**اده کاوی یک ابزار آنالیز مدیریتی**

**Data Mining an Analysis Implement Managemental**

استاد راهنما : مهندس عمادیارائه دهنده : قربان مقدم زرزری

چکیده

داده کاوی فرآیندی تحلیلی برای کاوش داده های طراحی شده است،که در جستجوی الگوهای سازگار، یا روابط سیستماتیک بین متغیرها است، و سپس به تائید این یافته ها با استفاده از الگوهای تشخیص داده شده می پردازد. استخراج اطلاعات مناسب از ميان انبوه داده‏ها و تبديل آنها به دانش مورد نياز سازمانها، بويژه در تصميم‌گيري‏هاي سازماني٬ نيازمند استفاده از روش‏هاي نوين در اين حوزه است. داده كاوي يكي از اين ابزار و رويكردهاست كه در فضاي مديريت دانش سازمان‏ها به كشف دانش از پایگاه داده ها كمك مي‏كند. این مقاله به بررسی ویژگی های منحصر به‏فرد این حوزه از فناوری و تکنیکهای استفاده از آن را نشان می دهد.

**مقدمه**

پيشينه طرح موضوع داده‏كاوي به دهه 1980 و به صورت جدي، به دهه 1990 برمي‏گردد. پيش از آن٬ از سيستم‏هاي جمع‏آوري و مديريت داده‏ها و اصطلاحاً لايروبي داده‏ها استفاده مي‏شد٬ اما به مرور زمان٬ استخراج و كشف سريع و دقيق اطلاعات با ارزش و پنهان از پايگاه داده‏ها به‏عنوان داده‏كاوي مورد توجه قرار گرفت. به اين شكل بود كه فرايند داده‏كاوي به عنوان فرايند آماري و تجزيه و تحليل درفرايند كشف دانش در پايگاه داده‏ها (KDD) پررنگ شد ،به حدي كه گاه٬ داده‏كاوي (DM) به‏عنوان مترادف كشف دانش در پايگاه داده‏ها(KDD) مورد استفاده قرار مي‏گرفت[2]. امروزه فرايند استخراج اطلاعات معتبر٬ از پيش ناشناخته٬ قابل فهم و قابل اعتماد از پايگاه داده‏هاي بزرگ و استفاده از آن در تصميم‏گيري و در فعاليت‏هاي تجاري داده‏كاوي ناميده مي‏شود[1]. در تعاريف متعدد و متنوع براي داده‏كاوي برموضوعاتي نظير: استخراج دانش كلان٬ كاوش در داده‏ها٬ تجزيه و تحليل داده‏ها و يافتن روابط و الگوهاي مطمئن بين داده‏ها تاكيد مي‏شود. هدف نهايي داده‏كاوي٬ ايجاد سيستم‏هاي پشتيباني تصميم‏گيري سازماني است. داده‏كاوي به استخراج اطلاعات مفيد و دانش از حجم زياد داده‏ها مي‏پردازد. داده‏كاوي٬ الگوهاي حاوي اطلاعات را در داده‏هاي موجود جست‌وجو مي‌كند. اين الگوها و الگوريتم‏ها، مي‏توانند توصيفي باشند يعني داده‏ها را توصيف كنند و يا جنبه پيش‏بيني داشته باشند، يعني از متغيرها براي پيش‏بيني ارزش‏هاي ناشناخته ساير متغيرها به‏كار روند. داده‏كاوي توصيفي، به‏دنبال يافتن اگرها در فعاليت‏ها يا اقدامات گذشته است و داده‏كاوي پيش‏بينانه با نگاه به سابقه٬ رفتار آینده را پيش‏بيني مي‏كند[1].

**حوزه فعالیتهای داده کاوی**

هدف داده‏كاوي٬ تجزيه و تحليل اكتشافي داده‏ها٬ كشف الگوها و قواعد و الگوريتم‏ها٬ مدل‏سازي پيش‏بينانه و جست‌وجوي انحرافات است. براي انجام اين هدف٬ فرايند داده‏كاوي درجهت كشف دانش درمراحل مختلف انجام مي‌شود كه عبارت است از:

1.        شناسايي هدف و فهم حوزه كاربرد آن است و مشخص مي‏كند كه چه‏كاري٬ در چه حوزه‏اي انجام خواهد شد.

2.        انتخاب داده‏ها يعني تعيين اهداف براي تجزيه و تحليل و كشف آن

3.         آماده‏سازي داده‏ها شامل تميزسازي داده‏ها

4.        اتخاذ بهترين روش داده‏كاوي براي دست‏يابي به اهداف

5.         اجراي داده‏كاوي يعني به‏كارگيري الگوريتم

6.         ارزيابي و اعتبارسنجي يافته‏ها

7.        استفاده از نتايج و تثبيت و تحكيم دانش كشف شده

8.        تصميم گيري براساس دانش كشف شده

اکتشاف در این مرحله معمولا با آماده سازی داده ها که ممکن است شامل تمیز کردن داده ها، تبدیل داده ها ، زیر مجموعه های انتخاب آثار ضبط شده و انجام برخی از عملیات اولیه انتخاب شروع می شود . سپس بسته به ماهیت تحلیلی ، این مرحله از فرایند استخراج داده ها ممکن است شامل هر انتخاب ساده و سر راست برای یک مدل رگرسیون استادانه درست شده را به تجزیه و تحلیل اکتشافی با استفاده از طیف گسترده ای از روش های گرافیکی و آماری به منظور شناسایی متغیرهای مربوطه و تعیین پیچیدگی از طبیعت مدل ها باشد. البته ناگفته نماند که داده کاوی معمولا با نوشتن مقدار زیادی گزارش و تحقیق و استعلام در آنها اشتباه گرفته می شود. اما در واقع داده کاوی هیچ کدام از اینها را شامل نمی شود. داده کاوی توسط تجهیزات خاصی صورت می پذیرد، که عملیات کاوش را بر اساس تجزیه و تحلیل مکرر داده ها انجام می دهد. داده کاوي با آنالیز های متداول آماری نیز متفاوت است؛در زیرمی توان برخی ازاصلی ترین تفاوت های داده کاوي و آنالیز آماری را مشاهدهنمود:

**آنالیز آماری:**

•          آمار شناسان همیشه با یک فرضیه شروع به کار می کنند

•          آنها از داده های عددی استفاده می کنند

•          آمارشناسان باید رابطه هایی را ایجاد کنند که به فرضیه آنها مربوط است

•          آنها می توانند داده های نابجا و نادرست را در طول آنالیز مشخص کنند

•          آنها می توانند نتایج کار خودرا تفسیر و برای مدیران بیان کنند

**داده کاوی :**

•          به فرضیه احتیاجی ندارد

•          ابزارهای داده کاوی از انواع مختلف داده ، نه تنها عددی می توانند استفاده کنند

•          الگوریتمهای داده کاوی به طور اتوماتیک روابط را ایجاد میکنند

•          داده کاوی به داده های صحیح و درست نیاز دارد

•          نتایج داده کاوی نسبتا پیچیده می باشد و نیاز به متخصصانی جهت بیان آنها به مدیران دارد

جهت درک بهتر تفاوت داده کاوی و آنالیزهای آماری به مثال زیر که در مورد شناخت کلاهبرداری های شرکت بیمه می باشد، توجه کنید.

**روش آنالیز آماری :**

یک مفسر ممکن است متوجه الگوی رفتاری شود که سبب کلاهبرداری بیمه گردد. براساس این فرضیه، مفسر به طرح یک سری سوال می پردازد تا این موضوع را بررسی کند. اگر نتایج حاصله مناسب نبود، مفسر فرضیه را اصلاح می کند و یا با انتخاب فرضیه دیگری مجددا شروع می کند. این روش نه تنها وقت گیر است بلکه به قدرت تجزیه و تحلیل مفسرنیز بستگی دارد. مهمتر از همه اینکه این روش هیچ وقت الگوهای کلاهبرداری دیگری را که مفسر به آنها مظنون نشده و در فرضیه جا نداده ، پیدا نمی کند.

**روش داده کاوی :**

 یک مفسر سیستم های داده کاوی را ساخته و پس از طی مراحلی ازجمله جمع آوری داده ها، یکپارچه سازی و اخلاص داده ها به انجام عملیات دادهکاوی می پردازد. داده کاوی تمام الگوهای غیرعادی را که از حالت عادی و نرمالانحراف دارند و ممکن است منجر به کلاهبرداری شوند را پیدا می کند. نتایج دادهکاوی حالت های مختلفی را که مفسر باید در مراحل بعدی تحقیق کند، نشان می دهند. درنهایت مدل های به دست آمده می توانند مشتریانی را که امکان کلاهبرداری دارند، پیشبینی نمایند.

تحلیلهاي داده‏کاوي به دو روش با ناظر و بدون ناظر و از طریق الگوریتمهایی چون شبکه هاي عصبی(NN)، طبقه بندي و درخت تصمیم (C&RT)، ژنتیک، تحلیل سبد خرید، شبکه کوهونن قابل اجراست. علاوه بر این الگوریتمهاي رایج، همچنان الگوریتمهاي جدیدي براي اهداف تحقیقات علمی یا تجاري از طریق طرحهاي پژوهشی دانشگاهی، تولید می‏شود. ویژگیهاي منحصر بفرد داده کاوي را می توان به صورت زیر برشمرد[3]:

·         نه تنها بر فاز تحلیل، بلکه بر طراحی مطالعه و جمع آوري داده نیز تاثیر می گذارند

·         امکان جستجوي پاسخ سؤالات دقیق و با پیچیدگی بالا را در دادههاي جمع‏آوري شده فراهم می کنند.

·         قادرند که به سؤالات بطور واضح و مشخص پاسخ دهند. مزیت اصلی و تفاوت آنها با سایر تکنیکها نیز در همین است که بجاي ارائه صرف استراتژي کلان، پاسخهاي دقیق در اختیار محقق قرار می دهند.

·         امکان سنجش اثر متغیرهاي مختلف بر روي متغیرهاي وابسته را فراهم می کنند.

·         به مدیران کمک می‏کنند که تأثیر سناریوهاي آتی را مورد ارزیابی قرار دهند و با مدلسازي گزینه هاي متعدد و کمک به تصمیم گیري در شرایط عدم قطعیت به انتخاب مسیر حرکت بپردازند.

محققینی که تنها روابط دو به دو را در نظر میگیرند و از داده کاوی استفاده نمی کنند، ابزار قدرتمندی را از دست می دهند که می تواند اطلاعات سودمندی را در اختیار آنان قرار دهد. در مسائل واقعی چندین متغیر به طور همزمان بر روي پاسخ تاثیر می گذارند، از این رو آنالیزهاي چندمتغیره جواب هاي دقیقتر و نزدیک به واقع تري را فراهم می کند. موفقیت در این مرحله کاملا متاثر از سه گام قبل است بگونه اي که اگر هر کدام از مراحل قبلی به درستی انجام نپذیرد، نتایج حاصل از داده کاوي نه تنها مفید نبوده ممکن است گمراه کننده نیز باشد.

تکنیکهاي داده کاوي از جمله تکنیکهاي نوین علمی هستند که در توصیف، تشریح، پیش بینی و کنترل پدیدهها به کار می روند[3]. این تکنیکها به اندازه‏گیري، تشریح و پیش‏بینی درجه وابستگی میان متغیرها میپردازند. روشهاي داده‏کاوي نه تنها بر جنبه هاي تحلیلی مطالعات، بلکه در طراحی و ابزارهاي جمع آوري داده براي تصمیم‏گیري و حل مسائل نیز تأثیر می‏گذارند. موفق‏ترین پروژههاي داده‏کاوي، در چارچوب فرآیند استانداردي اجرا می شودکه توسط یک تیم کاري در شرکت SPSS در قالب پروژهاي به نام CRISP-DM ارائه شده است[5]. برطبق CRISP-DM یک پروژه داده‏کاوی معین شامل چرخه حیاط شش مرحله‏ای است که توالی مراحل را نشان می دهد

حلی برای مسئله تحقیقی یا تجاری با سوالات اضافی جالب توجه می شود. در زیر مراحل کاری در داده‏کاوی را توضیح می‏دهیم:

مرحله درک پروژه و فهم حوزه کاربرد: اولین مرحله پردازش استاندارد CRISP–DM   می باشد که به صورت آشکار اهداف و نیازمندیها آن مشخص می شود. ترجمه اهداف و محدودیت آن در قاعده‏سازی، تعریف مسئله داده‏کاوی و مهیا کردن استراتژی اولیه برای نائل شدن به اهداف تعریف می‏شود.

مرحله انتخاب دادها : این مرحله شامل جمع آوری دادها برای استفاده از تحلیل اکتشافی و مشخص کردن اطلاعات اولیه برای ارزیابی داده‏های با کیفیت و انتخاب دادهای مفید و مورد نیاز می باشد.

مرحله آماده سازی داده‏ها: آماده کردن داده‏های اولیه خام به داده‏های نهایی، این دادها در کلیه مراحل بعدی استفاده می شود و از این نظر این مرحله تحلیل و تلاش بیشتری را می طلبد. انتخاب عناصر و شناسه‏های تحلیل شده را برای کاوش داده‏ها اختصاص می دهیم. و با تمیز کردن دادهای خام آن را برای ابزارهای مدلسازی آماده می‏کنیم.

مرحله مدلسازی: با انتخاب و به‏کار بستن تکنیکهای مدلسازی مناسب و روش داده‏کاوی معین نتایج مدلسازی را بهینه می‏کنیم که در صورت نیاز می توانیم با برگشت به عقب تحلیل مدلسازی را بهینه تر نماییم.

مرحله ارزیابی: مشخص کردن اینکه آیا مدل انتخابی، ما را به اهدافمان که در اولین مرحله تعیین کردیم می‏رساند. اتخاذ تصمیم راجع به استفاده از نتایج داده‏کاوی برای اعتبارسنجی نیز در این مرحله انجام می‏شود.

مرحله تحکیم و گسترش :استفاده کردن از مدل ایجاد شده، برای مثال می تواند تولید یک گزارش ساده از خروجیها را نام برد، و برای یک مثال پیچیده تکمیل کردن پردازش داده‏کاوی موازی در سایر حوزه‏ها می باشد که این الگو‏ها به یک دانش مفید و قابل استفاده تبدیل می شوند و پس از بهبود آنها، الگوهایی که کارا محسوب می‏شوند در یک سیستم اجرایی به کار گرفته خواهند شد.

نتیجه گیری

بررسی اجمالی پژوهشهاي صورت گرفته در حوزه دانش ابزراهای داده‏کاوی نشان می دهد که تحقیقات عمیق و اساسی در این باره خصوصاً در ایران اندك شمار است. از سوي دیگر با افزایش سرعت تحول در علوم، ضرورت استفاده از دانشهاي نوین بیش از پیش محرز شده است . داده کاوي به عنوان یک رشته علمی نوین در زمینه بازیابی و استخراج اطلاعات می تواند نقش مهمی در جهت دستیابی به این اهدافداشته باشد. امروزه اکثر نرم افرار های پایگاه داده ای مثل ORACLE و SQL Server نیز شامل ابزارهایی داده کاوی شده اند ولی نرم افزار های تخصصی داده‏کاوی همچون Intelligent Miner , Darwin , Mine Set, Knowledge Studio, Data Mind از مهمترین ابزار های داده کاوی به شمار می روند. در این مقاله قابلیتهاي دادهکاوي و مراحل کاری آن معرفی شد که در گامهای بعد می توان تأثیر آن را در عمل آزمود.