

What Is Data Mining and Why Do It?

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. One of those businesses, a hair braider, no longer braids the hair of the little girl. In the years since the first edition, the little girl grew up, and moved away, and no longer wears her hair in cornrows. Her father, one of the authors, 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 the wines of Italy and Germany.

Decades later, the Wine Cask still has a loyal customer. That loyalty is no accident. The staff learns the tastes of their customers and their price ranges. When asked for advice, the response is based on 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, their intimate knowledge of each customer is what keeps customers coming back. Another wine shop could open across the street and hire a staff of expert oenophiles, but achieving the same level of intimate customer knowledge would take them months or years.

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. 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.

What Is Data Mining?

Although some data mining techniques are quite new, data mining itself is not a new technology, in the sense that people have been analyzing data on computers since the first computers were invented — and without computers for centuries before that. Over the years, data mining has gone by many different names, such as knowledge discovery, business intelligence, predictive modeling, predictive analytics, and so on. The definition of data mining as used by the authors is:

Data mining is a business process for exploring large amounts of data to discover meaningful patterns and rules.

This definition has several parts, all of which are important.

Data Mining Is a Business Process

Data mining is a business process that interacts with other business processes. In particular, a process does not have a beginning and an end: it is ongoing. Data mining starts with data, then through analysis informs or inspires action, which, in turn, creates data that begets more data mining.

The practical consequence is that organizations who want to excel at using their data to improve their business do not view data mining as a sideshow. Instead, their business strategy must include collecting data, analyzing data for long-term benefit, and acting on the results.

At the same time, data mining readily fits in with other strategies for understanding markets and customers. Market research, customer panels, and other techniques are compatible with data mining and more intensive data analysis. The key is to recognize the focus on customers and the commonality of data across the enterprise.

Large Amounts of Data

One of the authors regularly asks his audiences, “How much is a lot of data?” when he speaks. Students give answers such as, “all the transactions for 10 million customers” or “terabytes of data.” His more modest answer, “65,356 rows,” still gets sighs of comprehension even though Microsoft has allowed more than one million rows in Excel spreadsheets since 2007.

A tool such as Excel is incredibly versatile for working with relatively small amounts of data. It allows a wide variety of computations on the values in each row or column; pivot tables are amazingly practical for understanding data and trends; and the charts offer a powerful mechanism for data visualization.

In the early days of data mining (the 1960s and 1970s), data was scarce. Some of the techniques described in this book were developed on data sets containing a few hundred records. Back then, a typical data set might have had a few attributes about mushrooms, and whether they are poisonous or edible. Another might have had attributes of cars, with the goal of estimating gas mileage. Whatever the particular data set, it is a testament to the strength of the techniques developed in those days that they still work on data that no longer fits in a spreadsheet.

Because computing power is readily available, a large amount of data is not a handicap; it is an advantage. Many of the techniques in this book work better on large amounts of data than on small amounts — you can substitute data for cleverness. In other words, data mining lets computers do what computers do best — dig through lots and lots of data. This, in turn, lets people do what people do best, which is set up the problem and understand the results.

That said, some case studies in this book still use relatively small data sizes. Perhaps the smallest is a clustering case study in Chapter 13. This case study finds demographically similar towns, among just a few hundred towns in New England. As powerful as Excel is, it does not have a built-in function that says “group these towns by similarity.”

That is where data mining comes in. Whether the goal is to find similar groups of New England towns, or to determine the causes of customer attrition, or any of a myriad of other goals sprinkled throughout the chapters, data mining techniques can leverage data where simpler desktop tools no longer work so well.

Meaningful Patterns and Rules

Perhaps the most important part of the definition of data mining is the part about meaningful patterns. Although data mining can certainly be fun, helping the business is more important than amusing the miner.

In many ways finding patterns in data is not tremendously difficult. The operational side of the business generates the data, necessarily generating patterns at the same time. However, the goal of data mining — at least as the authors

use the term — is not to find just any patterns in data, but to find patterns that are useful for the business.

This can mean finding patterns to help routine business operations. Consider a call center application that assigns customers a color. “Green” means be very nice, because the caller is a valuable customer, worth the expense of keeping happy; “yellow” means use some caution because the customer may be valuable but also has signs of some risk; and “red” means do not give the customer any special treatment because the customer is highly risky. Finding patterns can also mean targeting retention campaigns to customers who are most likely to leave. It can mean optimizing customer acquisition both for the short-term gains in customer numbers and for the medium- and long-term benefit in customer value.

Increasingly, companies are developing business models centered around data mining — although they may not use that term. One company that the authors have worked with helps retailers make recommendations on the web; this company only gets paid when web shoppers click on its recommendations. That is only one example. Some companies aggregate data from different sources, bringing the data together to get a more complete customer picture. Some companies, such as LinkedIn, use information provided by some people to provide premium services to others — and everyone benefits when recruiters can find the right candidates for open job positions. In all these cases, the goal is to direct products and services to the people who are most likely to need them, making the process of buying and selling more efficient for everyone involved.

Data Mining and Customer Relationship Management

This book is not about data mining in general, but specifically about data mining for customer relationship management. 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 articles, books, and conferences. Everything from lead-tracking software to campaign management software to call center software gets labeled 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 company’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 (and more profitable) 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 so 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, managing investments, and maintaining their loss ratio. Turning a product-focused organization into a customer-centric one takes more than data mining. 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 or induce customers to be more profitable in the long term).

In a 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 analytic tools and techniques. For data mining to be effective, the other requirements for analytic CRM must also be in place. To form a learning relationship with its customers, a company 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.

Data mining, to repeat the earlier definition, is a business process for exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. This book assumes that the *goal* of data mining is to allow a company 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 in this book are equally applicable in fields as varied as law enforcement, radio astronomy, medicine, and industrial process control.

Why Now?

Most data mining techniques have existed, at least as academic algorithms, for decades (the oldest, survival analysis, actually dates back centuries). Data mining has caught on in a big way, increasing dramatically since the 1990s. This is due to the convergence of several factors:

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

The combination of these factors means that data mining is increasingly appearing as a foundation of business strategies. Google was not the first search engine, but it was the first search engine to combine sophisticated algorithms for searching with a business model based on maximizing the value of click-through revenue. Across almost every business domain, companies are discovering that they have information — information about subscribers, about Web visitors, about shippers, and payment patterns, calling patterns, friends and neighbors. Companies are increasingly turning to data analysis to leverage their information.

Data Is Being Produced

Data mining makes the most sense where large volumes of data are available. In fact, most data mining algorithms *require* somewhat large amounts of data to build and train models.

One of the underlying themes of this book is that data is everywhere and available in copious amounts. This is especially true for companies that have customers — and that includes just about all of them. A single person browsing a website can generate tens of kilobytes of data in a day. Multiply that by millions of customers and prospects and data volumes quickly exceed the size of a single spreadsheet.

The Web is not the only producer of voluminous data. Telephone companies and credit card companies were the first to work with terabyte-sized databases, an exotically large size for a database as recently as the late 1990s. That time has passed. Data is available, and in large volumes, but how do you make any sense out of it?

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 is such an important part of the data mining story that Chapter 17 is devoted to this topic. Data warehousing brings together data from many different sources in a common format with consistent definitions for keys and fields. Operational systems are designed to deliver results quickly to the end user, who may be a customer at a website or an employee doing her job. These systems are designed for the task at hand, and not for the task of maintaining clean, consistent data for analysis. 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 algorithms are also computationally intensive. The continuing dramatic decrease in prices for disk, memory, processing power, and network bandwidth has brought once-costly techniques that were used only in a few government-funded laboratories into the reach of ordinary businesses.

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. 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 (among other things). 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 a phrase such as “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, help lines, and training classes.

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 and open-source software products, although no single product incorporates all of them.

Skills for the Data Miner

Who can be a data miner? The answer is not everyone, because some specific skills are needed. A good data miner needs to have skills with numbers and a basic familiarity with statistics (and a stronger knowledge of statistics is always useful). Chapters 4 and 6 cover many of the key statistical concepts required for data mining. Having a good working knowledge of Excel is also very useful, because it is the predominant spreadsheet in the business world. Spreadsheets such as Excel are very useful for analyzing smallish amounts of data and for presenting the results to a wide audience.

Of course, familiarity with data mining techniques is critical for a data miner. The bulk of this book is devoted to various techniques. Understanding the techniques themselves is important; more important is understanding when and how they are useful. Perhaps as important as the technical details is the demystification of data mining techniques. Although many are quite sophisticated, they are often based on a very accessible foundation. These techniques are not magic. Even when you cannot explain exactly how they arrive at an answer, it is possible to understand them, without a Ph.D. in mathematics or statistics. The techniques are better than magic, because they are useful and help solve real-world problems.

Another very important skill for a data miner is really an attitude: lack of fear of large amounts of data and the complex processing that might be needed to squeeze out results. Working with large data sets, data warehouses, and analytic sandboxes is key to successful data mining.

Finally, data mining is not just about producing technical results. No data mining model, for instance, ever really did anything more than shift bits around inside a computer. The results have to be used to help people (or increasingly, automated processes) make more informed decisions. Producing the technical results is the end of the beginning of the data mining process. Being able to work with other people, communicate results, and recognize what is really needed are critical skills for a good data miner. Throughout this book are many examples of data mining in the business context, both in the next two chapters and throughout the technical chapters devoted to each technique. Data mining is a learning process based on data, as described in the next sections, and any good data miner must be open to new ideas.

The Virtuous Cycle of Data Mining

In the first part of the nineteenth century, textile mills were the industrial success stories. These mills sprang up in the growing towns and cities along rivers in England and New England to harness hydropower. Water, running over water

wheels, drove spinning, knitting, and weaving machines. For a century, the symbol of the industrial revolution was water pouring over wheels providing the power for textile machines.

The business world has changed. Old mill towns are now quaint historical curiosities. Long mill buildings alongside rivers are warehouses, shopping malls, artist studios, and sundry other businesses. Even manufacturing companies often provide more value in services than in goods. The authors were struck by an ad campaign by a leading international cement manufacturer, Cemex, that presented concrete as a service. Instead of focusing on the quality of cement, its price, or availability, the ad pictured a bridge over a river and sold the idea that “cement” is a service that connects people by building bridges between them. Concrete as a service? Welcome to the twenty-first century.

The world has changed. Access to electrical or mechanical power is no longer the criterion for business success. For mass-market products, data about customer interactions is the new waterpower; knowledge drives the turbines of the service economy and, because the line between service and manufacturing is getting blurry, much of the manufacturing economy as well. Information from data focuses sales and marketing efforts by targeting customers, improves product designs by addressing real customer needs, and enhances resource allocation by understanding and predicting customer preferences.

Data is at the heart of many core business processes. It is generated by transactions in operational systems regardless of industry — retail, telecommunications, manufacturing, health care, utilities, transportation, insurance, credit cards, and financial services, for example. Adding to the deluge of internal data are external sources of demographic, lifestyle, and credit information on retail customers; credit, financial, and marketing information on business customers; and demographic information on neighborhoods of all sizes. The promise of data mining is to find the interesting patterns lurking in all these billions and trillions of bits lying on disk or in computer memory. Merely finding patterns is not enough. *You must respond to the patterns and act on them, ultimately turning data into information, information into action, and action into value.* This is the virtuous cycle of data mining in a nutshell.

To achieve this promise, data mining needs to become an essential business process, incorporated into other processes including marketing, sales, customer support, product design, and inventory control. The virtuous cycle places data mining in the larger context of business, shifting the focus away from the discovery mechanism to the actions based on the discoveries. This book emphasizes *actionable* results from data mining (and this usage of “actionable” should definitely not be confused with its definition in the legal domain, where it means that some action has grounds for legal action).

Marketing literature makes data mining seem so easy. Just apply the automated algorithms created by the best minds in academia, such as neural networks, decision trees, and genetic algorithms, and you are on your way to untold successes.

Although algorithms are important, the data mining solution is more than just a set of powerful techniques and data structures. The techniques must be applied to the right problems, on the right data. The virtuous cycle of data mining is an iterative learning process that builds on results over time. Success in using data will transform an organization from reactive to proactive. This is the virtuous cycle of data mining, used by the authors for extracting maximum benefit from the techniques described later in the book. Before explaining the virtuous cycle of data mining, take a look at a case study of data mining in practice.

A Case Study in Business Data Mining

Once upon a time, there was a bank with a business problem. One particular line of business, home equity lines of credit, was failing to attract enough good customers. There are several ways the bank could attack this problem.

The bank could, for instance, lower interest rates on home equity loans. This would bring in more customers and increase market share at the expense of lowered margins. Existing customers might switch to the lower rates, further depressing margins. Even worse, assuming that the initial rates were reasonably competitive, lowering the rates might bring in the worst customers — the disloyal. Competitors can easily lure them away with slightly better terms. The sidebar “Making Money or Losing Money” talks about the problems of retaining loyal customers.

MAKING MONEY OR LOSING MONEY?

Home equity loans generate revenue for banks from interest payments on the loans, but sometimes companies grapple with services that lose money.

As an example, Fidelity Investments once put its bill-paying service on the chopping block because this service consistently lost money. Some last-minute analysis saved it, by showing that Fidelity’s most loyal and most profitable customers used the service. Although it lost money, Fidelity made much more money on these customers’ other accounts. After all, customers that trust their financial institution to pay their bills have a very high level of trust in that institution. Cutting such value-added services may inadvertently exacerbate the profitability problem by causing the best customers to look elsewhere for better service.

Even products such as home equity loans offer a conundrum for some banks. A customer who owns a house and has a large amount of credit card debt is a good candidate for a home equity line-of-credit. This is good for the customer, because the line-of-credit usually has a much lower interest rate than the original credit card. Should the bank encourage customers to switch their debt from credit cards to home equity loans?

Continued

MAKING MONEY OR LOSING MONEY? (continued)

The answer is more complicated than it seems. In the short term, such a switch is good for the customer, precisely because it is bad for the bank: Less interest being paid by the customer means less revenue for the bank. Within the bank, such a switch also causes a problem. The credit card group may have worked hard to acquire a customer who would pay interest every month. That group doesn't want to lose its good customers.

On the other hand, switching the customer over may build a lifetime relationship that will include many car loans, mortgages, and investment products. When the focus is on the customer, the long-term view is sometimes more important, and it can conflict with short-term goals.

In this particular example, the bank was Bank of America (BofA), which was anxious to expand its portfolio of home equity loans after several direct mail campaigns yielded disappointing results. The National Consumer Assets Group (NCAG) decided to use data mining to attack the problem, providing a good introduction to the virtuous cycle of data mining. (The authors would like to thank Lounette Dyer, Larry Flynn, and Jerry Modes who worked on this problem and Larry Scroggins for allowing us to use material from a Bank of America case study.)

Identifying BofA's Business Challenge

BofA needed to do a better job of marketing home equity loans to customers. Using common sense and business consultants, it came up with these insights:

- People with college-age children want to borrow against their home equity to pay tuition bills.
- People with high but variable incomes want to use home equity to smooth out the peaks and valleys in their income.

These insights may or may not have been true. Nonetheless, marketing literature for the home equity line product reflected this view of the likely customer, as did the lists drawn up for telemarketing. These insights led to the disappointing results mentioned earlier.

Applying Data Mining

BofA worked with data mining consultants from Hyperparallel (then a data mining tool vendor that was subsequently absorbed into Yahoo!) to bring a range of data mining techniques to bear on the problem. There was no shortage of data. For many years, BofA had been storing data on its millions of retail customers in a large relational database on a powerful parallel computer from

Teradata. Data from 42 *systems of record* was cleansed, transformed, aligned, and then fed into the corporate data warehouse. With this system, BofA could see all the relationships each customer maintained with the bank.

This historical database was truly worthy of the name — some records dated back to 1914! More recent customer records had about 250 fields, including demographic fields such as income, number of children, and type of home, as well as internal data. These customer attributes were combined into a customer signature, which was then analyzed using Hyperparallel's data mining tools.

Decision trees (a technique discussed in Chapter 7) derived rules to classify existing bank customers as likely or unlikely to respond to a home equity loan offer. The decision tree, trained on thousands of examples of customers who had obtained the product and thousands who had not, eventually learned rules to tell the difference between them. After the rules were discovered, the resulting model was used to add yet another attribute to each prospect's record. This attribute, the "good prospect for home equity lines of credit flag" flag, was generated by a data mining model.

Next, a sequential pattern-finding technique (such as the one described in Chapter 15 on market basket analysis and sequential pattern analysis) was used to determine *when* customers were most likely to want a loan of this type. The goal of this analysis was to discover a sequence of events that had frequently preceded successful solicitations in the past.

Finally, a clustering technique (described in Chapter 13) was used to automatically segment the customers into groups with similar attributes. At one point, the tool found fourteen clusters of customers, many of which did not seem particularly interesting. Of these fourteen clusters, though, one had two intriguing properties:

- 39 percent of the people in the cluster had both business and personal accounts.
- This cluster accounted for more than a quarter of the customers who had been classified by the decision tree as likely responders to a home equity loan offer.

This result suggested to inquisitive data miners that people might be using home equity loans to start businesses.

Acting on the Results

With this new insight, NCAG (the business unit for home equity lines of credit) teamed with the Retail Banking Division and did what banks do in such circumstances: They sponsored market research to talk to customers. Four times a year, BofA would circulate a survey to the bank branches to find out what was actually happening on the frontline. With the knowledge gained from data mining, the bank had one more question to add to the list: "Will the proceeds

of the loan be used to start a business?” The result from the data mining study was one question on an in-house survey.

The results from the survey confirmed the suspicions aroused by data mining. As a result, NCAG changed the message of its campaign from “use the value of your home to send your kids to college” to something more on the lines of “now that the house is empty, use your equity to do what you’ve always wanted to do.”

Incidentally, market research and data mining are often used for similar ends — to gain a better understanding of customers. Although powerful, market research has some shortcomings:

- Responders may not be representative of the population as a whole. That is, the set of responders may be biased, particularly by the groups targeted by past marketing efforts (forming what is called an *opportunistic sample*).
- Customers (particularly dissatisfied customers and former customers) have little reason to be helpful or honest.
- Any given action may be the culmination of an accumulation of reasons. Banking customers may leave because a branch closed, the bank bounced a check, and they had to wait too long at ATMs. Market research may pick up only the proximate cause, although the sequence is more significant.

Despite these shortcomings, talking to customers and former customers provides insights that cannot be provided in any other way. This example with BofA shows that the two methods are compatible.

TIP When doing market research on existing customers, using data mining to take into account what is already known about them is a good idea.

Measuring the Effects of Data Mining

As a result of a marketing campaign focusing on a better message, the response rate for home equity campaigns increased from 0.7 percent to 7 percent. According to Dave McDonald, vice president of the group, the strategic implications of data mining are nothing short of the transformation of the retail side of the bank from a mass-marketing institution to a learning institution. “We want to get to the point where we are *constantly* executing marketing programs — not just quarterly mailings, but *programs* on a consistent basis.” He has a vision of a closed-loop marketing process where operational data feeds a rapid analysis process that leads to program creation for execution and testing, which in turn generates additional data to rejuvenate the process. In short, the virtuous cycle of data mining.

Steps of the Virtuous Cycle

The BofA example shows the virtuous cycle of data mining in practice. Figure 1-1 shows the four stages:

1. Identifying business opportunities.
2. Mining data to transform the data into actionable information.
3. Acting on the information.
4. Measuring the results.

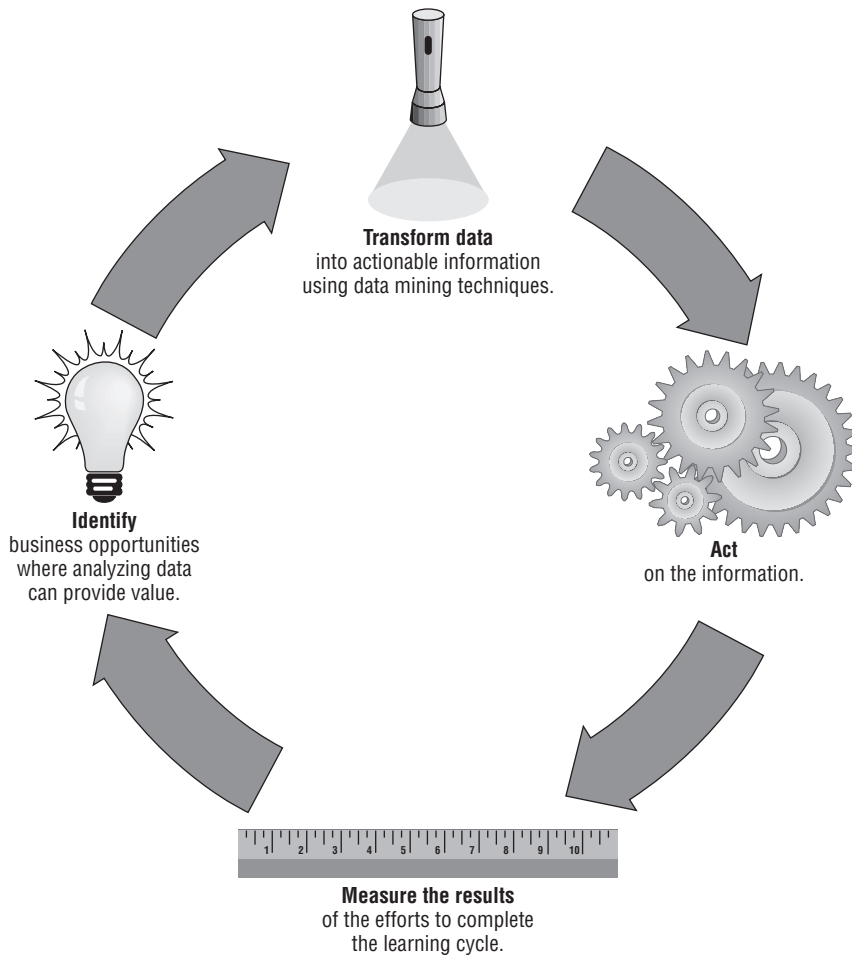


Figure 1-1: The virtuous cycle of data mining focuses on business results, rather than just exploiting advanced techniques.

As these steps suggest, the key to success is incorporating data mining into business processes and being able to foster lines of communication between the technical data miners and the business users of the results.

Identify Business Opportunities

The virtuous cycle of data mining starts by identifying the right business opportunities. Unfortunately, there are too many good statisticians and competent analysts whose work is essentially wasted because they are solving problems that don't help the business. Good data miners want to avoid this situation.

Avoiding wasted analytic effort starts with a willingness to act on the results. Many normal business processes are good candidates for data mining:

- Planning for a new product introduction
- Planning direct marketing campaigns
- Understanding customer attrition/churn
- Evaluating results of a marketing test
- Allocating marketing budgets to attract the most profitable customers

These are examples of where data mining can enhance existing business efforts, by allowing business managers to make more informed decisions — by targeting a different group, by changing messaging, and so on.

To avoid wasting analytic effort, it is also important to measure the impact of whatever actions are taken in order to judge the value of the data mining effort itself. As George Santayana said (in his full quote, of which only the last sentence is usually remembered):

Progress, far from consisting in change, depends on retentiveness. When change is absolute, there remains no being to improve and no direction set for possible improvement: and when experience is not retained, as among savages, infancy is perpetual. Those who do not learn from the past are condemned to repeat it.

In the data mining context, this also applies: If you cannot measure the results of mining the data, then you cannot learn from the effort and there is no virtuous cycle.

Measurements of past efforts and ad hoc questions about the business also suggest data mining opportunities:

- What types of customers responded to the last campaign?
- Where do the best customers live?
- Are long waits at automated tellers a cause of customer attrition?
- Do profitable customers use customer support?
- What products should be promoted with Clorox bleach?

Interviewing business experts is another good way to get started. Because people on the business side may not be familiar with data mining, they may not understand how to act on the results. By explaining the value of data mining to an organization, such interviews provide a forum for two-way communication.

One of the authors once participated in a series of meetings at a telecommunications company to discuss the value of analyzing call detail records (records of completed calls made by each customer). During one meeting, the participants were slow in understanding how this could be useful. Then, a colleague pointed out that lurking inside their data was information on which customers used fax machines at home (the details of the resulting project are discussed in Chapter 16 on link analysis). This observation got the participants thinking. Click! Fax machine usage would be a good indicator of who was working from home. For the work-at-home crowd, the company already had a product bundle tailored for their needs. However, without prodding from the people who understood the data and the techniques, this marketing group would never have considered searching through data to find a work-at-home crowd. Joining the technical and the business highlighted a very valuable opportunity.

TIP When talking to business users about data mining opportunities, make sure they focus on the business problems and not on technology and algorithms. Let the technical experts focus on the technology and let the business experts focus on the business.

Transform Data into Information

Data mining, the focus of this book, transforms data into actionable results. Success is about making business sense of the data, not using particular algorithms or tools. Numerous pitfalls interfere with the ability to use the results of data mining:

- Bad data formats, such as not including the zip code in the customer address.
- Confusing data fields, such as a delivery date that means “planned delivery date” in one system and “actual delivery date” in another system.
- Lack of functionality, such as a call-center application that does not allow annotations on a per-customer basis.
- Legal ramifications, such as having to provide a legal reason when rejecting a loan (and “my neural network told me so” is not acceptable).
- Organizational factors, because some operational groups are reluctant to change their operations, particularly without incentives.
- Lack of timeliness, because results that come too late may no longer be actionable.

Data comes in many forms, in many formats, and from multiple systems, as shown in Figure 1-2. Identifying the right data sources and bringing them together are critical success factors. Every data mining project has data issues: inconsistent systems, table keys that don't match across databases, records overwritten every few months, and so on. Complaints about data are the number one excuse for not doing anything. Chapters 17, 18, and 19 discuss various issues involving data, starting with data warehousing and working through the transformations into a format suitable for data mining. The real question is, "What can be done with available data?" This is where the techniques described later in this book come in.

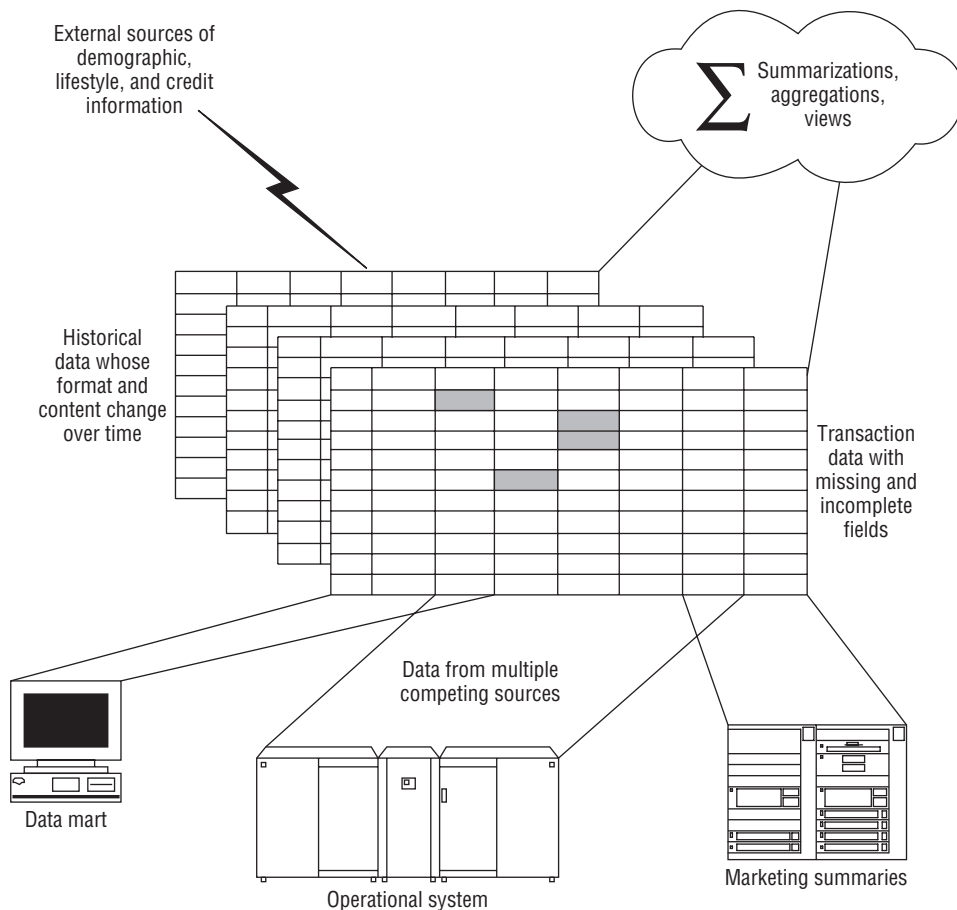


Figure 1-2: Data is never clean. It comes in many forms, from many sources both internal and external.

A wireless telecommunications company once wanted to put together a data mining group after having already acquired a powerful server and a data mining

software package. At this late stage, the company contacted the authors to help investigate data mining opportunities. One opportunity became apparent. A key factor for customer attrition was overcalls: new customers using more minutes than allowed by their rate plan during their first month. Customers would learn about the excess usage when the first bill arrived — sometime during the middle of the second month. By that time, the customers had run up large bills for the second month as well as the first and were even more unhappy. Unfortunately, the customer service group also had to wait for the same billing cycle to detect the excess usage. There was no lead time to be proactive.

However, the nascent data mining group had resources and had identified and investigated the appropriate data feeds. With some relatively simple programming, the group was able to identify these customers within days of their first overcall. With this information, the customer service center could contact at-risk customers and move them onto appropriate billing plans even before the first bill went out. This simple system was a big win, and a showcase for data mining. Simply having a data mining group — with the skills, hardware, software, and access — was the enabling factor for putting together the appropriate triggers to save at-risk customers.

Act on the Information

Taking action is the purpose of the virtuous cycle of data mining. As already mentioned, action can take many forms. Data mining makes business decisions more informed. Over time, better-informed decisions should lead to better results.

Sometimes, the “action” is simply doing what would have been done anyway — but with more (or less) confidence that the action will work. Even this is a success for data mining, because reducing the level of worry is a good thing.

More typically, actions are in line with what the business is doing anyway:

- Incorporating results into automated recommendation systems, when customers appear online
- Sending messages to customers and prospects via direct mail, e-mail, telemarketing, and so on; with data mining, different messages may go to different people
- Prioritizing customer service
- Adjusting inventory levels
- And so on

The results of data mining must feed into business processes that touch customers and affect the customer relationship.

Measure the Results

The importance of measuring results has already been highlighted, although this is the stage in the virtuous cycle most likely to be overlooked. The value of measurement and continuous improvement is widely acknowledged, and yet less attention than it deserves, because it has no immediate return-on-investment. How many business cases are implemented without anyone going back to see how well reality matched the plans? Individuals improve their own efforts by comparing and learning, by asking questions about why plans match or do not match what really happened, and by being willing to learn when and how earlier assumptions were wrong. What works for individuals also works for organizations.

Commonly, marketing efforts are measured based on financial measures — and these are very important. However, modeling efforts should also be measured. Consider what happened once at a large Canadian bank that had a plan to cross-sell investment accounts to its customers. This marketing message was all over the bank: in television and radio advertisements, in posters in the branch, in messages printed on the back of ATM receipts, in messages while customers were on hold for customer service, and so on. Customers could not miss the messages.

This story, though, concerns a different channel, direct mail. A data mining effort identified customers most likely to respond to an investment campaign offer. A marketing campaign was designed and targeted at customers who were likely to respond. In this case, though, the bank included a special holdout group: This group was predicted to respond well, but did not receive the direct mail. (The sidebar “Data Mining and Marketing Tests” discusses this idea in more detail.) Holding out potential responders is a rather controversial action for the direct mail manager. The data miners are saying, “This is a group that we think will respond, but don’t contact all of them; leave some out so we can learn from this test.”

What was learned was quite worth the cost of not contacting some good customers. Among customers who scored high for the investment account offer, the same proportion opened accounts regardless of whether they received the offer or not. The model did, indeed, find customers who would open the accounts. However, the marketing test also found that the marketing communication was superfluous. Given all the other marketing efforts, this particular direct mail campaign was not needed.

The time to start thinking about measurement is at the beginning when identifying the business problem. How can results be measured? A company that sends out coupons to encourage sales of its products will no doubt measure the coupon redemption rate. However, coupon-redeemers may have purchased the product anyway. Another appropriate measure is increased sales in particular stores or regions, increases that can be tied to the particular marketing effort. Such measurements may be difficult to make, because they require more detailed sales information. However, if the goal is to increase sales, there needs to be a way to measure this directly or indirectly. Otherwise, marketing efforts may be all “sound and fury, signifying nothing.”

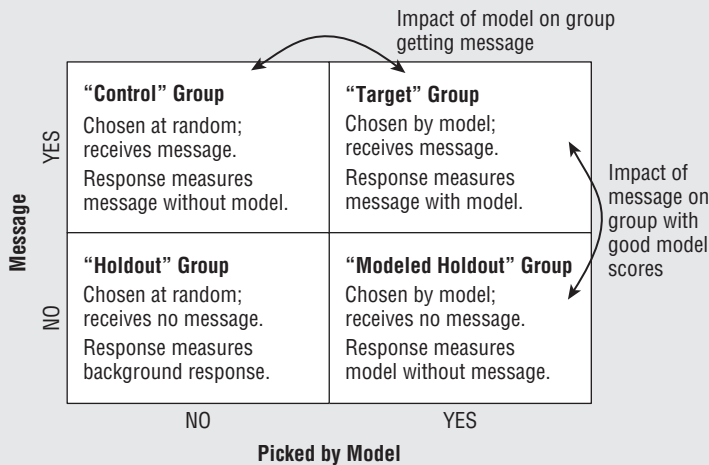
DATA MINING AND MARKETING TESTS

Marketing tests are an important part of analytic marketing, as is data mining. The two often complement each other, and marketing tests are an important part of understanding whether data mining efforts are working. Typically two things should be tested when using data mining for a marketing treatment. First, is the marketing message working? Second, is the data mining modeling working?

The key is to use holdout groups intelligently to understand these two factors. In practice, four potential groups exist:

- **Target Group:** Receives the treatment and has model scores indicating response.
- **Control Group:** Receives the treatment and is chosen either at random or based on lower model scores.
- **Holdout Group:** Does not receive the treatment and is chosen either at random or based on lower model scores.
- **Modeled Holdout Group:** Does not receive the treatment and has model scores indicating response.

These four groups are indicated in the following figure:



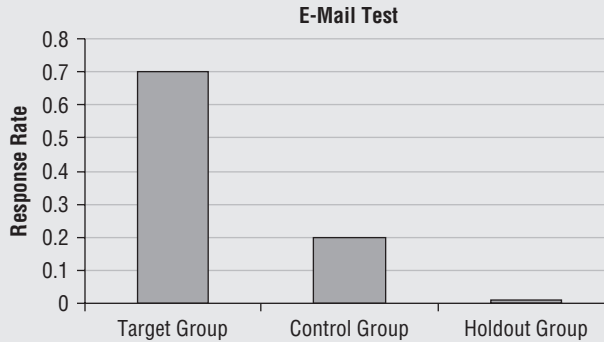
These four groups are used for measuring the effectiveness of both the message and the modeling effort.

The responses from these four groups then provide useful information. Using these groups for modeling is called *incremental response modeling* and is discussed in more detail in Chapter 5.

In the example where the Canadian bank learned that the direct mail effort was unnecessary, the response rates for the Modeled Holdout were the same as for the Target Group. This indicates that the treatment is not having an effect. The difference between the Target Group and the Control Group measures whether or not the modeling is working.

DATA MINING AND MARKETING TESTS (continued)

The following chart is an example from another bank that uses such charts to measure the effectiveness of its campaigns. This chart is the actual chart from the bank:



This chart readily shows the difference in response to determine whether the treatment works and whether the modeling works.

The first two bars show that the Target Group has a higher response rate than the Control Group, indicating that the modeling is working. The second two bars show that the Control Group has a higher response rate than the Holdout Group, indicating that the marketing treatment is working.

Just measuring these four groups is really the beginning of measuring the effectiveness of data mining. For instance, model scores are often broken into deciles. In such cases, it is important to include a sample from all deciles in the campaign to be sure that the model is working. Of course, everyone in the top deciles gets included in the effort (because this achieves the business goal). For the lower deciles, only a sample is included. The sample should be big enough to determine whether the deciles are really working – something that is quite important when using models. Chapter 4 explains the statistics background for determining the right size for such tests.

Standard reports, which may arrive weeks or months after marketing interventions have occurred, contain summaries. Marketing managers may not have the technical skills to glean important findings from such reports, even if the information is there. Understanding the impact on customer retention means tracking old marketing efforts for even longer periods of time. Well-designed reporting applications can be a big help for marketing groups and marketing analysts. However, for some questions, even more detail is needed.

Thinking of every marketing effort as a small business case is a good idea. Comparing expectations to actual results makes it possible to recognize promising

opportunities to exploit on the next round of the virtuous cycle. You are often too busy tackling the next problem to devote energy to measuring the success of current efforts. This is a mistake. Every data mining effort, whether successful or not, has lessons that can be applied to future efforts. The question is what to measure and how to approach the measurement so it provides the best input for future use.

As an example, let's start with what to measure for a targeted acquisition campaign. The canonical measurement is the response rate: How many people targeted by the campaign actually responded? This leaves a lot of information lying on the table. For an acquisition effort that uses a model score (where a high score indicates a higher likelihood of response), some examples of questions that have future value are:

- Did this campaign reach and bring in profitable customers?
- Did a higher model score indicate a higher response rate?
- Were these customers retained as well as would be expected?
- What are the characteristics of the most loyal customers reached by this campaign?
- Did the newly acquired customers purchase additional products?
- Did some messages or offers work better than others?
- Did customers reached by the campaign respond through alternate channels?

All of these measurements provide information for making more informed decisions in the future. Data mining is about connecting the past — through learning — to future actions.

One particular measurement is *lifetime customer value*. As its name implies, this is an estimate of the value of a customer during the entire course of his or her relationship (or perhaps for some fixed period in the future, such as for the next two years). In some industries, quite complicated models have been developed to estimate lifetime customer value. Even without sophisticated models, shorter-term estimates, such as value after one month, six months, and one year, can prove to be quite useful. Customer value is discussed in more detail in the next chapter.

Data Mining in the Context of the Virtuous Cycle

Consider a large telecommunications company in the United States. Such a company has millions of customers. It owns hundreds or thousands of switches located in central offices, which are typically in several states in multiple time zones. Each switch can handle thousands of calls simultaneously — including

advanced features such as call waiting, conference calling, call-forwarding, voice mail, and digital services. Switches, among the most complex computing devices yet developed, are available from a handful of manufacturers. A typical telephone company has multiple versions of several switches from each of the vendors. Each of these switches provides volumes of data in its own format on every call and attempted call — volumes measured in tens of gigabytes each day. In addition, each state has its own regulations affecting the industry, not to mention federal laws and regulations that are subject to rather frequent changes. To add to the confusion, the company offers thousands of different billing plans to its customers, which range from occasional residential users to Fortune 100 corporations.

How does this company — or any similar company with large volumes of data and large numbers of customers — manage its billing process, the bread and butter of its business, responsible for its revenue? The answer is simple: very carefully! Companies have developed detailed processes for handling standard operations; they have policies and procedures. These processes are robust. Bills go out to customers, even when the business reorganizes, even when database administrators are on vacation, even when computers are temporarily down, even as laws and regulations change, even when switches are upgraded, and when hurricanes strike. If an organization can manage a process as complicated as getting accurate bills out every month to millions of residential, business, and government customers, surely incorporating data mining into decision processes should be fairly easy. Is this the case?

Large companies have decades of experience developing and implementing mission-critical applications for running their business. Data mining is different from the typical operational system (see Table 1-1). The skills needed for running a successful operational system do not necessarily lead to successful data mining efforts.

Problems addressed by data mining differ from operational problems — *a data mining system does not seek to replicate previous results exactly*. In fact, replication of previous efforts can lead to disastrous results. It may result in marketing campaigns that target the same people over and over. You do not want to learn from analyzing data that a large cluster of customers fits the profile of the customers contacted in the previous campaign. Data mining processes need to take such issues into account, unlike typical operational systems that want to reproduce the same results over and over — whether completing a telephone call, sending a bill, authorizing a credit purchase, tracking inventory, or other countless daily operations.

Data mining is a creative process. Data contains many obvious correlations that are either useless or simply represent current business policies. For example,

Table 1-1: Data Mining Differs from Typical Operational Business Processes

TYPICAL OPERATIONAL SYSTEM	DATA MINING SYSTEM
Operations and reports on historical data	Analysis on historical data often applied to most current data to determine future actions
Predictable and periodic flow of work, typically tied to calendar	Unpredictable flow of work depending on business and marketing needs
Focus on individual items, one at a time (the needle in the haystack)	Focusing on larger groups at one time, trying to make sense of the haystack
Limited use of enterprise-wide data	The more data, the better the results (generally)
Focus on line of business (such as account, region, product code, minutes of use, and so on), not on customer	Focus on actionable entity, product, customer, sales region
Response times often measured in seconds/milliseconds (for interactive systems) while waiting weeks/month for reports	Iterative processes with response times often measured in minutes or hours
System of record for data	Copy of data
Descriptive and repetitive	Creative

analysis of data from one large retailer revealed that people who buy maintenance contracts are also very likely to buy large household appliances. Unless the retailer wanted to analyze the effectiveness of sales of maintenance contracts with appliances, such information is worse than useless because the maintenance contracts in question are only sold with large appliances. Spending millions of dollars on hardware, software, and data miners to find such results is a waste of resources that can better be applied elsewhere in the business. Analysts must understand what is of value to the business and how to arrange the data to bring out the nuggets.

Data mining results change over time. Models expire and become less useful as time goes on. One cause is that data ages quickly. Markets and customers change quickly as well.

Data mining provides feedback into other processes that may need to change. Decisions made in the business world often affect current processes and interactions with customers. Often, looking at data finds imperfections in operational systems, imperfections that should be fixed to enhance future customer understanding.

Lessons Learned

Data mining is an important part of customer relationship management. The goal of customer relationship management is to re-create, 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 generate 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. This 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. To be successful, data mining must become an integral part of a larger business process, *the virtuous cycle of data mining*.

The virtuous cycle of data mining is about harnessing the power of data and transforming it into actionable business results. Just as water once turned the wheels that drove machines throughout a mill, data must be gathered and disseminated throughout an organization to provide value. If data is water in this analogy, then data mining is the wheel, and the virtuous cycle spreads the power of the data to all the business processes.

The virtuous cycle of data mining is a learning process based on customer data. It starts by identifying the right business opportunities for data mining. The best business opportunities are those that will be acted upon. Without action, little or no value is to be gained from learning about customers. Also very important is measuring the results of the action. This completes the loop of the virtuous cycle, and often suggests further data mining opportunities.

The next chapter puts data mining in the context of customers themselves, starting with the customer lifecycle and following with several examples of the virtuous cycle in action.