدیه است. این شرکت در اواسط دهه ۹۰ سیستم خود را از سفارش تلفنی به سفارش آنلاین تغییر داد. از آن زمان، علیرغم رقابت شدید در این بازار، درآمد شرکت به بیش از ۱ میلیارد دلار و تعداد کارکنان آن به بیش از ۴,۰۰۰ نفر افزایش یافته است. در دنیای کسب‌وکار که تحت سلطه‌ی غول‌های آنلاین مثل آمازون. کام و وال‌مارت. کام و صدها شرکت دیگر است فروش آنلاین گل و هدیه کار چندان آسانی نیست. این شرکت از سه استراتژی کلیدی زیر استفاده می‌کند:

- ارتقاء و بهبود تجربه‌ی مشتری
- تحریک مؤثر تقاضا
- ایجاد یک نیروی کار برای پشتیبانی از محصولات و نوآوری خدمات (فرهنگ نوآوری).

¹ Dougherty and Wilson

² Batra et al.

³ Uzialko

⁴ Andronic

این شرکت به طور گسترده‌ای از فناوری‌های هوشمند برای ساخت یک زنجیره‌ی تأمین فوق‌العاده و تسهیل همکاری‌ها استفاده کرده است. اخیراً، این شرکت شروع به استفاده از سیستم‌های هوشمن برای ارتقاء استراتژی‌های رقابتی خود نمود. در این کتاب به چند فناوری که این شرکت استفاده کرده است اشاره کرده‌ایم.

۱. **بهینه‌سازی تجربه‌ی مشتری.** این خرده‌فروشی با بهره‌گیری از محصولات مدیریت داده‌ها و خودکارسازی بازاریابی ساس^۱ به گردآوری اطلاعات مربوط به نیازهای مشتریان و تحلیل آنها می‌پردازد. این اطلاعات شرکت را قادر می‌سازد تا بهترین گل و هدایا را برای هر مناسبی یافته و ارسال کنند. از آنجائیکه ارسال‌کنندگان هدایا می‌خواهند دریافت‌کنندگان را خوشحال کنند، ارائه‌ی توصیه‌های مناسب بسیار حیاتی است. شرکت از ابزارهای تحلیلی پیشرفته و داده‌کاوی ساس برای پیش‌بینی نیازهای مشتریان استفاده می‌کند. بازاریاب‌های شرکت می‌توانند ارتباطات مؤثرتری با مشتریان برقرار کنند. دانشمندان علم داده‌ها و تحلیلگران بازاریابی شرکت با استفاده از جدیدترین ابزارها می‌توانند داده‌کاوی‌های اثربخش‌تری انجام دهند. امروزه انتظارات مشتریان فراتر از قبل است زیرا مشتریان راحت‌تر می‌توانند محصولات و خدمات عرضه شده را به صورت آنلاین مقایسه کنند. ابزارهای تحلیل و هوش مصنوعی شرکت را قادر می‌سازد تا عقاید و احساسات مشتریان را درک کند. امروزه شرکت قادر است تا رفتارهای استدلال احساسی برای تصمیم‌گیری خرید و وفاداری مشتری را درک کند. این تغییر منجر به توصیه‌هایی برای محصول می‌شود که بعداً توضیح خواهیم داد.

۲. **چت‌بات‌ها.** **1-800-Flowers.com** یک بات بر روی فیسبوک مسنجر دارد. همانطور که در فصل ۱۲ توضیح داده شد، اینگونه بات‌ها می‌توانند به عنوان یک منبع اطلاعاتی و وسیله‌ای برای گفتگو و صحبت مفید باشند. این شرکت همچنین چت‌بات خود را بر روی وبسایت خود به صورت آنلاین و صوتی قرار داده است. علاوه بر این، مشتریان سراسر می‌توانند از دستیار گوگل برای سفارش‌دهی صوتی استفاده کنند. شرکت همچنین از الکسا صوتی با قابلیت «one-shot intent» برای امکان سفارش‌گذاری متنوع‌تر استفاده می‌کند.

۳. **خدمات مشتری.** این شرکت یک پورتال و امکان خرید وان-استاپ شاپینگ^۲ را مشابه آمازون. کام ارائه می‌کند و امکان پرداخت سلف-سرویس نیز وجود دارد. همین قابلیت در زمان خرید با کمک بات شرکت در فیسبوک مسنجر برای تکمیل سفارش وجود دارد. نیازی نیست که مشتریان برای تکمیل سفارش از فیسبوک خارج شوند.

۴. **توصیه‌های مبتنی بر هوش مصنوعی.** همانطور که از فصل ۱۲ به خاطر دارید، خرده‌فروشان تجارت الکترونیک با ارائه‌ی توصیه‌هایی برای محصول (مثلاً آمازون، نتفلیکس) متعالی شده‌اند. **1-800-Flowers.com** نیز با ارائه‌ی توصیه و مشاوره در مورد برندهای هدایای خود در وبسایتش (مثلاً هری آند دیوید^۳) همین کار را می‌کند. توصیه‌ها توسط آی‌بی‌ایم واتسون ایجاد شده و به عنوان یک «دستیار مهمان شناختی»^۴ ارائه می‌شوند که تجربه‌ی خرید آنلاین را به یک تجربه‌ی خرید از داخل فروشگاه تبدیل می‌کنند. این سرویس مبتنی بر هوش مصنوعی جی‌دبلیووای‌ان^۵ نام دارد. با استفاده از سیستم پردازش زبان طبیعی واتسون، مکالمات بین انسان و ماشین به راحتی امکان‌پذیر می‌شود.

۵. **شخصی‌سازی.** دپارتمان بازاریابی شرکت به کمک ابزارهای تحلیل پیشرفته‌ی ساس به بخش‌بندی مشتریان در قالب گروه‌هایی با ویژگی‌های مشابه پرداخته است. شرکت سپس تبلیغات هدفمند را برای هر بخش از مشتریان ارسال می‌کند.

¹ SAS Marketing Automation and Data Management

² One-Stop Shopping

³ Harry & David

⁴ Cognitive Concierge

⁵ GWYN (Gifts When You Need)

علاوه بر ایمیل، از کمپین‌های تبلیغاتی ویژه نیز استفاده می‌شود. شرکت بر اساس بازخوردها می‌تواند استراتژی بازاریابی خود را طرح‌ریزی یا بازنگری کند. ساس نیز به شرکت در تحلیل «پسندها» و «ناپسندها» مشتریان کمک می‌کند. از همه‌ی اینها گذشته، سیستم‌های هوشمند به شرکت و مشتریانش در اتخاذ تصمیمات آگاهانه کمک می‌کنند.

سوالاتی برای مورد کاربردی ۱۴.۱

۹. چرا امروزه ارائه‌ی تجربه‌ی بهتر به مشتری ضروری است؟

۱۰. چرا داده‌ها نیازمند ابزارهای تحلیلی پیچیده هستند؟

۱۱. مقاله‌ی “Key Benefits of SAS Marketing Automation” را مطالعه کنید. به نظر شما **1-800-Flowers.com**

از کدام مزایای بهره برده است و چرا؟

۱۲. ارتباط آی‌بی‌ام واتسون با «شخصی‌سازی» چیست؟

۱۳. ارتباط قابلیت‌های «ابزارهای تحلیلی پیشرفته‌ی ساس» با کاربرد آنها در این مورد کاربردی چیست؟

۱۴. از «ساس اینترپرایز ماینر^۱» برای داده‌کاوی استفاده می‌شود. توضیح دهید چه کاری و چگونه انجام می‌شود؟

۱۵. ساس محصولی به نام «اینترپرایز گاید^۲» دارد که **1-800-Flowers.com** از آن استفاده می‌کند. بر اساس قابلیت‌های این ابزار ببینید چگونه از آن استفاده می‌کند.

Sources: Compiled from J. Keenan. (2018, February 13). “**1-800-Flowers.com** Using Technology to Win Customers’ Hearts This Valentine’s Day.” *Total Retail*; S. Gaudin. (2016, October 26). “1-800-Flowers Wants to Transform Its Business with A.I.” *Computer World*; SAS. (n.d.). “Customer Loyalty Blossoms with Analytics.” *SAS Publication*, sas.com/en_us/customers/1-800-flowers.html/ (accessed July 2018).

یکی از حوزه‌های نوظهور پژوهشی و کاربردی، استفاده از فناوری‌های علم داده‌ها برای مطالعه پویایی‌های سازمانی، رفتار پرسنل و طراحی مجدد سازمان برای دستیابی بهتر به اهداف است. در واقع، چنین برنامه‌هایی تحت عنوان *ابزارهای تحلیل مردم*^۳ شناخته می‌شوند. به عنوان مثال، ابزارهای تحلیلی دپارتمان‌های منابع انسانی برای شناسایی نامزدهای ایده‌آل از مجموعه‌ای از رزومه‌های ارسالی به سازمان یا حتی از مجموعه‌های گسترده‌تری مانند لینکدین استفاده می‌کنند. توجه داشته باشید که با هوش مصنوعی و ابزارهای تحلیل، مدیران می‌توانند گستره بیشتری داشته باشند به عنوان مثال، به دلیل توصیه‌هایی که مدیران و کارکنان می‌توانند از دستیاران مجازی دریافت کنند، می‌توانند کنترل بیشتری داشته باشند. افزایش دامنه کنترل می‌تواند منجر به ساختارهای سازمانی تحت شود. همچنین، شرح وظایف مدیران ممکن است تغییر کند.

یک حوزه کاربردی جالب‌تر و جدیدتر به درک رفتار کارکنان با نظارت بر حرکات آنها در سازمان و استفاده از آن اطلاعات برای طراحی مجدد چیدمان یا تیم‌ها برای دستیابی به عملکرد بهتر مربوط می‌شود. یک شرکت به نام هیومنایز^۴ (قبلاً به عنوان سوشیومتریک سولوشنز^۴ شناخته می‌شد) دارای نشان‌هایی مجهز به جی‌پی‌اس و حسگر می‌باشد. وقتی کارمندان این نشان‌ها را بر لباس‌شان نصب می‌کنند، تمام حرکت آنها ثبت می‌شود. طبق گزارش‌ها هیومنایز توانسته است به شرکت‌ها در پیش بینی

¹ SAS Enterprise Miner

² Enterprise Guide

³ People Analytics

⁴ Sociometric Solutions

اینکه کدام کارمندان بر اساس تعامل با سایر کارمندان ممکن است در شرکت بمانند یا شرکت را ترک کنند، کمک کند. به عنوان مثال، آن دسته از کارمندانی که در اتافک خود می‌مانند نسبت به افرادی که به طور گسترده در حال حرکت هستند و با سایر کارمندان تعامل دارند، احتمال کمتری برای ترقی و پیشرفت در نردبان شرکت را دارند. جمع آوری و تجزیه و تحلیل داده‌های مشابه به سایر شرکت‌ها کمک کرده است تا اندازه اتافک‌های کنفرانس مورد نیاز یا حتی چیدمان دفتر را برای دستیابی به حداکثر بهره‌وری تعیین کنند. طبق وب سایت هیومنایز، شرکتی به دنبال درک بهتر ویژگی‌های رهبران خود بود. با تجزیه و تحلیل داده‌های این نشان‌ها، شرکت توانست تشخیص دهد که رهبران موفق در واقع شبکه‌های ارتباطی گسترده‌تری دارند که با آن تعامل داشته و زمان بیشتری را صرف تعامل با دیگران می‌کنند و همچنین از نظر فیزیکی فعال هستند. اطلاعات جمع‌آوری‌شده از میان رهبران تیم برای طراحی مجدد فضای کار و کمک به بهبود عملکرد سایر رهبران استفاده شد. روشن است که این روش می‌تواند ملاحظات حریم خصوصی را بوجود آورد، اما در یک سازمان، چنین مطالعاتی ممکن است قابل قبول باشد. وب سایت هیومنایز چندین مطالعه موردی جالب دیگر دارد که مثال‌هایی از نحوه استفاده از فناوری‌های کلان‌داده‌ها برای توسعه ساختارهای تیمی و طراحی سازمانی کارآمدتر را ارائه می‌دهند.

اگرچه فناوری‌های هوشمند ممکن است موجب غنی‌تر شدن بسیاری از مشاغل شوند، اما این تأثیر و رضایت ممکن است در مشاغل دیگر چندان زیاد نباشد. برخی معتقدند که سیستم‌های اطلاعاتی رایانه‌ای به طور کلی ممکن است اختیارات مدیریتی را در تصمیم‌گیری کاهش دهند و مدیران را به نارضایتی سوق دهند. با این حال، مطالعات سیستم‌های تصمیم‌گیری خودکار نشان داده است که کارکنانی که از چنین سیستم‌هایی استفاده می‌کنند، به‌ویژه آنهایی که این سیستم‌ها موجب توانمندی آنها شده‌اند از شغل خود رضایت بیشتری داشتند. اگر بتوان از یک سیستم هوش مصنوعی برای انجام وظایف روزمره و روتین استفاده کرد، وقت آزاد بیشتری را برای مدیران و کارکنان دانش برای انجام وظایف چالش برانگیزتر فراهم می‌کند. مهمترین وظیفه‌ی مدیران تصمیم‌گیری است. فناوری‌های هوشمند می‌توانند شیوه تصمیم‌گیری‌های بسیاری و در نتیجه مسئولیت‌های شغلی مدیران را تغییر دهند. به عنوان مثال، برخی از محققان دریافته‌اند که یک سیستم پشتیبانی تصمیم موجب بهبود عملکرد مدیران فعلی و جدید و همچنین سایر کارکنان شده است. این امر به مدیران کمک کرد تا دانش، تجربه و تخصص بیشتری کسب کنند و در نتیجه کیفیت تصمیم‌گیری آنها افزایش می‌یابد. بسیاری از مدیران گزارش می‌دهند که سیستم‌های هوشمند نهایتاً به آنها زمان داده‌اند تا از دفتر خارج شده و وارد میدان شوند. آنها همچنین دریافته‌اند که می‌توانند زمان بیشتری را صرف برنامه‌ریزی فعالیت‌ها کنند زیرا می‌توانند به لطف فناوری‌های سیستم هوشمند پیشاپیش از مشکلات احتمالی آگاه شوند (به پرونده‌ی آغازین فصل ۱ مراجعه کنید).

یک جنبه دیگر چالش مدیریتی در توانایی فناوری‌های هوشمند برای حمایت از فرایند تصمیم‌گیری به طور عام و برنامه‌ریزی استراتژیک و تصمیمات کنترلی به طور خاص نهفته است. سیستم‌های هوشمند می‌توانند فرایند و حتی سبک‌های تصمیم‌گیری را تغییر دهند به عنوان مثال، جمع‌آوری اطلاعات برای تصمیم‌گیری زمانی که الگوریتم‌ها در حال استفاده هستند، بسیار سریع‌تر انجام می‌شود. تحقیقات نشان می‌دهد که اکثر مدیران تمایل دارند تا روی تعداد زیادی از مشکلات به طور همزمان کار کنند و از یک مشکل به مشکل دیگری تمرکز کنند در حالی که منتظر اطلاعات بیشتر در مورد مشکل فعلی خود هستند. فناوری‌های هوشمند تمایل دارند با ارائه دانش و اطلاعات، زمان لازم برای انجام وظایف را در فرایند تصمیم‌گیری کاهش دهند و برخی از زمان‌های انتظار غیرمولد را حذف کنند. در زیر برخی از تأثیرات احتمالی سیستم هوشمند بر مشاغل مدیران آورده شده است:

- تصمیم‌گیری‌های زیاد به تخصص (تجربه) کمی نیاز دارد.
- به دلیل در دسترس بودن اطلاعات و خودکارسازی بخشی از گام‌های فرایند تصمیم‌گیری، تصمیم‌گیری سریع‌تر امکان‌پذیر است.
- برای پشتیبانی از مدیران ارشد کمتر به کارشناسان و تحلیلگران اتکار می‌شود. امروزه آنها می‌توانند خودشان به کمک سیستم‌های هوشمند تصمیم‌گیری کنند.
- قدرت بین مدیران مجدداً توزیع می‌شود (هرچه از اطلاعات و قابلیت تحلیل بیشتری برخوردار باشند، قدرت بیشتری در اختیار خواهند داشت).
- پشتیبانی برای تصمیمات پیچیده امکان توسعه‌ی سریع‌تر و با کیفیت‌تر راه‌حل‌ها را فراهم می‌کند.
- اطلاعات مورد نیاز برای تصمیم‌گیری سطح بالا به خودی خود تولید می‌شود.
- اتوماسیون تصمیمات یا فازهای روتین در فرایند تصمیم‌گیری (مثلاً برای تصمیم‌گیری خط مقدم و استفاده از تصمیم‌گیری خودکار) ممکن است منجر به حذف برخی مدیران شود.

Source: Decision Support And Business Intelligence Systems, Pearson Education India, 2008.

به طور کلی مشخص شده است که شغل مدیران میانی محتمل‌ترین شغل برای خودکار شدن است. مدیران سطح متوسط تصمیمات نسبتاً معمولی می‌گیرند که می‌تواند کاملاً خودکار باشد. مدیران سطوح پایین‌تر زمان زیادی را صرف تصمیم‌گیری نمی‌کنند.

در عوض، آنها بر سطوح غیر مدیریت نظارت کرده و آنها را آموزش داده و تشویق می‌کنند. برخی از تصمیمات معمول آنها، مانند برنامه‌ریزی، می‌تواند خودکار باشد. تصمیمات دیگری که شامل جنبه‌های شناختی هستند ممکن است خودکار نباشند. با این حال، حتی اگر نقش تصمیم‌گیری مدیران کاملاً خودکار باشد، بسیاری از فعالیت‌های دیگر آنها را نمی‌توان خودکار کرد یا فقط می‌توان تا حدی خودکار کرد.

در سراسر کتاب، ما نشان می‌دهیم که چگونه فناوری‌های هوشمند موجب بهبود یا خودکارسازی تصمیم‌گیری می‌شوند. البته این فناوری‌ها بر شغل مدیران تأثیر می‌گذارند. یکی از جنبه‌های آن تأثیر فناوری‌های هوشمندی است که توسط «ابر» پشتیبانی می‌شوند. یک مثال در فصل ۹، شکل ۹.۱۲ نشان داده شده است. این جریان داده‌ها را از منابع داده و خدمات از طریق یک سرویس اطلاعاتی به سرویس‌های تحلیلی برای انواع مختلف تصمیم‌گیری که توسط ابزارهای تحلیلی پشتیبانی می‌شود، نشان می‌دهد. اوزبالکو (۲۰۱۷) توضیح می‌دهد که چگونه انسان‌ها می‌توانند از هوش مصنوعی برای پیش‌بینی و تجزیه و تحلیل عواقب راه‌حل‌های بالقوه مختلف استفاده کنند و فرایند تصمیم‌گیری را ساده‌تر کنند. همچنین با استفاده از یادگیری ماشینی و یادگیری عمیق می‌توان تصمیمات بیشتری را به صورت خودکار

درآورد. یکی از تأثیرات سیستم‌های هوشمند، پشتیبانی بلادرنگ از تصمیم‌گیری است. یک ابزار محبوب برای انجام این کار، ساس دیسیژن منیجر^۱ است که در بینش فناوری ۱۴.۲ توضیح داده شده است.

بینش فناوری ۱۴.۲ ساس دیسیژن منیجر

ساس ریپل تایم دیسیژن منسجر (آرتی‌دی‌ام)^۲ یک محصول یکپارچه مبتنی بر تجزیه و تحلیل است که برای پشتیبانی بلادرنگ از تصمیم‌گیری در طراحی شده است که برای کمک به شرکت‌ها در واکنش سریع به تغییرات بازاریابی، نیازهای مشتریان، فناوری و سایر محیط‌های کسب‌وکار ضروری است. ساس به سوالات زیر پاسخ می‌دهد:

۱. **ساس آرتی‌دی‌ام چه کاری انجام می‌دهد؟** ساس، ابزار تحلیل را با منطق کسب‌وکار و استراتژی‌ها ترکیب می‌کند تا توصیه‌ها و تصمیمات بلادرنگ بهبودیافته‌ای را به کانال‌های تعامل با مشتریان مانند وب سایت‌ها، مراکز پاسخگویی، محل پایانه‌های فروش و دستگاه‌های خودپرداز ارسال کند.
۲. **چرا ساس آرتی‌دی‌ام مهم است؟** این به شما کمک می‌کند تا با خودکارسازی و بکارگیری ابزارهای تحلیل در فرایند تصمیم‌گیری در طول تعاملات آنی با مشتری، تصمیمات هوشمندانه‌تری بگیرید. با برآورده کردن موفق نیازهای خاص هر مشتری در زمان مناسب، مکان مناسب و در شرایط مناسب، کسب‌وکار شما می‌تواند سودآورتر شود.
۳. **ساس آرتی‌دی‌ام برای چه کسانی طراحی شده است؟** قابلیت‌های متمایزی را برای بازاریابانی که استراتژی‌های ارتباطی را تعریف می‌کنند، مدیرانی که به گزارش‌هایی اثربخشی بازاریابی نیاز دارند، تحلیلگران کسب‌وکار که رفتار مشتری را مدل‌سازی و پیش‌بینی می‌کنند، و مدیران کمپین که بخش‌های هدف مشتری را ایجاد می‌کنند، ارائه می‌کند.

مزایای کلیدی آرتی‌دی‌ام به شرح زیر است:

- هر بار و همیشه تصمیمات درستی می‌گیرد.
- نیازهای مشتری را با پیشنهاد مناسب، در زمان مناسب، از طریق کانال مناسب برآورده می‌کند.
- منابع ارزشمند فناوری اطلاعات را به شکل بهتری تخصیص می‌دهد.

ویژگی‌های کلیدی طبق شرکت ساس عبارتند از:

- تجزیه و تحلیل بلادرنگ
- ساخت سریع فرایند تصمیم‌گیری
- داده‌های سازمانی در سراسر سازمان
- تست کمپین‌ها
- فرایند تحلیلی خودآموز خودکار
- قابلیت اتصال

¹¹ SAS Decision Manager

² SAS Real-Time Decision Manager (RTDM)

برای جزئیات، به "SAS Real-Time Decision Manager" مراجعه کنید و متن آن را بخوانید. همچنین می‌توانید یک مقاله در مورد آر تی دی ام بارگذاری کنید

۱. اساس آر تی دی ام چه بهبودهایی در فرایند تصمیم‌گیری ایجاد می‌کند؟
۲. کدامیک از محصولات ساس به آر تی دی ام متصل شده یا در آن تعبیه شده‌اند؟
۳. ارتباط محصول با قابلیت توصیه‌ی محصول چگونه است؟

Source: SAS® Real-Time Decision Manager Make context-based marketing decisions during your real-time customer interactions. Copyright © 2018 SAS Institute Inc., Cary, NC, USA. All Rights Reserved. Used with permission.

تنی چند از نویسندگان شروع به گمانه زنی در مورد تأثیر هوش مصنوعی، تجزیه و تحلیل، و محاسبات شناختی بر آینده صنعت کرده‌اند. چند منبع جالب برای مشورت عبارتند از اوتور^۱ (۲۰۱۶) و رنسبوتهام^۲ (۲۰۱۶). یک گزارش ویژه توسط اکونومیست (استاندریج ۲۰۱۶) و یک کتاب از برینیولفسون و مک‌آفی^۳ (۲۰۱۶). گزارش اکونومیست بسیار جامع بوده و ابعاد بسیاری از تأثیر تحولات جاری بر صنعت و جامعه را در بر می‌گیرد. بحث‌های اصلی این است که فناوری در حال حاضر امکان انجام کارهای بیشتر و بیشتری را که قبلاً توسط انسان انجام می‌شدند توسط رایانه‌ها فراهم می‌کند. خودکارسازی کار البته قبلاً از زمان انقلاب صنعتی آغاز شده است. چیزی که این بار تغییر را به طور قابل توجهی گسترده‌تر می‌کند این است که این فناوری امکان انجام بسیاری از وظایف شناختی که باید توسط ماشین‌ها انجام شوند را میسر می‌کند و سرعت تغییر آنقدر ریشه‌ای و بنیادی است که تأثیر احتمالی آن بر سازمان‌ها و جامعه بسیار چشمگیر و در مواقعی غیرقابل پیش‌بینی خواهد بود.

این نویسندگان البته در پیش‌بینی‌های خود اتفاق نظر ندارند. اجازه دهید ابتدا روی تأثیرات سازمانی تمرکز کنیم. رنسبوتهام (۲۰۱۶)، استدلال می‌کند که محاسبات شناختی بسیاری از کارهای انجام شده توسط انسان را تبدیل به انجام رایانه‌ای می‌کند و در نتیجه هزینه‌های سازمان‌ها را کاهش می‌دهد. کیفیت خروجی ممکن است در کار شناختی نیز افزایش یابد که این امر در چندین مطالعه که عملکرد یک انسان را با یک ماشین مقایسه می‌کند، نشان داده شده است.

همه از پیروزی آی‌بی‌ام واتسون در مسابقه‌ی جنوپاردی! مطلع هستند یا سیستم گوگل در بازی «گو» برابر قهرمانان انسانی برنده شد. اما بسیاری از مطالعات دیگر در حوزه‌های خاص مانند تشخیص گفتار و تفسیر تصویر پزشکی که کار بسیار تخصصی و در عین حال روتین یا تکرارپذیر هستند نیز برتری مشابهی را در سیستم‌های خودکار نشان داده‌اند. از آنجایی که ماشین‌ها در هر زمان و مکانی در دسترس هستند، ممکن است دسترسی سازمان افزایش یافته و در نتیجه، مقیاس‌پذیری آسان‌تر و در نتیجه رقابت بین سازمان‌ها افزایش یابد. این تأثیرات سازمانی به این معنی است که سازمان‌های برتر دیروز ممکن است برای همیشه در صدر باقی نمانند، زیرا محاسبات شناختی و اتوماسیون می‌توانند بازیکنان فعلی را به چالش بکشند. این مورد در صنعت خودروسازی رخ داده است. اگرچه شرکت‌های خودروسازی سنتی در تلاش هستند تا به سرعت خودشان را بالا بکشند، گوگل،

¹ Autor

² Ransbotham

³ Brynjolfsson and McAfee

تسلا و سایر شرکت‌های فناوری با به چالش کشیدن رهبران عصر خودروسازی، ساختار صنعت را دچار نوآوری تحول‌آفرینی کرده‌اند. ابزارهای تحلیل و هوش مصنوعی عوامل ایجاد بسیاری از این تغییرات هستند.

سوالاتی برای مرور بخش ۱۴.۵

۱۴. تأثیرات سیستم‌های هوشمند بر وظایف مدیریتی را نام ببرید.
۱۵. واحدهای سازمانی جدید که به واسطه‌ی سیستم‌های هوشمند ایجاد می‌شوند را شرح دهید.
۱۶. مثال‌هایی از برنامه‌های کاربردی تجزیه و تحلیل و هوش مصنوعی که برای طراحی مجدد فضای کاری یا رفتار تیم استفاده می‌شوند را نام ببرید.
۱۷. محاسبات شناختی چگونه بر ساختار صنعت و رقابت در آن تأثیر می‌گذارد؟
۱۸. اثرات سیستم‌های هوشمند بر رقابت را شرح دهید.
۱۹. درباره تأثیر سیستم‌های هوشمند بر تصمیم‌گیری بحث کنید.

یکی از پربحث‌ترین موضوعات در بررسی تأثیرات سیستم‌های هوشمند تأثیر آنها بر مشاغل و کار است. در این خصوص یک اتفاق نظر کلی وجود دارد که:

- سیستم‌های هوشمند مشاغل جدید بسیاری را مانند اتوماسیون همیشه ایجاد خواهند کرد.
- نیاز به آموزش مجدد افراد زیادی وجود خواهد داشت
- ماهیت کار تغییر خواهد کرد.

بحث‌ها، مناظرات و اختلاف نظرهای مربوط به زمان، میزان و چگونگی برخورد با این پدیده‌ها، بسیاری از محققین را به خود مشغول کرده و از موضوعات این بخش می‌باشد.

طبق رنسبوتهام (۲۰۱۶)، مشاوره مالی به طور معمول یک کار دانشی سنگین و پیچیده محسوب می‌شود. از آنجایی که مشاوران ربات (روبو) از افراد پشتیبانی شخصی می‌کنند، هزینه‌های چنین خدماتی کاهش می‌یابد. این امر باعث می‌شود افراد بیشتری تقاضای چنین خدماتی را داشته باشند و در نهایت نیروی انسانی بیشتری را برای رسیدگی به مسائل مالی پیچیده‌تر و دشوارتر آزاد می‌کند. مشاوران روبو همچنین ممکن است باعث شوند برخی افراد شغل خود را از دست بدهند.

برخی نویسندگان استدلال می‌کنند که بخش خودکارسازی، که به محاسبات شناختی و هوش مصنوعی مربوط می‌شود، آنچه را که قطبی شدن بازار کار نامیده می‌شود در آینده تسریع خواهد کرد. این امر مستلزم رشد شغلی قابل توجهی در سطوح بالا و پایین بازار کار است اما در سطوح میانی موجب از دست رفتن مشاغل خواهد شد. مشاغلی که به مهارت‌های کم اما تخصصی نیاز دارند، مانند مراقبت شخصی، همچنان رد خواهند کرد. به همین ترتیب، مشاغلی که نیاز به مهارت بسیار بالایی دارند، مانند کارهای طراحی گرافیک و ... نیز در حال رشد هستند. اما مشاغلی که به «مهارت‌های متوسط» نیاز دارند، مانند دانش تخصصی

که بارها و بارها با تغییرات و سازگاری‌هایی مورد استفاده قرار می‌گیرند در معرض نابودی هستند. گاهی فناوری خودش را از بین می‌برد! به عنوان مثال، آی‌بی‌ام واتسون آنالیتیکز اکنون شامل قابلیت‌های پرس و جو برای شروع پرسیدن سؤالاتی است که قبلاً یک متخصص سیستم هوشمند قبلاً مطرح کرده است و بدیهی است که پاسخ‌هایی را ارائه می‌دهد. سایر محصولات تحلیل به عنوان یک سرویس با خدمات مشابه ممکن است منجر به نیاز افرادی شود که مهارت کمتری در استفاده از نرم‌افزارهای تحلیلی داشته باشند.

در گزارشی از اکونومیست اشاره می‌کند که حتی اگر هوش مصنوعی مستقیماً جایگزین کارگران نشود، مطمئناً کارمندان را ملزم به کسب مهارت‌های جدید برای حفظ شغل خود می‌کند. اختلال در بازار همیشه ناراحت کننده است چند سال آینده فرصت‌های بسیار خوبی را برای متخصصان فناوری هوشمند فراهم می‌کند تا آینده را شکل دهند.

ایلان ماسک از شرکت تسلا، کامیون‌های خودران مبتنی بر هوش مصنوعی را در سراسر جهان ظرف ۱۰ سال پیش‌بینی می‌کند. ناوگان‌هایی از چنین کامیون‌هایی وجود خواهد داشت که هر کدام از آنها یک کامیون سرگروه را دنبال می‌کنند. کامیون‌ها برقی، اقتصادی و بدون آلودگی خواهند بود. علاوه بر این، تصادفات کمتری رخ خواهد داد - عالی به نظر می‌رسد! اما در مورد هزاران راننده ای که شغل خود را از دست می‌دهند چه می‌شود؟ در مورد هزاران کارمند در ایستگاه‌های کامیون که آنها نیز شغل خود را از دست خواهند داد، چگونه؟ همین سناریو ممکن است در بسیاری از صنایع دیگر نیز اتفاق بیفتد. آمازون اولین فروشگاه فیزیکی «گو» را افتتاح کرده است. آنها در چند سال آینده ۳۰۰۰ مورد دیگر را برنامه‌ریزی می‌کنند. اداره پست در برخی از کشورها در حال حاضر نامه را با استفاده از وسایل نقلیه خودران توزیع می‌کند. به طور خلاصه، یک شانس بیکاری گسترده وجود دارد به عنوان

مثال: خلبانان در فِدِکس

فِدِکس^۱ ناوگانی متشکل از ۱۰۰۰ هواپیما دارد که در سراسر جهان پرواز می‌کنند. به گفته فرانک تود، سردبیر و ناشر ربات ریپورت، فدکس امیدوار است که در حدود سال ۲۰۲۰ این شرکت یک مرکز خلبانی جهانی با سه یا چهار خلبان داشته باشد که کل ناوگان فدکس را اداره کنند.

فاکس کام^۲، یک تولید کننده آیفون در تایوان، برنامه‌ریزی کرده بود که تقریباً تمام کارمندان خود (۶۰۰۰۰) در تایوان را با ربات‌ها جایگزین کند (باتن ۲۰۱۶) این شرکت تاکنون ۱۰۰۰۰ ربات برای این منظور تولید کرده است.

بحث در مورد اشتغال‌زایی فناوری از ابتدای انقلاب صنعتی در جریان بوده است. موضوع مربوط به سیستم‌های هوشمند به دلایل موارد زیر به شدت مورد بحث است:

- خیلی سریع حرکت می‌کنند

¹ Fedex

² Foxcom

- آنها ممکن است مشاغل متنوعی را انتخاب کنند، از جمله بسیاری از مشاغل یقه سفید و غیر فیزیکی
- مزیت نسبی آنها نسبت به کار ییدی بسیار زیاد است و به سرعت در حال رشد است (شکل ۲.۲ را در فصل ۲ ببینید)
- آنها در حال حاضر برخی از مشاغل حرفه ای را از چنگ مشاوران مالی، حقوقی، و متخصصان پزشکی در می آورند.
- قابلیت های هوش مصنوعی به سرعت در حال رشد است
- در روسیه، ربات ها در حال حاضر در مدارس ریاضی تدریس می کنند (بعضی از آنها کارشان را بهتر از انسان ها انجام می دهند). فقط به این فکر کنید که چه اتفاقی ممکن است برای حرفه معلمی بیفتد.

برای تأثیر بالقوه هوش مصنوعی بر مشاغل، دورمهیل^۱ (۲۰۱۷) را ببینید که امکان ماشین های هوشمند خلاق را بررسی می کند. به عنوان مثال، مطالعه مک کنزی تخمین می زند که هوش مصنوعی در آینده نزدیک ۳۰ درصد از مشاغل بانکی را در اختیار خواهد گرفت. این مطالعه همچنین پیش بینی می کند که ربات ها تا سال ۲۰۳۰ حدود ۸۰۰ میلیون شغل در سراسر جهان را در اختیار خواهند گرفت (اخبار مدیریت اطلاعات، ۲۰۱۷).

برای تحقیق در مورد خطرات بالقوه از دست دادن شغل، مک کنزی آند کامپنی مشاغل را به ۲۰۰۰ فعالیت کاری مجزا تقسیم کرده است از قبیل خوشامدگویی به مشتریان و پاسخ به سؤالات در مورد محصولات، که خرده فروشان انجام می دهند. پژوهشگران آن (نگاه کنید به چویی و همکاران، ۲۰۱۵) دریافتند خودکارسازی ۴۵ درصد از تمام ۲۰۰۰ فعالیت می تواند از نظر اقتصادی و فیزیکی توجیه پذیر باشد. فعالیت ها شامل انواع فیزیکی، شناختی و اجتماعی است. اگرچه وسایل نقلیه خودران مشاغل را از بین نمی برند، اما موجب از دست دادن مشاغل رانندگان تاکسی، اوبر و رانندگان شرکت های مشابه خواهند شد. همچنین رانندگان اتوبوس ممکن است شغل خود را از دست بدهند. مطالعه موردی ۱۴.۲ به بررسی سایر مشاغلی که سیستم های هوشمند جایگزین آنها شده اند می پردازد.

مشاغل یقه سفیدی که ربات ها قبلاً جایگزین آنها شده اند

مورد کاربردی ۱۴.۲

اگرچه استفاده از هواپیماهای بدون خلبان توسط فِدِکس و جایگزین شدن تمام معلم ها با ربات ها ممکن است مدتی طول بکشد طبق گفته شرمین (۲۰۱۵)، ربات ها در برخی مشاغل کاملاً جایگزین انسان ها شده اند از جمله:

- **بازاریاب های آنلاین.** با استفاده از پردازش زبان طبیعی، شرکت ها به طور خودکار تبلیغات بازاریابی و ایمیل هایی را ایجاد می کنند که افراد را برای خرید تحت تأثیر قرار می دهد (روبو بازاریاب ها). اینها بر اساس گفتگو با خریداران بالقوه و جستجوی خودکار پایگاه داده موارد تاریخی کار می کنند. چه کسی به یک بازاریاب آنلاین که دانش ضعیف، مغرضانه یا ناقصی داشته باشد نیاز دارد؟
- **تحلیلگران و مشاوران مالی.** همانطور که در فصل ۱۲ توضیح داده شد، مشاوران روبو در تمامی عرصه ها حضور دارند. این مشاوران با بهره گیری از توانایی مدیریت بلادرنگ کلان داده ها و انجام تجزیه و تحلیل و پیش بینی در عرض چند ثانیه، مورد توجه

¹ Dormehl

سرمایه‌گذارانی قرار می‌گیرند که حدود یک دهم هزینه‌ای را که مشاوران انسانی پرداخت می‌کنند می‌پردازند. علاوه بر این، مشاوران رباتی می‌توانند توصیه‌ها را شخصی‌سازی کنند

• **متخصصین بیهوشی، تشخیص بیماری و جراحان.** به نظر می‌رسد حوزه پزشکی از هوش مصنوعی مصون است. اما اینگونه نیست. سیستم‌های خبره ۴۰ سال است که برای تشخیص بیماری‌ها وجود دارند. اف‌دی‌ای^۱ قبلاً سیستم جی‌اندجی سِداسیس^۲ را برای انجام بیهوشی سطح پایین در جراحی‌هایی مانند کولونوسکوپی تأیید کرده است. آی‌بی‌ام واتسون تشخیص بسیار دقیق‌تری را نسبت به انسان در این زمینه نشان داده است موارد بیماری ریوی (۹۰٪ در مقابل ۵۰٪). در نهایت، جراحان از ماشین‌های خودکار در برخی از روش‌های تهاجمی استفاده می‌کنند.

• **گزارشگران مالی و ورزشی.** این مشاغل شامل جمع‌آوری اطلاعات، مصاحبه با افراد، پاسخ به سوالات، تجزیه و تحلیل مطالب و گزارش‌نویسی است. آسوشیتدپرس از سال ۲۰۱۴ بر روی ماشین‌های هوش مصنوعی آزمایش کرده است. نتایج تا کنون تقریباً بدون خطا و سوگیری هستند (و بدون اخبار جعلی!).

پالمر (۲۰۱۷) پنج شغل دیگر را که در معرض خطر هستند گزارش کرد که شامل مدیریت میانی، فروشندگان کالا، گزارش نویسان، حسابداران و حسابداران و برخی از انواع پزشکان می‌شد. مک‌فارلند^۳ (۲۰۱۷) صندوق‌دار، اپراتور باجه عوارض، کارمندان فست فود و رانندگان را جزو مشاغل پر خطر نام می‌برد. مشاغل کم‌خطر شامل پرستاران، پزشکان، دندانپزشکان، مربیان ورزشی جوانان و مددکاران اجتماعی است.

سوالاتی برای مورد کاربردی ۱۴.۲

۸. ویدئوی مصاحبه با پالمر را در [linkedin.com/pulse/5-jobsrobots-take-first-shelly-palmer/](https://www.linkedin.com/pulse/5-jobsrobots-take-first-shelly-palmer/) تماشا کنید. در مورد برخی از اظهارات مطرح شده در مورد پزشکان بحث کنید.
۹. در مورد امکان معاینه خود توسط یک ربات-تشخیص دهنده صحبت کنید. چه احساسی خواهید داشت؟
۱۰. با بمباران اخبار جعلی و سازندگان مغرضانه آنها، شاید عاقلانه باشید که همه آنها را با ماشین‌های هوشمند جایگزین کنیم. در مورد چنین احتمالی بحث کنید.
۱۱. شما متهم جرمی هستید که مرتکب نشده‌اید. آیا یک وکیل سنتی را ترجیح می‌دهید یا یک وکیل مجهز به دستگاه کشف الکترونیکی هوش مصنوعی؟ چرا؟

Sources: Compiled from E. Sherman. (2015, February 25). "5 White-Collar Jobs Robots Already Have Taken."

Fortune.com.fortune.com/2015/02/25/5-jobs-that-robots-already-aretaking (accessed April 2018); S. Palmer. (2017, February 26).

"The 5 Jobs Robots Will Take First." Shelly Palmer.

اجازه دهید به برخی از مطالعات دیگر نگاه کنیم. مطالعه‌ای که در سال ۲۰۱۶ در بریتانیا انجام شد، پیش‌بینی کرد که ربات‌ها تا سال ۲۰۲۶ حدود ۵۰ درصد از مشاغل را به خود اختصاص خواهند داد. ایگان (۲۰۱۵) گزارش می‌دهد که ربات‌ها در حال حاضر مشاغل زیر را تهدید می‌کنند: بازاریاب‌ها، اپراتور عوارضی و صندوق‌دارها، خدمات مشتری، کارگزاران مالی، روزنامه‌نگاران،

¹ FDA

² J&J Sedasys

³ McFarland

- طبق مطالعه‌ی پرایس‌واترهاوس کوپرز (پی‌دبلیوسی)^۱، پیش‌بینی می‌شود که ربات‌ها موجب افزایش رشد اقتصادی بریتانیا خواهند شد. بنابراین، حتی اگر ربات‌ها بتوانند حدود ۷ میلیون شغل را در بریتانیا از بین ببرند، حداقل ۷ میلیون شغل جدید و احتمالاً بیشتر در طول ۲۰ سال ایجاد خواهند کرد (بردن ۲۰۱۸).
- سرویس یادگیری عمیق جدید آی‌بی‌ام ممکن است به نجات مشاغل فناوری اطلاعات کمک کند
- ما با کمبود میلیون‌ها کارگر ماهر روبرو هستیم (به عنوان مثال، حدود ۵۰۰۰۰ راننده کامیون در ایالات متحده)، بنابراین اتوماسیون مشاغل می‌تواند منجر به خالی شدن میلیون‌ها فرصت و موقعیت شغلی شود.
- کورولوف^۲ (۲۰۱۶) مدعی است که کار زیادی وجود دارد، به خصوص برای افرادی که با تکنولوژی همگام هستند و مهارت‌های خود را توسعه می‌دهند.
- شرکت گارتنر پیش‌بینی می‌کند که تا سال ۲۰۲۰، هوش مصنوعی بیش از آنکه موجب حذف مشاغل شود، شغل ایجاد خواهد کرد (سینگ ۲۰۱۷).
- ویلسون و همکاران^۳ (۲۰۱۷) گزارشی در مورد گروه‌های جدیدی از مشاغل انسانی که توسط هوش مصنوعی ایجاد شده است ارائه کرده‌اند.
- برخی معتقدند که به دلیل نوآوری‌های ناشی از هوش مصنوعی، در مجموع مشاغل افزایش خواهد یافت.
- تخمین زده می‌شود که در سال ۲۰۱۸ بیش از ۴۹۰،۰۰۰ شغل برای دانشمندان داده باشد، اما تنها ۲۰۰،۰۰۰ دانشمند در دسترس خواهند بود. با این حال، در بلندمدت، هوش مصنوعی و یادگیری ماشینی ممکن است جایگزین اکثر دانشمندان علم داده شوند (پرز، ۲۰۱۷)^۴.

نکته: زمانی که این کتاب به چاپ رسید، ما با کمبود کارمند فناوری اطلاعات (چند میلیون در ایالات متحده) مواجه بودیم. اتوماسیون می‌تواند این کمبود را کاهش دهد. توجه داشته باشید که مطالعه‌ای که توسط ولدون (۲۰۱۷) گزارش شده است نشان داد که اکثر کارگران در واقع از تأثیر هوش مصنوعی و اتوماسیون بر مشاغل استقبال می‌کنند. به عنوان نکته پایانی، گووها^۵ (۲۰۱۷) دیدگاهی از کار و هوش مصنوعی به عنوان چشم اندازی از "نامیدی، امید و رهایی" ارائه می‌دهد. او نتیجه می‌گیرد که هوش مصنوعی می‌تواند کار را آزاد کند - این یک فرصت تاریخی است.

در حالی که ممکن است شغل خود را از دست ندهید، برنامه‌های هوشمند ممکن است موجب تغییر شغل‌تان شوند. یکی از جنبه‌های این تغییر این است که مشاغل کم مهارت توسط ماشین آلات انجام می‌شود، اما مشاغل با مهارت بالا ممکن است نشوند. بنابراین، مشاغل ممکن است بازطراحی شده یا با مهارت کم به منظور خودکار شدن یا با مهارت بالا طراحی شوند تا

¹ PricewaterhouseCoopers (PWC)

² Korolov

³ Wilson et al.

⁴ Perez

⁵ Guha

منحصراً توسط انسان‌ها اجرا شوند. علاوه بر این، مشاغل زیادی وجود خواهد داشت که در آن انسان‌ها و ماشین‌ها به عنوان یک تیم با هم کار خواهند کرد. تغییرات در مشاغل و فرایندهای کسب‌وکار بر آموزش، نوآوری، دستمزد و ماهیت خودکار تأثیر می‌گذارد. مانیکا (۲۰۱۷) و مانیکا و همکاران (۲۰۱۷)) شرکت مک کینزی تغییراتی که می‌توانند اساسی باشند را تجزیه و تحلیل کرد و به موارد و نتیجه‌گیری‌های زیر رسید.

نکته: در زمان انتشار این کتاب، ما با کمبود کارمندان فناوری اطلاعات (چند میلیون در ایالات متحده) مواجه بودیم. اتوماسیون می‌تواند این کمبود را کاهش دهد. توجه داشته باشید که مطالعه‌ای که توسط ولدون (۲۰۱۷) گزارش شده است، نشان داد که اکثر کارگران در واقع از تأثیر هوش مصنوعی و اتوماسیون بر مشاغل استقبال می‌کنند. به عنوان یادداشت پایانی گووها (۲۰۱۷) یک دیدگاه در مورد کار و هوش مصنوعی را به عنوان چشم اندازی از "یأس، امید و رهایی". او نتیجه می‌گیرد که هوش مصنوعی می‌تواند کار را آزاد کند - این یک فرصت تاریخی است.

اگرچه ممکن است شغل خود را از دست ندهید، اما برنامه‌های هوشمند ممکن است شغل‌تان را تغییر دهند. یکی از جنبه‌های این تغییر این است که مشاغلی که به مهارت کمی نیاز دارند توسط ماشین آلات انجام می‌شود، اما مشاغلی که به مهارت بالایی نیاز دارند ممکن است چنین نشوند.

از این رو، مشاغل ممکن است بازطراحی شوند یا سطح مهارت مورد نیاز آنها کاهش یابد تا بتوان آنها را خودکارسازی نمود یا اینکه با سطح مهارتی مورد نیاز بالایی طراحی شوند تا صرفاً توسط انسان‌ها انجام شوند. علاوه بر این، مشاغل زیادی وجود خواهد داشت که در آن افراد و ماشین‌ها به عنوان یک تیم با هم کار خواهند کرد. تغییرات در مشاغل و فرایندهای کسب‌وکار بر آموزش، نوآوری، دستمزد و ماهیت خود کار تأثیر می‌گذارد. مانیکا (۲۰۱۷) و مانیکا و همکاران (۲۰۱۷) از مک‌کنزی آند کامپانی تغییرات بنیادی را تحلیل کرده و به نتایج زیر دست یافته‌اند:

- بسیاری از فعالیت‌هایی که انسان‌ها انجام می‌دهند پتانسیل خودکار شدن را خواهند داشت.
- رشد بهره‌وری حاصل از رباتیک، هوش مصنوعی و یادگیری ماشینی نسبت به پیش از سال ۲۰۱۵ سه برابر خواهد شد.
- هوش مصنوعی مشاغل جدید زیادی را با پرداخت حقوق بالا ایجاد خواهد کرد.
- از آنجاییکه بیش از نیمی از جهان هنوز آفلاین است، تغییرات خیلی سریع نخواهد بود.

مثال: مهارت‌های دانشمندان علم داده‌ها تغییر خواهد کرد

طبق توسو (۲۰۱۷) از گروه مطالعاتی موسسه جهانی مک کینزی، تا سال ۲۰۲۴ ما با کمبود ۲۵۰,۰۰۰ دانشمند علم داده‌ها مواجه خواهیم بود و نیاز به بازآموزی یا آموزش دانشمندان وجود خواهد داشت تا بتوانند با فناوری‌های هوشمند و تغییرات علم داده‌ها و حل مسائل مربوط به دنیای واقعی کنار بیایند. بنابراین، آموزش صحیح باید تکامل یابد. الزامات شغلی دانشمندان

علم داده‌ها در حال حاضر در حال تغییر هستند. دانشمندان باید بدانند که چگونه از یادگیری ماشینی و فناوری‌های هوشمند برای ساخت اینترنت اشیاء و سایر سیستم‌های مفید بهره ببرند. الگوریتم‌های جدید بهبود عملیات و امنیت، و پلتفرم‌های داده در حال تغییر هستند تا متناسب با مشاغل جدید باشند.

اسنایدر (۲۰۱۷) دریافت که ۸۵ درصد از مدیرانی که با این فناوری‌های هوشمند آشنا هستند در عرض پنج سال بر نیروی کار آنها تأثیر خواهد گذاشت و ۷۹ درصد انتظار دارند که ساختار مجموعه مهارت‌های جاری آنها تغییر یابد.

- پیش از حرکت به سمت هوش مصنوعی باید قابلیت‌های دیجیتالی را ایجاد کرده و بکار گرفت.
- یادگیری ماشینی قدرتمند است، اما راه حل همه مشکلات نیست.
- فقط تیم‌های فناوری را مسئول فناوری‌های هوشمند قرار ندهید.
- افزودن یک شریک تجاری ممکن است به پروژه‌های مبتنی بر هوش مصنوعی کمک کند.
- یک رویکرد پورتفولیویی را برای ابتکار عمل‌های هوش مصنوعی اولویت‌بندی کنید.
- بزرگترین چالش‌های پیش‌رو، افراد و فرایندهای کسب‌وکار خواهند بود.
- همه کسب‌وکارها از سیستم‌های هوشمند استفاده نمی‌کنند، اما تقریباً همه آنها می‌کنند که استفاده می‌کنند درآمد و سودشان افزایش می‌یابد.
- پشتیبانی رهبری ارشد سازمان برای حرکت و تغییر به سمت هوش مصنوعی ضروری است.

مانیکا (۲۰۱۷) توصیه‌های زیر را برای سیاست‌گذاران ارائه کرده

است:

۱۸. برای تسهیل تغییرات از یادگیری و آموزش استفاده کنید.
۱۹. بخش خصوصی را در افزایش آموزش و بازآموزی درگیر کنید
۲۰. دولت‌ها مشوق‌هایی را برای بخش خصوصی فراهم می‌کنند تا کارکنان بتوانند در سرمایه‌های انسانی بهبود یافته سرمایه‌گذاری کنند.
۲۱. بخش‌های خصوصی و دولتی را برای ایجاد زیرساخت‌های دیجیتال مناسب تشویق کنید.
۲۲. طرح‌های نوآورانه برای درآمد و دستمزدها باید توسعه یابد.
۲۳. انتقال به کار جدید را به دقت برنامه‌ریزی کنید. با کارمندان که جابجا شده‌اند مدیریت و رفتار مناسبی داشته باشید.
۲۴. از فناوری‌های جدید به درستی استفاده کنید..
۲۵. بر ایجاد مشاغل جدید به ویژه مشاغل دیجیتالی تمرکز کنید.
۲۶. فرصت‌های افزایش بهره‌وری را به دستی جذب کنید.

- **ایلان ماسک:** «ما در مورد هوش مصنوعی باید بسیار محتاط باشیم. به صورت بالقوه خطرناک‌تر از سلاح‌های هسته‌ای است.» ویدیوی ۱۰ دقیقه‌ای را در [youtube.com/watch?v=SYqCbJ0AqR4](https://www.youtube.com/watch?v=SYqCbJ0AqR4) ببینید. ماسک پیش‌بینی می‌کند که جنگ جهانی سوم به دلیل هوش مصنوعی آغاز خواهد شد. او در چندین سخنرانی خود گفت: «ربات‌ها یک روز همه ما را خواهند کشت.»
- **بیل گیتس:** «من در گروهی هستم که نگران هوش برتر است. ماسک و برخی دیگر در این گروه هستند و من نمی‌دانم چرا برخی از مردم نگران نیستند.» (نظراتی که در تلویزیون و مصاحبه‌ها چندین بار داده شده است). او همچنین پیشنهاد کرد که از تولیدکنندگان و کاربران ربات‌ها و سایر ماشین‌های هوش مصنوعی مالیات گرفته شود.
- **استیون هاوکینز** دانشمند فقید اظهار داشت: «توسعه هوش مصنوعی کامل می‌تواند پایان نسل بشر باشد.»

بسیاری از مردم از هوش مصنوعی می‌ترسند زیرا معتقدند رایانه‌ها از ما باهوش‌تر می‌شوند. ویدیوی بوستروم از پرزنت معروف تی‌ا‌ی‌دی او را در [youtube.com/watch?v=MnT1xgZgkpk](https://www.youtube.com/watch?v=MnT1xgZgkpk) ببینید. همچنین به مگوایر (۲۰۱۷) برای بحث در مورد ربات‌های یادگیرنده و خطر ربات‌های یادگیرنده مراجعه کنید. برای اینکه چگونه ربات‌ها می‌توانند مهارت‌های حرکتی را از طریق آزمون و خطا بیاموزند، ویدیو را در [youtube.com/watch?v=JeVppkoloXs/](https://www.youtube.com/watch?v=JeVppkoloXs/) ببینید. برای اطلاعات بیشتر، پام (۲۰۱۸) را ببینید.

یک مکان خوب برای شروع اطلاعات در مورد این وضعیت، تماشای ویدئوی ۲۶ دقیقه‌ای مستند در مورد آینده هوش مصنوعی در [youtube.com/watch?v=UzT3Tkwx17A](https://www.youtube.com/watch?v=UzT3Tkwx17A) است. این ویدیو بر سهم هوش مصنوعی در کیفیت زندگی متمرکز است. یک مثال، مبارزه با جرم و جنایت در سانتا کروز، کالیفرنیا است، جایی که هوش مصنوعی توانست مکان و زمان وقوع جنایات را پیش‌بینی کند. پس از پیش‌بینی‌ها، اداره پلیس راهبردهای کاری خود را برنامه‌ریزی کرده است. نتیجه این کار، کاهش ۲۰ درصدی جرم است.

مثال دوم، پیش‌بینی احتمال موفقیت یک آهنگ خاص است. این پیش‌بینی به هنرمندان و مدیران کمک می‌کند تا فعالیت‌های خود را برنامه‌ریزی کنند. موفقیت بزرگی حاصل شده است. پیش‌بینی می‌شود در آینده، هوش مصنوعی آهنگ‌های برتر را بسازد.

در نهایت، یک داستان در مورد دوستیابی وجود دارد. قابلیت‌های هوش مصنوعی به دانشمندان این امکان را می‌دهد تا در میان جمعیت ۳۰۰۰۰ کاندیدای بالقوه، مورد تطبیقی بی‌عیب و نقص را پیدا کند.

یک بحث اساسی طرفداران نظریه‌ی آرمان‌شهر در مصاحبه‌ها، سخنرانی‌های تلویزیونی و غیره مطرح شده این است که هوش مصنوعی از انسان‌ها حمایت می‌کند و نوآوری‌ها را امکان‌پذیر می‌کند. هوش مصنوعی همچنین با انسان‌ها شریک خواهد شد. طرفداران آرمان‌شهرها بر این باورند که با گسترش هوش مصنوعی، انسان‌ها بهره‌ورتر می‌شوند و زمان بیشتری برای انجام کارهای خلاقانه‌تر خواهند داشت. در عین حال، وظایف بیشتری به طور کامل خودکار خواهند شد. قیمت محصولات و خدمات کاهش می‌یابد و کیفیت زندگی افزایش می‌یابد. در یک زمان، ممکن است به یک اقتصاد کاملاً خودکار و خودپایدار دست یابیم. در نهایت، مردم برای امرار معاش اصلاً مجبور به کار نخواهند بود. یکی از حامیان اصلی مزایای هوش مصنوعی، مارک زاکربرگ از فیسبوک است. او با ایلان ماسک (مدیر عامل شرکت تسلا)،

رهبر غیررسمی کمپ طرفداران ویران شهر در حال مناظره داغ است. زاکربرگ از کسانی که معتقدند هوش مصنوعی باعث «رستاخیز» خواهد شد انتقاد کرد (به بخش بعدی مراجعه کنید). ماسک ادعا کرد که زاکربرگ "درک محدودی" از هوش مصنوعی دارد و زاکربرگ با اشاره به مقاله خود در مورد هوش مصنوعی که در "کنفرانس برتر بینایی رایانه‌ای" جایزه دریافت کرد، پاسخ داد. برای جزئیات، به وانیان (۲۰۱۷) مراجعه کنید.

چندین موضوع به موضع طرفداران آرمان شهرها مربوط می‌شود. در اینجا سه مثال آورده شده است:

۱. هوش مصنوعی به قدری عالی خواهد بود که مشکل مردم این خواهد بود که با اوقات فراغت خود چه کنند. اگر هنوز فیلم وال-ایبی دیزنی را ندیده‌اید، بروید و آن را ببینید. این فیلم نشان می‌دهد که چگونه ربات‌ها به انسان‌ها خدمات می‌دهند. دنیس حاسابیس، یکی از حامیان قوی آرمان شهر از شرکت دیپ مایند که یک شرکت هوش مصنوعی است معتقد است که هوش مصنوعی روزی به مردم کمک می‌کند تا با درک اینکه چه چیزی انسان‌ها را منحصر به فرد می‌کند و با درک اسرار ذهن زندگی بهتری داشته باشند و از خلاقیت لذت ببرند.
۲. راه رسیدن به آرمان شهر هوش مصنوعی می‌تواند سنگلاخ باشد، برای مثال، تأثیراتی بر مشاغل و کار خواهد داشت. تثبیت و تنظیم کار و زندگی با ربات‌ها، ربات‌های محاوره‌ای و سایر برنامه‌های هوش مصنوعی زمان می‌برد.
۳. زمانی فرا می‌رسد که ما دیگر رانندگی نخواهیم کرد و ممکن است نیازی به مشاور مالی انسانی نباشد. همه چیز متفاوت خواهد بود، و تغییرات ممکن است سریع و متلاطم باشد و حتی ممکن است با فجایع روبرو شویم، همانطور که توسط گروه طرفدار ویران شهر پیش‌بینی شده است.

برای آماده شدن در برابر اقدامات ناخواسته رباتیک و هوش مصنوعی، ایلان ماسک و دیگران یک سازمان غیرانتفاعی به نام اپن‌ای‌آی^۱ را ایجاد کردند. با در نظر گرفتن خطر بالقوه ناخواسته، ماسک و دیگران یک شرکت تحقیقاتی غیرانتفاعی هوش مصنوعی را با سرمایه‌ی ۱ میلیارد دلار ایجاد کردند. هدف اصلی این شرکت، ایجاد مسیری برای هوش جامع مصنوعی ایمن است. همانطور که از فصل ۱ به یاد دارید، هوش جامع مصنوعی هنوز شکل نگرفته است اما در راه است.

برنامه اپن‌ای‌آی ایجاد هوش جامع مصنوعی ایمن و اطمینان از توزیع یکنواخت مزایای آن است. نتایج تحقیق در مجلات برتر منتشر شده است. علاوه بر این، اپن‌ای‌آی ابزارهای نرم افزاری متن باز ایجاد می‌کند. این سازمان یک وبلاگ دارد و اخبار مهم هوش مصنوعی را منتشر می‌کند. برای جزئیات، به openai.com مراجعه کنید.

ایلیزر یودکوفسکی^۲ یکی از بنیان‌گذاران مؤسسه تحقیقاتی هوش ماشینی، ایده هوش مصنوعی دوستانه را توسعه داد که بر اساس آن ماشین‌های هوش مصنوعی باید به گونه‌ای طراحی شوند که به جای آسیب رساندن به انسان‌ها به آنها منفعت

¹ OpenAI

² Elizier Yudkowsky

برسانند. برای جزئیات بیشتر به شرم^۱ (۲۰۱۸) مراجعه کنید و ویدئوی جذاب ۱:۲۹:۵۵ دقیقه‌ای را در [youtube.com/watch?v=EUjc1WuyPT8](https://www.youtube.com/watch?v=EUjc1WuyPT8) تماشا کنید.

دشوار است که بدانیم در آینده چه اتفاقی خواهد افتاد. اما برخی از اقدامات در حال حاضر برای جلوگیری از یک فاجعه انجام شده است. به عنوان مثال، چندین شرکت بزرگ اعلام کرده اند که ربات‌های قاتل تولید یا پشتیبانی نمی‌کنند.

مدیران و متخصصان علم داده‌ها باید از اثرات اجتماعی و بلندمدت مدل‌ها و الگوریتم‌های ریاضی آگاه باشند. کتی آنیل^۲، دکترای ریاضیات هاروارد که در امور مالی و صنعت علم داده‌ها کار می‌کرد، تجربیات خود را بیان کرد و مشاهدات خود را در کتاب معروف *سلاح‌های مخرب ریاضی: چگونه داده‌های بزرگ نابرابری را افزایش می‌دهد و دموکراسی را تهدید می‌کند* ارائه کرد. پیشنهاد می‌کنیم کتاب را بخوانید یا حداقل به سایت وبلاگ نویسنده به آدرس mathbabe.org/ مراجعه کنید. سایت وبلاگ موضوعات اجتماعی مرتبط با تجزیه و تحلیل را برجسته می‌کند. خلاصه‌ای از کتاب در knowledge.wharton.upenn.edu/article/rogue-algorithms-dark-side-big-data/ موجود است.

آنیل (۲۰۱۶) در کتاب خود استدلال می‌کند که مدل‌ها باید سه شرط را برآورده کنند. ابتدا باید شفاف باشند. یعنی اگر مدل قابل درک نباشد، کاربرد آن می‌تواند منجر به عواقب ناخواسته شود.

ثانیاً، مدل باید دارای اهداف قابل اندازه‌گیری واضح باشد. به عنوان مثال، کاربرد مشهور تجزیه و تحلیل در کتاب و فیلم *مانی‌بال*^۳ شامل مدلی است که با هدف افزایش تعداد بردهای مالی انجام شده است. و اقدامات ورودی پیشنهادی به خوبی قابل درک بودند. به جای استفاده از معیار (RBI) «run base in»، تحلیل‌گر در *مانی‌بال* از درصد مبتنی بر پایه و سایر معیارها استفاده کرد (که برای هر کسی با مهارت‌های ریاضی پایه به راحتی قابل محاسبه و درک است). از سوی دیگر، مدل‌هایی که برای ارزیابی ریسک اوراق بهادار با پشتوانه وام مسکن ساخته شده‌اند، زمانی که هیچ‌کس مفروضات اساسی اوراق بهادار وثیقه‌شده را کاملاً درک نمی‌کرد، معامله‌گران مالی در حال معامله بودند، به‌خاطر رهبری بحران مالی ۲۰۰۸ مقصر شناخته شده‌اند.

شرط سوم این است که مدل‌ها باید دارای مکانیزم خود تصحیح‌کننده و فرایندی باشند تا مرتباً تمیزی شوند و ورودی‌ها و خروجی‌های جدید مدام در نظر گرفته شوند. این موضوع سوم به ویژه در به کارگیری مدل‌ها در محیط‌های اجتماعی بسیار مهم است. در غیر این صورت، مدل‌ها مفروضات معیوب ذاتی را تداوم می‌بخشند مرحله مدل‌سازی اولیه اونیل چندین وضعیت را مورد بحث قرار می‌دهد که در آنها چنین است. به عنوان مثال، او مدل‌های ساخته شده در ایالات متحده را برای شناسایی معلمان ضعیف و پاداش دادن به معلمان بهتر توضیح می‌دهد. برخی از این مدل‌ها از نمرات آزمون دانش آموزان برای ارزیابی معلمان استفاده کردند. اونیل مثال‌های متعددی را ذکر کرد که در آن از مدل‌ها برای اخراج معلمان «کم‌کار» استفاده می‌شد، حتی اگر آن معلمان مورد علاقه دانش‌آموزان و والدین بودند. به طور مشابه، مدل‌ها برای بهینه‌سازی زمان‌بندی کارگران استفاده می‌شوند. در بسیاری از سازمان‌ها این برنامه‌ها ممکن است برای پاسخگویی به تغییرات تقاضای فصلی و روزانه ایجاد شده باشند، اما این مدل‌ها تأثیرات مخرب چنین تغییراتی را در برنامه‌ها بر خانواده‌های این کارگران معمولاً کم‌درآمد در نظر نمی‌گیرند. نمونه‌های دیگری از این دست شامل مدل‌های ارزیابی امتیاز اعتباری است که بر اساس نمایه‌های تاریخی است و بنابراین ممکن است بر

¹ Sherman

² Cathy O'Neil

³ Moneyball

اقلیت‌ها تأثیر منفی بگذارد. بدون مکانیسم‌هایی برای ممیزی چنین مدل‌هایی و اثرات ناخواسته آن‌ها، می‌توانند در بلند مدت زیان بیشتری نسبت به سود داشته باشند. بنابراین، سازندگان مدل باید چنین نگرانی‌هایی را در نظر بگیرند.

نکته: در می ۲۰۱۸، مقررات عمومی حفاظت از داده‌ها اتحادیه اروپا لازم الاجرا شد. این قانون شامل نیاز به توضیح داده‌ها است. طبق گفته سیوین (۲۰۱۸)، یک هوش مصنوعی قابل توضیح می‌تواند تأثیر الگوریتم‌های جانبدارانه را کاهش دهد.

اظهار نظر: شواهدی وجود دارد که در برخی موارد ادعاهای اونیل معتبر است و بنابراین سازندگان و مجریان مدل باید به مسائل توجه کنند. با این حال، به طور کلی، تجزیه و تحلیل به درستی طراحی شده‌اند و مزایای قابل توجهی برای جامعه به همراه دارند. علاوه بر این، مدل‌های تحلیلی، رقابت‌پذیری شرکت‌ها و کشورها را افزایش می‌دهند و مشاغل پردرآمد زیادی ایجاد می‌کنند. در بسیاری از موارد، شرکت‌ها سیاست‌های مسئولیت اجتماعی را بر عهده دارند که سوگیری‌ها و نابرابری‌ها را به حداقل می‌رساند. در نهایت، همانطور که ولدون (۲۰۱۷) بیان می‌کند، الگوریتم‌ها و هوش مصنوعی را می‌توان به‌عنوان یک تقویت‌کننده عالی در ارائه خدماتی که به‌طور سنتی برای عده معدودی از افراد خاص و برجسته محفوظ بود، برای همه در نظر گرفت.

سوالاتی برای مرور بخش ۱۴.۷

۱۱. بحث‌های مهم طرفداران آرمان شهر را خلاصه کنید.
۱۲. بحث‌های مهم طرفداران ویران شهر را خلاصه کنید.
۱۳. هوش مصنوعی دوستانه چیست؟
۱۴. آپن‌ای‌آی چیست و ارتباط آن با چشم‌انداز ویران شهر چیست؟
۱۵. ریسک‌های بالقوه‌ی بکارگیری مدلسازی و ابزارهای تحلیلی چیست؟

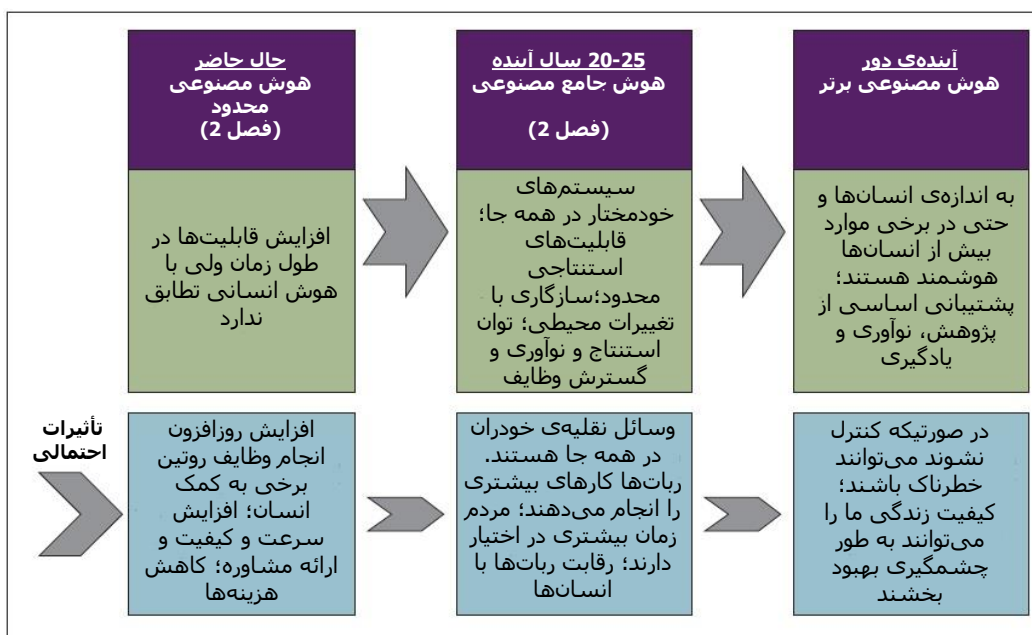
با نزدیک شدن به آخرین بخش این کتاب که برخی از جنبه‌های آینده سیستم‌های هوشمند را مورد بحث قرار می‌دهد، ارزش دارد که برخی از روندهای فناوری که این آینده را شکل خواهند داد، شرح دهیم. متأسفانه صدها گرایش فناوری مرتبط با محتوای این کتاب وجود دارد. دلیل آن این است که ابزارهای تجزیه و تحلیل، ابزارهای کلان‌داده‌ها، هوش مصنوعی، یادگیری ماشینی، رباتیک، اینترنت اشیا و سایر سیستم‌های هوشمند دارای تنوع زیادی هستند. بنابراین، ما در اینجا تنها مثال‌هایی از روندهای فناوری را ارائه می‌دهیم. ما این بخش را به زیر بخش‌های زیر تقسیم می‌کنیم:

- فهرست‌های گارتنر در سال ۲۰۱۸ و ۲۰۱۹.
- فهرست روندهای فناوری در سیستم‌های هوشمند
- رایانش محیطی.

شرکت گارتنر یک سازمان تحقیقاتی و مشاور برتر فناوری و همچنین برگزارکننده سمپوزیوم سالانه فناوری با حضور بیش از ۲۳۰۰۰ نفر است. این شرکت سالانه پیش‌بینی‌هایی در حوزه فناوری که به نظر می‌رسد بیشتر سازمان‌ها را تحت تأثیر قرار می‌دهند ارائه می‌کند. فهرست روندها در سال‌های ۲۰۱۸ و ۲۰۱۹ شامل ۱۰ مورد است که بیشتر آنها مستقیماً به محتوای کتاب ما مربوط می‌شوند. خلاصه فهرست ۲۰۱۸ در شکل ۱۴.۳ نشان داده شده است. این فهرست از نشریه‌ی مطبوعاتی گارتنر در ۴ اکتبر ۲۰۱۷ استخراج شده است که در gartner.com/newsroom/id/3812063 در دسترس است. مبنای و اصول آن در یک ویدیو (۵:۳۶ دقیقه) در youtube.com/watch?v=TPbKyD2bAR4 ارائه شده است.

موارد زیر از gartner.com/newsroom/id/3812063 برای سال ۲۰۱۸، و از ولدون (۲۰۱۸)، برای سال ۲۰۱۹ استخراج شده است:

۱. **بنیاد و توسعه هوش مصنوعی.** سیستم‌های هوش مصنوعی پیشرفته تصمیم‌گیری پشتیبانی می‌کنند که برخی از آنها مستقل هستند و سایر سیستم‌های هوش مصنوعی در ارتباط با تجزیه و تحلیل و علم داده‌ها توسعه یافته‌اند.
۲. **اپلیکیشن‌ها و ابزارهای تحلیل هوشمند.** تقریباً تمام سیستم‌های فناوری اطلاعات در چند سال آینده شامل هوش مصنوعی خواهند شد [gartner.com/smarterwithgartner/the-cios-journey-to-artificial-intelligence/] را ببینید.
۳. **اشیاء هوشمند و مستقل.** با استفاده از قابلیت‌های اینترنت اشیا، انفجاری در رشد وسایل نقلیه خودران و افزایش قابل توجهی در سایر اشیاء هوشمند رخ خواهد داد (مانند خانه‌های هوشمند و کارخانجاتی که ربات‌ها در حال مونتاژ ربات‌ها هستند).
۴. **دوقلوهای دیجیتال.** برای یک دوقلو دیجیتال که به نمایش دیجیتال اشیا و سیستم‌های دنیای واقعی اشاره دارد به [gartner.com/smarterwithgartner/preparefor-the-impact-of-digital-twins/] مراجعه کنید. این عمده‌تاً شامل سیستم‌های اینترنت اشیا با ۲۰ میلیارد چیز متصل در ۲ تا ۳ سال آینده می‌شود.
۵. **اُبر قدرتمند.** در رایانش پیشرفته، جمع‌آوری، پردازش و تحویل اطلاعات نزدیک‌تر به منابع اطلاعات انجام می‌شود.



شکل ۱۴.۳ پیش‌بینی آینده‌ی هوش مصنوعی

۶. پلتفرم‌های مکالمه انسان و ماشین. این پلتفرم‌ها در حال حاضر تعاملات زبان طبیعی را تسهیل می‌کنند و منجر به بهبود همکاری می‌شوند. اینها شامل فضاهای مشارکتی هوشمند است.
۷. تجربه همه‌جانبه. این سیستم‌ها روشی را که مردم می‌توانند جهان را ببینند و درک کنند (به عنوان مثال واقعیت افزوده) تغییر می‌دهند [\[gartner.com/smarterwithgartner/transform-business-outcomes-with-immersive-technology/\]](https://gartner.com/smarterwithgartner/transform-business-outcomes-with-immersive-technology/) را ببینید.
۸. بلاک چین. فناوری‌های بلاک چین [\[gartner.com/smarterwithgartner/areyou-ready-for-blockchain-infographic/\]](https://gartner.com/smarterwithgartner/areyou-ready-for-blockchain-infographic/) یک پلتفرم مهم و اساسی برای افزایش امنیت و اعتماد ارائه می‌دهد و به طور قابل توجهی معاملات تجاری را بهبود می‌بخشد.
۹. تجزیه و تحلیل افزوده. استفاده از یادگیری ماشینی این فناوری را قادر می‌سازد تا بر تحول تجزیه و تحلیل تمرکز کند، بنابراین بهتر به اشتراک گذاشته و مصرف می‌شود. این امر مدیریت آماده‌سازی داده‌ها و تجزیه و تحلیل را برای بهبود پشتیبانی تصمیم تسهیل می‌کند.
۱۰. موارد دیگر. اینها شامل فضای همکاری هوشمند، محاسبات کوانتومی، حریم خصوصی دیجیتال و اخلاقی، و پذیرش ریسک و اعتماد است.

- انجمن رایانه‌ی آی-تریپل-ئی نیز دارای ۱۰ پیش‌بینی برتر برای سال ۲۰۱۸ است. computer.org/web/pressroom/top-technology-trends-2018 این لیست شامل یادگیری عمیق، اینترنت اشیاء صنعتی، رباتیک، حمل و نقل کمکی، واقعیت افزوده، بلاک چین، و ارزهای دیجیتال می‌شود.
- نیومن (۲۰۱۸) فهرستی از ۱۸ روند فناوری در سی‌ایی‌اس ۲۰۱۸ را ارائه می‌دهد. اینها به نمایشگرها در سی‌ایی‌اس مربوط می‌شوند.
- کاربرد و ارزش تجاری بالقوه‌ی فناوری تجزیه و تحلیل و هوش مصنوعی بر اساس مطالعات ۴۰۰ مورد واقعی انجام شده در مک‌کنزی آند کامپانی به عنوان مصورسازی داده‌های تعاملی از آدرس mckinsey.com/featured-insights/artificial-intelligence/visualizing-the-uses-and-potential-impact-of-ai-and-other-analytics/ در دسترس است.
- ۱۰ روند برتر برای تجزیه و تحلیل در سال ۲۰۱۸ توسط اسمیت (۲۰۱۸) ارائه شده است. این فهرست ماهیت نسبتاً فنی دارد. این شامل «گرانث داده‌ها به ابر شتاب می‌گیرد»، «بینش‌ها به عنوان یک سرویس برخوانند خواست» و «ابزارهای تحلیلی ابری آند-تو-آند پدیدار خواهند شد».

● ۱۰ روند برتر فناوری هوش مصنوعی برای سال ۲۰۱۸ همانطور که توسط راثو و همکاران (۲۰۱۷) پیش بینی شده است شامل «یادگیری تقویت‌شده‌ی عمیق: تعامل با محیط برای حل مشکلات کسب‌وکار» و «هوش مصنوعی قابل توضیح: درک جعبه سیاه» است.

● برای هفت روند داده‌ها و تحلیل‌ها به datameer.com/blog/seven-dataanalytics-trends-2018/ مراجعه کنید.

● رایانه‌ها فکر کردن را یاد می‌گیرند و برای یادگیری فکر می‌کنند.

● ربات‌ها در نقش‌های غیر فیزیکی و شناختی بیشتر جایگزین انسان‌ها خواهند شد.

● افزودگی هوشمند بخشی از هوش مصنوعی پیشرفته (فصل ۱) است و به کنترل برنامه‌های جدید هوش مصنوعی ادامه خواهد داد.

● رایانش پیشرفته توسط گارتنر ذکر شده است، اما ارزش بسیار بیشتری دارد که ممکن است به «ابر» مرتبط نباشد. این فناوری تاثیر زیادی بر آینده مراکز داده خواهد داشت. برای جزئیات به اسکایز (۲۰۱۸) مراجعه کنید. توجه داشته باشید که بیشتر قابلیت‌های جدید برای «ابر» در استفاده از «اج» وجود دارد. برای اطلاعات بیشتر به ویکیپدیا مراجعه کنید. پیشرفت‌های هوش مصنوعی اج در پشتیبانی از یادگیری ماشین و واقعیت افزوده برتر خواهند بود.

سامر (۲۰۱۷) موارد زیر را فهرست می‌کند:

● سطح سواد در مورد داده‌ها هم در سازمان‌ها و هم در جامعه گسترش می‌یابد.

● نقاط اطلاعاتی از طریق سیستم‌های ترکیبی چند ابری به هم متصل خواهند شد.

● رمز و راز شبکه‌های روستایی توسط نظریه یادگیری عمیق آشکار خواهد شد.

● سیستم‌های سلف سرویس از کاتالوگ داده‌ها به عنوان خط مقدم خود استفاده خواهند کرد.

● نیاز به تمرکز بر رابط‌های برنامه نویسی کاربردی.

● ابزارهای تجزیه و تحلیل تبدیل به حالت مکالمه‌ای می‌شوند (به عنوان مثال، از طریق چت‌بات).

● ابزارهای تجزیه و تحلیل شامل قابلیت‌های همه جانبه خواهد بود.

● کاربران با استفاده از هوش افزوده به شرکت‌کنندگان تبدیل خواهند شد.

● برای ۱۱ روند برتری که محرک هوش کسب‌وکار در سال ۲۰۱۸ خواهند شد به سامر (۲۰۱۷) مراجعه کنید.

● برای شش روند تجزیه و تحلیل داده‌ها در سال ۲۰۱۸، اولوسرود^۱ (۲۰۱۸) را ببینید.

● برای روند رباتیک در سال ۲۰۱۸، چپمن (۲۰۱۸) را ببینید.

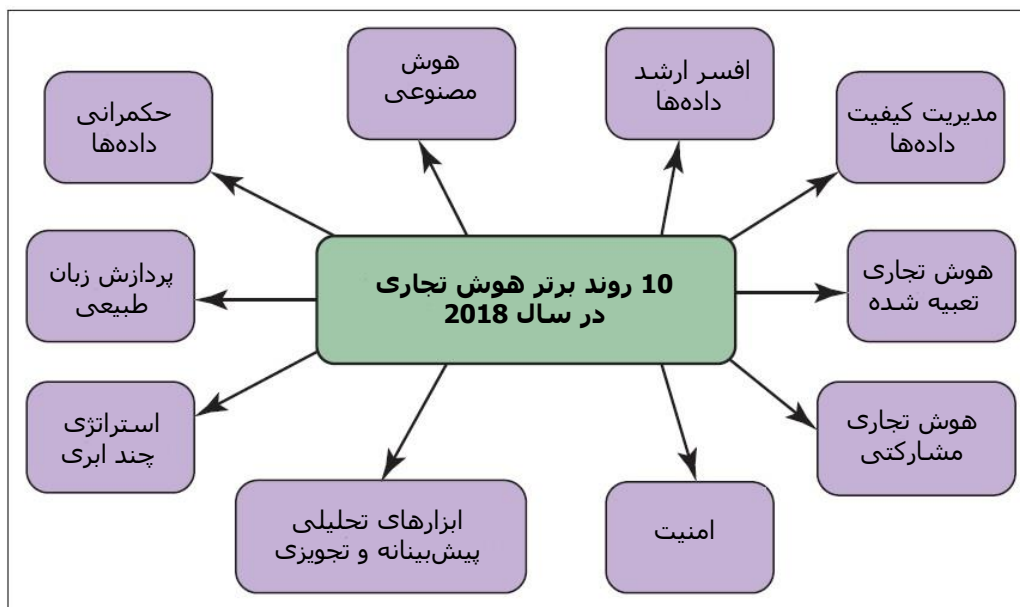
● برای ۱۰ پیش‌بینی سیستم‌های هوشمند، به پرس (۲۰۱۷) مراجعه کنید.

¹ Olavsrud

اکنون که روندهای بسیاری از فناوری‌ها را برای آینده مشاهده کرده‌اید، ممکن است بخواهید ببینید چه زمانی بر هوش مصنوعی تأثیر می‌گذارند. شکل ۱۴.۳ پیش‌بینی بلندمدت را نشان می‌دهد. هوش مصنوعی آینده به سه بخش تقسیم می‌شود: زمان حال، در حدود ۲۰ سال آتی و آینده‌ای دور.

آینده هوش تجاری و ابزارهای تحلیل در شکل ۱۴.۴ نشان داده شده است. برخی از پیش‌بینی‌های دیگر عبارتند از ابزارهای تحلیل هوشمند، بینش به عنوان یک سرویس، و فهرست‌نویسی داده‌ها. در نهایت، یک فناوری که ممکن است بر تجزیه و تحلیل و هوش مصنوعی تأثیر بگذارد را با جزئیات بیشتر توضیح می‌دهیم.

موضوعاتی که ارتباط نزدیکی با اینترنت اشیا، چت‌بات‌ها، خانه‌های هوشمند، ابزارهای تحلیل، حسگرها و «اشیاء» دارند در مفهوم **رایانش محیطی**^۱ (یا رایانش پارادایم) گنجانده شده‌اند. رایانش محیی تعاریف متعددی دارد، اما اساساً به محیط‌های الکترونیکی (مثلاً دستگاه‌های شبکه مانند حسگرها) اطلاق می‌شود که نسبت به افراد و محیط آنها حساس و پاسخگو هستند. بنابراین دستگاه‌های محیطی می‌توانند از افراد در هر کاری که انجام می‌دهند پشتیبانی کنند. هنگامی که دستگاه‌ها محیط اطراف خود را حس می‌کنند، روش‌های ورودی/خروجی متفاوتی را ارائه می‌کنند که به پیکربندی موقعیت‌ها بستگی دارد (به عنوان مثال، آنچه افراد در یک زمان خاص انجام می‌دهند).



شکل ۱۴.۴ آینده‌ی ابزارهای تحلیل

Source: "Analytics and BI Trends", Datapine, in Top 10 Analytics and Business Intelligence Trends for 2018, Business Intelligence, Dec 13th 2017, © 2017, Used with permission.

¹ Ambient Computing

در مجموع، همه چیز در زندگی ما رایانه‌ای و هوشمند خواهد بود. این مفهوم بر اساس تحقیقات قبلی در زمینه‌های رایانش فراگیر، تعامل انسان و ماشین، آگاهی از بافتار، پروفایل، شخصی‌سازی و طراحی تعامل است. برای جزئیات، به راهنمای en.wikipedia.org/wiki/Ambient_intelligence_and_Charara (2018) مراجعه کنید.

در حالی که این مفهوم عمدتاً آینده نگر است، ویژگی‌ها و مزایای آن قبلاً پیش‌بینی شده است. دستگاه‌های شبکه می‌توانند:

- افراد و سایر «اشیاء» و زمینه آنها را در هر زمان و مکان معین شناسایی کنند.
- با محیط و سیستم‌های موجود ادغام شوند.
- خواسته‌ها و نیازهای افراد را بدون درخواست پیش‌بینی کنند.
- خدمات هدفمند را بر اساس نیازهای مردم ارائه کنند.
- انعطاف پذیر باشند (یعنی می‌توانند اقدامات خود را در پاسخ به نیازها یا فعالیت‌های مردم تغییر دهند).
- نامرئی باشند

بسیاری از دستگاه‌ها و خدماتی که در این کتاب توضیح داده شده‌اند برخی از قابلیت‌های محاسبات محیطی را نشان می‌دهند. ال‌کسای آمازون احتمالاً در حال حاضر نزدیکترین فناوری به مفهوم محیطی است. برای جزئیات، به کوچ (2018) مراجعه کنید. برای اطلاعات بیشتر در مورد رایانش محیطی و ارتباط آن با اینترنت اشیا و شهرهای هوشمند، کونومی و روسوس (2016) را ببینید.

سوالاتی برای مرور بخش ۱۴.۸

۷. سه مورد از فهرست ۱۰ تایی گارتنر را که بیشترین ارتباط را با ابزارهای تحلیل و علم داده‌ها دارند شناسایی کنید.
۸. سه مورد از فهرست ۱۰ تایی گارتنر را که بیشترین ارتباط را با هوش مصنوعی و یادگیری ماشینی دارند شناسایی کنید.
۹. سه مورد از فهرست ۱۰ تایی گارتنر را که بیشترین ارتباط را با اینترنت اشیا، حسگرها و اتصال دارند شناسایی کنید.
۱۰. سه فناوری مرتبط با ابزارهای تحلیل را از لیست پیش‌بینی‌های دیگر شناسایی کنید و آنها را با جزئیات بیشتر بررسی کنید. گزارش بنویسید.
۱۱. سه فناوری مرتبط با علم داده‌ها را از فهرست طولانی شناسایی کنید و آنها را با جزئیات بیشتر بررسی کنید. گزارش بنویسید.
۱۲. سه فناوری مرتبط با هوش مصنوعی را از فهرست طولانی شناسایی کنید و آنها را با جزئیات بیشتر بررسی کنید. گزارش بنویسید.
۱۳. رایانش محیطی و سهم بالقوه آن در سیستم‌های هوشمند را شرح دهید.

یک توافق کلی میان کارشناسان هوش مصنوعی وجود دارد که هوش مصنوعی همه چیز را در دنیای ما به سمت بهتر شدن تغییر خواهد داد، (به عنوان مثال، لو-رام^۱ (۲۰۱۷) و ویولینو (۲۰۱۷) را ببینید). با این حال، در مورد زمان وقوع چنین تغییراتی و تأثیر آنها اختلاف نظر وجود دارد. تحقیقات هوش مصنوعی به دلیل پیشرفت در فناوری‌های رایانه‌ای مختلف مرتبط (مانند تراشه‌ها، اینترنت اشیا)، بهبود روش‌ها و ابزارهای هوشمند، افزایش فعالیت‌ها در شرکت‌های با فناوری پیشرفته که در تلاش برای به دست آوردن رهبری در حوزه‌های خاص سیستم‌های هوشمند هستند و شرکت‌هایی که میلیاردها دلار در هوش مصنوعی سرمایه‌گذاری می‌کنند، توسعه ابزارها و متدولوژی‌های هوش مصنوعی و موارد دیگر شتاب می‌گیرد. در این بخش، ابتدا به ارائه‌ی کارهایی که برخی از شرکت‌های بزرگ در زمینه فناوری‌های هوشمند انجام شده است می‌پردازیم.

یکی از راه‌های پیش‌بینی آینده‌ی هوش مصنوعی این است که به آنچه شرکت‌های بزرگ در حال حاضر انجام می‌دهند نگاه کنیم.

گوگل از پردازش زبان طبیعی در مترجم گوگل و همچنین در فرایندهای جستجوی خود استفاده می‌کند. از شبکه‌های عصبی در پایگاه داده‌های خود (برای تشخیص الگو) و برای تصمیم‌گیری در مورد آنها استفاده می‌کند. علاوه بر این، گوگل از الگوریتم‌های یادگیری ماشینی دیگر برای شخصی سازی تصمیمات تبلیغاتی استفاده می‌کند. دستیار گوگل و هوم دو پروژه‌ی کاربردی هستند که در سی‌ای‌اس ۲۰۱۸ توجه قابل توجهی را به خود جلب کردند. دستیار گوگل در تلاش است تا الکسا را از سلطنت خلع کند. علاوه بر این، گوگل بیشترین فعالیت را در زمینه خودروهایی خودران دارد. گوگل چندین شرکت هوش مصنوعی را خریداری کرد و در حال انجام تحقیقات گسترده در این زمینه است. گوگل یک تیم ویژه دارد که سعی دارد تا گفتگوی صوتی با یک شخصیت را در آن ارائه کند (ایدیسیکو ۲۰۱۷). گوگل دیپ‌مایند آلفاگو^۲ ماشینی است که قهرمانان بازی «گو» را شکست داد. گوگل از یادگیری ماشینی برای مدیریت پایگاه داده‌های عظیم و استراتژی‌های جستجوی خود استفاده می‌کند. در نهایت، گوگل با نمایش کلیپ‌های فیلم به ماشین‌های هوش مصنوعی خود می‌آموزد که چگونه رفتار کنند (مانند آشپزی) (گرشگورن^۳، ۲۰۱۷).

اپل به طور مخفیانه بر روی چندین پروژه هوش مصنوعی کار می‌کند. شناخته شده‌ترین آنها چت‌بات سیری است که در چندین محصول آن (به عنوان مثال، آیفون) تعبیه شده است. در سال ۲۰۱۶، اپل یک شرکت یادگیری ماشینی به نام توری^۴ را خریداری کرد. اگرچه اپل از گوگل، آمازون و مایکروسافت عقب افتاده است، ولی با استفاده از تملک‌ها و تحقیق و توسعه گسترده، به سرعت فاصله خود را کم می‌کند. اپل شرکت‌هایی را در زمینه تشخیص گفتار (ووکال!^۵)، تشخیص تصویر (پرسپشن^۶)، و تشخیص حالت چهره (ایموشن^۷) خریداری کرد. بنابراین، اپل در حال تبدیل شدن است به یک رهبر در عرصه‌ی هوش مصنوعی است. با چند صد میلیون کاربر سیری و خریدهای جدید در هوش مصنوعی، اپل به سرعت در حال پیشرفت است.

¹ Lev-Ram

² Google's DeepMind AlphaGo

³ Greshgorn

⁴ Turi

⁵ Vocal!

⁶ Perception

⁷ Emotion

مارک زاکربرگ، مدیر عامل فیسبوک، به آینده هوش مصنوعی معتقد است. او علاوه بر سرمایه‌گذاری شخصی خود در هوش مصنوعی، یان لی‌کان^۱، یک پیشگام در یادگیری عمیق را برای رهبری تحقیقات هوش مصنوعی در شرکت استخدام کرد. لی‌کان یک واحد ویژه فیسبوک را ایجاد کرد که پیشرفت‌های مهم هوش مصنوعی را شناسایی کرده و آنها را در محصولات فیسبوک ادغام می‌کند. فیسبوک میلیاردها دلار در هوش مصنوعی سرمایه‌گذاری کرد. با فیسبوک، هوش مصنوعی به جریان اصلی تبدیل می‌شود. فیسبوک با بیش از ۲ میلیارد کاربر خود، برنامه‌های کاربردی هوش مصنوعی خود را در سطح جهانی گسترش می‌دهد.

مایکروسافت در کلیه‌ی حوزه‌های تحقیقاتی فناوری هوش مصنوعی بسیار فعال است. در سال ۲۰۱۷، مالوبا^۲ یک شرکت نوپا در زمینه یادگیری عمیق و پردازش زبان طبیعی را خریداری کرد. برخی بر این باورند که این خرید به مایکروسافت کمک می‌کند تا از فیسبوک و گوگل در زمینه‌های مختلف بهتر عمل کند تشخیص گفتار و تصویر مالوبا در خواندن و درک متن از توانایی‌های انسانی در دستیار شخصی مجازی خود، کورتانا، برتر است. این دستیار به رفع مشکلات مردم در خصوص ایمیل و پیام‌رسانی کمک می‌کند. هوش مصنوعی محتوای پیام‌ها و هرگونه سند ذخیره شده و توصیه‌هایی را برای اقداماتی که باید انجام دهد را بررسی می‌کند. برای یک ویدیوی جامع درباره هوش مصنوعی امروز و فردا توسط دانشگاه استنفورد، [youtube.com/watch?v=wJqf17bZQsY](https://www.youtube.com/watch?v=wJqf17bZQsY) را تماشا کنید

آی‌بی‌ام در اوایل سال ۱۹۷۳ وارد حوزه رباتیک شد. در سال ۱۹۸۰، کیواس-وان را توسعه داد. در سال ۱۹۷۷، دیپ بلو را توسعه داد. و در سال ۲۰۱۴، یک آی‌بی‌ام بالغ وارد صحنه شد. آی‌بی‌ام همچنین به خاطر پروژه مغز مصنوعی خود شناخته شده است. (در خصوص بلو برین، به artificialbrains.com/blue-brain-project مراجعه کنید. آی‌بی‌ام همچنین به خاطر پروژه‌ی دیپ کیوای خود شناخته شده است.

آی‌بی‌ام در تحقیقات هوش مصنوعی، به ویژه در زمینه رایانش شناختی، بسیار فعال است. به فصل ۶ و research.ibm.com/ai/ مراجعه کنید. آی‌بی‌ام واتسون با همکاری آزمایشگاه‌های هوش مصنوعی ام‌آی‌تی توسعه یافته است. برخی دیگر از پروژه‌های فعلی بر روی نرم‌افزارهای یادگیری عمیق توزیع شده، ایجاد تریلرهای موسیقی و فیلم توسط ماشین‌ها، تشخیص حرکات، ترکیب هوش مصنوعی و اینترنت اشیا (به عنوان مثال، شناخت تجسم یافته)، و برنامه‌های کاربردی پزشکی پشتیبانی شده توسط واتسون (مراقبت شناختی، به عنوان مثال، تشخیص سرطان، مراقبت از سلامت روان، و افراد کم بینا) تمرکز دارند. آی‌بی‌ام واتسون در حال حاضر قوی‌ترین برند کاربرد هوش مصنوعی محسوب می‌شود. انتظار می‌رفت که یک میلیارد کاربر در سال ۲۰۱۸ از آن استفاده کنند و مزایای قابل توجهی از برنامه‌های کاربردی آن به دست آورند.

تحقیقات هوش مصنوعی در بسیاری از کشورها، به ویژه آلمان، ژاپن، فرانسه، بریتانیا و هند انجام می‌شود. اما بیشتر تحقیقات در خارج از ایالات متحده در چین انجام می‌شود. چین قصد دارد رهبر جهانی هوش مصنوعی باشد و دولت آن به شدت از فعالیت‌های بسیاری از شرکت‌های هوش مصنوعی حمایت می‌کند. همانطور که ممکن است از فصل ۱ به یاد داشته باشید، ولادیمیر پوتین گفته است که هر کسی که هوش مصنوعی را رهبری کند، جهان را کنترل خواهد کرد و در واقع، چین قصد دارد تا سال ۲۰۳۰ رهبر جهان باشد. در میان بسیاری از شرکت‌هایی که در زمینه هوش مصنوعی فعالیت می‌کنند، سه شرکت میلیاردها دلار سرمایه‌گذاری کرده، هزاران

¹ Yann LeCun

² Maluuba

متخصص هوش مصنوعی و مهندسان رباتیک را جذب کرده و استعدادهای جهانی در زمینه هوش مصنوعی را بکار گرفته‌اند. این سه شرکت گروه *علی بابا*، *تِنسِنِت* و *بایدو* هستند. هوش مصنوعی در حال حاضر اولویت دولت چین است. در داستان روی جلد مجله‌ی فورچون، لاشینسکی (۲۰۱۸) رقابت بین *تِنسِنِت* و *علی بابا* را توضیح داده و تحلیل می‌کند.

این غول تجارت الکترونیک یک آزمایشگاه بزرگ هوش مصنوعی برای مدیریت فعالیت‌های هوش مصنوعی خود ایجاد کرده است. هدف از اینکار بهبود قابلیت‌های هوش مصنوعی و پشتیبانی از تصمیم‌گیری در زمینه‌های زیر است: بینایی رایانه‌ای، پردازش زبان طبیعی، تشخیص گفتار، یادگیری ماشینی و چت‌بات‌ها. هوش مصنوعی در حال حاضر در بیش از ۱۰۰ محصول *تِنسِنِت* از جمله وی‌چَت و کیوکیو^۲ موجود است. شعار معروف هوش مصنوعی در چین "Juey, GO AI" است. *تِنسِنِت* از شرکت رباتیک یوبی‌تک آلفا^۳ پشتیبانی می‌کند. *تِنسِنِت* بزرگترین شرکت اینترنتی جهان است و هوش مصنوعی عملکرد آن را بهبود می‌بخشد. شعار دیگر آن « هوش مصنوعی در همه‌چیز » است. این شرکت یک آزمایشگاه در پلوو^۴، واشنگتن دارد. مراقبت‌های سلامت در آنجا اولویت اصلی تحقیقاتی است. برای اطلاعات بیشتر در مورد هوش مصنوعی در *تِنسِنِت*، نگاه کنید به مار^۵ (۲۰۱۸).

بایدو تحقیقات پردازش زبان طبیعی را پنج سال قبل از گوگل برای بهبود قابلیت‌های موتور جستجوی خود آغاز کرد. این شرکت در دره سیلیکون، سیاتل و پکن واقع شده است. *بایدو* چندین محصول دارد. یکی از آنها *دوئر اُس*^۶ است که یک دستیار صوتی است در بیش از ۱۰۰ برند لوازم خانگی در چندین کشور تعبیه شده است. این محصول اکنون برای گوشی‌های هوشمند بهینه شده است. *بایدو* همچنین بر روی وسایل نقلیه خودران کار می‌کند. در نهایت، این شرکت بکارگیری قابلیت تشخیص چهره را در سازمان‌ها ترویج می‌کند (به جای تگ‌های شناسایی). هوش مصنوعی *بایدو* در حال رشد است اما هنوز بسیار کوچکتر از هوش مصنوعی *علی‌بابا* است.

این *علی‌بابا* بزرگترین شرکت تجارت الکترونیک جهان و ارائه‌دهنده رایانش ابری و پلتفرم‌های اینترنت اشیا در پروژه‌های هوش مصنوعی فعال است و در شرکت‌های هوش مصنوعی، مانند *غول* تشخیص چهره *سِنس‌تایم*^۷ سرمایه‌گذاری کرده است. *علی‌بابا* روشی را برای انجام هوش مصنوعی توسعه داده است که در مورد کاربردی ۱۴.۳ توضیح داده شده است.

علی‌بابا. کام چگونه هوش مصنوعی را اجرا می‌کند

مورد کاربردی ۱۴.۳

علی بابا یک مدل مبتنی بر ابر به نام *ای‌تی برین*^۸ را alibabacloud.com/et توسعه داده است. منطق این است که امروز و در آینده نزدیک، ما در محیط رایانش ابری تجارت می‌کنیم و خواهیم کرد. محتوا، دانش و داده‌ها در فضای ابری هستند و *علی بابا* هم کاربر و هم ارائه دهنده آی‌کلاد^۹ است. مدل *ای‌تی برین* در شکل ۱۴.۶ نشان داده شده است.

¹ Tencent

² QQ

³ UBTech Alpha

⁴ Bellevue

⁵ Marr

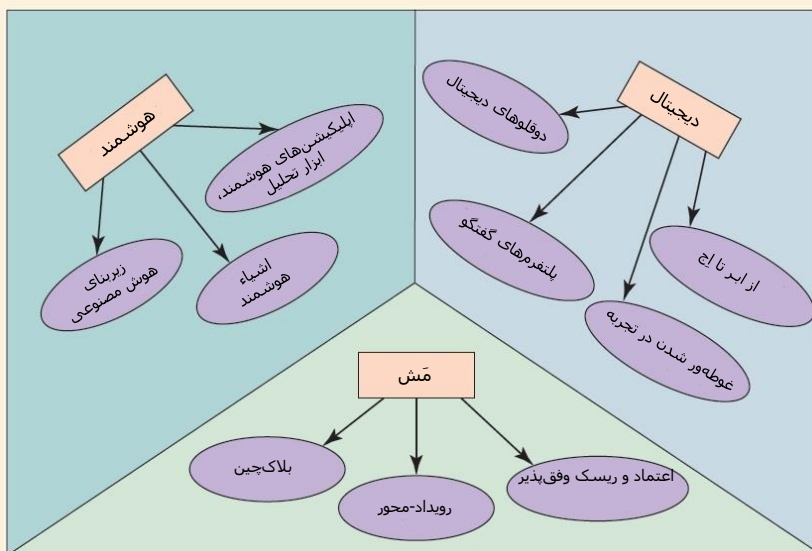
⁶ Duer OS

⁷ SenseTime

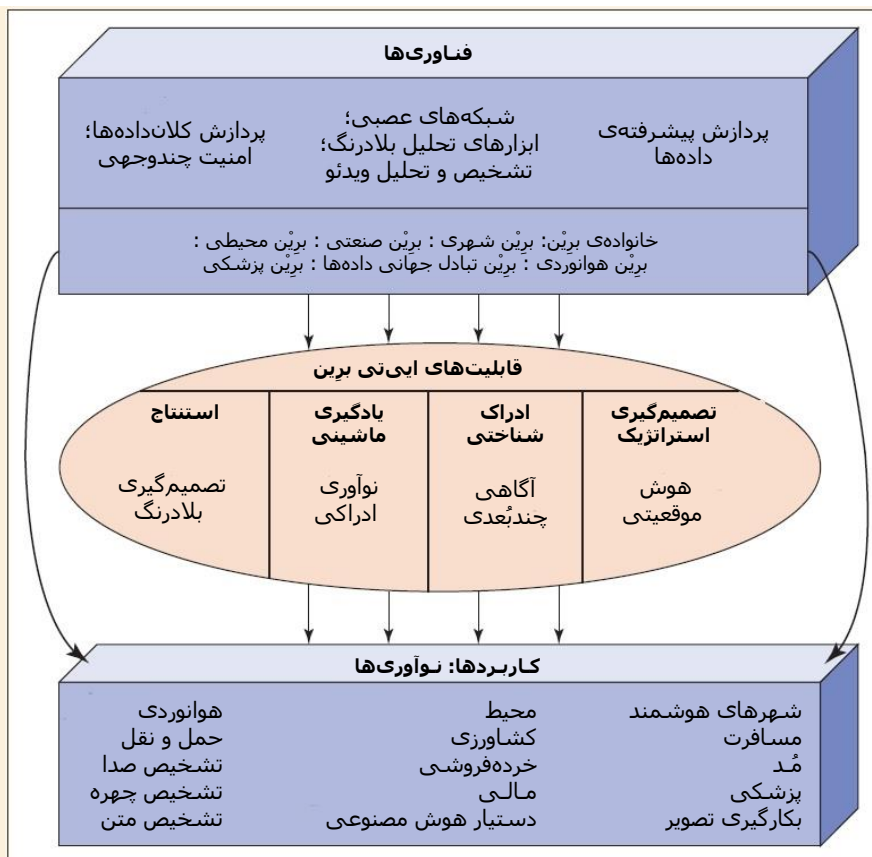
⁸ ET Brain

⁹ iCloud

ای تی برین از سه بخش تشکیل شده است: فناوری‌ها، قابلیت‌ها و برنامه‌ها. فناوری‌ها شامل کلان‌داده‌ها و پردازش تحلیلی، شبکه‌های عصبی، تحلیل تشخیص ویدیو و یادگیری ماشینی هستند. این فناوری‌ها چهار قابلیت عمده مانند ادراک شناختی، استدلال، تصمیم‌گیری بلادرنگ و یادگیری ماشین (بخش میانی تصویر را در شکل ببینید). این قابلیت‌ها حجم زیادی از برنامه‌های کاربردی مانند فعالیت‌های تجارت الکترونیک (کسب‌وکار به کسب‌وکار و کسب‌وکار به مصرف‌کنندگان)، مراقبت‌های پزشکی و سلامت، شهرهای هوشمند، کشاورزی، مسافرت، امور مالی و هوانوردی پشتیبانی کرده و محرک آنها هستند. در مجموع، این یک پلتفرم هوش مصنوعی فوق‌هوشمند است. ای تی برین در یک ویدئوی ۲۹:۲۶ دقیقه‌ای در [youtube.com/watch?v=QmkPDtQTarY](https://www.youtube.com/watch?v=QmkPDtQTarY) نشان داده شده است.



شکل ۱۴.۵ پیش‌بینی گارتنر



شکل ۱۴.۶ مدل ای تی برین علی‌بابا

ماموریت علی بابا دستیابی به ۲ میلیارد مصرف‌کننده و کمک به ۱۰ میلیون کسب‌وکار در سراسر جهان است. برای دستیابی به این ماموریت، این شرکت در هفت آزمایشگاه تحقیقاتی سرمایه‌گذاری کرد که بر هوش مصنوعی، یادگیری ماشینی، پردازش زبان طبیعی، تشخیص چهره (تصویر)، و امنیت تمرکز دارند. شبکه علی‌بابا از هوش مصنوعی برای بهینه‌سازی زنجیره تامین، شخصی‌سازی توصیه‌ها و ارائه دستیارهای شخصی مجازی استفاده می‌کند. علی بابا بر روی چندین صنعت و خرید از فروشگاه‌های فیزیکی با پشتیبانی هوش مصنوعی متمرکز شده است. به عنوان مثال، این شرکت در دفتر هوش مصنوعی خود در هنگ کنگ، «فَشِنِ ای‌آی^۱» را افتتاح کرد که با شرکت گِس^۲ همکاری می‌کند و به خریداران کمک می‌کند تا زمانی که در یک فروشگاه فیزیکی هستند یک دست لباس را به صورت آنلاین ایجاد کنند. نگاه کنید به engadget.com/2018/07/04/guess-alibaba- **aifashion-** این شرکت قصد دارد جهان را با هوش مصنوعی بهم متصل کند (نایت، ۲۰۱۸) و ممکن است تجارت جهان را کنترل کند.

سوالاتی برای مورد کاربردی ۱۴.۳

¹ Fashion AI

² Guess Inc.

۱. ارتباط بین رایانش ابری و هوش مصنوعی را توضیح دهید.
۲. منطق مدل ای‌تی برین را توضیح دهید.
۳. برای یافتن فعالیت‌های اخیر علی‌بابا در زمینه هوش مصنوعی در وب جستجو کنید.
۴. لاشینسکی (۲۰۱۸) را مطالعه کنید. چرا علی‌بابا درگیر رقابت شدید با تسنت است؟

Sources: Compiled from W. Knight. (2018, March 7). "Inside the Chinese Lab That Plans to Rewire the World with AI." *MIT Technology Review*; Marr, B. (2018, June 4). "Artificial Intelligence (AI) in China: The Amazing Ways Tencent Is Driving Its Adoption." *Forbes*; A. Lashinsky. (2018, June 21). "Alibaba v. Tencent: The Battle for Supremacy in China." *Fortune*. alibabacloud.com/et.

در حال حاضر شرکت‌های آمریکایی از شرکت‌های چینی جلوتر هستند. با این حال، این وضعیت ممکن است در آینده به دلیل سرمایه‌گذاری‌های هنگفت بر روی هوش مصنوعی در چین و حمایت‌های دولت چین تغییر کند. توجه داشته باشید که موضوع اصلی مذاکرات تجاری ایالات متحده و چین در سال ۲۰۱۸ بر استفاده از فناوری توسط شرکت‌های چینی است که از دانش و اسرار تجاری ایالات متحده استفاده می‌کنند.

به گفته مک کراکن^۱ (۲۰۱۷)، فناوری‌های هوشمند بزرگترین فرصت را برای شرکت‌های فناوری از زمان رایانش موبایل فراهم می‌کنند. به همین دلیل است که غول‌های فناوری و شرکت‌های نوپا سعی می‌کنند از هوش مصنوعی سوء استفاده کنند دژاردن^۲ (۲۰۱۷) یک اینفوگرافیک در مورد تأثیر آینده هوش مصنوعی ارائه می‌دهد که شامل ۱۵.۷ تریلیون دلار تا سال ۲۰۳۰ به شکل افزایش بهره‌وری و افزایش هزینه‌گرد مصرف‌کننده است. تا سال ۲۰۱۸، غول‌های فناوری و سایر شرکت‌ها ۳۰ میلیارد دلار در تحقیق و توسعه و ۱۳.۲ میلیارد دلار در شرکت‌های نوپا سرمایه‌گذاری خواهند کرد. انتظار می‌رود بیشترین پیشرفت در محصولات تشخیص تصویر و گفتار انجام پذیرد.

توجه داشته باشید که فیسبوک، آمازون، گوگل، آی‌بی‌ام و مایکروسافت علیرغم رقابت‌شان، مشارکتی را برای تحقیق در مورد پیشرفت‌ها و بهترین شیوه‌های هوش مصنوعی راه‌اندازی کرده‌اند.

اکنون که خواندن این کتاب را به پایان رساندید، ممکن است بپرسید، "چه اتفاقی برای فناوری‌های هوشمند در آینده خواهد افتاد؟" تاثیر قابل توجهی بر کسب‌وکار و کیفیت زندگی خواهد داشت. تغییراتی بسیار شگرفت و چشمگیر بوجود خواهد آمد. با سرمایه‌گذاری‌های میلیارد دلاری که عمدتاً در حوزه‌ی هوش مصنوعی هستند، پیشرفت‌هایی بدست خواهد آمد. ماشین‌ها هوشمندتر و هوشمندتر می‌شوند. به عنوان مثال، دستگاه کپی رایتینگ علی‌بابا، که مبتنی بر یادگیری عمیق و پردازش زبان طبیعی است، می‌تواند ۲۰,۰۰۰ خط متن را در یک ثانیه تولید کند. ماشین آنقدر باهوش است که تست تورینگ (فصل ۲) را با موفقیت پشت سر گذاشت، به این معنی که مانند انسان باهوش است اما می‌تواند بسیار سریعتر از انسان کار کند. اکنون به دو حوزه خواهیم پرداخت: کسب‌وکار و کیفیت زندگی.

¹ McCracken

² Desjardins

طبق گفته کورزر^۱ (۲۰۱۷)، ممکن است چالش‌هایی وجود داشته باشد، اما انتظار می‌رفت که هوش مصنوعی تا سال ۲۰۱۸ شکوفا شود. به احتمال بسیار زیاد شاهد افزایش تجاری‌سازی هوش مصنوعی به ویژه در بازاریابی، خدمات مالی، تولید و پشتیبانی فناوری اطلاعات خواهیم بود. به عنوان مثال، کیفیت و تجربه مشتری را می‌توان بهبود بخشید، برنامه‌های کاربردی هوش مصنوعی و اینترنت اشیاء را تقویت کرد. کورزر همچنین پیش‌بینی کرد که فرایندهای کنسگرانه‌تری به جای فرایندهای واکنشی بین وجود خواهد داشت. همکاری انسان و ماشین بیش از پیش وجود داشته و بسیاری از مشاغل خودکار خواهند شد، مشاغل جدید زیادی ایجاد خواهند شد. به دلیل افزایش قابلیت‌های چت‌بات‌ها و دستیاران شخصی مانند الکسا، سیری و دستیار گوگل، هوش مصنوعی مکالمه‌ای بیشتری وجود خواهد داشت. گارتنر پیش‌بینی کرد که تا پایان این دهه، مکالمه مردم با ماشین‌ها بیشتر از گفتگو با اعضای خانواده خواهد بود. (gartner.com/smarterwithgartner/gartnerpredicts-a-virtual-world-of-exponential-change/). یکی دیگر از حوزه‌هایی که کاربردهای امیدوارکننده دارد، تشخیص تصویر است. گوگل یک عامل محرک اصلی در هوش مصنوعی محاوره‌ای و تشخیص تصویر است.

هوش مصنوعی تأثیراتی بر زندگی خواهد داشت که نحوه رانندگی، غذا خوردن، سرگرمی، دریافت خدمات، یادگیری و مبارزه را تغییر خواهد داد. حوزه مهمی که سیستم‌های هوشمند هوش مصنوعی قبلاً در آن پیشرفت کرده اند، حوزه مراقبت سلامت است. برنارد تایسون^۲، مدیر عامل کیزر پرفورمنس^۳، بیانیه عمومی زیر را بیان کرد: «من فکر نمی‌کنم هیچ پزشکی بدون کمک هوش مصنوعی بتواند کار کند. غیرممکن است بدون هوش مصنوعی الگوها را انتخاب کنید، روندها را انتخاب کنید و واقعاً بر فعالیت مراقبت سلامت نظارت داشته باشید.» طبق گزارش ادیتورز (۲۰۱۸) راه حل‌های هوشمند می‌توانند شاخص‌های کیفیت زندگی را ۱۰ تا ۳۰ درصد بهبود بخشند. (هرچه بیشتر منتظر بمانیم، درصد بالاتر خواهد بود.) از جمله شاخص‌هایی که آنها ذکر می‌کنند عبارتند از: داشتن زندگی طولانی‌تر و سالم‌تر، کاهش انتشار گازهای گلخانه‌ای، نجات جان ۲۰۰,۰۰۰ نفر در سراسر جهان طی ۱۰ سال (به لطف خودروهای خودران)، کاهش زمان رفت و آمد مردم (مشکلات ترافیکی کمتر)، افزایش تعداد مشاغل (به عنوان مثال، توسط فناوری‌های جدید و محیط‌های تجاری مولدتر)، و ارائه مسکن بهتر و مقرون به صرفه‌تر.

وسایل نقلیه خودران، از جمله هواپیماهای بدون سرنشین، به وضوح زندگی ما را تغییر خواهند داد و ربات‌ها می‌توانند به ما (به ویژه افراد مسن و بیماران) خدمت‌رسانی کرده، ما را سرگرم کرده و در صورت مدیریت صحیح، همراه ما باشند. برای تأثیر هوش مصنوعی در آینده در جامعه، ویدیو را در [youtube.com/watch?v=KZz6f-nCCN8/](https://www.youtube.com/watch?v=KZz6f-nCCN8/) تماشا کنید. نتایج ناخواسته چه خواهد بود؟ چه می‌شود اگر ربات‌ها همه ما را بکشند؟ خب، احتمالاً هرگز این اتفاق نخواهد افتاد. مردم آنقدر باهوش هستند که مطمئن شوند که فقط نتایج خوب از سیستم‌های هوشمند حاصل می‌شود.

سوالاتی برای مرور بخش ۱۴.۹

۱. فعالیت‌های هوش مصنوعی شرکت‌های بزرگ فناوری ایالات متحده را شرح دهید.

¹ Kurzer

² Bernard Tyson

³ Kaiser Performance

۲. کار شرکت های غول پیکر چینی را شرح دهید.
۳. رویکرد علی بابا به هوش مصنوعی (مدل ای تی برین) را توضیح دهید.

- سیستم‌های هوشمند می‌توانند سازمان‌ها را به طرق مختلف از قبیل سیستم‌های مستقل یا سیستم‌های یکپارچه یا با سایر سیستم‌های اطلاعات رایانه‌ای تحت تاثیر قرار دهند.
- تأثیر ابزارهای تحلیل بر افراد متفاوت است و می‌تواند مثبت، خنثی یا منفی باشد.
- ممکن است با معرفی سیستم‌های هوشمند مسائل حقوقی جدی ایجاد شود؛ مهمترین بخش‌های مشکل‌ساز مربوط به مسئولیت‌پذیری و حریم خصوصی هستند.
- سیستم‌های هوشمند می‌توانند پیامدهای اجتماعی مثبتی به همراه داشته باشند. این فرصت‌ها شامل رهبری مردم برای مبارزه با تروریسم نیز می‌شوند. کیفیت زندگی، چه در محل کار و چه در خانه، احتمالاً در نتیجه‌ی استفاده از این فناوری‌ها بهبود می‌یابد. البته، مسائل بالقوه منفی وجود دارد که باید نگران آن بود.
- رشد سیستم‌های هوشمند منجر به تغییرات عمده در ساختار صنعت و اشتغال آینده خواهد شد.
- یک نبرد بزرگ در مورد اینکه چه کسی مالک داده‌های کاربرانی است که از تلفن‌های هوشمند، اتومبیل‌ها و غیره استفاده می‌کنند در جریان است.
- در استقرار سیستم‌های هوشمند، رعایت مسائل حقوقی، حریم خصوصی و اخلاقی ضروری است.
- قرار دادن ربات‌ها به عنوان همکار در محل کار مسائل حقوقی و اخلاقی را مطرح می‌کند.
- فناوری‌های هوشمند ممکن است بر فرایندهای کسب‌وکار، ساختار سازمانی و شیوه‌های مدیریت تأثیر بگذارد.
- ممکن است ایجاد سازمان مستقل و واحدهایی که سیستم‌های هوشمند را جاری‌سازی و مدیریت می‌کنند ضرورت داشته باشد.
- سیستم‌های هوشمند ممکن است مزیت رقابتی قابل توجهی را برای کاربران خود فراهم کنند.
- سیستم‌های هوشمند ممکن است موجب از بین رفتن مشاغل معمولی و متوسط شوند.
- در نهایت سیستم هوشمند ممکن است باعث بیکاری حتی در مشاغل نیازمند مهارت شود. بنابراین ممکن است نیاز به آموزش مجدد باشد.
- سیستم‌های هوشمند ممکن است منجر به تجدیدساختار بسیاری از مشاغل به ویژه از طریق همکاری انسان و ماشین شوند.
- سیستم‌های هوشمند مشاغل جدید بسیاری ایجاد خواهند کرد که نیاز به آموزش تخصصی دارند.
- استفاده از اتوماسیون سیستم‌های هوشمند ممکن است منجر به یک هفته کاری کوتاه‌تر و نیاز به جبران خسارت افرادی که شغل‌شان را از دست داده‌اند شود.
- برخی افراد از عواقب ناخواسته‌ی داشتن هوش مصنوعی و ربات‌ها می‌ترسند. ماشین‌ها یاد می‌گیرند و ممکن است به انسان آسیب برسانند.

--	--	--

۱. برخی می‌گویند که تحلیل به طور کلی مدیریت فعالیت‌ها را غیرانسانی می‌کند و دیگران می‌گویند که اینطور نیست. در مورد استدلال برای هر دو دیدگاه بحث کنید.
۲. تشخیص عفونت و تجویز دارو نقاط ضعف بسیاری از پزشکان شاغل است. بنابراین، به نظر می‌رسد که در صورتیکه پزشکان در جامعه از سیستم‌های تشخیص مبتنی بر تحلیل بهره ببرند بسیار بهتر خواهد بود. به سوالات زیر پاسخ دهید:
 - . به نظر شما چرا پزشکان از چنین سیستم‌هایی کم استفاده می‌کنند؟
 - . فرض کنید شما مدیر یک بیمارستانی هستید که پزشکان حقوق می‌گیرند و به شما گزارش می‌دهند. برای اینکه آنها را متقاعد کنید تا از سیستم‌های هوشمند استفاده کنند چکار می‌کنید؟
 - . اگر مزایای بالقوه‌ی سیستم‌های هوشمند بسیار زیاد هستند آیا جامعه می‌تواند کاری کند تا پزشکان از چنین سیستم‌هایی استفاده کنند؟
۳. برخی از نگرانی‌های اصلی حفظ حریم خصوصی در بکارگیری سیستم‌های هوشمند بر روی داده‌های تلفن همراه چیست؟
۴. برخی از موارد نقض حریم خصوصی کاربران را در ادبیات فعلی شناسایی کرده و تأثیر آنها بر علم داده‌ها را به عنوان یک حرفه بنویسید.
۵. برخی می‌ترسند که ربات‌ها و هوش مصنوعی همه ما به قتل برسانند. برخی دیگر مخالف این عقیده هستند. در این خصوص بحث کنید.
۶. برخی ادعا می‌کنند که به موضوع هوش مصنوعی بیش از حد انتظار پرداخته شده است. در این خصوص بحث کنید. یک سوال از کوئورا پرسیده و پنج پاسخ آن را تجزیه و تحلیل کنید.
۷. برخی ادعا می‌کنند که هوش مصنوعی ممکن است به یک موضوع حقوق بشر تبدیل شود (صفیه نوبل^۱ را جستجو کنید). در این خصوص بحث و مناظره کنید.
۸. درباره تأثیر بالقوه جی‌دی‌پی‌آر^۲ بر حریم خصوصی امنیت و تبعیض بحث کنید.
۹. در مورد اخلاق و عدالت در یادگیری ماشینی بحث کنید. (پاکزاد^۳ ۲۰۱۸ را مطالعه کنید).
۱۰. آیا ربات‌ها باید مانند کارگران مالیات بدهند؟ موریس (۲۰۱۷) را بخوانید و در مورد جوانب مثبت و منفی موضوع بنویسید.

¹ Safiya Noble

² GDPR

³ Pakzad

۳۴. مسائل اخلاقی مرتبط با تصمیم‌گیری مدیریتی را شناسایی کنید. در اینترنت جستجو کنید، به گروه‌های بحث/وبلاگ‌ها بپیوندید و مقالاتی از اینترنت بخوانید. گزارشی در مورد یافته‌هایتان تهیه کنید.
۳۵. در اینترنت جستجو کنید تا مثال‌هایی از اینکه سیستم‌های هوشمند چگونه می‌توانند فعالیت‌هایی مانند تفویض اختیار، سفارشی‌سازی انبوه و کار تیمی را تسهیل کنند پیدا کنید.
۳۶. در مورد مرکز منابع فناوری انجمن وکلای آمریکا تحقیق و بررسی کنید
americanbar.org/groups/departments_offices/legal_technology_resources.html و nolo.com
 دغدغه‌های مهم حقوقی و اجتماعی در مورد سیستم‌های هوشمند کدامها هستند؟ چگونه به آنها رسیدگی می‌شود؟
۳۷. در چندین سایت مرتبط با مراقبت سلامت کاوش کنید (مانند WebMD.com و who.int). مسائل مربوط به هوش مصنوعی و حریم خصوصی را پیدا کنید. گزارشی در مورد اینکه چگونه این سایت‌ها پیشنهادهای را بهبود حریم خصوصی ارائه می‌کنند بنویسید.
۳۸. به وب سایت Humanyze.com بروید. مطالعات موردی مختلف را مرور کنید و یک کاربرد جالب از حسگرها در درک مبادلات اجتماعی در سازمان‌ها را خلاصه‌سازی کنید.
۳۹. در مورد موضوع دستیار صوتی و حفاظت از حریم خصوصی تحقیق کنید. اینکار را با مطالعه‌ی کالینز^۱ (۲۰۱۷) و هاف^۲ (۲۰۱۷) شروع کنید.
۴۰. آیا اعطای حقوق به ربات‌های پیشرفته ایده خوبی است یا خیر؟ برای شروع کوتاسووا (۲۰۱۸) را مطالعه کنید
۴۱. اپلیکیشن‌های تشخیص چهره و صدا در حال رشد هستند. در مورد وضعیت مقررات آنها در یک کشور به انتخاب خودتان تحقیق کنید.
۴۲. در مورد مسائل اخلاقی خودروهای خودران تحقیق کنید. شروع با مطالعه‌ی هیمل‌رایش (۲۰۱۸) اینکار را شروع کنید.
۴۳. آیا سازمان شما برای هوش مصنوعی آماده است؟ در مورد این موضوع تحقیق کنید و تمام فعالیت‌های اصلی که شامل می‌شود را بیابید.
۴۴. در مورد نقش اینترنت اشیاء به عنوان ابزاری برای اتصال بین حسگرها و ابزارهای تحلیل تحقیق کرده و گزارشی بنویسید
۴۵. برخی افراد می‌گویند که ربات‌ها و چت‌بات‌ها ممکن است ریسک و کارمزد بیمه را افزایش دهند. در این مورد تحقیق کنید و گزارشی بنویسید.
۴۶. ویدیویی را در youtube.com/watch?v=wwuovuCfDU/ تماشا کنید و در مورد تأثیرات بالقوه‌ی ربات نظر دهید.
۴۷. در مورد موضوعی که در گیومه آمده است تحقیق کنید: «آیا هوش مصنوعی کنترل جهان را به دست خواهد گرفت؟»، «ربات‌ها چه زمان شورش خواهند کرد؟» و «آیا هوش مصنوعی کنترل کارخانجات را در دست خواهد گرفت؟» با خواندن مگوایر^۳ (۲۰۱۷) شروع کنید و advancedmp.com/artificial-intelligence/ را بخوانید و گزارشی بنویسید.

¹ Collins

² Huff

³ Maguire

۴۸. چویی و همکاران را بخوانید (۲۰۱۶) و در مورد حوزه‌هایی که در آنها ماشین‌ها می‌توانند و نمی‌توانند (فعلاً) جایگزین انسان‌ها شوند تحقیق کنید. تغییرات را از سال ۲۰۱۶ پیدا کنید. گزارشی بنویسید.
۴۹. ویدئوی ۳:۳۸ دقیقه‌ای را در [youtube.com/watch?v=78-1Mlkxyql/](https://www.youtube.com/watch?v=78-1Mlkxyql/) تماشا کرده و ارتباط آن را با پیش‌بینی‌های ماسک در مورد سلطه‌ی ربات‌ها بر جهان توضیح دهید.
۵۰. گزارش ساس در مورد اخلاق هوش مصنوعی را در sas.com/en_us/insights/articles/analytics/artificial-intelligence-ethics.html مطالعه کنید. در مورد هر یک از سه مرحله پیشنهادی نظر دهید. همچنین در مورد همکاری انسان و ماشین در حل مسئله نظر دهید.

<http://canoepoapp.com/> (accessed October 2018).

<http://amazon.org.br/en/imprensa/mapping-changein-the-amazon-how-satellite-images-are-haltingdeforestation/> (accessed October 2018).

<http://www.earthcastdemo.com/2018/07/bloombergearthcast-customizing-weather/> (accessed October 2018)

<https://www.worldbank.org/en/news/pressrelease/2018/02/22/world-bank-supports-sierraleones-efforts-in-landslide-recovery> (accessed October 2018)

Siemens.com. About Siemens. siemens.com/about/en/ (accessed September 2018).

Silvaris.com. Silvaris [overview.silvaris.com](http://silvaris.com/overview) (accessed September 2018).

Agah, A. (2017). *Medical Applications of Artificial Intelligence*. Boca Raton, FL: CRC Press.

Anthony, R. N. (1965). *Planning and Control Systems: A Framework for Analysis*. Cambridge, MA: Harvard University Graduate School of Business.

Baker, J., and M. Cameron. (1996, September). "The Effects of the Service Environment on Affect and Consumer Perception of Waiting Time: An Integrative Review and Research Propositions." *Journal of the Academy of Marketing Science*, 24, pp. 338–349.

Bean, R. (2017, May 8). "How Big Data Is Empowering AI and Machine Learning at Scale." *MIT Sloan Management Review*.

Boddington, P. (2017, January 3). "Case Study: Robot Camel Jockeys. Yes, really." *Ethics for Artificial Intelligence*.

Brainspace. (2016, June 13). "Augmenting Human Intelligence." *MIT Technology Review Insights*.

Butner, K. (2018, January 8). "Combining Artificial Intelligence with the Internet of Things Could Make Your Business Smarter." IBM Consulting Blog. CDC.gov. (2015, September 21). "Important Facts about Falls." cdc.gov/homeandrecreational/safety/falls/adultfalls.html (accessed September 2018).

Charles, T. (2018, May 21). "Influence of the External Environment on Strategic Decision." *Azcentral*. **yourbusiness**. azcentral.com/influence-external-environment-strategicdecisions-17628.html/ (accessed October 2018).

Chiguluri, V., Guthikonda, K., Slabaugh, S., Havens, E., Peña, J., & Cordier, T. (2015, June). Relationship Between Diabetes Complications and Health Related Quality of Life Among an Elderly Population in the United States. Poster presentation at the American Diabetes Association Seventy- Fifth Annual Scientific Sessions. Boston, MA.

Chongwatpol, J., & R. Sharda. (2010, December). "SNAP: A DSS to Analyze Network Service Pricing for State Networks." *Decision Support Systems*, 50(1), pp. 347–359.

Chung, C. (2016). "Dubai Camel Race Ride-Along." **YouTube.com**. youtube.com/watch?v=xFCRhk4GYds (accessed September 2018). Cordier, T., Slabaugh, L., Haugh, G., Gopal, V., Cusano, D., Andrews, G., & Renda, A. (2015, September). Quality of Life Changes with Progressing Congestive Heart Failure. Poster presentation at the Nineteenth Annual Scientific Meeting of the Heart Failure Society of America, Washington, DC.

Corea, F. (2017, December 1). "The Convergence of AI and Blockchain: What's the Deal?" **Medium.com**. medium.com/@Francesco_AI/the-convergence-of-ai-andblockchain-whats-the-deal-60c618e3acce (accessed September 2018).

- Davenport, T., & SAS Institute Inc. (2014, February). "Analytics in Sports: The New Science of Winning." sas.com/content/dam/SAS/en_us/doc/whitepaper2/iia-analyticsin-sports-106993.pdf (accessed September 2018).
- Davenport, T. H., & Patil, D. J. (2012). "Data Scientist." *Harvard Business Review*, 90, 70–76.
- Deahl, D. (2018, January 7). "This Automatic Feeder Can Tell the Difference Between Your Pets." *The Verge*.
- De Smet, A., et al. (2017, June). "Untangling Your Organization's Decision Making." *McKinsey Quarterly*.
- Duncan, A. (2016). "The BICC Is Dead." <https://blogs.gartner.com/alan-duncan/2016/03/11/the-bicc-isdead/> (accessed October 2018).
- Dundas.com. "How Siemens Drastically Reduced Cost with Managed BI Applications." www.dundas.com/Content/pdf/siemens-case-study.pdf (accessed September 2018).
- Eckerson, W. (2003). *Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions*. Seattle, WA: The Data Warehousing Institute.
- eMarketer. (2017, May). "Artificial Intelligence: What's Now, What's New and What's Next?" EMarketer Inc. Emc.com. (n.d.). "Data Science Revealed: A Data-Driven Glimpse into the Burgeoning New Field." emc.com/collateral/about/news/emc-data-science-study-wp.pdf (accessed September 2018)
- Emmert, Samantha. (2018, March 19). "Fighting Illegal Fishing." Global Fishing Watch. globalfishingwatch.org/research/fighting-illegal-fishing/ (accessed October 2018).
- Faggela, D. (2017, August 24). "Artificial Intelligence Plus the Internet of Things (IoT): 3 Examples Worth Learning From." *TechEmergence*.
- Faggela, D. (2018, March 29). "Artificial Intelligence Industry: An Overview by Segment." *TechEmergence*.
- Fernandez, J. (2017, April). "A Billion People a Day. Millions of Elevators. No Room for Downtime." IBM developer Works Blog. developer.ibm.com/dwblog/2017/konewatson-video/ (accessed September 2018).
- FinTech Futures. (2017, October 11). "IBM Combining Data Science and AI for Analytics Advance." **BankingTech.com**.
- Gartner, Inc. (2004). Using Business Intelligence to Gain a Competitive Edge. A special report.
- Gates, S., Smith, L. A., Fisher, J. D., et al. (2008). Systematic Review of Accuracy of Screening Instruments for Predicting Fall Risk Among Independently Living Older Adults. *Journal of Rehabilitation Research and Development*, 45(8), pp. 1105–1116.
- Gill, T. M., Murphy, T. E., Gahbauer, E. A., et al. (2013). "Association of Injurious Falls with Disability Outcomes and Nursing Home Admissions in Community Living Older Persons." *American Journal of Epidemiology*, 178(3), pp. 418–425.
- Gorry, G. A., & Scott-Morton, M. S. (1971). "A Framework for Management Information Systems." *Sloan Management Review*, 13(1), pp. 55–70.
- Havens, E., Peōa, J., Slabaugh, S., Cordier, T., Renda, A., & Gopal, V. (2015, October). Exploring the Relationship Between Health-Related Quality of Life and Health Conditions, Costs, Resource Utilization, and Quality Measures. Podium presentation at the ISOQOL Twenty-Seventh Annual Conference, Vancouver, Canada.
- Havens, E., Slabaugh, L., Peōa, J., Haugh, G., & Gopal, V. (2015, February). Are There Differences in Healthy Days Based on Compliance to Preventive Health Screening Measures? Poster presentation at Preventive Medicine 2015, Atlanta, GA.
- Healthcare IT News. (2017, November 9). "How AI Is Transforming Healthcare and Solving Problems in 2017." Slideshow.healthcareitnews.com/slideshow/howai-transforming-healthcare-and-solving-problems-2017?page=4/ (accessed September 2018).

Hesse, R., & G. Woolsey. (1975). *Applied Management Science: A Quick and Dirty Approach*. Chicago, IL: SRA Inc. Humana. 2016 Progress Report. populationhealth.humana.com/wp-content/uploads/2016/05/BoldGoal2016ProgressReport_1.pdf (accessed September 2018).

INFORMS. Analytics Section Overview. informs.org/Community/Analytics (accessed September 2018).

Jacquet, F. (2017, July 4). "Exploring the Artificial Intelligence Ecosystem: AI, Machine Learning, and Deep Learning." DZone.

Jarrett, C. (2018, January 21). "Amazon Set to Open Doors on AI-Powered Grocery Store." [Venturebeat.com. venturebeat.com/2018/01/21/amazon-set-to-open-doors-on-ai-powered-grocery-store/](http://venturebeat.com/2018/01/21/amazon-set-to-open-doors-on-ai-powered-grocery-store/) (accessed September 2018).

Keen, P. G. W., & M. S. Scott-Morton. (1978). *Decision Support Systems: An Organizational Perspective*. Reading, MA: Addison-Wesley.

Kemper, C., and C. Breuer. (2016). "How Efficient Is Dynamic Pricing for Sports Events? Designing a Dynamic Pricing Model for Bayern Munich." *International Journal of Sports Finance*, 11, pp. 4–25.

Kranz, M. (2017, December 27). "In 2018, Get Ready for the Convergence of IoT, AI, Fog, and Blockchain." *Insights*.

Liberto, D. (2017, June 29). "Artificial Intelligence Will Add Trillion to the Global Economy: PwC." *Investopedia*.

Lollato, R., Patrignani, A., Ochsner, T. E., Rocatelli, A., Tomlinson, P. & Edwards, J. T. (2015). "Improving Grazing Management Using a Smartphone App." www.bookstore.ksre.ksu.edu/pubs/MF3304.pdf (accessed October 2018).

Lopez, J. (2017, August 11). "Smart Farm Equipment Helps Feed the World." IQintel.com.

Metz, C. (2018, February 12). "As China Marches Forward on A.I., the White House Is Silent." *The New York Times*.

Nadav, S. (2017, August 9). "Business Intelligence Is Failing; Here Is What Is Coming Next." *Huffington Post*.

Norman, A. (2018, January 31). "Your Future Doctor May Not Be Human. This Is the Rise of AI in Medicine." *Futurism.com*.

Olshansky, C. (2017, August 24). "Coca-Cola Is Bringing Artificial Intelligence to Vending Machines." *Food & Wine*.

Opfer, C. (2016, June 22). "There's One Terrific Reason to Race Camels Using Robot Jockeys." *Howstuffworks.com*.

Padmanabhan, G. (2018, January 4). "Industry-Specific Augmented Intelligence: A Catalyst for AI in the Enterprise." *Forbes*.

Pajouh Foad, M., Xing, D., Hariharan, S., Zhou, Y., Balasundaram, B., Liu, T., & Sharda, R. (2013). Available-to-Promise in Practice: An Application of Analytics in the Specialty Steel Bar Products Industry. *Interfaces*, 43(6), pp. 503–517. [dx.doi.org/10.1287/inte.2013.0693](https://doi.org/10.1287/inte.2013.0693) (accessed September 2018).

Patrignani, A., & Ochsner, T. E., (2015). Canopeo: A Powerful New Tool for Measuring Fractional Green Canopy Cover. *Agronomy Journal*, 107(6), pp. 2312–2320;

PE Report. (2017, July 29). "Satellite-Based Advance Can Help Raise Farm Output by 20 Percent Experts." *Financial Express*.

PricewaterhouseCoopers Report. (2011, December). "Changing the Game: Outlook for the Global Sports Market to 2015." pwc.com/gx/en/hospitality-leisure/pdf/changing-the-game-outlook-for-the-global-sportsmarket-to-2015.pdf (accessed September 2018).

Quain, S. (2018, June 29). "The Decision-Making Process in an Organization." *Small Business Chron*.

Reisinger, D. (2018, February 22). "Here Are the Next Cities to Get Amazon Go Cashier-Less Stores." *Fortune*.

Sharda, R., Asamoah, D., & Ponna, N. (2013). "Research and Pedagogy in Business Analytics: Opportunities and Illustrative

Examples." *Journal of Computing and Information Technology*, 21(3), pp. 171–182.

Sharda, R., Delen, D., & Turban, E. (2016). *Business Intelligence, Analytics, and Data Science: A Managerial Perspective on Analytics*. 4th ed. NJ: Pearson.

Sharda, R., & P. Kalgotra. (2018). "The Blossoming Analytics Talent Pool: An Overview of the Analytics Ecosystem." In James J. Cochran (ed.). *INFORMS Analytics Body of Knowledge*. John Wiley, Hoboken, NJ

Silk, R. (2017, November). "Biometrics: Facial Recognition Tech Coming to an Airport Near You." *Travel Weekly*, 21.

Simon, H. (1977). *The New Science of Management Decision*. Englewood Cliffs, NJ: Prentice Hall.

Slade, L. (2017, December 21). "Meet the Jordanian Camel Races Using Robot Jockeys." **Sbs.com.au**.

Slowey, L. (2017, February 16). "Look Who's Talking: KONE Makes Elevator Services Truly Intelligent with Watson IoT." IBM Internet of Things Blog. **ibm.com/blogs/internetof-things/kone/** (accessed September 2018).

"Sports Analytics Market Worth by 2021." (2015, June 25). Wintergreen Research Press Release. Covered by PR Newswire at **http://www.prnewswire.com/newsreleases/sports-analytics-market-worth-47-billionby-2021-509869871.html**.

Srikanthan, H. . (2018, January 8). "KONE Improves 'People Flow' in 1.1 Million Elevators with IBM Watson IoT." Generis. **https://generisgp.com/2018/01/08/ibm-casestudy-kone-corp/** (accessed September 2018).

Staff. "Assisted, Augmented and Autonomous: The 3Flavours of AI Decisions." (2017, June 28). *Software and Technology*.

Tableau.com. Silvaris Augments Proprietary Technology Platform with Tableau's Real-Time Reporting Capabilities. **tableau.com/sites/default/files/case-studies/silvarisbusiness-dashboards_0.pdf** (accessed September 2018).

Tartar, Andre, et al. (2018, 26 July). "All the Things Satellites Can Now See from Space." **Bloomberg.com**. **www.bloomberg.com/news/features/2018-07-26/all-thethings-satellites-can-now-see-from-space** (accessed October 2018).

TeradataUniversityNetwork.com. (2015, Fall). "BSI: Sports Analytics—Precision Football" (video). **teradatauniversitynetwork.com/About-Us/Whats-New/BSI-Sports-Analytics-Precision-Football/** (accessed September 2018).

Thibodeaux, W. (2017, June 29). "This Artificial Intelligence Kiosk Is Designed to Spot Liars at Airports." Inc.com.

Turck, Matt. "Is Big Data Still a Thing? (The 2016 Big Data Landscape)." **http://mattturck.com/2016/02/01/bigdata-landscape/** (accessed September 2018).

Watson, H. (2005, Winter). Sorting Out What's New in Decision Support. *Business Intelligence Journal*.

Weldon, D. (2018, March 6). "Nearly Half of CIOs Now Plan to Deploy Artificial Intelligence." Information Management.

Wikipedia.org. On-base Percentage. **wikipedia.org/wiki/On_base_percentage**. (accessed September 2018).

Wikipedia.org. Sabermetrics. **wikipedia.org/wiki/Sabermetrics** (accessed September 2018).

Wikipedia.org. SIEMENS. **wikipedia.org/wiki/Siemens** (accessed September 2018).

YouTube.com. (2013, December 17). CenterPoint Energy Talks Real Time Big Data Analytics. **youtube.com/watch?v=s7CzeSIIIEfl** (accessed September 2018).

Yue, P. (2017, August 24). "Baidu, Beijing Airport Launch Facial Recognition for Passenger Check-In." China Money Network. **https://www.chinamoneynetwork.com/2017/08/24/baidu-capital-airport-launch-facialrecognition-system-airport** (accessed October 2018).

Zane, E. B. (2016). *Effective Decision-Making: How to Make Better Decisions Under Uncertainty And Pressure*. Kindle ed. Seattle, WA: Amazon Digital Services.

فصل ٢

Agrawal, V. "How Successful Investors Are Using AI to Stay Ahead of the Competition." *ValueWalk*, January 28, 2018.

Alpaydin, E. *Machine Learning: The New AI (The MIT Press Essential Knowledge Series)*. Boston, MA: MIT Press, 2016.

Beauchamp, P. "Artificial Intelligence and the Insurance Industry: What You Need to Know." *The Huffington Post*, October 27, 2016.

Blog. "Welcome to the Future: How AI Is Transforming Insurance."

Blog.metlife.com, October 1, 2017.

Bollard, A., et al. "The next-generation operating model for the digital world." *McKinsey & Company*, March 2017.

BrandStudio. "Future-Proof: How Today's Artificial Intelligence Solutions Are Taking Government Services to the Next Frontier." *Washington Post*, August 22, 2017.

Carey, S. "US Bank Doubles Its Conversion Rate for Wealth Customers Using Salesforce Einstein." *Computerworld UK*, November 10, 2017.

Carney, P. "Pat Carney: Artificial Intelligence versus Human Intelligence." *Vancouver Sun*, April 7, 2018.

Celentano, D. "Kraft Foods iPhone Assistant Appeals to Time Starved Consumers." *The Balance*, September 18, 2016.

Chandi, N. "How AI is Reshaping the Accounting Industry." **Forbes.com**, July 20, 2017.

Clozel, L. "IBM Unveils New Watson tools to Help Banks Manage Compliance, AML." *American Banker*, June 14, 2017.

Consultancy.uk. "How Artificial Intelligence Is Transforming the Banking Industry." September 28, 2017.
consultancy.uk/news/14017/how-artificial-intelligence-istransforming-the-banking-industry/ (accessed June 2018).

Coward, J. "Artificial Intelligence Is Unshackling DHL's Supply Chain Potential." *IoT Institute*, April 18, 2017. **ioti.com/industrial-iot/artificial-intelligence-unshacklingdhls-supply-chain-potential** (accessed June 2018).

Crosman, P. "U.S. Bank Bets AI Can Finally Deliver 360-Degree View." *American Banker*, July 20, 2017.

Davis, B. "15 Examples of Artificial Intelligence in Marketing." *Econsultancy*, April 19, 2016.

Dickson, B. "How Artificial Intelligence Optimizes Recruitment." *The Next Web*, June 3, 2017.

Dodge, J. "Artificial Intelligence in the Enterprise: It's On." *Computerworld*, February 10, 2016.

Dormehl, L. *Thinking Machines: The Quest for Artificial Intelligence—and Where It's Taking Us Next*. New York, NY: Tarcher-Perigee, 2017.

Essex, D. "AI in HR: Artificial Intelligence to Bring Out the Best in People." *TechTargetEssential Guide*, April 2017.

E. V. Staff. "Artificial Intelligence Used to Predict Short-Term Share Price Movements." *The Economic Voice*, June 22, 2017.

Finlay, S. *Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies*. 2nd ed. Seattle, WA: Relativistic, 2017.

Forrest, C. "7 Companies That Used Machine Learning to Solve Real Business Problems." *Tech Republic*, March 8, 2017.

Fuller, D. "LG Claims Its Roboking Vacuum Is As Smart As a Child." *Androidheadlines.com*, July 18, 2017.

Gagliardi, N. "Softbank Leads \$120M Investment in AI-Based Insurance Startup Lemonade." *ZDNET*, December 19, 2017.

Gangwani, T. "3 Ways to Improve Customer Experience Using A.I." *CIO Contributor Network*, October 12, 2016.

Gaudin, S. "White House: A.I. Will Be Critical Driver of U.S. Economy." *Computerworld*, October 12, 2016.

Gitlin, J. M. "Watch Out, Waze: INRIX's New Traffic App Is Coming for You." *Ars Technica*, March 30, 2016. arstechnica.com/cars/2016/watch-out-waze-inrixs-new-trafficapp-is-coming-for-you/ (accessed June 2018).

Greengard, S. "Delving into Gartner's 2016 Hype Cycle." *Baseline*, September 7, 2016.

Greig, J. "Gartner: AI Business Value Up 70% in 2018, and These Industries Will Benefit the Most." *Tech Republic*, April 25, 2018.

Haines, D. "Is Artificial Intelligence Making It Easier and Quicker to Get a New Job?" *Huffington Post UK*, December 4, 2017.

Hauari, G. "InsurersLeverage AI to Unlock Legacy Claims Data." *Information Management*, July 3, 2017.

Huang, G. "Why AI Doesn't Mean Taking the 'Human' Out of Human Resources." *Forbes.com*, September 27, 2017.

Hughes, T. "Google DeepMind's Program Beat Human at Go." *USA Today*, January 27, 2016.

ICAEW. "Artificial Intelligence and the Future of Accountancy." artificial-intelligence-report.ashx/, 2017.

Kaplan, J. *Artificial Intelligence: What Everyone Needs to Know*. London, UK: Oxford University Press, 2016.

Kharpal, A. "A.I. Is in a 'Golden Age' and Solving Problems That Were Once in the Realm of Sci-Fi, Jeff Bezos Says." *CNBC News*, May 8, 2017.

Kiron, D. "What Managers Need to Know About Artificial Intelligence?" *MIT Sloan Management Review*, January 25, 2017.

Knight, W. "Walmart's Robotic Shopping Carts Are the Latest Sign That Automation Is Eating Commerce." *Technology Review*, June 15, 2016.

Kolbjørnsrud, V., R. Amico, and R. J. Thomas. "How Artificial Intelligence Will Redefine Management." *Harvard Business Review*, November 2, 2016.

Korosec, K. "Inrix Updates Traffic App to Learn Your Daily Habits." *Fortune Tech*, March 30, 2016.

Liao, P-H., et al. "Applying Artificial Intelligence Technology to Support Decision-Making in Nursing: A Case Study in Taiwan." *Health Informatics Journal*, June 2015.

Marr, B., "The Key Definitions of Artificial Intelligence That Explain Its Importance." *Forbes*, February 14, 2018.

Marr, B. "What Everyone Should Know About Cognitive Computing." *Forbes.com*, March 23, 2016.

Martin, J. "10 Things Marketers Need to Know about AI." *CIO.com*, February 13, 2017.

McPherson, S.S. *Artificial Intelligence: Building Smarter Machines*. Breckenridge, CO: Twenty-First Century Books, 2017.

Meister, J. "The Future of Work: How Artificial Intelligence Will Transform the Employee Experience." *Forbes.com*, November 9, 2017.

Metz, C. "Facebook's Augmented Reality Engine Brings AI Right to Your Phone." *Wired*, April 19, 2017.

Mittal, V. "Top 15 Deep Learning Applications That Will Rule the World in 2018 and Beyond." **Medium.com**, October 3, 2017.

Narayan, K. "Leverage Artificial Intelligence to Build your Sales Pipeline." *LinkedIn*, February 14, 2018.

Ng, A. "What Artificial Intelligence Can and Can't Do Right Now." *Harvard Business Review*, November 9, 2016.

Nordrum, A. "Hedge Funds Look to Machine Learning, Crowdsourcing for Competitive Advantage." *IEEE Spectrum*, June 28, 2017.

Ovaska-Few, S. "How Artificial Intelligence Is Changing Accounting." *Journal of Accountancy*, October 9, 2017.

Padmanabhan, G. "Industry-Specific Augmented Intelligence: A Catalysts for AI in the Enterprise." *Forbes*, January 4, 2018.

Pennington, R. "Artificial Intelligence: The New Tool for Accomplishing an Old Goal in Marketing." *Huffington Post*, January 16, 2018.

Press, G. "Top 10 Hot Artificial Intelligence (AI) Technologies." *Forbes*, January 23, 2017.

Pyle, D., and C. San José. "An Executive's Guide to Machine Learning." McKinsey & Company, June 2015.

Reinharz, S. *An Introduction to Artificial Intelligence: Professional Edition: An Introductory Guide to the Evolution of Artificial Intelligence*. Kindle Edition. Seattle, WA: Simultaneous Device Usage (Amazon Digital Service), 2017.

Sample, I. "AI Watchdog Needed to Regulate Automated Decision-Making, Say Experts." *The Guardian*, January 27, 2017.

Santana, D. "Metromile Launches AI Claims Platform." *Digital Insurance*, July 25, 2017.

Savar, A. "3 Ways That A.I. Is Transforming HR and Recruiting." **INC.com**, June 26, 2017.

Schrage, M. "4 Models for Using AI to Make Decisions." *Harvard Business Review*, January 27, 2017.

Shah, J. "Robots Are Learning Complex Tasks Just by Watching Humans Do Them." *Harvard Business Review*, June 21, 2016.

Sharma, G. "China Unveils Multi-Billion Dollar Artificial Intelligence Plan." *International Business Times*, July 20, 2017. ibtimes.co.uk/china-unveils-multi-billion-dollar-artificial-intelligence-plan-1631171/ (accessed January 2018).

Sincavage, D. "How Artificial Intelligence Will Change Decision-Making for Businesses." *Business 2 Community*, August 24, 2017.

Singh, H. "How Artificial Intelligence Will Transform Financial Services." *Information Management*, June 6, 2017.

SMBWorld Asia Editors. "Hays: Artificial Intelligence Set to Revolutionize Recruitment." *Enterprise Innovation*, August 30, 2017.

Smith, J. *Machine Learning: Machine Learning for Beginners. Can Machines Really Learn Like Humans? All About Artificial Intelligence (AI), Deep Learning and Digital Neural Networks*. Kindle Edition. Seattle, WA: Amazon Digital Service, 2017.

Staff. "Assisted, Augmented and Autonomous: The 3 Flavours of AI Decisions." *Software and Technology*, June 28, 2017. tgdaily.com/technology/assisted-augmented-and-autonomous-the-3-flavours-of-ai-decisions

Steffi, S. "List of 50 Unique AI Technologies." *Hacker Noon.com*, October 18, 2017.

Taylor, P. "Welcome to the Machine – Learning." *Forbes Brand-Voice*, June 3, 2016. forbes.com/sites/sap/2016/06/03/welcome-to-the-machine-learning/#3175d50940fe (accessed June 2017).

Theobald, O. *Machine Learning for Absolute Beginners: A Plain English Introduction*. Kindle Edition. Seattle, WA, 2017.

Tiwan, R. "Artificial Intelligence (AI) in Banking Case Study Report 2017." *iCrowd Newswire*, July 7, 2017.

USC. "AI Computer Vision Breakthrough IDs Poachers in Less Than Half a Second." *Press Release*, February 8, 2018.

- Violino, B. "Most Firms Expect Rapid Returns on Artificial Intelligence Investments." *Information Management*, November 1, 2017.
- Warawa, J. "Here's Why Accountants (Yes, YOU!) Should Be Driving AI Innovation." *CPA Practice Advisor*, November 1, 2017.
- Waxer, C. "Get Ready for the BOT Revolution." *Computerworld*, October 17, 2016.
- Wellers, D., et al. "8 Ways Machine Learning Is Improving Companies' Work Processes." *Harvard Business Review*, May 31, 2017.
- Wislow, E. "5 Ways to Use Artificial Intelligence (AI) in Human Resources." *Big Data Made Simple*, October 24, 2017.
bigdata-madesimple.com/5-ways-to-use-artificialintelligence-ai-in-human-resources/.
- Yurcan, B. "TD's Innovation Agenda: Experiments with Alexa, AI and Augmented Reality." *Information Management*, December 27, 2017.
- Zarkadakis, G. *In Our Own Image: Savior or Destroyer? The History and Future of Artificial Intelligence*. New York, NY: Pegasus Books, 2016.
- Zhou, A. "EY, Deloitte and PwC Embrace Artificial Intelligence for Tax and Accounting." *Forbes.com*, November 14, 2017.

فصل ٣

- Abela, A. (2008). *Advanced Presentations by Design: Creating Communication That Drives Action*. New York, NY: Wiley.
- Annas, G. (2003). "HIPAA Regulations—A New Era of Medical-Record Privacy?" *New England Journal of Medicine*, 348(15), 1486–1490.
- Ante, S., & J. McGregor. (2006). "Giving the Boss the Big Picture: A Dashboard Pulls Up Everything the CEO Needs to Run the Show." *Business Week*, 43–51.
- Carotenuto, D. (2007). "Business Intelligence Best Practices for Dashboard Design." WebFOCUS.
www.datawarehouse.inf.br/papers/information_builders_dashboard_best_practices.pdf (accessed August 2016).
- Dell Customer Case Study. "Medical Device Company Ensures Product Quality While Saving Hundreds of Thousands of Dollars." <https://software.dell.com/documents/instrumentation-laboratory-medicaldevice-companyensures-product-quality-whilesaving-hundreds-ofthousands-of-dollars-casestudy-80048.pdf> (accessed August 2016).
- Delen, D. (2010). "A Comparative Analysis of Machine Learning Techniques for Student Retention Management." *Decision Support Systems*, 49(4), 498–506.
- Delen, D. (2011). "Predicting Student Attrition with Data Mining Methods." *Journal of College Student Retention* 13(1), 17–35.
- Delen, D. (2015). *Real-World Data Mining: Applied Business Analytics and Decision Making*. Upper Saddle River, NJ: Financial Times Press (A Pearson Company).
- Delen, D., D. Cogdell, & N. Kasap. (2012). "A Comparative Analysis of Data Mining Methods in Predicting NCAA Bowl Outcomes." *International Journal of Forecasting*, 28, 543–552.
- Eckerson, W. (2006). *Performance Dashboards*. New York: Wiley.
- Few, S. (2005, Winter). "Dashboard Design: Beyond Meters, Gauges, and Traffic Lights." *Business Intelligence Journal*, 10(1).
- Few, S. (2007). "Data Visualization: Past, Present and Future." Perceptualedge.com/articles/Whitepapers/Data_Visualization.pdf (accessed July 2016).
- Fink, E., & S. J. Moore. (2012). "Five Best Practices for Telling Great Stories with Data." Tableau Software, Inc.
www.tableau.com/whitepapers/telling-data-stories (accessed May 2016).

Freeman, K., & R. M. Brewer. (2016). "The Politics of American College Football." *Journal of Applied Business and Economics*, 18(2), 97–101.

Gartner Magic Quadrant. (2016, February 4). gartner.com (accessed August 2016).

Grimes, S. (2009a, May 2). "Seeing Connections: Visualizations Makes Sense of Data. *Intelligent Enterprise*." i.cmpnet.com/intelligententerprise/next-era-businessintelligence/Intelligent_Enterprise_Next_Era_BI_Visualization.pdf (accessed January 2010).

Grimes, S. (2009b). Text "Analytics 2009: User Perspectives on Solutions and Providers." Alta Plana. altaplana.com/TextAnalyticsPerspectives2009.pdf (accessed July, 2016).

Hardin, M. Hom, R. Perez, & Williams L. (2012). "Which Chart or Graph Is Right for You?" Tableau Software. http://www.tableau.com/sites/default/files/media/which_chart_v6_final_0.pdf (accessed August 2016).

Hernández, M., & S. J. Stolfo. (1998, January). "Real-World Data Is Dirty: Data Cleansing and the Merge/Purge Problem." *Data Mining and Knowledge Discovery*, 2(1), 9–37.

Hill, G. (2016). "A Guide to Enterprise Reporting." Ghill. customer.netspace.net.au/reporting/definition.html (accessed July 2016).

Kim, W., B. J. Choi, E. K. Hong, S. K. Kim, & D. Lee. (2003). "A Taxonomy of Dirty Data." *Data Mining and Knowledge Discovery*, 7(1), 81–99.

Kock, N. F., R. J. McQueen, & J. L. Corner. (1997). "The Nature of Data, Information and Knowledge Exchanges in Business Processes: Implications for Process Improvement and Organizational Learning." *The Learning Organization*, 4(2), 70–80.

Kotsiantis, S., D. Kanellopoulos, & P. E. Pintelas. (2006). "Data Preprocessing for Supervised Learning." *International Journal of Computer Science*, 1(2), 111–117.

Lai, E. (2009, October 8). "BI Visualization Tool Helps Dallas Cowboys Sell More Tony Romo Jerseys." *ComputerWorld*.

Quinn, C. (2016). "Data-Driven Marketing at SiriusXM," Teradata Articles & News. <http://bigdata.teradata.com/US/Articles-News/Data-Driven-Marketing-At-SiriusXM/> (accessed August 2016); "SiriusXM Attracts and Engages a New Generation of Radio Consumers." <http://assets.teradata.com/resourceCenter/downloads/CaseStudies/EB8597.pdf?processed=1> (accessed August 2018).

Novell. (2009, April). "Executive Dashboards Elements of Success." Novell white paper. www.novell.com/docrep/documents/3rkw3etfc3/Executive%20Dashboards_Elements_of_Success_White_Paper_en.pdf (accessed June 2016).

Radha, R. (2008). "Eight Best Practices in Dashboard Design." *Information Management*. www.informationmanagement.com/news/columns/-10001129-1.html (accessed July 2016).

SAS. (2014). "Data Visualization Techniques: From Basics to Big Data." http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/data-visualizationtechniques-106006.pdf (accessed July 2016).

Thammasiri, D., D. Delen, P. Meesad, & N. Kasap. (2014). "A Critical Assessment of Imbalanced Class Distribution Problem: The Case of Predicting Freshmen Student Attrition." *Expert Systems with Applications*, 41(2), 321–330.

فصل ٤

Chan, P., Phan, W., Prodromidis, A., & Stolfo, S. (1999). "Distributed Data Mining in Credit Card Fraud Detection." *IEEE Intelligent Systems*, 14(6), 67–74.

- CRISP-DM. (2013). "Cross-Industry Standard Process for Data Mining (CRISP-DM)." <http://crisp-dm.orgwww.the-odelingagency.com/crisp-dm.pdf> (accessed February 2, 2013).
- Davenport, T. (2006, January). "Competing on Analytics." *Harvard Business Review*, 99–107.
- Delen, D. (2009). "Analysis of Cancer Data: A Data Mining Approach." *Expert Systems*, 26(1), 100–112.
- Delen, D. (2014). *Real-World Data Mining: Applied Business Analytics and Decision Making*. Upper Saddle River, NJ: Pearson.
- Delen, D., Cogdell, D., & Kasap, N. (2012). "A Comparative Analysis of Data Mining Methods in Predicting NCAA Bowl Outcomes." *International Journal of Forecasting*, 28, 543–552.
- Delen, D., Sharda, R., & Kumar, P. (2007). "Movie Forecast Guru: A Web-Based DSS for Hollywood Managers." *Decision Support Systems*, 43(4), 1151–1170.
- Delen, D., Walker, G., & Kadam, A. (2005). "Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods." *Artificial Intelligence in Medicine*, 34(2), 113–127.
- Dunham, M. (2003). *Data Mining: Introductory and Advanced Topics*. Upper Saddle River, NJ: Prentice Hall.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). "From Knowledge Discovery in Databases." *AI Magazine*, 17(3), 37–54.
- Hoffman, T. (1998, December 7). "Banks Turn to IT to Reclaim Most Profitable Customers." *Computerworld*.
- Hoffman, T. (1999, April 19). "Insurers Mine for Age-Appropriate Offering." *Computerworld*.
- Kohonen, T. (1982). "Self-Organized Formation of Topologically Correct Feature Maps." *Biological Cybernetics*, 43(1), 59–69.
- Nemati, H., & Barko, C. (2001). "Issues in Organizational Data Mining: A Survey of Current Practices." *Journal of Data Warehousing*, 6(1), 25–36.
- Nisbet, R., Miner, G., & Elder IV, J. (2009). "Top 10 Data Mining Mistakes." *Handbook of Statistical Analysis and Data Mining Applications*. Academic Press, pp. 733–754.
- Quinlan, J. (1986). "Induction of Decision Trees." *Machine Learning*, 1, 81–106.
- Saltelli, A. (2002). "Making Best Use of Model Evaluations to Compute Sensitivity Indices." *Computer Physics Communications*, 145, 280–297.
- Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity Analysis in Practice – A Guide to Assessing Scientific Models*. Hoboken, NJ: John Wiley.
- SEMMA. (2009). "SAS's Data Mining Process: Sample, Explore, Modify, Model, Assess." sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html (accessed August 2009).
- Sharda, R., & Delen, D. (2006). "Predicting Box-Office Success of Motion Pictures with Neural Networks." *Expert Systems with Applications*, 30, 243–254.
- Shultz, R. (2004, December 7). "Live from NCDM: Tales of Database Buffoonery." directmag.com/news/ncdm-12-07-04/index.html (accessed April 2009).
- Skalak, D. (2001). "Data Mining Blunders Exposed!" *DB2 Magazine*, 6(2), 10–13.
- Thongkam, J., Xu, G., Zhang, Y., & Huang, F. (2009). "Toward Breast Cancer Survivability Prediction Models Through Improving Training Space." *Expert Systems with Applications*, 36(10), 12200–12209.
- Wald, M. (2004, February 21). "U.S. Calls Release of JetBlue Data Improper." *The New York Times*.

Wright, C. (2012). "Statistical Predictors of March Madness: An Examination of the NCAA Men's Basketball Championship." <http://economics-files.pomona.edu/GarySmith/Econ190/Wright%20March%20Madness%20Final%20Paper.pdf> (accessed February 2, 2013).

Zaima, A. (2003). "The Five Myths of Data Mining." *What Works: Best Practices in Business Intelligence and Data Warehousing*, Vol. 15. Chatsworth, CA: The Data Warehousing Institute, pp. 42–43.

Zolbanin, H., Delen, D., & Zadeh, A. (2015). "Predicting Overall Survivability in Comorbidity of Cancers: A Data Mining Approach." *Decision Support Systems*, 74, 150–161.

فصل ٥

Abbott, D. (2014). *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst*. Hoboken, NJ: John Wiley.

Aizerman, M., E. Braverman, & L. Rozonoer. (1964). "Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning." *Automation and Remote Control*, Vol. 25, pp. 821–837.

American Heart Association, "Heart Disease and Stroke Statistics," heart.org (accessed May 2018).

Boiman, E. S., & M. Irani. (2008). "In Defense of Nearest-Neighbor Based Image Classification," *IEEE Conference on Computer Vision and Pattern Recognition, 2008 (CVPR)*, 2008, pp. 1–8.

Bouzembrak, Y., & H. J. Marvin. (2016). "Prediction of Food Fraud Type Using Data from Rapid Alert System for Food and Feed (RASFF) and Bayesian Network Modelling." *Food Control*, 61, 180–187.

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24(2), 123–140.

Breiman, L. (2001). "Random Forests." *Machine Learning*, 45 (1), 5–32.

Chow, C., & C. Liu (1968). "Approximating Discrete Probability Distributions with Dependence Trees." *IEEE Transactions on Information Theory*, 14(3), 462–473.

Delen, D., & R. Sharda. (2010). "Predicting the Financial Success of Hollywood Movies Using an Information Fusion Approach." *Indus Eng J*, 21 (1), 30–37.

Delen, D., L. Tomak, K. Topuz, & E. Eryarsoy (2017). Investigating Injury Severity Risk Factors in Automobile Crashes with Predictive Analytics and Sensitivity Analysis Methods. *Journal of Transport & Health*, 4, 118–131.

Delen, D., A. Oztekin, & L. Tomak. (2012). "An Analytic Approach to Better Understanding and Management of Coronary Surgeries." *Decision Support Systems*, 52 (3), 698–705.

Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 1189–1232.

Freund, Y., & R. E. Schapire. (1996, July). "Experiments with a New Boosting Algorithm." In *Icml* (Vol. 96, pp. 148–156).

Friedman, D. (2014). "Oral Testimony before the House Committee on Energy and Commerce, by the Subcommittee on Oversight and Investigations," April 1, 2014, www.nhtsa.gov/Testimony (accessed October 2014).

Friedman, N., D. Geiger, & M. Goldszmidt. (1997). "Bayesian Network Classifiers." *Machine Learning*, Vol. 29, No. 2–3, 131–163.

Haykin, S. (2009). *Neural Networks and Learning Machines*, 3rd ed. Upper Saddle River, NJ: Prentice Hall.

- Hopfield, J. (1982, April). "Neural Networks and Physical Systems with Emergent Collective Computational Abilities." *Proceedings of National Academy of Science*, Vol. 79, No. 8, 2554–2558.
- Koller, D., & N. Friedman. (2009). *Probabilistic Graphical Models: Principles and Techniques*. Boston, MA: MIT Press.
- McCulloch, W., & W. Pitts. (1943). "A Logical Calculus of the Ideas Imminent in Nervous Activity." *Bulletin of Mathematical Biophysics*, Vol. 5.
- Medsker, L., & J. Liebowitz. (1994). *Design and Development of Expert Systems and Neural Networks*. New York, NY: Macmillan, p. 163.
- Meyfroidt, G., F. Güiza, J. Ramon, & M. Bruynooghe. (2009). "Machine Learning Techniques to Examine Large Patient Databases." *Best Practice & Research Clinical Anaesthesiology*, 23(1), 127–143.
- Minsky, M., & S. Papert. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Neural Technologies. "Combating Fraud: How a Leading Telecom Company Solved a Growing Problem." neuralt.com/iqs/dlsfa.list/dlcpti.7/downloads.html (accessed March 2009).
- NHTSA (2018) National Highway Traffic Safety Administration (NHTSA's) General Estimate System (GES), www.nhtsa.gov (accessed January 20, 2017).
- Pearl, J. (1985). "Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning." *Proceedings of the Seventh Conference of the Cognitive Science Society*, 1985, pp. 329–334.
- Pearl, J. (2009). *Causality*. Cambridge University Press, Cambridge, England.
- Principe, J., N. Euliano, & W. Lefebvre. (2000). *Neural and Adaptive Systems: Fundamentals Through Simulations*. New York, NY: Wiley.
- Reddy, B. K., D. Delen, & R. K. Agrawal. (2018). "Predicting and Explaining Inflammation in Crohn's Disease Patients Using Predictive Analytics Methods and Electronic Medical Record Data." *Health Informatics Journal*, 1460458217751015.
- Reagan, R. (1982). "Remarks on Signing Executive Order 12368, Concerning Federal Drug Abuse Policy Functions," June 24, 1982. Online by Gerhard Peters and John T. Woolley, The American Presidency Project. <http://www.presidency.ucsb.edu/ws/?pid=42671>.
- Surowiecki, J. (2006). *The Wisdom of Crowds*. New York, NY: Penguin Random House.
- Topuz, K., F. Zengul, A. Dag, A. Almehti, & M. Yildirim (2018). "Predicting Graft Survival Among Kidney Transplant Recipients: A Bayesian Decision Support Model." *Decision Support Systems*, 106, 97–109.
- Vorhies, W. (2016). "Want to Win Competitions? Pay Attention to Your Ensembles." Data Science Central Web Portal, www.datasciencecentral.com/profiles/blogs/wantto-win-at-kaggle-pay-attention-to-your-ensembles (accessed July 2018).
- Wang, G., T. Xu, T. Tang, T. Yuan, & H. Wang. (2017). A "Bayesian Network Model for Prediction of Weather-Related Failures in Railway Turnout Systems." *Expert Systems with Applications*, 69, 247–256.
- Wolpert, D. (1992). "Stacked Generalization." *Neural Networks*, 5(2), 241–260.
- Zahedi, F. (1993). *Intelligent Systems for Business: Expert Systems with Neural Networks*, Wadsworth, Belmont, CA.
- Zhang, H., A. C. Berg, M. Maire, & J. Malik. (2006). "SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition." In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* (Vol. 2, pp. 2126–2136). IEEE.

- Abad, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, . . . M. Isard. (2016). "TensorFlow: A System for Large-Scale Machine Learning." *OSDI, 16*, pp. 265–283.
- Altman, E. I. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *The Journal of Finance*, 23(4), pp. 589–609.
- Bahdanau, D., K. Cho, & Y. Bengio. (2014). "Neural Machine Translation by Jointly Learning to Align and Translate." ArXiv Preprint ArXiv:1409.0473.
- Bengio, Y. (2009). "Learning Deep Architectures for AI." *Foundations and Trends® in Machine Learning*, 2(1), pp. 1–127.
- Bergstra, J., O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, . . . Y. Bengio. (2010). "Theano: A CPU and GPU Math Compiler in Python." *Proceedings of the Ninth Python in Science Conference*, Vol. 1.
- Bi, R. (2014). "When Watson Meets Machine Learning." www.kdnuggets.com/2014/07/watson-meets-machinelearning.html (accessed June 2018).
- Boureau, Y.-L., N. Le Roux, F. Bach, J. Ponce, & Y. LeCun (2011). "Ask the Locals: Multi-Way Local Pooling for Image Recognition." *Proceedings of the International Computer Vision (ICCV'11) IEEE International Conference*, pp. 2651–2658.
- Boureau, Y.-L., J. Ponce, & Y. LeCun. (2010). "A Theoretical Analysis of Feature Pooling in Visual Recognition." *Proceedings of International Conference on Machine Learning (ICML'10)*, pp. 111–118.
- Buchanan, B. G., & E. H. Shortliffe. (1984). *Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley.
- Cognitive Computing Consortium. (2018). <https://cognitivecomputingconsortium.com/resources/cognitivecomputing-defined/#1467829079735-c0934399-599a> (accessed July 2018).
- Chen, T., M. Li, Y. Li, M. Lin, N. Wang, M. Wang, . . . Z. Zhang. (2015). "Mxnet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems." ArXiv Preprint ArXiv:1512.01274.
- Collobert, R., K. Kavukcuoglu, & C. Farabet. (2011). "Torch7: A Matlab-like Environment for Machine Learning." Big-Learn, NIPS workshop.
- Cybenko, G. (1989). "Approximation by Superpositions of a Sigmoidal Function." *Mathematics of Control, Signals and Systems*, 2(4), 303–314.
- DeepQA. (2011). "DeepQA Project: FAQ, IBM Corporation." https://researcher.watson.ibm.com/researcher/view_group.php?id=2099 (accessed May 2018).
- Delen, D., R. Sharda, & M. Bessonov, M. (2006). "Identifying Significant Predictors of Injury Severity in Traffic Accidents Using a Series of Artificial Neural Networks." *Accident Analysis & Prevention*, 38(3), 434–444.
- Denyer, S. (2018, January). "Beijing Bets on Facial Recognition in a Big Drive for Total Surveillance." *The Washington Post*. https://www.washingtonpost.com/news/world/wp/2018/01/07/feature/in-china-facial-recognition-is-sharp-end-of-a-drive-for-totalsurveillance/?noredirect=on&utm_term=.e73091681b31.

- Feldman, S., J. Hanover, C. Burghard, & D. Schubmehl. (2012). "Unlocking the Power of Unstructured Data." IBM White Paper. <http://www-01.ibm.com/software/ebusiness/jstart/downloads/unlockingUnstructuredData.pdf>. (accessed May 2018).
- Ferrucci, D., E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, Kalyanpur, A. A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, & C. Welty. (2010). "Building Watson: An Overview of the DeepQA Project." *AI Magazine*, 31(3), pp. 59–79.
- Goodfellow, I., Y. Bengio, & A. Courville. (2016). "Deep Learning." Cambridge, MA: MIT Press.
- Goodfellow, I. J., D. Warde-Farley, P. Lamblin, V. Dumoulin, M. Mirza, R. Pascanu, . . . Y. Bengio. (2013). "Pylearn2: A Machine Learning Research Library." ArXiv Preprint ArXiv: 1308.4214.
- Graves, A. (2013). "Generating Sequences with Recurrent Neural Networks." ArXiv Preprint ArXiv:1308.0850.
- Graves, A., & N. Jaitly. (2014). "Towards End-to-End Speech Recognition with Recurrent Neural Networks." *Proceedings on International Conference on Machine Learning*, pp. 1764–1772.
- Graves, A., N. Jaitly, & A. Mohamed. (2013). "Hybrid Speech Recognition with Deep Bidirectional LSTM." IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 273–278.
- Graves, A., A. Mohamed, & G. Hinton. (2013). "Speech Recognition with Deep Recurrent Neural Networks." IEEE Acoustics, Speech and Signal Processing (ICASSP) International Conference, pp. 6645–6649.
- Graves, A., & J. Schmidhuber. (2009). "Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks." *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, pp. 545–552.
- Gualtieri, M. (2017). "Cognitive Search Is the AI Version of Enterprise Search, Forrester." go.forrester.com/blogs/17-06-12-cognitive_search_is_the_ai_version_of_enterprise_search/ (accessed July 2018).
- Haykin, S. S. (2009). *Neural Networks and Learning Machines*, 3rd ed. Upper Saddle River, NJ: Prentice Hall.
- He, K., X. Zhang, S. Ren, & J. Sun. (2015). "Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification." *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034.
- Hinton, G. E., S. Osindero, & Y.-W. Teh. (2006). "A Fast Learning Algorithm for Deep Belief Nets." *Neural Computation*, 18(7), 1527–1554.
- Hochreiter, S., & J. Schmidhuber (1997). "Long Short-Term Memory." *Neural Computation*, 9(8), 1735–1780.
- Hornik, K. (1991). "Approximation Capabilities of Multilayer Feedforward Networks." *Neural Networks*, 4(2), 251–257.
- IBM. (2011). "IBM Watson." www.ibm.com/watson/ (accessed July 2017).
- Jia, Y. (2013). "Caffe: An Open Source Convolutional Architecture for Fast Feature Embedding." <http://Goo.Gl/Fo9YO8> (accessed June 2018).
- Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, . . . T. Darrell, T. (2014). "Caffe: Convolutional Architecture for Fast Feature Embedding." *Proceedings of the ACM International Conference on Multimedia*, pp. 675–678.
- Keysers, D., T. Deselaers, H. A. Rowley, L.-L. Wang, & V. Carbune. (2017). "Multi-Language Online Handwriting Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), pp. 1180–1194.
- Krizhevsky, A., I. Sutskever, & G. Hinton. (2012). "Imagenet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems*, pp. 1097–1105S.
- Kumar, S. (2017). "A Survey of Deep Learning Methods for Relation Extraction." <http://arxiv.org/abs/1705.03645>. (accessed June 2018)
- LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, & L. D. Jackel. (1989). "Backpropagation Applied

to Handwritten ZIP Code Recognition." *Neural Computation*, 1(4), 541–551.

Liang, X., X. Shen, J. Feng, L. Lin, & S. Yan. (2016). "Semantic Object Parsing with Graph LSTM." *European Conference on Computer Vision*. New York, NY: Springer, pp. 125–143.

Mahajan, D., R. Girshick, V. Ramanathan, M. Paluri, & L. van der Maaten. (2018). "Advancing State-of-the-Art Image Recognition with Deep Learning on Hashtags." <https://code.facebook.com/posts/1700437286678763/advancing-state-of-the-art-image-recognition-with-deep-learning-on-hashtags/>. (accessed June 2018)

Mikolov, T., K. Chen, G. Corrado, & J. Dean. (2013). "Efficient Estimation of Word Representations in Vector Space." ArXiv Preprint ArXiv:1301.3781.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, & J. Dean. (2013). "Distributed Representations of Words and Phrases and Their Compositionality" *Advances in Neural Information Processing Systems*, pp. 3111–3119.

Mintz, M., S. Bills, R. Snow, & D. Jurafsky. (2009). "Distant Supervision for Relation Extraction Without Labeled Data." *Proceedings of the Joint Conference of the Forty-Seventh Annual Meeting of the Association for Computational Linguistics and the Fourth International Joint Conference on Natural Language Processing of the AFNLP*, Vol. 2, pp. 1003–1011.

Mozur, P. (2018, June 8). "Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras." *The New York Times*, issue June 8, 2018.

Nguyen, T. H., & R. Grishman. (2015). "Relation Extraction: Perspective from Convolutional Neural Networks." *Proceedings of the First Workshop on Vector Space Modeling for Natural Language Processing*, pp. 39–48.

Olson, D. L., D. Delen, and Y. Meng. (2012). "Comparative Analysis of Data Mining Models for Bankruptcy Prediction." *Decision Support Systems*, 52(2), pp. 464–473.

Principe, J. C., N. R. Euliano, and W. C. Lefebvre. (2000). *Neural and Adaptive Systems: Fundamentals Through Simulations*. New York: Wiley. Reynolds, H., & S. Feldman. (2014, July/August). "Cognitive Computing: Beyond the Hype." *KM World*, 23(7), p. 21.

Riedel, S., L. Yao, & A. McCallum. (2010). "Modeling Relations and Their Mentions Without Labeled Text." *Joint European Conference on Machine Learning and Knowledge Discovery in Databases.*, New York, NY: Springer, pp. 148–163

Robinson, A., J. Levis, & G. Bennett. (2010, October). "Informs to Officially Join Analytics Movement." *ORMS Today*.

Royyuru, A. (2014). "IBM's Watson Takes on Brain Cancer: Analyzing Genomes to Accelerate and Help Clinicians Personalize Treatments." Thomas J. Watson Research Center, www.research.ibm.com/articles/genomics.shtml (accessed September 2014).

Rumelhart, D. E., G. E. Hinton, & R. J. Williams. (1986). "Learning Representations by Back-Propagating Errors." *Nature*, 323(6088), pp. 533.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, . . . M. Bernstein. (2015). "Imagenet Large Scale Visual Recognition Challenge." *International Journal of Computer Vision*, 115(3), 211–252.

Sato, K., C. Young, & D. Patterson. (2017). "An In-Depth Look at Google's First Tensor Processing Unit (TPU)." <https://cloud.google.com/blog/big-data/2017/05/an-indepth-look-at-googles-first-tensor-processing-unittpu>. (accessed June 2018)

Scherer, D., A. Müller, & S. Behnke. (2010). "Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition." *International Conference on Artificial Neural Networks.*, New York, NY: Springer, 92–101.

Slowey, L. (2017a, January 25). "Winning the Best Picture Oscar: IBM Watson and Winning Predictions." <https://www.ibm.com/blogs/internet-of-things/best-picture-oscarwatson-predicts/>(accessed August 2018).

Slowey, L. (2017b, May 10). "Watson Predicts the Winners: Eurovision 2017." <https://www.ibm.com/blogs/internet-of-things/eurovision-watson-tone-predictions/> (accessed August 2018).

Sutskever, I., O. Vinyals, & Q. V. Le. (2014). "Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems*, pp. 3104–3112.

Ung, G. M. (2016, May). "Google's Tensor Processing Unit Could Advance Moore's Law 7 Years into the Future." *PCWorld*. <https://www.pcworld.com/article/3072256/google-io/googles-tensor-processing-unit-said-toadvance-moores-law-seven-years-into-the-future.html> (accessed July 2018).

Vinyals, O., L. Kaiser, T. Koo, S. Petrov, I. Sutskever, & G. Hinton, G. (2015). "Grammar As a Foreign Language." *Advances in Neural Information Processing Systems*, pp. 2773–2781.

Vinyals, O., A. Toshev, S. Bengio, & D. Erhan. (2015). "Show and Tell: A Neural Image Caption Generator." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3156–3164.

Vinyals, O., A. Toshev, S. Bengio, & D. Erhan. (2017). "Show and Tell: Lessons Learned from the 2015 MSCOCO Image Captioning Challenge." *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 652–663.

Wilson, R. L., & R. Sharda. (1994). "Bankruptcy Prediction Using Neural Networks." *Decision Support Systems*, 11(5), 545–557.

Wu, Y., M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, & K. Macherey. (2016). "Google's Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation." ArXiv Preprint ArXiv:1609.08144.

Xu, K., J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, & Y. Bengio. (2015). "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention." *Proceedings of the Thirty-Second International Conference on Machine Learning*, pp. 2048–2057.

Zeng, D., K. Liu, S. Lai, G. Zhou, & J. Zhao (2014). "Relation Classification via Convolutional Deep Neural Network." <http://doi.org/http://aclweb.org/anthology/C/C14/C14-1220.pdf>. (accessed June 2018).

Zhou, Y.-T., R. Chellappa, A. Vaid, & B. K. Jenkins. (1988). "Image Restoration Using a Neural Network." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7), pp. 1141–1151

فصل ٧

Bond, C. F., & B. M. DePaulo. (2006). "Accuracy of Deception Judgments." *Personality and Social Psychology Reports*, 10(3), pp. 214–234.

Brogan, C., & J. Bastone. (2011). "Acting on Customer Intelligence from Social Media: The New Edge for Building Customer Loyalty and Your Brand." SAS white paper.

Chun, H. W., Y. Tsuruoka, J. D. Kim, R. Shiba, N. Nagata, & T. Hishiki. (2006). "Extraction of Gene-Disease Relations from MEDLINE Using Domain Dictionaries and Machine Learning." *Proceedings of the Eleventh Pacific Symposium on Biocomputing*, pp. 4–15.

Coussement, K., & D. Van Den Poel. (2008). "Improving Customer Complaint Management by Automatic Email Classification Using Linguistic Style Features as Predictors." *Decision Support Systems*, 44(4), pp. 870–882.

Coussement, K., & D. Van Den Poel. (2009). "Improving Customer Attrition Prediction by Integrating Emotions from Client/Company Interaction Emails and Evaluating Multiple Classifiers." *Expert Systems with Applications*, 36(3), pp. 6127–6134.

Cutts, M. (2006, February 4). "Ramping Up on International Webspam." mattcutts.com/blog. mattcutts.com/blog/rampingup-on-international-webspam (accessed March 2013).

Delen, D., & M. Crossland. (2008). "Seeding the Survey and Analysis of Research Literature with Text Mining." *Expert Systems with Applications*, 34(3), pp. 1707–1720.

- Esuli, A., & F. Sebastiani. (2006, May). SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. *Proceedings of LREC*, 6, pp. 417–422.
- Etzioni, O. (1996). “The World Wide Web: Quagmire or Gold Mine?” *Communications of the ACM*, 39(11), pp. 65–68.
- EUROPOL. (2007). EUROPOL Work Program 2005. statewatch.org/news/2006/apr/europol-workprogramme-2005.pdf (accessed October 2008).
- Feldman, R., & J. Sanger. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Boston, MA: ABS Ventures.
- Fuller, C. M., D. Biro, and D. Delen. (2008). “Exploration of Feature Selection and Advanced Classification Models for High-Stakes Deception Detection.” *Proceedings of the Forty-First Annual Hawaii International Conference on System Sciences (HICSS)*. Big Island, HI: IEEE Press, pp. 80–99.
- Ghani, R., K. Probst, Y. Liu, M. Krema, and A. Fano. (2006). “Text Mining for Product Attribute Extraction.” *SIGKDD Explorations*, 8(1), pp. 41–48.
- Goodman, A. (2005). “Search Engine Showdown: Black Hats Versus White Hats at SES. SearchEngineWatch.” searchenginewatch.com/article/2066090/Search-Engine-Showdown-Black-Hats-vs.-White-Hats-at-SES (accessed February 2013).
- Han, J., & M. Kamber. (2006). *Data Mining: Concepts and Techniques*, 2nd ed. San Francisco, CA: Morgan Kaufmann. *Harvard Business Review*. (2010). “The New Conversation: Taking Social Media from Talk to Action.” A SAS–Sponsored Research Report by Harvard Business Review Analytic Services. sas.com/resources/whitepaper/wp_23348.pdf (accessed March 2013).
- Kaplan, A. M., & M. Haenlein. (2010). “Users of the World, Unite! The Challenges and Opportunities of Social Media.” *Business Horizons*, 53(1), pp. 59–68.
- Kim, S. M., & E. Hovy. (2004, August). “Determining the Sentiment of Opinions.” *Proceedings of the Twentieth International Conference on Computational Linguistics*, p. 1367.
- Kleinberg, J. (1999). “Authoritative Sources in a Hyperlinked Environment.” *Journal of the ACM*, 46(5), pp. 604–632.
- Lin, J., & D. Demner-Fushman. (2005). “Bag of Words” Is Not Enough for Strength of Evidence Classification.” *AMIA Annual Symposium Proceedings*, pp. 1031–1032. pubmedcentral.nih.gov/articlerender.fcgi?artid=1560897.
- Liu, B., M. Hu, & J. Cheng. (2005, May). “Opinion Observer: Analyzing and Comparing Opinions on the Web.” *Proceedings of the Fourth International Conference on World Wide Web*, pp. 342–351.
- Mahgoub, H., D. Rosner, N. Ismail, and F. Torkey. (2008). “A Text Mining Technique Using Association Rules Extraction.” *International Journal of Computational Intelligence*, 4(1), pp. 21–28.
- Manning, C. D., & H. Schütze. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- McKnight, W. (2005, January 1). “Text Data Mining in Business Intelligence.” *Information Management Magazine*. information-management.com/issues/20050101/1016487-1.html (accessed May 22, 2009).
- Mejova, Y. (2009). “Sentiment Analysis: An Overview.” Comprehensive exam paper. <http://www.cs.uiowa.edu/~ymejova/publications/CompsYelenaMejova.pdf> (accessed February 2013).
- Miller, T. W. (2005). *Data and Text Mining: A Business: Applications Approach*. Upper Saddle River, NJ: Prentice Hall.
- Morgan, N., G. Jones, & A. Hodges. (2010). “The Complete Guide to Social Media from the Social Media Guys.” thesocialmediaguys.co.uk/wp-content/uploads/downloads/2011/03/CompleteGuidetoSocialMedia.pdf (accessed February 2013).

Nakov, P., A. Schwartz, B. Wolf, and M. A. Hearst. (2005). "Supporting Annotation Layers for Natural Language Processing." *Proceedings of the ACL, Interactive Poster and Demonstration Sessions*. Ann Arbor, MI: Association for Computational Linguistics, pp. 65–68.

Paine, K. D., & M. Chaves. (2012). "Social Media Metrics." SAS white paper. sas.com/resources/whitepaper/wp_19861.pdf (accessed February 2013).

Pang, B., & L. Lee. (2008). *OPINION Mining and Sentiment Analysis*. Hanover, MA: Now Publishers; available at <http://books.google.com>.

Ramage, D., D. Hall, R. Nallapati, & C. D. Manning. (2009, August). "Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-Labeled Corpora." *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1*, pp. 248–256.

Schmidt, L.-H. (1996). "Commonness Across Cultures." In A. N. Balslev (ed.), *Cross-Cultural Conversation: Initiation* (pp. 119–132). New York: Oxford University Press.

Scott, W. R., & G. F. Davis. (2003). "Networks in and Around Organizations." *Organizations and Organizing*. Upper Saddle River, NJ: Pearson Prentice Hall.

Shatkay, H., A. Höglund, S. Brady, T. Blum, P. Dønnes, and O. Kohlbacher. (2007). "SherLoc: High-Accuracy Prediction of Protein Subcellular Localization by Integrating Text and Protein Sequence Data." *Bioinformatics*, 23(11), pp. 1410–1415.

Snyder, B., & R. Barzilay. (2007, April). "Multiple Aspect Ranking Using the Good Grief Algorithm." *HLT-NAACL*, pp. 300–307.

Strapparava, C., & A. Valitutti. (2004, May). "WordNet Affect: An Affective Extension of WordNet." *LREC*, 4, pp. 1083–1086.

The Westover Group. (2013). "20 Key Web Analytics Metrics and How to Use Them." <http://www.thewestovergroup.com> (accessed February 2013).

Thomas, M., B. Pang, & L. Lee. (2006, July). "Get Out the Vote: Determining Support or Opposition from Congressional Floor-Debate Transcripts." In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 327–335.

Weng, S. S., & C. K. Liu. (2004). "Using Text Classification and Multiple Concepts to Answer E-Mails." *Expert Systems with Applications*, 26(4), pp. 529–543.

فصل ٨

"Canadian Football League Uses Frontline Solvers to Optimize Scheduling in 2016." www.solver.com/news/canadianfootball-league-uses-frontline-solvers-optimizescheduling-2016 (accessed September 2018).

"Risk-Based Planning and Scheduling (RPS) with Simio." www.simio.com, Simio LLC, www.simio.com/aboutsimio/why-simio/simio-RPS-risk-based-planningand-scheduling.php (accessed September 2018).

Arsham, H. (2006a). "Modeling and Simulation Resources." home.ubalt.edu/ntsbarsh/Business-stat/RefSim.htm (accessed November 2018).

Arsham, H. (2006b). "Decision Science Resources." home.ubalt.edu/ntsbarsh/Business-stat/Refop.htm (accessed November 2018).

Arthur, Molly. "Simulation-Based Production Scheduling System." www.simio.com, Simio LLC, 2014, www.simio.com/case-studies/A-Steel-Tubing-Manufacturer-Expects-More-Market-Share/A-Steel-Tubing-Manufacturer-Expects-More-Market-Share.pdf (accessed September 2018).

Bailey, M. J., J. Snapp, S. Yetur, J. S. Stonebraker, S. A. Edwards, A. Davis, & R. Cox. (2011). "Practice Summaries: American

Airlines Uses Should-Cost Modeling to Assess the Uncertainty of Bids for Its Full-Truckload Shipment Routes." *Interfaces*, 41(2), 194–196.

Banks, J., & Gibson, R. R. (2009). Seven Sins of Simulation Practice." *INFORMS Analytics*, 24–27.
www.analyticsmagazine.org/summer-2009/193-strategic-problemsmodeling-the-market-space (accessed September 2018).

Bowers, M. R., C. E. Noon, W. Wu, & J. K. Bass. (2016). "Neonatal Physician Scheduling at the University of Tennessee Medical Center." *Interfaces*, 46(2), 168–182.

Chongwatpol, J., & R. Sharda. (2013). "RFID-Enabled Track and Traceability in Job-Shop Scheduling Environment." *European Journal of Operational Research*, 227(3), 453–463. <http://dx.doi.org/10.1016/j.ejor.2013.01.009>.

Christiansen, M., K. Fagerholt, G. Hasle, A. Minsaas, & B. Nygreen. (2009, April). "Maritime Transport Optimization: An Ocean of Opportunities." *OR/MS Today*, 36(2), 26–31.

Clemen, R. T., & Reilly, T. (2000). *Making Hard Decisions with Decision Tools Suite*. Belmont, MA: Duxbury Press.

Dilkina, B. N., & W. S. Havens. "The U.S. National Football League Scheduling Problem. Intelligent Systems Lab," www.cs.cornell.edu/~bistra/papers/NFLsched1.pdf (accessed September 2018).

Farasyn, I., K. Perkoz, & W. Van de Velde. (2008, July/August). "Spreadsheet Models for Inventory Target Setting at Procter & Gamble." *Interfaces*, 38(4), 241–250.

Goodwin, P., & Wright, G. (2000). *Decision Analysis for Management Judgment*, 2nd ed. New York: Wiley.

Hurley, W. J., & M. Balez. (2008, July/August). "A Spreadsheet Implementation of an Ammunition Requirements Planning Model for the Canadian Army." *Interfaces*, 38(4), 271–280. ingrammicrocommerce.com, "CUSTOMERS," <https://www.ingrammicrocommerce.com/customers/> (accessed July 2016).

Kearns, G. S. (2004, January–March). "A Multi-Objective, Multicriteria Approach for Evaluating IT Investments: Results from Two Case Studies." *Information Resources Management Journal*, 17(1), 37–62.

Kelly, A. (2002). *Decision Making Using Game Theory: An Introduction for Managers*. Cambridge, UK: Cambridge University Press.

Knight, F. H. (1933). *Risk, Uncertainty and Profit: With an Additional Introductory Essay Hither to Unpublished*. London school of economics and political science.

Koksalan, M., & S. Zionts. (Eds.). (2001). *Multiple Criteria Decision Making in the New Millennium*. Berlin: Springer-Verlag.

Kontoghiorghes, E. J., B. Rustem, & S. Siokos. (2002). *Computational Methods in Decision Making, Economics, and Finance*. Boston: Kluwer.

Kostuk, Kent J., and K. A. Willoughby. (2012). "A Decision Support System for Scheduling the Canadian Football League." *Interfaces*, 42(3), 286–295.

Manikas, A. S., J. R. Kroes, & T. F. Gattiker. (2016). Metro Meals on Wheels Treasure Valley Employs a Low-Cost Routing Tool to Improve Deliveries. *Interfaces*, 46(2), 154–167.

Mookherjee, R., J. Martineau, L. Xu, M. Gullo, K. Zhou, A. Hazlewood, X. Zhang, F. Griarte, & N. Li. (2016). "End-to-End Predictive Analytics and Optimization in Ingram Micro's Two-Tier Distribution Business." *Interfaces*, 46 (1),49–73.

Ovchinnikov, A., & J. Milner. (2008, July/August). "Spreadsheet Model Helps to Assign Medical Residents at the University of Vermont's College of Medicine." *Interfaces*, 38(4), 311–323.

Simio.com. "Cosan Case Study—Optimizing Agricultural Logistics Operations." <http://www.simio.com/case-studies/Cosan-agricultural-logistics-simulation-softwarecase-study/index.php> (accessed September 2018).

Slaugh, V. W., M. Akan, O. Kesten, & M. U. Unver. (2016). "The Pennsylvania Adoption Exchange Improves Its Matching Process." *Interfaces*, 462, 133–154.

Solver.com. "Optimizing Vendor Contract Awards Gets an A+." solver.com/news/optimizing-vendor-contractawards-gets (accessed September 2018).

Turban, E., & J. Meredith. (1994). *Fundamentals of Management Science*, 6th ed. Richard D. Irwin, Inc.

Wikipedia.com. Cosan. <https://en.wikipedia.org/wiki/Cosan> (accessed November 2018).

فصل ٩

Adapted from Alteryx.com. Great Clips. alteryx.com/sites/default/files/resources/files/case-study-great-chips.pdf (accessed September 2018).

Adapted from Snowflake.net. (n.d.). "Chime Delivers Personalized Customer Experience Using Chime." www.snowflake.net/product (accessed September 2018).

Adshead, A. (2014). "Data Set to Grow 10-fold by 2020 as Internet of Things Takes Off." www.computerweekly.com/news/2240217788/Data-set-to-grow-10-fold-by-2020-as-internet-of-things-takes-off (accessed September 2018).

Altaweel, Mark. "Accessing Real-Time Satellite Imagery and Data." GIS Lounge, 1 Aug. 2018, www.gislounge.com/accessing-real-time-satellite-imagery/.

Amodio, M. (2015). Salesforce adds predictive analytics to Marketing Cloud. Cloud Contact Center. cloudcontactcenterzone.com/topics/cloud-contact-center/articles/413611-salesforce-adds-predictive-analytics-marketingcloud.htm (accessed September 2018).

Asamoah, D., & R. Sharda. (2015). Adapting CRISP-DM process for social network analytics: Application to healthcare. *In AMCIS 2015 Proceedings*. aisel.aisnet.org/amcis2015/BizAnalytics/GeneralPresentations/33/ (accessed September 2018).

Asamoah, D., R. Sharda, A. Zadeh, & P. Kalgotra. (2016). "Preparing a Big Data Analytics Professional: A Pedagogic Experience." *In DSI 2016 Conference*, Austin, TX.

Awadallah, A., & D. Graham. (2012). "Hadoop and the Data Warehouse: When to Use Which." teradata.com/whitepapers/Hadoop-and-the-Data-Warehouse-When-to-Use-Which (accessed September 2018).

Blogs.ESRI.com. "From Customers to CxOs, Starbucks Delivers World-Class Service." (2014). <https://blogs.esri.com/esri/ucinsider/2014/07/29/starbucks/> (accessed September 2018).

Broniatowski, D. A., M. J. Paul, & M. Dredze. (2013). "National and Local Influenza Surveillance through Twitter: An Analysis of the 2012–2013 Influenza Epidemic." *PloS One*, 8(12), e83672.

Cisco. (2016). "The Zettabyte Era: Trends and Analysis." cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivitywp.pdf (accessed October 2018).

ComputerWeekly.com. (2016). "Big-Data and Open Source Cloud Technology Help Gulf Air Pin Down Customer Sentiment." www.computerweekly.com/news/450297404/Big-data-and-open-source-cloudtechnology-help-Gulf-Air-pin-down-ustomersentiment (accessed September 2018).

CxOtoday.com. (2014). "Cloud Platform to Help Pharma Co Accelerate Growth." www.cxotoday.com/story/mankind-pharma-to-drive-growth-with-softlayerscloud-platform/ (accessed September 2018).

Dalininaa, R., "Using Natural Language Processing to Analyze Customer Feedback in Hotel Reviews," www.datascience.com/resources/notebooks/data-sciencesummarize-hotel-reviews (Accessed October 2018).

DataStax. "Customer Case Studies." datastax.com/resources/casestudies/eBay (accessed September 2018).

Davis, J. (2015). "Salesforce Adds New Predictive Analytics to Marketing Cloud. Information Week." informationweek.com/big-data/big-data-analytics/salesforce-addsnew-predictive-analytics-to-marketing-cloud/d/did/1323201 (accessed September 2018).

Dean, J., & S. Ghemawat. (2004). "MapReduce: Simplified Data Processing on Large Clusters." research.google.com/archive/mapreduce.html (accessed September 2018).

Delen, D., M. Kletke, & J. Kim. (2005). "A Scalable Classification Algorithm for Very Large Datasets." *Journal of Information and Knowledge Management*, 4(2), 83–94.

Demirkan, H., & D. Delen. (2013, April). "Leveraging the Capabilities of Service-Oriented Decision Support Systems: Putting Analytics and Big Data in Cloud." *Decision Support Systems*, 55(1), 412–421.

Digit.HBS.org. (2015). "Starbucks: Brewing up a Data Storm!" <https://digit.hbs.org/submission/starbucksbrewing-up-a-data-storm/> (accessed September 2018).

Dillow, C. (2016). "What Happens When You Combine Artificial Intelligence and Satellite Imagery." fortune.com/2016/03/30/facebook-ai-satellite-imagery/ (accessed September 2018).

Ekster, G. (2015). "Driving Investment Performance with Alternative Data." integrity-research.com/wp-content/uploads/2015/11/Driving-Investment-Performance-With-Alternative-Data.pdf (accessed September 2018).

Henschen, D. (2016). "Salesforce Reboots Wave Analytics, Preps IoT Cloud." *ZD Net*. zdnet.com/article/salesforce-rebootswave-analytics-preps-iot-cloud/ (accessed September 2018).

Higginbotham, S. (2012). "As Data Gets Bigger, What Comes after a Yottabyte?" gigaom.com/2012/10/30/as-datagets-bigger-what-comes-after-a-yottabyte (accessed September 2018).

Hope, B. (2015). "Provider of Personal Finance Tools Tracks Bank Cards Sells Data to Investors." *Wall Street Journal*. wsj.com/articles/provider-of-personal-finance-toolstracks-bank-cards-sells-data-to-investors-1438914620 (accessed September 2018).

Jonas, J. (2007). "Streaming Analytics vs. Perpetual Analytics (Advantages of Windowless Thinking)." jeffjonas.typepad.com/jeff_jonas/2007/04/streaming_analy.html (accessed September 2018).

Kalgotra, P., & R. Sharda. (2016). "Rural Versus Urban Comorbidity Networks." Working Paper, Center for Health Systems and Innovation, Oklahoma State University.

Kalgotra, P., R. Sharda, & J. M. Croff. (2017). "Examining Health Disparities by Gender: A Multimorbidity Network Analysis of Electronic Medical Record." *International Journal of Medical Informatics*, 108, 22–28.

Kelly, L. (2012). "Big Data: Hadoop, Business Analytics, and Beyond." wikibon.org/wiki/v/Big_Data:_Hadoop,_Business_Analytics_and_Beyond (accessed September 2018).

Moran, P. A. (1950). "Notes on Continuous Stochastic Phenomena." *Biometrika*, 17–23.

"Overstock.com: Revolutionizing Data and Analytics to Connect Soulfully with their Customers," at <https://www.teradata.com/Resources/Videos/Overstock-com-Revolutionizing-data-and-analy> (accessed October 2018).

"Overstock.com Uses Teradata Path Analysis To Boost Its Customer Journey Analytics," March 27, 2018, at <https://www.retailtinsights.com/doc/overstock-comuses-teradata-path-analysis-boost-customer-journeyanalytics-0001> (accessed October 2018).

Palmucci, J., "Using Apache Spark for Massively Parallel NLP," at <http://engineering.tripadvisor.com/using-apachespark-for-massively-parallel-nlp/> (accessed October 2018).

- RedHat.com. (2016). "Gulf Air's Big Data Innovation Delivers Deeper Customer Insight." <https://www.redhat.com/en/success-stories> (accessed September 2018).
- RedHat.com. (2016). "Gulf Air Builds Private Cloud for Big Data Innovation with Red Hat Technologies." <https://www.redhat.com/en/about/press-releases/gulf-airbuilds-private-cloud-big-data-innovation-red-hattechnologies> (accessed September 2018).
- Russom, P. (2013). "Busting 10 Myths about Hadoop: the Big Data Explosion." *TDWI's Best of Business Intelligence*, 10, 45–46.
- Sarasohn-Kahn, J. (2008). *The Wisdom of Patients: Health Care Meets Online Social Media*. Oakland, CA: California HealthCare Foundation.
- Shaw, C. (2016). "Satellite Companies Moving Markets." quandl.com/blog/alternative-data-satellite-companies (accessed September 2018).
- Steiner, C. (2009). "Sky High Tips for Crop Traders" (accessed September 2018).
- St Louis, C., & G. Zorlu. (2012). "Can Twitter Predict Disease Outbreaks?" *BMJ*, 344.
- Tableau white paper. (2012). "7 Tips to Succeed with Big Data in 2013." cdnlarge.tableausoftware.com/sites/default/files/whitepapers/7-tips-to-succeed-with-big-data-in-2013.pdf (accessed September 2018).
- Tartar, Andre, et al. "All the Things Satellites Can Now See From Space." Bloomberg.com, Bloomberg, 26 July 2018, www.bloomberg.com/news/features/2018-07-26/all-the-things-satellites-can-now-see-from-space (accessed October 2018).
- Thusoo, A., Z. Shao, & S. Anthony. (2010). "Data Warehousing and Analytics Infrastructure at Facebook." In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data* (p. 1013).
- Turner, M. (2015). "This Is the Future of Investing, and You Probably Can't Afford It." businessinsider.com/hedgefunds-are-analysing-data-to-get-an-edge-2015-8 (accessed September 2018).
- Watson, H. (2012). "The Requirements for Being an Analytics-Based Organization." *Business Intelligence Journal*, 17(2), 42–44.
- Watson, H., R. Sharda, & D. Schrader. (2012). "Big Data and How to Teach It." *Workshop at AMCIS*, Seattle, WA.
- Wheeler, C. (2014). "Going Big with GIS." www.esri.com/esri-news/arcwatch/0814/going-big-with-gis (accessed October 2018).
- White, C. (2012). "MapReduce and the Data Scientist." Teradata Vantage White Paper. teradata.com/white-paper/MapReduce-and-the-Data-Scientist (accessed September 2018).
- Wikipedia.com. "Petabyte." en.wikipedia.org/wiki/Petabyte (accessed September 2018).
- Zadeh, A. H., H. M. Zolbanin, R. Sharda, & D. Delen. (2015). "Social Media for Nowcasting the Flu Activity: Spatial-Temporal and Text Analysis." *Business Analytics Congress, Pre-ICIS Conference*, Fort Worth, TX.
- Zikopoulos, P., D. DeRoos, K. Parasuraman, T. Deutsch, D. Corrigan, & J. Giles. (2013). *Harness the Power of Big Data*. New York: McGraw-Hill.
- "Zion China uses Azure IoT, Stream Analytics, and Machine Learning to Evolve Its Intelligent Diabetes Management Solution," www.codeproject.com/Articles/1194824/Zion-China-uses-Azure-IoTStream-Analytics-and-M (accessed October 2018) and <https://microsoft.github.io/techcasestudies/iot/2016/12/02/IoT-ZionChina.html> (accessed October 2018).

"A Brief History of Robotics since 1950." Encyclopedia.com. <http://www.encyclopedia.com/science/encyclopediasalmanacs-transcripts-and-maps/brief-history-robotics-1950> (accessed September 2018).

Ackerman, E. (2016). *IEEE Spectrum*. <http://spectrum.ieee.org/automaton/robotics/home-robots/tegami-latest-friendly-squishable-social-robot> (March 5, 2017).

"Adidas's High-Tech Factory Brings Production Back to Germany." (2017, January 14). *The Economist*. <https://www.economist.com/business/2017/01/14/adidas-high-tech-factory-brings-production-back-to-germany> (accessed September 2018).

Allinson, M. (2017, March 4). "BMW Shows Off Its Smart Factory Technologies at Its Plants Worldwide." Robotics and Automation. <https://roboticsandautomationnews.com/2017/03/04/bmw-shows-off-its-smartfactory-technologies-at-its-plants-worldwide/11696/> (accessed September 2018).

Aoki, S., et al. (1999). "Automatic Construction Method of Tree-Structural Image Conversion Method ACTIT." *Journal of the Institute of Image Information and Television Engine*, 53(6), pp. 888–894 (in Japanese).

"A Robot Cooks Burgers at Startup Restaurant Creator." (2018). Techcrunch. <https://techcrunch.com/video/arobot-cooks-burgers-at-startup-restaurant-creator/> (accessed September 2018).

Ayres, R., & S. Miller. (1981, November). "The Impacts of Industrial Robots." Report CMU-RI-TR-81-7. Pittsburgh, PA: The Robotics Institute at Carnegie Mellon University.

Bereznak, A. (2015, January 7). "This Robot Can Comfort Children Through Chemotherapy." Yahoo Finance. <https://finance.yahoo.com/news/this-robot-cancomfort-children-through-107365533404.html> (accessed September 2018).

"Berry Picking at Its Best with Sensor Technology." (2018). Pepperl+Fuchs. <https://www.pepperl-fuchs.com/usa/en/27566.htm> (accessed September 2018).

Bogue, R. (2016). "Robots Poised to Revolutionise Agriculture." *Industrial Robot: An International Journal*, 43(5), pp. 450–456

Broekens, J., M. Heerink, & H. Rosendal. (2009). "Assistive Social Robots in Elderly Care: A Review." *Gerontechnology*, 8, pp. 94–103 doi: 10.4017/gt.2009.08.02.002.00.

Carlsson, B. (1998) "The Evolution of Manufacturing Technology and Its impact on Industrial Structure: An International Study." IUI Working Paper 203. International Joseph A. Schumpeter Society Conference on Evolution of Technology and Market in an International Context. The Research Institute of Industrial Economics (IUI), Stockholm, May 24–28, 1988.

"Case Study Pepper, Courtyard Marriott." SoftBank Robotics. <https://www.softbankrobotics.com/us/solutions/pepper-marriott> (accessed September 2018).

Chirgwin, R. (2018, May 29). "Softbank's 'Pepper' Robot Is a Security Joke." The Register. https://www.theregister.co.uk/2018/05/29/softbank_pepper_robot_multiple_basic_security_flaws/ (accessed September 2018).

Coxworth, B. (2018, May 29). "Restaurant Keeps Its Prices Down – With a Robotic Kitchen." New Atlas. <https://newatlas.com/spyce-restaurant-robotickitchen/54818/> (accessed September 2018).

"Da Vinci Robotic Prostatectomy – A Modern Surgery Choice!" Robotic Oncology. <https://www.roboticoncology.com/da-vinci-robotic-prostatectomy/> (accessed September 2018).

Drummond, K. (2012, March 8). "Navy's Newest Robot Is a Mechanized Firefighter." wired.com. <https://www.wired.com/2012/03/firefight-robot/> (accessed September 2018).

- Dupont, T. (2015, October 15). "The MAARS Military Robot." Prezi. <https://prezi.com/fsrlswo0qklp/the-maars-militaryrobot/> (accessed September 2018).
- Engel, J. (2018, May 3). "Spyce, MIT-Born Robotic Kitchen Startup, Launches Restaurant: Video." Xconomy. <https://www.xconomy.com/boston/2018/05/03/spyce-mitborn-robotic-kitchen-startup-launches-restaurantvideo/> (accessed September 2018).
- Fallon, S. (2015). "A Blue Robotic Bear to Make Sick Kids Feel Less Blue." <https://www.wired.com/2015/03/bluerobotic-bear-make-sick-kids-feel-less-blue/> (accessed August 2018).
- Forrest, C. (2015). "Chinese Factory Replaces 90% of Humans with Robots, Production Soars." TechRepublic. <https://www.techrepublic.com/article/chinese-factoryreplaces-90-of-humans-with-robots-productionsoars/> (accessed September 2018).
- France, A. (2014, December 1). "Nestlé Employs Fleet of Robots to Sell Coffee Machines in Japan." *The Guardian*. <https://www.theguardian.com/technology/2014/dec/01/nestle-robots-coffee-machines-japan-george-clooneypepper-android-softbank> (accessed September 2018).
- Gandhi, A. (2013, February 23). "Basics of Robotics." Slideshare. <https://www.slideshare.net/AmeyaGandhi/basics-ofrobotics> (accessed September 2018).
- Goris, K., et al. (2010, September). "Mechanical Design of the Huggable Robot Probo." Robotics & Multibody Mechanics Research Group. Brussels, Belgium: Vrije Universiteit Brussels.
- Green, D. (2018). "Adidas Just Opened a Futuristic New Factory – and It Will Dramatically Change How Shoes Are Sold." Business Insider. <http://www.businessinsider.com/adidas-high-tech-speedfactory-begins-production-2018-4> (accessed September 2018).
- Hernandez, D. (2018). "Seven Jobs Robots Will Create – or Expand." *The Wall Street Journal*. <https://www.wsj.com/articles/seven-jobs-robots-will-createorexpand-1525054021> (accessed September 2018).
- History of Robots. (n.d.). Wikipedia. https://en.wikipedia.org/wiki/History_of_robots (accessed September 2018).
- "Huggable Robot Befriends Girl in Hospital." YouTube video. <https://youtu.be/UaRCCA2rRR0> (accessed August 2018).
- "Innovative Human-Robot Cooperation in BMW Group Production." (2013, October 9). BMW Press Release. <https://www.press.bmwgroup.com/global/article/detail/T0209722EN/innovative-human-robotcooperation-in-bmw-group-production?language=en> (accessed September 2018).
- Javelosa, J., & K. Houser. (2017). "Production Soars for Chinese Factory Who Replaced 90% of Employees with Robots." Future Society. <https://futurism.com/2-production-soarsfor-chinese-factory-who-replaced-90-of-employeeswith-robots/> (accessed September 2018).
- Jeong, S., et al. (2015). "A Social Robot to Mitigate Stress, Anxiety, and Pain in Hospital Pediatric Care." *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*.
- Jeong, S., et al. (2015). "Designing a Socially Assistive Robot for Pediatric Care." *Proceedings of the Fourteenth International Conference on Interaction Design and Children. ACM*.
- Jeong, S., & D. Logan. (2018, April 21–26). "Huggable: The Impact of Embodiment on Promoting Socio-emotional Interactions for Young Pediatric Surgeons." MIT Media Lab, Cambridge, MA, CHI 2018, Montréal, QC, Canada.
- Jiji. (2017, November 21). "SoftBank Upgrades Humanoid Robot Pepper." *The Japan Times*. <https://www.japantimes.co.jp/news/2017/11/21/business/tech/softbankupgrades-humanoid-robot-pepper/#.W6B3qPZFzIV> (accessed September 2018).
- Joshua, J. (2013, February 24). "The 3 Types of Robots." Prezi. <https://prezi.com/iifjw387ebum/the-3-types-of-robots/>

(accessed September 2018).

Kelly, M. (2018, July 16). "This Adorable Robot Wants to Make Air Travel Less Stressful." The Verge. <https://www.theverge.com/2018/7/16/17576334/klm-royal-dutch-airlinesrobot-travel-airport> (accessed September 2018).

Kelly, S. M. (2017, August 10). "A Robotic Crib Rocked My Baby to Sleep for Months." CNN Tech. <https://money.cnn.com/2017/08/10/technology/gadgets/snooreview/index.html> (accessed September 2018).

Lee, K. F. (2018). "The Human Promise of the AI Revolution." *The Wall Street Journal*. <https://www.wsj.com/articles/the-human-promise-of-the-ai-revolution-1536935115> (accessed September 2018).

Mayank. (2012, June 18). "Basic Parts of a Robot." max-Embedded.com. <http://maxembedded.com/2012/06/basic-parts-of-a-robot/> (accessed September 2018).

McHugh, R., & J. Rascon. (2015, May 23). "Meet MEDi, the Robot Taking Pain Out of Kids' Hospital Visits." NBC News. <https://www.nbcnews.com/news/us-news/meet-medirobot-taking-pain-out-kids-hospital-visits-n363191> (accessed September 2018).

Meister, J. (2017), "The Future Of Work: How Artificial Intelligence Will Transform The Employee Experience," <https://www.forbes.com/sites/jeannemeister/2017/11/09/the-future-of-work-how-artificial-intelligence-will-transform-the-employee-experience/> (accessed November 2018).

Modular Advanced Armed Robotic System. (n.d.). Wikipedia. https://en.wikipedia.org/wiki/Modular_Advanced_Armed_Robotic_System (accessed September 2018).

Nagato, T., H. Shibuya, H. Okamoto, & T. Koezuka. (2017, July). "Machine Learning Technology Applied to Production Lines: Image Recognition System." *Fujitsu Scientific & Technical Journal*, 53(4).

O'Kane, S. (2018, May 17). "Raden Is the Second Startup to Bite the Dust After Airlines Ban Some Smart Luggage." Circuit Breaker. <https://www.theverge.com/circuitbreaker/2018/5/17/17364922/raden-smart-luggage-airline-ban-bluesmart> (accessed September 2018).

Park, H. W., et al. (2017). "Growing Growth Mindset with a Social Robot Peer." *Proceedings of the Twelfth ACM/IEEE International Conference on Human Robot Interaction*.

Personal Robots Group. (2016). <https://www.youtube.com/watch?v=sF0tRCqvyT0> (accessed March 5, 2017).

Personal Robots Group, MIT Media Lab. (2017). "Growing Growth Mindset with a Social Robot Peer." *Proceedings of the Twelfth ACM/IEEE International Conference on Human Robot Interaction*.

Prasad, C. (2018, January 22). "Fabio, the Pepper Robot, Fired for 'Incompetence' at Edinburgh Store." *IBN Times*. <https://www.ibntimes.com/fabio-pepper-robot-firedincompetence-edinburgh-store-2643653> (accessed September 2018).

Ro, L. (2016, October 18). "World's First Smart Crib SNOO Will Help Put Babies to Sleep." Curbed. <https://www.curbed.com/2016/10/18/13322582/snoo-smart-cribives-behar-dr-harvey-karp-happiest-baby> (accessed September 2018).

"Robotics Facts." Idaho Public Television. <http://idahoptv.org/sciencetrek/topics/robots/facts.cfm> (accessed September 2018).

"Robots in Agriculture." (2015, July 6). Intorobotics. <https://www.intorobotics.com/35-robots-in-agriculture/> (accessed September 2018).

"Robotics: Types of Robots." ElectronicsTeacher.com. <http://www.electronicsteacher.com/robotics/typeof-robots.php> (accessed September 2018).

Rosencrance, L. (2018 May 31). "Tabletop Grapes to Get Picked by Robots in India, with Help from Virginia Tech." Robotics Business Review. <https://www.roboticsbusinessreview.com/agriculture/tabletop-grapespicked-robots-india-virginia-tech/> (accessed September 2018).

Schuster, W. M. (2018). "Artificial Intelligence and Patent Ownership." *Washington & Lee L. Rev.*, 75.
Shadbolt, P. (2015, February 15). "U.S. Navy Unveils Robotic Firefighter." CNN. <https://www.cnn.com/2015/02/12/tech/mci-saffir-robot/index.html> (accessed September 2018).

"Shipboard Autonomous Firefighting Robot – SAFFiR." (2015, February 4). YouTube. https://www.youtube.com/watch?time_continue=252&v=K4OtS534oYU (accessed September 2018).

Simon, M. (2018, May 17). "The Wired Guide to Robots." Wired. <https://www.wired.com/story/wired-guide-torobots/> (accessed September 2018).

"Tabletop Grapes to Get Picked by Robots in India." Agtechnews.com. <http://agtechnews.com/Ag-Robotics-Technology/Tabletop-Grapes-to-Get-Picked-by-Robotsin-India.html> (accessed September 2018).

"The da Vinci® Surgical System." (2015, September). Da Vinci Surgery. <http://www.davincisurgery.com/da-vinci-surgery/da-vinci-surgical-system/> (accessed September 2018).

"Types of Robots." (2018). RoverRanch. <https://prime.jsc.nasa.gov/ROV/types.html> (accessed September 2018).

Westlund, J. K., J. M. Lee, J. Plummer, L. Faridia, F. Gray, J. Berlin, M. Quintus-Bosz, H. Harmann, R. Hess, M. Dyer, S. dos Santos, K. Adalgeirsson, S. Gordon, G. Spaulding, S. Martinez, M. Das, M. Archie, M. Jeong, & C. Breazeal, C. (2016).

"Tega: A Social Robot." Video Presentation. *Proceedings of the Eleventh ACM/IEEE International Conference on Human Robot Interaction*.

White, T. (2015, February 4). "Making Sailors 'SAFFiR' –Navy Unveils Firefighting Robot Prototype at Naval Tech EXPO." America's Navy. https://www.navy.mil/submit/display.asp?story_id=85459 (accessed September 2018).

Zimberoff, L. (2018, June 21). "A Burger Joint Where Robots Make Your Food." <https://www.wsj.com/articles/a-burger-joint-where-robots-make-your-food-1529599213> (accessed September 2018).

فصل ١١

Babbar-Sebens, M., et al. "A Web-Based Software Tool for Participatory Optimization of Conservation Practices in Watersheds." *Environmental Modelling & Software*, 69, 111–127, July 2015.

Basco-Carrera, L., et al. "Collaborative Modelling for Informed Decision Making and Inclusive Water Development." *Water Resources Management*, 31:9, July 2017.

Bhandari, R., et al. "How to Avoid the Pitfalls of IT Crowdsourcing to Boost Speed, Find Talent, and Reduce Costs." *McKinsey & Company*, June 2018.

Bridgwater, A. "Governments to Tap IoT for 'Collective Intelligence.'" *Internet of Business*, January 2, 2018.

Carter, R. "The Growing Power of Artificial Intelligence in Workplace Collaboration." *UC Today*, June 28, 2017.

Chiu, C-M., T. P. Liang, and E. Turban. "What Can Crowdsourcing Do for Decision Support?" *Decision Support Systems*, September 2014.

Coleman, D. "10 Components of Collaborative Intelligence." *CMS Wire*, November 21, 2011.

de Lares Norris, M. A. "Collaboration Technology Is the Driving Force for Productivity and Businesses Need to Embrace It . . . Now." *IT ProPortal*, January 4, 2018.

DeSanctis, G., and R. B. Gallupe. "A Foundation for the Study of Group Decision Support Systems." *Management Science*, 33:5, 1987.

Dewhurst, M., and P. Willmott. "Manager and Machine: The New Leadership Equation." *McKinsey & Company*, September 2014.

Dignan, L. "A Sweet Idea: Hershey Crowdsourcing for Summer Chocolate Shipping Concepts." *ZDNet*, January 14, 2016.

Digneo, C. "49 Online Collaboration Tools to Help Your Team Be More Productive." *Time Doctor*, 2018. biz30.timedoctor.com/online-collaboration-tools/(accessed July 2018).

50Minutes.com. *The Benefits of Collective Intelligence: Make the Most of Your Team's Skills*. Brussels, Belgium: **50Minutes.com** (Lemaitre Publishing), 2017.

Finnegan, M. "Cisco Shakes Up Collaboration Efforts; Morphs Spark into Webex." *Computer World*, May 2, 2018.

Goecke, J. "Meet Cisco Spark Assistant, Your Virtual Assistant for Meetings." *Cisco Blogs*, November 2, 2017.

Goldstein, P. "How Can AI Improve Collaboration Technology?" *Biztech Magazine*, June 5, 2017.

Grant, R. P. "Why Crowdsourcing Your Decision-Making Could Land You in Trouble." *The Guardian*, March 10, 2015.

Howe, J. *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*. New York: Crown Business, 2008.

Hughes, C., et al. *Build Your Own Teams of Robots with LEGO® Mindstorms® NXT and Bluetooth®*. New York, NY:McGraw-Hill/Tab Electronic, 2013.

Kang, C-K. "Marsbee—Swarm of Flapping Wing Flyers for Enhanced Mars Exploration." *NASA.gov*, March 30, 2018.

Keith, E. "Here's How a New Crowd-Sourced Map Is Making Canadian Streets Safer for Cyclists." *Narcity.com*, June 2018.

Kurzer, R. "Meet Suzy: The New Crowd Intelligence Platform with the Cute Name." *MarTech Today*, March 27, 2018.

Loten, A. "The Morning Download: AI-Enabled Sales Tools Spotlight Data Needs." *The Wall Street Journal*, March 27, 2018.

Marr, B. "Are You Ready to Meet Your Intelligent Robotic Co-Worker?" *Forbes.com*, September 8, 2017.

McMahon, K., et al. "Beyond Idea Generation: The Power of Groups in Developing Ideas." *Creativity Research Journal*, 28, 2016.

Microsoft. "Hendrick Motorsports Uses Microsoft Teams to Win Productivity Race." *Customers.Microsoft.com*, April 27, 2017.

Moran, C. "How Should Your Company Prepare for Robot Coworkers?" *Fast Company*, February 13, 2018.

Morar HPI. "A Global Survey Reveals Employee Perception of Advanced Technologies and Virtual Assistants in the Workplace." *Cisco.com*, October 2017.

Mulgan, G. *Big Mind: How Collective Intelligence Can Change Our World*. Princeton, NJ: Princeton University Press, 2017.

Nizri, G. "Shaping the Future of Work: A Collaboration of Humans and AI." *Forbes.com*, August 17, 2017.

Pena, S. "12 Benefits of a Collaborative Workspace." *Creator*, June 14, 2017. wework.com/creator/start-yourbusiness/12-benefits-of-a-collaborative-workspace/ (accessed July 2018).

Power, B. "Improve Decision-Making with Help from the Crowd." *Harvard Business Review*, April 8, 2014.

Reese, H. "New Research Shows That Swarm AI Makes More Ethical Decisions Than Individuals." *Tech Republic*, June 8, 2016.

Ruiz-Hopper, M. "Hendrick Motorsports Gains Competitive Advantage on the Race Track." *Microsoft.com*, September 26, 2016.

Staff Writers. "Scientists Simulate a Space Mission in Mars-Analogue Utah Desert." *Mars Daily*, October 19, 2016.

Stewart, C. "The 18 Best Tools for Online Collaboration." *Creative Blog*, March 7, 2017.

Thorn, C., and J. Huang. "How Carnegie Is Using Technology to Enable Collaboration in Networks." *Carnegie Foundation Blog*, September 9, 2014.

Tobe, F. "Why Co-Bots Will Be a Huge Innovation and Growth Driver for Robotics Industry." *IEEE Spectrum*, December 30, 2015.

Unanimous AI. "XPRIZE Uses Swarm AI Technology to Optimize Visioneers Summit Ideation." *Unanimous AI*, 2018. [UAI_case_study_xprize_0601_0601.pdf](#) (accessed July 2018).

Wladawsky-Berger, I. "Building an Effective Human-AI Decision System." *The Wall Street Journal*, December 1, 2017.

Xia, L. "Improving Group Decision-Making by Artificial Intelligence." In C. Sierra, Editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI*, 2017.

Yazdani, M., et al. "A Group Decision Making Support System in Logistics and Supply Chain Management." *Expert Systems with Applications*, 88, December 1, 2017.

Yoon, Y., et al. "Preference Clustering-Based Mediating Group Decision-Making (PCM-GDM) Method for Infrastructure Asset Management." *Expert Systems with Applications*, 83, October 15, 2017.

Yurieff, K. "Robot Predicts Boston Will Win Amazon HQ2." *CNN Tech*, March 13, 2018a.

Yurieff, K. "Robot Co-Workers? 7 Cool Technologies Changing the Way We Work." *CNN Tech*, May 4, 2018b.

فصل ١٢

Afaq, O. "Developing a Chatbot Using Microsoft's Bot Framework, LUIS and Node.js (Part 1)." *Smashing Magazine*, May 30, 2017. [smashingmagazine.com/2017/05/chatbot-microsoft-bot-framework-luis-nodejs-part1/](#) (accessed April 2018).

Aggarwal, C. *Recommended Systems: The Textbook*. [eTextbook]. New York, NY: Springer, 2016.

Arora, S. "Recommendation Engines: How Amazon and Netflix Are Winning the Personalization Battle." *Martech Advisor*, June 28, 2016.

Arthur, R. "Sephora Launches Chatbot on Messaging App Kik." *Forbes*, March 30, 2016.

Bae, J. "Development and Application of a Web-Based Expert System Using Artificial Intelligence for Management of Mental Health by Korean Emigrants." *Journal of Korean Academy of Nursing*, April 2013.

Beaver, L. "Chatbots Explained: Why Businesses Should Be Paying Attention to the Chatbot Revolution." *Business Insider*, March 4, 2016.

CBS News. "LinkedIn Adding New Training Features, News Feeds and 'Bots.'" *CBS News*, September 22, 2016. [cbsnews.com/news/linkedin-adding-new-trainingfeatures-news-feeds-and-bots](#) (accessed April 2018).

Clark, D. "IBM: A Billion People to Use Watson by 2018." *The Wall Street Journal*, October 26, 2016.

Cognizant. "Bot Brings Transavia Airlines Closer to Customers." *Cognizant Services*, 2017. [https://www.cognizant.com/content/dam/Cognizant_Dotcom/landing-pageresources/transavia-case-study.pdf](#) (accessed April 2018).

Costa, A., et al. (eds.) *Personal Assistants: Emerging Computational Technologies (Intelligent Systems Reference Library)*. New York, NY: Springer, 2018.

Crist, R. "How to Get Started with an Alexa Smart Home." *CNET*, July 5, 2017. [cnet.com/how-to/how-to-get-started-withan-alexa-smart-home/](#) (accessed April 2018).

Crook, J. "WeWork Has Big Plans for Alexa for Business." *TechCrunch*, November 30, 2017.

Davydova, O. "25 Chatbot Platforms: A Comparative Table." *Chatbots Journal*, May 11, 2017.

De Aenlle, C. "A.I. Has Arrived in Investing. Humans Are Still Dominating." *The New York Times*, January 12, 2018.

Demmissie, L. "Robo Advisors: The Future of Finance." *The Ticker Tape*, March 13, 2017.

Ell, K. "ETFs Powered by Artificial Intelligence Are Getting Smarter, Says Fund Co-Founder." *CNBC News*, January 23, 2018.

Eule, A. "Rating the Robo-Advisors." *Barron's*, July 29, 2017.

Ferron, E. "Mobile 101: What Are Bots, Chatbots and Virtual Assistants?" *New Atlas*, February 16, 2017. newatlas.com/what-is-bot-chatbot-guide/47965/ (accessed April 2018).

Garg, N. "Case Study: How Kenyt Real Estate Chatbot Is Generating Leads." *Medium*, June 22, 2017.

Gikas, M. "What the Amazon Echo and Alexa Do Best." *Consumer Reports*, July 29, 2016. consumerreports.org/wireless-speakers/what-amazon-echo-and-alexa-dobest (accessed April 2018).

Gilani, S. "Your Perfectly Diversified Portfolio Could Be in Danger—Here's Why." *Money Morning @ Wall Street*, December 6, 2016.

Griffiths, T. "Using Chatbots to Improve CRM Data: A WeChat Case Study." *Half a World*, November 16, 2016.

Gwynn, J. "Zuckerberg's Facebook Messenger Launches 'Chat Bots' Platform." *USA Today*, April 12, 2016.

Hachman, M. "Microsoft Combines Cortana and Bing with Microsoft Research to Accelerate New Features." *PCWorld*, September 29, 2016.

Huang, N. "Robo Advisers Get the Human Touch." *Kiplinger's Personal Finance*, September 2017.

Hunt, M. "Enterprise Chatbots and the Conversational Commerce Revolutionizing Business." *Entrepreneur*, July 3, 2017.

Ignat, A. "Iggy—A Chatbot UX Case Study." *Chatbot's Life*, August 9, 2017. chatbotslife.com/iggy-a-chatbot-ux-casestudy-b5ac0379029c/ (accessed April 2018).

Ismail, K. "Top 14 Chatbot Building Platforms of 2014." *CMS Wire*, December 19, 2017.

Johnson, K. "Everything Amazon's Alexa Learned to Do in 2017." *Venturebeat.com*, December 29, 2017.

Kaya, E. *Bot Business 101: How to Start, Run & Grow Bot/AI Business*. Kindle Edition. Seattle, WA: Amazon Digital Services, 2017.

Kelly, H. "Amazon wants Alexa everywhere." *CNN Tech*, September 22, 2018.

Kelly, H. "Battle of the Smart Speakers: Google Home vs. Amazon Echo." *CNN Tech*, May 20, 2016. money.cnn.com/2016/05/20/technology/google-home-amazonecho/index.html?iid=EL (accessed April 2018).

Keppel, D. *Best Robo-Advisor: Ultimate Automatic Wealth Management*. North Charleston, SC: Create Space Pub., 2016.

Knight, K. "Expert: Bots May Be a Marketers New Best Friend." *BizReport*, December 7, 2017a.

Knight, K. "Expert: How to Engage Chatbots Without Losing the Human Touch." *BizReport*, February 13, 2017b.

Knight, K. "Report: Over Half of Millennials Have or Will Use Bots." *Biz Report*, February 24, 2017c.

Korosec, K. "Start Your Car from Inside Your Home Using Amazon's Alexa." *Fortune.com*, August 18, 2016.

Lachecca, D. "Conversational AI Creates New Dialogues for Government." *eGovInnovation*, October 24, 2017.

Larson, S. "Baidu Is Bringing AI Chatbots to Healthcare." *CNNTech*, October 11, 2016.

Lovett, L. "Chatbot Campaign for Flu Shots Bolsters Patient Response Rate by 30%." *Healthcareitnews.com*, January 24, 2018.

Mah, P. "The State of Chatbots in Marketing." *CMOInnovation*, November 4, 2016.

Makadia, M. "Benefits for Recommendation Engines to the Ecommerce Sector." *Business 2 Community*, January 7, 2018.

Mangalindan, J. P. "RBC: Amazon Has a Potential Mega-Hit on Its Hand." *Yahoo! Finance*, April 25, 2017.

Marino, J. "Big Banks Are Fighting Robo-Advisors Head On." *CNBC News*, June 26, 2016.

Matney, L. "Siri-Creator Shows Off First Public Demo of Viv, 'The Intelligent Interface for Everything.'" *Tech Crunch*, May 9, 2016. techcrunch.com/2016/05/09/siri-creator-shows-off-first-public-demo-of-viv-the-intelligent-interface-for-everything (accessed April 2018).

McClellan, J. "What the Evolving Robo Advisory Industry Offers." *AII Journal*, October 2016.

Morgan, B. "How Chatbots Improve Customer Experience in Every Industry: An Infograph." *Forbes*, June 8, 2017.

Newlands, M. "How to Create a Facebook Messenger Chatbot for Free Without Coding." *Entrepreneur*, March 14, 2017a.

Newlands, M. "10 Ways Enterprise Chatbots Empower CEOs." *MSN.com*, August 9, 2017b. msn.com/en-us/money/smallbusiness/10-ways-enterprise-chatbotsemPOWER-ceos/ar-AApMgU8 (accessed April 2018).

Noyes, K. "Watson's the Name, Data's the Game." *PCWorld*, October 7, 2016.

Nur, N. "Singapore's POSB Launches AI-Driven Chatbot on Facebook Messenger." *MIS Asia*, January 19, 2017.

O'Brien, M. "What Can Chatbots Do for Ecommerce?" *ClickZ.com*, April 11, 2016.

Oremus, W. "When Will Alexa Know Everything?" *Slate.com*, April 6, 2018.

O'Shea, A. "Best Robo-Advisors: 2016 Top Picks." *NerdWallet*, March 14, 2016.

O'Shea, A. "Betterment Review 2017." *NerdWallet*, January 31, 2017.

Perez, S. "Voice-Enabled Smart Speakers to Reach 55% of U.S. Households by 2022, Says Report." *Tech Crunch*, November 8, 2017.

Pohjanpalo, K. "Investment Bankers Are Hard to Replace with Robots, Nordea Says." *Bloomberg*, November 27, 2017.

Popper, B. "How Netflix Completely Revamped Recommendations for Its New Global Audience." *The Verge*, February 17, 2016. theverge.com/2016/2/17/11030200/netflixnew-recommendation-system-global-regional (accessed April 2018).

Quoc, M. "10 Ecommerce Brands Succeeding with Chatbots." *A Better Lemonade Stand*, October 23, 2017.

Radu, M. "How to Pay Less for Advertising? Use Baro—An Ad Robot for Campaigns Optimization." *150sec.com*, August 18, 2016.

Rayome, A. "How Sephora Is Leveraging AR and AI to Transform Retail and Help Customers Buy Cosmetics." *TechRepublic*, February, 2018, "If This Model Is Right." *Bloomberg Business*, June 18, 2015.

Reisinger, D. "10 Reasons to Buy the Amazon Echo Virtual Personal Assistant." Slide Show. *eWeek*, February 9, 2016.

Schlicht, M. "The Complete Beginner's Guide to Chatbots." *Chatbots Magazine*, April 20, 2016. chatbotsmagazine.com/the-complete-beginner-s-guide-to-chatbots-8280b7b906ca (accessed April 2018).

StartUp. "How Netflix Uses Big Data." *Medium.com*, January 12, 2018. medium.com/swlh/how-netflix-uses-bigdata-20b5419c1edf (accessed April 2018).

Sun, Y. "Alibaba's AI Fashion Consultant Helps Achieve Record-Setting Sales." *MIT Technology Review*, November 13, 2017.

TalKing. "Top Useful Chatbots for Health." *Chatbots Magazine*, February 7, 2017.

Taylor, S. "Very Human Lessons from Three Brands That Use Chatbots to Talk to Customers." *Fast Company*, October 21, 2016. fastcompany.com/3064845/human-lessonsfrom-brands-using-chatbots (accessed April 2018).

Ulanoff, L. "Mark Zuckerberg's AI Is Already Making Him Toast." *Mashable*, July 22, 2016.

فصل ١٣

Ashton, K. *How to Fly a Horse: The Secret History of Creation, Invention and Discovery*. New York City, NY: Doubleday, January 2015.

Bhaskar, R., and J. Dias "How a Digital Factory Can Transform Company Culture." *McKinsey & Company*, September 2017.

Bordo, M. "Israeli Air Force Works on Battlefield IoT Technology." *ReadWrite.com*, June 21, 2016.

Bray, E. "Are Consumers, Automakers and Insurers Really for Self-Driving Cars?" *Tech Crunch*, August 10, 2016.

Brokaw, L. "Six Lessons from Amsterdam's Smart City Initiative." *MIT Sloan Management Review*, May 25, 2016.

Bughin, J., M. Chui, and J. Manyika. "An Executive's Guide to the Internet of Things." *McKinsey Quarterly*, August 2015.

Bui, T. "To Succeed in IoT, Hire a Chief Data Officer." *Tech Crunch*, July 11, 2016.

Burkacky, O., et al. "Rethinking Car Software and Electronics Architecture." *McKinsey & Company*, February 2018.

Burt, J. "IoT to Have Growing Impact on Businesses, Industries, Survey Finds." *eWeek*, May 4, 2016.

Chui, M., et al. "What It Takes to Get an Edge in the Internet of Things?" *McKinsey Quarterly*, September 2018.

Coumau, J., et al. "A Smart Home Is Where the Bot Is." *McKinsey Quarterly*, January 2017.

Deichmann, J., M. Roggendorf, and D. Wee. "Preparing IT Systems and Organizations for the Internet of Things." *McKinsey & Company*, November 2015.

Diaz, J. "CES 2017: LG's New Smart Fridge Is Powered by Alexa." *Android Headlines*, January 4, 2017. androidheadlines.com/2017/01/ces-2017-lgs-new-smart-fridgepowered-alexa.html/ (accessed August 2018).

Donaldson, J. "Is the Role of RFID in the Internet of Things Being Underestimated?" *Mojix*, May 2, 2017.

Durrios, J. "Four Ways IoT Is Driving Marketing Attribution." *Enterprise Innovation*, April 8, 2017.

Editors. "Smart Cities Will Use 1.6B Connected Things in 2016." *eGov Innovation*, December 22, 2015.

Editors. "Global Smart Cities IoT Technology Revenues to Exceed US\$60 Billion by 2026." *Enterprise Innovation*, January 23, 2018.

Estopace, E. "French National Railway Operator Taps IoT for Rail Safety." *eGov Innovation*, February 21, 2017a.

Estopace, E. "Consortium to Build a Smart Mobility System for Hong Kong." *Enterprise Innovation*, March 26, 2017b.

Fenwick, N. "IoT Devices Are Exploding on the Market." *Information Management*, January 19, 2016.

Fitzgerald, M. "Data-Driven City Management: A Close Look at Amsterdam's Smart City Initiative." *MIT Sloan Management Review*, May 19, 2016.

Freeman, M. "Connected Cars: The Long Road to Autonomous Vehicles." *San Diego Union Tribune*, April 3, 2017.

Gemelli, M. "Smart Sensors Fulfilling the Promise of the IoT." *Sensors Magazine*, October 13, 2017.

Greengard, S. "How AI Will Impact the Global Economy." *CIO Insight*, October 7, 2016.

Hamblen, M. "Smart City Tech Connects Cars and Bikes with Big Data at MCW: Innovators Can Put Air Quality Sensors on Bicycles, While Wireless Connections Help Pave the Way for Driverless Cars." *Computerworld*, February 22, 2016.

Hawkins, A. "Intel Is Working with Waymo to Build Fully Self-Driving Cars." *The Verge*, September 18, 2017.

Hedge, Z. "Case Study: Athens International Airport Uses EXM and Libelium's IoT Platform to Enhance Environmental Monitoring." *IoT Now*, September 1, 2017.

Henderson, P. "10 Ways Analytics Can Make Your City Smarter." *InfoWorld and SAS Report AST = 0182248*, June 6, 2017.

Hu, F. *Security and Privacy in Internet of Things (IoTs): Models, Algorithms, and Implementations*. Boca Raton, FL: CRC Press, 2016.

Hupfer, S. "AI Is the Future of IoT." *IBM Blog*, December 15, 2016. ibm.com/blogs/internet-of-things/ai-future-iot/ (accessed July 2018).

IBM. "Embracing the Internet of Things in the New Era of Cognitive Buildings." White Paper. *IBM Global Business Services*, 2016.

Jamthe, S. *The Internet of Things Business Primer*. Stanford, CA: Sudha Jamthe, 2015.

Jamthe, S. *IoT Disruptions 2020: Getting to the Connected World of 2020 with Deep Learning IoT*. Seattle, WA: Create Space Independent Publishing Platform, 2016.

Kastrenakes, J. "Nest Can Now Use Your Phone to Tell When You've Left the House." *The Verge*, March 10, 2016. theverge.com/2016/3/10/11188888/nest-nowuses-location-for-home-away-states-launches-familyaccounts (accessed April 2018).

Khoury, A. "You Can Now Hail a Ride in a Fully Autonomous Vehicle, Courtesy of Waymo." *Digital Trends*, February 17, 2018.

Korosec, K. "Toyota Is Using Nvidia's Supercomputer to Bring Autonomous Driving to the Masses." *The Verge*, May 10, 2017.

Koufopoulos, J. "9 Examples of the Internet of Things That Aren't Nest." *Percolate*, January 23, 2015.

Kvitka, C. "Navigate the Internet of Things." January/February 2014. oracle.com/technetwork/issue-archive/2014/14-jan/o14interview-utzschneider-2074127.html (accessed April 2018).

Lacey, K. "Higher Ed Prepares for the Internet of Things." *University Business*, July 27, 2016. universitybusiness.com/article/higher-prepares-internet-things (accessed April 2018).

Libelium. "Smart Factory: Reducing Maintenance Costs and Ensuring Quality in the Manufacturing Process." *Libelium World*, March 2, 2015. technology.ihs.com/531114/theinternet-of-everything-needs-a-fabric (accessed April 2018).

Manyika, J., M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin, and D. Aharon. "Unlocking the Potential of the Internet of Things." *McKinsey Global Institute*, June 2015.

Marcus, J. "CNH Industrial Halves Product Downtime with IoT." *Product Lifecycle Report*, May 6, 2015.

Martin, E. "AI May Have Your Health and Finances on Record Before the Year Is Out." *FutureFive*, July 20, 2017. futurefive.co.nz/story/five-ways-ai-machine-will-affect-your-lifeand-business-year/ (accessed April 2018).

McCafferty, D. "How the Internet of Things Is Changing Everything." *Baseline*, June 16, 2015.

McGrath, J. "Haier Wants You to Live Smaller and Smarter with Its New Appliances." *Digital Trends*, January 5, 2016. digitaltrends.com/home/haier-shows-off-u-smartappliances-at-ces-2016 (accessed April 2018).

McLellan, C. "Internet of Things in the Enterprise: The State of Play." *ZDNet.com*, February 1, 2017a. zdnet.com/article/enterprise-iiot-in-2017-the-state-of-play/ (accessed April 2018).

McLellan, C. "Cybersecurity in an IoT and Mobile World." Special Report. *ZDNet*, June 1, 2017b.

Meola, A. "What Is the Internet of Things (IoT)? Meaning & Definition." *Business Insider*, May 10, 2018.

Miller, M. *The Internet of Things: How Smart TVs, Smart Cars, Smart Homes, and Smart Cities Are Changing the World*. Indianapolis, IN: Que Publishing, 2015.

Miller, R. "IoT Devices Could Be Next Customer Data Frontier." *TechCrunch*, March 30, 2018.

Morris, C. "Ordinary Home Appliances Are About to Get Really Sexy." *Fortune.com*, January 6, 2016. fortune.com/2016/01/06/home-appliances-ces-2016 (accessed April 2018).

Morris, S., D. Griffin, and P. Gower. "Barclays Puts in Sensors to See Which Bankers Are Their Desks." *Bloomberg*, August 18, 2017.

Murray, M. "Intel Lays Out Its Vision for a Fully Connected World." *PC Magazine*, August 16, 2016.

Ohnsman, A. "Our Driverless Future Begins as Waymo Transitions to Robot-Only Chauffeurs." *Forbes*, November 7, 2017.

Park, H. "The Connected Customer: The Why Behind the Internet of Things." *Blue Hill Research*. White Paper. January 2017.

Perkins, E. "Securing the Internet of Things." Report G00300281. *Gartner Inc.*, May 12, 2016.

Pitsker, K. "Put Smart Home Technologies to Work for You." *Kiplinger's Personal Finance*, October 2017.

PTC, Inc. "Internal Transformation for IoT Business Model Reshapes Connected Industrial Vehicles." *PTC Transformational Case Study*, November 12, 2015. ptc.com/~media/Files/PDFs/IoT/J6081_CNH_Industrial_Case_Study_Final_11-12-15.pdf?la=e (accessed April 2018).

Pujari, A. "Becoming a Smarter Manufacturer: How IoT Revolutionizes the Factory." *Enterprise Innovation*, June 5, 2017.

Rainie, L., and J. Anderson. "The Internet of Things Connectivity Binge: What Are the Implications?" *PewInternet.com*, June 6, 2017.

SAS. "SAS Analytics for IoT: Smart Cities." *SAS White Paper 108482_G14942*, September 2016.

SAS. "5 Steps for Turning Industrial IoT Data into a Competitive Advantage." *SAS White Paper 108670_G456z_0117.pdf*, January 2017.

Scannell, B. "High Performance Inertial Sensors Propelling the Internet of Moving Things." Technical Article. *Analog Devices*, 2017.

Schwartz, S. *Street Smart: The Rise of Cities and the Fall of Cars*. Kindle Edition. New York, NY: Public Affairs, 2015.

Shah, S. "HPE, Tata to Build 'World's Largest' IoT Network in India." *Internet of Business*, February 27, 2017. internetofbusiness.com/hpe-tata-largest-iot-network-india/ (accessed April 2018).

Sharda, R., et al. *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*. 4th ed. New York, NY: Pearson, 2018.

Sinclair, B. *IoT Inc.: How Your Company Can Use the Internet of Things to Win in the Outcome Economy*. Kindle Edition, New York, NY: McGraw-Hill Education, 2017.

Solomon, S. "Israel Smart-Roads Startup Nabs Prestigious EY Journey Prize." *The Times of Israel*, October 26, 2017.

Sorkin, A. "Larry Page's Flying Taxis Now Exiting Stealth Mode." *The New York Times*, March 12, 2018.

Staff. "Study Reveals Dramatic Increase in Capabilities for IoT Services." *Information Management*, May 5, 2017.

Technavio. "Smart Sensors for the Fourth Industrial Revolution: Molding the Future of Smart Industry with Advanced Technology." *Technavio.com*, September 12, 2017.

Tokuoka, D. *Emerging Technologies: Autonomous Cars*. Raleigh, NC: Lulu.com, 2016.

Tombs, J. "Smart Factory Tech Defining the Future of Production Processes." *RCR Wireless News*, March 28, 2016.

Townsend, A. *Smart Cities: Big Data, Civic, Hackers and the Quest for a New Utopia*. New York, NY: W. W. Norton, 2013.

Twentyman, J. "Athens International Airport Turns to IoT for Environmental Monitoring." *Internet of Business*, September 4, 2017.

Venkatakrishnan, K. "Are Connected Consumers Driving Smart Homes?" *Enterprise Innovation*, May 31, 2017.

Violino, B. "19 Top Paying Internet-of-Things Jobs." *Information Management*, October 25, 2017.

Weinreich, A. "The Future of the Smart Home: Amazon, Walmart, & the Home That Shops for Itself." *Forbes*, February 1, 2018.

Weldon, D. "Steps for Getting an IoT Implementation Right." *Information Management*, October 30, 2015.

فصل ١٤

Ainsworth, M. B. (2017, October). "Artificial Intelligence for Executives." *SAS White Paper, ai20for20executives.pdf*, October 2018.

Andronic, S. (2017, September 18). "5 Ways to Use Artificial Intelligence as a Competitive Advantage." *Moonoia.com*. Anon. (2017, February 20). "Big Data and Data Sharing: Ethical Issues." *UK Data Service. ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethicalissues.pdf* (accessed July 2018).

Autor, D. H. (2016, August 15). "The Shifts—Great and Small—in Workplace Automation." *MIT Sloan Review. sloanreview.mit.edu/article/the-shifts-great-and-small-inworkplace-automation/* (accessed July 2018).

Baird, Z. et al. (2017, August). "The Evolution of Employment and Skills in the Age of AI." *McKinsey Global Institute*.

Baroudy, K., et al. (2018, March). "Unlocking Value from IoT Connectivity: Six Considerations for Choosing a Provider." *McKinsey & Company*.

Batra, G., A. Queirolo, & N. Santhanam. (2018, January). "Artificial Intelligence: The Time to Act Is Now." *McKinsey & Company*.

Bloomberg News. (2017, November 29). "Ethical Worries Are Marring Alphabet's AI Healthcare Initiative." *Information Management*.

- Bossmann, J. (2016). "Top 9 Ethical Issues in Artificial Intelligence." *World Economic Forum*.
- Botton, J. (2016, May 28). "Apple Supplier Foxconn Replaces 60,000 Humans with Robots in China." *Market Watch*.
- Brynjolfsson, E., & A. McAfee. (2016). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. Boston, MA: W.W. Norton.
- Bughin, J., B. McCarthy, & M. Chui. (2017, August 28). "A Survey of 3,000 Executives Reveals How Businesses Succeed with AI." *Harvard Business Review*.
- Burden, E. (2018, July 16). "Robots Will Bolster U.K. Growth and Create New Jobs, PwC says." *Bloomberg News*.
- Catliff, C. (2017, August 15). "Three Ways Your Business Can Leverage Artificial Intelligence." *The Globe and Mail*.
- Chapman, S. (2018, January 16). "The Robotics Trends of 2018, According to Tharsus." *Global Manufacturing*.
- Charara, S. (2018, January 4). "A Quick and Dirty Guide to Ambient Computing (and Who Is Winning So Far)." *Theambient.com*.
- Chui, M., K. George, & M. Miremadi. (2017, July). "ACEO Action Plan for Workplace Automation." *McKinsey Quarterly*.
- Chui, M., J. Manyika, & M. Miremadi. (2015, November). "Four Fundamentals of Workplace Automation." *McKinsey Quarterly*.
- Chui, M., J. Manyika, & M. Miremadi. (2016, July). "Where Machines Could Replace Humans—and Where They Can't (Yet)." *McKinsey Quarterly*.
- Civin, D. (2018, May 21). "Explainable AI Could Reduce the Impact of Biased Algorithms." *Ventura Beat*.
- Clozel, L. (2017, June 30). "Is Your AI Racist? This Lawmaker Wants to Know." *American Banker*.
- Cokins, G. (2017, March 22). "Opinion Could IBM's New Deep Learning Service Tool Help Save IT Jobs?" *Information Management*.
- Collins, T. (2017, December 18). "Google and Amazon Really DO Want to Spy on You: Patent Reveals Future Version of Their Voice Assistants Will Record Your Conversations to Sell You Products." *Daily Mail*.
- Crespo, M. (2017, July 31). "The Future of Work in the Era of Artificial Intelligence." *Equal Times*.
- Crosman, P. (2017, August 17). "Why Cybercriminals Like AI As Much As Cyberdefenders Do." *American Banker*.
- Daugherty, P. R., & J. Wilson. (2018). *Human + Machine: Reimagining Work in the Age of AI*. Boston, MA: Business Review Press.
- Desjardins, J. (2017, August 21). "Visualizing the Massive \$15.7 Trillion Impact of AI." *Visual Capitalist*.
- de Vos, B. (2018, July 11). "Opinion: These 3 Business Functions Will Be the First to Benefit from Artificial Intelligence." *Information Management*.
- DiCamillo, N. (2018, July 12). "Morgan Stanley Draws from 'Hundreds of Conversations' with Experts to Build Its AI." *American Banker*.
- Dickson, B. (2017, July 28). "What Is the Future of Artificial Intelligence?" *Tech Talk*.
- Donahue, L. "A Primer on Using Artificial Intelligence in the Legal Profession." *Jolt Digest*, January 3, 2018.
- Dormehl, L. (2017). *Thinking Machines: The Quest for Artificial Intelligence—and Where It's Taking Us Next*. New York, NY: TarcherPerigee.

- Eadicicco, L. (2017, October 13). "Google Searches for Its Voice." *Time for Kids*.
- Editors. (2018, July 12). "Smart Solutions Can Help ASEAN Cities Improve Quality-of-Life Indicators by 10-30%." *eGov Innovation*.
- Egan, M. (2015, May 13). "Robots Threaten These 8 Jobs." CNNMoney.com.
- Ekster, G. (2015). Driving Investment Performance with Alternative Data. [integrity-research.com/wp-content/uploads/2015/11/Driving-Investment-Performance-With-Alternative-Data.pdf](https://www.integrity-research.com/wp-content/uploads/2015/11/Driving-Investment-Performance-With-Alternative-Data.pdf) (accessed July 2018).
- Elgan, M. (2017, April 29). "How the Amazon Echo Look Improves Privacy?" *Computer World*.
- Elson, R. J., & LeClerc, R. (2005). Security and Privacy Concerns in the Data Warehouse Environment. *Business Intelligence Journal*, 10(3), 51.
- Gaudin, S. (2016, October 26). "1-800-Flowers Wants to Transform Its Business with A.I." *Computer World*.
- Gershgorn, D. (2017, October 22). "Google Is Teaching Its AI How Humans Hug, Cook and Fight." *Quartz*. qz.com/1108090/google-is-teaching-its-ai-how-humans-hug-cook-and-fight/ (accessed April 2018).
- Goldman, S. (2018, March 22). "The Ethics of Legal Analytics." Law.com.
- Guha, A. (2017, June 5). "Labour and Artificial Intelligence: Visions of Despair, Hope and Liberation." Hindustan Times.com.
- Himmelreich, J. (2018, March 27). "The Ethical Challenges Self-Driving Cars Will Face Every Day." *Smithsonian*.
- Hu, F. (2016). *Security and Privacy in Internet of Things (IoT): Models, Algorithms, and Implementations*. Boca Raton, FL: CRC Press.
- Huff, E. (2017, January 17). "Proof That Amazon Devices Are Spies in Your Own Home: Alexa Automatically Orders Product after 'Hearing' Audio in Private Homes." *Natural News*.
- Kahn, J. (2017, November 29). "Legal AI Gains Traction as U.K. Startup Targets U.S." *Bloomberg Technology*.
- Kaplan, J. (2017). *Startup Targets. Artificial Intelligence: What Everyone Needs to Know*. London, United Kingdom: Oxford University Press.
- Kassner, M. (2017, January 2). "5 Ethics Principles Big Data Analysts Must Follow." *Tech Republic*.
- Keenan, J. (2018, February 13). "1-800-Flowers.com Using Technology to Win Customers' Hearts This Valentine's Day." *Total Retail*.
- Kelly, H. (2018, January 29). "Robots Could Kill Many Las Vegas Jobs." Money.CNN.com.
- Kiron, D. (2017, January 25). "What Managers Need to Know About Artificial Intelligence." *MIT Sloan Management Review*.
- Knight, W. (2018, March 7). "Inside the Chinese Lab That Plans to Rewire the World with AI." *MIT Technology Review*.
- Kokalitcheva, K. (2017, May 9). "The Full History of the Uber-Waymo Legal Fight." *Axio*.
- Konomi, S., & G. Roussos (ed.). (2016). *Enriching Urban Spaces with Ambient Computing, the Internet of Things, and Smart City Design (Advances in Human and Social Aspects of Technology)*. Hershey, PA: GI Global.
- Korolov, M. (2016, December 2). "There Will Still Be Plenty of Work to Go Around So Job Prospects Should Remain Good." *IT World*.
- Kottasova, I. (2018, April 12). "Experts Warn Europe: Don't Grant Rights." Money.CNN.com.

- Kovach, S. (2018, January). "Amazon Has Created a New Computing Platform That Will Future-Proof Your Home." *Business Insider*. [businessinsider.com/amazon-alexabest-way-future-proof-smart-home-2018-1/](https://www.businessinsider.com/amazon-alexabest-way-future-proof-smart-home-2018-1/) (Accessed July 2018).
- Krauth, O. (2018, January 23). "Robot Gender Gap: Women Will Lose More Jobs Due to Automation Than Men, WEF Finds." *Tech Republic*.
- Krigsman, M. (2017, January 30). "Artificial Intelligence: Legal, Ethical, and Policy Issues." *ZDNet*.
- Kurzer, R. (2017, December 21). "What Is the Future of Artificial Intelligence?" *Martechology Today*.
- Lashinsky, A. (2018, June 21). "Alibaba v. Tencent: The Battle for Supremacy in China." *Fortune*.
- Lawson, K. (2017, May 2). "Do You Need a Chief Artificial Intelligence Officer?" *Information Management*.
- Leggatt, H. (2017, June 7). "Biggest Stressor in U.S. Workplace Is Fear of Losing Jobs to AI, New Tech." *Biz Report*.
- Lev-Ram, M. (2017, September 26). "Tech's Magic 8 Ball Says Embrace the Future." *Fortune*.
- Maguire, J. (2017, February 3). "Artificial Intelligence: When Will the Robots Rebel?" *Datamation*. [datamation.com/data-center/artificial-intelligence-when-will-the-robots-rebel.html](https://www.datamation.com/data-center/artificial-intelligence-when-will-the-robots-rebel.html) (accessed April 2018).
- Manyika, J. (2017, May). "Technology, Jobs, and the Future of Work." *McKinsey Global Institute*.
- Manyika, J., M. Chi, M. Miremadi, J. Bughin, K. George, P. Willmott, & M. Dewhurst. (2017, January). "Harnessing Automation for a Future That Works." *Report from the McKinsey Global Institute*. [mckinsey.com/global-themes/digitaldisruption/harnessing-automation-for-a-future-that-works/](https://www.mckinsey.com/global-themes/digitaldisruption/harnessing-automation-for-a-future-that-works/) (accessed April 2018).
- Marr, B. (2018, June 4). "Artificial Intelligence (AI) in China: The Amazing Ways Tencent Is Driving Its Adoption." *Forbes*.
- Marshall, A., & A. Davies. (2018, February 9). "The End of Waymo v. Uber Marks a New Era for Self-Driving Cars: Reality." *Wired*.
- Mason, R., F. Mason, & M. Culnan. (1995). *Ethics of Information Management*. Thousand Oaks, CA: Sage.
- McCracken, H. (2017, October 10). "How to Stop Worrying and Love the Great AI War of 2018." *Fast Company*.
- McFarland, M. (2017a, April 28). "Robots Hit the Streets—and the Streets Hit Back." *CNN Tech*.
- McFarland, M. (2017b, September 15). "Robots: Is Your Job At Risk?" *CNN News*.
- Morgan, B. (2017, June 13). "Ethics and Artificial Intelligence with IBM Watson's Rob High." *Forbes*.
- Morris, D. (2017, February 18). "Bill Gates Says Robots Should Be Taxed Like Workers." *Fortune.com*.
- Newman, D. (2018, January 16). "Top 18 Tech Trends at CES 2018." *Forbes.com*.
- Olavsrud, T. (2018, March 15). "6 Data Analytics Trends That Will Dominate 2018." *CIO*.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Publishing).
- Pakzad, R. (2018, January 21). "Ethics in Machine Learning." *Medium.com*.
- Palmer, S. (2017, February 26). "The 5 Jobs Robots Will Take First." *Shelly Palmer*.
- Perez, A. (2017, May 31). "Opinion Will AI and Machine Learning Replace the Data Scientist?" *Information Management*.

- Pham, S. (2018, February 21). "Control AI Now or Brace for Nightmare Future, Experts Warn." **Money.cnn.com** (News).
- Press, G. (2017, November 9). "10 Predictions for AI, Big Data, and Analytics in 2018." *Forbes.com*.
- Provazza, A. (2017, May 26). "Artificial Intelligence Data Privacy Issues on the Rise." *Tech Target* (News).
- Rainie, L., & J. Anderson. (2017, June 6). "The Internet of Things Connectivity Binge: What Are the Implications?" *Pew Research Center*.
- Ransbotham, S. (2016). "How Will Cognitive Technologies Affect Your Organization?" sloanreview.mit.edu/article/how-will-cognitive-technologies-affect-your-organization/ (accessed July 2018).
- Rao, A., J. Voyles, & P. Ramchandani. (2017, December 5). "Top 10 Artificial Intelligence (AI) Technology Trends for 2018." *USBlogs PwC*.
- Rayo, E. A. "AI in Law and Legal Practice – A Comprehensive View of 35 Current Applications." *Techemergence*, September 19, 2018.
- Rikert, T. (2017, September 25). "Using AI and Machine Learning to Beat the Competition." *NextWorld*. insights.nextworldcap.com/ai-machine-learning-b01946a089b2 (accessed July 2018).
- Ross, J. (2017, July 14). "The Fundamental Flaw in AI Implementation." *MIT Sloan Management Review*. sloanreview.mit.edu/article/the-fundamental-flaw-in-aiimplementation/ (accessed July 2018).
- Sage, A. et al. (2018, February 9). "Waymo Accepts \$245 Million and Uber's 'Regret' to Settle Self-Driving Car Dispute." *Reuters* (Business News).
- SAS. (n.d.). "Customer Loyalty Blossoms with Analytics." *SAS Publication*, sas.com/en_us/customers/1-800-flowers.html (accessed July 2018).
- SAS. (2018). "Artificial Intelligence for Executives." *White Paper*.
- Sharma, K. (2017, June 28). "5 Principles to Make Sure Businesses Design Responsible AI." *Fast Company*.
- Shchutskaya, V. (2017, March 20). "3 Major Problems of Artificial Intelligence Implementation into Commercial Projects." *InData Labs*. <https://indatalabs.com/blog/data-science/problems-of-artificial-intelligenceimplementation/> (accessed April 2018).
- Sherman, E. (2015, February 25). "5 White-Collar Jobs Robots Already Have Taken." *Fortune.com* fortune.com/2015/02/25/5-jobs-that-robots-already-aretaking (accessed April 2018).
- Sherman, J. (2018, October 16). "Human-Centered Design for Empathy Values and AI." *AlMed*.
- Singh, G. (2017a, September 20). "Opinion: 5 Components That Artificial Intelligence Must Have to Succeed." *HealthDataManagement*.
- Singh, S. (2017b, December 13). "By 2020, Artificial Intelligence Will Create More Jobs Than It Eliminates: Gartner." *The Economic Times* (India).
- Smith, Ms. (2018, March 12). "Ransomware: Coming to a Robot Near You Soon?" *CSO, News*.
- Smith, N. (2018, January 3). "Top 10 Trends for Analytics in 2018." *CIO Knowledge*.
- Snyder, A. (2017, September 6). "Executives Say AI Will Change Business, But Aren't Doing Much About It." *Axios.com*.
- Sommer, D. (2017, December 20). "Opinion Predictions 2018: 11 Top Trends Driving Business Intelligence." *Information Management*.
- Spangler, T. (2017, November 24). "Self-Driving Cars Programmed to Decide Who Dies in a Crash." *USA Today*.

Standage, T. (2016) "The Return of the Machinery Question." Special Report. *The Economist*. economist.com/sites/default/files/ai_mailout.pdf (accessed July 2018).

Steinberg, J. (2017, April 26). "Echo Lock: Amazon's New Alexa Device Provide Fashion Advice." *INC*.

Straus, R. (2014, May 31). "Will You Be Replaced by a Robot? We Reveal the 100 Occupations Judged Most and Least at Risk of Automation." ThisisMoney.com. thisismoney.co.uk/money/news/article-2642880/Table-700-jobsreveals-professions-likely-replaced-robots.html (accessed April 2018).

Sykes, N. (2018a, March 27). "Opinion: Edge Computing and the Future of the Data Center." *Information Management*.

Sykes, N. (2018b, January 17). "Opinion: 9 Top Trends Impacting the Data Center in 2018." *Information Management*.

Thusoo, A. (2017, September 27). "Opinion: AI Is Changing the Skills Needed of Tomorrow's Data." *Information Management*.

Uzialko, A. (2017, October 13). "AI Comes to Work: How Artificial Intelligence Will Transform Business." *Business News Daily*.

Vanian, J. (2017, July 26). "Mark Zuckerberg Argues Against Elon Musk's View of Artificial Intelligence. . . Again." *Fortune*.

Violino, B. (2017, June 27). "Artificial Intelligence Has Potential to Drive Large Profits." *Information Management*.

Violino, B. (2018, February 21). "Most Workers See Smart Robots As Aid to Their Jobs, Not Threat." *Information Management*.

WallStreetJournal.com. (2016). "What They Know." wsj.com/public/page/what-they-know-digital-privacy.html (accessed April 2018).

Welch, D. (2016, July 12). The Battle for Smart Car Data. *Bloomberg Technology*. bloomberg.com/news/articles/2016-07-12/your-car-s-been-studying-you-closelyand-everyone-wants-the-data (accessed April 2018).

Weldon, D. (2017a, May 5). "AI Seen as Great 'Equalizer' in Bringing Services to the Masses." *Information Management*.

Weldon, D. (2017a, August 11). "Majority of Workers Welcome Job Impacts of AI, Automation." *Information Management*.

Weldon, D. (2017c, July 21). "Smarter Use of Analytics Offers Top Competitive Advantage." *Information Management*.

Weldon, D. (2018, February 28). "Knowing When It's Time to Appoint a Chief Data Officer." *Information Management*.

Weldon, D. (2018, October 18) "Gartner's top 10 strategic technology trends for 2019." *Information Management*.

West, D. (2018). *The Future of Work: Robots, AI, and Automation*. Washington, DC: Brookings Institute Press.

Wilson, H. et al. (2017, March 23). "The Jobs That Artificial Intelligence Will Create." *MIT Sloan Management Review*.

Yudkowsky (2016, May 5) [youtube.com/watch?v=EUjc1WuyPT8](https://www.youtube.com/watch?v=EUjc1WuyPT8).