

Data mining: On the trail to marketing gold

Shawn Thelen

Assistant Professor of Marketing and International Business, Hofstra University, Hempstead, New York (mktstt@hofstra.edu)

Sandra Mottner

Assistant Professor of Marketing, College of Business and Economics, Western Washington University (sandra.mottner@wwu.edu)

Barry Berman

Walter H. "Bud" Miller Distinguished Professor of Business, Hofstra University, Hempstead, New York (mktbxb@hofstra.edu)

What is data mining, and how does it differ from traditional statistical modeling? Along with finding the answers here, managers can take a look at important recent developments in data mining, examine some of its marketing-related applications, and learn how to establish and maintain a data mining system. Armed with this information, they can then determine their firm's level of commitment to the process. Companies that do not want to commit the financial and personnel resources to data mining can still secure many of its benefits through outsourcing.

Most marketing managers are comfortable dealing with samples, using Excel spreadsheets, and evaluating output from sales analyses. However, few have had much exposure to data mining. A study by Calderon, Chen, and Kim (2003) of the use of data mining among *Fortune* 500 companies found that 35 percent did not employ it. Four reasons were cited by 73 percent of the nonusers: faced with a more pressing business problem (36 percent), not seen as cost effective (18 percent), lack of familiarity (9 percent), and lack of skilled personnel (9 percent).

Fayyad (2003) reports on a recent study from The Cutter Corporation, an IT consulting firm, which found that only 15 percent of the companies surveyed said their data warehousing projects have been a success. And a business strategist for WWAV Rapp Collins, the UK's largest direct marketing agency, reports that clients that have implemented customer relationship management (CRM) data mining programs are using only 18 percent of the potential of their databases, according to Reed (2002).

Nevertheless, data are the digital currency of the new century, so knowing how to mine them and extract value from them is critical for business—and will become more so in the future. Not all companies are equipped to deal with the enormous volume or complexity of data that can be acquired and stored today, or to fully understand or appreciate some of the new statistical methods for using those data.

Besides defining data mining here and comparing it to traditional statistical modeling, we shall describe important recent developments, illustrate marketing-related applications, and discuss how a firm can establish and maintain a data mining system. We will also focus on successful data mining strategies and show managers how to avoid potential pitfalls. Much of our discussion is also applicable to those firms that cannot devote adequate financial and personnel resources to data mining, yet want to secure its benefits for themselves through outsourcing some or all of their data mining operations.

Data mining vs. traditional statistical modeling

A sophisticated type of decision support system, data mining uses automated statistical analyses to sift through a company's business and customer databases to uncover hidden problems and opportunities. Many large companies today have terabytes of data, within which they could probably find more information about their customers, markets, and competition than they would ever need. Data mining enables marketers to better

Data mining is a potential solution to a big problem facing many firms: an overabundance of data and a relative dearth of staff, technology, and time to transform numbers and notes into meaningful information about existing and prospective customers.

extract valuable business information from the "mountains of data" in a firm's systems. It is a potential solution to a big problem facing many companies: an overabundance of data and a relative dearth of staff, technology, and time to transform numbers and notes into meaningful information about existing and prospective customers.

Data mining works with data from disparate sources that may be located in several departments of an organization (sales, credit, inventory) and in different platforms (mainframe, UNIX, PC). Besides providing the tools to extract, transform, and load this disparate information into a common database repository, it enables a firm to measure consumer behavior on the basis of 100 or more attributes, instead of the three or four associated with traditional statistical modeling. The more attributes a firm uses, reports Jones (2002), the greater the complexity of the data and the greater the need for data mining tools.

Whereas traditional statistical analysis relies on hypothesis testing, data mining focuses on machine-driven model building, identifying relationships and interdependencies that affect a marketing-related problem or opportunity.

Some of the techniques are closely related to those in traditional statistics, such as linear regression and time series analysis, but the models are often richer and more flexible. Two data mining techniques, neural networks and decision trees, can each handle up to 200 predictor variables. In contrast, a traditional technique such as multiple regression cannot cope with this level of complexity. The ability to handle a large number of variables enables data mining to determine which variables from a much larger set are most closely associated with the problem or opportunity.

With traditional statistical modeling, an analyst would pose a question such as: "Are higher-income people prone to be more loyal to a warehouse club than those with lower income levels?" and the hypothesis would either be supported or unsupported. Data mining, on the other hand, could potentially provide more insight by pointing out other factors contributing to store loyalty that the analyst would not otherwise have been able to consider testing. The process might group customers who have a warehouse club's credit cards, who live within 10 miles of the club, who own more than one car, and who have a separate freezer. And although managers using data mining need to understand some basic statistical principles, highly specialized training in statistics is not necessary.

Data mining models typically consist of responses and predictors. For example, an auto insurance company collects information about its clients such as family size, households with excellent credit ratings, and so on. Known as *predictors*, these factors can be used to estimate the loss ratio among selected consumer groups, or to determine which type of auto customer is most likely to purchase other products—the *responses* to the predictors. As King and Linden (2002) explain, those employing data mining techniques seek to build models that can reliably estimate the response, given that the appropriate predictors have been collected.

Recent developments in data mining

With the exception of large companies that have separate market research departments staffed with data mining personnel, many firms have been ill-equipped to deal with the enormous volume or complexity of data that can be acquired and stored today. But advances in both hardware and the capabilities of database management systems are making data mining much more economically feasible. For example, the lower cost of disk storage enables firms to store more data per dollar of equipment. Microprocessors are becoming more powerful, and advances in symmetrical multiprocessor technology have removed a lot of the overhead that once

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات