

داده کاوی چیست؟

داده کاوی، مدیریت دانش، نسبت مدیریت دانش و داده کاوی، حوزه های داده کاوی، کاربرد داده کاوی، داده کاوی چیست؟

تغییر روند توجه سازمان ها از منابع مالی به سرمایه های انسانی، فراهم آمدن زمینه های تراکم و حتی انفجار اطلاعات از طریق فن آوری اطلاعات و ارتباطات در سازمان ها، الزام و اهمیت مدیریت دانش را بیش از گذشته به رخ کشیده است. استخراج اطلاعات مناسب از میان انبوه داده ها و تبدیل آنها به دانش مورد نیاز سازمان، به ویژه در تصمیم گیری های سازمانی و دیدن آینده درست نیازمند استفاده از روش های نوین در این حوزه است. داده کاوی یکی از این ابزار و رویکردهاست که در فضای مدیریت دانش سازمان ها به کشف دانش از پایگاه داده ها کمک می کند.

● مدیریت دانش

امروزه در عصر یادگیری، کارکنان از ارزشمندترین منابع سازمان به شمار می آورند. کارکنان، حجم عظیم داده ها، اطلاعات و دانش سازمان را تولید کرده و به کار می بندند. از این رو در سازمان های نوین، همکاری مدیریت دانش و مدیریت منابع انسانی، به امری ضروری تبدیل شده است. مدیریت دانش، رویکردی سیستمی و یکپارچه است که تمامی دارایی های اطلاعاتی سازمان، اعم از: مستندات، پایگاه داده ها، خطی مشی ها، رویه ها و تجربیات کارکنان و سازمان را در برمی گیرد و به اشتراک می گذارد.

مدیریت دانش، سکوی تسهیل کننده استخراج، ذخیره سازی، یکپارچه سازی، انتقال، مشاهده، تجزیه و تحلیل و استفاده از دانش در سازمان است. مدیریت دانش به منظور حفظ، تجزیه و تحلیل، سازماندهی، بهبود و تسهیم تجارب کسب و کار، به عنوان ابزار، روش و راهبرد در سازمان به کار می رود. در واقع، مدیریت دانش متدولوژی استخراج سیستمی و استفاده از دانش در سازمان است که از طریق همین دانش و سرمایه فکری، ثروت و ارزش می آفریند. در نگرش سیستمی به مدیریت دانش، سه مولفه اصلی: منابع انسانی، فرایندها و فن آوری دخالت و با یکدیگر همپوشانی دارند.

● نسبت مدیریت دانش و داده کاوی

برای به کارگیری مدیریت دانش در سازمان، از ابزار متفاوتی استفاده می شود که یکی از مهم ترین آنها، داده کاوی است. هنگامی که طراحی مجدد فرآیندهای کسب و کار (BRP) مورد نظر است، داده کاوی به عنوان تکنیک، در خدمت مدیریت دانش قرار می گیرد. در هر بنگاه صنعتی، اقتصادی و تجاری نیز دانش با ارزش در مورد مشتری، محصول و بازار را می توان از طریق داده کاوی به دست آورد. با بزرگ تر شدن حجم داده ها و اطلاعات سازمان ها، اهمیت این مسئله روز به روز بیشتر می شود. هنگامی که حجم داده ها زیاد است، برای مدیریت آنها از سیستم های پایگاه داده ها

هنگامی آید می کار به دانش آوردن دست به و استخراج برای کاوی داده ها، حجم بودن زیاد حالت در .شود می استفاده (DBS DBMS) که حجم دانش زیاد است، از سیستم مدیریت دانش (KMS) استفاده می شود.

● داده کاوی

پیشینه طرح موضوع داده کاوی به دهه ۱۹۸۰ و به صورت جدی، به دهه ۱۹۹۰ برمی گردد. پیش از آن از سیستم های جمع آوری و مدیریت داده ها و اصطلاحاً لایروبی داده ها استفاده می شد، اما به مرور زمان، استخراج و کشف سریع و دقیق اطلاعات با ارزش و پنهان از پایگاه داده ها، به عنوان داده کاوی مورد توجه قرار گرفت. به این شکل بود که فرایند داده کاوی به عنوان فرایند آماری و تجزیه و تحلیل در فرایند کشف دانش در پایگاه داده ها (KDD) پیرنگ شد، به حدی که گاه، داده کاوی (DM) به عنوان مترادف کشف دانش در پایگاه داده ها (KDD) مورد استفاده قرار می گرفت. امروزه فرآیند استخراج اطلاعات معتبر، از پیش ناشناخته، قابل فهم و قابل اعتماد از پایگاه داده های بزرگ و استفاده از آن در تصمیم گیری و در فعالیتهای تجاری «داده کاوی» نامیده می شود. در تعاریف متعدد و متنوع برای داده کاوی بر موضوعاتی نظیر: استخراج دانش کلان، کاوش در داده ها، تجزیه و تحلیل داده ها و یافتن روابط و الگوهای مطمئن بین داده ها تاکید می شود. هدف نهایی داده کاوی، ایجاد سیستم های پشتیبانی تصمیم گیری سازمانی است. داده کاوی، به استخراج اطلاعات مفید و دانش از حجم زیاد داده ها می پردازد. داده کاوی، الگوهای حاوی اطلاعات را در داده های موجود جست و جو می کند. این الگوها و الگوریتم ها، می توانند توصیفی باشند، یعنی داده ها را توصیف کنند و یا جنبه پیش بینی داشته باشند، یعنی متغیرها برای پیش بینی ارزش های ناشناخته سایر متغیرها به کار روند. داده کاوی توصیفی، به دنبال یافتن اگرها در فعالیتهای یا اقدامات گذشته است و داده کاوی پیش بینانه با نگاه به سابقه، رفتار آینده را پیش بینی می کند.

● داده کاوی چیست و چه نیست

موارد زیر در حوزه داده کاوی قرار می گیرند:

استخراج یا کاوش دانش از میان حجم عظیم داده ها، استخراج اطلاعات و مدل کردن الگوهای پنهانی در میان انبوه داده ها، استخراج اطلاعات

غیر منتظره، ناشناخته و بالقوه مفید از داده ها، استخراج اطلاعات یا الگوهای مفید و جالب از داده ها در پایگاه داده های بزرگ.

موارد زیر در محدوده داده کاوی قرار نمی گیرند:

انبارش داده ها، پردازش قیاسی داده ها، سیستم های خبره (ES)، فرایند تجزیه و تحلیل مستقیم، ابزار تجزیه و تحلیل آماری، مشاهده گری داده ها.

● حوزه های داده کاوی

داده کاوی در سه حوزه مستقل به کار می رود و در آنها ریشه دوانده است:

(۱) آمار کلاسیک و الگوهای آماری

(۲) هوش مصنوعی

(۳) یادگیری خودکار و شبکه های عصبی

در داده کاوی، هوش مصنوعی، یادگیری خودکار، تئوری پایگاه داده ها و علم آمار در هم آمیخته شده است. برای انجام داده کاوی از ابزار مختلف نظیر: تفکیک کردن، دسته بندی، درخت تصمیم گیری، تحلیل قواعد وابستگی و تحلیل خوشه ها و الگوریتم های عمومی استفاده می شود.

● مراحل فرآیند

هدف داده کاوی، تجزیه و تحلیل اکتشافی داده ها، کشف الگوها و قواعد و الگوریتم ها، مدل سازی پیش بینانه و جست و جوی انحرافات است. برای انجام این هدف، فرآیند داده کاوی در جهت کشف دانش در مراحل مختلف انجام می شود که عبارت است از:

(۱) اولین گام در داده کاوی، شناسایی هدف و فهم حوزه کاربرد آن است و مشخص می کند که چه کاری، در چه حوزه ای انجام خواهد شد.

(۲) انتخاب داده ها یعنی تعیین اهداف برای تجزیه و تحلیل و کشف آن

(۳) آماده سازی داده ها شامل تمیز سازی داده ها

(۴) اتخاذ بهترین روش داده کاوی برای دست یابی به اهداف

(۵) اجرای داده کاوی یعنی به کارگیری الگوریتم

(۶) ارزیابی و اعتبارسنجی یافته ها

(۷) استفاده از نتایج و تثبیت و تحکیم دانش کشف شده

(۸) تصمیم گیری براساس دانش کشف شده.

● موانع و چالش ها

با وجود مزایای فراوان که داده کاوی برای مدیریت دانش سازمان ها و دست یابی به دانش برای تصمیم گیری سازمانی دارد، این حوزه با موانع و چالش هایی روبه روست که عبارتند از:

(۱) فقدان داده برای پشتیبانی تجزیه و تحلیل

(۲) قدرت محدود محاسبه برای به دست آوردن محاسبات ریاضی مورد نیاز الگوریتم های داده کاوی

(۳) عدم مطلوبیت و جذابیت بیشتر الگوها

(۴) خطر وجود داده های آلوده و کسب نتایج کاملاً غلط

(۵) تمرکز بیش از حد بر الگوریتم ها

(۶) هزینه نسبتاً گران سرمایه گذاری در همه حوزه ها

(۷) عدم پوشش کامل همه حوزه ها

● کاربرد داده کاوی

امروزه داده کاوی در حوزه های بسیار متنوع و متفاوت استفاده می شود:

■ حوزه علم: شیمی، فیزیک، داروسازی، تجزیه و تحلیل تصاویر پزشکی، تعیین نوع رفتار با بیماران و پیشگویی میزان موفقیت های اعمال پزشکی، تعیین میزان موفقیت روش های درمانی در برخورد با بیماری های سخت، بیوشیمی، حسگرهای کنترل در اقمار مصنوعی، بیو علم شامل توصیف ژن ها و تقسیم بندی گروه پروتئین ها و توسعه داروها. خرده فروشی: تجزیه و تحلیل سبد خرید بازار، تعیین الگوهای خرید مشتریان.

■ بانکداری: پیش بینی الگوهای کلاهبرداری از طریق کارت های اعتباری، تعیین میزان استفاده از کارت های اعتباری براساس گروه های اجتماعی فروش و بازاریابی: تجزیه و تحلیل سهام و سرمایه، تعیین مشتریان وفادار، مدیریت ریسک و پیش بینی فروش بیمه: تجزیه و تحلیل دعاوی و پیشگویی میزان جریمه بیمه نامه های جدید توسط مشتریان دیگر حوزه ها: ورزش و سرگرمی، فضاوردی و دیگر حوزه های علم.