Important Data Mining Techniques

Data mining is the process of looking at large banks of information to generate new information. Intuitively, you might think that data "mining" refers to the extraction of new data, but this isn't the case; instead, data mining is about extrapolating patterns and new knowledge from the data you've already collected.

Relying on techniques and technologies from the intersection of database management, statistics, and machine learning, specialists in data mining have dedicated their careers to better understanding how to process and draw conclusions from vast amounts of information. But what are the techniques they use to make this happen?

Data Mining Techniques

Data mining is highly effective, so long as it draws upon one or more of these techniques:

1. Tracking patterns. One of the most basic techniques in data mining is learning to recognize patterns in your data sets. This is usually a recognition of some aberration in your data happening at regular intervals, or an ebb and flow of a certain variable over time. For example, you might see that your sales of a certain product seem to spike just before the holidays, or notice that warmer weather drives more people to your website.

2. Classification. Classification is a more complex data mining technique that forces you to collect various attributes together into discernable categories, which you can then use to draw further conclusions, or serve some function. For example, if you're evaluating data on individual customers' financial backgrounds and purchase histories, you might be able to classify them as "low," "medium," or "high" credit risks. You could then use these classifications to learn even more about those customers.

3. Association. Association is related to tracking patterns, but is more specific to dependently linked variables. In this case, you'll look for specific events or attributes that are highly correlated with another event or attribute; for example, you might notice that when your customers buy a specific item, they also often buy a second, related item. This is usually what's used to populate "people also bought" sections of online stores.

4. Outlier detection. In many cases, simply recognizing the overarching pattern can't give you a clear understanding of your data set. You also need to be able to identify anomalies, or outliers in your data. For example, if your purchasers are almost exclusively male, but during one strange week in July, there's a huge spike in female purchasers, you'll want to investigate the spike and see what drove it, so you can either replicate it or better understand your audience in the process.

5. Clustering. Clustering is very similar to classification, but involves grouping chunks of data together based on their similarities. For example, you might choose to cluster different demographics of your audience into different packets based on how much disposable income they have, or how often they tend to shop at your store.

6. Regression. Regression, used primarily as a form of planning and modeling, is used to identify the likelihood of a certain variable, given the presence of other variables. For example, you could use it to project a certain price, based on other factors like availability, consumer demand, and competition. More specifically, regression's main focus is to help you uncover the exact relationship between two (or more) variables in a given data set.

7. Prediction. Prediction is one of the most valuable data mining techniques, since it's used to project the types of data you'll see in the future. In many cases, just recognizing and understanding historical trends is enough to chart a somewhat accurate prediction of what will happen in the future. For example, you might review consumers' credit histories and past purchases to predict whether they'll be a credit risk in the future.

Data Mining Tools

So do you need the latest and greatest machine learning technology to be able to apply these techniques? Not necessarily. In fact, you can probably accomplish some cutting-edge data mining with relatively modest database systems, and simple tools that almost any company will have. And if you don't have the right tools for the job, you can always create your own.

However you approach it, data mining is the best collection of techniques you have for making the most out of the data you've already gathered. As long as you apply the correct logic, and ask the right questions, you can walk away with conclusions that have the potential to revolutionize your enterprise.