

DATA MINING: A COMPETITIVE WEAPON FOR BANKING AND RETAIL INDUSTRIES

Amir M. Hormozi and Stacy Giles

Data mining is proving to be a valuable tool, by identifying potentially useful information from the large amounts of data collected, and enabling an organization to gain an advantage over its competitors. This article defines what data mining is and covers the operations of data mining as have been classified throughout literature. It then focuses on the important uses of data mining, which include but are not limited to marketing, risk management, fraud detection, and customer acquisition and retention. Examples of how the banking and retail industries have been effectively utilizing data mining in these areas are provided.

TODAY, ORGANIZATIONS ARE REALIZING the advantages that come with the utilization of data mining. Data mining is proving to be a valuable tool, by identifying potentially useful information from the large amounts of data collected, and enabling an organization to gain an advantage over its competitors. A report from Forrester Research stated that of the Fortune 1000 companies, 52 percent were planning to utilize data mining in their marketing strategies in 2001, which is an increase from 18 percent in 1999. Also, 48 percent were planning to use data mining to improve customer service, which is an increase from 16 percent in 1999 (Le Beau, 2000).

The increase in the amount of data being collected is not the only reason for companies to look to data mining. According to Mitchell (1999), several recent trends have increased the interest in data mining, including the declining cost of data storage and the increasing ease of collecting data, the development of robust and efficient machine learning algorithms to process data, and the declining cost

of computational power. With greater data storage capabilities and declining costs, data mining has offered organizations a new way of doing business. Data mining can help organizations better understand their business, be able to better serve their customers, and increase the effectiveness of the organization (Chopoorian et al., 2001).

Today, data mining is being used throughout many industries, including financial institutions, retail, insurance, and telecommunications. According to Crook et al. (2001), in the past few years many organizations, especially retailers and banks, have recognized the importance of the information they have on their customers.

This article begins with what data mining is, the operations of data mining, and the uses of data mining. It then specifically focuses on the banking and retail industries and how they have been effectively utilizing data mining.

WHAT IS DATA MINING?

Data mining has become a widely accepted process for organizations to enhance their

AMIR M. HORMOZI is a professor of operations management at Texas A&M University — Corpus Christi.

STACY GILES is an Internet marketing specialist at USDM.net, a leading marketing and technology company for the tourism industry. She specializes in online marketing and account management.

TABLE 1 Data Mining Defined throughout the Literature

Fayyad et al. (1996a)	Data mining is a step in the knowledge discovery in databases (KDD) process and refers to algorithms that are applied to extract patterns from the data. The extracted information can then be used to form a prediction or classification model, identify trends and associations, refine an existing model, or provide a summary of the database being mined.
Cabena et al. (1998)	Data mining is defined as "the process of extracting previously unknown, valid, and actionable information from large databases and then using the information to make crucial business decisions."
Fabris (1998)	Data mining is described as the automated analysis of large amounts of data to find patterns and trends that may have otherwise gone undiscovered.
Chung and Gray (1999)	"The objective of data mining is to identify valid, novel, potentially useful, and understandable correlations and patterns in existing data."
Hui and Jha (1999)	Data mining is the process of discovering interesting knowledge from large amounts of data that can be used to help companies make better decisions and remain competitive in the marketplace.
Berry and Linoff (2000)	Data mining is "the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules."

organizational performance and gain a competitive advantage. Because the data mining process is a relatively new concept, it has been defined in various ways by various authors in the recent past. Table 1 provides a few definitions from the literature on what data mining is and how it can be beneficial to an organization.

The definitions of data mining given by these authors are somewhat different but all have the same idea: to extract important information from existing data and enable better decision making throughout an organization. Not only can data mining improve decision making by searching for relationships and patterns from the extensive data collected by organizations, but it can also reduce information overload (Zhu et al., 2001). Data mining enables an organization to focus on the most important information in the database, which allows managers to make more knowledgeable decisions by predicting future trends and behaviors (Chopoorian et al., 2001). Before data mining was utilized, managers were not as capable of making such informed decisions because searching through large amounts of data was too expensive and time-consuming.

DATA MINING OPERATIONS

Throughout the literature, data mining operations are also classified in different ways. The following are a few categorizations of data mining operations, including (1) clustering/segmentation, (2) visualization, (3) predictive modeling, (4) link analysis, (5) deviation detection, (6) dependency modeling, and (7) summarization.

- *Clustering/segmentation.* Clustering, or segmentation, is a method of grouping data into classes (or clusters) that share similar patterns or characteristics (Fayyad et al., 1996b; Cabena et al., 1998; Groth, 1998; Bradley et al., 1999; Yoon, 1999; Berry and Linoff, 2000; Chye and Gerry, 2002). The clusters are determined from the data, instead of relying on predefined classes as in classification (Fayyad et al., 1996b; Groth, 1998; Bradley et al., 1999; Yoon, 1999; Berry and Linoff, 2000; Chye and Gerry, 2002). The miner is the one who must determine if there is any meaning to attach to the resulting clusters (Berry and Linoff, 2000). According to Groth (1998), retailers can use clustering to discover where similarities exist in their customer base so they can create and understand different groups to which they sell and market.
- *Visualization.* Visualization is the graphical representation of data (Groth, 1998; Chye and Gerry, 2002). It can mean a lot more than two-dimensional charts and maps by bringing out points that you might not normally see (Groth, 1998). According to Chye and Gerry (2002), visualization can contribute toward a greater understanding of a data set and help to detect hidden patterns in data, especially intricate data containing complex and nonlinear interactions. Berry and Linoff (2000) add that because humans are extremely practiced at extracting meaning from visual scenes, visualization can be very helpful.
- *Predictive modeling.* According to Chye and Gerry (2002), the most common and important applications in data mining involve prediction. Bradley et al. (1999) state that

A good model can enable a better understanding of the customers and allow prediction of which customers will stay and which will leave.

predictive modeling has the goal of predicting a specific attribute based on other attributes in the data. According to Cabena et al. (1998), predictive modeling is similar to the human learning experience, where observations are used to form a model of the essential, underlying characteristics of an occurrence. In data mining, predictive modeling is used to analyze a database to determine some essential characteristics about the data. The predictive modeling approach is applicable across many industries. Examples of business applications that it supports are customer retention management, credit approval, cross selling, and target marketing. Cabena et al. (1998), Bradley et al. (1999), and Chye and Gerry (2002) include classification as a part of their predictive modeling category. They conclude that classification has the objective of predicting the most likely state of a variable of interest when given the values of other variables. According to Groth (1998), a company can use classification to help predict which customers they are more likely to lose to competitors by constructing a model derived from historical data of customers that are loyal versus those who have left. A good model can enable a better understanding of the customers and allow prediction of which customers will stay and which will leave. Yoon (1999) states that regression can also be used for prediction. Regression maps a data item into a real-valued prediction variable using the databases (Fayyad et al., 1996b; Yoon, 1999). Estimation is another part of predictive modeling. According to Chye and Gerry (2002), estimation refers to the prediction of a target variable that is metric in nature (e.g., predicting the amount spent, duration of a call, or the account balance).

□ *Link analysis.* The link analysis operation has the objective of identifying links or connections between the records in a database (Fayyad et al., 1996b; Cabena et al., 1998; Yoon, 1999). According to Cabena et al. (1998), there are three specializations of link analysis: associations discovery, sequential pattern discovery, and similar time sequence discovery. Associations discovery can be used to analyze goods purchased in a visit to a shop to determine which products tend to sell well together, which is called market basket analysis (MBA) or product affinity analysis. Sequential pattern discovery is used to understand long-term customer buying behavior by identifying associations across

related purchase transactions over time that reveal information about the sequence in which consumers purchase goods and services. Similar time sequence discovery is the discovery of links between two time-dependent sets of data and is based on the degree of similarity between the patterns that both time series demonstrate. Retailers can determine whether a product with a specific pattern of sales over time matches the sales curve of other products, even if the pattern match is some time behind. Yoon (1999) also states that market basket analysis is a typical application of link analysis in which the technique is applied to analyze point-of-sales transaction data to identify product affinities. Retailers are able to analyze transactions with the use of scanners and find out what items sell well together, such as baby's diapers and formula, so they can determine what items to display together for effective marketing.

□ *Deviation detection.* According to Cabena et al. (1998) and Yoon (1999), deviation detection is often the source of true discovery because outliers express deviation from an expectation and norm that was previously known. Yoon (1999) gives various types of deviation, including unusual patterns that do not fit into a previously measured class, significant changes in data from one time period to the next, outlying points in a dataset or records that do not belong to any particular cluster, and discrepancies between an observation and a reference. One of the distinctive characteristics of deviation detection noted by Bradley et al. (1999) is that the ordering of observations is significant and must be accounted for. Cabena et al. (1998) maintain that deviation detection supports several business applications, including fraud detection in the use of credit cards, insurance claims, and telephone cards; quality control; and defects tracing.

□ *Dependency modeling.* Fayyad et al. (1996b) describe dependency modeling as significant dependencies among variables, which exist at two levels: structured and quantitative. The structured level specifies which variables are locally dependent, while the quantitative level specifies the strengths of the dependencies using a numerical scale. According to Bradley et al. (1999), in dependency modeling we try to gain insight into the data by deriving some causal structure within the data. Models of causality can be probabilistic or deterministic as in deriving

TABLE 2 Literature Coverage by Selected Authors of Data Mining Operations

	Fayyad et al. (1996b)	Cabena et al. (1998)	Groth (1998)	Bradley et al. (1999)	Yoon (1999)	Berry and Linoff (2002)	Chye and Gerry (2002)
Clustering/ segmentation	✓	✓	✓	✓	✓	✓	✓
Visualization			✓			✓	✓
Predictive modeling		✓		✓		✓	✓
Link analysis	✓	✓			✓		
Deviation detection		✓		✓	✓		
Dependency modeling	✓			✓			
Summarization	✓			✓			

functional dependencies between fields in the data. Density estimation methods, in general, fall under this category.

- *Data summarization.* Data summarization provides a summary about a subset of data (Fayyad et al., 1996b; Bradley et al., 1999). Bradley et al. (1999) state that there are two classes of methods that correspond to taking horizontal (cases) or vertical (fields) slices of the data. In the former, one would like to produce summaries of subsets, and in the latter, one would like to predict relations between fields. This class of methods is different from other methods in that rather than predicting a specified field (e.g., classification) or grouping cases together (e.g., clustering), the goal is to find relations between fields. Associations rules are one method that can be used. Associations are rules that state that certain combinations of values occur with other combinations of values with a certain frequency and certainty. One example of this is market basket analysis, where one would like to summarize which products are bought together.

Table 2 shows how the data mining operations have been classified throughout the literature by selected authors.

USES OF DATA MINING

Data mining has many uses in industries today that can be beneficial to the bottom line of a company. Mitchell (1999) reports that the field of data mining "has already produced practical applications in such areas as analyzing medical outcomes, detecting credit card fraud, predicting customer purchase behavior, predicting the personal interests of Web users, and optimizing manufacturing processes. It has also led to a set of fascinating scientific questions about how computers might automatically learn from past experience." The uses of data mining discussed

in this article include (1) marketing, (2) risk management, (3) fraud detection, and (4) customer acquisition and retention.

Marketing

Marketing is one of the foremost areas where data mining techniques can be applied. Data mining enables an organization to sort through vast amounts of customer data to target the right customers. This is of vital importance to the marketing department of any organization. Substantial amounts of time and money can be saved if an organization knows who their customers are and are able to predict what their spending patterns will be. According to Cabena et al. (1998), the objective of database marketing is to analyze corporate databases so that effective marketing and promotional campaigns can be accomplished. The sales organization builds a database of customer product preferences and lifestyles from sources such as credit card transactions, loyalty cards, warranty cards and discount coupons, entries to free prize drawings, and customer complaint calls. This information is then mixed with publicly available information. Data mining algorithms then sort through the data and look for clusters of "model" consumers who share the same characteristics — for example, interests, income level, and spending habits. This group is then a target for marketing efforts.

Targeting potential customers is a marketing department's main objective. Data mining helps those in the marketing department better understand what drives a particular customer behavior, product activity, or event, and therefore enables the organization to be proactive rather than reactive to the constantly changing conditions in the business environment (Chopoorian et al., 2001). By better understanding customers, an organization is able to target those customers who have a higher chance of responding favorably.

Methods must be found to determine the number of customers who are likely to be lost to competitors so that a business can be better prepared.

Peacock (1998) identifies several potential uses that data mining has in the area of marketing, including:

- *Customer acquisition.* Marketers use data mining methods to discover attributes that can predict customer responses to offers and communications programs. Then the attributes of customers that are found to be most likely to respond are matched to corresponding attributes appended to rented lists of noncustomers. The objective is to select only noncustomer households most likely to respond to a new offer.
- *Customer retention.* Data mining helps to identify customers who contribute to the company's bottom line but who may be likely to leave and go to a competitor. The company can then target these customers for special offers and other inducements.
- *Customer abandonment.* Customers who cost more than they contribute should be encouraged to take their business elsewhere. Data mining can be used to reveal whether a customer has a negative impact on the company's bottom line.
- *Market basket analysis.* Retailers and direct marketers can spot product affinities and develop focused promotion strategies by identifying the associations between product purchases in point-of-sale transactions.

Risk Management

Risk management covers not only risks involving insurance, but also business risks from competitive threat, poor product quality, and customer attrition. Customer attrition, the loss of customers, is an increasing problem and data mining is used in the finance, retail, and telecommunications industries to help predict the possible losses of customers (Cabena et al., 1998). Losing customers to competitors is a major concern for industries today, with the increasing amount of competition businesses are facing. Therefore, methods must be found to determine the number of customers who are likely to be lost to competitors so that a business can be better prepared. One approach that can be used is to build a model of customers who are likely to leave and go to a competitive company. An analysis of customers who have recently left can often show nonintuitive patterns, such as after a customer has a change of address or a recent protracted exchange with one of the agents of the company (Cabena et al., 1998).

Fraud detection

Fraud detection must also be dealt with across all industries. The sectors where many transactions are made are more vulnerable, such as health care, retail, credit card services, and telecommunications (Cabena et al., 1998). The pioneers of the use of data mining techniques to prevent fraud were the telephone companies and insurance companies, with banks following close behind (Decker, 1998). Fraud can result in a business losing substantial amounts of money. Being able to protect a business from the chance of fraud is an important concern for an organization and data mining can help. To detect fraudulent actions, a model can be built using fraudulent behavior (or potentially fraudulent behavior) that has been done in the past and then use data mining to identify behavior that is similar (Cabena et al., 1998).

Customer Acquisition and Retention

Customer acquisition and retention is primarily a marketing effort but is discussed as a separate area because of its vital importance to industries. Customer acquisition, the task of acquiring new customers, is a major objective for any organization because without customers, a business cannot thrive. Customer retention involves retaining those customers the business already has. Although acquiring customers is a much easier task than retaining customers, it is less expensive for a business to try to retain the existing customers rather than developing additional marketing initiatives to attract new customers. Marple and Zimmerman (1999) conclude that "it costs 33 to 50 percent less to sell to existing customers than it does to sell to new ones and that the odds of an existing customer repeating business with you are exponentially higher than converting someone else's customer." Organizations can use data mining to help in acquiring new customers and retaining existing customers.

Regarding customer acquisition, Weir (1998) states that customer profiling is used to help identify the characteristics of good customers with the objectives of predicting who will become one and helping the marketing department target new prospects. Customer acquisition can be targeted appropriately using data mining to find patterns in a customer database. By only targeting persons who have the potential to become customers, an organization reduces the amount of time and money spent on customer acquisition efforts.

Utilizing data mining, a business is better able to understand the needs of its customers and, therefore, can improve and extend the existing relationship.

To retain an organization's existing customers, it is important for an organization to know what factors indicate whether a customer is likely to leave the business and go to competitors. Data mining can assist in identifying customers who are likely to leave and the business can offer those customers incentives to stay. Berger (1999) points out that an organization can create models that offer detailed "profiles" or characteristics of customers likely to churn, be profitable, or respond to a specific retention offer. When an organization knows the customers who are most profitable and most likely to respond favorably to an offer, there is less chance of losing those customers to competitors. Utilizing data mining, a business is better able to understand the needs of its customers and, therefore, can improve and extend the existing relationship.

THE BANKING INDUSTRY

The banking industry is recognizing the importance of the information it has on its customers. This industry has high information demands and uses information technology not only to improve the quality of service, but also to gain a competitive advantage (Hwang et al., 2002). The enormous amount of data that banks have been collecting over the years can greatly influence the success of data mining efforts.

According to Fabris (1998), by using data mining to analyze patterns and trends, bank executives can predict with increased accuracy how customers will react to adjustments in interest rates, which customers will be likely to accept new product offers, which customers will be at a higher risk for defaulting on a loan, and how to make each customer relationship more profitable. Data mining is proving itself very useful in the banking industry. In the following sections, examples are given of how the banking industry has been effectively utilizing data mining in the areas of (1) marketing, (2) risk management, (3) fraud detection, and (4) customer acquisition and retention.

Marketing

One of the most widely used areas of data mining for the banking industry is in marketing. The bank's marketing department can use data mining to analyze customer databases and develop statistically sound profiles of individual customer preferences for products and services. By offering only those products and services the customer really wants, the bank saves

money on promotions and offerings that would otherwise be unprofitable (Decker, 1998). Bank marketers need to focus on their customers by learning more about them. Bank of America uses database marketing to improve customer service and increase profits. By consolidating five years of customer history records, the bank is able to market and sell targeted services to their customers (Cabena et al., 1998). By revealing patterns of customer behavior, their profitability can be determined and the bank can also expand its business by offering each individual customer other products and services.

Cross selling is a marketing area where data mining can be used. Cross selling is when a service provider makes it attractive for a customer to buy additional products or services with the same business (Cabena et al., 1998). The more products and services a bank can provide for customers, the more likely the bank is to retain those customers. Cabena et al. (1998) give the example of Mellon Bank, which uses data mining to find customers with demand deposit accounts who may be interested in a home equity loan. A model is built of the customers who already have home equity loans and this model is then used to pinpoint other customers who may also be interested. Another example is the Bank of America, which has recently completed a project with IBM's data mining tools to search its database of corporate clients and try to figure out what products the clients may need next (Lach, 1999).

Risk Management

Data mining is also used for risk management in the banking industry. Bank executives need to know whether the customers they are dealing with are reliable. Offering new customers credit cards, extending existing customers lines of credit, and approving loans can be risky decisions for banks if they do not know anything about their customers. Data mining can be used to reduce the risk of banks that issue credit cards by determining those customers who are likely to default on their accounts. An example given by Kuykendall (1999) is of a bank discovering that cardholders who withdrew money at casinos had higher rates of delinquency and bankruptcy.

Crook et al. (2001) report that credit scoring was one of the earliest financial risk management tools developed. Credit scoring can be valuable to lenders in the banking industry when making lending decisions. Lenders

By knowing what the chances of default are for a customer, the bank is in a better position to reduce its risks.

would not have expanded the number of loans they give out without having an accurate, objective, and controllable risk assessment tool (Crook et al., 2001). The examples of both a good and a bad loan applicant's histories can be used to develop a profile for a good and bad new loan applicant (Cabena et al., 1998). Data mining can derive the credit behaviors of individual borrowers with installment, mortgage, and credit card loans, using characteristics such as credit history, length of employment, and length of residency (Decker, 1998). A score is produced that allows a lender to evaluate the customer and decide whether the person is a good candidate for a loan or if there is a high risk of default.

Customers who have been with the bank for longer periods of time, have remained in good standing, and have higher wages are more likely to receive a loan than a new customer who has no history with the bank or who earns low wages. By knowing what the chances of default are for a customer, the bank is in a better position to reduce its risks. The Bank of Montreal analyzes customers' transactions in checking, savings, and other accounts for insight into customers who may be at risk of defaulting (Fabris, 1998).

Fraud Detection

Another area where data mining can be used in the banking industry is in fraud detection. Groth (1998) reports that, in banking, the most widespread use of data mining is in the area of fraud detection. Being able to detect fraudulent actions is an increasing concern for many businesses; and with the help of data mining, more fraudulent actions are being detected and stopped. According to Decker (1998), two different approaches have been developed by financial institutions to detect fraud patterns. In the first approach, a bank taps the data warehouse of a third party (potentially containing transaction information from many companies) and uses data mining programs to identify fraud patterns. The bank can then cross-reference those patterns with its own database for signs of internal trouble. In the second approach, fraud pattern identification is based strictly on the bank's own internal information.

One system that has been successful in detecting fraud is the Falcon fraud assessment system developed by HNC, Inc.; it was built using a neural network shell and is used by many banks to detect suspicious credit card transactions (Brachman et al., 1996). Kuykendall

(1999) reports that "Falcon is used by nine of the top ten credit card issuing banks, where it examines the transactions of 80 percent of cards held in the U.S." Mellon Bank also uses data mining for fraud detection and is able to better protect itself and its customers' funds from potential credit card fraud (Cabena et al., 1998).

Customer Acquisition and Retention

Not only can data mining help the banking industry gain new customers, but it can also help in retaining existing customers. Customer acquisition and retention are important concerns for any industry, especially the banking industry. Today, customers have so many options with regard to where they can choose to do their business. Executives in the banking industry must be aware that if they are not giving each customer their full attention, the customer can simply find another company that will. Data mining can help in targeting new customers for products and services and in discovering a customer's previous purchasing patterns so that the bank will be able to retain its existing customers by offering incentives that are individually tailored to the customer's needs.

When Chase Manhattan Bank in New York began to lose customers to competitors, it began using data mining to analyze customer accounts and make changes in its account requirements, thus allowing the bank to retain its profitable customers (Fabris, 1998). Data mining is also being used by Fleet Bank in Boston to identify the best candidates for mutual fund offerings. The bank mines customer demographics and account data along different product lines to determine which customers may be likely to invest in a mutual fund, and then this information is used to target those customers (Fabris, 1998). Bank of America's West Coast customer service call center has its representatives ready with customer profiles gathered from data mining to pitch new products and services that are the most relevant to each individual caller (Fabris, 1998).

Mortgage bankers are also concerned with retaining customers. Mortgage Guaranty Insurance Corporation (MGIC), in Milwaukee, began a customer retention program for lenders, with the goal of helping lenders retain their customers' lender of choice. The program uses leading-edge Internet technologies, predictive models, and consumer-direct marketing to enable lenders to identify new customers and retain those that they already have (Marple and

Zimmerman, 1999). For lenders to remain competitive, they must understand the needs of their customers; otherwise, the customers will be lost to competitors.

For lenders to remain competitive, they must understand the needs of their customers; otherwise, the customers will be lost to competitors.

THE RETAIL INDUSTRY

The retail industry is also realizing that it is possible to gain a competitive advantage utilizing data mining. Retailers have been collecting enormous amounts of data throughout the years, just as the banking industry, and now have the tool that is needed to sort through this data and find useful pieces of information. For retailers, data mining can be used to provide information on product sales trends, customer buying habits and preferences, supplier lead times and delivery performance, seasonal variations, customer peak traffic periods, and similar predictive data for making proactive decisions (Savage, 2002). According to Groth (1998), retailers are interested in creating data mining models to answer questions such as:

- How much are customers likely to spend over long periods of time?
- What is the frequency of customer purchasing behavior?
- What are the best types of advertisements to reach certain segments?
- What advertising mediums are most effective in reaching customers?
- What is the optimal timing to send mailers?

In the following sections, examples of how the retail industry has also been effectively utilizing data mining in the areas of (1) marketing, (2) risk management, (3) fraud detection, and (4) customer acquisition and retention are shown.

Marketing

One of the most widely used areas of data mining for the retail industry, like the banking industry is in marketing. Groth (1998) suggests that customer identification is critical to most retailers, and is likely to become more so. Therefore, it is crucial that retailers identify with their customers by offering promotions and incentives that are tailored to fit each individual customer's needs.

Market basket analysis is a marketing method used by many retailers to determine the optimal locations to promote products. Brachman et al. (1996) defines market basket analysis as "the study of retail stock movement data recorded at a point of sale — to support decisions on shelf-space allocation, store layout, and product location and promotion effectiveness."

As indicated by Peacock (1998), market basket analysis uses information about products already purchased by customers to predict which products they would be likely to buy if given special offers or even if they are just made aware of the products. Knowing where to locate products and promote them effectively can increase the stores' sales.

Another marketing tactic employed by many retail stores is the use of loyalty cards. Rewarding those customers who are frequent buyers encourages them to do even more of their shopping at that store and makes them less likely to buy from competitor stores (Cabena et al., 1998). Rewarding customers for purchasing products gives them the incentive to purchase even more. Supermarkets can use loyalty cards so that when the store has a new service to offer, such as ready-to-eat frozen meals, it goes only to the people who have already shown interest in frozen meals. They can also be used to increase spending, where shoppers could be offered a free dinner when they have spent over \$400 during a month (Berry and Linoff, 2000).

Coupon printers at the checkout stands of supermarkets provide an additional way to target customers. These printers are beneficial to brand managers who may not know which customers to target for their brand of products. The coupon printer at the checkout stand can be programmed to print out a coupon for their particular brand when certain products are purchased. This ensures that an incentive to try the particular brand is offered to the people who are currently buying products in certain categories (Berry and Linoff, 2000). For example, if shoppers buy a certain kind of laundry detergent, they will be given a coupon for other laundry care items of the desired brand.

There are an increasing number of retail companies utilizing data mining for marketing purposes and benefiting from its use. Fingerhut Corp., a catalog retailer in Minnetonka, Minnesota, sends out 130 different catalogs and has a six-terabyte data warehouse holding information on more than 65 million customers. Users can query the database daily and crunch more than 3000 variables on the company's 12 million most active customers in the past four years. The company uses more than 300 predictive models to search the data warehouse, with one or more models being used on each potential recipient of every catalog that the company sends out. In addition, data mining has also enabled Fingerhut to create new catalogs. One example is the "mover's catalog," which was

The managers can look at the videotape to see exactly what happened and do not even have to be anywhere near the store when they do it.

determined to be successful when targeted to customers who had changed their residence. The catalog contained targeted products for this segment of customers and saved the company money by not sending these customers other catalogs right after they had relocated (Lach, 1999). Another example given by Lach (1999) is of the marketing department at Eddie Bauer in Redmond, Washington, which is able to run queries from its desktops to determine which promotions to offer in each store and which catalog to send to a customer who has not made any recent purchases.

Risk Management

Risk management is another area where data mining is used in the retail industry; however, there has not been as much research done in this area as has been done in the other areas. Cabena et al. (1998) report that retail organizations use data mining to understand which products may be vulnerable to competitive offerings or changing consumer purchasing patterns. The previous purchasing patterns of customers are analyzed to identify those customers with low product or brand loyalty. Data mining enables retailers to remain competitive and reduce their risks by helping them understand what their customers are really doing. Retailers can then target those customers who are more likely to buy a certain brand or product and also be able to promote products in stores where and when they are needed.

Fraud Detection

Retail industries must also be aware that fraud detection is necessary. Fraud occurring at POS terminals is a concern for retailers but can be reduced using data mining. It is estimated that 38 percent of retail shrink occurs because of dishonest employees (Cabena et al., 1998). Weir (2001) reports that the average supermarket is almost twice as likely to lose a dollar to shrink as it is to earn a dollar in profit. And with about 25 cents of every shrink dollar traceable to POS fraud, it is no wonder retailers continue looking for ways to reduce the number of dishonest cashiers.

According to Weir (2001), some supermarkets have begun to make use of digitized closed-circuit television (CCTV) systems along with POS data mining to enable retail loss prevention managers to expose cashier stealing and sweethearting, assemble convincing evidence, and deal with these situations as a matter of routine. The managers decide what

constitutes suspicious behavior and send their software out to detect it. This is called exception-based reporting. The system flags POS transactions that are the most susceptible to fraud — refunds, credits, discounts, no-sale rings, and the like — and compiles them in a report that identifies the date, time, and check-out lanes where they took place. The managers can then look at the videotape to see exactly what happened and do not even have to be anywhere near the store when they do it.

Customer Acquisition and Retention

Data mining can also help in acquiring and retaining customers in the retail industry. The retail industry must deal with an enormous amount of competition and can use data mining to better understand what their customers' needs are. A retailer can study customers' past purchasing histories and know with what kinds of promotions and incentives to target customers. Also, if a store has seen a number of people leave the store and go to competitors, data mining can be used to study the customers' past purchasing histories and then use this information to keep other customers from going to a competitor store.

CONCLUSION

This article has defined data mining as a tool used to extract important information from existing data and enable better decision making throughout an organization. An organization uses a data warehouse to combine various data from its databases into an acceptable format so that the data can be mined. The data is then analyzed using data mining and the information that is captured is used throughout the organization to support decision making. The operations of data mining that have been classified throughout the literature were also discussed, including (1) clustering/segmentation, (2) visualization, (3) predictive modeling, (4) link analysis, (5) deviation detection, (6) dependency modeling, and (7) summarization.

This research has shown that many industries, including the banking and retail industries, have been effectively using data mining. In marketing, data mining enables an organization to sort through vast amounts of customer data to target the right customers. In risk management, data mining helps an organization predict the number of customers who are likely to be lost to competitors so that the business can be better prepared. Data mining is useful for fraud detection because, by detecting these

fraudulent actions, an organization can keep from losing substantial amounts of money. Data mining is also important for customer acquisition and retention, enabling an organization to target new customers and retain the customers it already has. Those industries that effectively utilize data mining have the ability to develop or enhance a competitive advantage. ▲

References

1. Berger, C. August 1999. Data Mining to Reduce Churn. *Target Marketing*, 22(8), 26-27.
2. Berry, M.J.A. and Linoff, G.S. 2000. *Mastering Data Mining*. John Wiley & Sons, New York.
3. Brachman, R.J., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, G., and Simoudis, E. November 1996. Mining Business Databases. *Communications of the ACM*, 39(11).
4. Bradley, P.S., Fayyad, U.M., and Mangasarian, O.L. Summer 1999. Mathematical Programming for Data Mining: Formulations and Challenges. *Journal on Computing*, 11(3), 217-238.
5. Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., and Zanasi, A. 1998. *Discovering Data Mining: From Concept to Implementation*. Prentice Hall, New Jersey.
6. Chopoorian, J.A., Witherell, R., Khalil, O.E.M., and Ahmed, M. Spring 2001. Mind Your Business by Mining Your Data. *SAM Advanced Management Journal*, 66(2).
7. Chung, H. and Gray, M., Summer 1999. Special Section: Data Mining. *Journal of Management Information Systems*, 16(1).
8. Chye, K.H. and Gerry, C.K.L. July-Dec. 2002. Data Mining and Customer Relationship Marketing in the Banking Industry. *Singapore Management Review*, 24(2), 1-27.
9. Crook, J.N., Edelman, D.B., and Thomas, L.C. 2001. Editorial Review. *Journal of the Operational Research Society*, 52(9), 972-973.
10. Decker, P. March/April 1998. Data Mining's Hidden Dangers. *Banking Strategies*, 6-14.
11. Fabris, P. 1998. Advanced Navigation. *CIO*, May 15.
12. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. 1996a. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, Cambridge, MA.
13. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. November 1996b. The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, 39(11), 27-34.
14. Groth, R. 1998. *Data Mining: A Hands-On Approach for Business Professionals*. Prentice Hall, New Jersey.
15. Hui, S.C. and Jha, G. October 2000. Data Mining for Customer Service Support. *Information & Management*, 38(1), 1-13.
16. Hwang, H., Ku, C., Yen, D.C., and Cheng, C. July 2002. Critical Factors Influencing the Adoption of Data Warehouse Technology: A Study of the Banking Industry in Taiwan. *Decision Support Systems*, (1039), 1-21.
17. Kempster, M. June 1998. Digging for Data. *Inform*, 12(6), 22-30.
18. Kuykendall, L. September 1999. The Data-Mining Toolbox. *Credit Card Management*, 12(7).
19. Lach, J. July 1999. Data Mining Digs In. *American Demographics*, 21(7).
20. Le Beau, C. 2000. Mountains to Mine. *American Demographics*, 22(8).
21. Marple, M. and Zimmerman, M. August 1999. A Customer Retention Strategy. *Mortgage Banking*, 59(11).
22. Mitchell, T.M. November 1999. Machine Learning and Data Mining. *Communications of the ACM*, 42(11).
23. Peacock, P.R. Winter 1998. Data Mining in Marketing: Part 1. *Marketing Management*, 6(4).
24. Savage, T. March 2002. Future Trends in Retail IT Systems. *European Retail Digest*, (33), 17-19.
25. Weir, J. Fall 1998. Data Mining: Exploring the Corporate Asset. *Information Systems Management*, 15(4), 68-71.
26. Weir, T. November 15, 2001. Caught on Camera. *Supermarket Business Magazine*, 56(11), 11-12.
27. Yoon, Y. Spring 1999. Discovering Knowledge in Corporate Databases. *Information Systems Management*, 16(2), 64-71.
28. Zhu, D., Premkumar, G., Zhang, X., and Chu, C. Fall 2001. Data Mining for Network Intrusion Detection: A Comparison of Alternative Methods. *Decision Sciences*, 32(4), 635-660.

A vertical bar on the left side of the page, consisting of a series of colored squares (red, orange, yellow, green, blue) stacked vertically, with a small red diamond at the top.

COPYRIGHT INFORMATION

TITLE: Data Mining: A Competitive Weapon for Banking and
Retail Industries

SOURCE: Information Systems Management 21 no2 Spr 2004

PAGE(S): 62-71

WN: 0410802780008

The magazine publisher is the copyright holder of this article and it is reproduced with permission. Further reproduction of this article in violation of the copyright is prohibited. To contact the publisher:
<http://www.auerbach-publications.com/>

Copyright 1982-2004 The H.W. Wilson Company. All rights reserved.