

فرآیند استخراج اطلاعات در داده‌های بزرگ با استفاده از الگوریتم‌های ریاضی داده کاوی نام دارد .

راز داده کاوی این است که با داشتن شناخت از گذشته می توان آینده را پیش بینی کرد چرا که گذشته شباهت زیادی به آینده دارد.

داده کاوی از بین حجم زیادی از داده ها به دنبال الگوهای پنهان و ارتباط بین آنها می باشد تا در نهایت اطلاعاتی اجرایی در اختیار قرار دهد .

داده کاوی تلفیقی از دانش آمار ، هوش مصنوعی ، یادگیری ماشین و مدیریت سیستم های اطلاعاتی می باشد .

الگوریتم های مورد استفاده در داده کاوی شامل موارد زیر می باشند :

درخت تصمیم

(Rule base)

روش های قانون

پایه

- نزدیکترین همسایه
- شبکه بیزین
- شبکه عصبی
- در پایگاه داده های

OLAP

جهت بررسی دقیق تر داده ها از تکنیک های داده کاوی استفاده می شود .

Support Vector Machines

ارائه مدل بخشی از فرایند داده کاوی می باشد که به معنی طرح سوال در خصوص داده ها و ایجاد و پیاده سازی مدل در فضای عملیاتی می باشد :

- تعریف مساله
- آماده سازی داده ها
- شناسایی داده ها
- ارائه مدل
- شناسایی و تایید مدل ها
- پیاده سازی مدل ها

قبل از انجام عملیات داده کاوی نیاز است تا پردازشی بر روی داده ها به منظور تفکیک داده های ناکارآمد از داده های مفید صورت گیرد :

حذف داده های تکراری .

جایگزینی داده های مفقود شده .

تصمیم گیری در رابطه با داده های پراکنده .

فشرده سازی داده ها .

کاهش ابعاد داده ها .

کاربرد های داده کاوی بسیار وسیع هستند که تعدادی از آنها به شرح زیر می باشند :

سیستم های اقتصادی ، بانکی ، مدیریتی ، نرم افزارهای امنیتی ، علوم سیاسی و پزشکی و ...

برای مثال به کاربری تکنیک های داده کاوی (

Data Mining

) در یک سیستم مدیریتی نظیر مدیریت ارتباط با مشتری (

CRM

(

به سازمان کمک می کند تا در برقراری ارتباط موثر با مشتریان موفق عمل نماید چرا که مشتری محور اصلی هر سازمان محسوب می شود .

- می توان رفتار مشتریانی که از کمپین های تخفیف استقبال نموده اند را تحلیل کرده و به دلیل عدم استقبال سایر مشتریان پی برد .
- علت پیروزی یک کمپین را شناسایی نمود .

تعدادی از نرم افزار کاربردی حوزه داده کاوی شامل موارد زیر می باشند :

Weka

RapidMiner

۳

Orange

علاوه بر نرم افزارها ، از زبان‌های برنامه نویسی مانند R نیز می‌توان استفاده نمود استفاده از آن‌ها قابلیت‌های بیشتری را در اختیار قرار می‌دهد که می‌توان پیاده سازی مدل‌های مختلف را توسط آن انجام داد.

