

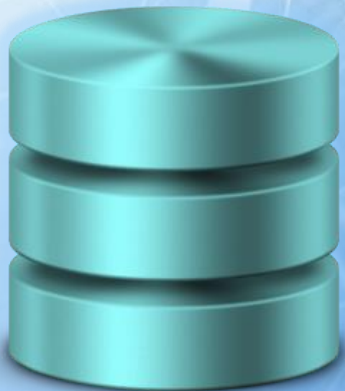
سیستم های اطلاعاتی یکپارچه حسابداری مدیریت





فصل ششم

مفاهیم داده کاوی



❖ حتماً تاکنون بارها عبارت "انقلاب دیجیتال" به گوشتان خورده است و احتمالاً درباره واژه هایی مانند انقلاب دیجیتال، انفجار اطلاعات، عصر رایانه، عصر اطلاعات و ارتباطات و واژه های مشابه، اطلاعاتی نیز دارید. اما چقدر به کاربرد عملی و ملموس این عبارات و مخصوصاً فایده انقلاب دیجیتال در زندگی فکر کرده اید؟

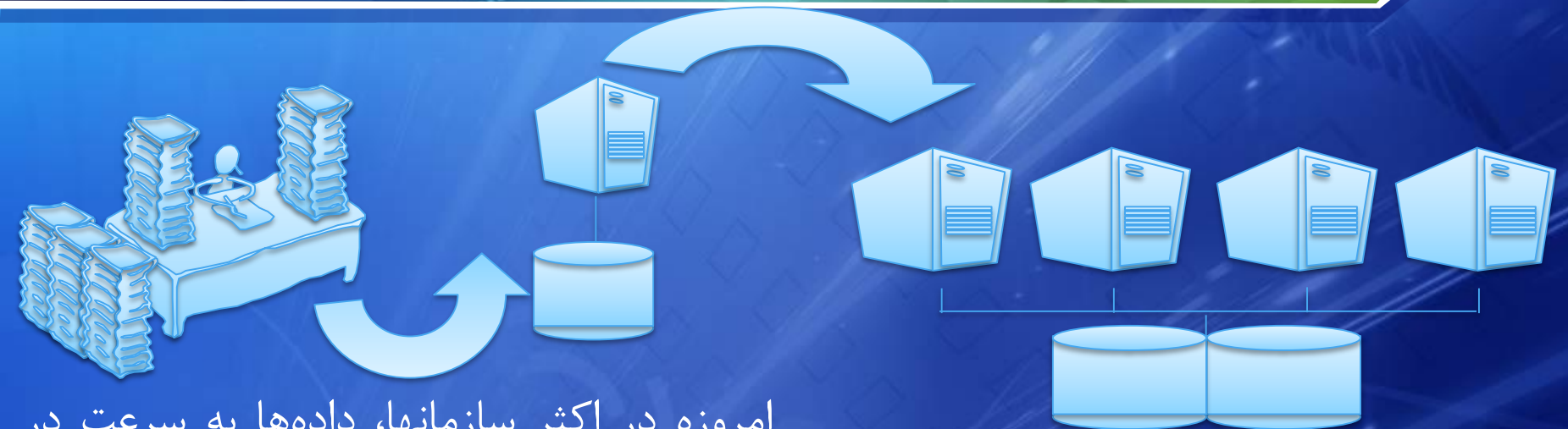


❖ از سال ۱۹۵۰ به بعد که رایانه ، در تحلیل و ذخیره سازی داده ها به کار رفت ، حجم اطلاعات ذخیره شده در آن پس از حدود ۲۰ سال دو برابر شد و همزمان با پیشرفت فناوری اطلاعات در دهه اخیر ، حجم داده ها در پایگاه داده ها هر دو سال یک بار ، دو برابر شد و همچنان با سرعت بیش تری نسبت به گذشته حجم اطلاعات ذخیره شده بیش تر و بیش تر می شود



- در این میان سازمانهایی موفقند که بتوانند حداقل ۷۰٪ داده هایشان را تحلیل کنند. تحقیقات انجام یافته نشان داده است که سازمانها کمتر از درصد داده هایشان را برای تحلیل استفاده می کنند. به عبارت دیگر در حالی که غرق در داده ها هستند تشنه دانش می باشند.





امروزه در اکثر سازمانها، داده‌ها به سرعت در حال جمع آوری و ذخیره شدن می‌باشند.

کم‌کم سازمانها علاقه‌مند شدند تا از این منبع بزرگ و ارزشمند برای رشد و توسعه خود استفاده کنند.



مقدمه

❖ با وجود شبکه جهانی وب ، سیستم های یکپارچه اطلاعاتی ، سیستم های یکپارچه بانکی ، تجارت الکترونیکی و ... لحظه به لحظه به حجم داده ها در پایگاه داده ها اضافه شده و باعث به وجود آمدن انبارهای (توده های) عظیمی از داده ها شده است ، به طوری که ضرورت کشف و استخراج سریع و دقیق دانش از این پایگاه داده ها را بیش از پیش نمایان کرده است



❖ نیاز به طراحی سیستم هایی که قادر به اکتشاف سریع اطلاعات مورد علاقه کاربران با تاکید بر حداقل مداخله انسانی باشند از یک سو و روی آوردن به روش های تحلیل متناسب با حجم داده های حجیم از سوی دیگر ، به خوبی احساس می شود . در حال حاضر ، داده کاوی مهم ترین فناوری برای بهره وری موثر ، صحیح و سریع ازداده های حجیم است و اهمیت آن رو به فزونی است .



تاریخچه

❖ با توجه به وجود اطلاعات ارزشمند در پایگاه های داده ای در اواخر دهه ۸۰ میلادی ، تلاش برای استخراج و استفاده از اطلاعات پایگاه های داده ای شروع شد .



❖ داده کاوی فرایندی است که در آغاز دهه ۹۰ پا به عرصه ظهور گذاشته و با نگرشی نو ، به مسئله استخراج اطلاعات از پایگاه داده ها می پردازد .

تاریخچه

- ❖ در سال ۱۹۸۹ و ۱۹۹۱ کارگاه های کشف دانش از پایگاه داده ها توسط پیاتتسکی و همکارانش و در فاصله سال های ۱۹۹۱ تا ۱۹۹۴ کارگاه های فوق ، توسط فایاد و پیاتتسکی و دیگران برگزار شد .
- ❖ به طور رسمی اصطلاح داده کاوی برای اولین بار توسط « فیاض » در اولین کنفرانس بین المللی « کشف دانش و شناخت داده » در سال ۱۹۹۵ مطرح شد .
- ❖ از سال ۱۹۹۵ داده کاوی به صورت جدی وارد مباحث آمار شد و در سال ۱۹۹۶ ، اولین شماره مجله کشف دانش از پایگاه داده ها منتشر شد .

داده کاوی چیست؟

بطور کلی داده کاوی پل ارتباطی میان علم آمار ، علم کامپیوتر ، هوش مصنوعی ، الگوشناسی ، فراگیری ماشین و بازنمایی بصری داده می باشد. و مانند درختی است که در تکنولوژی های دیگر ریشه دارد



مسائل کامپیوتری
(نموداری) اعداد
visualization

statistics آمار

هوش مصنوعی

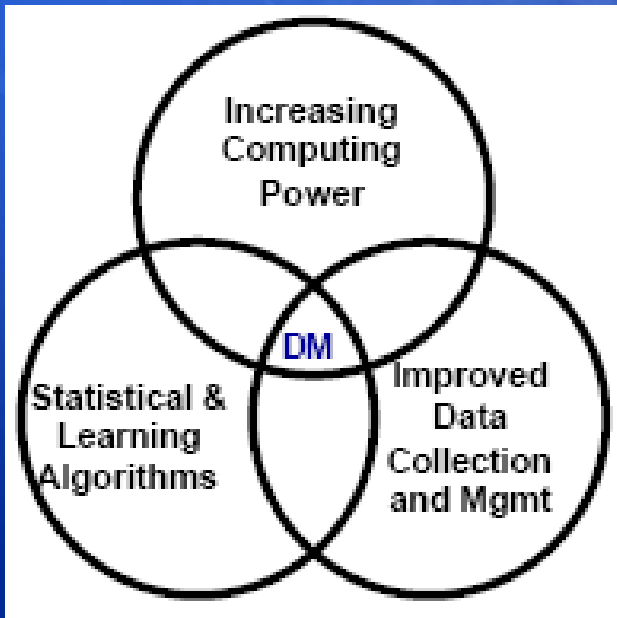
AI(machine Learning)

تکنولوژی داده

Database Technology

داده کاوی چیست؟

❖ داده کاوی یک رشته نسبتاً جدید علمی است که از انجام پژوهش در رشته های مختلف آمار، یادگیری ماشین ، علوم رایانه به خصوص مدیریت پایگاه داده ها شکل گرفته است . البته مرزهای این رشته ها در داده کاوی مبهم و بعضی وقت ها دارای اشتراک های فراوانی هستند



داده کاوی چیست؟

❖ داده کاوی فرآیندی پیچیده جهت شناسایی الگوها و مدل های صحیح، جدید و به صورت بالقوه مفید، در حجم وسیعی از داده می باشد، به طریقی که این الگوها و مدلها برای انسانها قابل درک باشند. داده کاوی به صورت یک محصول قابل خریداری نمی باشد، بلکه یک رشته علمی و فرآیندی است که بایستی به صورت یک پروژه پیاده سازی شود.



تعاریف داده کاوی

❖ نگاهی به ترجمه لغوی داده کاوی به ما در درک بهتر این واژه کمک می کند . واژه لاتین Mine به معنای استخراج از منابع نهفته و با ارزش زمین اطلاق می شود .

❖ ادغام این کلمه با Data به معنی داده

بر جستجوی عمیق از داده های قابل

دسترس با حجم زیاد برای یافتن

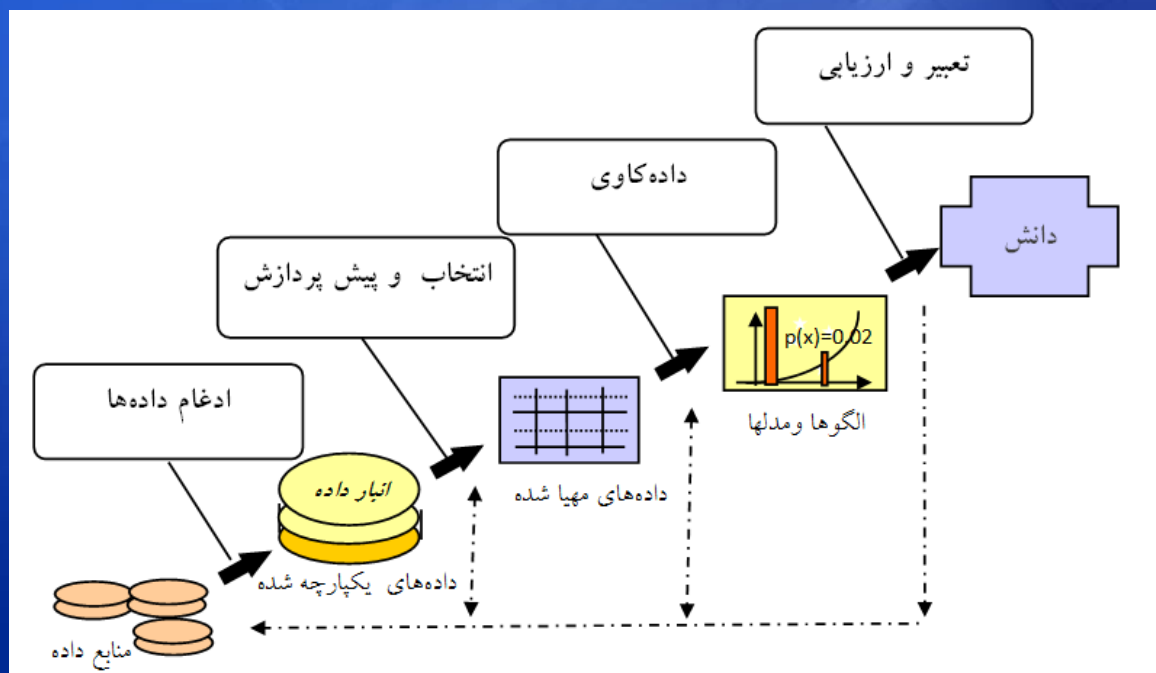
اطلاعات مفید که قبلا نهفته بودند ،

تاکید دارد



تعاریف داده کاوی

❖ داده کاوی عبارت از تحلیل و آنالیز مجموعه داده های بزرگ به منظور یافتن ارتباطات غیر قابل انتظار اما قابل فهم و مفید برای صاحب داده میباشد. این ارتباطات پیدا شده را اصطلاحاً مدل میگویند.



تعاریف داده کاوی

❖ داده ها اغلب حجیم می باشند و به تنهایی قابل استفاده نیستند، اما دانش نهفته در داده ها با داده کاوی قابل استفاده می باشد.

❖ بنابراین بهره گیری از قدرت فرآیند داده کاوی جهت شناسایی الگوها و مدلها و نیز ارتباط عناصر مختلف در پایگاه داده جهت کشف دانش نهفته در داده ها و نهایتا تبدیل داده به اطلاعات، روز به روز ضروری تر می شود.



کاربردهای داده کاوی

❖ CRM مدیریت ارتباط با مشتری (Customer Relationship Management): از کاربردهای کلاسیک داده کاوی است که می توان به



موارد زیر اشاره کرد :

- شناخت مشتریان
- جذب مشتریان
- بقا مشتریان
- توسعه مشتریان

کاربردهای داده کاوی

❖ بطور مثال یک شرکت مخابراتی را در نظر بگیرید که قصد دارد مشتریان خودش را خوشه بندی کند:

خوشه	طول حساب کاربری	پیام صوتی	مکالمه در روز	مکالمه در عصر	مکالمه در شب	مکالمات بین المللی	تعداد
1	62	0	189.2	200.2	209	10.2	852
2	100.7	29.2	178.6	202.3	201.5	10.1	830
3	101.2	0	179.2	202	199.4	10.3	520
4	107.1	31.2	188.8	208.3	204	11	83
5	140.9	0	168	198.4	194.1	10.1	808
6	103	0.5	187.7	202.4	193.8	10.5	240

تحلیل نتایج:

- خوشه ۱: مشتریان جدید
- خوشه ۲: مشتریان پیام صوتی
- خوشه ۳: مشتریان میانگین
- خوشه ۴: مشتریان قدرتمند
- خوشه ۵: مشتریان کم استفاده
- خوشه ۶: مشتریان بین المللی

کاربردهای داده کاوی

بیمه :



تجزیه و تحلیل دعاوی



پیشگویی میزان خرید بیمه نامه های جدید توسط مشتریان



کاربردهای داده کاوی



- ❖ پزشکی :
- ❖ تخمین هزینه بیماران
- ❖ پیش بینی احتمال مبتلا شدن شخصی به بیماری خاص
- ❖ پیش بینی تجویز داروی مناسب برای افراد
- ❖ خوشه بندی بیماران به منظور کشف گروههای بیماری

کاربردهای داده کاوی

○ مثال (تخمین هزینه بیماران): بیمارستانی برای تخمین هزینه درمانی بیماران سرطانی، اطلاعات ۱۰۰۸ نفر از بیماران را جمع آوری نمود. این مجموعه داده شامل ۱۲ متغیر است. این بیمارستان قصد دارد هزینه درمانی بیماران را پیش بینی کند.

○ هدف: تخمین هزینه درمان

○ ابزار: کلاسه بندی و پیش بینی

○ داده ها: جنسیت، وضعیت تاهل، میزان فعالیت، تعداد دفعات شیمی درمانی، هزینه درمانی و ...

○ نتایج:

○ دقت شبکه های عصبی در این پیش بینی ۹۷٪ گردید. شبکه های عصبی به منظور پیش بینی هزینه های درمانی، از دقت بالاتری نسبت به روش هایی مانند درخت تصمیم بهره مند می باشند.

کاربردهای داده کاوی



❖ حوزه کاربردی فضا و سفرهای فضائی

✓ حجم بسیار زیادی از اطلاعات

✓ نویز بسیار بالا

✓ ارزش بسیار زیاد دانش قابل استخراج

✓ پردازش اطلاعات جمع آوری شده از فضا

✓ پردازش اطلاعات مربوط به سفینه های فضائی

✓ ارائه دانش مفید برای اتخاذ تصمیم نهائی جهت پرتاب یا عدم پرتاب یک

سفینه به فضا

کاربردهای داده کاوی



❖ حوزه اطلاعات جغرافیائی و اقلیمی

✓ کشف پدیده های اقلیمی جدید

✓ تکنیکهای بصری سازی و بازنمایی اطلاعات

✓ پردازش انواع اطلاعات (تصاویر، اطلاعات به دست آمده از سنجشگرها)

کاربردهای داده کاوی

❖ بانکداری :

■ پیش بینی الگوهای کلاهبرداری از طریق کارتهای اعتباری، تشخیص مشتریان ثابت،

تعیین میزان استفاده از کارتهای اعتباری، بر اساس گروههای اجتماعی

❖ وضعیت : شخصی برای یک وام اقدام می کند

❖ وظیفه : آیا بانک اعطای وام را تصویب نماید؟

■ توجه : افرادی که بالاترین موجودی را دارند
نیازی به وام ندارند، و افرادی که پایین ترین
موجودی را دارند احتمالاً قدرت بازپرداخت را
ندارند. بهترین مشتریان بانک در میانه این دو
هستند.



کاربردهای داده کاوی

❖ مثال: پیش بینی میزان ریسک اعتباری در بانک

❖ برای مدیران بانک ها در بسیاری مواقع این امر مهم می باشد که بتوانند میزان ریسک اعتباری مشتریان خود را پیش بینی کنند. برای این منظور ممکن است مشخصه های مربوط به مشتریان جمع آوری شوند.

شماره	عنوان
۱	سن
۲	میزان تحصیلات
۳	تعداد سالهایی که در مکان فعلی مشغول به کار است
۴	تعداد سالهایی که در مکان فعلی سکونت دارد
۵	درآمد سالانه
۶	نسبت بدهی ها به درآمد
۷	میزان بدهی کارت اعتباری
۸	قصور در پرداخت های پیشین

کاربردهای داده کاوی

❖ بانک مرکزی آمریکا در طی ۲ سال ۴.۸ میلیارد دلار صرفه جویی نمود.



❖ این بانک با استفاده از سیستم مدیریت ریسک اعتباری بر پایه داده کاوی که توسط شرکت **SAS** طراحی شده بود، سیستم پیگیری برخطی را برای تخصیص برخط میزان ریسک اعتباری مشتریان خود طراحی نمود.

کاربردهای داده کاوی

- پیش بینی میزان تاثیر تبلیغات بانکی
- بانک ها سالانه هزاران دلار صرف بازاریابی مستقیم خود می نمایند که این تبلیغات شامل موارد زیر هستند:
 - تبلیغات خیابانی
 - ارسال کاتالوگ ها
 - ارسال نامه های تبلیغاتی
 - پخش تراکت
- بعد از اعمال الگوریتم های داده کاوی نتایج زیر بدست می آید:
 - به بانک کمک کرد تا مشتریان خود را بهتر بشناسد.
 - روش های تبلیغاتی را پیشنهاد کرد که هم کارا تر بودند و هم سود بیشتری را برای بانک فراهم ساختند.
 - نرخ واکنش مشتریان ۵٪ افزایش پیدا کرد و هزینه ارسال تبلیغات به میزان ۲۰٪ کاهش پیدا کرد.

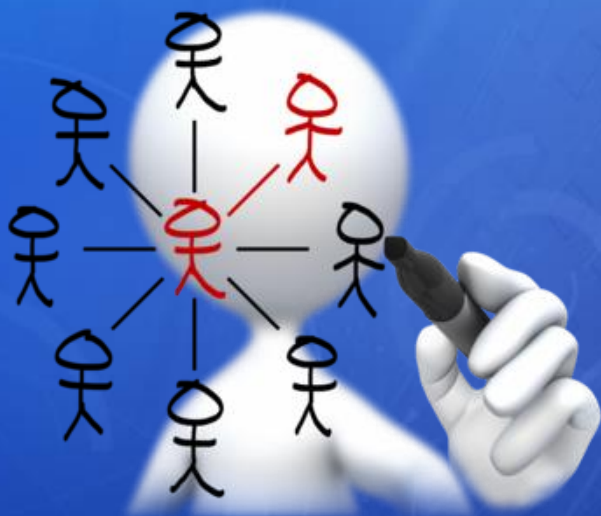
کاربردهای داده کاوی: جمع بندی

- ✓ حوزه های اصلی شامل کاربردهای علمی، تجاری و امنیتی می باشد.
- ✓ در تمام حوزه ها با حجم بسیار زیاد اطلاعات و خصایص متعدد مواجهیم.
- ✓ در تمام حوزه ها با انواع اطلاعات مفید روبرو هستیم.
- ✓ کاهش شدید هزینه ها، افزایش درآمدها و نجات زندگی انسانها از دستاوردهای داده کاوی در هریک از حوزه های کاربردی آن است.
- ✓ کاربردهای تجاری: تشخیص صحت ادعای خسارت در بیمه، تشخیص سوء استفاده از کارتهای اعتباری، تحلیل اطلاعات مشتریان یک سازمان،...
- ✓ کاربردهای علمی: حوزه های پزشکی، جغرافیائی و اقلیمی، فضا و سفرهای فضائی
- ✓ کاربردهای امنیتی: مبارزه با تروریسم، مقابله با نفوذگران به شبکه های کامپیوتری



هدف داده کاوی

❖ داده کاوی یک فرایند است و هدفش این است که بتواند یکسری الگو هائی از اعداد بدست بیاورد که خاصیت های زیر را داشته باشند:



❖ **Valid** معتبر باشد

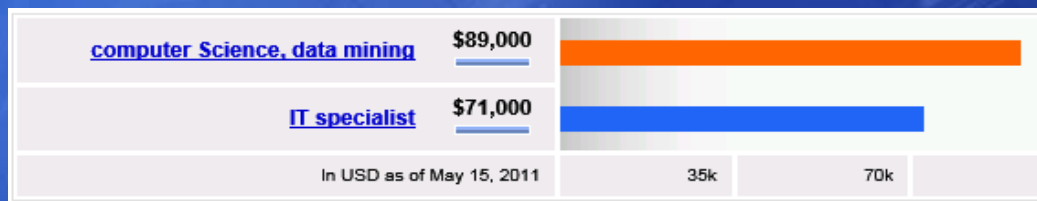
❖ **Useful** مفید باشد

❖ **Novel** جدید باشد

❖ **Understandable** قابل فهم باشد

جایگاه شغلی داده کاوی

- ❖ درآمد شغل مربوط به داده کاوی به طور میانگین ۲۱٪ از سایر شغل ها بیشتر می باشد.
- ❖ درآمد شغل داده کاوی ۲۵٪ بیشتر از میانگین درآمد شغل های مربوط به فناوری اطلاعات می باشد.
- ❖ درآمد شغل های مربوط به فناوری اطلاعات ۴٪ زیر میانگین درآمد سایر شغل ها می باشد.



جایگاه علم داده کاوی

❖ بنابر اعلام دانشگاه MIT دانش نوین داده کاوی (Data mining) یکی از ده دانش در حال توسعه ای

است که دهه آینده را با انقلاب تکنولوژیکی مواجه می سازد. این تکنولوژی امروزه دارای کاربرد بسیار

وسعی در حوزه های مختلف است به گونه ای که امروزه حد و مرزی برای کاربرد این دانش در نظر

نگرفته و زمینه های کاری این دانش را از ذرات کف اقیانوسها تا اعماق فضا می دانند .



**Massachusetts
Institute of
Technology**