

داده کاوی (Data Mining) چیست؟ +

مراحل و کاربردهای آن

در دنیای امروزی تحولات و پیشرفت‌های چشمگیر در صنایع مختلف به کمک راهکارها و سامانه‌های دیجیتال صورت گرفته است. این سامانه‌ها از علم داده بهره گرفته‌اند و فرایندهای داده کاوی را انجام داده‌اند تا به نتایج موثر در اعمال سیاست‌های پیشرفت و توسعه کسب‌وکارها و سازمان‌ها رسیده‌اند.

داده کاوی یا Data Mining در کنار هوش مصنوعی نقش مهمی را در کسب بینش و برنامه‌ریزی‌های کسب‌وکارها بر عهده گرفته است. پدیده استخراج بینش از داده‌های خام چندان جدید نیست و قدمتی به اندازه یک قرن دارد. آلن تورینگ اولین نمونه دیتا ماینینگ را با معرفی ماشین محاسباتی خود که شبیه کامپیوتر بود در دهه 1930 میلادی انجام داد. پیش از آنکه با داده‌کاوی آشنا شویم لازم است تا تعریف مختصری از داده و انواع آن داشته باشیم.

کلیک کنید

برای دریافت دمو



داده چیست؟

داده یا data کوچک‌ترین و ساده‌ترین واحدی است که به عنوان محتوا در علم داده به کار می‌رود. هر آنچه توسط افراد و سیستم‌ها به عنوان آمار، ارقام یا گزارش جمع‌آوری می‌شود، داده است. نباید داده و اطلاعات را با هم یکی فرض کنیم. داده، محتوایی خام و تفسیر نشده است اما اطلاعات را می‌توان نتیجه یا نسخه‌ای تکامل‌یافته و پردازش‌شده از داده‌ها دانست. به عبارت دیگر داده باعث به وجود آمدن اطلاعات می‌شود. از یک سری داده می‌توان اطلاعات مختلفی به دست آورد.

انواع منابع داده

داده‌ها از منابع مختلف و متعددی در یک پایگاه داده جمع می‌شوند تا بر روی آنها عملیات مختلف صورت گیرد. داده‌ها به شکل‌های مختلفی وارد هر سیستم می‌شوند. همچنین داده‌ها ممکن است از منابع مختلفی نیز جمع‌آوری شوند. برخی از این منابع عبارتند از:

- **تراکنش‌های کسب‌وکار:** تراکنش‌های کلی هر کسب‌وکار شامل خرید و فروش، مبادلات بانکی، سهام و موارد دیگر.
- **داده‌های علمی:** این داده‌ها در فرمت‌ها و اشکال مختلف متنی، بصری یا کدنویسی، عدد و آمار و یا هر گونه دیگر از انواع اسناد در سراسر جهان تولید می‌شوند.
- **داده‌های پزشکی:** اطلاعات مربوط به سلامت و سابقه بیماری افراد و همچنین داده‌های علمی مرتبط با کشف و شناسایی زود هنگام بیماری‌ها، تجویز دارو و موارد

مختلف در قالب‌های مختلف.

- **تصاویر:** حجم زیادی از تصاویر و ویدئوهایی که گونه‌ای سند و داده هستند و نیاز به نگهداری و تحلیل دارند.
- **گزارش‌ها و اسناد متنی:** در هر زمینه‌ای هنوز هم نوشته‌ها و متون تعیین‌کننده هستند و به عنوان سند در سیستم‌های دیجیتال ثبت می‌شوند که تبدیل به داده خواهند شد.

داده‌های برخی منابع دارای پیچیدگی‌های بیشتری نسبت به بقیه منابع هستند. کسب‌وکارها در حال حاضر برای مدیریت امور خود نیازمند بررسی و ارزیابی داده‌های سازمانی خود هستند و از همین روی امکانات و برنامه‌های نرم‌افزاری متعددی به کمک هوش مصنوعی و علم داده جهت همراهی و موفقیت کسب‌وکارها طراحی و تولید شده است.



تعریف داده کاوی

علم داده شاخه‌های زیادی دارد، که داده کاوی یکی از آنهاست و به معنای کاوش اطلاعات در یک پایگاه داده بزرگ است. دو اصطلاح Data Mining و Knowledge Discovery in Data برای کاوش داده به کار می‌رود که KDD به معنای کشف دانش در داده‌هاست. این دانش یک دانش میان‌رشته‌ای محسوب می‌شود. به زبان ساده‌تر می‌توان گفت:

“داده کاوی مجموعه‌ای از روش‌هاست که بر داده‌های موجود در پایگاه داده‌های بزرگ و پیچیده اعمال می‌شود تا الگوها و روابط پنهان در داده‌ها را کشف کند.”

از زمانی که دیتا ماینینگ شکل گرفت ارزش داده‌ها مشخص شد. در دنیای امروزی نقش داده مانند نقش نفت در دوره‌های گذشته است. به همین دلیل کاوش داده با عنوان دیتا ماینینگ معرفی شد که به فرایند کشف الگوهای منظم و روندهای معنادار داده‌ها گفته می‌شود و به کمک الگوریتم‌های علم آمار و ریاضی انجام می‌شود. کشف مدل‌ها و الگوها، همبستگی در داده‌ها و ناهنجاری‌ها در حل مسائل در حوزه‌های مختلف از جمله کارهایی است که در دیتا ماینینگ انجام می‌شود.

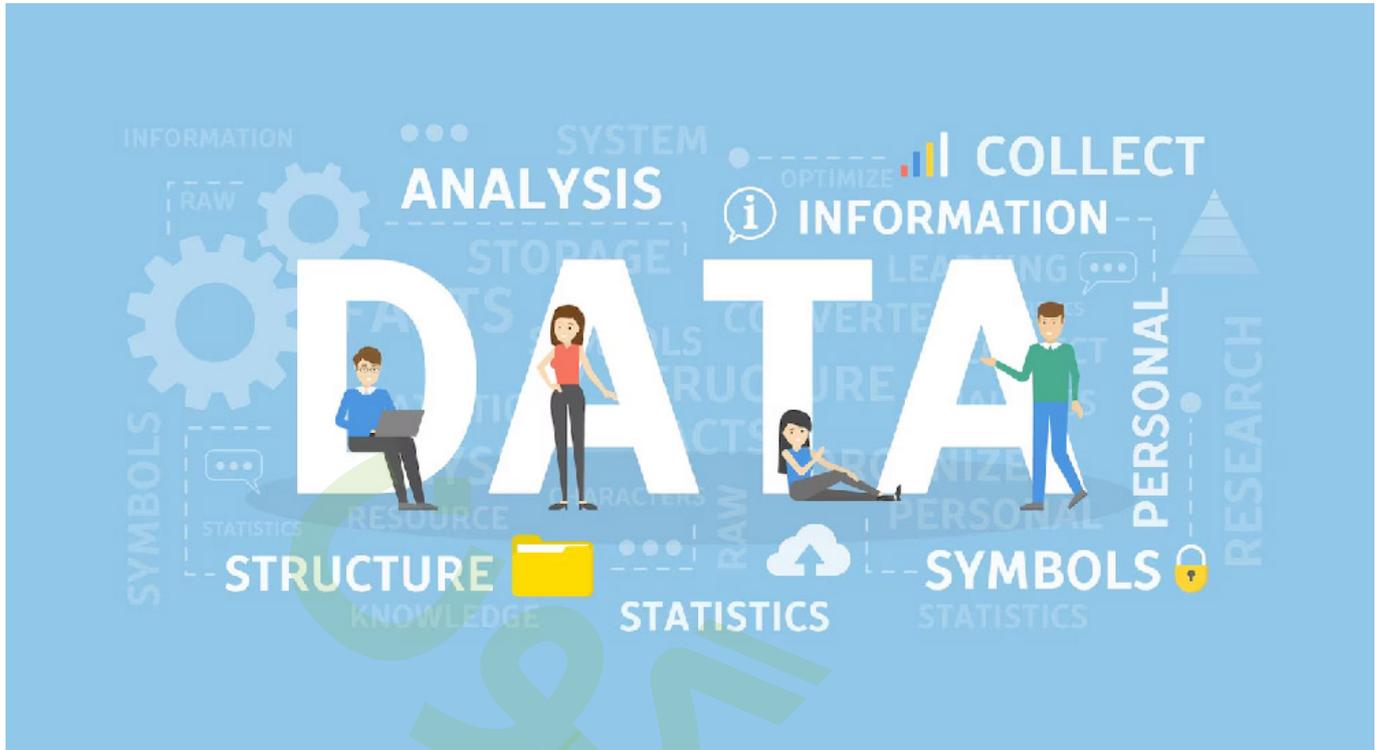
داده کاوی در سازمان‌ها

داده‌کاوی در سازمان‌ها برای حل مشکلات، پیش‌بینی روندها، کاهش خطرات و یافتن فرصت‌های نوآوری و افزایش بهره‌وری و بهینه‌سازی سیستم عملکردی به کار می‌رود.

هوش مصنوعی، یادگیری ماشین، مدیریت پایگاه داده و علم ریاضی و آمار نیز مکمل فرایندهای دیتا ماینینگ هستند.

در ابتدای مسیر شکل‌گیری و تکامل این دانش، حجم داده‌هایی که تولید می‌شد چندان زیاد نبود و ذهن انسان به راحتی قادر بود تا آنها را تحلیل کند و نتایج درستی را از بررسی و ارزیابی داده‌ها به دست آورد. اما با افزایش حجم داده‌ها نیاز به وجود یک راهکار مؤثر و هوشمند هر روز بیشتر و بیشتر احساس می‌شد. در نهایت راهکارهای دیجیتال در این زمان گسترش بیشتری یافتند و پا به عرصه گذاشتند. در حال حاضر این عملیات در ابعاد بسیار بزرگ و به کمک هوش مصنوعی انجام می‌شود، چون دیگر ذهن بشر قابلیت چنین تحلیل‌هایی را ندارد و اگر هم داشته باشد زمان پرداختن به این همه داده مدتی بسیار طولانی خواهد بود.

تحلیل‌ها و پژوهش‌های داده کاوی در شاخه‌های مختلفی چون علم آمار، علوم کامپیوتر، یادگیری ماشین، مدیریت پایگاه داده، پردازش سیگنال، بصری‌سازی داده و موارد دیگری انجام می‌شود.



تفاوت علم داده و داده کاوی

علم داده Data Science با طیف گسترده‌تری از عملیات ارتباط دارد به شکلی که دیتا ماینینگ، تجسم داده، هوش مصنوعی و یادگیری ماشین در زیر آن قرار می‌گیرند. علم داده چرخه کامل گردش داده در تمام مراحل و کلیه فرایندها از تولید، جمع‌آوری، آنالیز و تجسم داده را شامل می‌شود. هدف علم داده استخراج بینش از داده است. در حالی که دیتا ماینینگ بخشی متمرکز و مجموعه کوچکی از علم داده است و از الگوریتم‌ها و تکنیک‌های آماری برای کشف الگوها و روابط در بیگ دیتا استفاده می‌کند.

- نظارت دقیق بر عملکردهای اجرایی

- کشف الگوهای پنهان در داده‌ها

- کشف الگوهای دست اول و مطمئن از داده‌ها
- افزایش سرعت آنالیز بیگ دیتا
- افزایش سرعت و اطمینان از پیش‌بینی رفتار و روندهای منظم
- شناسایی گپ‌های اطلاعاتی و ریسک‌های امنیتی
- تسهیل تحلیل مسائل و تصمیم‌گیری

اهمیت داده کاوی

نتایجی که در داده کاوی به دست می‌آید ضریب اطمینان بالا و ریسک کمی دارند، زیرا بر اساس تحلیل‌های دقیق و جامع و بر اساس الگوهای صحیح انجام شده‌اند. این نتایج به اتخاذ تصمیم‌هایی منجر می‌شوند که احتمال عدم موفقیت آنها بسیار کم است. در تصمیم‌گیری‌های شهودی و مبتنی بر احساس هر گونه سیاست‌گذاری یا تصمیم‌گیری ممکن است با شکست یا خطا مواجه شود. اما در تصمیم‌گیری‌های مبتنی بر داده از الگوهای استفاده می‌شود که احتمال شکست تا حد بسیار زیادی کاهش می‌یابد و خطای انسانی نیز از آن حذف می‌شود. تصمیم‌گیری نادرست به معنای هدررفت سرمایه و خلاف آن به معنای سود بیشتر خواهد بود.

فرایند داده کاوی

مجموعه اقداماتی که از مرحله جمع‌آوری داده تا استخراج اطلاعات و دانش از داده را تحت پوشش قرار می‌دهد در این فرایند جای می‌گیرند. داده‌کاوی برای داده‌های بزرگ یا به اصطلاح بیگ دیتا انجام می‌شود و هدف آن شناخت همبستگی‌های داده‌ها و استخراج

الگوها و تحلیل‌هایی است که به حل مسائل و برطرف کردن نیازهای مدیریتی کمک کنند.

فرایند داده‌کاوی شامل گام‌هایی است که عبارتند از:

- **گام اول، پاک‌سازی داده (Data Cleaning):** در این گام داده‌های نامرتب از مجموعه حذف می‌شوند و برای داده‌های ناموجود نیز تدابیری اندیشیده می‌شود.
- **گام دوم، یکپارچه‌سازی داده (Data Integration):** داده‌های به‌دست‌آمده از منابع گوناگون، ارزیابی و تحلیل می‌شوند و در یک انبار داده مستقر می‌شوند.
- **گام سوم، انتخاب داده (Data Selection):** داده‌های مورد نیاز برای تحلیل در این فاز انتخاب می‌شوند.
- **گام چهارم، تبدیل داده (Data Transformation):** تبدیل داده روشی برای تثبیت داده محسوب می‌شود که در آن فرم داده‌ها تغییر می‌کند. داده‌ها به شکلی استاندارد و ساده‌تر برای تحلیل تبدیل می‌شوند.
- **گام پنجم، کاوش داده (Data Mining):** استخراج الگوهای مهم به شکل هوشمند انجام می‌شود.
- **گام ششم، ارزیابی الگو (Pattern Evaluation):** دقت، صحت و قابلیت الگوهای به‌دست‌آمده بررسی می‌شوند.
- **گام هفتم، ارائه دانش (Knowledge Representation):** نتیجه به‌دست آمده یا به اصطلاح دانش، به شیوه‌ای که برای کاربر قابل درک باشد ارائه می‌شود. در این مرحله فاز بصری‌سازی نیز دخیل خواهد بود که نقش بسیار مهمی در درک و تفسیر نتایج دارد.

چالش‌های داده کاوی



چالش‌های داده کاوی

داده‌های هر سازمان و کسب‌وکاری متشکل از اطلاعات ارزشمند و متنوعی هستند که در انجام تحلیل بر روی آنها ممکن است چالش‌هایی وجود داشته باشد. برخی از این چالش‌ها در ادامه معرفی می‌شوند و پس از آن توضیحات تکمیلی در مورد آنها آورده خواهد شد.

- حفظ حریم خصوصی و مسایل امنیتی
- چالش‌های روش‌شناسی
- داده‌های ناقص و پراکندگی داده
- مقیاس‌پذیری الگوریتم‌ها

- پیچیدگی برخی داده‌ها
- انتخاب درست روش تحلیل مناسب برای داده‌ها
- دشواری ارائه مفاهیم قابل درک از داده‌های پیچیده
- کارایی
- منابع داده

حفظ امنیت داده یکی از اولویت‌ها در هر سازمان است که تعیین نقش و میزان دسترسی به داده با توجه به این قضیه پدیده‌ای سخت و چالش‌برانگیز خواهد شد. اینکه کاربران بتوانند از داده‌های مورد نیاز خود استفاده کنند اما در عین حال حریم خصوصی و امنیت داده‌ها حفظ و تضمین شود نیازمند طراحی‌های خاصی است.

چالش‌های روش‌شناختی مربوط به روش‌ها و محدودیت‌هایی نظیر تطبیق‌پذیری در داده کاوی است. در روش‌شناسی سعی بر آن است که روش‌هایی با پیچیدگی کم و قابل تعمیم به مسائل مختلف که توانایی کار با حجم بالای داده را داشته باشد ارائه شوند.

همچنین ممکن است با انتخاب و به‌کارگیری ابزار نامناسب در تحلیل داده‌ها نتایج به‌دست‌آمده اعتبار کافی را نداشته باشند و نتایج با خطاهایی مواجه باشند که سازمان را مختل و دچار خسارت کنند. از سوی دیگر جهت انتقال نتایج به‌دست‌آمده باید از ابزار مناسب استفاده کرد تا مفاهیم را به درستی و کامل منتقل کنند.

بیشتر روش‌های آماری و هوش مصنوعی برای بیگ دیتا طراحی نشده‌اند. امروزه حجم داده‌ها به قدری زیاد است که در تحلیل آنها مسائل مقیاس‌پذیری تبدیل به چالش شده‌اند.

تنوع داده‌های کاربردی که از منابع مختلف به سمت پایگاه داده سرآزیر می‌شوند ممکن است حتی منجر به انباشتگی داده شوند و مدیریت آنها سخت شود.

مزایای ابزار هوش تجاری در داده‌کاوی

یکی از چالش‌های اصلی که در کاوش داده به وجود می‌آید انتقال مفاهیم و نتایج به‌دست‌آمده از فرایندهای داده کاوی است. کاربران مدیران و مسئولان هر سازمان و کسب‌وکاری برای درک نتایج به‌دست‌آمده نیاز به یک ابزار تصویرسازی دارند. هوش تجاری شرایطی را فراهم می‌آورد که این نتایج به اشکال مختلف قابل نمایش و مفاهیم قابل درک باشند.

داشبوردهای مدیریتی مختلف علاوه بر قابلیت نمایش نتایج، امکان تنظیماتی را فراهم می‌آورد که در آنها می‌توان داده‌کاوی را به صورت کاملاً هدفمند و شخصی‌سازی‌شده انجام داد و نتایجی به شکل گزارش‌ها و نمودارهای مختلف دریافت کرد. هوش تجاری چارگون تسهیلات فراوانی را برای مدیریت بهتر و نظارت بیشتر بر داده‌های سازمانی فراهم می‌آورد. به کمک این داشبوردهای مدیریتی دسترسی سریع به داده‌ها و اطلاعات امکان‌پذیر می‌شود. گلوگاه‌ها و نقاط قوت فرایندها شناسایی می‌شوند و قدرت پیش‌بینی نیز برای اتخاذ تصمیم‌های مهم بالا می‌رود. همچنین امکان پایش دائمی فرایندهای سازمانی فراهم می‌شود و با استفاده از آن می‌توان برنامه‌های مؤثری را در راستای افزایش بهره‌وری در سازمان شناسایی و به کار گرفت.



کاربردهای داده کاوی

از علم داده تقریباً در تمامی صنایع و فعالیتهای تولیدی، خدماتی، آموزش، اجرایی و اقتصادی استفاده می‌شود. حوزه‌های مختلفی چون بهداشت و درمان، سلامت، کسب‌وکارها، تجارت، بیمه، بانکداری، امور مالی، جامعه‌شناسی، دانشگاه‌ها و تقریباً تمامی بنگاه‌های اقتصادی از کاربردهای داده کاوی بهره‌مند خواهند شد. هر فعالیتی که با دنیای دیجیتال در ارتباط باشد، حجم زیادی از داده را در تبادلات و تراکنش‌های مختلف خود تولید می‌کند که جمع‌آوری و ارزیابی این داده‌ها به شناخت بهتر ابعاد قابل پیشرفت فعالیت مؤثر خواهد بود. در ادامه برخی از کاربردهای داده کاوی را در زمینه‌های مختلف معرفی می‌کنیم.

• سلامت عمومی: گسترش فرهنگ بهداشت و حفظ سلامت عموم مردم جهان

- پیشگیری از حملات سایبری: شناسایی الگوهای امنیتی و الگوریتم‌های مهاجم
- ساخت و عمران: تسهیل و ایجاد الگوهای بهینه شهری و راه‌سازی
- آموزش: بهبود کیفیت سیستم آموزشی
- تحقیقات جنایی و جرم‌شناسی: شناخت الگوهای نقض قانون و امنیت عمومی
- مدیریت ارتباط با مشتریان: افزایش بهره‌وری از طریق بهبود روابط با مشتریان
- تحقیقات بازار: کاربردهای مدیریتی داده کاوری جهت شناسایی فرصت‌ها و تحلیل کسب‌وکار

مزایای داده کاوی در مدیریت

به طور کل می‌توان گفت تصمیم‌گیری‌ها و سیاست‌گذاری به کمک نتایج به‌دست‌آمده از داده کاوی مسیر درست‌تر و نتایج بهتری را به دنبال خواهد داشت. برخی از مزایای داده کاوی در مدیریت عبارتند از:

- ایجاد دیدگاه متناظر با واقعیت موجود در داده‌ها برای مدیران و مسئولان
- قابلیت پیش‌بینی و کشف الگوهای احتمالی وقایعی که در آینده رخ می‌دهند
- فراهم کردن شرایط تصمیم‌گیری مبتنی بر داده
- افزایش کارایی سازمانی به کمک نتایج تحلیل عملکرد
- شناخت بهتر و بیشتر فرصت‌ها و ایرادها
- کاهش هزینه‌ها و افزایش بهره‌وری

معایب داده کاوی در مدیریت

- روش‌های داده کاوی به شکل کامل و صد در صد درست نیستند و ممکن است در برخی از شرایط عواقب و خسارات جبران‌ناپذیری را برای کسب‌وکار به وجود آورند.
- کار با سیستم‌ها و روش‌های کاوش داده نیاز به دانش سطح بالایی در علم داده دارد و به راحتی امکان‌پذیر نیست.
- در برخی از موارد ممکن است حریم خصوصی و امنیت کاربر تأمین نشود.
- هزینه مورد نیاز برای فراهم آوردن یک شرایط داده کاوی ایده‌آل زیاد خواهد بود.



سخن پایانی

در حال حاضر سازمان‌هایی که داده‌محورسازی فعالیت‌های خود را در اولویت قرار داده و از مزایای آن استفاده می‌کنند، هر روز گام‌های بیشتری را نسبت به رقبای خود در مسیر دستیابی به اهداف طی می‌کنند. داده کاوی و ابزارهای هوش تجاری در کنار یکدیگر به بررسی و تحلیل تمام داده‌های داخلی سازمان می‌پردازند و امکانات رشد و توسعه را از تمام ابعاد سنجیده و گزارش‌هایی کاملاً مبتنی بر داده از فرصت‌ها و شکاف‌های ساختاری و عملکردی هر کسب‌وکار و سازمانی را ارائه می‌کنند.