

How Data Mining Works

Using existing and relevant data, data mining builds models to identify patterns among the attributes presented in the data set. Models are the mathematical representations (simple linear relationships and/or complex highly nonlinear relationships) that identify the patterns among the attributes of the objects (e.g., customers) described in the data set. Some of these patterns are explanatory (explaining the interrelationships and affinities among the attributes), whereas others are predictive (foretelling future values of certain attributes). In general, data mining seeks to identify four major types of patterns:

1. *Associations* find the commonly co-occurring groupings of things, such as beer and diapers going together in market-basket analysis.
2. *Predictions* tell the nature of future occurrences of certain events based on what has happened in the past, such as predicting the winner of the Super Bowl or forecasting the absolute temperature of a particular day.

Application Case 4.2

Harnessing Analytics to Combat Crime: Predictive Analytics Helps Memphis Police Department Pinpoint Crime and Focus Police Resources

When Larry Godwin took over as director of the Memphis Police Department (MPD) in 2004, crime across the metro area was surging, and city leaders were growing impatient. "The mayor told me I want this crime problem fixed," recalls Godwin, a 38-year veteran of the MPD. But the new director understood that a business-as-usual approach to crime fighting would no longer be good enough. Early on in his tenure, Godwin convened a meeting of top law enforcement experts to formulate a fresh strategy to turn the tide in the city's crime war. Among the participants in this mini-summit was Dr. Richard Janikowski, a professor of criminology at the University of Memphis, who specialized in using predictive analytics to better understand patterns.

Fighting Crime with Analytics

Janikowski proposed the idea of mining MPD's crime data banks to help zero in on where and when criminals were hitting hardest and then "focus police resources intelligently by putting them in the right place, on the right day, at the right time." By doing so, he said, "you'll either deter criminal activity or you're going to catch people." The idea made sense to Godwin and in short order the MPD and the University of Memphis—along with Project Safe Neighborhoods—teamed up in a pilot program that later became known as Operation Blue CRUSH, or Crime Reduction Utilizing Statistical History.

The data-driven pilot was wildly successful. During one 2-hour operation, officers arrested more criminals than they normally apprehended over an entire weekend. But for Blue CRUSH to be successful on a citywide scale, the MPD would need to align its resources and operations to take full advantage of the power of predictive analytics. If done right, a citywide rollout of Blue CRUSH had the potential to save money through efficient deployments—a big plus in a city facing serious budget pressures—even as the intelligence-based approach would help drive down overall crime rates. Shortly after, all precincts embraced Blue CRUSH, and predictive analytics has become one of the most potent weapons in MPD's crime-fighting arsenal. At the heart of the system is a versatile statistical analysis tool—IBM SPSS Modeler—that enables officers to unlock the intelligence hidden in the department's huge digital library of crime records and police reports going back nearly a decade.

Safer Streets

All indications are that Blue CRUSH and its intelligence-driven crime-fighting techniques are putting a serious dent in Memphis-area crime. Since the program was launched, the number of Part One crimes—a category of serious offenses including homicide, rape, aggravated assault, auto theft, and larceny—has plummeted, dropping 27 percent from

(Continued)

Application Case 4.2 (Continued)

2006 to 2010. Intelligent positioning of resources has been a major factor in the decline, helping to deter criminal activity by having more officers patrolling the right area at the right time on the right day.

More intelligent deployments also lead to faster reaction time, since officers are likely to be better positioned to respond to an unfolding crime. In addition, MPD's organized crime units are using data from the predictive analytics solution to run special details that lead to successful multi-agency drug busts and other criminal roundups. Not surprisingly, arrest rates have been steadily improving across the Memphis area, which has a population of 680,000.

Today, the MPD is continuing to explore new ways to exploit statistical analysis in its crime-fighting mission. Of course, predictive analytics and data mining is just one part of MPD's overall strategy for keeping Memphis residents safe. Effective liaisons

with community groups and businesses, strong partnerships with regional and federal law enforcement agencies, and intelligent organizational and operational structures all play a part in continuing MPD's success story. "At the end of the day, everybody wants to reduce crime," says Godwin. "Everybody wants a safe community because without it, you don't have economic growth, you don't have anything."

QUESTIONS FOR DISCUSSION

1. How did the Memphis Police Department use data mining to better combat crime?
2. What were the challenges, the proposed solution, and the obtained results?

Source: public.dhe.ibm.com/common/ssi/ecm/en/imc14541usen/IMC14541USEN.PDF (accessed September 2013).

3. *Clusters* identify natural groupings of things based on their known characteristics, such as assigning customers in different segments based on their demographics and past purchase behaviors.
4. *Sequential relationships* discover time-ordered events, such as predicting that an existing banking customer who already has a checking account will open a savings account followed by an investment account within a year.

These types of patterns have been *manually* extracted from data by humans for centuries, but the increasing volume of data in modern times has created a need for more automatic approaches. As data sets have grown in size and complexity, direct manual data analysis has increasingly been augmented with indirect, automatic data processing tools that use sophisticated methodologies, methods, and algorithms. The manifestation of such evolution of automated and semiautomated means of processing large data sets is now commonly referred to as *data mining*.

Generally speaking, data mining tasks can be classified into three main categories: prediction, association, and clustering. Based on the way in which the patterns are extracted from the historical data, the learning algorithms of data mining methods can be classified as either supervised or unsupervised. With supervised learning algorithms, the training data includes both the descriptive attributes (i.e., independent variables or decision variables) as well as the class attribute (i.e., output variable or result variable). In contrast, with unsupervised learning the training data includes only the descriptive attributes. Figure 4.3 shows a simple taxonomy for data mining tasks, along with the learning methods, and popular algorithms for each of the data mining tasks.

PREDICTION **Prediction** is commonly referred to as the act of telling about the future. It differs from simple guessing by taking into account the experiences, opinions, and other relevant information in conducting the task of foretelling. A term that is commonly associated with prediction is *forecasting*. Even though many believe that these two terms are