Michael J. A. Berry and Gordon S. Linoff
Data Mining Techniques
2nd Edition, Wiley, 2004, Chapter 1

CHAPTER
1

Why and What Is Data Mining?

In the first edition of this book, the first sentence of the first chapter began with the words "Somerville, Massachusetts, home to one of the authors of this book, . . ." and went on to tell of two small businesses in that town and how they had formed learning relationships with their customers. In the intervening years, the little girl whose relationship with her hair braider was described in the chapter has grown up and moved away and no longer wears her hair in corn-rows. Her father has moved to nearby Cambridge. But one thing has not changed. The author is still a loyal customer of the Wine Cask, where some of the same people who first introduced him to cheap Algerian reds in 1978 and later to the wine-growing regions of France are now helping him to explore Italy and Germany.

After a quarter of a century, they still have a loyal customer. That loyalty is no accident. Dan and Steve at the Wine Cask learn the tastes of their customers and their price ranges. When asked for advice, their response will be based on their accumulated knowledge of that customer's tastes and budgets as well as on their knowledge of their stock.

The people at The Wine Cask know a lot about wine. Although that knowledge is one reason to shop there rather than at a big discount liquor store, it is their intimate knowledge of each customer that keeps people coming back. Another wine shop could open across the street and hire a staff of expert oenophiles, but it would take them months or years to achieve the same level of customer knowledge.

Well-run small businesses naturally form learning relationships with their customers. Over time, they learn more and more about their customers, and they use that knowledge to serve them better. The result is happy, loyal customers and profitable businesses. Larger companies, with hundreds of thousands or millions of customers, do not enjoy the luxury of actual personal relationships with each one. These larger firms must rely on other means to form learning relationships with their customers. In particular, they must learn to take full advantage of something they have in abundance—the data produced by nearly every customer interaction. This book is about analytic techniques that can be used to turn customer data into customer knowledge.

Analytic Customer Relationship Management

It is widely recognized that firms of all sizes need to learn to emulate what small, service-oriented businesses have always done well—creating one-to-one relationships with their customers. Customer relationship management is a broad topic that is the subject of many books and conferences. Everything from lead-tracking software to campaign management software to call center software is now marketed as a customer relationship management tool. The focus of this book is narrower—the role that data mining can play in improving customer relationship management by improving the firm's ability to form learning relationships with its customers.

In every industry, forward-looking companies are moving toward the goal of understanding each customer individually and using that understanding to make it easier for the customer to do business with them rather than with competitors. These same firms are learning to look at the value of each customer so that they know which ones are worth investing money and effort to hold on to and which ones should be allowed to depart. This change in focus from broad market segments to individual customers requires changes throughout the enterprise, and nowhere more than in marketing, sales, and customer support.

Building a business around the customer relationship is a revolutionary change for most companies. Banks have traditionally focused on maintaining the spread between the rate they pay to bring money in and the rate they charge to lend money out. Telephone companies have concentrated on connecting calls through the network. Insurance companies have focused on processing claims and managing investments. It takes more than data mining to turn a product-focused organization into a customer-centric one. A data mining result that suggests offering a particular customer a widget instead of a gizmo will be ignored if the manager's bonus depends on the number of gizmos sold this quarter and not on the number of widgets (even if the latter are more profitable).

In the narrow sense, data mining is a collection of tools and techniques. It is one of several technologies required to support a customer-centric enterprise. In a broader sense, data mining is an attitude that business actions should be based on learning, that informed decisions are better than uninformed decisions, and that measuring results is beneficial to the business. Data mining is also a process and a methodology for applying the tools and techniques. For data mining to be effective, the other requirements for analytic CRM must also be in place. In order to form a learning relationship with its customers, a firm must be able to:

- *Notice* what its customers are doing
- *Remember* what it and its customers have done over time
- *Learn* from what it has remembered
- *Act* on what it has learned to make customers more profitable

Although the focus of this book is on the third bullet—learning from what has happened in the past—that learning cannot take place in a vacuum. There must be transaction processing systems to capture customer interactions, data warehouses to store historical customer behavior information, data mining to translate history into plans for future action, and a customer relationship strategy to put those plans into practice.

The Role of Transaction Processing Systems

A small business builds relationships with its customers by noticing their needs, remembering their preferences, and learning from past interactions how to serve them better in the future. How can a large enterprise accomplish something similar when most company employees may never interact personally with customers? Even where there is customer interaction, it is likely to be with a different sales clerk or anonymous call-center employee each time, so how can the enterprise notice, remember, and learn from these interactions? What can replace the creative intuition of the sole proprietor who recognizes customers by name, face, and voice, and remembers their habits and preferences?

In a word, nothing. But that does not mean that we cannot try. Through the clever application of information technology, even the largest enterprise can come surprisingly close. In large commercial enterprises, the first step—noticing what the customer does—has already largely been automated. Transaction processing systems are everywhere, collecting data on seemingly everything. The records generated by automatic teller machines, telephone switches, Web servers, point-of-sale scanners, and the like are the raw material for data mining.

These days, we all go through life generating a constant stream of transaction records. When you pick up the phone to order a canoe paddle from L.L.

Bean or a satin bra from Victoria's Secret, a call detail record is generated at the local phone company showing, among other things, the time of your call, the number you dialed, and the long-distance company to which you have been connected. At the long-distance company, similar records are generated recording the duration of your call and the exact routing it takes through the switching system. This data will be combined with other records that store your billing plan, name, and address in order to generate a bill. At the catalog company, your call is logged again along with information about the particular catalog from which you ordered and any special promotions you are responding to. When the customer service representative that answered your call asks for your credit card number and expiration date, the information is immediately relayed to a credit card verification system to approve the transaction; this too creates a record. All too soon, the transaction reaches the bank that issued your credit card, where it appears on your next monthly statement. When your order, with its item number, size, and color, goes into the cataloger's order entry system, it spawns still more records in the billing system and the inventory control system. Within hours, your order is also generating transaction records in a computer system at UPS or FedEx where it is scanned about a dozen times between the warehouse and your home, allowing you to check the shipper's Web site to track its progress.

These transaction records are not generated with data mining in mind; they are created to meet the operational needs of the company. Yet all contain valuable information about customers and all can be mined successfully. Phone companies have used call detail records to discover residential phone numbers whose calling patterns resemble those of a business in order to market special services to people operating businesses from their homes. Catalog companies have used order histories to decide which customers should be included in which future mailings—and, in the case of Victoria's secret, which models produce the most sales. Federal Express used the change in its customers' shipping patterns during a strike at UPS in order to calculate their share of their customers' package delivery business. Supermarkets have used point-of-sale data in order to decide what coupons to print for which customers. Web retailers have used past purchases in order to determine what to display when customers return to the site.

These transaction systems are the customer touch points where information about customer behavior first enters the enterprise. As such, they are the eyes and ears (and perhaps the nose, tongue, and fingers) of the enterprise.

The Role of Data Warehousing

The customer-focused enterprise regards every record of an interaction with a client or prospect—each call to customer support, each point-of-sale transaction, each catalog order, each visit to a company Web site—as a learning opportunity. But learning requires more than simply gathering data. In fact,

many companies gather hundreds of gigabytes or terabytes of data from and about their customers without learning anything! Data is gathered because it is needed for some operational purpose, such as inventory control or billing. And, once it has served that purpose, it languishes on disk or tape or is discarded.

For learning to take place, data from many sources—billing records, scanner data, registration forms, applications, call records, coupon redemptions, surveys—must first be gathered together and organized in a consistent and useful way. This is called *data warehousing*. Data warehousing allows the enterprise to remember what it has noticed about its customers.

TIP **Customer patterns become evident over time. Data warehouses need to support accurate historical data so that data mining can pick up these critical trends.**

One of the most important aspects of the data warehouse is the capability to track customer behavior over time. Many of the patterns of interest for customer relationship management only become apparent over time. Is usage trending up or down? How frequently does the customer return? Which channels does the customer prefer? Which promotions does the customer respond to?

A number of years ago, a large catalog retailer discovered the importance of retaining historical customer behavior data when they first started keeping more than a year's worth of history on their catalog mailings and the responses they generated from customers. What they discovered was a segment of customers that only ordered from the catalog at Christmas time. With knowledge of that segment, they had choices as to what to do. They could try to come up with a way to stimulate interest in placing orders the rest of the year. They could improve their overall response rate by not mailing to this segment the rest of the year. Without some further experimentation, it is not clear what the right answer is, but without historical data, they would never have known to ask the question.

A good data warehouse provides access to the information gleaned from transactional data in a format that is much friendlier than the way it is stored in the operational systems where the data originated. Ideally, data in the warehouse has been gathered from many sources, cleaned, merged, tied to particular customers, and summarized in various useful ways. Reality often falls short of this ideal, but the corporate data warehouse is still the most important source of data for analytic customer relationship management.

The Role of Data Mining

The data warehouse provides the enterprise with a memory. But, memory is of little use without intelligence. Intelligence allows us to comb through our memories, noticing patterns, devising rules, coming up with new ideas, figuring out the right questions, and making predictions about the future. This book describes tools and techniques that add intelligence to the data warehouse. These techniques help make it possible to exploit the vast mountains of data generated by interactions with customers and prospects in order to get to know them better.

Who is likely to remain a loyal customer and who is likely to jump ship? What products should be marketed to which prospects? What determines whether a person will respond to a certain offer? Which telemarketing script is best for this call? Where should the next branch be located? What is the next product or service this customer will want? Answers to questions like these lie buried in corporate data. It takes powerful data mining tools to get at them.

The central idea of data mining for customer relationship management is that data from the past contains information that will be useful in the future. It works because customer behaviors captured in corporate data are not random, but reflect the differing needs, preferences, propensities, and treatments of customers. The goal of data mining is to find patterns in historical data that shed light on those needs, preferences, and propensities. The task is made difficult by the fact that the patterns are not always strong, and the signals sent by customers are noisy and confusing. Separating signal from noise—recognizing the fundamental patterns beneath seemingly random variations—is an important role of data mining.

This book covers all the most important data mining techniques and the strengths and weaknesses of each in the context of customer relationship management.

The Role of the Customer Relationship Management Strategy

To be effective, data mining must occur within a context that allows an organization to change its behavior as a result of what it learns. It is no use knowing that wireless telephone customers who are on the wrong rate plan are likely to cancel their subscriptions if there is no one empowered to propose that they switch to a more appropriate plan as suggested in the sidebar. Data mining should be embedded in a corporate customer relationship strategy that spells out the actions to be taken as a result of what is learned through data mining. When low-value customers are identified, how will they be treated? Are there programs in place to stimulate their usage to increase their value? Or does it make more sense to lower the cost of serving them? If some channels consistently bring in more profitable customers, how can resources be shifted to those channels?

Data mining is a tool. As with any tool, it is not sufficient to understand how it works; it is necessary to understand how it will be used.

DATA MINING SUGGESTS, BUSINESSES DECIDE

This sidebar explores the example from the main text in slightly more detail. An analysis of attrition at a wireless telephone service provider often reveals that people whose calling patterns do not match their rate plan are more likely to cancel their subscriptions. People who use more than the number of minutes included in their plan are charged for the extra minutes—often at a high rate. People who do not use their full allotment of minutes are paying for minutes they do not use and are likely to be attracted to a competitor's offer of a cheaper plan.

This result suggests doing something proactive to move customers to the right rate plan. But this is not a simple decision. As long as they don't quit, customers on the wrong rate plan are more profitable if left alone. Further analysis may be needed. Perhaps there is a subset of these customers who are not price sensitive and can be safely left alone. Perhaps any intervention will simply hand customers an opportunity to cancel. Perhaps a small "rightsizing" test can help resolve these issues. Data mining can help make more informed decisions. It can suggest tests to make. Ultimately, though, the business needs to make the decision.

What Is Data Mining?

Data mining, as we use the term, is the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. For the purposes of this book, we assume that the *goal* of data mining is to allow a corporation to improve its marketing, sales, and customer support operations through a better understanding of its customers. Keep in mind, however, that the data mining techniques and tools described here are equally applicable in fields ranging from law enforcement to radio astronomy, medicine, and industrial process control.

In fact, hardly any of the data mining algorithms were first invented with commercial applications in mind. The commercial data miner employs a grab bag of techniques borrowed from statistics, computer science, and machine learning research. The choice of a particular combination of techniques to apply in a particular situation depends on the nature of the data mining task, the nature of the available data, and the skills and preferences of the data miner.

Data mining comes in two flavors—directed and undirected. Directed data mining attempts to explain or categorize some particular target field such as income or response. Undirected data mining attempts to find patterns or similarities among groups of records without the use of a particular target field or collection of predefined classes. Both these flavors are discussed in later chapters.

Data mining is largely concerned with building models. A model is simply an algorithm or set of rules that connects a collection of inputs (often in the form of fields in a corporate database) to a particular target or outcome. Regression, neural networks, decision trees, and most of the other data mining techniques discussed in this book are techniques for creating models. Under the right circumstances, a model can result in insight by providing an explanation of how outcomes of particular interest, such as placing an order or failing to pay a bill, are related to and predicted by the available facts. Models are also used to produce *scores*. A score is a way of expressing the findings of a model in a single number. Scores can be used to sort a list of customers from most to least loyal or most to least likely to respond or most to least likely to default on a loan.

The data mining process is sometimes referred to as *knowledge discovery* or KDD (knowledge discovery in databases). We prefer to think of it as *knowledge creation*.

What Tasks Can Be Performed with Data Mining?

Many problems of intellectual, economic, and business interest can be phrased in terms of the following six tasks:

- Classification
- Estimation
- Prediction
- Affinity grouping
- Clustering
- Description and profiling

The first three are all examples of directed data mining, where the goal is to find the value of a particular target variable. Affinity grouping and clustering are undirected tasks where the goal is to uncover structure in data without respect to a particular target variable. Profiling is a descriptive task that may be either directed or undirected.

Classification

Classification, one of the most common data mining tasks, seems to be a human imperative. In order to understand and communicate about the world, we are constantly classifying, categorizing, and grading. We divide living things into phyla, species, and general; matter into elements; dogs into breeds; people into races; steaks and maple syrup into USDA grades.

Classification consists of examining the features of a newly presented object and assigning it to one of a predefined set of classes. The objects to be classified are generally represented by records in a database table or a file, and the act of classification consists of adding a new column with a class code of some kind.

The classification task is characterized by a well-defined definition of the classes, and a training set consisting of preclassified examples. The task is to build a model of some kind that can be applied to unclassified data in order to classify it.

Examples of classification tasks that have been addressed using the techniques described in this book include:

- Classifying credit applicants as low, medium, or high risk
- Choosing content to be displayed on a Web page
- Determining which phone numbers correspond to fax machines
- Spotting fraudulent insurance claims
- Assigning industry codes and job designations on the basis of free-text job descriptions

In all of these examples, there are a limited number of classes, and we expect to be able to assign any record into one or another of them. Decision trees (discussed in Chapter 6) and nearest neighbor techniques (discussed in Chapter 8) are techniques well suited to classification. Neural networks (discussed in Chapter 7) and link analysis (discussed in Chapter 10) are also useful for classification in certain circumstances.

Estimation

Classification deals with discrete outcomes: yes or no; measles, rubella, or chicken pox. Estimation deals with continuously valued outcomes. Given some input data, estimation comes up with a value for some unknown continuous variable such as income, height, or credit card balance.

In practice, estimation is often used to perform a classification task. A credit card company wishing to sell advertising space in its billing envelopes to a ski boot manufacturer might build a classification model that put all of its cardholders into one of two classes, skier or nonskier. Another approach is to build a model that assigns each cardholder a "propensity to ski score." This might be a value from 0 to 1 indicating the estimated probability that the cardholder is a skier. The classification task now comes down to establishing a threshold score. Anyone with a score greater than or equal to the threshold is classed as a skier, and anyone with a lower score is considered not to be a skier.

The estimation approach has the great advantage that the individual records can be rank ordered according to the estimate. To see the importance of this, imagine that the ski boot company has budgeted for a mailing of 500,000 pieces. If the classification approach is used and 1.5 million skiers are identified, then it might simply place the ad in the bills of 500,000 people selected at random from that pool. If, on the other hand, each cardholder has a propensity to ski score, it can send the ad to the 500,000 most likely candidates.

Examples of estimation tasks include:

- Estimating the number of children in a family
- Estimating a family's total household income
- Estimating the lifetime value of a customer
- Estimating the probability that someone will respond to a balance transfer solicitation.

Regression models (discussed in Chapter 5) and neural networks (discussed in Chapter 7) are well suited to estimation tasks. Survival analysis (Chapter 12) is well suited to estimation tasks where the goal is to estimate the time to an event, such as a customer stopping.

Prediction

Prediction is the same as classification or estimation, except that the records are classified according to some predicted future behavior or estimated future value. In a prediction task, the only way to check the accuracy of the classification is to wait and see. The primary reason for treating prediction as a separate task from classification and estimation is that in predictive modeling there are additional issues regarding the temporal relationship of the input variables or predictors to the target variable.

Any of the techniques used for classification and estimation can be adapted for use in prediction by using training examples where the value of the variable to be predicted is already known, along with historical data for those examples. The historical data is used to build a model that explains the current observed behavior. When this model is applied to current inputs, the result is a prediction of future behavior.

Examples of prediction tasks addressed by the data mining techniques discussed in this book include:

- Predicting the size of the balance that will be transferred if a credit card prospect accepts a balance transfer offer
- Predicting which customers will leave within the next 6 months
- Predicting which telephone subscribers will order a value-added service such as three-way calling or voice mail

Most of the data mining techniques discussed in this book are suitable for use in prediction so long as training data is available in the proper form. The choice of technique depends on the nature of the input data, the type of value to be predicted, and the importance attached to explicability of the prediction.

Affinity Grouping or Association Rules

The task of affinity grouping is to determine which things go together. The prototypical example is determining what things go together in a shopping cart at the supermarket, the task at the heart of *market basket analysis*. Retail chains can use affinity grouping to plan the arrangement of items on store shelves or in a catalog so that items often purchased together will be seen together.

Affinity grouping can also be used to identify cross-selling opportunities and to design attractive packages or groupings of product and services.

Affinity grouping is one simple approach to generating rules from data. If two items, say cat food and kitty litter, occur together frequently enough, we can generate two *association rules*:

- People who buy cat food also buy kitty litter with probability P1.
- People who buy kitty litter also buy cat food with probability P2.

Association rules are discussed in detail in Chapter 9.

Clustering

Clustering is the task of segmenting a heterogeneous population into a number of more homogeneous subgroups or *clusters*. What distinguishes clustering from classification is that clustering does not rely on predefined classes. In classification, each record is assigned a predefined class on the basis of a model developed through training on preclassified examples.

In clustering, there are no predefined classes and no examples. The records are grouped together on the basis of self-similarity. It is up to the user to determine what meaning, if any, to attach to the resulting clusters. Clusters of symptoms might indicate different diseases. Clusters of customer attributes might indicate different market segments.

Clustering is often done as a prelude to some other form of data mining or modeling. For example, clustering might be the first step in a market segmentation effort: Instead of trying to come up with a one-size-fits-all rule for "what kind of promotion do customers respond to best," first divide the customer base into clusters or people with similar buying habits, and then ask what kind of promotion works best for each cluster. Cluster detection is discussed in detail in Chapter 11. Chapter 7 discusses self-organizing maps, another technique sometimes used for clustering.

Profiling

Sometimes the purpose of data mining is simply to describe what is going on in a complicated database in a way that increases our understanding of the people, products, or processes that produced the data in the first place. A good enough *description* of a behavior will often suggest an *explanation* for it as well. At the very least, a good description suggests where to start looking for an explanation. The famous gender gap in American politics is an example of how a simple description, "women support Democrats in greater numbers than do men," can provoke large amounts of interest and further study on the part of journalists, sociologists, economists, and political scientists, not to mention candidates for public office.

Decision trees (discussed in Chapter 6) are a powerful tool for profiling customers (or anything else) with respect to a particular target or outcome. Association rules (discussed in Chapter 9) and clustering (discussed in Chapter 11) can also be used to build profiles.

Why Now?

Most of the data mining techniques described in this book have existed, at least as academic algorithms, for years or decades. However, it is only in the last decade that commercial data mining has caught on in a big way. This is due to the convergence of several factors:

- The data is being produced.
- The data is being warehoused.
- Computing power is affordable.
- Interest in customer relationship management is strong.
- Commercial data mining software products are readily available.

Let's look at each factor in turn.

Data Is Being Produced

Data mining makes the most sense when there are large volumes of data. In fact, most data mining algorithms *require* large amounts of data in order to build and train the models that will then be used to perform classification, prediction, estimation, or other data mining tasks.

A few industries, including telecommunications and credit cards, have long had an automated, interactive relationship with customers that generated many transaction records, but it is only relatively recently that the automation of everyday life has become so pervasive. Today, the rise of supermarket point-of-sale scanners, automatic teller machines, credit and debit cards, pay-per-view television, online shopping, electronic funds transfer, automated order processing, electronic ticketing, and the like means that data is being produced and collected at unprecedented rates.

Data Is Being Warehoused

Not only is a large amount of data being produced, but also, more and more often, it is being extracted from the operational billing, reservations, claims processing, and order entry systems where it is generated and then fed into a data warehouse to become part of the corporate memory.

Data warehousing brings together data from many different sources in a common format with consistent definitions for keys and fields. It is generally not possible (and certainly not advisable) to perform computer- and input/output (I/O)–intensive data mining operations on an operational system that the business depends on to survive. In any case, operational systems store data in a format designed to optimize performance of the operational task. This format is generally not well suited to decision-support activities like data mining. The data warehouse, on the other hand, should be designed exclusively for decision support, which can simplify the job of the data miner.

Computing Power Is Affordable

Data mining algorithms typically require multiple passes over huge quantities of data. Many are computationally intensive as well. The continuing dramatic decrease in prices for disk, memory, processing power, and I/O bandwidth has brought once-costly techniques that were used only in a few government-funded laboratories into the reach of ordinary businesses.

The successful introduction of parallel relational database management software by major suppliers such as Oracle, Teradata, and IBM, has brought the power of parallel processing into many corporate data centers for the first time. These parallel database server platforms provide an excellent environment for large-scale data mining.

Interest in Customer Relationship Management Is Strong

Across a wide spectrum of industries, companies have come to realize that their customers are central to their business and that customer information is one of their key assets.

Every Business Is a Service Business

For companies in the service sector, information confers competitive advantage. That is why hotel chains record your preference for a nonsmoking room and car rental companies record your preferred type of car. In addition, companies that have not traditionally thought of themselves as service providers are beginning to think differently. Does an automobile dealer sell cars or transportation? If the latter, it makes sense for the dealership to offer you a loaner car whenever your own is in the shop, as many now do.

Even commodity products can be enhanced with service. A home heating oil company that monitors your usage and delivers oil when you need more, sells a better product than a company that expects you to remember to call to arrange a delivery before your tank runs dry and the pipes freeze. Credit card companies, long-distance providers, airlines, and retailers of all kinds often compete as much or more on service as on price.

Information Is a Product

Many companies find that the information they have about their customers is valuable not only to themselves, but to others as well. A supermarket with a loyalty card program has something that the consumer packaged goods industry would love to have—knowledge about who is buying which products. A credit card company knows something that airlines would love to know—who is buying a lot of airplane tickets. Both the supermarket and the credit card company are in a position to be knowledge brokers or *infomediaries*. The supermarket can charge consumer packaged goods companies more to print coupons when the supermarkets can promise higher redemption rates by printing the right coupons for the right shoppers. The credit card company can charge the airlines to target a frequent flyer promotion to people who travel a lot, but fly on other airlines.

Google knows what people are looking for on the Web. It takes advantage of this knowledge by selling sponsored links. Insurance companies pay to make sure that someone searching on "car insurance" will be offered a link to their site. Financial services pay for sponsored links to appear when someone searches on the phrase "mortgage refinance."

In fact, any company that collects valuable data is in a position to become an information broker. The *Cedar Rapids Gazette* takes advantage of its dominant position in a 22-county area of Eastern Iowa to offer direct marketing services to local businesses. The paper uses its own obituary pages and wedding announcements to keep its marketing database current.

Commercial Data Mining Software Products Have Become Available

There is always a lag between the time when new algorithms first appear in academic journals and excite discussion at conferences and the time when commercial software incorporating those algorithms becomes available. There is another lag between the initial availability of the first products and the time that they achieve wide acceptance. For data mining, the period of widespread availability and acceptance has arrived.

Many of the techniques discussed in this book started out in the fields of statistics, artificial intelligence, or machine learning. After a few years in universities and government labs, a new technique starts to be used by a few early adopters in the commercial sector. At this point in the evolution of a new technique, the software is typically available in source code to the intrepid user willing to retrieve it via FTP, compile it, and figure out how to use it by reading the author's Ph.D. thesis. Only after a few pioneers become successful with a new technique, does it start to appear in real products that come with user's manuals and help lines.

Nowadays, new techniques are being developed; however, much work is also devoted to extending and improving existing techniques. All the techniques discussed in this book are available in commercial software products, although there is no single product that incorporates all of them.

How Data Mining Is Being Used Today

This whirlwind tour of a few interesting applications of data mining is intended to demonstrate the wide applicability of the data mining techniques discussed in this book. These vignettes are intended to convey something of the excitement of the field and possibly suggest ways that data mining could be profitably employed in your own work.

A Supermarket Becomes an Information Broker

Thanks to point-of-sale scanners that record every item purchased and loyalty card programs that link those purchases to individual customers, supermarkets are in a position to notice a lot about their customers these days.

Safeway was one of the first U.S. supermarket chains to take advantage of this technology to turn itself into an information broker. Safeway purchases address and demographic data directly from its customers by offering them discounts in return for using loyalty cards when they make purchases. In order to obtain the card, shoppers voluntarily divulge personal information of the sort that makes good input for actionable customer insight.

From then on, each time the shopper presents the discount card, his or her transaction history is updated in a data warehouse somewhere. With every trip to the store, shoppers teach the retailer a little more about themselves. The supermarket itself is probably more interested in aggregate patterns (what items sell well together, what should be shelved together) than in the behavior of individual customers. The information gathered on individuals is of great interest to the *manufacturers* of the products that line the stores' aisles.

Of course, the store assures the customers that the information thus collected will be kept private and it is. Rather than selling Coca-Cola a list of frequent Pepsi buyers and vice versa, the chain sells *access* to customers who, based on their known buying habits and the data they have supplied, are likely prospects for a particular supplier's product. Safeway charges several cents per name to suppliers who want their coupon or special promotional offer to reach just the right people. Since the coupon redemption also becomes an entry in the shopper's transaction history file, the precise response rate of the targeted group is a matter of record. Furthermore, a particular customer's response or lack thereof to the offer becomes input data for future predictive models.

American Express and other charge card suppliers do much the same thing, selling advertising space in and on their billing envelopes. The price they can charge for space in the envelope is directly tied to their ability to correctly identify people likely to respond to the ad. That is where data mining comes in.

A Recommendation-Based Business

Virgin Wines sells wine directly to consumers in the United Kingdom through its Web site, www.virginwines.com. New customers are invited to complete a survey, "the wine wizard," when they first visit the site. The wine wizard asks each customer to rate various styles of wines. The ratings are used to create a profile of the customer's tastes. During the course of building the profile, the wine wizard makes some trial recommendations, and the customer has a chance to agree or disagree with them in order to refine the profile. When the wine wizard has been completed, the site knows enough about the customer to start making recommendations.

Over time, the site keeps track of what each customer actually buys and uses this information to update his or her customer profile. Customers can update their profiles by redoing the wine wizard at any time. They can also browse through their own past purchases by clicking on the "my cellar" tab. Any wine a customer has ever purchased or rated on the site is in the cellar. Customers may rate or rerate their past purchases at any time, providing still more feedback to the recommendation system. With these recommendations, the web

site can offer customers new wines that they should like, emulating the way that stores like the Wine Cask have built loyal customer relationships.

Cross-Selling

USAA is an insurance company that markets to active duty and retired military personnel and their families. The company attributes information-based marketing, including data mining, with a doubling of the number of products held by the average customer. USAA keeps detailed records on its customers and uses data mining to predict where they are in their life cycles and what products they are likely to need.

Another company that has used data mining to improve its cross-selling ability is Fidelity Investments. Fidelity maintains a data warehouse filled with information on all of its retail customers. This information is used to build data mining models that predict what other Fidelity products are likely to interest each customer. When an existing customer calls Fidelity, the phone representative's screen shows exactly where to lead the conversation.

In addition to improving the company's ability to cross-sell, Fidelity's retail marketing data warehouse has allowed the financial services powerhouse to build models of what makes a loyal customer and thereby increase customer retention. Once upon a time, these models caused Fidelity to retain a marginally profitable bill-paying service that would otherwise have been cut. It turned out that people who used the service were far less likely than the average customer to take their business to a competitor. Cutting the service would have encouraged a profitable group of loyal customers to shop around.

A central tenet of customer relationship management is that it is more profitable to focus on "wallet share" or "customer share," the amount of business you can do with each customer, than on market share. From financial services to heavy manufacturing, innovative companies are using data mining to increase the value of each customer.

Holding on to Good Customers

Data mining is being used to promote customer retention in any industry where customers are free to change suppliers at little cost and competitors are eager to lure them away. Banks call it attrition. Wireless phone companies call it churn. By any name, it is a big problem. By gaining an understanding of *who* is likely to leave and *why*, a retention plan can be developed that addresses the right issues and targets the right customers.

In a mature market, bringing in a new customer tends to cost more than holding on to an existing one. However, the incentive offered to retain a customer is often quite expensive. Data mining is the key to figuring out which customers should get the incentive, which customers will stay without the incentive, and which customers should be allowed to walk.

Weeding out Bad Customers

In many industries, some customers cost more than they are worth. These might be people who consume a lot of customer support resources without buying much. Or, they might be those annoying folks who carry a credit card they rarely use, are sure to pay off the full balance when they do, but must still be mailed a statement every month. Even worse, they might be people who owe you a lot of money when they declare bankruptcy.

The same data mining techniques that are used to spot the most valuable customers can also be used to pick out those who should be turned down for a loan, those who should be allowed to wait on hold the longest time, and those who should always be assigned a middle seat near the engine (or is that just our paranoia showing?).

Revolutionizing an Industry

In 1988, the idea that a credit card issuer's most valuable asset is the information it has about its customers was pretty revolutionary. It was an idea that Richard Fairbank and Nigel Morris shopped around to 25 banks before Signet Banking Corporation decided to give it a try.

Signet acquired behavioral data from many sources and used it to build predictive models. Using these models, it launched the highly successful balance transfer program that changed the way the credit card industry works. In 1994, Signet spun off the card operation as Capital One, which is now one of the top 10 credit card issuers. The same aggressive use of data mining technology that fueled such rapid growth is also responsible for keeping Capital One's loan loss rates among the lowest in the industry. Data mining is now at the heart of the marketing strategy of all the major credit card issuers.

Credit card divisions may have led the charge of banks into data mining, but other divisions are not far behind. At Wachovia, a large North Carolina-based bank, data mining techniques are used to predict which customers are likely to be moving soon. For most people, moving to a new home in another town means closing the old bank account and opening a new one, often with a different company. Wachovia set out to improve retention by identifying customers who are about to move and making it easy for them to transfer their business to another Wachovia branch in the new location. Not only has retention improved markedly, but also a profitable relocation business has developed. In addition to setting up a bank account, Wachovia now arranges for gas, electricity, and other services at the new location.

And Just about Anything Else

These applications should give you a feel for what is possible using data mining, but they do not come close to covering the full range of applications. The data mining techniques described in this book have been used to find quasars, design army uniforms, detect second-press olive oil masquerading as "extra virgin," teach machines to read aloud, and recognize handwritten letters. They will, no doubt, be used to do many of the things your business will require to grow and prosper for the rest of the century. In the next chapter, we turn to how businesses make effective use of data mining, using the virtuous cycle of data mining.

Lessons Learned

Data Mining is an important component of analytic customer relationship management. The goal of analytic customer relationship management is to recreate, to the extent possible, the intimate, learning relationship that a well-run small business enjoys with its customers. A company's interactions with its customers generates large volumes of data. This data is initially captured in transaction processing systems such as automatic teller machines, telephone switch records, and supermarket scanner files. The data can then be collected, cleaned, and summarized for inclusion in a customer data warehouse. A well-designed customer data warehouse contains a historical record of customer interactions that becomes the memory of the corporation. Data mining tools can be applied to this historical record to learn things about customers that will allow the company to serve them better in the future. The chapter presented several examples of commercial applications of data mining such as better targeted couponing, making recommendations, cross selling, customer retention, and credit risk reduction.

Data mining itself is the process of finding useful patterns and rules in large volumes of data. This chapter introduced and defined six common data mining tasks: classification, estimation, prediction, affinity grouping, clustering, and profiling. The remainder of the book examines a variety of data mining algorithms and techniques that can be applied to these six tasks. To be successful, these techniques must become integral parts of a larger business process. That integration is the subject of the next chapter, *The Virtuous Cycle of Data Mining*.