CDi

The Center for Data Insight

at

Northern Arizona University

A Perspective on Data Mining

July 1998

Authors:

Dr. Kenneth Collier
Dr. Bernard Carey
Ms. Ellen Grusy
Mr. Curt Marjaniemi
Mr. Donald Sautter

Table of Contents

Executive Summary 1

1.0 What is Data Mining? 2

1.1 Data Mining and Knowledge Discovery in Databases (KDD) 2

1.2 Relevance to Corporate Enterprises 5

2.0 The Emergence of Data Mining: Hype or substance? 6

2.1 The Emergence of an Industry 6

2.2 Recovery from the Artificial Intelligence Hype of the 80's 7

3.0 The Proof 0f Concept Period 9

4.0 What the Critics Say 11

4.1 The View of the Business Analyst 11

4.2 Views of the Press 12

4.3 Views of Those Who Have Used Data Mining 14

5.0 Early Adopters: An Industry Overview 17

6.0 Data Mining Algorithms 19

7.0 Data Mining Companies: Pieces of a Larger Puzzle 27

7.1 History of Data Mining - Software Vendors 27

7.2 History of Data Mining - Service Providers 31

8.0 A Move Toward Vertical Applications 33

8.1 The Tendency of Companies to Verticalize applications 33

8.2 Who is Specializing in What? 33

9.0 The Analysts' view of the future market 34

10.0 Business Models for an Integrated Solutions Company or Consortium..... 36

10.1 A Collection of Software Offerings..... 36

10.2 Full Service Data Mining..... 36

10.3 Vertical Solutions Providers..... 37

Bibliography..... 38

Executive Summary

This document presents an overview of the subject of data mining. Data mining is a stage in the overall process of Knowledge Discovery in Large Databases (KDD). Data mining is a semi-automated process for finding undiscovered patterns and/or relationships in large databases. Data mining finds its roots in the Machine Learning community whereby academicians invented and developed artificial intelligence algorithms as a basis for machine learning.

Such algorithms were the basis for the interest in artificial intelligence in the 1980s, a disappointment to the corporate world due to an overselling of an immature technology at the time. The good news is that such technologies kept maturing and became the basis for an industry with solid technological and business foundations in the 1990s. In parallel with the development of data mining products, very powerful computers, networks, and database systems also came into existence that permit storing, formatting, accessing, and analyzing large operational data stores to improve decision support operations in businesses. Such systems, in combination with data mining tools, now permit the development of new knowledge management processes to apply to meeting both corporate and scientific objectives.

This document first describes data mining and the overall knowledge discovery process. It presents opinions of industry analysts concerning the benefits of data mining. The document then gives examples from corporations who have used data mining technologies to meet a variety of business objectives. For those interested in the actual algorithms embedded in the data mining tools, a section is provided that summarizes the main data mining algorithms now being employed.

KDD is a new industry. There are a number of companies that are providing data mining tools and other products to support the KDD process. A section of this report presents a summary of companies providing products in this industry as of the date of the writing of this report.

Last, some thoughts are presented as to the future strategies for the use of such technologies.

1.0 What is Data Mining?

1.1 Data Mining and Knowledge Discovery in Databases (KDD)

"Data mining is the exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules." [4] While there are many other accepted definitions of data mining, this one captures the notion that data miners are searching for meaningful patterns in large quantities of data. The implied goal of such an effort is the use of these meaningful patterns to improve business practices including marketing, sales, and customer management. Historically the finding of useful patterns in data has been referred to as knowledge extraction, information discovery, information harvesting, data archeology, and data pattern processing in addition to data mining. In recent years the field has settled on data mining to describe these activities. [9] Statisticians have commonly used the term data mining to refer to the patterns in data that are discovered through multivariate regression analyses and other statistical techniques.

As the evolution of data mining has matured, it is widely accepted to be a single phase in a larger life cycle known as Knowledge Discovery in Databases or KDD for short. The term KDD was coined in 1989 to refer to the broad process of finding knowledge in data stores. [10] The field of KDD is particularly focused on the activities leading up to the actual data analysis and including the evaluation and deployment of results. KDD nominally encompasses the following activities (see ***Figure 1***):

1) Data Selection – The goal of this phase is the extraction from a larger data store of only the data that is relevant to the data mining analysis. This data extraction helps to streamline and speed up the process.

2) Data Preprocessing – This phase of KDD is concerned with data cleansing and preparation tasks that are necessary to ensure correct results. Eliminating missing values in the data, ensuring that coded values have a uniform meaning and ensuring that no spurious data values exist are typical actions that occur during this phase.

3) Data Transformation – This phase of the lifecycle is aimed at converting the data into a two-dimensional table and eliminating unwanted or highly correlated fields so the results are valid.

4) Data Mining – The goal of the data mining phase is to analyze the data by an appropriate set of algorithms in order to discover meaningful patterns and rules and produce predictive models. This is the core element of the KDD cycle.

5) Interpretation and Evaluation – While data mining algorithms have the potential to produce an unlimited number of patterns hidden in the data, many of these may not be meaningful or useful. This final phase is aimed at selecting those models that are valid and useful for making future business decisions.

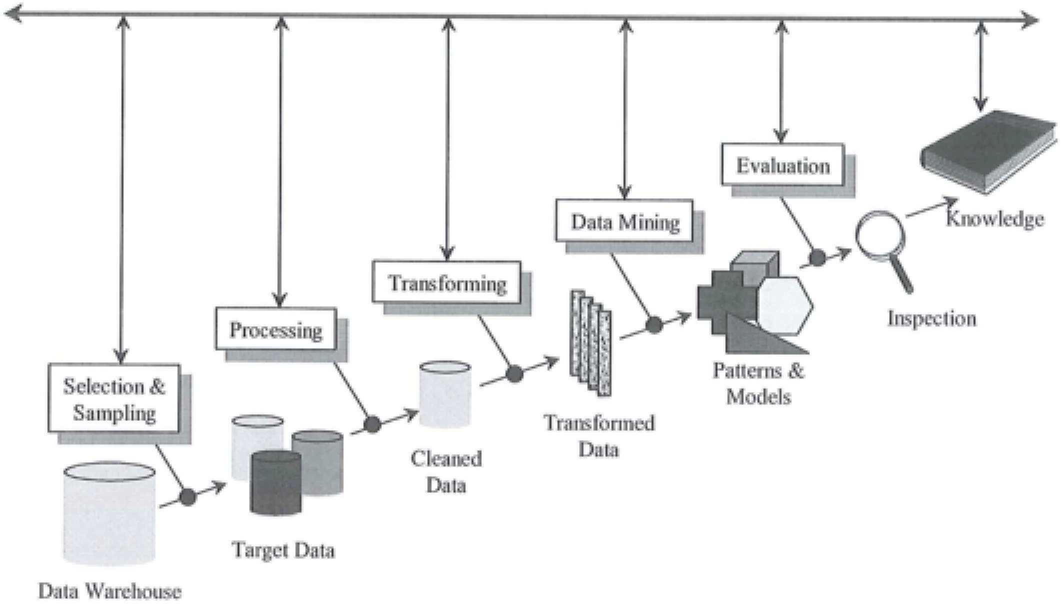


Figure 1 The Traditional KDD Paradigm

The result of this process is newly acquired knowledge formerly hidden in the data. This new knowledge may then be used to assist in future decision making.

The Center for Data Insight has extended this process model in the following ways:

Framing the Question(s) - One of the common misconceptions about data mining is that one can blindly set the algorithms loose on the data to find all interesting patterns that may be present. Data mining is not a magic panacea for curing all business ills. Rather it is a decision support tool that, when used in conjunction with business understanding, can provide a valuable means of gaining new business insights. Therefore, the first step in the KDD cycle must be to determine between one and three questions or goals to help direct the KDD focus.

Actionable Results - The current KDD cycle ends with the evaluation and validation of analytical results. The difficulty with this conclusion is it does not provide a prescription for what to do with these results in business decision support. For example, a market basket analysis which tells you that people who purchase eggs tend to also purchase bread does not tell you what to do with this newly gained information. Should a supermarket put bread and eggs side by side or should it place them at opposite ends of the store to send the buyer past other goods on his/her trip from the eggs to the bread? Therefore, we feel that the important final phase in the KDD cycle is to identify a set of actionable results based on the validated models.

Iteration - While the current KDD cycle supports the return to previous phases to improve the data mining results, experience has shown that iteration is much more an integral element in the cycle than implied by the traditional model. The CDI has adapted for the KDD process an concept widely recognized in the software engineering community. That process is shown in Figure 2.

Note that the model illustrated in Figure 2 also incorporates the additional phases of "framing the questions" and "producing actionable results". Under this model a prototype study is conducted to determine if the data will support the objectives. Successive iterations serve to refine and adjust the data and the algorithms to enhance the outcome.

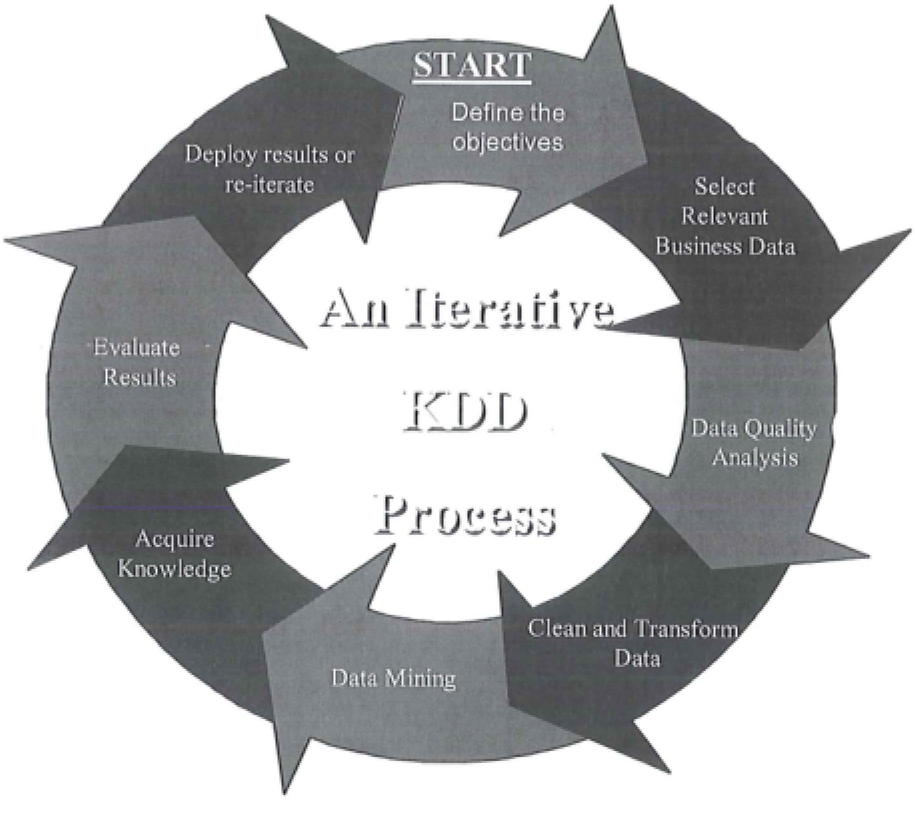


Figure 2 A Refined KDD Paradigm

1.2 Relevance to Corporate Enterprises

Small businesses rely on their knowledge of the customer to inspire loyalty and confidence. For example, a businessman might continue to purchase business suits from the same tailor since that tailor has grown to understand his particular body shape, features, fabric likes and dislikes. Similarly, one returns to the hairstylist who has developed a personalized understanding of the client's preferences. While the small tailor or hairstylist do not necessarily have more knowledge about clothing or hair than larger clothing or beauty chains, it is their personal knowledge about their customers that gives them a business edge.

Companies of all sizes can benefit by emulating what small, service-oriented businesses have always done well - creating one-to-one relationships with their customers. [4] In every industry, companies are trying to improve their business by individualizing its customer interactions. Small businesses build these customer relationships by NOTICING customer needs, REMEMBERING their preferences, and LEARNING from past interactions. Large enterprises have great difficulty accomplishing something similar since customers may never interact personally with company employees. Or if there is such interaction, it may be with different sales people or an anonymous call-center. While nothing can replace the ability of the small business owner to recognize customer preferences and patterns, data mining provides a mechanism that simulates this ability for larger enterprises. [4]

Most large companies have collected customer and transactional data for many years into operational data stores for billing and auditing purposes. Some companies have discovered that this transactional data is one of their most valuable assets. Consumers create a constant stream of data during daily business transactions. Consider placing a single telephone order to a catalog retailer. A transaction record is created by the local telephone company showing the time of the call, the number dialed, and the long distance company used to connect the call. The long distance company creates its own similar transaction record as well. Once connected, the catalog company generates a transaction record including all customer information, the catalog number, products purchased, etc. When the customer provides the requisite credit card information to pay for the order, the credit card company logs the transaction in its own operational data store. Finally, when the order is shipped the shipping company records the source and destination points of the package as well as its weight and other salient information. The amount of data collected during any given business day is staggering.

Although this data is a rich source of potential knowledge about customers and business practices, many companies make the mistake of discarding their transactional data after a certain period of time. During the 1990s many companies reached the conclusion that their data is a valuable asset. These companies moved quickly to build data warehouses and data marts. Furthermore, companies such as Wal-Mart recognized the benefit of applying data mining to these rich stores of historical data.

2.0 The Emergence of Data Mining: Hype or substance?

2.1 The Emergence of an Industry

Is there a real data mining industry? Or is data mining a lot of hype as unfortunately occurred with artificial intelligence in the 1980s? While the AI "boom" of the 1980s turned out to be a "bust", there were many beneficial outcomes of those efforts. Among these were the advancements in machine learning technologies and algorithms. Through these efforts, along with advances in database technologies, data storage capabilities, and parallel computing, a data mining industry has emerged. First as academic research, then as a collection of algorithms on the public domain, finally as a set of robust commercial products, the industry has matured in the past fifteen years to provide real advantages to those who employ it. This maturation has taken on a number of forms.

- The past twenty years has seen our economy make a transition into the information age. Computers, data and information have become the basis for decision making in many industries.
- Companies have and are collecting very large amounts of information about their customers, their products, their markets, their employees, their manufacturing processes, their distribution processes, and their marketing processes. This historical information can be "mined" to develop predictive models to guide future decision-making.
- The field of machine learning has continued to evolve in the academic communities. New concepts, new algorithms and new computer structures and systems have been invented and applied to real world scientific and business problems. These ideas are being transitioned to industry in the form of new products. They have also become the basis for start-up companies developing entire new businesses.
- Through experience, an understanding has developed that data mining is a step in a larger knowledge discovery process. A systematic methodology has evolved to take raw data and transform it into information and to take information and transform it into knowledge to help solve real business problems. It is now understood that the larger process requires the careful use of different computer technologies in different stages. Raw operational data can be transformed into predictive models that support meeting major business objectives. Data mining plays a critical role in this overall process.
- As the field of machine learning has developed, there has been a constant evolution of fragile AI technologies from the 1970s and 1980s, such as expert systems and neural computing, into mature products. These products correctly used can be successfully deployed into the business environment.
- There has been an evolution and development of very powerful and efficient data base systems and associated access and query tools.
- The development of fast, powerful, networked computer systems to provide massive amounts of storage and the rapid execution of data mining software has occurred in the 1990s.
- There is an intense intellectual ferment occurring now in the knowledge discovery and data mining field. New divisions of existing corporations such as Angoss and ISL, or new corporations such as Unica, Cognos and Evoke have been created. These companies are developing, adapting and extending AI and other algorithms to provide data mining products targeted at specific business applications. They are

also developing the tools necessary to efficiently perform each step of the knowledge discovery process. Examples of such tools for intermediate steps are data cleansing, data compression, data access, and data visualization.

- There is a significant commitment by high technology companies such as Thinking Machines and Silicon Graphics Incorporated to provide data mining products as extensions of their core expertise in parallel computing and advanced computer graphics systems.
- There is a significant commitment by large, stable corporations such as SAS, ORACLE and IBM to provide data warehousing and data mining products as extensions of their existing, core expertise in statistical data processing, on-line analytical processing, large data bases, data warehousing , powerful workstations and networked servers.
- There is a significant commitment by large, stable corporations such as KPMG Peat Marwick LLP to provide complete knowledge discovery systems (building upon their extensive core expertise in business practices and systems). Such a company can understand the business objectives for a client and develop an integrated system solution. That solution will cover all aspects of the knowledge discovery process. The knowledge discovery results are deployed into businesses such as financial services, transportation, insurance, medical and many others.
- Enough examples exist of the successful use of data mining to justify investing in the use of these technologies. It has been shown that data mining products have been put in an enterprise form and that very significant benefits accrue from their use.

In summary, data mining and the associated computer technologies have now reached a mature point. Very important competitive advantages are being gained by a variety of companies that have used data mining. However, it must be emphasized that data mining must be used in a thoughtful and careful manner.

2.2 Recovery from the Artificial Intelligence Hype of the 80's

There has been, and continues to be, a hype associated with data mining technologies that rightfully makes the prospective corporate user wary. The first cycle of hype for AI occurred in the late 1970s through the mid-1980s.

The unfulfilled hype for artificial intelligence based products originally promoted in the 1980s rightfully has a negative resonance to this day. As an example, in response to the push to more automated manufacturing plants artificial intelligence approaches were hyped as a panacea for corporations seeking profits through productivity. The hype said, buy a computer vision system and the quality of your product will be doubled. The hype said, buy twenty robots and your defect rate will be halved and you can also reduce your labor force and costs. The hype said, install such and such an intelligent computer material handling system and you can turn out the lights at night and find your finished product waiting for you the next day. Neural networks were presented as an analog for the human brain even with little understanding of basic, human cognitive functions. It was thought that parallel computing should be employed on an assumption that if the software ran faster it must be smarter. Enormous investments were made. Great expectations were raised. Enormous investments were lost. Great disappointments occurred.

However, even with these disappointments there were significant successes. The lesson learned in the 1980s was that success occurred when the expectations were reasonable and when the technologies were properly used. An example of such a success is the expert system. Given a set of conditions the human expert would make an informed decision. Careful methodologies were developed to capture the knowledge of experts in certain problem domains and then encode that knowledge in rule-based computer software. The expert system was able to make similar or identical decisions to the human expert when faced with those same conditions. Examples of this were successful computer systems developed for doing medical diagnostics and repair of complex equipment such as locomotives.

The data mining industry currently runs the risk of creating a second hype cycle. A data mining tools may be presented as a "magic box" in which you pour raw data and a business solution flows out. It does not work that way. Data mining can yield many misleading or unimportant findings. The user must understand which results have real meaning. Getting useful results from mining the data requires a careful understanding of the data. Data mining also requires a careful interpretation of the relationships or patterns uncovered in the data. Data mining therefore requires an extensive interaction between the person who is the business expert and the person who understands the data mining tools.

The business expert must be involved. This is the person who understands the information contained in the data and can evaluate whether the output of the analytical or mining stages truly makes sense in the specific business domain. The proper use of predictive models must be carefully integrated into the actual business processes so the models can be properly evaluated and updated. The various data mining tools have certain strengths and certain weaknesses. The tool and its use must be properly matched to the expertise of the user and that person's objectives in using the tool.

As part of the hype there is also a technical jargon used with data mining. Like all high tech products, such jargon makes it difficult for a new user to really understand how the data mining product really works, what it really does, what the practical results really are, and which product is best used in which way to successfully meet a given business objective. The confusion resulting from such jargon can result in a potential client not using data mining at all. Even worse, the client may make a very expensive investment in following false paths or learning various tools just to understand how to best employ data mining in a given business.

The Center for Data Insight at Northern Arizona University was created to provide a neutral environment where business clients can come and cut through the hype. They learn from staff already experienced in the knowledge discovery process and the leading data mining tools. In a low-risk way the new client considering the use of data mining can quickly get up to speed on data mining and gain insights into which tools are best to meet their business objectives.

The first step in such a process is a proof-of-concept period to develop hard evidence that data mining can be cost-effectively employed to meet given business objectives.

3.0 The Proof Of Concept Period

Data mining, knowledge discovery and computer systems technologies have matured in the past twenty years. The disappointments in AI technologies from the 1980s are now behind us. A more careful use of mature technologies is being employed. Real examples exist of the benefits of using data mining. Organizations such as KPMG and the Center for Data Insight exist to assist businesses in cutting through the present hype and learn and apply these new technologies in a cost-effective manner. A mature industry now exists to give businesses confidence that investments in new products will be supported by a stable industry.

All of these factors provide confidence that a business should invest in data mining to achieve a competitive advantage. However, the first step should be to conduct a proof-of-concept study. The purpose of this step is to ensure that the most important business objectives are being met and to ensure the investment in data mining is done in the most cost-effective manner.

The proof-of-concept period is used to answer the following questions.

- What is data mining?
- What do the data mining tools really do?
- How should my raw operational data be structured to be compatible with data mining?
- Which data-mining tool, or suite of tools, is best suited to meet my business objectives?
- Is there hard evidence that can be generated by mining my data that shows that my company should invest in data mining and deploy it in my business?

The proof-of-concept process is as follows.

- Define the business objectives. Start with at most three objectives in order to focus the study.
- Identify the corporate data that contains information related to those business objectives.
- Create a sample data set that contains all relevant information.
- Identify a domain expert(s) to work with a group experienced in knowledge discovery systems.
- Install the data in a facility that has the computational power to handle the size of the data being examined and which has a suite of knowledge discovery tools suitable to meet the business objectives.
- The domain expert(s) works with the data mining expert(s) to determine which data mining tool(s) are best suited to meet the business objectives.
- Extract relationships and patterns from the business data set.
- The domain expert(s) works with the data mining expert(s) to determine which patterns and relationships are really relevant to the business objectives. Experience in the CDI on a number of data mining projects has shown that surprising results may occur at this stage. Underlying assumptions about how a business works, how the market works, or how the customer behaves may change.

- Develop models to predict how data mining results can assist in meeting business objectives.
- The company then decides what level of investment to make in data mining consistent with their business plan.

At this point, a company will have significant evidence of how data mining can be employed to achieve a competitive advantage, training in data mining, and the skeleton of a development plan for using data mining in a cost effective manner.

4.0 What the Critics Say

It is useful to gain a sense of the public perception of this emerging technology. To assist in this we provide perspectives from three relevant communities: business analysts, the press, and data mining users and practitioners.

4.1 The View of the Business Analyst

The following is a collection of brief excerpts from some of the most influential members of the business and technology communities:

From the Aberdeen Group Enterprise Data Mining Buying Guide: 1997 Edition

Data mining offers technology to systematize the excavation of the value from databases. However, because of several significant differences in the types of tools, levels of expertise, and financial health among the suppliers, enterprises and lines of business will need to carefully evaluate prospective data mining suppliers and partners.

...To unleash the power of what Aberdeen calls scalable, commercial-grade data mining tools, enterprises must select, clean, and transform their data, perhaps integrate data brought in from outside sources, and essentially establish a recognizable environment for data mining algorithms.

Data mining is more about data preparation than the sometimes-wonderful possibilities of any given algorithm or set of algorithms. Approximately 75% of the work in data mining is in getting the data to the point where data mining tools can actually start running.

...By misapplying a scalable, commercial-grade data mining tool, an enterprise can squander potential and possibly millions of dollars.

From the Data Mining Systems web page at www.datagems.com:

The Gartner group predicts that data mining will be one of the five hottest technologies in the closing years of the century.

...Data mining is finally coming to the forefront of information management initiatives. Our research indicates that the majority of Global 2000 organizations will find data mining technologies to be critical to their business success by the year 2000. Historically, data mining referred to the process of applying artificial intelligence (often through an automated process) to look for patterns in large data sets. We believe the most successful information architectures incorporate both traditional data mining tools and end-user DSS tools.

From Herb Edelstein

This past year has been one of real, albeit not spectacular, growth in the data mining industry. A survey Two Crows Corp. recently conducted provided strong evidence of corporate satisfaction…

…Of those organizations far enough along to have formed an opinion, ALL plan to continue or expand their present use of data mining. The survey also revealed that data mining is very much in its early growth stages. While many companies are interested in the technology, only a few have active projects up and running. One of the most interesting findings was that while the process of implementing data mining is proving more difficult than expected, its results often exceed expectations.

…IT teams have bought into the myths of data mining's almost magical ease of use. The myth is that if you simply put a data mining tool on your terabyte-sized database, truth, beauty, and valuable information will emerge. The reality is that successful data mining requires an understanding of the business, the data, the tools, and the process of data analysis.

From David Stodder

…Many IT members grew up with the rise of AI - and many also remember what happened when the hype balloon popped and AI tumbled to the ground. Data miners know they can't afford another debacle. "Cautious optimism" defines the mood of the data mining industry.

[redacted] *Magazine, April 1997: "Scientific Data Miners Make Use of All Tools Available"*

Extraction and analysis of patterns, trends, and clusters in large databases, however, can provide invaluable business information and forecasts. The general reasoning in the marketplace is that if you don't do the data mining, your competitor surely will.

4.2 Views of the Press

Recently there has been much written in the general press regarding data mining. The following excerpts from newspapers around the nation provide some insight into the perceptions of the press.

From the San Jose Mercury News, October 6, 1997: "Dominating With Data - Data Mining Emerges as The Future Of Business Analysis."

Many…businesses built on data -- from the National Hockey League and its mostly paper-based statistics operation to retailers with their myriad daily transactions -- are beginning to realize that there are big dollars hidden in their bulging databanks. Dozens of small and large data mining companies are helping them go digging. While the industry is still young, practitioners believe that today's data mining trials will become tomorrow's essential tools.

…Now that data storage is relatively cheap and computers can digest hoards of data with scarcely a burp, business is ready to pay attention.

…A Dataquest analyst says, "It's a fragmented market without an 800-pound gorilla at this point," she said. As happened with small data warehousing companies in the early 1990s, small data mining companies are likely to consolidate and be bought out over the next few years by larger database companies. "IBM has figured it out. Oracle hasn't figured it out yet, but they will soon enough."

From the Minneapolis-St. Paul Star Tribune, August 17, 1997

More than ever, businesses want to know as much as they can about their customers and potential customers. Using cutting-edge technology, firms such as Damark International Inc. and U.S. Bancorp are learning more about their customers through ... data mining.

… Modern data mining … is more precise and less expensive to use than the mathematical computer models that preceded it, and as a result is more important to business.

… While the results are not dramatic, Damark's Voda says even a small increase in orders from the 100 million catalogs Damark mails each year can make data mining worthwhile. He said cost reductions and increased sales attributed to data mining have allowed Damark to recover last year's data mining investment. "We're not looking for the Holy Grail. It's a game of inches, and it does not take a huge improvement in orders to drop a lot of money to the bottom line," Voda says.

From The Washington Post, June 12, 1998: "Va. Tech Firm's Stock Surges 76% in Debut."

Another Washington area technology company, MicroStrategy Inc. of Vienna, made a splash yesterday in its stock market debut, rising 76 percent in its first day of trading to close at $21.12 a share.

Analysts said investors liked MicroStrategy because they believe it is pioneering a market that can only grow. The much-talked-about company, founded in 1989, specializes in "data mining" software that allows companies to synthesize scattered data into usable facts. With MicroStrategy software a clothing chain could, for instance, determine what color blouses sell best when it is cloudy in Washington.

From the Boulder Camera, February 27, 1998: IBM TAKING ON 'DATA MINING'

International Business Machines, the world's largest computer company, launched a company-wide effort Thursday to capture a bigger share of the booming market for "business intelligence" tools.

This market, which includes computers and software tools designed to enhance business decision-making, could grow to $70 billion by the turn of the century, according to analysts.

From the Arizona Republic, March 22, 1998: "NO MORE SECRETS IN 'DATA MINING,' YOU'RE ORE"

Acxiom is a leader among hundreds of companies around the country that now maintain vast electronic reservoirs. These companies include such retailers as Sears, Roebuck and Co., gift shop chains like Hallmark Cards Inc. and insurance companies like Allstate.

"Technology has been the enabler," said [Donald] Hinman [an Axciom executive], who likens the advances to the invention of the printing press. "Today it's almost unbounded, our ability to gather, sort and make sense of the vast quantities of information."

The number of data warehouses, large and small, using faster computers, the Internet and other networks now exceeds 1,000, a tenfold increase in five years.

"They have gone on an information collecting binge," said Charles Morgan Jr., Acxiom's chief executive, describing the data mining explosion.

4.3 Views of Those Who Have Used Data Mining

In addition to reports from the press and the opinions of analysts, there is an ever-increasing collection of case studies and anecdotal evidence that suggests that data mining technology has moved beyond the proof-of-performance phase of its evolution and is maturing into a mainstream business practice. The following are a few anecdotes from some significant corporations that have found value in data mining.

4.3.1 A Financial Institution

The Center for Data Insight recently hosted analysts from a major financial institution considering data mining as a possible analytical tool. The analysts brought their own company's data into the CDI to recreate an analysis that they had already conducted using traditional statistical methods. Unica's Model 1 produced a predictive model in less than one minute that had a lower margin of error than a similar model that had taken the analysts two weeks to build by hand. This result alone was significant enough to convince the company to adopt data mining as a standard business practice.

4.3.2 A Petroleum Company

In another CDI engagement, a major petroleum company sought to profile, among other things, its most profitable credit card customers, and those customers who purchase super unleaded gasoline. The results of this study concluded that the company's best customers are men with a middle income who drive trucks. The study also found that super unleaded customers tend to drive older cars and those driving newer cars tend to purchase regular unleaded. The evaluation of these results concluded that people buy super-unleaded gas because their cars need the higher octane to run properly. Since these customers will purchase super-unleaded in any case, the company was able to save over $500,000 by eliminating an unnecessary ad campaign.

4.3.3 A Financial Institution

The following example was reported by the San Jose Mercury News, October 6, 1997, "DOMINATING WITH DATA - DATA MINING EMERGES AS THE FUTURE OF BUSINESS ANALYSIS."

Mellon Bank, based in Pittsburgh, acknowledges using data mining techniques for everything from fraud detection to target marketing and loan underwriting. But officials tend to skirt talking about specific findings.

"The financial service industry as a whole has been using this technology for quite a while," said Peter Johnson, vice president of Mellon's technology group. "The primary difference between what we're doing now and what we were doing in the 1980s was that we used to have to give the computer rules to help it make a decision. Now we have machine learning techniques that study the data and learn the rules themselves."

"The consumer benefits from smarter, better packages at discounted rates that are attractive to them," said Susan Maharis, Bell Atlantic's staff director of marketing information systems. "Product managers are saying, 'Gee, I never would have thought of putting those products together.' And consumers are getting more of what they want."

Consumers also are the focus of the data mining efforts inside Safeway's grocery stores in Great Britain. Safeway discovered through data mining that only eight of the 27 different types of orange juice in the stores sold enough to be profitable; the company responded by reducing the number of orange juice brands available. But when data mining revealed that only a dozen of more than 100 different types of cheese were profitable, the computer also noted that Safeway's most profitable customers were reaching for the least profitable cheese. The company kept all the cheeses rather than frustrating -- and possibly driving away -- its most valued customers.

4.3.4 A Financial Institution and a Direct Mail Order Company

The following examples were reported by the Minneapolis-St. Paul Star Tribune, 9/17/97

At U.S. Bancorp, formerly First Bank System, in Minneapolis, computer specialists are mining half a trillion bytes of computer data to find ways to minimize credit risks, cut customer attrition and boost the success rate for selling new bank services. "I think data mining is going to help us," said Richard Payne, vice president and manager of the customer decision analysis group at U.S. Bancorp.

U S West, the 14-state regional telephone company that provides most of the local phone service in the Twin Cities, is sifting through a database of customer orders in Seattle to see if data mining can help predict how customers will respond to new advertising campaigns.

"By the end of the year, this should be deployed to 200 key people marketing U S West products," said Gloria Farler, executive director of market intelligence and decision support for U S West in Denver. "We're going to give people who are selling telephone services a more precise way to do it."

And at direct-mail catalog companies Fingerhut Cos. Inc. in Minnetonka and Damark International Inc. in Brooklyn Park, data miners are predicting consumer buying trends by segmenting millions of U.S. customers into groups who exhibit similar purchasing characteristics. Fingerhut uses the information to tailor mailings of the 130 different catalogs it sends to consumers each year. "There are about 3,500 variables we now study over the lifetime of a consumer's relationship with us," said Andy Johnson, senior vice president of marketing at Fingerhut. "I can predict in the aggregate which customers will do similar things." Corporations such as Fingerhut have found that, by combining years worth of business transaction data with U.S. Census Bureau and privately gathered demographic information, they can predict consumer behavior well enough to profit from their knowledge.

Fingerhut is mining a data warehouse with information about more than 10 million current customers to find out which are most likely to buy certain types of products, and therefore should be mailed certain catalogs. Fingerhut mails catalogs to demographic niches such as home textiles, outdoor, home handyman, and holiday products.

"Data mining is a low-cost way for us to assess the buying behavior of groups of customers," Johnson said. In a recent data mining effort, Fingerhut studied past purchases of customers who had changed residences, and found that they were three times more likely to buy items such as tables, fax machines, phones and decorative products, but were not more likely to purchase high-end consumer electronics, jewelry or footwear.

Mining also showed that those buying patterns persist for 12 weeks after the consumer moves, but that purchases peak in about the first four weeks. As a result, Fingerhut has created a catalog aimed at people who have recently moved, and carefully tailored it to their purchasing patterns, Johnson said. He hastened to say the data mining results don't mean that all consumers who move will avoid buying computers or shoes, but that "people will not buy at a rate that would justify our investment in printing catalogs."

Another data mining effort discovered a subset of Fingerhut customers who responded more favorably to catalogs printed in Spanish. Billing, customer service, and customer correspondence also are provided in Spanish. Aside from the language in which they were printed, the Spanish catalogs varied only slightly from a standard Fingerhut catalog of the same type: They contained a somewhat greater emphasis on fine jewelry, which data mining showed to be particularly appealing to people who like Spanish language catalogs, he said. The result: The Spanish catalog generates about 40 percent more orders than would normally be expected from those consumers, Johnson said.

Data mining also helps Fingerhut identify customers who are not likely to purchase or not likely to purchase enough to make the catalog mailing profitable. This group, dubbed "potential attriters," is the target of special promotions to win them back, such as discount certificates. Customers identified as unlikely to be profitable enough to justify continued catalog mailings are likely to be dropped from Fingerhut's current mailing lists.

...Within the direct-mail catalog industry, data mining's biggest contribution to date has been in cutting costs but allowing companies to focus mailings on the customers most likely to buy, Johnson said. "The big carrot is to develop a way to find additional revenue. That's what everyone is seeking." He declined to say how much Fingerhut has saved as a result of data mining or how much additional revenue it has obtained. But, he adds, "We are the second-largest catalog company in the United States, and we would not be in business without the market segmentation data that data mining produces."

5.0 Early Adopters: An Industry Overview

This section presents an overview of the first industries to embrace data mining such as banking, insurance retail, telecommunications and direct mail order companies.. Many companies within these industries have benefited from their experiences as indicated by the previous section.

5.1 Finance

In 1988 Richard Fairbank and Nigal Morris shopped the concept around to 25 banks before Signet Banking Corporation decided to give data mining a try. Signet acquired behavioral data from many sources and used it to build predictive models. Using these models it launched the highly successful balance transfer card product that changed the way the credit card industry works.

Now Data mining is at the heart of the marketing strategy of all the so-called monoline credit card banks: First USA, MBNA, Advanta, and Capital One. [5]

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

MasterCard member financial institutions worldwide have used standard neural network technology to successfully predict and spot potential fraud, based on transaction irregularities and predetermined parameters. Neural networks help members identify fraud in real time and on a delayed basis. MasterCard estimates its members have saved more than $50 million by using neural network technology. [1]

5.2 Retail

Safeway, like several other large chains, has turned itself into an information broker. The supermarket purchases demographic data directly from its customers by offering them discounts in return for using a Safeway savings club card. In order to obtain the card, shoppers voluntarily divulge personal information that is later used in predictive modeling.

From then on, each time the shopper presents the discount card, his or her transaction history is updated in a data warehouse somewhere. With every trip to the store, the shopper presents the discount card and his or her transaction history is updated in a data warehouse somewhere. With each trip the shoppers tell Safeway a little more about themselves. Patterns were identified such as items that are bought together should be shelved together. The information gathered on individuals is of great interest to the manufacturers of the products that line the store's aisles.

Safeway charges 5.5 cents per name to suppliers who want their coupon or promotion to reach just the right people. Since the coupon redemption is also logged into the data warehouse, the response rate can then be calculated. Furthermore the customer's response, or lack of response can than be entered into the data warehouse for future predictive models. [5]

5.3 Insurance

Health Insurance Commission HIC, an agency of the Australian federal government developed a data mining solution that analyzes data and detects unnecessary prescriptions or referrals by medical practitioners. With data mining HIC can put in place procedures to reduce these costly incidents. [7]

5.4 Telecommunications

MCI uses data mining to score their database, which are numerical ratings of customers along certain dimensions. Say that potential customers with a high propensity to travel might seem like a good market for cellular service, and that such customers seem likely to accept a direct mail offer. The propensity of a customer to travel could then be indicated by a numeric score, which is used to analyze this new market. This process of creating a model and then scoring -- a traditional database marketing technique -- used to take MCI six weeks. Now scoring can be done in hours. These speed and cost savings give MCI a competitive advantage [6].

Bell Atlantic began with a small data mining "proof of concept" project involving a relatively simple analysis. Bell Atlantic used an association algorithm to determine which products are likely to lead to sales of other products. Here is one example of what it learned. In 51 percent of the cases, customers who purchased caller ID, call return, and repeat dial as one product set and three-way calling as a second product also purchased touch tone, call waiting, and call forwarding as a third product set. What that means is the remaining 49 percent of the customers -- who purchased only the first two product sets -- are also likely candidates to purchase the third product set.

To get that answer, Bell Atlantic extracted customer data from its 500-gigabyte marketing data warehouse, which resides in a DB/2 database on an IBM 3090 mainframe. Using a dial-up connection, Bell Atlantic then fed the data to Intelligent Miner, which runs on an IBM RS/6000 server at Vision headquarters. Eric Vignola, senior business intelligence consultant at Vision, directs the mining exercises from Bell Atlantic's New York office, using a laptop loaded with Intelligent Miner's client software. [19]

6.0 Data Mining Algorithms

This section presents some of the industry-accepted algorithms for automated data mining. These algorithms are summarized in Table 1 and described in more detail below.

Table 1 Basic Types of Data Mining Algorithms

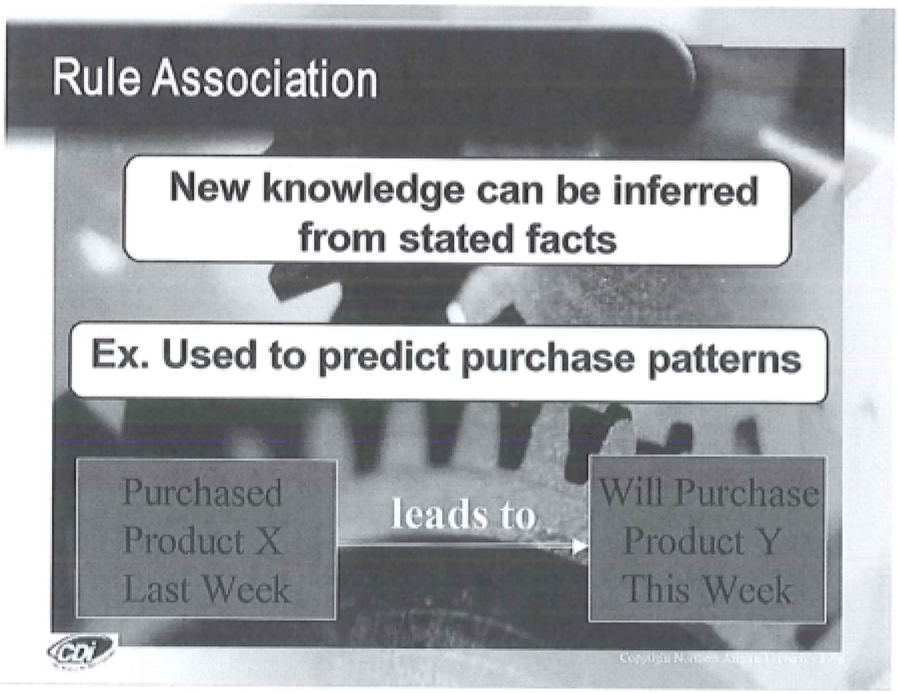
DATA MINING ALGORITHMS	
Rule Association	Identifies cause and effect relationships and assigns probabilities or certainty factors to support the conclusions. Rules are of the form "if <conditions>, then <conclusion>" and can be used to make predictions or estimate unknown values
Memory-based Reasoning (MBR) or Case-based Reasoning (CBR)	These algorithms find the closest past analogs to a present situation in order to estimate an unknown value or predict an unknown outcome.
Cluster Analysis	Separates heterogeneous data into homogeneous and semi-homogeneous subgroups. Based on the assumption that observations tend to be like their neighbors. Clustering increases the ability to make predictions.
Classification Algorithms and Decision Trees	Determines natural splits in the data based on a target variable. First splits occur on the most significant variables. A branch in a decision tree can be viewed as the conditional side of a rule. Algorithms such as Classification and regression trees (CART) or chi-squared automatic induction (CHAID) are the most common examples.
Artificial Neural Networks	Uses a collection of input variables, mathematical activation functions, and weightings of inputs to predict the value of target variable(s). Through an iterative training cycle, a neural network modifies its weights until the predicted output matches actual values. Once trained, the network is a model that can be used against new data for predictive purposes.
Genetic Algorithms	Uses a highly iterative process of selection, crossover, and mutation operations to evolve successive generations of models. A fitness function is used to keep certain members and discard others. Genetic algorithms are primarily used to optimize neural network topologies and weights. However, they can be used by themselves for modeling.

6.1 Rule Association Algorithms

Rule association algorithms generate rules of the form *lhs→ rhs* in which new knowledge, expressed on the right-hand side (rhs), can be inferred whenever the facts stated on the left-hand side (lhs) are true. An example of an association rule is "30% of transactions that contain beer also contain diapers."

Rule association algorithms are utilized in a number of ways. Such algorithms search for causal relationships in the database and generate a set of rules that can be incorporated into expert systems for deployment. Decision trees may be based upon rule association algorithms. There are additional algorithms that are grouped under the general category of rule association such as Apriori and affinity grouping techniques.

A typical application that uses rule association algorithms is Market Basket Analysis. Affinity grouping is often used in market basket analyses and groups items together with degrees of certainty.



Rule Association

New knowledge can be inferred from stated facts

Ex. Used to predict purchase patterns

Purchased Product X Last Week

leads to

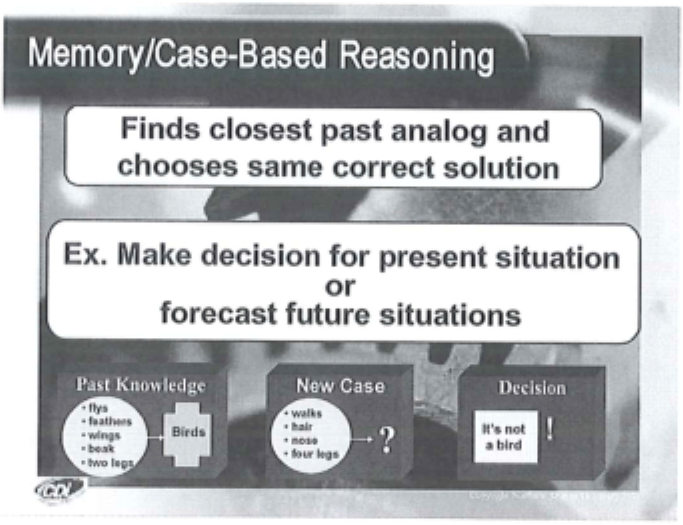
Will Purchase Product Y This Week

Copyright Northern Arizona University - [illegible]

Figure 3 Rule Association Algorithms

6.2 Memory-based Reasoning or Case-based Reasoning (CBR)

A more recent approach to resolving problems uses a database of previously solved cases as a repository for reusable solutions. Each case includes an identified problem and the solution that resolved it.



Memory/Case-Based Reasoning

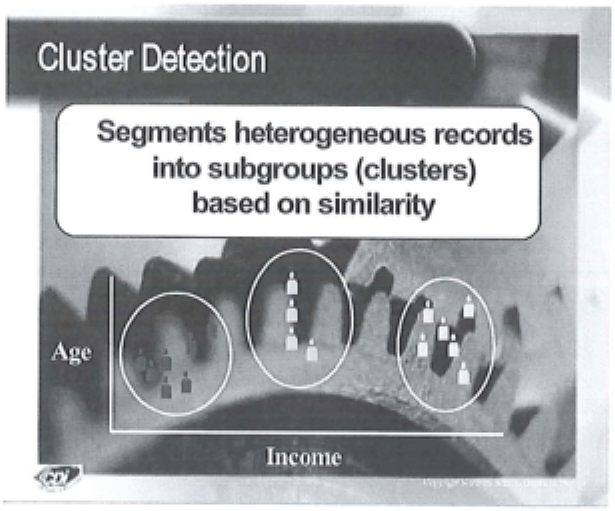
Finds closest past analog and chooses same correct solution

Ex. Make decision for present situation
or
forecast future situations

Figure 4 Case-Based Reasoning

6.3 Clustering

Clustering is the task of segmenting a heterogeneous population into a number of more homogenous subgroups or *clusters*. What distinguishes clustering from classification is that clustering does not rely on predefined classes. The records are grouped together on the basis of self-similarity.



Cluster Detection

Segments heterogeneous records into subgroups (clusters) based on similarity

Figure 5 Clustering

6.3.1 K-Means

In this algorithm, the first step is to choose the number of clusters (this is the K in K-Means). Next K "seeds" are chosen to be the initial guess at the centroids of the cluster. These seeds than divide up the data into K clusters. The mean of the data attributes in each cluster is then calculated. These new values denote the new seed or centroid of the data cluster. The process of recalculating the mean of each new data cluster is continued until the centroids stop moving (when the mean of the data attributes of that data cluster equals the centroid).

6.3.2 Nearest-Neighbor

Nearest Neighbor (more precisely k-nearest neighbor, also k-NN) is a predictive technique suitable for classification models.

Unlike other predictive algorithms, the training data is not scanned or processed to create the model. Instead, the training data is the model. When a new case or instance is presented to the model, the algorithm looks at all the data to find a subset of cases that are most similar to it and uses them to predict the outcome.

There are two principal drivers in the k-NN algorithm: the number of nearest cases to be used (k) and a metric to measure what is meant by nearest.

Each use of the k-NN algorithm requires that we specify a positive integer value for k. This determines how many existing cases are looked at when predicting a new case. k-NN refers to a family of algorithms that we could denote as 1-NN, 2-NN, 3-NN, and so forth. For example, 4-NN indicates that the algorithm will use the four nearest cases to predict the outcome of a new case.

6.4 Decision Trees and Rule Induction Algorithms

A decision tree is a technique that displays the results, usually in a graphical output, in the form of a tree. Decision trees are useful for directed data mining, particularly classification. Some of the popular algorithms for generating decision trees are described in further detail below.

A decision tree is generated by dividing records in a data set into disjoint subsets. Each subset of the original data set is described by a simple rule on one or more attributes. These disjoint subsets make up the leaves of the tree, with the root node containing all of the records. An example of a decision tree from Angoss' Knowledge Seeker is shown in Figure 6.

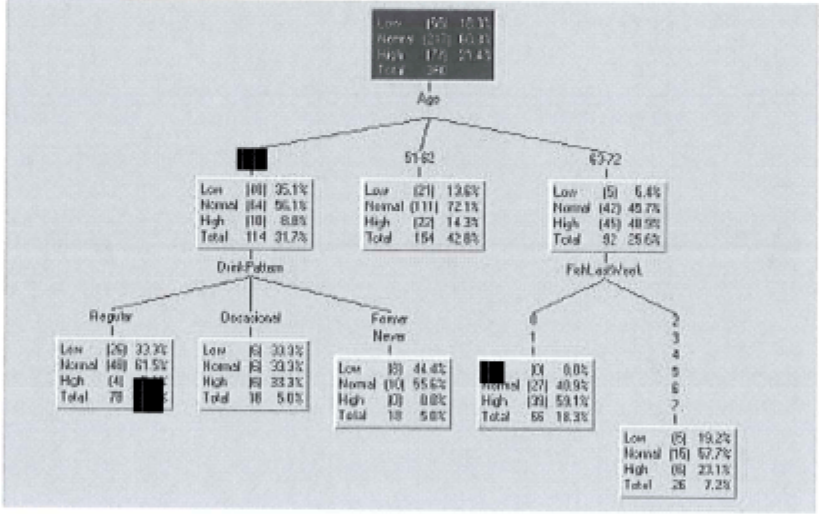


Figure 6 A decision tree as represented by Angoss' Knowledge Seeker..

6.4.1 Iterative Dichotomiser (ID3)

The ID3 algorithm (Quinlan '86) is a decision tree building algorithm which determines the classification of objects by testing the values of their properties. It builds the tree in a top down fashion, starting from a set of objects and a specification of properties. At each node of the tree, a property is tested and the results used to partition the object set. This process is recursively done until the set in a given subtree is homogeneous with respect to the classification criteria - in other words it contains objects belonging to the same category. This then becomes a leaf node. At each node, the property to test is chosen based on theoretic criteria that seek to maximize information gain and minimize entropy. In simpler terms, that property is tested which divides the candidate set into the most homogeneous subsets.

6.4.2 Classification and Regression Trees (CART)

The CART model is a predictive, statistical model used to classify remote sensing information. It employs statistical regression techniques (e.g. stepwise multiple regression) to construct dichotomous decision trees. Training set data is used as input to produce the decision trees.

CART builds a binary tree by splitting the records at each node according to a function of a single input field. The first task then is to find which of the independent fields makes the best splitter. The measure used to evaluate a potential splitter is called the *diversity*. A high index of diversity indicates an even distribution, while a low index indicates that the set contains an even distribution of classes. The best splitter is the one that decreases the diversity of the record sets by the greatest amount. In other words we want to maximize the difference between the diversity before split and the diversity of children after split. CART is able to classify on both continuous and categorical variables.

6.4.3 The C4.5 Algorithm

Quinlan proposed this algorithm in 1993. The C4.5 algorithm generates a classification-decision tree for the given data set by recursive partitioning of data. The decision tree is grown using a depth-first strategy. The algorithm considers all the possible tests that can split the data set and selects a test that gives the best information gain. C4.5 is very similar to CART. They both treat continuous variables in much the same way. They differ in that C4.5 treats categorical variables differently. When C4.5 assesses a categorical field's value as a splitter, its default behavior is to assume that there will be one branch for each value taken on by that variable. So for example if *color* is chosen as the best field on which to split the root node than there will be a child node for each color presented in the data.

6.4.4 Chi-squared Automatic Interaction Detection (CHAID)

J.A. Hartigan first published CHAID in 1975. Using tests of statistical significance, CHAID automatically sub-groups data. The principle difference between CART, C4.5, and CHAID is that CHAID is restricted to categorical variables. In order to develop a tree from CHAID, a continuous variable needs to be binned into a set of ranges such as *high*, *medium*, and *low*. CHAID also attempts to stop growing the tree before overfitting occurs.

As in the other two above algorithms, CHAID searches for a way to use the input variables to split the training data into two or more child nodes. The child nodes are chosen in such a way that the probability of the target field taking on a particular value differs from node to node.

The algorithm chooses the splitter based on the X^2 test. The X^2 test is defined as the sum of the squares of the standardized differences between the expected and observed frequencies of some occurrence in each sample. The test is a measure of the probability that an apparent association is due to chance or, inversely, that an observed difference between samples is due to chance. The predictor that generates the groupings that differ the most (according to this test) is chosen as the splitter for the current node.

6.4.5 Supervised Learning In Quest (SLIQ)

SLIQ (Supervised Learning In Quest) developed by IBM's Quest project team, is a decision tree classifier designed to classify large training data. It handles both numeric and categorical attributes. It uses a pre-sorting technique in the tree-growth phase. This helps avoid costly sorting at each node. SLIQ keeps a separate sorted list for each continuous attribute and a separate list called a class list. An entry in the class list corresponds to a data item. Each entry has a class label and a name of the node it belongs to in the decision tree. An entry in the sorted attribute list has an attribute value and the index of the data item in the class list.

SLIQ grows the decision tree in breadth-first manner. For each attribute, it scans the corresponding sorted list and simultaneously calculates entropy values for each distinct value of all the nodes in the frontier of the decision tree. After the entropy values have been calculated for each attribute, one attribute is chosen for a split for each node in the current frontier, and they are expanded to have a new frontier. Then one more scan of the sorted attribute list is performed to update the class list for the new nodes.

While SLIQ handles disk-resident data that is too large to fit in memory, it still requires some information to stay memory-resident. This grows in direct proportion to the number of input records, putting a hard limit on the size of training data. The Quest team has recently designed a new decision-tree-based classification algorithm, called SPRINT (Scalable PaRallelizable INduction of decision Trees), that for the moment removes all of the memory restrictions.

6.4.6 Scalable PaRallelizable INduction of Decision Trees (SPRINT)

Most of the current classification algorithms require that all or a portion of the entire data set remain permanently in memory. This limits their suitability for mining over large databases. SPRINT removes all of the memory restrictions, and is fast and scalable. The algorithm has also been designed to be easily parallelized, allowing many processors to work together to build a single consistent model. The combination of these characteristics makes SPRINT an ideal tool for data mining.

6.4.7 Naïve Bayes

Naïve-Bayes is a classification technique that is both predictive and descriptive. It analyzes the relationship between each independent variable and the dependent variable to derive a conditional probability for each relationship. When a new case is analyzed, combining the effects of the independent variables on the dependent variable makes a prediction variable (the outcome that is predicted). In theory, a Naïve-Bayes prediction will only be correct if all the independent variables are statistically independent of each other, which is frequently not true. For example, data about people will usually contain multiple attributes (such as weight, education, income, and so forth) that are all correlated with age. In such a case, using Naïve-Bayes would be expected to over-emphasize the effect of age. Notwithstanding these limitations, practice has shown that Naïve-Bayes produces good results, and its simplicity and speed make it an ideal tool for modeling and investigating simple relationships.

Naïve-Bayes requires only one pass through the training set to generate a classification model. This makes it the most efficient data mining technique. However, Naïve-Bayes does not handle continuous data, so any independent or dependent variables that contain continuous values must be binned or bracketed.

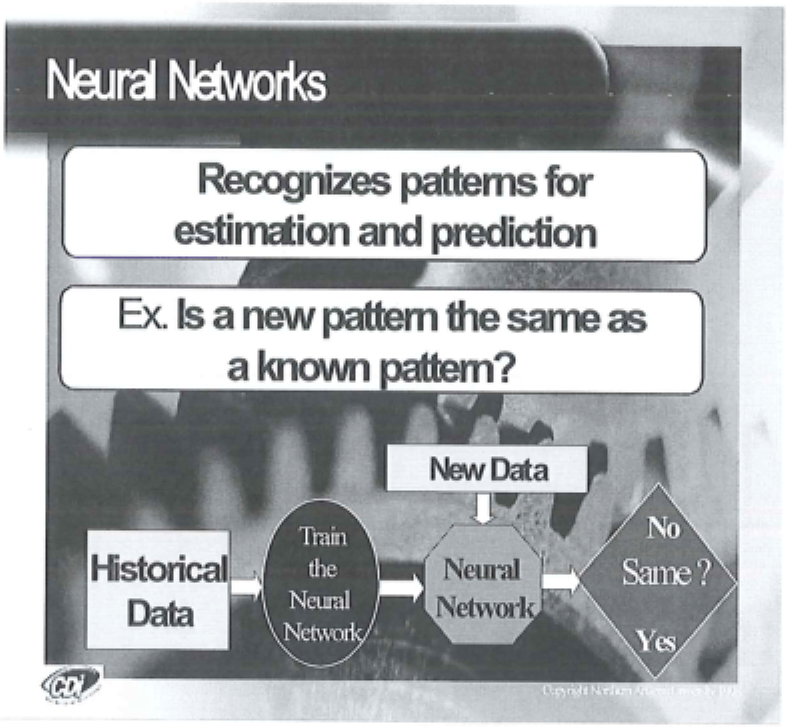
Using Naïve-Bayes for classification is a fairly simple process. During training, the probability of each outcome (dependent variable value) is computed by counting how many times it occurs in the training dataset. This is called the prior probability. For example, if the Good Risk outcome occurs 2 times in a total of 5 cases, then the prior probability for Good Risk is 0.4. You can think of the prior probability in the following way: "If I know nothing else about a loan applicant, there is a 0.4 probability that the applicant is a Good Risk." In addition to the prior probabilities, Naïve-Bayes also computes how frequently each independent variable value occurs in combination with each dependent (i.e. output) variable value. These frequencies are then used to compute conditional probabilities that are combined with the prior probability to make the predictions. In essence, Naïve-Bayes uses the conditional probabilities to modify the prior probabilities.

6.5 Neural Networks

Artificial neural networks mimic the pattern-finding capacity of the human brain and hence some researchers have suggested applying Neural Network algorithms to pattern mapping. Neural networks have been applied successfully in applications that involve classification. In their most common incarnation they learn from a training set. Patterns are learned by the neural that are then the basis for classification and prediction.

One of the chief advantages of neural networks is their wide applicability. Neural nets can be adapted to solve a wide variety of problems. Neural Nets also have the advantage that they scale very well to multiple processors. This is useful for very large data sets that are housed on multiprocessor servers.

Neural networks have two major drawbacks. The first is the difficulty in understanding the models they produce. The second is their particular sensitivity to the format of incoming data. Different data representations can produce different results; therefore, setting up the data is a significant part of the effort of using them.



Neural Networks

Recognizes patterns for estimation and prediction

Ex. Is a new pattern the same as a known pattern?

Figure 7 Neural Networks

6.6 Genetic Algorithms

Genetic algorithms are primarily used as a technique for the solution of various combinatorial or optimization problems. Genetic algorithms are also being used for data mining and are included in the present overview.

The name of the method derives from its similarity to the process of natural selection. The objective may be to find an optimal solution from many possible solutions based upon some defined criterion. For example, a task may be to find a fixed number of market parameters that have the most significant influence on the performance of a given market. Many combinations of parameters may be possible.

The underlying algorithms are driven by three mechanisms that have a relationship to genetic selection. First is a selection of the strongest possible solutions. Second is similar to cross breeding where new solutions are generated by mixing aspects of the parent sets. Third is similar to mutation where accidental changes get generated from the population. After a number of new generations of solutions are built a solution that cannot be improved any further is identified. This solution is taken as a final one.

Genetic algorithms have two weak points. First, the very way of formulating the problem deprives one of any opportunity to estimate the statistical significance of the obtained solution. Second, a specialist is needed to formulate the problem effectively. Genetic algorithms are more an instrument for scientific research rather than a tool for practical data analysis.

7.0 Data Mining Companies: Pieces of a Larger Puzzle

This section examines the key data mining companies that are providing products that support the advancement of this industry. This list does not necessarily include every company who purports to provide a data-mining product or service. However, this list is representative of products and services available in this industry at this point in time. These companies generally fall into one of two categories, data mining software vendors or data mining service providers. The companies in each category are presented in a chronological order as each emerged onto the market. The following excerpts describe a product or service provided by each company. These excerpts are based on material from the vendor web-sites.

7.1 History of Data Mining - Software Vendors

IDIS - Data Mining Suite (no date available)

Information Discovery, Inc. is a provider of large-scale data mining oriented decision support software and solutions. Decision support needs are served with the pattern discovery and data mining software, strategic consulting and warehouse architecture design. Information Discovery, Inc. provides data mining technologies and customized business solutions.

NeoVista – Decision Series (No Date available)

NeoVista specializes in advanced pattern detection techniques. Highly parallel computing solutions were employed in defense related industries. Now, NeoVista's Decision Series software suite brings the same advanced technology to commercial data mining. Decision Series software can be deployed on scalable, parallel platforms that are available from accepted, standard hardware providers and operate against data resident in popular databases or in legacy systems. The data mining solutions from NeoVista thereby augment existing decision support environments by integrating with installed standards based systems.

Angoss - Knowledge Seeker (1984)

ANGOSS Software Corporation is an international publisher of software with offices in North America, subsidiaries in Europe, and distributors in the countries of Australia, South Africa and Japan. ANGOSS was founded in 1984 and went public in 1993. KNOWLEDGE SEEKER is its data-mining product. KNOWLEDGE SEEKER has been successfully used by a large number of customers to date. ANGOSS has a solid reputation for excellence in Knowledge Engineering, this is the process of gathering business value from organization data using Knowledge Discovery tools.

Pilot - Discovery Server (1984)

Pilot Software, a subsidiary of Platinum Equity Holdings provides both OLAP and data mining products. More than 100,000 users in industries such as financial services, consumer packaged goods, telecommunications, health care and retail have used online analytical processing (OLAP) and data mining products since its founding in 1984. It is unknown to the writers of this report how many use OLAP versus data mining. Headquartered in Cambridge, Mass., Pilot Software has a strong international presence with offices worldwide.

Reduct & Lobbe (1986)

Reduct & Lobbe is a computer software company specializing in machine learning, decision-aid and idea generation technologies. Their primary expertise is the application of adaptive methods for decision support, performance improvement and automation.

ISoft - Alice (1988)

ISoft has been specializing in applying intelligent software technologies to the business world since 1988. Isoft has acquired experience in Machine Learning and Knowledge Discovery techniques through participation in major European research projects. ISoft has developed packages that allow software developers to build agent-based decision support systems.

Business Objects - Business Miner (1990)

Business Objects introduced the concept of the "semantic layer " in 1990, a patented technology that maps complex databases to a business representation understandable by non-technical end-users.

HNC (1990)

Robert Hecht-Nielsen and Todd W. Gutschow, who had worked together at TRW in a neurocomputing research and development program, founded HNC in 1986. At the beginning, HNC offered training in neural network technology, as well as neural network software and hardware tools for users who wished to design and build their own application software. In 1990, the company changed its business strategy from tools to solutions and began planning for the development and marketing of intelligent, decision-support application software - the kind of predictive software solutions that HNC offers today. HNC has grown along with its product offerings and, as of the end of fiscal year 1997, had over 700 employees and revenues of $113 million. It became a public company in June of 1995.

MIT - Management Intelligenter Technologien Inc. - DataEngine (1991)

MIT was founded in 1991. MIT provides a package of integrated services for fuzzy logic and neural networks. Presently MIT consists of an interdisciplinary team of 25 people in engineering, management and computer science.

Integral Solutions Ltd - Clementine, (1992)

ISL was formed in 1989 as a management buy-out of the AI Products Divisions from SD-SCICON (later to become EDS). For its first 5 years ISL researched, developed and supplied Artificial Intelligence (AI) and Knowledge Based Systems (KBS) technology. Products include PC-PACK and Poplog TM. ISL's research in Safety Critical Decision Support systems for Healthcare has led to a joint venture with another company, ICRF Infermed. In 1992, ISL began to develop Clementine, a data mining product. ISL claims that Clementine is as simple to learn as a spreadsheet, as it hides details of the advanced technologies involved while offering rich facilities and high productivity to sophisticated 'power' users.

Partek (1992)

A privately owned corporation headquartered near St. Louis Missouri has been developing and marketing analysis and modeling software since 1992.

Unica Technologies – PRW & Model 1(1992)

Massachusetts Institute of Technology alumni founded Unica Technologies, Inc. in 1992. The focus has been to apply advances in machine learning, statistical, and neural network technologies to solve data mining problems in business. Unica is a privately held company located in the Boston metropolitan area. The Unica team has over 70 years of combined industry experience in pattern recognition, neural networks, machine learning, artificial intelligence, and computer science. Unica offers the Pattern Recognition Workbench (PRW) and Model 1 software for data mining.

Megaputer Intelligence, Ltd. (1993)

Megaputer Intelligence, Ltd. (MPI) was founded in 1993 in Moscow, Russia. The research and development team of the company includes expertise in artificial intelligence, programming and analysis.

Polyanalyst 3.3, PolyAnalyst Lite (1993)

PolyAnalyst started as a research and development endeavor at the Moscow State University in 1989. When the new technology started delivering successful business predictive models, the project acquired a commercial aspect.

DataMind - DataCruncher (1994)

DataMind was founded in 1994 and has headquarters in San Mateo, California. DataMind provides a data mining product. Regional sales centers are located in Atlanta, Boston, Chicago, San Mateo, and Paris, France.

Hyperparallel - Discovery (1994)

Founded in 1994, Hyperparallel provides parallel computing technology that allows its Knowledge Discovery building blocks to exploit today's massively parallel computers.

Quadstone – Decisionhouse (1994)

Quadstone specializes in the production and use of software to perform large-scale analysis of customer databases. The company's Decisionhouse software provides a fully scalable suite of tools for integrated data access, visualization, analysis and mining.

Quadstone is a privately held, limited liability company which was founded in November 1994 and started operations in March 1995. The founding directors control the company with staff members and external venture capital companies Northern Venture Managers and Scottish Development Finance owning the balance of equity. There are no plans at present to become a quoted company.

Thinking Machines – Darwin (1995)

Thinking Machine started in the mid-1970's. It was originally in the business of producing a massively parallel computer titled the Connection Machine. This product developed an expertise in parallel processing and technologies for processing very large data sets. Thinking Machines determined that this core expertise could be adapted and entered the data mining industry in the early 1990s. Darwin®, TMC's high-end data mining software suite, enables users to extract meaningful information from large databases -- information that reveals hidden patterns, trends, and correlations -- and allows them to make predictions that solve business problems.

Forwiss - Delta Miner (1996)

The basic techniques of the Delta Miner were developed at FORWISS where, since 1993, a research group has investigated algorithms for Data Mining. Delta Miner was recognized as one of the best three products in the category "Business Management Solutions" at the Systems '96 trade show in Munich.

IBM - Intelligent Miner (1996)

Intelligent Miner is a new product area for IBM building upon its core expertise in databases and computing. Intelligent Miner supports mining of data stored in relational databases and flat files.

Intelligent Miner can be used to discover associations or patterns, to segment (or cluster) records based on similarity of attributes, to discover similar time sequences, or to create predictive or classification models.

SGI - MineSet (1996)

SGI has a strong core competency in high-speed computing and computer graphics. SGI developed MineSet as a data-mining product. MineSet supports a scalable client/server model, taking advantage of Silicon Graphics' high-performance ORIGIN handling large data sets. MineSet's client and server modules are supported on all Silicon Graphics platforms running Irix 6.2 or above. MineSet functionality can also be made available on PCs running Hummingbird Communication's Exceed 3D and other UNIX® X-servers supporting the industry-standard OpenGL®.

SRA - KDD Toolset (1996)

SRA International, Inc. was founded in 1978, and with major corporate locations in Arlington and Fairfax, Virginia, SRA serves clients from offices across the United States. SRA offers consulting and systems integration to both government and commercial organizations.

Cognos - Scenario (1997)

Founded in 1969, Cognos is an international corporation with corporate headquarters in Ottawa, Canada, and U.S. sales headquarters in Burlington, Massachusetts, USA. Cognos provides a data mining product. Cognos operates more than 32 offices in 12 countries around the world such as Australia, Belgium, France, Germany, Hong Kong, Italy, Japan, The Netherlands, Singapore, South Africa, Sweden, and the United Kingdom. The company employs more than 1,400 people worldwide. Cognos products are sold through a direct sales force, and an extensive network of 800 resellers and distributors.

Dialogis Software & Services Inc. - Kepler (1997)

Kepler is currently in Version 1.0 beta. In September (1997), it was presented at the ILP and KDD Summer School and at the annual conference of the GVC (Process Technology and Chemical Engineering Society) in Dresden.

Nuggets (1997)

Nuggets uses proprietary search algorithms to develop English "if - then" rules. These algorithms use genetic methods and learning techniques to "intelligently" search for valid hypotheses that become rules. In the act of searching, the algorithms "learn" about the training data as they proceed. The result is a very fast and efficient search strategy that does not preclude any potential rule from being found. The new and proprietary aspects include the way in which hypotheses are created and the searching methods. The user sets the criteria for valid rules. Nuggets also provides a suite of tools to use the rules for prediction of new data, under-standing, classifying and segmenting data. The user can also query the rules or the data to perform special studies.

SPSS (1997)

SPSS claims "in-house quantitative experts are continuing their preference for existing statistics-based data mining tools with SPSS and SAS (at 30 percent and 20 percent respectively) overwhelming the newer entrants." The SPSS web-site also states that in its January 1998 report "Data Warehouse Marketing Trends/Opportunities: An In-Depth Analysis of Key Market Trends," leading data mining industry analysts from the Meta Group solidified an increasing preference for statistical solutions, specifically SPSS, in data mining:

TriVida (1997)

TriVida is a new company. It claims that it has a product that performs "continuous data mining" that automatically discovers patterns and relationships in incoming data, and autonomously adapts those solutions as changing data alters these patterns and relationships. It has server-based applications to analyze the data streams of Internet, Intranet, Extranet and client/server enterprise applications

SAS - Enterprise Miner (1998)

SAS is a well established company with the largest market share for OLAP and statistical processing software. It has recently developed a data mining product titled Enterprise Miner which contains functions such as a Process Flow Diagram, a drag-and-drop graphical user interface (GUI), automating the entire data-mining process of sampling, exploring, modifying, modeling, and assessing customer data. The GUI provides the common ground for the three diverse groups - business analysis, quantitative professionals, and IT - who must collaborate to create successful customer-relationship management systems and perform other data-mining activities.

Red Brick - Data Mine (1998)

Red Brick Data Mine, a component of Red Brick DecisionScape(SM), is Red Brick's server-integrated data mining solution designed to help information systems staff and business analysts predict and understand the impact of decisions on their customers, businesses and markets.

7.2 History of Data Mining - Service Providers

Data Distilleries (1995)

Data Distilleries (DDI) was founded in September 1995 as a spin-off of the Dutch National Center for Mathematics and Computer Science (CWI) to shorten time to market for research results. In the two years of its existence, Data Distilleries has become a provider of data-mining solutions in the Netherlands, with some of the largest banks and insurance companies as reference sites.

Retrograde Data Systems (1994)

Retrograde Data Systems is a vendor neutral Data Mining consulting company offering products, services and programs that are specifically designed to bring business professionals and information technologists up to speed with Data Mining. Retrograde was founded in 1994 by Ms. Lelia Morrill who has worked with several Fortune 100 companies developing Information Management Strategies, Data Warehouses and Decision Support Systems since 1984. A primary objective of the company is to extend Data Mining to the general business community. Retrograde was established to develop vertical packages for insurance around a powerful Data Mining Engine.

Knowledge Discovery 1 (1996)

Knowledge Discovery One, Inc. (KD1) was founded in January 1996 to build complete, sophisticated, yet easy-to-use applications that allow retailers to better understand and predict their customers' buying habits. Employing advanced knowledge discovery and data-mining techniques, KD1's Retail Discovery Suite allows the retailer to operate a more profitable organization by providing a complete and detailed understanding of their advertising, merchandising, assortment, inventory, promotions, and vendor performance issues.

Exclusive Ore (No Date Available)

Rob Gerritsen and Estelle Brand, two experts in data mining and data base management founded exclusive Ore Inc. They have worked directly with over 15 of the top data mining products.

KPMG Peat Marwick LLP (1997)

Founded in 1897, KPMG Peat Marwick LLP is one of the major, world-wide companies providing a large variety of business services. KPMGs core services include Audit and Accounting, Tax, Management Consulting, Corporate Finance and Corporate Recovery. KPMG developed a new capability in the 1990's to provide services in data warehousing. It has now extended that core competency to also include data mining services. KPMG has funded the development of the Center of Data Insight where it has access to a complete facility for training clients in data mining technologies and to conduct proof-of-concept projects to structure a data mining solution for each client. KPMG will work with a client to define the business objectives, to conduct pilot projects to identify the best suite of products to implement an integrated knowledge discovery system and develop and deploy that system into the client's business environment.

Center for Data Insight (1997)

The CDI is a nonprofit research and development center for Knowledge Discovery in Databases issues. The CDI is embodied in a state-of-the-art laboratory which supports the best-in-class data mining tools with a staff that possesses expertise in both tools and data mining methods. The academicians in the CDI represent a collaborative effort between the colleges of engineering, business, and mathematics at Northern Arizona University in Flagstaff, Arizona. The CDI is funded by the KPMG Peat Marwick consulting firm and maintains partnerships with many of the leading software vendors in this field.

Two Crows (No Date Available)

The principals of Two Crows Corporation provide consulting in data mining, data warehousing, and building statistical models. They have analyzed data, built and marketed commercially successful software, created applications, and managed and analyzed large databases for major companies.

Figure Figure 8 contains a graph showing the growth of the data mining industry since 1980. Much of this growth has been stimulated by successes in artificial intelligence and parallel computing and from advances in data storage capacities.

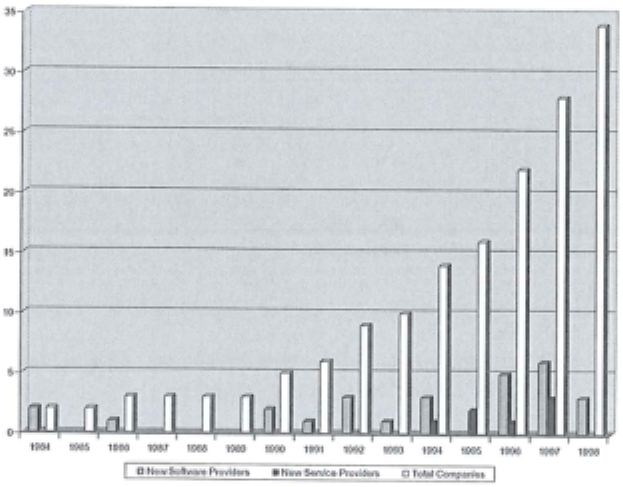


Figure 8 Growth of the data mining industry

8.0 A Move Toward Vertical Applications

There appears to be two fundamental paradigms for data mining software design: generalization or verticalization. Generalized applications are designed to solve a wide-variety of data mining problems regardless of the problem domain or business industry. Vertical applications present a usage environment and collection of algorithms that is highly focused on a particular problem domain (e.g., direct marketing) or a particular industry (e.g., retail banking). The benefit of verticalization is it provides a familiar environment to the user and offers the opportunity to guide the user through the complexities of the KDD process. Applications can be further customized to exactly fit the end user's needs. Generalized applications have the benefit of a broader customer base. However, due to their generality, the user is often left to determine his/her own methodology and interpretation of results. For many of the general-purpose tools this translates into steeper learning curves and the potential for spurious results.

8.1 The Tendency of Companies to Verticalize applications

Of the twenty-eight data mining software companies overviewed in this report only four are currently taking a verticalized approach to data mining. However, it is notable that these vendors are among the newer companies in the data mining game. Based on observations and experiences in the Center for Data Insight, it appears that the next stage of evolution in data mining will include increased verticalization of applications. While these applications may be based upon familiar underlying algorithms the products will be highly customized to provide a familiar environment for each user.

8.2 Who is Specializing in What?

Unica's Model 1 is specializing in marketing through modules such as: Response Modeler, Customer Segmenter/Profiler, Cross Seller, and Customer Valuater. Through extensive involvement with various business applications, Unica thoroughly understands the process of using pattern recognition and related technologies to solve data mining problems. MIT alumni founded Unica with a focus on applying advances in machine learning, statistical, and neural network technologies to solve data mining problems in business.

HNC's Database Mining Marksman is a predictive modeling system designed for analysis for direct marketers. Robert Hecht-Nielsen and Todd W. Gutschow founded HNC. They had worked together at TRW in a neurocomputing research and development program. At the beginning, HNC offered training in neural network technology, as well as neural network software and hardware tools for users who wished to design and build their own application software. In 1990, the company changed its business strategy from tools to solutions and began planning for the development and marketing of intelligent, decision-support application software

IDIS Information Discovery Inc. has verticalized through targeting Banking and Financial Services, Retail and Packaged Goods, Direct Marketing, Web-Log and Access Data Mining, and Manufacturing Quality and Warranty. They accomplish this through adapting their product to the particular business desired.

HyperParallel has verticalized within Retail, Banking, and Telecom. This company provides enabling technology, methodology, and practices to Global 1000 firms that are transforming themselves to customer relationship marketing organizations. Also provided is a portfolio of data mining algorithms, application recipes, and best practice models for the use of enterprise data in orchestrating customer relationship marketing initiatives.

9.0 The Analysts' view of the future market

Predicting information technology is quite difficult. Bill Gates thought in 1981 "640k of memory ought to be enough for anybody". Thomas Watson, IBM chairman fifty years ago, was also said to have commented that "there is a world for *maybe* 5 computers". In addition, Ken Olso the founder of Digital Equipment Corporation (1977) could not find "a reason why anyone would want a computer in their home". As it happens they were all wrong. However, for completeness in this report, listed below are some analysts' views of the future of data mining.

According to DataQuest, a market research arm of Gartner Group, license revenue estimates for data mining tools worldwide in 1997 was $40.6 million. This number was restricted in that they decided not to include SPSS and SAS in the survey since these companys' tools in 1997 did not provide data mining functionality. This number also does not include consulting services provided by vendors which could easily double the total revenue number.

DataQuest only counted tool sales in their estimate. These sales represent a decreasing portion of vendors' product lines. For example DataQuest only took into account Group1's Pattern Recognition Workbench (PRW) revenue, not their more vertical applications such as Model One. According to Scott Sassone, Unica's director of product development, "Vertical applications are the way to go. We see a limited market for the toolkit. It tends to sell only to a [redacted]. or a very sophisticated statistician."

DataQuest concluded that the data mining software market leader was IBM, with 15 percent of overall tool license revenue. Rounding out the top five are Information Discovery with 10 percent; Unica with 9 percent; Silicon Graphics at 6 percent; and Quadstone also with 6 percent. This leaves more than 54 vendors to fight over the other 54 percent of the tools on the market, as shown in the chart in Figure 9.

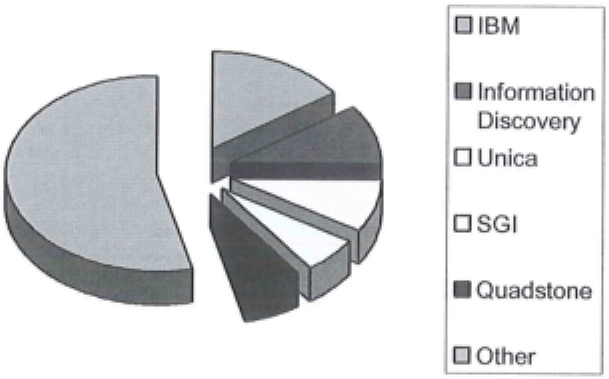


Figure 9 Market Share of Data Mining tool vendors for 1997

These numbers suggest that the market leader, IBM, had only $6.2 million in license revenue for IntelligentMiner in 1997. As report co-author Peggy O'Neill noted, "This really puts things into perspective when you consider the size of IBM's distribution channels." O'Neill offers several theories for why the market remains so small: the difficulty in using the software, the need for extensive consulting services, and the limited reach of many vendors.

DataQuest predicts tool sales using these measures will peak in 1999 at $75 million and fall to $53 million by 2002. The fact that revenue is expected to increasingly derive from applications explain the negative growth rate for tool sales over the next several years.

DataQuest agrees. According to the report, the successful companies are distancing themselves from the enabling technology and have started to develop applications for solving problems.

DataQuest predicts that many vendors will die if they don't make a successful transition into applications, which it sees as a quick and easy way to continue growth. [19]

A research group at Aston University at Birmingham called Group 9 on the future of Data Mining states that, "...as the size of databases continue to rapidly grow throughout the years, the use of data mining tools will become more and more vital. The next and most logical step for these tools is to make them a standard feature within a data warehouse/software package. Indeed, this has already been done with Braincel – a neural network driven data mining tool that has been incorporated in to the Excel and Lotus 1-2-3 software products. This has helped to create the first ever 'intelligent spreadsheet'. This means that complex, fuzzy and rapidly changing data can all now be analyzed with minimum user intervention. In fact, one of the main aims for data mining tools is to make them as automated as possible, thus limiting the amount of specialized knowledge a user needs to have in order to get the best results from the tools." [16]

"...To conclude, data mining tools are heading in the direction that they will have multi-media analysis abilities, and be able to analyze several types of databases simultaneously. The end-users or professional data miners will be the people operating the tools that are embedded in to standard software packages, most probably utilizing neural network technology." [16]

10.0 Business Models for an Integrated Solutions Company or Consortium

Data mining software tools are currently focused on particular phases in the knowledge discovery lifecycle. While many of these tools are very mature and well developed, collectively they do not integrate particularly well. Users of these software products face relatively steep learning curves that do not necessarily translate well from one tool to the next. This current state of the art provides golden opportunities to the company that brings all the pieces together into one, seamless integration. There are a variety of approaches that might be taken in order to advance the state of the art to the next level.

10.1 A Collection of Software Offerings

There are some OLAP software vendors such as Cognos and Business Objects who have ventured into the KDD arena with data mining products. Other data-mining vendors provide built-in utilities for handling some data cleansing and data transformation activities. If these companies grow their product line, they are uniquely poised to develop a collection of independent tools that encompass all of the key elements in the KDD process. The benefit of such a collection of tools would be a similar look-and-feel with distinctly different functionality. One might view this approach as similar to Microsoft's Office Suite, Lotus 1-2-3, and other similar software products. Each tool may be used by itself but the common usage environment makes it simple to transition from one tool to the next. Furthermore, the output from one tool can seamlessly be imported into other tools. A key element in such a tool suite is a tool for model deployment into the business environment. For example, if the model's purpose is to assist in direct marketing campaigns, then the tool might produce a desktop application for each member of the sales force to help determine what actions to take with which customers.

10.2 Full Service Data Mining

Currently none of the data mining software supports all of the KDD activities from data warehousing through data mining and knowledge deployment. While many tools provide mechanisms for data cleansing and some data transformation, these facilities are limited to the more routine data quality tasks. An ultimate data mining and decision support application might:

- Provide direct access to the data warehouse for on-line analytical processing
- Support rapid data extraction
- Provide for data quality verification and correction
- Allow data cleansing, transformation, and preprocessing
- Provide multidimensional data visualization capabilities
- Provide a broad variety of data mining algorithms
- Generate useful and comprehensive reports
- Export actionable deployment strategies

Furthermore such a tool would guide the analyst through each of the phases in the knowledge discovery cycle and enhance the effectiveness of the process. The first company that successfully brings the concept of a full-service data-mining tool to market will advance the current state of the art to the next generation of data mining.

10.3 Vertical Solutions Providers

Of the data mining software companies presented in this report only four provide a focused, vertical solution such as direct marketing or campaign management. All other products present generally applicable solutions. The benefit of providing a vertical solution for a particular industry or problem domain across industries is that the end user need not be an analyst to use the tool. Most of the generalized tools require some understanding of analytics in addition to business domain knowledge to produce valid models. Verticalized tools can more easily provide mechanisms for the average business user to generate and validate predictive models. Unica's Model 1 is an example of such a tool. It's focus on direct marketing allows an environment that guides the user from step to step and presents the results in a direct marketing context and in terms that marketing personnel are familiar with.

Opportunities are ripe for additional verticalized tools to emerge and be successful on the data mining market. Once again, model deployment should be provided with such a tool. Many current tools are capable of exporting C, Java, or C++ source code for further application development. Some tools export SQL queries for extracting a data subset from the database for focused campaigns. However, tools do not currently provide direction to the end user as to what actions might be taken based on the results of a mining analysis. Perhaps more than any other option, a verticalized decision support system can realistically implement this deployment concept.

Bibliography

[1] "Data Analysis for Forecasting, Fraud Detection, and Decision-Making at Los Alamos National Laboratory" Los Alamos Projects, http://www-xdiv.lanl.gov/XCM/research/genalg/members/hillol/datam_lanl.html

[2] "IBM Taking On Data Mining" Boulder Camera, February 27, 1988

[3] Alexander, Steve: Staff Writer "Bancorp Are Learning More About Their Customers Through Data Mining" Minneapolis-St. Paul Star Tribune, 17 August 1997, Section: Business Page: 1D

[4] Berry, Michael J.A. and Linoff, Gordon, "Data Mining Techniques: For Marketing, Sales, and Customer Support", John Wiley & Sons, Inc. 1997.

[5] Brand , Estelle and Gerritsen , Rob "DBMS, Data Mining Solutions Supplement" White paper from Exclusive Ore Inc., [redacted]

[6] Brooks, Peter "Targeting Customers" DBMS, December 1996

[7] Burkstrand, Beth: Staff Writer "Va. Tech Firm's Stock Surges 76% in Debut" Washington Post, June 12, 1998, Page F01

[8] Edelstein, Herb "Where the Data Mining Gold Rush Is Headed" Database Programming & Design, vol. 10 no. 12, December 1997, p. 78.

[9] Fayyad, Usama, Piatetsky-Shapiro, G. and Smyth, P., "From Data Mining to Knowledge Discovery: An Overview", Advances in Knowledge Discovery and Data Mining, AAAI Press, 1996B.

[10] Fayyad, Usama, Piatetsky-Shapiro, G. and Smyth, P., "Knowledge Discovery and Data Mining: Towards a Unifying Framework", Proceedings of The Second

[11] IBM Case Study, [redacted]

[12] International Conference on Knowledge Discovery and Data Mining, Edited by Evangelos Simoudis, Jiawei Han, and Usama Fayad, August 2-4, 1996A.

[13] J.C. Shafer, R. Agrawal, M. Mehta, "SPRINT: A Scalable Parallel Classifier for Data Mining", Proc. of the 22th Int'l Conference on Very Large Databases, Mumbai (Bombay), India, Sept. 1996.

[14] M. Mehta, R. Agrawal and J. Rissanen, "SLIQ: A Fast Scalable Classifier for Data Mining", Proc. of the Fifth Int'l Conference on Extending Database Technology, Avignon, France, March 1996.

[15] O'Harrow Jr., Robert: Washington Post Staff Writer "No More Secrets In Data Mining, You're Ore" Arizona Republic, March 22, 1998: Section: Business Page: D1

[16] Patel, L., Perry, S., Taylor, D., Brown, A. and Tike, S. "The Future of Data Mining" http://www.aston.ac.uk/~golderpa/CS342/grouppages/dssg9/fut.html

[17] Rae-Dupree, Janet: Staff Writer "Dominating With Data - Data Mining Emerges As The Future Of Business Analysis" San Jose Mercury News, 6 October 1997, Section: Business Monday, Page: 1E

[18] Stodder, David "Canary in a Data Mine" Database Programming & Design, vol. 10 no. 11, November 1997, p. 5.

[19] Schroder, Norma "Data Mining Market Share 1997", Dataquest, 8 June 1998, [redacted]

[20] Wilson, Linda "A Cautious Approach to Data Mining" Data Warehouse, December 1997