CHAPTER 8

Industry Applications of Data Mining

This chapter contains examples of how data mining is used in banking/finance, retailing, healthcare, and telecommunications. The purpose of this chapter is to give the user some ideas of the types of activities in which data mining is already being used and what companies are using them.

The chapter is organized as follows:

8.1 Data-Mining Applications in Banking and Finance

Data mining has been used extensively in the banking and financial markets. In the banking industry, data mining is heavily used to model and predict credit fraud, to evaluate risk, to perform trend analysis, and to analyze profitability, as well as to help with direct marketing campaigns.

In the financial markets, neural networks have been used in stock-price forecasting, in option trading, in bond rating, in portfolio management, in commodity price prediction, in mergers and acquisitions, as well as in forecasting financial disasters.

Several of the financial companies who use neural networks and have been referenced on the Internet are Daiwa Securities, NEC Corporation, Carl & Associates, LBS Capital Management, Walkrich Investment Advisors, and O'Sullivan Brothers Investments. The number of investment companies that data mine is much more extensive than this, but they are not willing to be referenced.

One interesting book in the area of global finance is *Neural Networks in the Capital Markets*, edited by Apostolos-Paul Refenes. The book explores equity applications, foreign exchange applications, bond applications, and macroeconomic and corporate performance. Most of the contributed chapters are from university professors, a group whose publishing in the areas of economic and capital market forecasting is most impressive; however, there are government and industry contributions from Citibank N.A., Daimler Benz AG, County NatWest Investment Management, and NeuroDollars.

8.1.1 Stock Forecasting

There are many software applications on the market that use data-mining techniques for stock prediction. One such application used for stock prediction is shown in Figure 8-1.

Figure 8-1 Stock forecasting.

NETPROPHET by Neural Applications Corporation is a stock-prediction application that makes use of neural networks. The two lines shown in the graph in Figure 8-1 represent the real and the predicted stock values.

In banking, the most widespread use of data mining is in the area of fraud detection. HNC's Falcon product specifically addresses this area. HNC comments that credit fraud detection is now in place to monitor more than 160 million payment-card accounts this year. They also claim a healthy return on investment. While fraud is decreasing, applications for payment card accounts are rising as much as 50% a year.

The widespread use of data mining in banking has not been unnoticed. In 1996, *Bank Systems & Technology* commented: "Data mining is the most important application in financial services in 1996."

Finding banking companies who use data mining is not easy, given their proclivity for silence. The following list of financial companies that use data mining required some digging into SEC reports from data mining vendors that are made available to the public. The list includes: Bank of America, First USA Bank, Headlands Mortgage Company, FCC National Bank, Federal Home Loan Mortgage Corporation, Wells Fargo Bank, Nations-Banc Services, Mellon Bank N.A., Advanta Mortgage Corporation, Chemical Bank, Chevy Chase Bank, U.S. Bancorp, and USAA Federal Savings Bank. Again it is reasonable to assume that most large banks are performing some sort of data mining, although many have policies not to discuss it.

8.1.2 Cross-Selling and Customer Loyalty in the Banking Industry

Most major financial institutions have statistics and data-mining groups. In fact, banks like Wells Fargo, Bank of America, Fleet Bank, and others have been the subject of many articles about their sophisticated data mining, and modeling of their customers' behavior. The next question to ask is: how well do financial institutions know their customers? A study published in DM News and conducted by Deluxe Corporation found that 43% of consumers surveyed said their financial service provider does not know their specific needs well at all; 60% said the offers they received were not relevant to their needs; and 39% said they did not receive offers at all.

The study by Deluxe Corporation demonstrates a significant problem with data mining: the inability to leverage data-mining studies into actionable results. For example, while a bank may know that customers meeting certain criteria are likely to close their accounts, it is another matter to figure out a strategy to do something about it. One vendor that has developed a suite of products designed at integrating predictive technologies with customer interaction points is RightPoint software. Other vendors are working on the same problem, particularly on the web, where predicting what a customer will best respond to is critical. Web banking companies like Security First and BroadVision, among others, are also trying to incorporate one-to-one marketing, using predictive technologies, to their banking sites.

The RightPoint Real-Time Marketing Suite takes data-mining models and leverages them within real-time interactions with customers. The RightPoint Real-Time Marketing Suite is designed to create, manage, and deliver 1:1 marketing campaigns for high-touch industries (such as banking, telecommunications, and retail sales) that rely on direct customer interaction to conduct business. For these and similar businesses, it is essential to ensure that each customer interaction seizes the opportunity to increase customer satisfaction, loyalty, and revenue-generation potential. Predictive models are used to evaluate the right marketing message to be delivered to customers. Dynamic learning technology also builds predictive models on the fly and calculates probabilities of acceptance, indicating which offers are being accepted by which types of customers. (See collaborative filtering as discussed in Chapter 4, for a discussion of one dynamic learning technology) These predictive models can also be used in conjunction with business rules to provide the right offer at the right time.

One aspect of pinpointing market opportunities is identifying high-value customers. In his book, *All Consumers are Not Created Equal*, author Garth Hallberg cites MediaMark Research, Inc. findings that about one-third of customers account for 68% of all purchases. Traditionally, marketers have focused on segmenting and courting high-value consumers. Where marketers have fallen short is in taking that understanding of high-value customers and using this information to predict the qualities that would raise the value of mid-level consumers, opening a large (and largely untapped) market opportunity.

Real-time marketing focuses on executing one-to-one campaigns that utilize predictive technologies to capture a sense of personalization. The idea is that by tailoring marketing options to consumers, companies get a better response rate for their campaigns.

Equally important, businesses now have an effective outlet for building loyalty and brand value, by tapping into customers' demands for personalized service, and their desire to escape the hassle of researching different service offerings. For example, a mortgage customer may tell the lending bank about an existing auto loan. An agent of the bank can add this information to the customer's profile, and present back a pre-approved refinance of the auto loan. This will save the customer money by consolidating the existing mortgage and auto loan with one bank. If the bank can calculate the savings on the fly, the customer can see a clear benefit.

Halifax Bank Using Real-Time Marketing

Halifax PLC, the second largest bank in the United Kingdom, has chosen its RightPoint Real-Time Marketing Suite as the foundation for a customer relationship initiative. RightPoint will enable Halifax customer service representatives to mobilize vital information about a customer and determine which campaigns, products or services to offer at the point of customer contact.

Halifax's direct customer service center receives more than 20 million customer calls per year and employs 800 customer service representatives. With the call center increasingly becoming the customer interaction center, a customer's decision to do business with a company is often based on whether a company is aware of that customer's preferences and acts upon them accordingly. Using RightPoint, Halifax representatives will have a valuable tool for reliably predicting and delivering on the requirements of their customers in real-time, thereby increasing customer satisfaction and loyalty, and attaining an important competitive advantage.

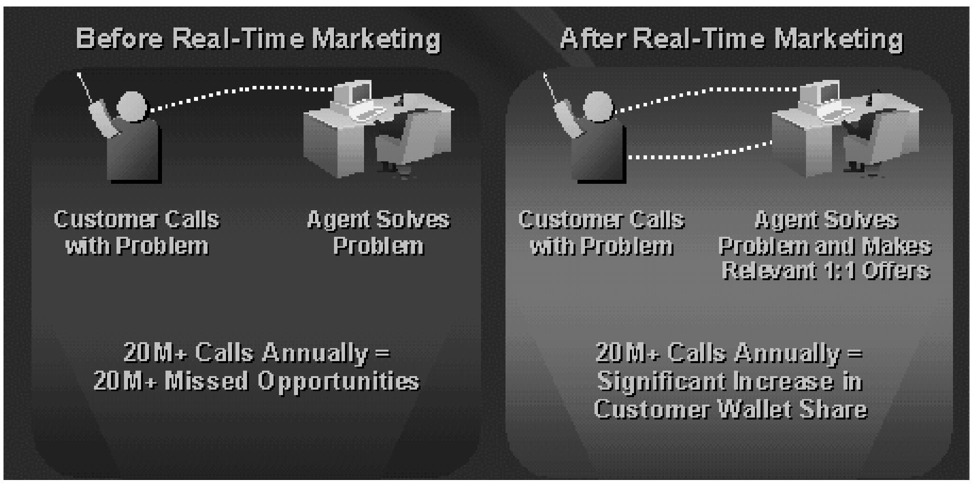


Figure 8-2 Halifax Bank engaging in real-time marketing.

"Our direct channel is playing an ever-increasing role in delivering customer contact, with call volumes predicted to grow to more than 50 million calls per year over the next three years," says Dick Spelman, director of distribution at Halifax. "We need to ensure that we can harness customer data at the point of contact so that a customized service is delivered in real-time. This is what the RightPoint solution will deliver to our agents. The other parallel challenge that call centers face is generating sales income. Rather than add more agents and use questionable handover techniques, RightPoint offers us the potential to convert inbound service calls into profitable sales. If organizations don't tackle the revenue-generation aspect of their call-center activities, they will not be able to afford the current unbridled growth in service traffic.

"RightPoint gives Halifax the ability to leverage each and every customer interaction and make one-to-one marketing a reality," Spelman continues. "By capitalizing on the untapped revenue potential present during each customer interaction, Halifax will be able to grow the lifetime value of its customers while also reaching out and building stronger relationships. Halifax is looking to see a significant increase in response rates to campaigns that will ultimately help increase its market share in this highly competitive industry."

Delivering Predictive Technologies in a Real-Time Environment

An analyst may have built a data-mining model that can predict that 30% of mortgage customers meeting a certain set of criteria would agree to taking out auto loans with you if you could make a compelling offer. The challenge is to take this knowledge and:

- Deliver it to customer-contact points.
- Put it in combination with business rules.
- Leverage information you may have gathered during customer interaction.
- Provide an immediate feedback loop on the effectiveness of an active marketing campaign.
- Allow marketers to fine-tune their campaigns on the fly.

By combining these capabilities in a closed-loop system, businesses have the ability to react quickly to market conditions and significantly improve customer-response rates.

Looking further into the architecture of any real-time marketing software solution, there should be three primary components: a tool for targeting marketing campaigns, a campaign server, and a suite of applications for moving the campaigns out to various customer touch points.

Figure 8-3 shows the components of this system that are required to deliver predictive technologies in a real-time environment:

- Predictive models for mass marketing, represented in the left-hand oval.
- Customer information, which may include transactional data, a customer marketing database, and information that the customer just gave you while interacting with you (represented in the right-hand oval).
- A predictive engine capable of delivering the predictions in real time (seconds on web or call center).
- Business rules, which state when to use which predictive models (i.e. this model is only used when the person calling in has a family, has over $30,000 with us, and has not been pitched a product before).
- A feedback loop of responses that will monitor the success of the technology as well as allow marketers to dynamically learn from it.

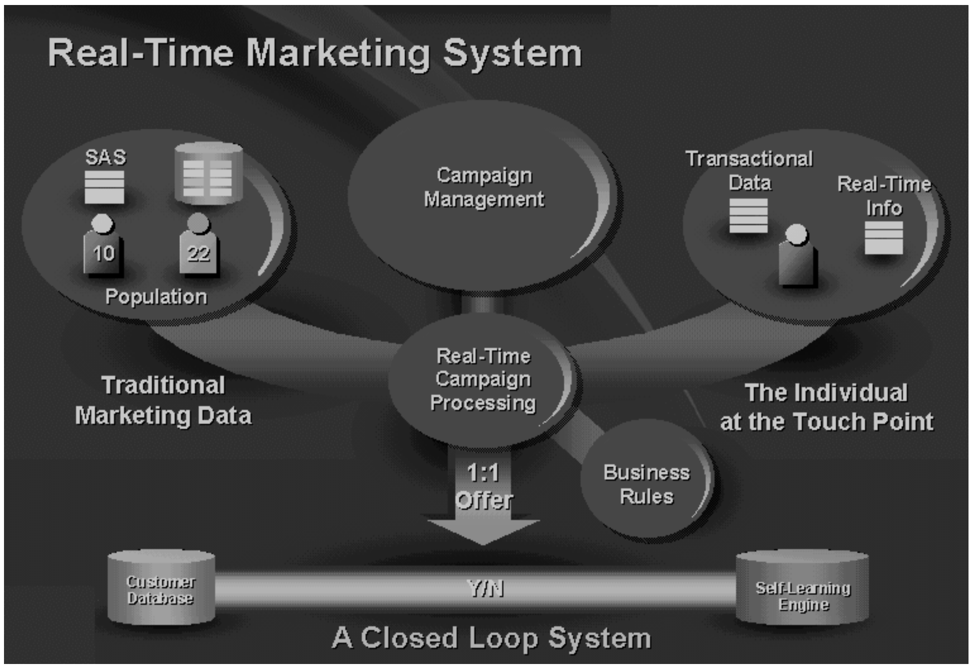


Figure 8-3 Using predictive technologies with a real-time marketing system.

The real-time campaign marketing tool is used by marketers to collect and capture market specifications for new campaigns. Because segmentation is at least half the effort of any marketing campaign, an important function of this tool is providing pinpoint accuracy when segmenting target markets. Campaign marketing tools with embedded data-mining technology provide advanced pattern recognition for high accuracy and predictive capabilities that enable marketers to proactively plan campaigns. Another important feature is the ability to support incremental updates, so that any time a customer's profile changes — for example, marital status — that data can be automatically factored into a marketer's customer segmentation. This ensures that customers are targeted to the appropriate campaigns at all times.

A campaign server administers and monitors active campaigns to enable one-to-one marketing. As customer information comes in from one of the touch point applications, the server should be able to take this up-to-date information and make on-the-fly predictions about the marketing campaigns that will be the most effective; for example, in building that customer's loyalty or obtaining additional revenue through up-selling and cross-selling. The campaign server then sends the appropriate campaigns up to the touch point application. Ideally, the campaign server will also store the results of active marketing campaigns, allowing marketers to evaluate their effectiveness and make any necessary adjustments in real-time.

Finally, touch point applications actually deliver marketing campaigns to the user. These applications transparently call down to the campaign server with the customer information, and then present the most appropriate marketing offer(s). For example, a touch point application for the call center will reside on each call center agent's computer and "pop up" up-selling or cross-selling opportunities on the fly, presenting the agent with a script. Similarly, when a consumer visits the Web site and types in an ID, the Web touch point application can present screens with offers or information of particular interest to that person.

Businesses cannot afford to wait months, or even weeks, to evaluate the effectiveness of a marketing campaign. Central to real-time marketing is having a closed-loop system that allows marketers to easily review the effectiveness of active marketing campaigns and then immediately adjust the campaign if it's not performing as expected. The ability to fine-tune a marketing campaign in a matter of days allows marketers to maximize response rates and focus resources on the highest value campaigns.

Emerging enterprise software for incorporating data mining technology with real-time marketing will play a central role in the ability of businesses to build both their brand value and the value of their customers. By providing pinpoint segmentation, companies can now conduct personalized, one-to-one customer marketing on a broad scale. Powerful capabilities for updating and reviewing campaign-related information on the fly means that marketers are serving today's customer needs, not yesterday's. It also means that businesses have a feedback loop for evaluating marketing campaigns and immediately adjusting them for greater effectiveness. The ability to extend marketing campaigns in real time across multiple touch points provides companies with unprecedented customer outreach. Most importantly, companies can now conduct effective, proactive marketing campaigns for the first time.

8.2 Data-Mining Applications in Retail

Slim margins have pushed retailers into embracing data warehousing earlier than other industries. Retailers have seen improved decision-support processes lead directly to improved efficiency in inventory management and financial forecasting. The early adoption of data warehousing by retailers has given them a better opportunity to take advantage of data mining. Large retail chains and grocery stores store vast amounts of point-of-sale data that is information rich. In the forefront of the applications that have been adopted in retail are direct marketing applications.

Chapter 5 used the example of a direct-mail campaign to show the usefulness of data mining in retail-marketing activities. The direct-mail industry is an area where data mining, or data modeling, is widely used. Almost every type of retailer, including catalogers, consumer retail chains, grocers, publishers, business-to-business marketers, and packaged goods manufacturers, uses direct marketing. There are many vertical applications that support direct-marketing campaigns, such as HNC's Marksman product. The claim could be made that every Fortune 500 company today has used data mining in a direct marketing campaign, usually through outsourcing lists to third parties like Harte-Hanks or The Polk Company.

Direct marketers are often concerned about customer segmentation, which is a clustering problem in data mining. Many vendors offer customer segmentation packages, like the one shown in Figure 8-4.

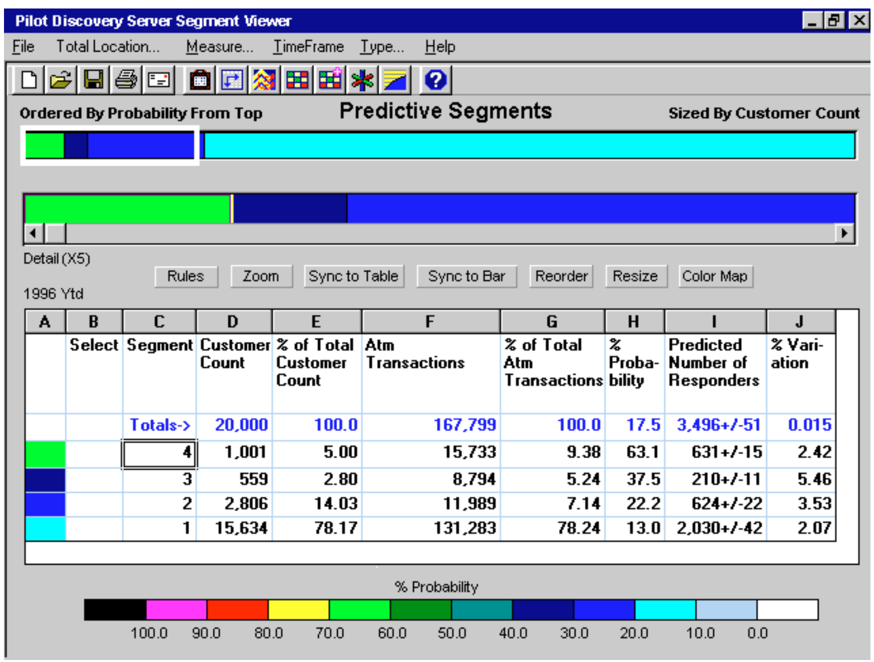


Figure 8-4 Customer segmentation software. Courtesy of Pilot Software.

Pilot Software also uses the customer segmentation to help in direct-mailing campaigns, as shown in Figure 8-5.

IBM has used data mining for several retailers to analyze shopping patterns within stores based on point of sale (POS) information. For example, one retail company with $2 billion in revenue, 300,000 UPC codes, and 129 stores in 15 states found some interesting results after analyzing its sales information. A store executive comments: "We found that people who were coming into the shop gravitated to the left-hand side of the store for promotional items and were not necessarily shopping the whole store." Such information is used to change promotional activity and provide a better understanding of how to lay out a store in order to optimize sales.

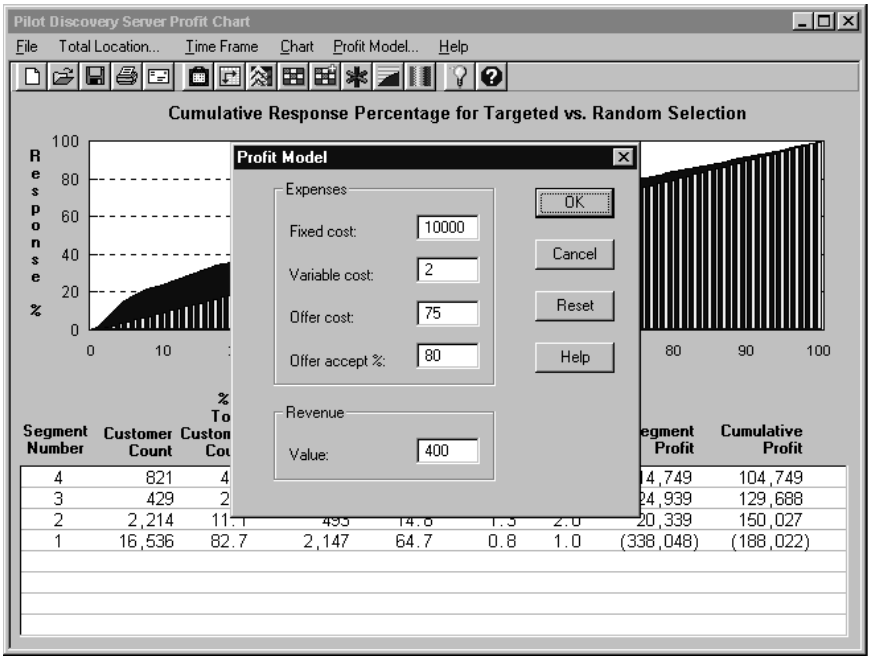


Figure 8-5 An application for a direct marketing campaign. Courtesy of Pilot Software.

Other Types of Retail Data-Mining Studies Retailers are interested in many different types of data-mining studies. In the area of marketing, retailers are interested in creating data-mining models to answer questions like:

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

Merchandisers are beginning to profile issues such as:

- What types of customers are buying specific products?

- What determines the best product mix to sell on a regional level?
- What are the latest product trends?
- When is a merchandise department saturated?
- What are the times when a customer is most likely to buy?
- What types of products can be sold together?

In discussing customer profitability, customers may wish to build models to answer questions like:

- How does a retailer retain profitable customers?
- What are the significant customer segments that buy products?

Customer identification is critical to successful retail organizations, and is likely to become more so. Data mining helps model and identify the traits of profitable customers and reveal the "hidden" relationship that standard query processes have not already found. For further reading on the area of customer management, one interesting work is the book *The One-to-One Future* by D. Peppers and M. Rogers.

8.2.1 An Example of Data Mining for Property Valuation

One application of data mining in real estate is the AREAS Property Valuation product from HNC Software, which performs property valuation as shown in Figure 8-6.

While some would not categorize the real estate market as a retail industry, the concept of using data mining to predict property valuations can be directly applied to any product or commodity. For example, the proper valuations of antique furniture, used cars, or clothing apparel could be predicted in the same manner.

Another application of data mining in the airline industry is a customer retention management package by SABRE Decision Technologies™. SABRE is a leader in working with the airline industry to use data warehousing to increase profitability, and make better business decisions.

Some companies that use data mining in retail, and that have been referenced in articles or by data-mining companies, are Victoria's Secret, National Car Rental, JOCKEY International, Marriott Ownership, the Reader's Digest, and WalMart. In Chapter 2, a sample figure from MapInfo Corporation shows a visualization application for locating optimal site locations for businesses.

AREAS ™ *Property Valuation*

Automated Real Estate Analysis System

Macro Model

October 18, 1995

Loan ID: 1234
Property ID: 0008

Subject Information

Borrower: Jones
Address: 13915 Oak St
City: San Diego
County: SAN DIEGO
State: CA
Zip Code: 92129 - 2185

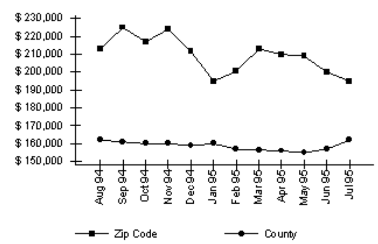
Appraisal Comparison

Last Appraisal: September 1995
Price: $200,000
% Difference: -1.1%
$ Difference: $-2,131

Valuation Range

High: $214,856
Predicted Value: **$197,869**
Low: $180,882
% Range: 8.6%

Median Price Chart



Neighborhood Medians

	Subject	*ZIP + 4*	*Neighborhood*
*Sale Price **	$197,869	$192,500	$208,000
Recent Sales	N/A	1	7
Living Area	1549	1490	1802
Room Count	7 - 3 - 2.00	6 - 3 - 2.00	8 - 4 - 2.50
Fireplaces	1	2	1
Garage Spaces	2	2	2
Year Built	1989	1988	1988
Lot Size	4,750	4,661	4,661

Neighborhood Sales

	Subject	*Sale 1*	*Sale 2*	*Sale 3*	*Sale 4*	*Sale 5*	*Sale 6*
*Sale Price **	**$197,869**	$194,000	$205,000	$208,000	$209,000	$185,500	$210,000
Sale Date	**N/A**	Sep-94	Jul-95	Dec-94	Jul-95	Jan-95	Apr-95
Distance	**N/A**	0.37	0.37	0.37	0.37	0.00	0.37
Nbhd Level	**B**	B	B	B	B	B	B
Property Type	**SFR**	SFR	SFR	SFR	SFR	SFR	SFR
Lot Size	**4,750**	4,661	4,403	4,312	4,312	7,100	6,706
View	**No**	No	No	No	No	No	No
Stories	**1**	2	2	2	2	1	2
Year Built	**1989**	1988	1988	1988	1988	1988	1988
Room Count	7 - 3 - 2.00	7 - 4 - 3.00	8 - 4 - 2.50	7 - 4 - 2.50	8 - 4 - 2.50	6 - 3 - 2.00	8 - 4 - 3.00
Living Area	**1549**	1894	1771	1802	1715	1553	1845
Central Air	**No**	No	Yes	No	No	No	No
Garage Spaces	**2**	2	2	2	2	2	0
Pool	**No**	No	No	No	No	No	No
Fireplaces	**1**	1	1	0	1	1	1

** Predicted Value*

Addresses

1 - 13442 WALNUT CT
SAN DIEGO
CA 92129-4403

2 - 13479 WALNUT CT
SAN DIEGO
CA 92129-4403

3 - 13458 CHESTNUT LN
SAN DIEGO
CA 92129-4403

4 - 13466 SPRUCE AVE
SAN DIEGO
CA 92129-4403

5 - 8818 PINE ST
SAN DIEGO
CA 92129-2172

6 - 13414 SPRUCE AVE
SAN DIEGO
CA 92129-4403

About this Report

This AREAS valuation and report is based on statistical processes derived from public record data. Accuracy of the data and the resulting conclusions of the statistical analysis cannot be guaranteed.

This report does not constitute an appraisal of the subject, nor should it be solely relied upon in place of competent appraisal or underwriting processes.

1995 HNC Software Inc.
5930 Cornerstone Court West, San Diego, CA 92121-3728
(619) 546-8877, FAX (619) 452-6524

4.0||M3|AUG94|JUL95

Figure 8-6 A data-mining application for property valuation from HNC Software, Inc.

8.2.2 An Example of Analyzing Customer Profitability in Retail

In the previous chapters we addressed methodologies that extract information from transactional data to produce "product-focused," actionable recommendations. In this section, we describe a methodology where the actionables are "customer-focused." This is an area in which Dovetail Solutions has significant expertise. Dovetail Solutions' methodology is the Value, Activity, and Loyalty™ Method, or VAL™ for short. VAL™ uses transactional data to extract information about customer activity, churn rate, and expected future purchases.

Customer value is not only determined by past revenues, but also by the customer's expected future purchasing behavior. This can be measured by customer activity and expected lifetime. Activity gauges the likelihood that a customer will purchase again, while lifetime measures how long a customer is expected to remain active. This would allow a retailer to measure the overall "health" of its customer base by determining profitability and "churn" rate. In turn, retailers could use these findings to determine customer acquisition goals needed to meet future revenue or profitability objectives.

Retail customers rarely ever tell a store that they have stopped shopping there (have become "inactive"). Therefore, it is highly advantageous for a store to know how many "active" customers reside in its customer base, how much sales revenue is expected from them, and what the customer churn or attrition rate is.

A common method of measuring churn rates is by looking at rules of thumb based on recency. For example, if a customer has not shopped for a long hiatus, say for the past year, she is considered inactive. While intuitive, this rule of thumb is overly simplistic in that it ignores differences in purchasing behavior across customer segments and individuals. For example, suppose that a customer shops twice a month. If this frequent shopper became inactive eight months ago, we may still count her as "active" because her last purchase is within the arbitrarily specified hiatus of one year. However, given her purchase habits, it is unlikely that this customer is still active. Likewise, infrequent shoppers might be incorrectly classified as inactive when in fact they are still active.

However, there is usually not enough data on individual customers to be able to adequately extract their purchasing patterns. The VAL methodology addresses this limitation by pooling the entire customer base to robustly estimate individual customer behavior based on limited individual purchase history. The underlying mathematical description of the overall customer population is based on "hazard rate" types of models, extracted from analogous processes in the natural sciences. This is validated by many studies that have shown that customer purchase patterns follow trends and regularities that can be accurately described using these types of models. An interesting reference, and one that influenced this discussion, is *Repeat Buying: Facts, Theory and Applications*, by A.S.C. Ehrenberg (Oxford University Press, 1988).

The VAL method uses transactional data to measure the probability that any given customer is active (the activity), gauges how many active customers reside in the customer database, determines the customer-base churn or attrition rate, and forecasts revenues from the currently active customers. It also extracts useful bellwether information, such as average customer lifetime (how long customers are expected to remain active), and average repurchase rates. This methodology is superior to the classic RFM (Recency, Frequency, and Monetary) analysis because it is forward-looking, as opposed to backward-looking, and produces more actionable results.

For example, segmenting the customer base by activity and value can suggest marketing strategies to stimulate those customers who are marginally active, but who have high expected value. Churn rates in conjunction with revenue forecasts can be used to determine what customer acquisition rate is required to meet revenue or profitability goals. Churn rates over time can also be used to identify and counteract seasonal periods that might trigger inactivity.

Transactional data can be a very valuable asset to retailers because of the actionable information it can generate if it is analyzed and mined carefully. With today's computing power and affordability, mining transactional data is no longer reserved for the large retailers. Mid-size and small retailers can now routinely collect and analyze transactional data. Moreover, there is less need to rely on outside vendors of panel data, since much of the information can be obtained directly from in-house transactional data. Market Basket Analysis, Assortment Optimization (discussed in earlier chapters), and the Value, Activity, and Loyalty methodology are examples of techniques that generate both product-focused and customer-focused actionables.

8.3 Data-Mining Applications in Healthcare

Chapter 3 discussed types of studies that can be done in the healthcare industry, as well as data-preparation issues. With the amount of information and issues in the healthcare industry, not to mention the information from medical research, biotechs, and the pharmaceutical industry, the types of studies listed in Chapter 3 are only the tip of the iceberg for data-mining opportunities.

Data mining has been used extensively in the medical industry already. For example, NeuroMedical Systems used neural networks to perform a pap smear diagnostic aid. Vysis uses neural networks to perform protein analysis for drug development. The University of Rochester Cancer Center and the Oxford Transplant Center use KnowledgeSEEKER, a decision tree technology, to help with their research. The Southern California Spinal Disorders Hospital uses Information Discovery to data mine. Information Discovery quotes one doctor as saying" "Today alone, I came up with a diagnosis for a patient who did not even have to go through a physical exam."

8.3.1 Uses of Data Visualization in the Medical Industry

Data visualization is one area that has built interest in the medical field. Belmont Research's CrossGraphs product has been used in many different applications. For example, Figure 8-7 shows a diagram for studying healthcare costs.

The graph shows the average cost-per-patient for fee-for-services patients, HMO patients, and other patients. For the categories 14 and 112, costs for "other" payer types varies widely.

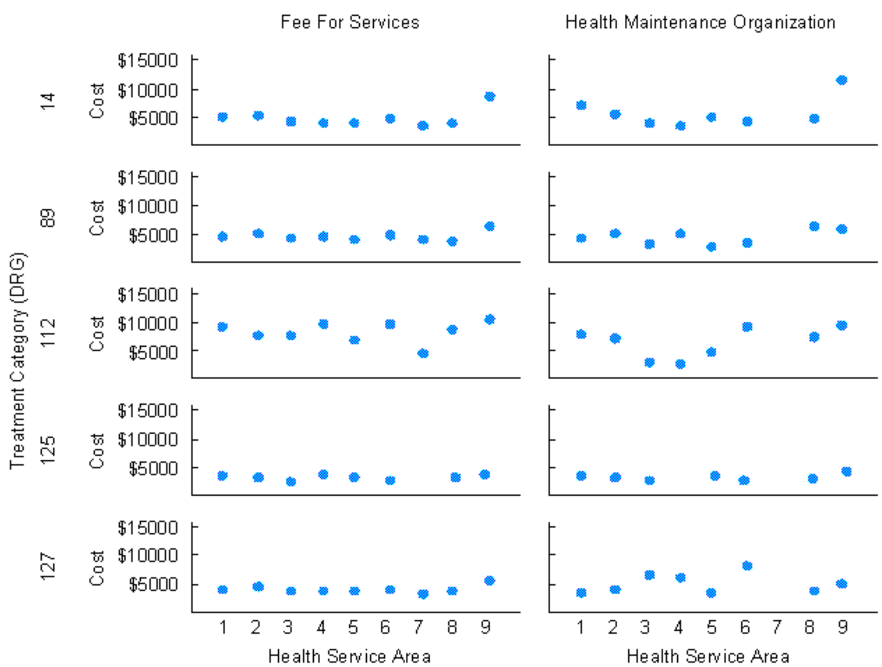


Figure 8-7 Average cost per patient by health service area, treatment category (DRG), and payer type (Belmont Research).

Another example, shown in Figure 8-8, is an array of graphs that show, side-by-side, a story of antibacterial activity of Cefdinir over time.

Figure 8-8 is useful for comparing the efficacy rates of different antibacterial pathogens over time. In this case, the antimicrobial agent, Cefdinir, is being studied against other agents for an eight-hour period.

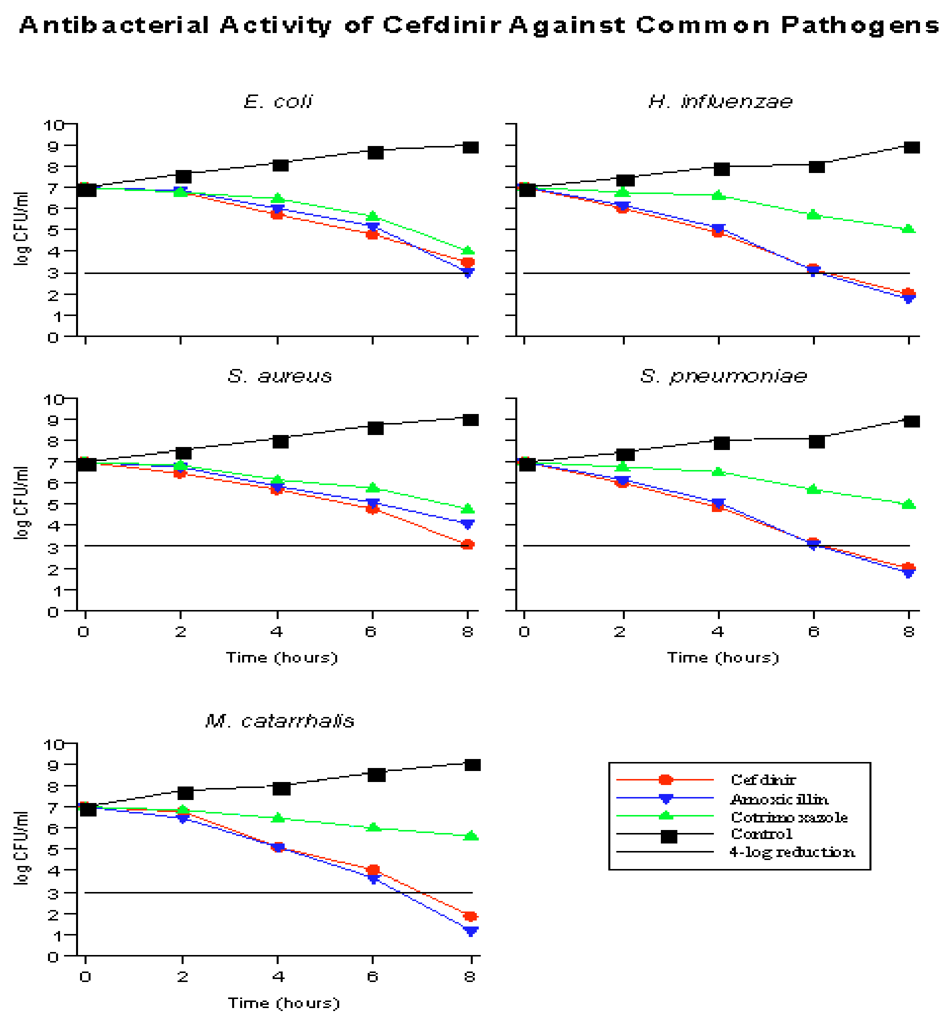


Figure 8-8 Efficacy of several antibacterial drugs over time (Belmont Research, Inc.).

Another example of a very useful application of data visualization is from MapInfo, using mapping technology to show patient location in order to deliver better service, as shown in Figure 8-9.

8.4 Data-Mining Applications in Telecommunications

In recent years, the telecommunications industry has undergone one of the most dramatic makeovers of any industry. The U.S. Telecommunications Act of 1996 allowed Regional Bell Operating Companies (RBOCS) to enter the long-distance market and offer “cable-like” services. The European Liberalization of Telecommunications Services, effective January 1, 1998, liberalized telecommunications services in Europe, and offers full competition among participating European nations. Sixty-eight nations liberalized their telecommunications market on January 1, 1998 to coincide with the European commitment based on the World Trade Organization’s Telecommunications Agreement.

Not only has there been massive deregulation, but in the United States, there has been a sell-off by the FCC of airwaves to companies pioneering new ways to communicate. The cellular industry is rapidly taking on a life of its own.

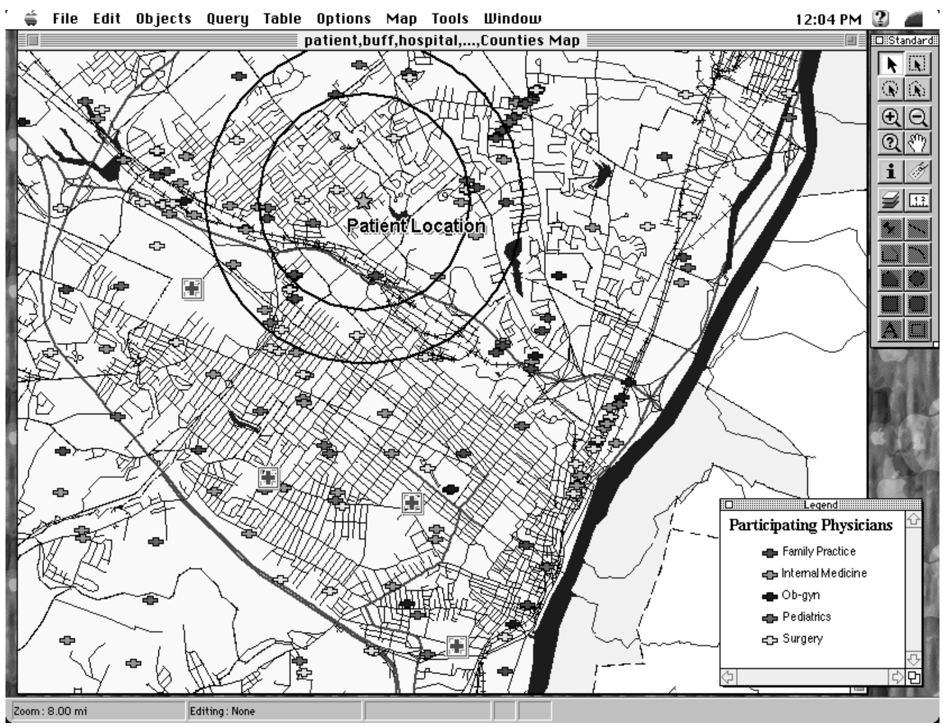


Figure 8-9 Mapping locations of physicians, patients, and patient care facilities.

With the hyper-competitive nature of this industry, a need to understand customers, to keep them, and to model effective ways to market new products to these customers is driving a demand for data mining in telecommunications where no demand existed in distant memory.

Companies like AT&T®, GTE Telecommunications®, and AirTouch® Communications have announced the use of data mining. American Management Systems® (AMS) Mobile Communications Industry Group has taken an active interest in data mining as well. AMS and AT&T offer consulting services around data mining, as do GTE and Cincinnati Bell Information Services®, among others.

Coral Systems® of Longmont, Colorado is a company that incorporates data-mining techniques in their FraudBuster™ product, which tracks known types of fraud by modeling subscriber usage patterns and predicting when a carrier is suspected of fraud. There are several companies looking at cellular fraud for telecommunications, including Lightbridge® and GTE.

Several other companies offer products to combat customer churn. For example, RightPoint Corporation focuses on data-mining issues in the telecommunications industry and, in particular, customer retention or churn. Industry experts have pointed out that the cellular telephone market experiences a 30% churn rate in the United States. A report by Digital Equipment Corporation®, produced by Evan Davies and Hossein Pakraven in September 1995, quantifies the cost of customer churn. In their report, they estimate that the cost of acquiring new customers is as high as $400 for each new subscriber.

Data visualization is another area with many strategic uses in telecommunications. Figure 8-10 shows a map, created by Empower Geographics® using MapInfo's technology, showing problem areas for a wireless telecommunications network.

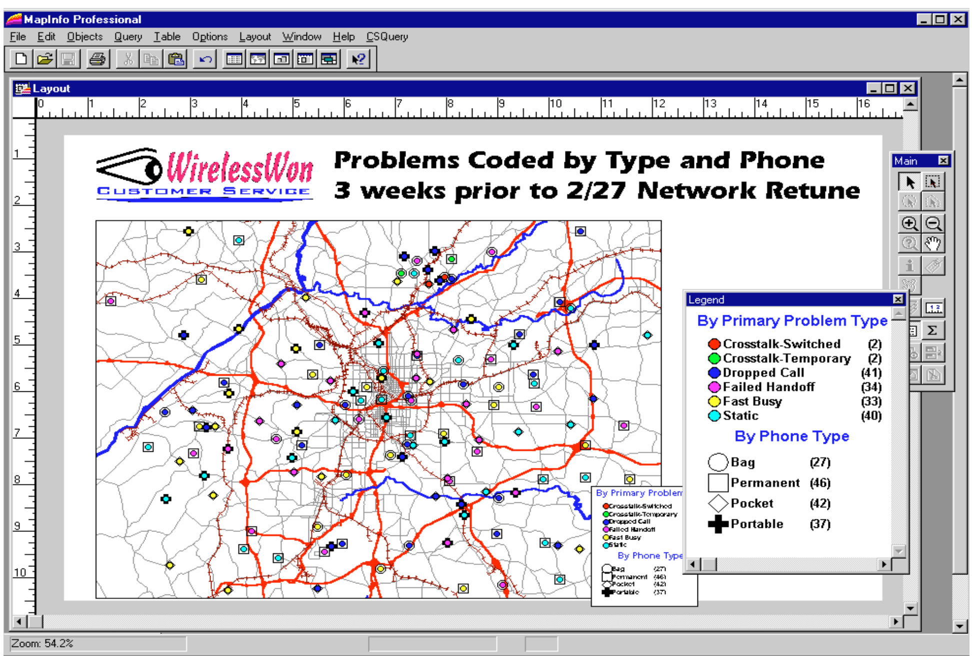


Figure 8-10 A map of a wireless telecommunications network pinpoints dropped calls.

8.4.1 Types of Studies in Telecommunications

The telecommunications industry is interested in answering a wide variety of questions with the help of data mining. For example:

- How does one recognize and predict when cellular fraud occurs?
- How does one retain customers and keep them loyal when competitors offer special offers and reduced rates?
- Which customers are most likely to churn?
- What characteristics make a customer likely to be profitable or unprofitable?
- How does one predict whether customers will buy additional products like cellular service, call waiting, or basic services?
- What are the factors that influence customers to call more at certain times?
- What characteristics indicate high-risk investments, such as investing in new fiber-optic lines?
- What products and services yield the highest amount of profit?
- What characteristics differentiate our products from those of our competitors?
- What set of characteristics indicates companies or customers who will increase their line usage?

8.5 Summary

This chapter covered industry examples of data mining in banking and finance, retail, healthcare, and telecommunications. While this is certainly not an inclusive list of all data mining activities, it does provide examples of how data mining is employed today. Chapter 8 will discuss specific data-mining studies for these industries, and will attempt to describe many of the data-preparation issues involved in performing these studies. More experienced users of data mining acknowledge that accumulation and preparation of data are the biggest hurdles to beginning the process of data mining.