

## عنوان:

### داده کاوی و ارتباط آن با صنعت

تهیه کننده:

محمد امین ادیبی – دانشجوی دکترای مهندسی صنایع

دانشگاه صنعتی امیر کبیر

### داده کاوی و ارتباط آن با صنعت

دانش با ارزش ترین دارائی هر سازمان خصوصی یا دولتی می باشد بطوریکه می توان دانش در اختیار یک سازمان را عامل پیشرفت یا تمایز آن با رقبایش دانست. دانش است که سازمان ها را قادر می سازد تا بتوانند با استفاده از توانمندی هایشان به پیشرفت دست یابد یا در محیط های رقابتی با رقبایشان بصورت موثر و کارا رقابت کنند. از طرفی، اگرچه دانش در حوزه های مختلف یک سازمان وجود دارد اما شناسائی، جمع آوری و مدیریت آن اغلب دشوار و چالش برانگیز است. از این رو مباحثی همچون کشف دانش، مدیریت دانش و مهندسی دانش از عناوین حائز اهمیت برای محققین و مدیران می باشد.

از سوی دیگر پیشرفتهای در تکنولوژی اطلاعات (IT)، سیستمهای جمع آوری داده و تکنولوژیهای ذخیره سازی داده در کنار توسعه روشهای علمی باعث شده است محققان به سوی کشف دانش از پایگاههای داده تمایل پیداکنند. در واقع داده ها از همه ی فعالیتهای یک سازمان ثبت می شوند که این داده های ذخیره شده پتانسیل بالائی به عنوان منبعی برای کشف دانش جدید دارند. لذا استفاده از داده های جمع آوری شده به موضوعی درخور توجه تبدیل شده است. داده کاوی (Data Mining) یک پاسخ طبیعی به نیاز تبدیل داده های ذخیره شده به دانش مفید می باشد که از این دانش استخراج شده می توان برای مدلسازی، دسته بندی و پیش بینی در کاربردهای مختلف استفاده کرد.

دانش داده کاوی امروزه با پیشرفت صنعت پردازندهها و امکان ذخیره سازی، بازیابی و تحلیل دادهها رونق بسیاری پیدا کرده و جوابگوی بسیاری از سوالات و نیازهای محققین و مدیران کسب و کارهای کوچک و بزرگ،

دستگاه‌های حکومتی و حتی افراد عادی شده است. از فردی که در اینترنت جستجو می‌کند تا کسی که به عنوان سیاستمدار سعی در پیش بینی وضع اقتصادی است، چه بدانند و چه ندانند از نتایج کاربردی شدن دانش داده‌کاوی بهره می‌برند.

داده‌کاوی ترکیبی از مفاهیم و الگوریتمها از آمار، هوش مصنوعی، یادگیری ماشینی و مدیریت داده است که در حوزه‌های مختلفی همچون امور حاکمیتی، بانکداری، مالی، توزیع و فروش، بازاریابی، بیمه، علوم، مهندسی و ... برای کشف روابط و الگوهای پنهان بکار می‌رود. داده‌کاوی یک حوزه در حال گسترش بوده که هر روزه بر اهمیت و جذابیت آن افزوده می‌شود. صنعت نیز یکی از حوزه‌هایی است که داده‌کاوی می‌تواند در آن مورد استفاده قرار گیرد و باعث بدست آمدن مزایای فراوان برای سازمان‌ها شود.

داده‌کاوی نیازمند شناسایی مساله، جمع‌آوری داده در خصوص آن مساله که در ضمن باعث شناخت بیشتر از مساله هم می‌شود و مدل‌های کامپیوتری که بوسیله آن تجزیه و تحلیل‌های آماری و سایر ابزارهای آنالیز قابل انجام هستند، می‌باشد.

به‌منظور هدایت سیستماتیک تجزیه و تحلیل داده‌کاوی، معمولاً از یک فرایند کلی پیروی می‌شود. این فرایندها، فرایندهای استاندارد هستند که به عنوان مثال می‌توان به فرایند داده‌کاوی CRISP-DM که یک فرایند استاندارد داده‌کاوی برای صنعت می‌باشد و SEMMA که بصورت ویژه برای نرم افزار SAS توسعه داده شده است اشاره کرد. گام‌های فرایند CRISP-DM بصورت زیر می‌باشد:

#### ۱. تعریف مساله

تعریف مساله شامل تعریف هدف مساله، ارزیابی شرایط فعلی، تعریف اهداف داده‌کاوی و ایجاد یک برنامه زمانبندی پروژه داده‌کاوی است.

#### ۲. تحلیل داده‌ها

وقتی که هدف مساله مشخص شد و برنامه ریزی پروژه انجام گرفت، گام تحلیل داده‌ها نیازمندی‌های داده‌ای را مورد مطالعه قرار می‌دهد. این مرحله شامل تهیه مجموعه داده‌های اولیه، توصیف داده‌ها، کنکاش در داده‌ها و ارزیابی کیفیت داده است. کنکاش داده‌ها شامل مشاهده‌ی پارامترهای آماری، می‌تواند در پایان این فاز نیز اتفاق بیافتد. مدل‌هایی مثل خوشه‌بندی می‌تواند در طول این گام به منظور شناسایی الگوها در داده‌ها انجام گیرد.

#### ۳. آماده‌سازی داده‌ها

زمانی که منابع در دسترس داده مشخص شدند، بایستی داده‌ها از آن انتخاب، پاک و در قالب موردنظر ریخته شوند. همچنین کنکاش عمیق‌تر داده‌ها نیز در طول این فاز می‌تواند انجام گیرد. به علاوه ممکن است مدل‌های دیگری نیز بکار روند تا بتوان الگوها را براساس تعریف مسأله استخراج کرد.

#### ۴. مدل‌سازی

ابزارهای نرم‌افزاری داده‌کاوی همچون تصویرسازی (یعنی ترسیم داده‌ها و تشکیل رابطه‌ها) و تحلیل خوشه‌بندی (یعنی تشخیص اینکه کدام داده‌ها و مشاهدات بهتر در کنار هم قرار می‌گیرند) برای تحلیل‌های اولیه مفید هستند. ابزارهایی مانند شناسائی قوانین عمومی می‌تواند قوانین همبستگی اولیه را استخراج کند. وقتی که فهم بیشتری از داده‌ها به وسیله‌ی شناسائی الگو که با توجه به خروجی‌های مدل‌های اولیه حاصل می‌شود، افزایش می‌یابد، مدل‌های تخصصی‌تر برحسب نوع داده می‌تواند مورد استفاده قرار بگیرد. در این مرحله تقسیم بندی داده‌ها به مجموعه داده‌های آموزشی و مجموعه داده‌های آزمون لازم است.

#### ۵. ارزیابی

نتایج مدل‌های استفاده شده در مراحل قبلی بایستی در بستر مساله تعریف شده و اهداف تعریف شده در مراحل اولیه مورد ارزیابی قرار بگیرد. این کار منجر به شناسایی نیازهای بعدی خواهد شد. این نیازها اغلب شامل بازگشت به مراحل قبلی در فرایند CRISP-DM است. تعریف مساله در داده‌کاوی فرایند تکراری با بازگشت به عقب است که در آن نتایج اقدامات مدل‌سازی مختلف و تکرار آنها روابط جدید بین داده‌ها را به کاربر نشان می‌دهد که باعث درک عمیق‌تر از مساله برای کاربر خواهد شد.

#### ۶. توسعه

داده‌کاوی هم می‌تواند برای بررسی صحت فرضیه‌هایی که قبلاً وجود داشته مورد استفاده قرار می‌گیرد و هم برای کشف دانش یعنی شناسایی روابط غیرمنتظره و مفید. به کمک دانش کشف شده در فازهای قبلی فرایند CRISP-DM مدل‌های درست و منطقی حاصل می‌شوند که می‌توان آن‌ها را برای اهداف مختلف بکار برد. این اهداف می‌توانند شامل پیش‌بینی یا شناسائی شرایط حساس و کلیدی باشند.

هرچند که تمامی گام‌های ذکر شده در این فرایندهای استاندارد ممکن است در یک مطالعه‌ی داده‌کاوی موردنیاز نباشد ولی این فرایندها باعث همگرا شدن گام‌های داده‌کاوی به نتیجه‌ی دلخواه می‌شوند

در ادامه به شرح کاربرد‌های داده‌کاوی در صنعت بصورت مختصر پرداخته می‌شود.

### • داده‌کاوی در طراحی مهندسی

طراحی مهندسی یک فرایند تصمیم سازی بین رشته ای، چند بعدی و غیر خطی است که طی آن پارامترها، فعالیت ها و اجزاء انتخاب می شوند، چنین انتخاب هایی معمولاً بر مبنای داده، اطلاعات و دانش تاریخی می باشد. لذا طراحی مهندسی می تواند یکی از حوزه های مهم و اولیه برای کاربرد داده کاوی باشد. از جمله کاربردهای داده کاوی در طراحی مهندسی به وارد زیر اشاره نمود:

- انتخاب قطعات مناسب
- بهینه سازی ویژگی های طراحی
- تخمین هزینه محصول
- طراحی تست های فنی
- توسعه محصول
- برنامه ریزی مواد
- انتخاب طرح بهینه

### داده کاوی در عملیات ساخت و تولید

جمع آوری داده در فرایند تولید ساخت بسیار معمول است اما استفاده از آنها معمولاً محدود به چند کاربرد خاص است. این در حالی است که ابزارهای داده کاوی پتانسیل بسیار عالی برای استفاده در کنترل بهتر سیستم های ساخت و تولیدی دارند. این کاربرد بخصوص در سیستم های تولیدی باعث ارتقاء بازدهی خطوط می شود. در ادامه به چند کاربرد داده کاوی در این حوزه اشاره شده است:

- توسعه استراتژی های تولید
- کنترل کیفیت محصول
- بهینه سازی پارامترهای ماشینکاری
- پیش بینی عملکرد و بازدهی فرایندها
- تشخیص خصوصیات مواد
- تعیین خواص مکانیکی محصولات

### داده کاوی در سیستم های پشتیبان تصمیم

باید توجه داشت که دانش با ارزش ترین دارایی هر سازمان است و تصمیم بر مبنای ترکیب دانش و قضاوت که از حوزه های مختلف در کنار هم جمع می شوند به وجود می آید. با دانستن این حقیقت می توان جایگاه داده کاوی در تصمیم گیری را درک کرد. در زیر کاربردهای داده کاوی در ایجاد سیستم های پشتیبان تصمیم در بخش های مختلف ذکر شده است:

- اصلاح برنامه زمانبندی
- درک داده های بدست آمده از فرایند صنعتی
- بهبود دقت و عملکرد سیستم تولیدی

- بررسی ارتباط بین پارامترهای تولید و کیفیت محصول
- کاربردی کردن شبیه سازی
- ارتقاء عملکرد سیستم های ERP

### داده کاوی در چیدمان و کنترل تسهیلات (کارخانه ها)

مسائل چیدمان و کنترل کارخانه با چگونگی استفاده موثر و کار از منابع سرو کار دارند. داده های بسیار زیادی در طی عملیات در کارخانه ضبط می شوند. این داده ها می توانند برای بهینه سازی فرایندها مورد استفاده قرار گیرند. داده های تاریخی جمع آوری شده از این طریق می توانند اطلاعات بسیار قوی از عملکرد کارخانه و فرایند جاری به دست دهد. نمونه هایی از کاربرد داده کاری در این حوزه عبارتند از:

- طراحی سلول های تولیدی
- ارزیابی طرح های چیدمان
- مهندسی همزمان

### داده کاوی در شناسایی عیوب و بهبود کیفیت

شناسایی عیوب یکی از نخستین حوزه هایی است که داده کاوی در آن بکار رفته است. ایده بکار گرفته شده در داده کاوی در این خصوص مطالعه ی رفتار گذشته سیستم و فهم فرایند و سپس پیش بینی و ارتقاء عملکرد آینده سیستم است. از داده کاوی می توان برای شناسایی الگوهایی که می توانند به خرابی منجر شوند استفاده کرد. چنین شیوه هایی نه تنها در شناسایی محصولات معیوب کمک می کنند بلکه بصورت فنی پارامتر های مهم و تاثیر گذار بر کیفیت را نیز تعیین می کنند. چند نمونه از کاربردها در ادامه ذکر شده است:

- کنترل کیفیت
- تعیین طرح های کنترل کیفی
- شناسایی نوع خرابی
- شناسایی مکانیزم خرابی قطعات مکانیکی
- بهبود کیفیت
- مهندسی کیفیت
- پیش بینی احتمال خرابی

### داده کاوی در نگهداری و تعمیرات

نگهداری پیشگیرانه از کلیدی ترین مباحث در مهندسی ساخت و فرایند است. پایگاه های داده شامل موارد خرابی ماشینالات و تجهیزات و رفتار آنها در زمان خرابی می تواند برای طراحی سیستم های مدیریت نگهداری و تعمیرات مورد استفاده قرار گیرند. رویکردهای مختلف استفاده از داده کاوی در نگهداری و تعمیرات عبارتند از:

- مطالعه روی داده های مربوط به زمان خرابی در گذشته و پیشنهاد برنامه ی مناسب نگهداری

- ایجاد شیوه های نگهداری بهینه
- پیش بینی خرابی بر اساس سیگنال های دریافتی از سنسورها
- شناسایی زیر سیستم های موثر بر کارکرد ضعیف سیستم

### داده کاوی در مدیریت روابط مشتری

در سالهای اخیر مدل های بازاریابی از رویکرد محصول محوری به رویکرد مشتری محوری تغییر کرده است. مدیریت روابط مشتری (CRM) در ارتباط با افزایش سود حاصل از تعامل با مشتری ها فعالیت می کند.

در محیط پر رقابت کسب و کار امروزی، کاربرد داده کاوی در CRM توجه بسیاری به خود جلب کرده است. در CRM نیز همانند تولید محصول با کیفیت بالا و هزینه پایین دارای اهمیت بوده و در ارتباط با مدیریت تقاضا از مشتری نهائی (بصورت عقبگرد) تا تامین کننده ی مواد خام گسترانیده شده است. به نمونه هایی از کاربرد داده کاوی در CRM در لیست زیر اشاره شده است:

- بخش بندی مشتریان برای شناسایی تمایلات آنها و تغییرات احتمالی
- بهبود کسب و کار با هدف گذاری روی مشتریان
- طراحی محصول