18

Data Mining in Sales Marketing and Finance

Objectives:

- Data mining enables companies to identify trends within the data warehouse.
- The recent advances in technology have enabled companies to couple these technologies (data mining and campaign management) more tightly, with the following benefits: increased speed with which they can plan and execute marketing campaigns, increased accuracy and response rates of campaigns, and higher overall marketing return on investment.
- The key to making a successful data mining software product is to embrace the business problems that the technology is meant to solve, not to incorporate the hottest technology.
- Intelligent information agents scan the Internet for relevant information and aggregate it into an information service to automate the process of monitoring and aggregate relevant online sources and capturing significant content for decision makers.
- The concept of a sector-oriented agent-based online information system is formed, which focuses on the external information needs of enterprises in individual sectors.
- Data mining helps the company to understand the patterns behind past purchase transactions, thereby helping in the design and launch of new offerings, in an expeditious and cost-effective manner.
- "Market basket" analysis is formed that suits the application of affinity rules to analyzing consumer purchases.
- Data mining applied in finance supports financial asset management and risk management. Its use in computational finance will have a major impact in the modeling of currency markets, in tactical asset allocation, bond and stock valuation, and portfolio optimization.
- Financial data collected in the banking and financial industries is often relatively complete, reliable, and of high quality, which facilitates systematic data analysis and data mining.

S. Sumathi and S.N. Sivanandam: *Data Mining in Sales Marketing and Finance*, Studies in Computational Intelligence (SCI) **29**, 411–438 (2006)
www.springerlink.com

Abstract. The data mining applications in various fields like sales database marketing, integrating customer value and campaign management software, and financial data analysis are described in this chapter.

The popular types of applications that leverage companies' investments in data warehousing are data mining and campaign management software. Data mining and campaign management have been successfully deployed by hundreds of Fortune 1000 companies around the world, with impressive results. But recent advances in technology have enabled companies to couple these technologies more tightly, with the following benefits: increased speed with which they can plan and execute marketing campaigns, increased accuracy and response rates of campaigns, and higher overall marketing return on investment.

The data mining process results in the creation of a model. A model embodies the discovered patterns and can be used to make predictions for records for which the true behavior is unknown. These predictions, usually called *scores*, are numerical values that are assigned to each record in the database and indicate the likelihood that the customer will exhibit a particular behavior. These numerical values are used to select the most appropriate prospects for a targeted marketing campaign. Campaign management software enables companies to deliver to customers and prospects timely, pertinent, and coordinated offers, and also manages and monitors customer communications across all channels.

The market for data mining will be billions of dollars by the turn of the century. Unfortunately, much of what is now considered data mining will be irrelevant, since it is disconnected from the business world. In general, marketing analysts predictions that the technology of data mining will be very relevant to businesses in the future are correct. The key to making a successful data mining software product is to embrace the business problems that the technology is meant to solve, not to incorporate the hottest technology. In Section. 18.2 we address some of the issues related to the development of data mining technology as it relates to business users.

Information agent technology is an approach, which is used to present to improve the worse situations occurring. Section 18.3 introduces the concept of a sector-oriented agent-based online information system that focuses on the external information needs of enterprises in individual sectors of the agrifood industry.

Data mining software can help find the "high-profit" gems buried in mountains of information. However, merely identifying the best prospects is not enough to improve customer value. We must somehow fit the data mining results into the execution of marketing campaigns that enhance the profitability of customer relationships. Unfortunately, data mining and campaign management technologies have followed separate paths – until now. The organization stands to gain a competitive edge by understanding and utilizing this new union. Section 18.4 describes how we can profit from the integration of data mining and campaign management technologies.

Section 18.5 discusses some of the problems of the consumer package goods industry, a case study of some of the challengers presented to data miners within this industry, and critiques current knowledge discovery research in these areas.

Data mining is being increasingly applied in finance, especially to support financial asset management and risk management. It is considered by many financial management institutions as an innovative technology to support conventional quantitative techniques. Its use in computational finance will have a major impact in the modeling of currency markets, in tactical asset allocation, bond and stock valuation, and portfolio optimization. In addition the application of data mining for scoring

tasks delivers valuable support for the management of client credit risk and fraud detection. The last subsection of this chapter discusses how data mining is applied for financial data analysis

18.1 Data Mining can Bring Pinpoint Accuracy to Sales

Data warehousing – the practice of creating huge, central stores of customer data that can be used throughout the enterprise – is becoming more and more commonplace. But data warehouses are useless if companies do not have the proper applications for accessing and using the data.

Two popular types of applications that leverage companies' investments in data warehousing are data mining and campaign management software. Data mining enables companies to identify trends within the data warehouse (such as "families with teenagers are likely to have two phone lines," in the case of a telephone company's data). Campaign management software enables them to leverage these trends via highly targeted and automated direct marketing campaigns (such as a telemarketing campaign intended to sell second phone lines to families with teenagers).

Data mining automates the detection of patterns in a database and helps marketing professionals improve their understanding of customer behavior, and then predict behavior. For example, a pattern might indicate that married males with children are twice as likely to drive a particular sports car than married males with no children. A marketing manager for an auto manufacturer might find this somewhat surprising pattern quite valuable.

The data mining process can model virtually any customer activity. The key is to find patterns relevant to current business problems. Typical patterns that data mining uncovers include those customers who are most likely to drop a service, who are likely to purchase merchandise or services, and who are most likely to respond to a particular offer.

Unfortunately, for most companies today, the use of data mining models within campaign management is a manual, time-intensive process. When a marketer wants to run a campaign based on model scores, he or she has to call a modeler (usually a statistician) to have a model run against a database so that a score file can be created. The marketer then has to solicit the help of an IT staffer to merge the scores with the marketing database. This disjointed process is fraught with problems and errors and can take weeks. Often, by the time the models are integrated with the database, either the models are outdated or the campaign opportunity has passed.

The solution is the tight integration of data mining and campaign management technologies. Under this scenario, marketers can invoke statistical models from within the campaign management application, score customer segments on the fly, and quickly create campaigns targeted to customer segments offering the greatest potential. Here is how it works:

Step 1: Creating the Model

A modeler creates a predictive model using the data mining application. He or she then exports the model to a campaign management application, possibly by simply by dragging and dropping the data from one application to the other. This process of exporting a model tells the campaign management software that the model exists and is available for later use.

Step 2: Dynamically Scoring the Data

Once a model has been put into the campaign management system, marketers can then reference the model's score just as they would reference any other piece of data. Records can be selected based on the score, in conjunction with other characteristics in the data. When the campaign is run, the records in the database are scored dynamically using the model.

Dynamic scoring avoids manual integration of scores with the database and eliminates the need to score an entire database. Instead, dynamic scoring marks only relevant customer subsets and only when needed. This shrinks marketing cycle times and assures fresh, up-to-date results. Once a model is in the campaign management system, the user can start to build marketing campaigns based upon it simply by choosing it from a menu of options.

Any company that is creating or has created a data warehouse should be considering the use of integrated data mining and campaign management applications, which unlock the data and put it to use. By discovering customer behavior patterns and then acting upon them quickly, companies can stave off competition; and increase customer retention, cross selling and up-selling, all of which ultimately contribute to higher overall revenues.

18.2 From Data Mining to Database Marketing

18.2.1 Data Mining vs. Database Marketing

The current state-of-the-art analysis of databases is done by high-tech analysts (typically statisticians) using sophisticated tools, e.g., SAS or S-Plus. In essence these analysts are manual data miners. In contrast, data mining software technology promises to automate that analysis, allowing business users (who do not have a Ph.D. in statistics) to develop a more accurate and sophisticated understanding of their data.

Before we go any further, it is probably a good idea to discuss the terminology found in much of the data mining literature. There seems to be a multitude of terms related to the process of analyzing information contained in a database: data mining, database mining, and database marketing. Is there a difference between these terms?

Let us start with the technology. The technology is *data mining*. Data mining is, in some ways, an extension of statistics, with a few artificial intelligence and machine learning twists thrown in. Like statistics, data mining is not a business solution – it is just the underlying technology. Statistics does not by itself solve business problems. Unfortunately, data mining is being touted as a business solution when it is simply the base technology upon which business solutions might be built.

Database mining, which incorporates the ability to access directly data stored in a database, is one step beyond the core technology of data mining. The distinction (database rather than data) might seem to be a trivial improvement, but like most transitions from technology to solutions it requires a major leap for developers. For example, at a recent data mining conference, only one presenter discussed how their work interacted with a database. All the other presenters assumed that the data was available in flat files or that any interaction with a database was so irrelevant as to be not worth mentioning. However, anyone familiar with commercial information processing knows the critical impact of interacting with data stored in relational databases (RDBMS).

Database marketing, on the other hand, supports a variety of business processes. It involves transforming a database into business decisions. For example, consider a catalog retailer who needs to decide who to send a new catalog to. The information incorporated into the database marketing process is the historical database of previous mailings and the features associated with the (potential) customers, such as age, zip code, their response in the past, etc. The database marking software would use this information to build a model of customer behavior that would generate a mailing list of customers most likely to respond to the new catalog. In the end, any models of the database the data mining software might create are irrelevant – what matters is the list of potential customers who receive the catalog and the accuracy of the list.

18.2.2 What Exactly is Data Mining?

OK, now we know that data mining, the technology, is not the solution to our problems. But what is the technology? How does it differ from statistics and other time-proven techniques? And what is the end product from the technology? In a field filled with hype, the answers to these questions can often be vague or misleading. In this section we ground some expectations.

The phrase *discover interesting patterns* often comes up during discussions of data mining. A pretty vague statement since "interesting" usually depends on a specific vertical market and "pattern" is irrelevant without some specific of business problem. For most problems, a pattern is some set of measurable characteristics that can be correlated with some other characteristic. For example, a pattern that might be discovered by a data mining application could be something like this: if the age is between 16 and 20 and the zip code is 90210, then one probably drives a car costing greater than $50,000. What this pattern does not say is everyone matching this pattern must drive an expensive car. Usually a pattern is associated with an "accuracy," which specifies the percentage of pattern matches where the correlated characteristic is correct. As far as "interesting" is concerned that would depend on the business problem. If we are trying to market luxury products, this sort of pattern might very well be interesting. But if we are trying to predict medical insurance fraud, this pattern is unlikely to be useful, and therefore uninteresting.

Coverage is also an important concept. In the previous example, the discovered pattern only applies to some fraction of people living in one zip code. If the business is national, a pattern that includes only one zip code is not enough. In that case the database marketing system would need to discover many more patterns. Coverage relates the total number of possible pattern matches to the number of records that do match some pattern for a desired characteristic. If a collection of patterns matches all records with the desired characteristic, the coverage is 100%. The tradeoff is between accuracy and coverage. A pattern that matches everyone in the US would naturally match all people who buy luxury cars. The pattern would have 100% coverage but very low accuracy.

Another word that often shows up in data mining is "model." A model is simply a collection of patterns for some desired characteristic (models usually come in a form more complicated than a simple list of characteristics to match). For example, one common model is known as ARMA (autoregressive moving average). Recently neural network and other models based on biological concepts have come into vogue. There are lots of model types out there, but in the end they are irrelevant to the business problem. A model should never be confused with a solution.

Given that the model and the business solution are two different things, how can a model be turned into a business solution? To start, there are some things that apply to nearly all database marketing applications. For instance, actionable characteristics, those characteristics that the business has some control over, are usually more important than those that are nonactionable. An example of an actionable characteristic is whether or not someone is sent a catalog. A nonactionable characteristic might be the amount of their last order. A business can decide to send or not send a catalog but it cannot control the amount of a customer's last order. This is especially important when targeting new customers. A pattern that says, "If someone is sent a catalog with a 10% off coupon, they will order $100 worth of merchandise" is much more useful than the pattern "if someone ordered $100 before, they will order $100 again." In the first case the catalog retailer can take action to target potential customers while in the second they must simply wait for the order to come in.

18.2.3 Who is Developing the Technology?

Researchers, primarily in the fields of computer science and statistics, have been responsible for the development of most of the data mining technology currently available. From a business standpoint, this has been a problem since (academic) researchers are good at developing and evaluating data mining technologies, but they tend to get caught up in minute details of the technology. They are not interested (nor, should they be) in the fact that the core technology is only a small part of delivering a business solution, and that compromise must be made in order to deliver a usable piece of software. Another group of data mining researchers are who we call *downsized data miners*. These are people, primarily with research backgrounds, who worked on data mining research until cutbacks and company downsizing forced them into product development. When downsized data miners develop software, the end product is usually a complex tool (as opposed to a problem-solving application) or intermediate software product. Lately some downsized data miners have claimed that they will be deploying business solutions; however most software is currently in some form of pre-release (beta, alpha, even prealpha!). These complex data mining tools compete with other high-end analysis tools (e.g., SAS or S-Plus) that require users to have sophisticated skills. Ultimately very few of these researchers will directly impact the development of database marketing as a business solution.

On the other side of the coin the researchers are the developers who are trying to create database marketing software applications for business users. Unlike data mining tools, these applications do not require users to know how to set up statistical experiments or build data models. The developers of database marketing applications start with the business problems and try to determine if some piece of data mining technology might be useful in solving the problem. The technology associated with a data mining software application, just one small part of the overall product, will be built using techniques developed by researchers. Although current software products could be more sophisticated, the future for these software companies is the future of data mining.

18.2.4 Turning Business Problems into Business Solutions

The technology commonly referred to, as data mining, already exists in at least cursory form. Unfortunately, for business users, the data mining community is currently focusing on refining the technology, without attempting to validate it in business applications. From a practical standpoint, who cares if some algorithm is a 5% improvement over the best data mining technique if it only works from a command line interface on some supercomputer? If it is not easily usable, it is irrelevant to most users.

To deliver data mining technology into the hands of business users, several changes from the current state of the technology will be required. These changes can be broken down into three key areas:

- A built-in understanding of business problems
- Ease of use (a.k.a. executive level)
- Integration with relational database products

The first point is the most important. A database marketing software product will not succeed if it does not start with an understanding of real-world business problems. Ultimately the transition between model and business solution will require a thorough understanding of the marketplace to formulate the problem in a way that will affect a business. The ability of a database marketing application to make use of this information will determine if it is truly useful to a business. Therefore, industry-specific value-added solution providers will probably have an important place in the field of database marketing. They should be able to contribute vertical market-specific templates and metadata that will guide the database mining technology toward solutions to the business problems.

Once the business problem has been taken into consideration, the process and results need to be conveyed to the businessperson who needs to make a decision. It cannot be assumed that the person who makes the decision will understand how to work with a neural network model or how to interpret the results from such a model. User-friendly graphical user interfaces (GUIs) are a necessity. These GUIs must integrate smoothly into the business user's overall decision support (DSS) application environment. This environment is usually client/server, with a PC running Windows as the preferred client platform. Technologically related input parameters must be avoided at all costs. A decision tree database mining application should not require the user to specify search width, search depth, amount of training records, etc. The user will not understand what these terms mean, let alone know what to provide as input values. Instead the user should be asked for things related to his or her world. How much time can the process take? How much "effort" should be dedicated to the problem? The application will need to translate between the user-specified parameters and the parameters required by the technology. A feedback process by which the application provides the user information related to their input parameters would be very useful. For example, the system might tell the user that when the "effort" knob is set to 5, the process will take about three hours and will look at 40% of the database. By increasing the setting to 7, the time might increase to five hours but 75% of the database will be analyzed. Such tradeoffs are within the scope of knowledge of business users.

Finally, database marketing applications must be smoothly integrated with standard relational database products. Business users do not want to deal with dumping an RDBMS as a flat file or translating between different data formats. Database marketing applications need to work with ODBC (Open Database Connectivity) and leading relational database interfaces so that they can interact directly with the databases. When an application speaks to a database, it will probably be in SQL, the standard for the relational database industry. These things would be obvious to developers of business software, but not necessarily to those in the research-oriented field of data mining.

18.2.5 A Possible Scenario for the Future of Data Mining

What does the future have in store for data mining? In the end, much of what is called data mining will likely end up as standard tools built into database or data warehouse software products. As a indication for this statement, we would like to use the field of spell checking software as an example. Just look back ten years to the infancy of computer word processing. Many companies made spell-checking software. We would usually buy a spell checker as a separate piece of software for use with whatever word processor we might have. Sometimes the spell checker would not understand a particular word processor's file format. Some spell checkers might have even required to dump the document as an ASCII file before it would check the spelling (on the ASCII file). In that case, we would have had to manually make corrections in the original document. Eventually the spell checkers became more user friendly and understood every possible document format. Functionality also increased. The future of spell checking probably looked pretty rosy.

So, where are the spell-checking companies today? Where is the spell-checking software? If we look at the local computer store we will not find much there. Instead we will find that the new word processor comes with a built-in spell checker. As word processor software increased in sophistication and functionality, it was a natural progression to include spell checking into the standard system.

The future of data mining may very well parallel the history of spell checking. The functionality of database marketing products will increase to integrate with relational database products (no more dumping a RDBMS into a flat file!) and with key DSS – Decision Support Systems application environments, it will stress the business problem rather than the technology, and present the process to the user in a friendly manner. Database marketing will start losing some of the hype and begin to provide real value to users. This will make database marketing an important business in and of itself. The larger RDBMS and data warehouse companies have already expressed an interest in integrating data mining into their database products. In the end, this new market and its business opportunities will drive mainstream database companies to database marketing. Ten years from now there may be only a few independent data mining companies left in existence. The real survivors will likely be the ones with the foresight to develop a strong relationship with the mainstream database industry.

18.3 Data Mining for Marketing Decisions

The Internet and the information available on it provide a new source for competitive market monitoring of agrifood companies. Intelligent information agents that scan the Internet for relevant information and aggregate it into an information service, offer the possibility to automate the process of monitoring and relevant online sources and capture significant content for decision makers. The combination of information agents for sector-specific information search on the Internet, with filtering techniques for enterprise-focused information retrieval from the search results, provides an efficient information system alternative for agrifood sectors with small-and medium-sized enterprises. Ongoing research deals with the formulation of appropriate search directives for the agents and the differentiation of appropriate filter templates for information retrieval.

In today's fast-paced world, the availability of relevant information to serve critical success factors of business management is a critical success factor in itself. The magnitude of information sources on the Internet could, in principle, greatly improve the information situation. However, the size of the Internet and the limits in the search efficiency of standard search engines reduce the success of processes in information search and information retrieval. Information agent technology is an approach, which could improve the situation. The section introduces the concept of a sector-oriented agent-based online information system that focuses on the external information needs of enterprises in individual sectors of the agrifood industry. This approach supports sectors with small- and medium-sized enterprises, which lack the resources to establish enterprise-specific systems. The sector orientation requires system features, which allow enterprises to "personalize" information retrieval from the system as far as possible. The discussion is separated into three parts. After an introduction into the principles of information agents (part 1), the section discusses the design of agent-based information retrieval systems (part 2) and, specifically, systems for sectors in the agrifood industry (part 3).

The business information circle. The focus of any management information system is the information needs of management. They are linked to management's information system are the information needs of management. They are linked to management's critical success factors and define the basic information complexes, which could serve management's information needs through appropriate delivery systems. This is a well-established and tested information circle and could involve information from internal and external information sources.

The Information circle. With the advent of the Internet with its magnitude of information sources, the delivery of business external information receives increased attention. The focus of interest includes the

1. Search for appropriate information in anonymous information sources and
2. Adjustment of information complexes (documents) to individual needs through information personalization.

In principle, personalization is the link between the specific information needs of any individual member of management and the delivery of information for management's critical success factors from anonymous information sources.

The integration of information search and personalization is one of the major challenges in the utilization of Internet information sources for management information needs. An approach to reach this integration is the utilization of intelligent information agents, which use knowledge about management's critical success factors to search information sources on the Internet for appropriate information complexes.

Intelligent Information Agents. Computer programs, which are able to carry out a given task in an open environment using certain human characteristics such as autonomy, social cooperation, intelligence, and learning capacity, may be referred to as "intelligent software agents." This is the definition we build on in this section. However, it is not the only definition in literature. As a fast growing area of research, the characteristics of intelligent software agents, and, in consequence, their definition, reflect the actual level of advancements in research.

"Intelligent information agents" represent a specific category of "intelligent software agents," which access heterogeneous and potentially geographically distributed information sources as, e.g., on the Internet, to proactively acquire, mediate, and maintain information with specific relevance for an interest group of users with predetermined information needs. The search usually builds on lists of prespecified keywords. Intelligent information agents act autonomously through the so-called robots, which

(a) communicate with Web servers without human intervention,
(b) monitor Internet sites for relevant contents or changes in contents, and
(c) may utilize the Internet's hypertext structure to reach beyond specified Internet sites and to search for information along identified hyperlinks.

To improve search efficiency, the space of robots could be confined to predefined information sources or sources that are characterized by certain industries or certain domain-relevant content. Intelligent information agents are being delivered through a number of software solutions. A direct comparison is difficult, as most differences are more gradual than principal. Differences concern primarily

(a) the search directive for the robots,
(b) the information retrieval from the documents identified through the Internet search, which may reach beyond keyword search and include pattern recognition or context search, and
(c) the presentation of results.

18.3.1 Agent-Based Information Retrieval Systems

Principles

The knowledge of users' critical success factors and the availability of robots with their advanced ability to search information sources on the Internet for appropriate documents are not sufficient for the development of an information retrieval system. It requires, in addition, the availability of

(a) a taxonomy model, which maps users' information needs and provides the links between users' Critical success factors (CSF), the information topics, which could serve the CSFs, and the search results,
(b) a search directive for the robots which is usually provided through lists of search keywords and Internet sites,
(c) functions (tools) for the extraction of relevant information items from collected documents, and
(d) functions for the classification of information and its linkage to the taxonomy model for users' retrieval support.

The taxonomy model directed toward information needs of business management employs a multilevel hierarchical structure in which, e.g., critical success factors (first level) are linked to one or several levels of information topics with relevance for the CSF. The taxonomy model provides a structural basis for the efficient organization and access of the magnitude of information documents that might be provided by the robots. In principle, the extraction of information form documents involves the attachment of an appropriate content descriptor to a document (indexing) and the elimination of nonrelevant information retrieval systems. The indexing follows a three-step procedure,

(a) the (automatic) identification of phrases of interest (keywords) in textual data,
(b) the aggregation of relevant information phrases in content descriptors that represent a document's topic and
(c) the attachment of the content descriptor to the documents for retrieval support.

The classification of information links documents to the information topics in the hierarchical taxonomy model. The "quality" of information extraction, and, subsequently, the effectiveness of an information retrieval system, is determined by the users' impression of the fit between the information need in a certain information domain, the formulation of the content descriptors, and the content of the document.

Retrieval Support

The taxonomy model based on users' critical success factors and their linkage to information topics and search results provides the basic infrastructure for information retrieval. However, users could have advanced retrieval knowledge, which might allow them to bypass the taxonomy model's search structure and to identify the required document more directly. As an example, a user might know that certain information should be in one of a few documents retrieved from a certain information source or include, as one of a few documents; a certain keyword. In these cases, retrieval systems based on source or keyword identification would be superior. To capture such situations, it is suggested to complement the "taxonomy retrieval system" by additional retrieval alternatives, which required their own (hierarchical) retrieval infrastructure and document indexing.

Personalization

With appropriately focused robots, agent-based information retrieval systems could be designed to best meet the information needs of any individual user. However, the broad-based realization of such an approach is not only limited by available resources for information search at least in small- and medium-sized enterprises (SMEs), but it disregards the efficiency potential provided by similarities in information needs among enterprises within any specific sector of the economy. It adds to efficiency in information search, if robots focus on a sector's aggregated information needs and leave further individualization (personalization) to a subsequent selection of classified documents. This approach requires the development of

(a) a taxonomy model, which best maps the joint user needs of all or at least the majority of enterprises in the sector and
(b) filters, which allow the separation of, personalized information subsets for individuals.

The personalization could be based on multi-tier templates, which provide filters for a hierarchical individualization scheme, which stepwise narrows down information accessibility. As an example, while the robots may have collected information for aggregated information needs of a certain sector of the economy, the filters may be directed toward the information needs of

(a) subsectors within the sector (first level of filter templates),
(b) enterprises within any of the subsectors (second level of filter templates), and
(c) functional departments within any of the enterprises (third level of filter templates).

Retrieval Systems for Agribusiness Supply Chains

In the development of agent-based retrieval systems for agribusiness production chains, the taxonomy models would need to capture user needs of enterprises on the same and on different stages of the chain (horizontal and vertical information needs). User needs ask for as many as possible individualized taxonomy models, efficiency considerations for as few as possible aggregated ones.

Research suggests that agribusiness production chains could build on a common list of critical success factors and information topics. Differences between stages do not affect the list but priorities within the list. The topics include competitors, products, personnel, market development, research, food law, regional news, and sector-related politics. For agent-based information search, the main differences in information interest concern

(a) the keyword list in the search directive and
(b) the selection of information sources.

As both subjects are not part of the taxonomy model, one could build an information retrieval system for agribusiness supply chains on a common taxonomy model, and consider these differences in the indexing or documents and the implementation of appropriate filter templates. Ongoing research deals with the identification and evaluation of keyword lists, information sources, and filter templates for different groups of users (e.g., in enterprises, functional departments, etc.) on different stages of the supply chains.

In summary, since the availability of information has become a critical success factor for enterprise management, access to competitive intelligence is essential for enterprises. In general, small- and medium-sized enterprises cannot afford to establish an individual information service that monitors their competitive environment.

18.3.2 Applications of Data Mining in Marketing

Data mining has been applied in many areas of marketing such as direct marketing, customer acquisition, customer retention, cross selling, trend analysis, affinity grouping, and customer lifetime value analysis.

Many firms have successfully implemented data mining methodologies to gain sustainable competitive advantage. For instance, Capital One designed a highly successful direct marketing campaign by investing in data mining technology. Through the user of these methodologies, the company was able to accurately predict the response patterns of potential customers and the types of offers best suited for specific types of customers. Numerous other companies have adopted data mining methodologies for customer relationship management. These include Marriott, *Reader's Digest*, Bank of America, Reuters, etc. Wal-Mart is considered among the pioneers in using data mining technology to analyze and predict market trends.

Data mining can provide companies with accurate information on which customers are likely to purchase new products, how much are they willing to spend, and what kinds of products are likely to be purchased together. It enables a company to understand the patterns behind past purchase transactions, thereby helping in the design and launch of new offerings, in an expeditious and cost-effective manner. Indeed, with data mining technology, company data that was previously perceive as "junk" suddenly becomes valuable. The use of data mining methodologies often results in a strategic marketing advantage that cannot be easily duplicated by the competition (unless they adopt similar methodologies themselves).

18.4 Increasing Customer Value by Integrating Data Mining and Campaign Management Software

As a database marketer, we understand that some customers present much greater profit potential than others. But, how will we find those high-potential customers in a database that contains hundreds of data items for each of millions of customers?

To be successful, database marketers must, first, identify market segments containing customers or prospects with high profit potential and, second, build and execute campaigns that favorably impact the behavior of these individuals.

The first task, identifying market segments, requires significant data about prospective customers and their buying behaviors. In theory, the more data, the better. In practice, however, massive data stores often impede marketers, who struggle to sift through the minutiae to find the nuggets of valuable information.

Recently, marketers have added a new class of software to their targeting arsenal; data mining applications automate the process of searching the mountains of data to find patterns that are good predictors of purchasing behaviors. After mining the data, marketers must feed the results into campaign management software that, as the name implies, manages the campaign directed at the defined market segments.

In the past, the link between data mining and campaign management software was mostly manual. In the worst cases, it involved "sneaker net," creating a physical file on tape or disk, which someone then carried to another computer, where they loaded it into the marketing database.

This separation of the data mining and campaign management software introduces considerable inefficiency and opens the door for human errors. Tightly integrating the two disciplines presents an opportunity for companies to gain competitive advantage.

18.4.1 Some Definitions

A Data warehouse is a repository for relevant business data. While traditional databases primarily store current operational data, data warehouses consolidate data from multiple operational and external sources in order to attain an accurate, consolidated view of customers and the business.

Database Marketing uses information in computerized databases to target offerings to customers and prospects.

Data Mining uses technologies such as neural networks, decision trees or standard statistical techniques to search large volumes of data. In doing so, data mining builds models for patterns that accurately predict customer behavior.

Scoring uses a model to predict future behavior. The score assigned to each individual in a database indicates that person's likelihood of exhibiting a particular customer behavior.

Campaign Management uses information in a data warehouse or marketing database to plan, manage, and assess marketing campaigns designed to impact customer behavior.

A Customer segment is a group of prospects or customers who are selected from a database based on characteristics they possess or exhibit.

Scoring on the fly or *dynamic scoring* is the ability to score an already defined customer segment within a campaign management tool. Rather than scoring an entire database, dynamic scoring works with only the required customer subsets, and only when needed.

Attrition, sometimes known as *churn*, occurs when a customer terminates his or her relationship with a service provider. Marketing efforts usually focus on minimizing churn because the cost of bringing a customer back is usually much greater than the cost retaining the customer in the first place.

18.4.2 Data Mining Defined

Data mining, by its simplest definition, automates the detection of relevant patterns in a database. For example, a pattern might indicate that married males with children are twice as likely to drive a particular sports car than married males with no children. If we are a marketing manager for an auto manufacturer, this somewhat surprising pattern might be quite valuable.

However, data mining is not magic. For many years, statisticians have manually "mined" databases looking for statistically significant patterns. Today, data mining uses well-established statistical and machine learning techniques to build models that predict customer behavior. The technology enhances the procedure by automating the mining process, integrating it with commercial data warehouses, and presenting it in a relevant way for business users.

The leading data mining products, such as those from companies like SAS and IBM, are now more than just modeling engines employing powerful algorithms. Instead, they address the broader business and technical issues, such as their integration into today's complex information technology environments.

In the past, the hyperbole surrounding data mining suggested that it would eliminate the need for statistical analysts to build predictive models. However, the value that an analyst provides cannot be automated out of existence. Analysts will still be needed to assess model results and validate the reasonability of the model predictions. Since data mining software lacks the human experience and intuition to recognize the difference between a relevant and an irrelevant correlation, statistical analysts will remain in high demand.

18.4.3 The Purpose of Data Mining

Data mining helps marketing professionals improve their understanding of customer behavior. In turn, this better understanding allows them to target marketing campaigns more accurately and to align campaigns more closely with the needs, wants and attitudes of customers and prospects.

If the necessary information exists in a database, the data mining process can model virtually any customer activity. The key is to find patterns relevant to current business problems.

Typical questions that data mining answers include:

- Which customers are most likely to drop their cell phone service?
- What is the probability that a customer will purchase at least $100 worth of merchandise from a particular mail-order catalog?
- Which prospects are most likely to respond to a particular offer?

Answers to these questions can help retain customers and increase campaign response rates, which, in turn, increase buying, cross selling and return on investment (ROI).

18.4.4 Scoring the Model

Data mining builds models by using inputs from a database to predict customer behavior. This behavior might be attrition at the end of a magazine subscription, cross-product purchasing, willingness to use an ATM card in place of a more expensive teller transaction, and so on.

The prediction provided by a model is usually called a *score*. A score (typically a numerical value) is assigned to each record in the database and indicates the likelihood that the customer whose record has been scored will exhibit a particular behavior.

For example, if a model predicts customer attrition, a high score indicates that a customer is likely to leave, while a low score indicates the opposite. After scoring a set of customers, these numerical values are used to select the most appropriate prospects for a targeted marketing campaign.

18.4.5 The Role of Campaign Management Software

Database marketing software enables companies to deliver to customers and prospects timely, pertinent, and coordinated messages and value propositions (offers or gifts perceived as valuable). Today's campaign management software goes considerably further. It manages and monitors customer communications across multiple touchpoints, such as direct mail, telemarketing, customer service, point-of-sale, e-mail, and the Web.

Campaign management automates and integrates the planning, execution, assessment, and refinement of possibly tens to hundreds of highly segmented campaigns running monthly, weekly, daily, or intermittently. The software can also run campaigns that are triggered in response to customer behavior or milestones – such as the opening of a new account.

Increasing Customer Lifetime Value

Consider, for example, customers of a bank who only use the institution for a checking account. An analysis reveals that after depositing large annual income bonuses, some customers wait for their funds to clear before moving the money quickly into their stock brokerage or mutual fund accounts outside the bank. This represents a loss of business for the bank.

To persuade these customers to keep their money in the bank, marketing managers can use campaign management software to immediately identify large deposits and trigger a response. The system might automatically schedule a direct mail or telemarketing promotion as soon as a customer's balance exceeds a predetermined amount. Based on the size of the deposit, the triggered promotion can then provide an appropriate incentive that encourages customers to invest their money in the bank's other products. Finally, by tracking responses and following rules for attributing customer behavior, the campaign management software can help measure the profitability and ROI of all ongoing campaigns.

Integrating Data Mining and Campaign Management

The closer integrating data mining and campaign management work together, the more business results. Today, campaign management software uses the scores generated by the data mining model to sharpen the focus of targeted customers or prospects, thereby increasing response rates and campaign effectiveness.

Unfortunately, the use of a model within campaign management today is often a manual, time-intensive process. When someone in marketing wants to run a campaign that uses model scores, he or she usually calls someone in the modeling group to get a file containing the database scores. With the file in hand, the marketer must then solicit the help of someone in the information technology group to merge the scores with the marketing database.

This disjointed process is fraught with problems:

- The large numbers of campaigns that run on a daily or weekly basis can be difficult to schedule and can swamp the available resources.
- The process is error prone; it is easy to score the wrong database or the wrong fields in a database.
- Scoring is typically very inefficient. Entire databases are usually scored, not just the segments defined for the campaign. Not only is effort wasted, but the manual process may also be too slow to keep up with campaigns run weekly or daily.

The solution to these problems is the tight integration of data mining and campaign management technologies. Integration is crucial in two areas:

First, the campaign management software must share the definition of the defined campaign segment with the data mining application to avoid modeling the entire database. For example, a marketer may define a campaign segment of high-income males between the ages of 25 and 35 living in the northeast. Through the integration of the two applications, the data mining application can automatically restrict its analysis to database records containing just those characteristics.

Second, selected scores from the resulting predictive model must flow seamlessly into the campaign segment in order to form targets with the highest profit potential.

18.4.6 The Integrated Data Mining and Campaign Management Process

This section examines how to apply the integration of data mining and campaign management to benefit the organization. The first step creates a model using a data mining tool. The second step takes this model and puts it to use in the production environment of an automated database marketing campaign.

Step 1: Creating the Model

An analyst or user with a background in modeling creates a predictive model using the data mining application. This modeling is usually completely separate from campaign creation. The complexity of the model creation typically depends on many factors, including database size, the number of variables known about each customer, the kind of data mining algorithms used, and the modeler's experience.

Interaction with the campaign management software begins when a model of sufficient quality has been found. At this point, the data mining user exports his or her model to a campaign management application, which can be as simple as dragging and dropping the data from one application to the other. This process of exporting a model tells the campaign management software that the model exists and is available for later use.

Step 2: Dynamically Scoring the Data

Dynamic scoring allows us to score an already-defined customer segment within the campaign management tool rather than in the data mining tool. Dynamic scoring both avoids mundane, repetitive manual chores and eliminates the need to score an entire database. Instead, dynamic scoring marks only relevant customer subsets and only when needed.

Scoring only the relevant customer subset and eliminating the manual process shrinks cycle times. Scoring data only when needed assures "fresh," up-to-date results. Once a model is in the campaign management system, a user (usually someone other than the person who created the model) can start to build marketing campaigns using the predictive models. Models are invoked by the campaign management system.

When a marketing campaign invokes a specific predictive model to perform dynamic scoring, the output is usually stored as a temporary score table. When the score table is available in the data warehouse, the data mining engine notifies the campaign management system and the marketing campaign execution continues.

Here is how a dynamically scored customer segment might be defined:

Where

Length_of_service = 9 And Average balance > 150 And In Model (promo9). score > 0.80

In this example:

Length of service = 9 limits the application of the model to those customers in the ninth month of their 12-month contracts, thus targeting customers only at the most vulnerable time. (In reality, there is likely a variety of contract lengths to consider this when formulating the selection criteria.)

Average balance > 150 selects only customers spending, on average, more than $150 each month. The marketer deemed that it would unprofitable to send the offer to less valuable customers.

Promo9 is the name of a logged predictive model that was created with a data mining application. This criterion includes a threshold score, 0.80, which a customer must surpass to be considered "in the model." This third criterion limits the campaign to just those customers in the model, i.e., those customers most likely to require an inducement to prevent them switching to a competitor.

18.4.7 Data Mining and Campaign Management in the Real World

Ideally, marketers who build campaigns should be able to apply any model logged in the campaign management system to a defined target segment. For example, a marketing manager at a cellular telephone company might be interested in high-value customers likely to switch to another carrier. This segment might be defined as customers who are nine months into a twelve-month contract, and whose average monthly balance is more than $150.

The easiest approach to retain these customers is to offer all of them a new high-tech telephone. However, this is expensive and wasteful since many customers would remain loyal without any incentive. Instead, to reduce costs and improve results, the marketer could use a predictive model to select only those valuable customers who would likely defect to a competitor unless they receive the offer.

18.4.8 The Benefits of Integrating Data Mining and Campaign Management

For Marketers

Improved campaign results through the use of model scores that further refine customer and prospect segments.

Records can be scored when campaigns are ready to run, allowing the use of the most recent data. "Fresh" data and the selection of "high" scores within defined market segments improve direct marketing results.

Accelerated marketing cycle times that reduce costs and increase the likelihood of reaching customers and prospects before competitors.

Scoring takes place only for records defined by the customer segment, eliminating the need to score an entire database. This is important to keep pace with continuously running marketing campaigns with tight cycle times.

Accelerated marketing "velocity" also increases the number of opportunities used to refine and improve campaigns. The end of each campaign cycle presents another chance to assess results and improve future campaigns.

Increased accuracy through the elimination of manually induced errors. The campaign management software determines those records to be scored and when.

For Statisticians

Less time spent on mundane tasks of extracting and importing files, leaving more time for creativity – building and interpreting models. Statisticians have greater impact on corporate bottom line.

18.5 Completing a Solution for Market-Basket Analysis – Case Study

Algorithms for finding rules or affinities between items in a database are well known and well documented in the knowledge discovery community. A prototypical application of such affinity algorithms is in "market basket" analysis – the application of affinity rules to analyzing consumer purchases. Such analyses are of particular importance to the consumer package goods industry. The retailers and wholesalers in the industry generated over 300 billion dollars of sales every year in the United States alone. Despite the economic importance of this industry, data mining solutions to the key business problems are yet to be developed. This section discusses some of the problems of the consumer package goods industry, notes a case study of some of the challengers presented to data miners within this industry, and critiques current knowledge discovery research in these areas.

18.5.1 Business Problem

The consumer package goods industry that exists within a complex economics is informational environment. Mass merchandizing or products is in decline; U.S. consumers are increasing recognized as belonging to fifty (or more) distinct segments, each with its own demographic profile, buying power, product preferences, and media access. The items being sold, consumer products are more diverse than ever before; a single category of food may easily contain hundreds of competing products. Within this highly differentiated environment, strong product brand names continue to offer a strong competitive advantage. By themselves temporary price reductions are not sufficient for establishing consumer loyalty to either a store or product. Consumers are knowledgeable, and mobile, enough to seek out the lowest possible prices for a product. Ultimately consumer value is gained by those retailers able to negotiate favorable terms with their suppliers. Retailers gain the requisite detailed knowledge of customers through the creation of consumer loyalty programs and the use of on-line transaction processing systems; this information about the consumer is a crucial component in retailer–supplier negotiations.

Consumer packaged goods is a mature industry in the United States. Profit is no longer merely a matter of opening more stores, and selling to increasing numbers of consumers; the market is becoming saturated, and the available consumer disposable income largely consumed.

Maintenance of an existing customer base is more important than growing entirely new customers; this new phase of retail growth is based upon selling more and a greater variety of products to pre-existing consumers. The most profitable retailers are those that are able to maintain or reduce their operating costs. Economics of scope, not scale, determine profitability. Data warehousing is one of the foremost technological means of increasing operational efficiency. Efficient consumer response systems, based upon data warehouses, are expected to save the industry $30 billion a year. Category management, an organizational strategy for enhancing retailer–wholesaler coordination, is another means of increasing operational efficiency. In the following two brief case studies we examine how data warehouses, category management, and data mining techniques show promise for answering the concerns of two large consumer package goods companies.

18.5.2 Case Studies

A major international food manufacturer, with significant equity and a wide variety of manufactured products, is interested in optimizing its product-advertising budget. Like many package goods retailers, this manufacturer has an extensive and rapidly growing advertising budget. Essential to the endeavor is the cooperation of their independent retail outlets in the creation and design of product promotions. The manufacturer sought to create a suite of software tools for the design of promotions, utilizing the newest data mining technology, and to make these tools available in real time to their category managers and to the managers of their retail outlets. The business case suggested that there would be at least three sources of return in the creation of this tool: Improved coordination with retailers, more effective cross sales across product categories, reduced promotional competition from other manufacturers, and enhanced promotional returns. NCR Corporation proposed and designed a state-of-the-art neural network for forecasting and optimizing planned promotions. The network met, or exceeded, industry standards for promotional forecasts (within 15% of actual sales, 85% of the time). Despite the statistical quality of the results the application was never put into production by the manufacturer, the software design necessary to implement the results was too complex. Part of the application complexity stemmed from the hierarchical data types necessitated by the varied products and markets; another component of the complexity was reconciling the different product world views of the manufacturer and the retailer.

A major regional food retailer, a grocer, sought analyses of its consumer transactions within its produce and salad dressing departments. The retailer anticipated improved design of store layouts, improved promotional design, and an insight into the market role of the various highly differentiated products within the category. The retailer clearly anticipated a causal analysis that would reveal the products, which, when purchased by consumers, would lead to additional add-on sales of other products. NCR produced a market basket analysis, which revealed the distinctive purchasing profiles that are associated with each major brand of interest. The NCR analysis revealed that the best selling brands were not those that resulted in the greatest amount of attendant sales. The NCR analysis supported the existing category management plans by the retailer, and also independently confirmed the results of a demographic panel survey. Despite these successes the market basket analysis, by itself, did not produce any new actionable results for the retailer. In the next section on data mining, key data mining algorithms and outputs are examined for their suitability for answering these, and other, consumer package goods questions.

18.5.3 Data Mining Solutions

Affinity algorithms are well understood and well documented by the data mining community. The quintessential application of affinity algorithms is in the area of market basket analysis. For instance, these algorithms when applied to market basket analysis produce rules such as "Those baskets producing product X are also 75% likely to contain product Y." Additional research has focused on optimizing the speed and efficiency with which these rules are found; however additional applied research is needed to the support decision-making needs of the consumer goods industry (and other relevant business groups).

First, affinity algorithms produce individual, isolated rules; associations between groups of products are not revealed. While the analysis can be repeated across all products in a category, or even a store, the number of rules produced grows exponentially. Not only is the computation complex, but the resulting welter of rules is hard to interpret as well.

Second, the output of affinity algorithms seems to suggest causal relationships between products. Yet the algorithms themselves embody no casual assumptions. The nature of product affinities needs to reconsidered, either a new and causal form of affinities analysis needs to be produced, or a thorough understanding of noncausal applications and use of affinity rules needs to be obtained.

Third, affinity algorithms lack robustness. The algorithms produce a point estimate of affinity; yet retailers need to understand how (and if) these rules apply large groups of transaction. A similar issue is the minimum sample size needed to produce robust results.

Fourth, market basket analyses carry implicit information about consumer preferences. Even when consumer identification is missing from transaction data, the data can still be grouped or segmented using data mining techniques to reveal distinct group of consumer preferences. Affinity algorithms imply that samples are taken from homogenous groups of customers; yet business knowledge suggests that consumers are highly varied in taste and expenditure.

Fifth, the market basket analyses, for some set of business questions, may require the rigor of a properly designed statistical experiment. Reasoning from standard to promotional pricing, as well as reasoning from standard display conditions to promotional display conditions, is unwarranted. Yet much of the potential of market basket analysis stems from the capacity of retailers to manipulate product pricing, display, or even attributes to meet consumer need.

Sixth, and finally, standard forecasting tools produce estimates of sales single goods across times. (This is not conventionally the domain of market basket or affinities analysis.) However retailers and manufacturers need to have forecasts for whole groups of products. Producing individual product forecasts, and then aggregating, will not produce optimum forecasts since sales of one product contains information about the potential sales of other products, indeed, the forecasts may not even aggregate correctly. Techniques such as "state space analysis," which combine forecasting with multivariate analysis may prove useful.

18.5.4 Recommendations

The consumer package goods industry is an important, and expansive, industrial segment of the economy. This industry is dependent upon information for its continued economic growth. It is therefore making great progress in collecting large databases of relevant data about its industry. The corresponding questions the industry has about its data are both interesting and economically fruitful. This section considered two case studies of applying standard data mining techniques to industrial questions in the area of consumer package goods. The examples discussed a wholesaler and a retailer who sought better management of product categories and a resulting improved economy of scope. Commercial success of data mining will in part be dependent upon the capacity of algorithms to model complex, hierarchical arrangements of goods and products.

18.6 Data Mining in Finance

Data mining has received much attention as companies and organizations started to ask how they can better utilize the huge data stores they built up over the past two decades. While some interesting progress has been achieved over the past few years, especially when it comes to techniques and scalable algorithms, very few organizations have managed to benefit from the technology. This paradoxical situation of having too much data, yet not be able to utilize it or mine it, arose because of technical and business challenges.

In many cases the desired target variable does not necessarily exist in the database. If the database includes information about customer purchases, a business user might only be interested in customers whose purchases were more than one hundred dollars. Obviously, it would be straightforward to add a new column to the database that contained this information. But this would involve database administrator and IT personnel, complicating a process that is probably complicated already. In addition, the database could become messy as more and more possible targets are added during an exploratory data analysis phase. The solution is to allow the user to interactively create a new target variable. Combing this with an application wizard (it can significantly improve the user's experience, besides simplifying the process, and can help prevent human error by keeping the user on track), it would be relatively simple to allow the user to create computed targets on the fly.

Mining financial data presents special challenges. For one, the rewards for finding successful patterns are potentially enormous, but so are the difficulties and sources of confusions. The efficient market theory states that it is practically impossible to predict long-term financial markets. However, there is good evidence that short-term trends do exist and programs can be written to find them. The data miner's challenge is to find the trends quickly while they are valid, as well as to recognize the time when the trends are number longer effective.

Additional challenges of financial mining are to take into account the abundance of domain knowledge that describes the intricately inter-related world of global financial markets and to deal effectively with time series and calendar effects. The relational data mining (RDM) is a learning method able to learn more expressive rules than other symbolic approaches. RDM is thus better suited for financial mining, because it is able to make better use of application underlying domain knowledge. Relational data mining also has a better ability to explain the discovered rules – ability critical for avoiding spurious patterns, which inevitably arise when the number of variables examined is very large. The earlier algorithms for relational data mining, also known as ILP – inductive logic programming, suffer from a well-known inefficiency. The researchers introduced a new approach, which combines relational data mining with the analysis of statistical significance of discovered rules. This reduces the search space and speeds up the algorithms. They also introduced a set of interactive tools for "mining" the knowledge from the experts. This helps to further reduce the search space.

Data mining does not operate in a vacuum. The results of the data mining process will drive efforts in areas such as marketing, risk management, and credit scoring. Each of these areas is influenced by financial considerations that need to be incorporated in the data mining modeling process. A business user is concerned with maximizing profit, not minimizing RMS error. The information necessary to make these financial decisions (costs, expected revenue, etc.) is often available and should be provided as an input to data mining application.

18.7 Data Mining for Financial Data Analysis

Most banks and financial institutions offer a wide variety of banking services (such as checking, savings, and business and individual customer transactions), credit (such as business, mortgage, and automobile loans), and investment services (such as mutual funds). Some also offer insurance services and stock investment services.

Financial data collected in the banking and financial industries is often relatively complete, reliable, and of high quality, which facilitates systematic data analysis and data mining. Here we present a few typical cases.

Design and construction of data warehouse for multidimensional data analysis and data mining: Like many other applications, data warehouses need to be constructed for banking and financial data. Multidimensional data analysis methods should be used to analyze the general properties of such data. For example, one may like to view the debt and revenue changes by month, region, sector, and other factors, along with maximum, minimum, total, average, trend, and other statistical information. Data warehouses, data cubes, multifeature and discovery-driven data cubes, characteristic and comparative analysis, and outlier analysis all play important roles in financial data analysis and data mining.

Loan payment prediction and customer credit policy analysis: Loan payment prediction and customer credit analysis are critical to the business of a bank. Many factors can strongly or weakly influence loan payment performance and customer credit rating. Data mining methods such as feature selection and attribute relevance ranking may help identify important factors and eliminate irrelevant once. For example factors related to the risk of loan payments include loan-to-value ratio, term of the loan, debt ratio, payment-to-income ratio, customer income level, education level, residence region, credit history, and so on. Analysis of the customer payment history may find that, say, payment-to-income ratio is a dominant factor, while education level and debt ratio are not. The bank may then decide to adjust its loan granting policy so as to grant loans to those whose application was previously denied but whose profile shows relatively low risks according to the critical factor analysis.

Classification and clustering of customers for targeted marketing: Classification and clustering methods can be used for customer group identification and targeted marketing. For example, customers with similar behaviors regarding banking and loan payments may be grouped together by multidimensional clustering techniques. Effective clustering and collaborative filtering methods (i.e., the use various techniques to filter out information, such as nearest neighbor classification, decision trees, and so on) can help identify customer groups, associate a new customer with an appropriate customer group, and facilitate targeted marketing.

Detection of money laundering and other financial crimes: To detect money laundering and other financial crimes, it is important to integrate information from multiple databases (like bank transaction databases and federal or state crime history databases), as long as they are potentially related to the study. Multiple data analysis tools can then be used to detect unusual patterns, such as large amounts of cash flow at certain periods, by certain groups of people, and so on. Useful tools include data visualization tools (to display transactions activities using graphs by time and by groups of people), linkage analysis tools (to identify links among different people and activities), classification tools (to filter unrelated attributes and rank the highly related ones), clustering tools (to group different cases), outlier analysis tools (to detect unusual amounts of fund transfers or other activities), and sequential pattern analysis tools (to characterize unusual access sequences). These tools may identify important relationships and patterns of activities and help investigators focus on suspicious cases for further detailed examination.

18.8 Summary

Database marketing software applications will have a tremendous impact on how business is done in the future. Although the core data mining technology is here today, developers need to take what already exists and turn it into something that business users can work with. The successful database marketing applications will combine data mining technology with a thorough understanding of business problems and present the results in a way that the user can understand. At that point the knowledge contained in a database will be understood by people who can turn what is known into, what can be done.

The section also suggests the implementation of an Internet-based online information retrieval system, which offers agrifood industry sector-specific information, which is directly, linked to management's critical success factors and allows personalization through the implementation of appropriate filtering techniques. The information retrieval system is based on information agents, which search for appropriate documents on distributed information sources. The utilization of information agents has left the early experimental phase and is in the early stages of professional use. Ongoing research on the implementation of agent-based intelligent information retrieval systems for the agrifood sector does focus on the selection of appropriate search directives for the robots, the formulation of taxonomy models that best map users' information needs, and the appropriate retrieval, individualization, and presentation of information from the search results.

In this section the method that integrates data mining and campaign management software to increase the customer value was discussed. The section also considered two case studies of applying standard data mining techniques to industrial questions in the area of consumer package goods. The examples discussed a wholesaler and a retailer who sought better management of product categories and a resulting improved economy of scope. Commercial success of data mining will in part be dependent upon the capacity of algorithms to model complex, hierarchical arrangements of goods and products. Also how data mining is applied in finance and finance data analysis has been described.

18.9 Review Questions

1. Compare data mining and database marketing.
2. Explain in detail data mining for marketing decisions.
3. What is agent-based information retrieval systems?
4. What are the applications of data mining in marketing and explain in detail.
5. Define scoring, campaign management, and customer segment attrition.
6. What is the role of campaign management software?
7. What are the steps involved in implementing integrated DM and campaign management process?
8. What are the benefits of integrating data mining and campaign management?
9. Explain in detail how data mining is used in finance.
10. Describe the design and construction of data warehouse for multidimensional data analysis and DM.
11. Explain data mining for financial data analysis.