

ده تکنیک کلیدی داده کاوی و نحوه استفاده کسب و کارها از آنها



ده تکنیک کلیدی داده کاوی و نحوه استفاده کسب و کارها از آنها

کسب و کارها حجم غیرقابل تصویری از داده را جمع آوری و ذخیره می کنند، اما چطور همه آن داده ها تبدیل به بینشی می شوند تا آنها را در ساخت کسب و کار بهتری یاری رسانند؟ داده کاوی فرآیند غربال مقادیر انبوه داده برای شناسایی روندها یا الگوهای پنهان کسب و کار است، که بینش های تجاری تحول آفرین را میسر می سازد.

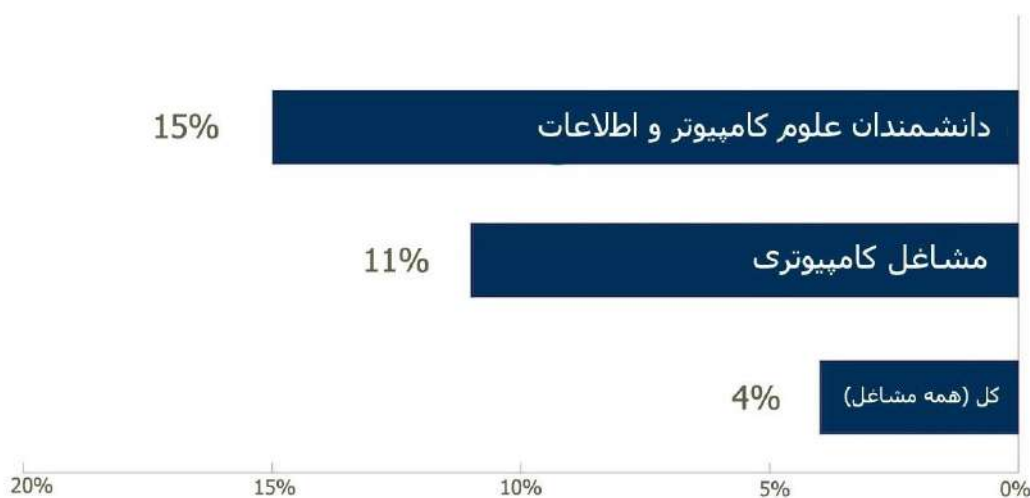
داده کاوی فناوری جدیدی نیست. به گفته هکر بیتز، ریشه های آن به دهه ۱۹۳۰ بر می گردد، اما این اصطلاح در دهه ۱۹۹۰ به طور گسترده تری مورد استفاده قرار گرفت، زیرا کسب و کارها تلاش کردند با حجم روزافزون تولید داده جامعه ما در کسب ارزش، دست و پنجه نرم کنند.

ظهور رایانه های مدرن و استفاده از تکنیک های داده کاوی به این معنا بود که کسب و کارها در نهایت می توانند مقادیر نمایی از داده ها را تجزیه و تحلیل کنند و بینش های غیر شهودی و ارزشمند را استخراج کنند. پیش بینی نتایج احتمالی کسب و کار، کاهش انواع ریسک و استفاده از فرصت های جدید شناسایی شده را فراهم می کرد.

به دلیل سودمندی آن در میان بسیاری از صنایع و نقش حیاتی آن در موفقیت کسب و کار، داده کاوی یک مسیر شغلی امیدوارکننده است. شرکت ها به دانشمندان داده ماهر در تکنیک های

استخراج نیازمندی‌ها که توانایی تعریف یافته‌هایشان از داده‌ها به روش‌های قابل فهم را داشته باشند. طبق آمار اداره آمار کار ایالات متحده، انتظار می‌رود که استخدام دانشمندان علوم کامپیوتر و اطلاعات تا سال ۲۰۲۹، ۱۵ درصد افزایش یابد.

پیش‌بینی رشد اشتغال دانشمندان علوم کامپیوتر و اطلاعات در ایالات متحده ۲۰۱۹-۲۰۲۹

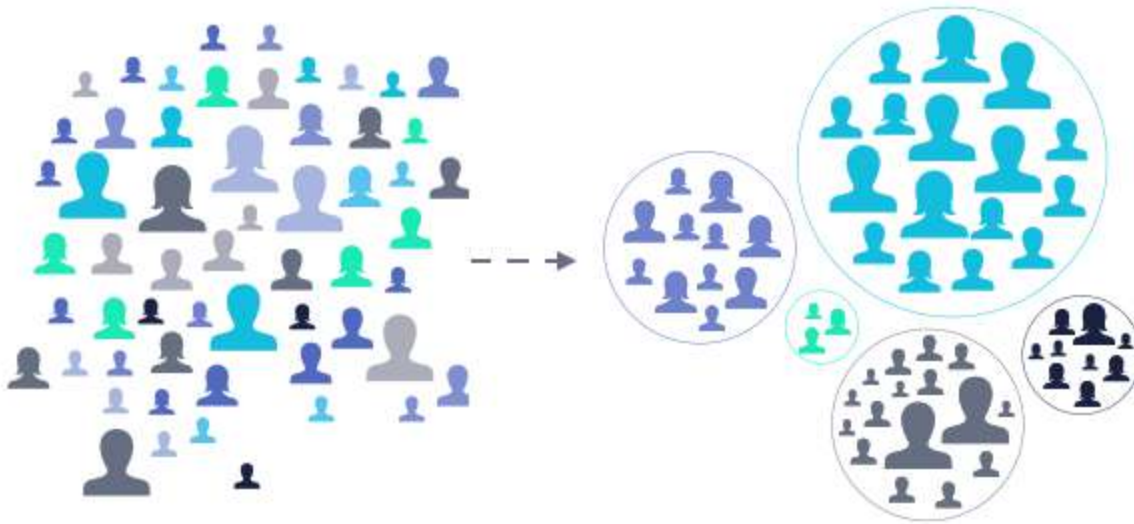


اداره آمار کار ایالات متحده، چشم‌انداز شغلی (2021)

تکنیک‌های کلیدی که داده‌کاوان مشتاق باید بدانند چیست؟ در اینجا ده تکنیک کلیدی داده‌کاوی را به تفصیل بررسی خواهیم کرد:

۱. خوشه‌بندی
۲. کشف قوانین انجمنی
۳. پاکسازی داده‌ها
۴. مصورسازی داده‌ها
۵. طبقه‌بندی
۶. یادگیری ماشینی
۷. پیش‌بینی
۸. شبکه‌های عصبی
۹. کشف داده‌های پرت و دورافتاده
۱۰. انبار داده‌ها

اخوشه بندی



خوشه بندی تکنیکی برای نمایش داده ها به صورت بصری است - مانند نمودارهایی که روند خرید یا جمعیت شناسی فروش برای یک محصول خاص را نشان می دهند.

خوشه بندی در داده کاوی چیست؟

خوشه بندی به فرآیند گروه بندی مجموعه ای از نقاط داده مختلف بر اساس ویژگی های آنها اشاره دارد. با انجام این کار، استخراج کنندگان داده می توانند به طور یک پارچه داده ها را به زیرمجموعه هایی تقسیم کنند، و امکان تصمیم گیری آگاهانه تر را از نظر جمعیت شناختی گسترده (مانند مصرف کنندگان یا کاربران) و رفتارهای نسبی شان فراهم کنند.

روش های خوشه بندی داده ها

روش خوشه بندی تفکیکی: این روش شامل تقسیم یک مجموعه داده به گروهی از خوشه های خاص برای ارزیابی بر اساس معیارهای هر خوشه به صورت جداگانه است. در این روش، نقاط داده فقط به یک گروه یا خوشه تعلق دارند.

روش سلسله مراتبی: با روش سلسله مراتبی، نقاط داده یک خوشه واحد هستند که بر اساس شباهت ها گروه بندی می شوند. سپس این خوشه های ایجاد شده جدید را می توان به صورت جداگانه از یکدیگر تجزیه و تحلیل کرد.

روش مبتنی بر چگالی: یک روش یادگیری ماشینی که در آن خوشه‌ها مناطقی با چگالی بیشتر هستند (نقاط داده متراکم‌تر) که با نواحی دارای چگالی کم‌تر (تراکم داده کم) از هم جدا شده‌اند. در این روش‌ها، نقاطی که در یک محدوده معین (یک شعاع همسایگی خاص) از هم قرار دارند در یک خوشه قرار می‌گیرند

روش مبتنی بر شبکه: دسته‌های از روش‌های مبتنی بر چگالی هستند که در آن‌ها هر منطقه مجزا در فضای داده که جست‌وجو می‌شود، در ساختار شبکه مانندی قرار می‌گیرد.

نمونه‌هایی از خوشه‌بندی در کسب و کار

خوشه‌بندی به کسب و کارها کمک می‌کند تا داده‌های خود را به طور مؤثرتری مدیریت کنند. برای مثال، خرده‌فروشان می‌توانند از مدل‌های خوشه‌بندی برای تعیین مشتریان محصولات خاص استفاده کنند، نظر به اینکه این مشتریان در چه روزهایی و با چه دفعاتی محصولات خاصی را خریداری می‌کنند. این می‌تواند به خرده‌فروشان کمک کند تا محصولات و خدمات را برای مشتریان در یک جمعیت یا منطقه خاص هدف قرار دهند.

خوشه‌بندی به فروشگاه‌های مواد غذایی کمک می‌کند تا محصولات را بر اساس ویژگی‌های مختلف (برند، اندازه، هزینه، طعم و غیره) گروه‌بندی کنند و تمایلات فروش‌شان را بهتر درک کنند. همچنین به شرکت‌های بیمه خودرو کمک می‌کند تا مشتریانانی که معمولاً مطالبات سالانه بالایی دارند را شناسایی کنند تا سیاست‌های قیمت‌گذاری مؤثرتری داشته باشند. به علاوه بانک‌ها و مؤسسات مالی ممکن است از خوشه‌بندی برای درک بهتر نحوه ارائه خدمات برای مشتریان حضوری در مقابل مجازی در برنامه‌ریزی بهتر ساعات کاری شعب و کارکنان استفاده کنند.

۲. کشف قوانین انجمنی



کشف قوانین انجمنی برای یافتن همبستگی یا ارتباط بین نقاط یک مجموعه داده استفاده می شود.

کشف قوانین انجمنی در داده کاوی چیست؟

داده کاوی ها از کشف قوانین انجمنی برای کشف روابط منحصر به فرد یا جالب بین متغیرها در پایگاه داده استفاده می کنند. انجمن پذیری اغلب برای کمک به شرکت ها در تعیین تحقیقات بازاریابی و استراتژی استفاده می شود.

روش های کشف قوانین انجمنی در داده کاوی

دو رویکرد اولیه با استفاده از کشف قوانین انجمنی در داده کاوی، روش های تک بعدی و چند بعدی هستند.

کشف قوانین انجمنی تک بعدی: این روش شامل جستجوی یک نمونه تکرار شونده از یک نقطه داده یا صفت است. به عنوان مثال، یک خرده فروش در پایگاه داده خود نمونه هایی که یک محصول خاص خریداری شده است را جستجو می کند.

کشف قوانین انجمنی چند بعدی: این روش شامل جستجوی بیش از یک نقطه داده در یک مجموعه داده است. همان خرده فروش ممکن است بخواهد اطلاعات بیشتری نسبت به جزئیات خرید مشتری یا خرید مشتری بداند - مانند سن، روش خرید (نقد یا کارت اعتباری)، یا سن.

نمونه هایی از کشف قوانین انجمنی در کسب و کار

تجزیه و تحلیل رفتار خرید نمونه ای از کشف قوانین انجمنی است - یعنی خرده فروشان در مطالعات داده متوجه می شوند که والدینی که لوازم مراقبتی کودک خرید می کنند، با احتمال بیشتری مواد غذایی یا نوشیدنی خاصی را نیز برای خود خریداری می کنند. این خریدها را می توان با کشف قوانین انجمنی تجزیه و تحلیل کرد.

تجزیه و تحلیل قوانین انجمنی کاربردهای بسیار دیگری در تجارت دارد. برای خرده فروشان در ارائه پیشنهادات خرید مفید است. به عنوان مثال، اگر مشتری یک گوشی هوشمند، تبلت یا دستگاه بازی ویدیویی بخرد، تجزیه و تحلیل قوانین انجمنی می تواند موارد مرتبط مانند کابل ها، نرم افزارهای قابل اجرا و قاب های محافظ را پیشنهاد دهد.

علاوه بر این، دولت ها می توانند از قوانین انجمنی برای به کارگیری داده های سرشماری و برنامه ریزی در خدمات عمومی استفاده کنند. همچنین توسط پزشکان برای تشخیص موثرتر بیماری ها و شرایط مختلف استفاده می شود.



پاکسازی داده ها فرآیند آماده سازی داده ها برای استخراج است.

پاکسازی داده ها در داده کاوی چیست؟

پاکسازی داده ها شامل سازماندهی داده ها، حذف داده های تکراری یا خراب، و پر کردن مقادیر از دست رفته است. هنگامی که این فرآیند کامل شد، مفیدترین اطلاعات را می توان برای تجزیه و تحلیل جمع آوری کرد.

روش های پاکسازی داده ها

تأیید داده ها: این روش شامل بررسی این موضوع است که هر رکورد از مجموعه داده در فرمت و قالب مناسب می باشد یا خیر (به عنوان مثال، شماره تلفن، شماره تامین اجتماعی).
تبدیل انواع داده: تضمین می کند که داده ها در سراسر مجموعه داده یکنواخت هستند. به عنوان مثال، متغیرهای عددی فقط شامل اعداد هستند، در حالی که متغیرهای رشته ای می توانند شامل حروف، اعداد و کاراکترها باشند.

حذف داده‌های نامرتب: این کار داده‌های بی‌فایده یا غیرقابل اجرا را پاک می‌کند، بنابراین می‌توان بر روی داده‌های ضروری تأکید کامل داشت.

حذف داده‌های تکراری: این کار با افزایش کارایی و کاهش خطاها به سرعت فرآیند استخراج کمک می‌کند.

حذف خطاها: اشتباهات تایپی، غلط‌های املائی و خطاهای ورودی را که می‌تواند بر نتایج تحلیل تاثیر منفی بگذارد، حذف می‌کند.

تکمیل مقادیر از دست رفته: این یک مقدار تخمینی را برای همه داده‌ها ارائه می‌دهد و مقادیر از دست رفته که می‌تواند منجر به نتایج نا اریب (نا متوازن) یا نادرست شود را کاهش می‌دهد.

نمونه‌هایی از پاکسازی داده‌ها در کسب و کار

به گفته اکسپرتین، ۹۵ درصد از کسب و کارها می‌گویند که تحت تأثیر کیفیت پایین داده‌ها قرار گرفته‌اند. کار با داده‌های نادرست زمان و منابع را هدر می‌دهد، هزینه‌های تجزیه و تحلیل را افزایش می‌دهد (زیرا مدل‌ها باید تکرار شوند) و اغلب منجر به تجزیه و تحلیل نادرست و غیر قابل اعتماد می‌شود.

در نهایت، مهم نیست که مدل‌ها یا الگوریتم‌های آن‌ها چقدر عالی باشند، زمانی که داده‌هایشان نادرست، ناقص یا خراب هستند کسب و کارها متضرر می‌شوند.

۴. مصورسازی داده ها



مصورسازی داده ها تفسیر داده ها به شکل گرافیکی برای نشان دادن معنای آن برای ذینفعان کسب و کار است.

مصورسازی داده ها در داده کاوی چیست؟

داده ها را می توان به روش های بصری از طریق جداول، نمودارها، نقشه ها، طرح ها و موارد دیگر ارائه کرد. این روش اولیه ای است که در آن دانشمندان داده یافته های خود را نمایش می دهند.

روش های مصورسازی داده ها

روش های زیادی برای نمایش داده ها به صورت بصری وجود دارد. چند روش به این ترتیب است:

نمودارهای مقایسه: نمودارها و جداول روابط موجود در داده ها را بیان می کنند، مانند فروش ماهانه محصول در یک دوره یک ساله.

نقشه ها: نقشه های داده برای تجسم داده های مربوط به مکان های جغرافیایی خاص استفاده می شود. به وسیله نقشه ها می توان از داده ها برای نشان دادن تراکم جمعیت و تغییرات استفاده کرد؛ مقایسه جمعیت ایالت ها مجاور، شهرستان ها و کشورهای هم سایه؛ تشخیص

چگونگی پراکندگی جمعیت در مناطق جغرافیایی؛ و ویژگی های یک منطقه را با سایر مناطق مقایسه کرد.

نقشه های حرارتی (هیت مپ): این یک تکنیک م صور سازی رایج است که داده ها را از طریق رنگ های مختلف و سایه زدن برای نشان دادن الگوها و محدوده ها در داده ها استفاده می شود. می توان از آن برای ردیابی همه چیز استفاده کرد، از تغییرات دمای یک منطقه گرفته تا روندهای غذایی و فرهنگ عامه.

نمودارهای هیستوگرام: هیستوگرام متغیر را به دسته هایی تقسیم می کند، نقاط داده ای را در هر دسته می شمارد و دسته ها را روی محور X نمایش داده و تعداد نقاط را روی محور Y نشان می دهد.

نمودارهای چگالی: نمودار چگالی نسخه هموار و پیوسته ای از هیستوگرام است که از روی داده ها تخمین زده می شود. نمودارهای چگالی نمایش نقاط منفرد داده ای در طول زمان را آسان می کند (به عنوان مثال، ماه، سال، دهه).

نمودارهای شبکه: این نمودارها نحوه ارتباط نقاط داده با یکدیگر را با استفاده از یک سری خطوط (یا پیوندها) برای اتصال اشیا به یکدیگر نشان می دهند.

نمودارهای پراکندگی: این نمودارها روابط نقطه داده را بر روی یک محور دو متغیره نشان می دهند. از نمودارهای پراکنده می توان برای مقایسه متغیرهای منحصر به فرد مانند امید به زندگی یک کشور یا مقدار پولی که سالانه صرف مراقبت های بهداشتی می شود، استفاده کرد.

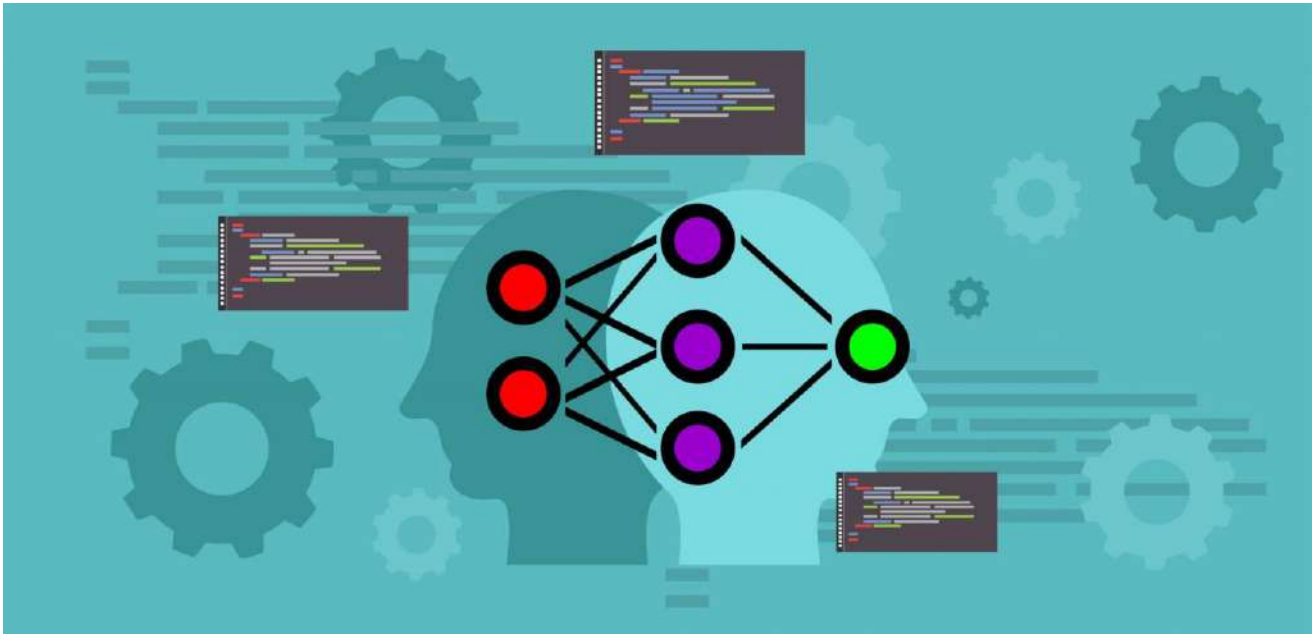
نمونه هایی از مصورسازی داده ها در کسب و کار

نمایش داده ها به صورت بصری مهارتی مهم است زیرا درک راحت تر داده ها برای مدیران، کارمندان و مشتریان میسر می سازد. بر اساس گزارش مارکت اند مارکت، انتظار می رود اندازه بازار برای ابزارهای مصورسازی داده های جهانی تا سال ۲۰۲۶ تقریباً دو برابر شود (به ۱۰/۲ میلیارد دلار).

شرکت ها زمانی که از داده ها استفاده می کنند چون درک و تفسیرشان آسان تر است، می توانند سریع تر و آگاهانه تر تصمیم بگیرند. امروزه، این امر معمولاً از طریق رسانه های مؤثر و قابل دسترسی بصری مانند نمودارها، مدل های سه بعدی و حتی واقعیت افزوده انجام می شود. در

نتیجه، برای متخصصان مشتاق داده ایده خوبی است که یادگیری چنین مهارت هایی را در نظر بگیرند.

۵. طبقه بندی



طبقه بندی یک تکنیک اساسی در داده کاوی است و تقریباً در هر صنعتی قابل اعمال است. فرآیندی است که در آن نقاط داده از مجموعه داده های بزرگ بر اساس نحوه استفاده از آنها در دسته ها ارجاع داده می شوند.

طبقه بندی در داده کاوی چیست؟

در داده کاوی، طبقه بندی برای استخراج نقاط قابل مقایسه داده ها و برای تجزیه و تحلیل های مقایسه ای مفید است. طبقه بندی همچنین برای تعیین گروه های گسترده در یک پایگاه جمعیتی، مخاطبان هدف یا کاربران استفاده می شود که از طریق آنها کسب و کارها می توانند درک قوی تری به دست آورند.

روش های طبقه بندی داده کاوی

رگرسیون لجستیک: رگرسیون لجستیک تکنیکی آماری است برای نشان دادن تاثیر متغیرهای کمی یا کیفی بر متغیر وابسته دو مقداری (دو طبقه ای).

به عنوان مثال، یک سرویس ایمیل می تواند از رگرسیون لجستیک برای پیش بینی هزینه بودن یا نبودن یک ایمیل استفاده کند.

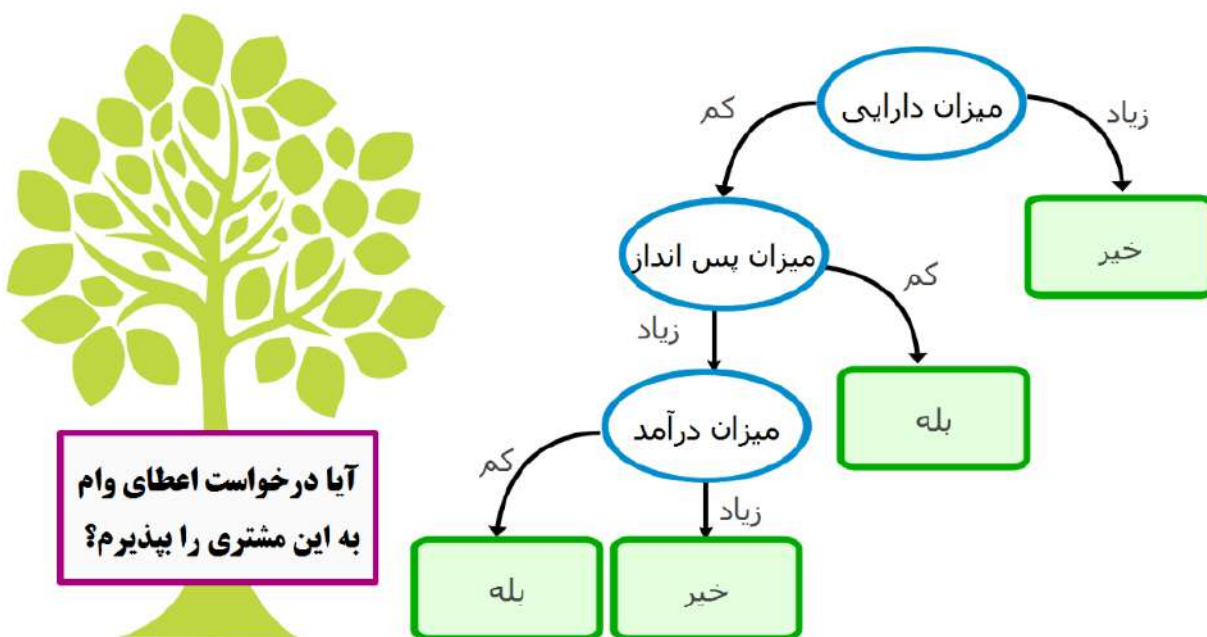
آدرس سایت: modamcrm.com

شماره تماس: ۰۲۱-۲۸۴۲۱۶۰۸

شماره تماس و واتس آپ: ۰۹۰۳۹۱۵۱۰۷۵

درختان تصمیم: پس از طبقه بندی داده ها، می توان نتایج را در نموداری به نام درخت تصمیم ترسیم کرد. درخت های تصمیم نوع بسیار مهمی از الگوریتم های داده کاوی هستند که برای مدل سازی نظارت شده مورد استفاده قرار می گیرند.

به عنوان مثال، اگر مسؤل اعتبارات یک بانک بخواهد در مورد اعطای وام به مشتری جدید تصمیم گیری کند، احتمالاً این سوال را خواهد پرسید که به مشتری جدید وام اعطا شود یا خیر؟ در شکل زیر یک درخت تصمیم نشان داده شده است. همان طور که مشاهده می شود، داده ها در درخت های تصمیم به صورت «بله» و «خیر» طبقه بندی می شوند و سپس سؤالات دیگری نیز به روشی مشابه پرسیده می شوند.



کا-نزدیکترین همسایه: این الگوریتمی است که سعی دارد یک شی ناشناخته را با مقایسه آن با دیگران شناسایی کند. برای مثال، زنجیره های خواربارفروشی ممکن است از الگوریتم کا نزدیکترین همسایه ها برای تصمیم گیری در مورد گنجاندن یک ایستگاه غذای گرم در چیدمان فروشگاه جدیدش بر اساس عادات مصرف کننده در بازار محلی استفاده کنند.

بیز ساده: بر اساس قضیه احتمال بیز این الگوریتم از داده های تاریخی برای پیش بینی اینکه آیا رویدادهای مشابه بر اساس مجموعه ای متفاوت از داده ها رخ خواهند داد یا خیر، استفاده می کند.

ماشین بردار پشتیبان: این الگوریتم یادگیری ماشینی اغلب برای تعریف خطی استفاده می شود که به بهترین نحو یک مجموعه داده را به دو کلاس تقسیم کند SVM. می تواند به طبقه بندی تصاویر کمک کند و در نرم افزارهای تشخیص چهره و دست خط استفاده می شود.

نمونه هایی از طبقه بندی در کسب و کار

مؤسسات مالی مصرف کنندگان را بر اساس متغیرهای زیادی برای بازاریابی وام های جدید یا پروژه ریسک کارت اعتباری طبقه بندی می کنند. در همین حال، برنامه های هواشناسی داده ها را برای پیش بینی کل بارش برف و سایر فرم های مشابه طبقه بندی می کنند. فرو شگاه های مواد غذایی زمانی که از روش طبقه بندی برای گروه بندی محصولات توسط مصرف کنندگانی که آنها را خریداری می کنند استفاده می کنند در واقع به پیش بینی الگوهای خرید کمک می کنند.

۶. یادگیری ماشینی



یادگیری ماشینی فرآیندی است که در آن کامپیوترها از الگوریتم ها برای یادگیری استفاده می کنند. یادگیری ماشینی علمی است که باعث می شود رایانه ها بدون نیاز به یک برنامه صریح در مورد یک موضوع خاص یاد بگیرند. به عنوان زیر مجموعه ای از هوش مصنوعی، الگوریتم های

یادگیری ماشینی یک مدل ریاضی بر اساس داده‌های نمونه یا "داده‌های آموزش" به منظور پیش‌بینی یا تصمیم‌گیری بدون برنامه‌ریزی آشکار، ایجاد می‌کنند.

یادگیری ماشینی در داده کاوی چیست؟

در داده کاوی، کاربردهای یادگیری ماشین بسیار گسترده است. یادگیری ماشین و داده کاوی زیر چتر علم داده قرار می‌گیرند اما اصطلاحات قابل تعویض نیستند. به عنوان مثال، رایانه‌ها داده کاوی را به عنوان بخشی از عملکردهای یادگیری ماشین خود انجام می‌دهند.

روش‌هایی برای یادگیری ماشینی

یادگیری با نظارت: در این روش، الگوریتم‌ها ماشین‌ها را آموزش می‌دهند تا با استفاده از داده‌های از پیش برچسب‌گذاری شده با مقادیر صحیح، که ماشین‌ها آن‌ها را طبقه‌بندی کردند، یاد بگیرند. این روش‌ها تحت نظارت (با ناظر) نامیده می‌شود زیرا این فرآیند رایانه‌ها را برای طبقه‌بندی داده‌ها و پیش‌بینی نتایج آموزش می‌دهد یا "نظارت" می‌کند. یادگیری ماشینی با نظارت در طبقه‌بندی داده کاوی به کار می‌رود.

یادگیری بدون نظارت: هنگامی که رایانه‌ها داده‌های بدون برچسب را مدیریت می‌کنند، وارد یادگیری بدون ناظر می‌شوند. در این حالت، کامپیوتر خودش داده‌ها را طبقه‌بندی می‌کند و سپس به دنبال الگوها می‌گردد. مدل‌های بدون نظارت برای انجام خوشه‌بندی و انجمن‌پذیری استفاده می‌شوند.

یادگیری نیمه نظارتی: یادگیری نیمه نظارتی از ترکیبی داده‌های برچسب دار و بدون برچسب استفاده می‌کند و آن را ترکیبی از مدل‌های فوق می‌کند.

یادگیری تقویتی: این یک فرآیند لایه‌ای تر است که در آن رایانه‌ها تصمیم‌گیری را بر اساس بررسی داده‌ها در یک محیط خاص یاد می‌گیرند. به عنوان مثال، یک کامپیوتر ممکن است با بررسی داده‌های هزاران بازی آنلاین بازی شطرنج را بیاموزد.

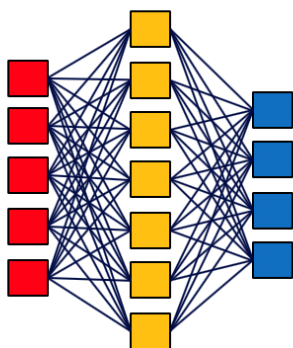
نمونه‌هایی از یادگیری ماشین در کسب و کار

با یادگیری ماشینی، شرکت‌ها می‌توانند از رایانه‌ها برای شناسایی سریع انواع الگوهای داده (در فروش، استفاده از محصول، عادت‌های خرید و غیره) استفاده کنند و با استفاده از این دید، طرح‌های تجاری توسعه دهند. این یک نیاز رو به رشد در بسیاری از صنایع است.

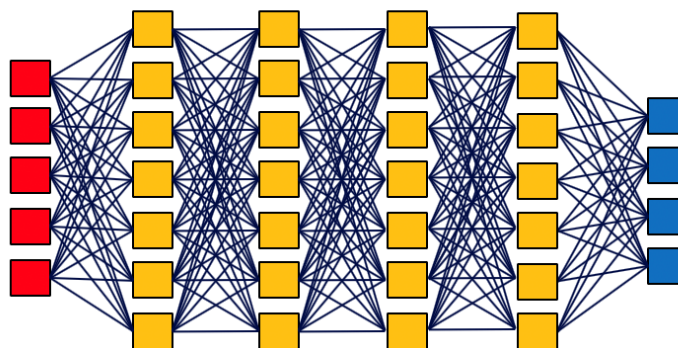
بر اساس تحقیقات میکرو استراتژی (MicroStrategy)، ۱۸ درصد از متخصصان تحلیلگر گفتند که یادگیری ماشین و هوش مصنوعی بیشترین تأثیر را بر استراتژی های آنها در پنج سال آینده خواهند داشت. بنابراین یادگیری موضوعات پیشرفته تر مانند یادگیری ماشین برای دانشمندان داده ضروری است.

شبکه های عصبی

Simple Neural Network



Deep Neural Network



■ Input layer

■ Hidden layer

■ Output layer

رایانه ها حجم زیادی داده را بسیار سریعتر از مغز انسان پردازش می کنند، اما هنوز ظرفیت اعمال عقل و تخیل را در کار با داده ها ندارند. شبکه های عصبی یکی از راه هایی است که به کامپیوترها کمک می کند بیشتر شبیه انسان ها استدلال کنند.

شبکه های عصبی در داده کاوی چیست؟

شبکه های عصبی مصنوعی در تلاشند تا به صورت دیجیتالی شیوه عملکرد مغز انسان را تقلید کنند. شبکه های عصبی بسیاری از پردازنده های رایانه ای (مشابه روشی که مغز از نورون ها استفاده می کند) را برای پردازش داده ها، تصمیم گیری و یادگیری آن گونه که یک انسان انجام می دهد - یا حداقل تا حد امکان شبیه به آن، ترکیب می کند.

روش های شبکه عصبی

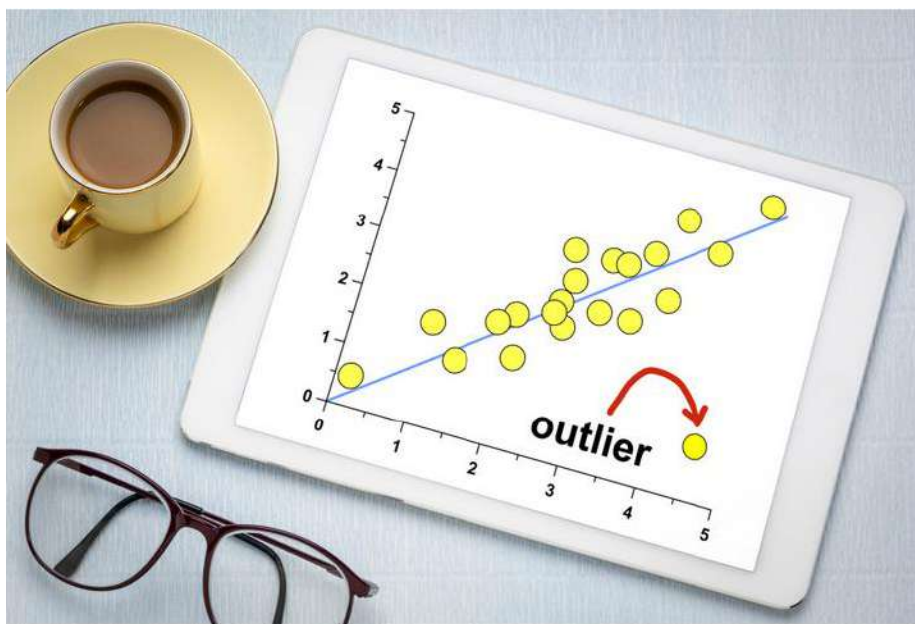
شبکه های عصبی از سه لایه اصلی تشکیل شده اند: ورودی، پنهان و خروجی. داده ها از طریق لایه ورودی وارد می شوند، در لایه پنهان پردازش می شوند و در لایه خروجی اسقرار پیدا می کنند،

جایی که هر اقدام مرتبط بر اساس داده ها انجام می شود. لایه پنهان می تواند از لایه های پردازشی زیادی تشکیل شده باشد که بسته به میزان داده های مورد استفاده و یادگیری دارد. یادگیری با ناظر و بدون ناظر نیز برای شبکه های عصبی اعمال می شود. شبکه های عصبی از این نوع الگوریتمها برای «آموزش» خود برای عملکردی مشابه مغز انسان استفاده می کنند.

نمونه هایی از شبکه های عصبی در کسب و کار

شبکه های عصبی کاربردهای گسترده ای دارند. آنها می توانند به کسب و کارها کمک کنند تا الگوهای خرید مصرف کننده را پیش بینی کنند و کمپین های بازاریابی را بر روی جمعیت شناسی خاص متمرکز کنند. آنها همچنین می توانند به خرده فروشان کمک کنند تا پیش بینی های دقیق فروش را انجام دهند و نحوه استفاده از قیمت گذاری پویا را درک کنند. علاوه بر این، آنها به بهبود روش های تشخیصی و درمانی در مراقبت های بهداشتی، اصلاح مراقبت و عملکرد کمک می کنند.

۸. تشخیص داده های پرت



تشخیص داده های پرت یک جزء کلیدی در حفظ پایگاه داده ایمن است. شرکت ها از آن برای آزمایش تراکنش های تقلبی، مانند استفاده غیرعادی از کارت اعتباری که احتمالاً نشان از سرقت است، استفاده می کنند.

تشخیص داده های پرت در داده کاوی چیست؟

در حالی که سایر روش های داده کاوی به دنبال شناسایی الگوها و روندها هستند، تشخیص داده های پرت به دنبال موارد منحصر به فرد است: نقطه یا نقاط داده ای که با بقیه متفاوت هستند یا واگرا از نمونه کلی هستند. تشخیص داده های پرت خطاها را پیدا می کند، مانند داده هایی که به اشتباه وارد شده یا از نمونه اشتباه استخراج شده اند. انحرافات طبیعی داده ها نیز می تواند آموزنده باشد.

روش هایی برای تشخیص داده های پرت

داده پرت عددی: در این روش از میانه (Median) و دامنه چارکی (Inter-quartile Range) برای شناسایی داده پرت استفاده کنیم. هر یک از این شاخص ها به ترتیب به عنوان برآورد نقطه تمرکز و پراکندگی در روش های ناپارامتری به کار می روند. بنابراین می توانیم قاعده ای برای شناسایی داده پرت به این ترتیب پیدا کنیم که اگر نقطه ای از سه برابر دامنه چارکی از چارک اول یا سوم دور باشد، آن را داده پرت بشناسیم.

نمره Z: نشان می دهد که یک نقطه داده چند انحراف استاندارد از میانگین نمونه فاصله دارد. همچنین به عنوان آنالیز مقادیر حدی نیز شناخته می شود.

روش DBSCAN: الگوریتم خوشه بندی DBSCAN می تواند گروه ها را بر اساس غلظت دسته بندی کند و این دسته ها در کنار هم خوشه ها را تشکیل می دهند. داده هایی که به هیچ گروهی (بر اساس غلظت و تراکم) متصل نیستند، توسط الگوریتم DBSCAN به عنوان یک داده ی پرت (outlier) شناسایی می شود.

نمونه هایی از تشخیص داده های پرت در کسب و کار

تقریباً هر کسب و کاری می تواند از درک ناهنجاری ها در خطوط تولید یا توزیع خود و نحوه رفع آنها بهره مند شود. خرده فروش ها می توانند از تشخیص داده های پرت استفاده کنند تا بدانند چرا فروشگاه های شان شاهد افزایش خریدهای عجیب هستند، مانند خرید بیل های برفی در تابستان، و نحوه واکنش به چنین یافته هایی.

به طور کلی، تشخیص داده های پرت برای بهبود مدیریت انتقال یک محصول از محل تولید تا محل مصرف مشتری و ایجاد یک محیط راحت تر برای مشتریان، کاربران و سایر گروه های کلیدی استفاده می شود.



مدل سازی پیش بینی به دنبال تبدیل داده ها به پیش بینی کنش یا رفتار آینده است. این مدل ها مجموعه داده ها را برای یافتن الگوها و روندها بررسی می کنند، سپس احتمالات یک برآمد در آینده را محاسبه می کنند.

پیش بینی در داده کاوی چیست؟

مدل سازی پیش بینی کننده یکی از رایج ترین کاربردهای داده کاوی است و در مجموعه داده های بزرگی که حجم نمونه وسیعی را نشان می دهند بهترین کارایی را دارد.

روش های پیش بینی

مدل سازی پیش بینی از برخی تکنیک ها و اصطلاحات مشابه سایر فرآیندهای داده کاوی استفاده می کند. در اینجا چهار مثال آورده شده است:

مدل سازی پیش بینی: (Forecast modeling) تکنیکی رایج است که در آن رایانه با تجزیه و تحلیل داده های تاریخی به سؤال پاسخ می دهد (مثلاً یک فروشگاه در روز دوشنبه چقدر شیر باید در انبار داشته باشد؟).

مدل‌سازی طبقه‌بندی (Classification modeling) این روش، داده‌ها را در گروه‌هایی قرار می‌دهد که می‌توان از آن‌ها برای پاسخ به سؤالات مستقیم استفاده کرد.

مدل‌سازی خوشه‌ای (cluster modeling) با خوشه‌بندی داده‌ها در گروه‌هایی با ویژگی‌های مشترک، می‌توان از یک مدل پیش‌بینی‌کننده برای مطالعه آن مجموعه داده‌ها و تصمیم‌گیری استفاده کرد.

مدل سازی سری زمانی (Time series modeling) این مدل داده‌ها را بر اساس زمانی که داده‌ها وارد شده است، تجزیه و تحلیل می‌کند. مطالعه روند فروش در طول یک سال نمونه‌ای از مدل سازی سری های زمانی است.

نمونه‌هایی از پیش‌بینی در کسب و کار

مدل‌سازی پیش‌بینی‌کننده یک الزام تجاری است که تقریباً هر گوشه از بخش‌های دولتی و خصوصی را تحت تأثیر قرار می‌دهد. بر اساس مایکرو استراتژی، ۵۲ درصد از کسب و کارهای جهانی، مدل‌سازی پیشرفته و پیش‌بینی‌کننده را اولویت اصلی خود در تجزیه و تحلیل در نظر می‌گیرند.

برای تعیین پیش‌بینی‌های فروش و پیش‌بینی عادات خرید مصرف‌کننده می‌توان مدل‌های پیش‌بینی‌کننده ایجاد کرد. این مدل‌ها به تولیدکنندگان کمک می‌کنند تا توزیع نیازها را پیش‌بینی کنند و برنامه‌های نگهداری را مشخص سازند.

سازمان‌های دولتی از داده‌های سرشماری برای ترسیم روند جمعیت و هزینه‌های پروژه استفاده می‌کنند در حالی که تیم‌های بیس‌بال از مدل‌های پیش‌بینی برای تعیین قراردادهای و ساخت فهرست وظایف استفاده می‌کنند.



انبارسازی داده فرآیندی است که توسط آن داده ها قبل از ارزیابی، جمع آوری و ذخیره می شوند.

انبار داده در داده کاوی چیست؟

داده کاوی داده ها را قبل از اینکه بتوان از آن در تجزیه و تحلیل کسب و کار استفاده کرد از منابع متعدد در یک آرشیو مشترک جمع آوری می کنند. این فرآیند که انبار سازی داده نامیده می شود، معمولاً قبل از فرآیند داده کاوی اتفاق می افتد.

روش های ذخیره سازی داده ها

داده ها قبل از بارگیری در انبار داده از طریق یک فرآیند سه مرحله ای به نام ETL می گذرند ETL. مخفف عبارت استخراج، تبدیل و بارگذاری (extract, transform, and load) است:

استخراج: داده ها کپی شده و از منبع خود به یک محل اسقرار انبار منتقل می شوند. داده ها می توانند ساختاری باشند (نام، تاریخ، شماره کارت اعتباری و غیره) یا بدون ساختار (عکس ها، فیلم ها، فایل های صوتی، پست های رسانه های اجتماعی).

تبدیل: در این مرحله داده ها فیلتربندی و پاکسازی می شوند - خطاها حذف می شوند و داده ها اعتبار سنجی می شوند. داده ها نیز به گونه ای قالب بندی شده اند که متناسب با انبار باشند.

بارگذاری: در مرحله آخر، داده های تبدیل شده در انبار داده آپلود می شوند. این مراحل را می توان با به روز رسانی داده ها تکرار کرد.

نمونه هایی از انبار داده ها در کسب و کار

انبارهای داده، کار با داده های کلان را آسان تر می کنند - مخصوصاً برای مشاغلی که با پایگاه های بزرگ مشتریان، گزارش های فروش و صورت حساب، و طرح های منابع سرکار دارند. از طریق انبارسازی داده ها، کسب و کارها می توانند مشتریان را از مجموعه های وسیعی تقسیم بندی کرده و هدف قرار دهند. مانند مجموعه های سفارش های فروش، جستجوی محصول، یا نام نویسی در برنامه های وفاداری. همچنین می توانند طیف گسترده ای از نقاط داده، حتی پست های رسانه های اجتماعی در مورد محصولات و مشاغل را ذخیره و تجزیه و تحلیل کنند.

انبارسازی داده همچنین منابع مختلف داده را در یک مکان ادغام می کند و استخراج و تصمیم گیری را کارآمدتر می کند و باعث صرفه جویی در وقت و هزینه کسب و کار می شود.

نتیجه گیری

کسب و کارهایی که به دنبال مزیت رقابتی هستند، اغلب داده ها را در میان بهترین منابع خود می یابند و تکنیک های داده کاوی برای به ثمر رساندن این منبع حیاتی هستند. فرآیند استخراج به کسب و کارها اجازه می دهد تا قدرت داده ها را مهار کنند، بصیرتی کسب کنند، الگوها و ناهنجاری ها را شناسایی کنند و راه هایی برای بهره وری بیشتر بیابند.

همانطور که به تولید حجم فزاینده ای از داده های متنوع ادامه می دهیم، توانایی استخراج این داده ها برای درک بهتر اهمیت فزاینده ای پیدا می کند. سازمان ها عموماً به دنبال راه های سریع تر و کارآمدتری برای کار با داده های خود، روش های بیشتری برای تجسم داده ها و سیستم های محاسباتی هستند که بتوانند تصمیمات انسانی بیشتری بگیرند.

در نتیجه، بسیاری از شرکت ها انتظار دارند سرمایه گذاری خود را در طرح های تحلیلی، که شامل داده کاوی نیز می شود، افزایش دهند. بر اساس گزارش تجزیه و تحلیل وضعیت سازمانی جهانی مایکرو ستراتیژی ۲۰۱۸، ۷۱ درصد از شرکت های جهانی می گویند که قصد دارند پول بیشتری را برای تجزیه و تحلیل هزینه کنند (۷۳ درصد از شرکت های آمریکایی قصد دارند بودجه تجزیه و تحلیل خود را افزایش دهند).