**معرفی داده کاوی**

  تهيه و تنظيم : علي محرمي

داده کاوی یا کشف دانش در پايگاه داده ها  ( **KDD** ) علم نسبتا تازه ای است که با توجه پیشرفت کشور در زمینه IT و نگاه های ویژه به دولت الکترونیک و نفوذ استفاده از سیستم های رایانه ای در صنعت و ایجاد بانک های اطلاعاتی بزرگ توسط ادارات دولتی، بانک ها و بخش خصوصی نیاز به استفاده از آن به طور عمیقی احساس می شود. داده کاوی یعنی کشف دانش و اطلاعات معتبر پنهان در پایگاه های داده. یا به بیان بهتر تجزیه و تحلیل ماشینی داده ها برای پیدا کردن الگوهای مفید و تازه و قابل استناد در پایگاه داده های بزرگ ، داده کاوی نامیده می شود. داده کاوی در پایگاه های داده کوچک نیز بسیار پرکاربرد است و از نتایج و الگوهای تولید شده بوسیله آن در تصمیم گیری های استراتژیک تجاری شرکتهای کوچک نیز می توان بهره های فراوان برد. کاربرد داده کاوی در یک جمله را این گونه می توان بیان کرد : " داده کاوی اطلاعاتی می دهد ، که شما برای گرفتن تصمیم هوشمندانه ای درباره مشکلات سخت شغلتان به آنها نیاز دارید".

**مثالی کلاسیک از کاربرد داده کاوی**

اغلب تجارت ها به تصميم گيريهاي استراتژيك و يا اتخاذ خط مشي هاي جديد براي خدمت رساني بهتر به مشتريان نياز دارند. به عنوان مثال فروشگاهها آرايش مغازه خود را براي ايجاد ميل بيشتر به خريد مجدداً طراحي مي كنند. اين  مثال به داده هايي در مورد رفتار مصرفي گذشته مشتريان براي تعيين الگوهايي به وسيله داده كاوي، نياز دارند.

برای روشن تر شدن مسئله می توان مثال را اینگونه بیان کرد که در یک فروشگاه زنجیره ای پس از داده کاوی مشخص میشود که درصدی از مشتریان خرید تلویزیون، میز تلوزیون و گلدان کریستالی را هم در همان روز و بعد از خرید تلوزیون میخرند. مدیر فروشگاه می تواند بلافاصله دستوراتی صادر کند که براساس مدلهای تلويزیون موجود میزهایی و براساس مدل میزها گلدانهای کریستالی برای فروش سفارش داده شود و غرفه های جنبی غرفه تلویزیون را به میز و گلدان کریستالی اختصاص دهد. مطمئنا حتی پس از مدت کوتاهی سود حاصل از این بخش از فروشگاه به طور قابل ملاحضه ای ترقی خواهد کرد.

در واقع ابزار داده كــــاوي، داده را مي گيرد و يك تصوير از واقعيت به شكل مدل مي سازد، اين مدل روابط موجود در داده ها را شرح مي دهد.

برای بهبود بهره وری از یک فروشگاه داده کاوی از داده های انبار داده، مدل هایی را ارائه میدهد که بیانگر این هستند که  چه محصولات يا خدماتي، به چه مشترياني، در چه زماني و از طريق چه كانالي عرضه شود.

بيشتر شركتها، بانكهاي داده اي عظيمي شامل داده هاي بازاريابي، منابع انساني و مالي را دارا هستند. بنابراين، سرمايه گذاري در زمينه انبار داده، يكي از اجزاي حياتي در استراتژي مديريت ارتباط با مشتري است.

رابطه مشتري با زمان تغيير مي كند و چنانچه تجارت و مشتري درباره يكديگر بيشتر بدانند اين رابطه تكامل و رشد مي يابد. چرخه زندگي مشتري چارچوب خوبي براي به كارگيري داده كاوي در مديريت ارتباط با مشتري فراهم مي كند. در بخش ورودي داده كاوي، چرخه زندگي مشتري مي گويد چه اطلاعاتي در دسترس است و در بخش خروجي آن، چرخه زندگي مي گويد چه چيزي احتمالاً جالب توجه است و چه تصميماتي بايد گرفته شود. داده كاوي مي تواند سودآوري مشتري هاي بالقوه را كه مي توانند به مشتريان بالفعل تبديل شوند، پيش بيني كند و اينكه تا چه مدت به صورت مشتريان وفادار خواهند ماند و چگونه احتمالاً ما را ترك خواهند كرد.

بعضي از مشتريان مرتباً مراجعاتشان را به شركتها براي كسب مزيتهايي كه طي رقابت ميان آنها به وجود مي آيد، تغيير مي دهند. در اين صورت شركتها مي توانند هدفشان را روي مشترياني متمركز كنند كه سودآوري بيشتري دارند.

بنابراين مي توان از طريق داده كاوي ارزش مشتريان را تعيين، رفتار آينده آنها را پيش بيني و تصميمات آگاهانه اي را در اين رابطه اتخاذ كرد.

**از کاربرد های داده کاوی می توان به نمونه های زیر اشاره کرد :**

1.        بانکداری :

**·** از جالب توجه ترین کاربرد های داده کاوی می توان به کشف پول شویی اشاره کرد.

**·** تشخیص مشتریان ثابت و همیشگی

**·** تعیین مشتریان استفاده کننده از یک سرویس خاص

2.        بیمه :

**·** پیش گویی میزان استقبال از بیمه نامه های جدید

**·** تشخیص کلاهبرداری ها و مشخص کردن رفتار های نا متناسب

**·** تشخیص نیاز مشتریان و خواسته های آنها

**·** تشخیص تخلفات پزشکی

واضح است که زمینه استفاده از داده کاوی بی نهایت گسترده است. و دو مثال فوق به خاطر درک راحت تر انتخاب شده اند.

داده کاوی شباهت زیادی به تحلیل های آماری دارد. ولی داده کاوی از جهات زیادی با آمار متفاوت است و مزیت های زیادی نسبت به آمار دارد. جالب ترین تفاوت داده کاوی با تحلیل های آماری این است که در آمار ما فرضیه ای طرح می کنیم و با استفاده از تحلیل های آماری به اثبات یا رد فرضیه می پردازیم اما داده کاوی به فرضیه احتیاجی ندارد. در واقع  ابزار داده کاوی فرض می کند که شما خود هم نمی دانید به دنبال چه می گردید. و این نکته ای  است که باعث می شود کار آمدی داده کاوی در مواقع بروز مشکل نمایان شود . برای مثال ما در آمار فرض می کنیم که دو گروه فاصله ای باهم ارتباط دارند سپس با استفاده از ضریب هم بستگی پیرسون مشخص می کنیم که ارتباط وجود دارد یا خیر. ولی داده کاوی بدون توجه به اینکه ما اینگونه فرضی داشته باشیم یا نه با کاوش میان داده ها اگر ارتباطی مخفی معنی داری وجود داشته باشد آن را به اطلاع ما می رساند. تفاوت بعدی آمار و داده کاوی در این است که آمار فقط می تواند از داده های عددی استفاده کند ولی داده کاوی از داده های غیر عددی هم استفاده می کند. تفاوت های دیگری هم میان آمار و داده کاوی وجود دارد که بحث در مورد آنها در حوصله این مقاله نمی گنجد.

اما برای اولین بار در سال 1950 از رایانه برای تحلیل و ذخیره پایگاه داده ها استفاده شد. ولی حجم اطلاعات و میزان رشد آنها به قدری زیاد بوده است که هم اکنون کسی از میزان اطلاعات ذخیره شده در پایگاه داده های سراسر دنیا به صورت دقیق اطلاعی ندارد ولی مطمئنا حجم اطلاعات و مخصوصا سرعت رشد آنها به قدری زیاد شده که آمار شناسان و تحلیل گران در بررسی و تحلیل پایگاههای داده در زمینه های مختلف ناتوانند. بعضی از پایگاه داده ها به قدری بزرگ و پیچیده شده اند که تحلیل روابط و استخراج اطلاعات مفید پنهان شده در آنها واقعا از ظرفیت ذهنی بشری فراتر رفته است. از زمانی که رشد  پایگاه های داده و حجم اطلاعات ، سرعت گرفت و میزان داده ها افزایش یافت ، نیاز به تحلیل ماشینی داده ها و استخراج سریع و دقیق دانش نهفته در آنها احساس شد. شايد بتوان لوول(1983)را اولين شخصي دانست که گزارشي در مورد داده کاوي تحت عنوان «شبيه سازي فعاليت داده كاوي » ارائه نمود.

عمل داده کاوی از یک پایگاه داده به چند مرحله مشخص تقسیم می شود که ما در این مقاله به معرفی و توضیحی مختصر در مورد هر یک از این مراحل اکتفا می کنیم :

**1.** **مرحله اول : تشکیل انبار داده**

    با توجه به عنوان ، این مرحله برای تشکیل محیطی پیوسته و یک پارچه جهت انجام مراحل بعدی و  داده کاوی در آن، انجام می گیرد.در حالت کلی انبار داده مجموعه پیوسته و طبقه بندی شده است که دائما در حال تغییر بوده و دینامیک است که برای کاوش  آماده می شود.

**2.** **مر حله دوم : انتخاب داده ها**

در این مرحله برای کم کردن هزینه های عملیات داده کاوی، داده هایی از پایگاه داده انتخاب می شوند که مورد مطالعه هستند و هدف داده کاوی دادن نتایجی در مورد آنهاست.

**3.** **مرحله سوم : تبدیل داده ها**

مشخص است برای انجام عملیات داده کاوی لزوما باید تبدیلات خاصی روی داده ها انجام گیرد ممکن است این تبدیلات خیلی راحت و مختصر مثل تبدیل byte به  integer باشد یا خیلی پیچیده و زمان بر و با هزینه های بالا مثل تعریف صفات جدید و  یا تبدیل و استخراج داده ها از مقادیر رشته ای و ... باشد.

**4.** **مرحله چهارم : کاوش در داده ها**

در این مرحله است که داده کاوی انجام می شود.در این مرحله با استفاده از تکنیک های داده کاوی داده ها مورد کاوش قرار گرفته ، دانش نهفته در آنها استخراج شده و الگو سازی صورت می گیرد.

**5.** **مر حله پنجم : تفسیر نتیجه**

در این مرحله نتایج و الگو های ارائه شده توسط ابزار داده کاو مورد بررسی قرار گرفته و نتایج مفید معیین می شود.

طرز کار ابزار داده کاو  اینگونه است که ابزار به دنبال اثبات این است که وجود چیزی به معنای وجود چیز دیگری است و  سعی می کند در درجه اول از توالی ارتباطات برای کشف یک الگو بهره بگیرد و در نهایت اطلاعات بدست آمده را دسته بندی کند تا به الگوی خاصی برسد که بتواند آن را براساس فاکتورهای داخی به مخاطبش ارائه دهد.

همچنین در داده کاوی از الگوریتم های ژنتیک و شبکه های عصبی هم استفاده می شود.شبکه های عصبی به علت کار آمدی در حل مسائل پیچیده و بزرگ مورد استفاده اند و کاربرد الگوریتم های ژنتیک در داده کاوی برای جستجو و ساختن یک مدل بهینه در میان مدل های بدست آمده است ، به این گونه که مدل های اولیه روی کرومزوم هایی قرار می گیرند و با رقابت بر سر انتقال صفات به نسل بعد ، بهترین مدل و لایق ترین آنها به کاربر ارائه می شوند.

   داده کاوی امروز گسترش زیادی یافته است به طوری که اکثر نرم افرار های پایگاه داده ای مثل SQL Server  و  ORACLE نیز شامل ابزارهایی داده کاوی شده اند ولی هنوز نرم افزار های تخصصی داده کاوی همچون Intelligent Miner , Darwin , Mine Set, Knowledge Studio, Data Mind از مهمترین ابزار های داده کاوی اند.