Role of Data Mining and Its Utilities For Business Intelligence

Najaf Shan Fatima
(Jaipuria Institute of Management)
Ghaziabad

Abstract

In this paper we argue that data mining can make a significant contribution to knowledge management initiatives. We use two studies to show how data mining can make the difference during the knowledge management process. First, we describe how data mining was used as part of the knowledge management initiative in a major company. Currently, banks and other financial institutions are maintaining huge electronic data repositories. Valuable bits of information are embedded in these data repositories. The huge size of these data sources make it impossible for a human analyst to come up with interesting information that will helpful in the decision making process. Most of commercial enterprises have been quick to recognize the value of this concept, as a consequence of which the software market itself for data mining is expected to be in excess of 10 billion USD. In this note, I have discusses broad areas of application, like risk management, portfolio management, trading, customer profiling and customer care, where data mining techniques can be used in banks and other financial institutions to enhance their business performance. We demonstrate how implementation of data mining for generating new knowledge about credit risk improved the company performance. Secondly, we carry out a feasibility study of a knowledge management project in a major Finnish engineering company.
We explore how data mining can improve the quality and responsiveness of reporting in corporate communications unit. The results of the studies allow us to conclude that data mining has strong functional capabilities to support knowledge management initiatives.

1. Introduction

Data mining is a term that covers a broad range of techniques being used in a variety of industries. Due to increased competition for profits and market share in the marketing arena, data mining has become an essential practice for maintaining a competitive edge in every phase of the customer lifecycle. Historically, one form of data mining was also known as "data dredging." This was considered beneath the standards of a good researcher. It implied that a researcher might actually search through data without any specific predetermined hypothesis. Recently, however, this practice has become much more acceptable, mainly because this form of data mining has led to the discovery of valuable nuggets of information. In corporate America, if a process uncovers
information that increases profits, it quickly gains acceptance and respectability.
Another form of data mining began gaining popularity in the marketing arena in the late 1980s and early 1990s. A few cutting edge credit card banks saw a form of data mining, known as data modeling, as a way to enhance acquisition efforts and improve risk management. The high volume of activity and unprecedented growth provided a fertile ground for data modeling to flourish. The successful and profitable use of data modeling paved the way for other types of industries to embrace and leverage these techniques. Today, industries using data modeling techniques for marketing include insurance, retail and investment banking, utilities, telecommunications, catalog, energy, retail, resort, gaming, pharmaceuticals, and the list goes on and on.

2. Data Mining – conceptual over view

Data mining is often set in the broader context of *knowledge discovery in databases,* or KDD. This term originated in the artificial intelligence (AI) research field. The KDD process involves several stages: selecting the target data, preprocessing the data, transforming them if necessary, performing data mining to extract patterns and relationships, and then interpreting and assessing the discovered structures. Once again the precise boundaries of the data mining part of the process are not easy to state; for example, to many people data transformation is an intrinsic part of data mining. In this text we will focus primarily on data mining algorithms rather than the overall process. For example, we will not spend much time discussing data preprocessing issues such as data cleaning, data verification, and defining variables. Instead we focus on the basic principles for modeling data and for constructing algorithmic processes to fit these models to

data.The process of seeking relationships within a data set of seeking accurate, convenient, and useful summary representations of some aspect of the data involves a number of steps:

- Determining the nature and structure of the representation to be used;
- Deciding how to quantify and compare how well different representations fit the data (that is, choosing a "score" function);
- Choosing an algorithmic process to optimize the score function; and deciding what principles of data management are required to implement the algorithms efficiently.

The goal of this text is to discuss these issues in a systematic and detailed manner Figure 1 depicts a generic data mining process.

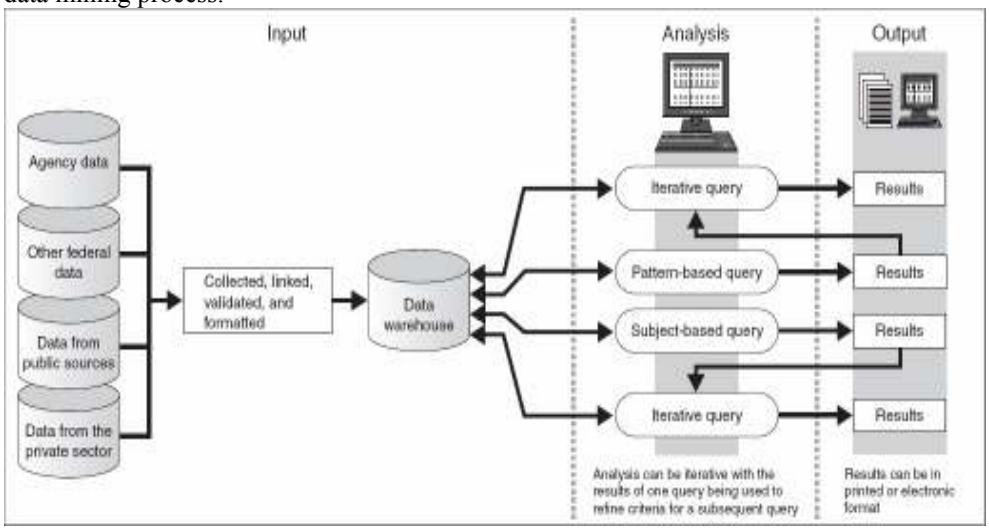

Figure 1: An Overview of the Data Mining Process

3. Tasks Can Be Performed with Data Mining

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

Classification
Estimation
Prediction
Affinity grouping
Clustering
Description and profiling

3.1 Classification

Classification, one of the most common data mining tasks, seems to be a human imperative. In order to understand and communicate about the world, we are constantly classifying, categorizing, and grading. Classification consists of examining the features of a newly presented object and assigning it to one of a predefined set of classes. The objects to be classified are generally represented by records in a database table or a file, and the act of classification consists of adding a new column with a class code of some kind. The classification task is characterized by a well-defined definition of the classes, and a training set consisting of pre classified examples. The task is to build a model of some kind that can be applied to unclassified data in order to classify it.

Examples of classification tasks that have been addressed using the techniques include:

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

3.2 Estimation

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

In practice, estimation is often used to perform a classification task. Acredit card company wishing to sell advertising space in its billing envelopes to a ski boot manufacturer might build a classification model that put all of its cardholders into one of two classes, skier or nonskier. Another approach is to build a model that assigns each cardholder a "propensity to ski score." This might be a value from 0 to 1 indicating the estimated probability that the cardholder is a skier. The classification task now comes down to establishing a threshold score. Anyone with a score greater than or equal to the threshold is classed as a skier, and anyone with a lower score is considered not to be a skier.The estimation approach has the great advantage that the individual records can be rank ordered according to the estimate. To see the importance of this, imagine that the ski boot company has budgeted for a mailing of 500,000 pieces. If the classification approach is used and 1.5 million skiers are identified, then it might simply place the ad in the bills of 500,000 people selected at random from that pool. If, on the other hand, each cardholder has a propensity to ski score, it can send the ad to the 500,000 most likely candidates.

Examples of estimation tasks include:

Estimating the number of children in a family

Estimating a family's total household income

Estimating the lifetime value of a customer

Estimating the probability that someone will respond to a balance transfer solicitation.

3.3 Prediction

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

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

Examples of prediction tasks addressed by the data mining techniques include:

Predicting the size of the balance that will be transferred if a credit card prospect accepts a balance transfer offer

Predicting which customers will leave within the next 6 months

Predicting which telephone subscribers will order a value-added service such as three-way calling or voice mail

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

3.4 Affinity Grouping or Association Rules

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

Affinity grouping can also be used to identify cross-selling opportunities and to design attractive packages or groupings of product and services. Affinity grouping is one simple approach to generating rules from data. If two items, say cat food and kitty litter, occur together frequently enough, we can generate two *association rules*:

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

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

3.5 Clustering

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

developed through training on preclassified examples. In clustering, there are no predefined classes and no examples. The records are grouped together on the basis of self-similarity. It is up to the user to determine what meaning, if any, to attach to the resulting clusters. Clusters of symptoms might indicate different diseases. Clusters of customer attributes might indicate different market segments.
Clustering is often done as a prelude to some other form of data mining or modeling. For example, clustering might be the first step in a market segmentation effort: Instead of trying to come up with a one-size-fits-all rule for "what kind of promotion do customers respond to best," first divide the customer base into clusters or people with similar buying habits, and then ask what kind of promotion works best for each cluster.

3.6 Profiling

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

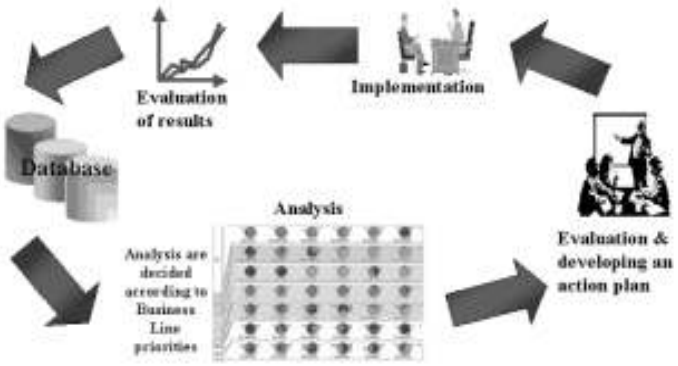

Figure 2. The process of Relational Marketing

4. Why Now?

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

- The data is being produced.
- The data is being warehoused.
- Computing power is affordable.
- Interest in customer relationship management is strong.
- Commercial data mining software products are readily available.
- Let's look at each factor in turn.
- Data Is Being Produced

Data mining makes the most sense when there are large volumes of data. In fact, most data mining algorithms *require* large amounts of data in order to build and train the models that will then be used to perform classification, prediction, estimation, or other data mining tasks. A few industries, including telecommunications and credit cards, have long had an automated, interactive relationship with customers that generated many transaction records, but it is only relatively recently that the automation of everyday life has become so pervasive. Today, the rise of supermarket point of-sale scanners, automatic teller machines, credit and debit cards, paper-view television, online shopping, electronic funds transfer, automated order processing, electronic ticketing, and the like means that data is being produced and collected at unprecedented rates.

5. Data Is Being Warehoused

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

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

The data warehouse, on the other hand, should be designed exclusively for decision support, which can simplify the job of the data miner.

6. Computing Power Is Affordable

Data mining algorithms typically require multiple passes over huge quantities of data. Many are computationally intensive as well. The continuing dramatic decrease in prices for disk, memory, processing power, and I/O bandwidth has brought once-costly techniques that were used only in a few government funded laboratories into the reach of ordinary businesses. The successful introduction of parallel relational database management software by major suppliers such as Oracle, Teradata, and IBM, has brought the power of parallel processing into many corporate data centers for the first time. These parallel database server platforms provide an excellent environment for large-scale data mining.

7. Interest in Customer Relationship Management Is Strong

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

Every Business Is a Service Business

For companies in the service sector, information confers competitive advantage. That is why hotel chains record your preference for a nonsmoking room and car rental companies record your preferred type of car. In addition, companies that have not traditionally thought of themselves as service providers are beginning to think differently. Does an automobile dealer sell cars or transportation?

If the latter, it makes sense for the dealership to offer you a loaner car whenever your own is in the shop, as many now do. Even commodity products can be enhanced with service. A home heating oil company that monitors your usage and delivers oil when you need more, sells a better product than a company that expects you to remember to call to arrange a delivery before your tank runs dry and the pipes freeze. Credit card companies, long-distance providers, airlines, and retailers of all kinds often compete as much or more on service as on price.

Information Is a Product

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

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

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

8. How Data Mining Is Being Used Today

This whirlwind tour of a few interesting applications of data mining is intended to demonstrate the wide applicability of the data mining techniques discussed in this paper These vignettes are intended to convey

something of the excitement of the field and possibly suggest ways that data mining could be profitably employed in your own work.

8.1 A Supermarket Becomes an Information Broker

Thanks to point-of-sale scanners that record every item purchased and loyalty card programs that link those purchases to individual customers, supermarkets are in a position to notice a lot about their customers these days. Safeway was one of the first U.S. supermarket chains to take advantage of this technology to turn itself into an information broker. Safeway purchases address and demographic data directly from its customers by offering them discounts in return for using loyalty cards when they make purchases. In order to obtain the card, shoppers voluntarily divulge personal information of the sort that makes good input for actionable customer insight.

From then on, each time the shopper presents the discount card, his or her transaction history is updated in a data warehouse somewhere. With every trip to the store, shoppers teach the retailer a little more about themselves. The supermarket itself is probably more interested in aggregate patterns (what items sell well together, what should be shelved together) than in the behavior of individual customers. The information gathered on individuals is of great interest to the *manufacturers* of the products that line the stores' aisles. Of course, the store assures the customers that the information thus collected will be kept private and it is. Rather than selling Coca-Cola a list of frequent Pepsi buyers and vice versa, the chain sells *access* to customers who, based on their known buying habits and the data they have supplied, are likely prospects for a particular supplier's product. Safeway charges several cents per name to suppliers who want their coupon or special promotional offer to reach just the right people. Since the coupon redemption also becomes an entry in the shopper's transaction history file, the precise response rate of the targeted group is a matter of record. Furthermore, a particular customer's response or lack thereof to the offer becomes input data for future predictive models. American Express and other charge card suppliers do much the same thing,

selling advertising space in and on their billing envelopes. The price they can charge for space in the envelope is directly tied to their ability to correctly identify people likely to respond to the ad. That is where data mining comes in.

8.2 A Recommendation-Based Business

Virgin Wines sells wine directly to consumers in the United Kingdom through its Web site, www.virginwines.com. New customers are invited to complete a survey, "the wine wizard," when they first visit the site. The wine wizard asks each customer to rate various styles of wines. The ratings are used to create a profile of the customer's tastes. During the course of building the profile, the wine wizard makes some trial recommendations, and the customer has a chance to agree or disagree with them in order to refine the profile. When the wine wizard has been completed, the site knows enough about the customer to start making recommendations. Over time, the site keeps track of what each customer actually buys and uses this information to update his or her customer profile. Customers can update their profiles by redoing the wine wizard at any time. They can also browse through their own past purchases by clicking on the "my cellar" tab. Any wine a customer has ever purchased or rated on the site is in the cellar. Customers may rate or rerate their past purchases at any time, providing still more feedback to the recommendation system. With these recommendations, the web site can offer customers new wines that they should like, emulating the way that the stores like the Wine Cask have built loyal customer relationships.

8.3 Cross-Selling

USAA is an insurance company that markets to active duty and retired military personnel and their families. The company attributes information-based marketing, including data mining, with a doubling of the number of products held by the average customer. USAA keeps detailed records on its customers and uses data mining to predict where they are in their life cycles and what products they are likely to need. Another company that has used data mining to improve its cross-selling ability is Fidelity Investments. Fidelity maintains a data warehouse filled with information on all of its retail customers. This information is used to build data mining models that predict what other Fidelity products are likely to interest each customer. When an existing customer calls Fidelity, the phone representative's screen shows exactly where to lead the conversation.In addition to improving the company's ability to cross-sell, Fidelity's retail marketing data warehouse has allowed the financial services powerhouse to build models of what makes a loyal customer and thereby increase customer retention. Once upon a time, these models caused Fidelity to retain a marginally profitable bill-paying service that would otherwise have been cut. It turned out that

people who used the service were far less likely than the average customer to take their business to a competitor. Cutting the service would
have encouraged a profitable group of loyal customers to shop around.A central tenet of customer relationship management is that it is more profitable to focus on "wallet share" or "customer share," the amount of business you can do with each customer, than on market share. From financial services to heavy manufacturing, innovative companies are using data mining to increase the value of each customer.

8.4 Holding on to Good Customers

Data mining is being used to promote customer retention in any industry where customers are free to change suppliers at little cost and competitors are eager to lure them away. Banks call it attrition. Wireless phone companies call it churn. By any name, it is a big problem. By gaining an understanding of *who* is likely to leave and *why*, a retention plan can be developed that addresses the
right issues and targets the right customers. In a mature market, bringing in a new customer tends to cost more than holding on to an existing one. However, the incentive offered to retain a customer
is often quite expensive. Data mining is the key to figuring out which customers should get the incentive, which customers will stay without the incentive, and which customers should be allowed to walk.

8.5 Weeding out Bad Customers

In many industries, some customers cost more than they are worth. These might be people who consume a lot of customer support resources without buying much. Or, they might be those annoying folks who carry a credit card they rarely use, are sure to pay off the full balance when they do, but must still be mailed a statement every month. Even worse, they might be people who owe you a lot of money when they declare bankruptcy. The same data mining techniques that are used to spot the most valuable customers can also be used to pick out those who should be turned down for a loan, those who should be allowed to wait on hold the longest time, and those who should always be assigned a middle seat near the engine (or is that just our paranoia showing?).

8.6 Revolutionizing an Industry

In 1988, the idea that a credit card issuer's most valuable asset is the information it has about its customers was pretty revolutionary. It was an idea that Richard Fairbank and Nigel Morris shopped around to 25 banks before Signet Banking Corporation decided to give it a try.Signet acquired behavioral data from many sources and used it to build predictive models. Using these models, it launched the highly successful balance transfer program that changed the way the credit card industry works. In 1994, Signet spun off the card operation as Capital One, which is now one of the top 10 credit card issuers. The same aggressive use of data mining technology that fueled such rapid growth is also responsible for keeping Capital One's loan loss rates among the lowest in the industry. Data mining is now at the heart of the marketing strategy of all the major credit card issuers.Credit card divisions may have led the charge of banks into data mining, but other divisions are not far behind. At Wachovia, a large North Carolina-based bank, data mining techniques are used to predict which customers are likely to be moving soon. For most people, moving to a new home in another town
means closing the old bank account and opening a new one, often with a different company. Wachovia set out to improve retention by identifying customers who are about to move and making it easy for them to transfer their business to another Wachovia branch in the new location. Not only has retention improved markedly, but also a profitable relocation business has developed. In addition to setting up a bank account, Wachovia now arranges for gas, electricity, and other services at the new location.

Figure 3 shows general structure of Relational Marketing Activity.

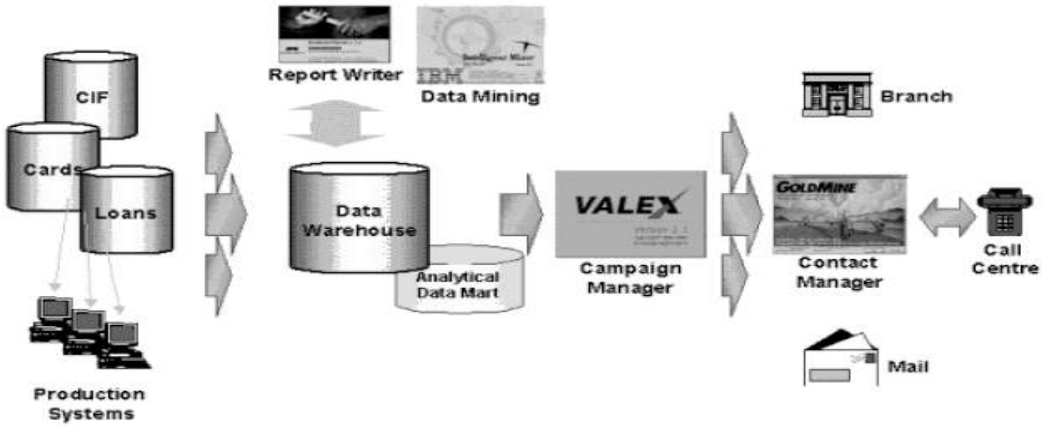

Figure 3.**The Relational Marketing process is supported by a computing infrastructure where many software packages are integrated with the bank's information system**.

9. Methodology

The organisations selected for the survey utilize computers in their daily operations. They were assured that all information given would be treated with the strictest confidence. The questionnaires were posted or sent by email to 100 organisations, and they were given a period of one month to respond. Some of the responses were collected via face-to-face interview with the company operation/management staff of companies. The returned questionnaires were checked for consistency of the answers and for completeness. The data were coded and analysed using the statistical software package, SPSS version 11.
Figure 4 represents the conceptual model used in this study. The model for this study tests certain factors that apply data mining tools to improve the response times of queries. In this model, factors that affect the implementation of data mining tools include end-users (data warehouse administrators or decision makers) involvement and non-end-users involvement. To implement data mining successfully, factors that hinder this need to be resolved by the data warehouse administrators or decision makers.

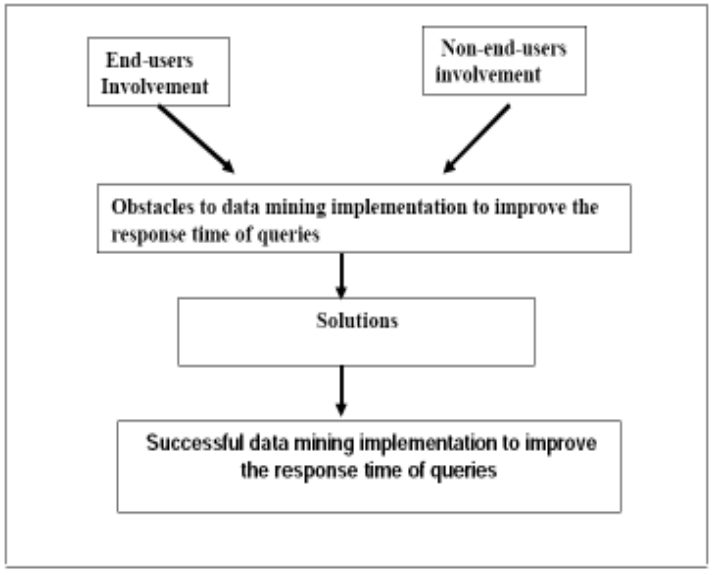

Figure. 4: The Conceptual Model

End-users (data warehouse administrators or decision makers) involvement plays an important role in

the successful implementation of information technology project. Data mining implementation is no exception.
End-users who have been given adequate training on data mining can contribute to its successful implementation. Developing countries, such as Malaysia, rely heavily on data mining experts to give training to end- users. With effective technology transfer and systematic training, data mining can be successfully implemented in the country.
Apart from end-user involvement in data mining implementation, there are other matters that relate to software cost, user-interface and database (DB) issues that affect implementation. To determine the actual reasons to why data mining is not implementation to improve the response time of queries, the fo llowing hypothesis is formulated:

H1. Data mining is not implemented to improve the response time of queries is due more to the end-users related rather non-end users related problems.

Rick (2002) states that within the next five year many businesses will be managing a petabyte (1015) of data, which is equivalent to 250 billion pages of text, or enough to fill 20 million four-drawer filing cabinets.
Business transactions in electronic format are continuously growing. The efficient processing of information is crucial for the business to function well and profitably. SQL processing usually accounts for 60 to 90 percent of the computer resources of a relational database server. Normally, over 60 percent of performance problems in database applications are caused by poorly performing SQL statements, and the performance of at least 30 percent of all SQL statements can be extensively improved. Most data Most data warehouse administrators, when asked, would say that 90 percent or more of data warehouse application performance problems are due to poor SQL performance (Frank & Richard, 2001).
Data warehouse performance improvement through SQL tuning, can result in vast savings by delaying expensive hardware upgrades, avoiding time-consuming (the refore costly) data warehouse redesign or reconfiguration, and improving business productivity (Frank & Richard, 2001).
This research investigated the current tools that can be applied to address the problems relating to the response time of queries. The objective is to find out whether the decision makers or data warehouse administrators know how some of the existing tools help to improve the response time of a query. New and existing data mining techniques were integrated to select relevant attributes, relevant tuples, or both relevant attributes and tuples to form redundant data structure as a complete solution to improve the response time of queries. The next question is, what are the critical success factors? Hence, the following hypothesis is formulated:

H2. Factors that contribute to the success of data mining implementation to improve the response time of queries are more the end-users related rather than non- end-users involvement.

In the survey, the questionnaire developed used questions, which are open ended and having multiplechoice answers and using Likert Scale answers. Data collection has been constituted by the response to the questionnaire and mostly sent by e- mail. Some of the responses were collected via face-to-face interviewing with company operation and management staff.

10. Discussion of Results

The questionnaires were sent to 100 Malaysian business organisations, randomly chosen. Forty-two questionnaires were returned – a response rate of 42%. Eight business organisations, among those that responded, do not have any digital data transmission, but, they utilise computers in their daily operation
All the 42 business organisations, a great percentage (80.95%) use electronic data transmission for business purpose. This is a positive sign indicating that the business community is keen to adopt electronic communication as a major business medium. It has been reported that *e-business revenues will increase from $61 billion in 2001 to $148 billion in 2005* (James, 2001). Sixteen (38.1 %) business organisation indicated their current data warehouse capacity for their decision-support systems to be more than or equal to 1 GB. In the next five years, fifteen (35.71 %) organisations indicated their data warehouse capacity for their decision-support systems to be more than or equal to 1 GB. Based on a recent survey conducted by META group (Dave, 2002) *over 90 percent of Global 2000 companies reported that having less than 10 GB of data.* After 12 months of implementing a data warehouse, 43% of the companies project that the size of the data warehouse will be between 10 GB to 250 GB (Dave, 2002).

Data warehouse performance tuning is an important part of the management and administration of a decision support system. It helps data warehouse administrators or decision makers to improve the response time of a query. Table 2 shows that most responding companies (71.43%) understand the processes of data warehouse performance tuning.

Table 1: Companies and Their Understanding of Data Warehouse Performance Tuning Process

	Yes	No
Understand data warehouse performance tuning process	30 (71.43%)	12 (28.57%)

Table 2, indicates that 59.52% of responding companies, do not use any tool to improve the response time of queries. The data warehouse server takes care of most of the tuning work (Auto-configuring, self-tuning). Most of the responding companies (64.29%) suggested having a better tool for the datawarehouse administrators or decision makers to optimise the performances of their data warehouse.

Most of the responding companies (85.71%) are not sure and do not use any data mining tools in their decision-support system. Only a few companies (14.29%) use any data mining tools in their decisionsupport system. Most of responding companies (95.24%) are not sure and do not use any data mining tools to improve the response time of queries. One of responding company states that they use SQL Server Data Transformation Services (DTS) to improve the response time of queries. As is known, the function of SQL Server DTS are data- manipulation utility services in SQL Server 7.0, and provide import, export, and data-manipulating services between OLE DB, ODBC, and ASCII data stores. DTS is not a data mining tool. This highlights the need to have data mining tools to help the data warehouse administrators or decision makers.

Table 2 Understanding Data Mining Tools

The responding companies do not use any tools to improve the response time of queries	25 (59.52%)
The responding companies use any tools to improve the response time of queries	17 (40.48%)
The responding companies suggest having a better tool for data warehouse administrators or decision makers to optimise the performance of their database	27 (64.29%)
The responding companies think that the current tools good are enough for data warehouse administrators or decision makers to optimise the performance of the database	15 (35.71%)
The responding companies have never used data mining tools in decision-support systems	21 (50%)
The responding companies are not sure whether any data mining tools have been used in decision-support systems	15 (35.71%)
The responding companies never used data mining tools in decision-support systems	6 (14.29%)
The responding companies have never used data mining tools to improve the response time of queries	28 (66.67%)
The responding companies are not sure whether any data mining tools have been used to improve the response time of queries	12 (28.57%)
The responding companies have used data mining tools to improve the response time of queries	2 (4.76%)

Reasons for Not Implementing Data Mining to Improve the Response Time of Queries

Table 3 lists the reason for not implementing data mining to improve the response time of queries. Three reasons have mean values higher than 3: *Lack of required expertise, high software cost* and *lack of knowledge about data mining*. Other reasons are not considered significant.

Reasons	Mean
1. Lack of required expertise	3.44
2. High software cost	3.31
3. Lack of knowledge about data mining	3.06
4. High training cost	2.88
5. It is difficult to use	2.69
6. Difficult to improve the response time of queries	2.56
7. Data Mining only handles the logical level of a data warehouse rather than physical level of a data warehouse	2.56
8. Data warehouse self-tuning takes care of improving the response time of queries	2.31
9. Small data warehouse capacity	2.31
10. Lack of enthusiasm	0.31
11. Advancement in DB technology	0.31

Table 3: Reasons for Not Implementing Data Mining to improve the response time of queries

10.1 Obstacles to Data Mining Implementation to Improve the Response Time of Queries

The respondents were asked to express in their opinions concerning the difficulty in implementing data mining to improve the response time of queries. Table 4 summarizes the data collected.

There are 3 reasons with mean value above 3 and the important reasons given in order of significance are:

1. Lack of required expertise
2. High software cost
3. High training cost

Reasons	Mean
1. Lack of required expertise	3.81
2. High software cost	3.31
3. High training cost	3.31
4. Data warehouse's self tuning takes care of improving the response time of queries	2.88
5. Lack of knowledge about data mining	2.75
6. Not user friendly interface	2.5
7. Lack of enthusiasm	0.31
8. Advancement in technology in DB	0.31

Table 4: Obstacles to data mining implementation to improve the response time of queries

10.2 Factors that Contribute Towards the Success of Data Mining Implementation to Improve the Response Time of Queries

Table 5 lists 9 factors that contribute towards the success of data mining implementation to improve the response time of queries.

Factors	Mean
1. Sufficient knowledge about data mining	3.94
2. Availability of the required expertise	3.69
3. Support from top level management	3.56
4. Low software cost	3.43
5. Low training cost	3.31
6. User-friendly interface	3.25
7. Data mining is able to interact with multiple database platforms	3.25
8. Lack of enthusiasm	0.31
9. Advancement in technology in DB	0.31

Table 5: Factors that contribute towards the success of data mining implementationto improve the response time of queries

Seven (7) factors have a mean value of higher than 3. The following top five factors in the order of significance:

1. Ensure sufficient knowledge about data mining
2. The availability of required expertise
3. Support from top level management
4. Low software cost
5. Low training cost

11.Conclusion of survey

To test hypothesis H1, the reasons were classified into end-users and non-end-users related reasons. The statistical package SPSS version 11 was then used to compute the mean for each group.
The end-users related reasons are:
1. High training cost
2. Lack of required expertise
3. Lack of knowledge about data mining
4. Lack of enthusiasm
The non-end-users related reasons are:
1. High software cost
2. Data warehouse self tuning takes care of improving the response time of queries
3. No user-friendly interface
4. Advancement in DB technology
In Table 6 shows the mean ranking given to end-users related reasons significantly higher than the mean ranking of the non-end-users related reasons. This infers that the reasons why *data mining is not carried out to improve the response time of queries are due more to end-users involvement rather than non-endusers involvement.*

Pair	**Total Mean**	**Mean of Group**	**Mean Difference** End-users related reasons – Non-end-users related reasons
End-users related reasons	10.19	2.55	0.3
Non-end-users related reasons	9	2.25	

Table 6: Paired sample test between end-users and non-end-users related reasons to why data mining is not carried out to improve the response time of queries

The results of the survey showed that there are 9 factors considered important enough to contribute to the success of data mining implementation to improve the response time of queries (Table 5). To test the hypothesis H2, the success factors were classified into end-users related reasons and non-end-users related reasons. SPSS was used to compute the mean for each group. The results are summarised in Table 7.
The end-user related reasons are:
1. Ensure sufficient knowledge about data mining
2. The availability of required expertise
3. Support from top-level management
4. Low training cost
5. Lack of enthusiasm
The non-end-user related reasons are:
1. Data mining is able to interact with multiple database platforms
2. Low software cost
3. User- friendly interface
4. Advancement in DB technology

Pair	**Total Mean**	**Mean of Group**	**Mean Difference** End-users related reasons – Non-end-users related reasons
End-users related reasons	14.81	2.96	0.4
Non-end-users related reasons	10.25	2.56	

Table 7: Paired sample test between end-users related reasons and non-end-users related reasons success factors of data mining implementation to improve the response time of queries

On the average, end-user related reasons received higher ranking than the non-end-user related reasons.

This infers that success factors of data mining to improve the response time of queries are more endusers related reasons rather than non-end-users related reasons. The conceptual model can now be refined and filled in with the answers for the research que stions. The result is depicted in Figure 5

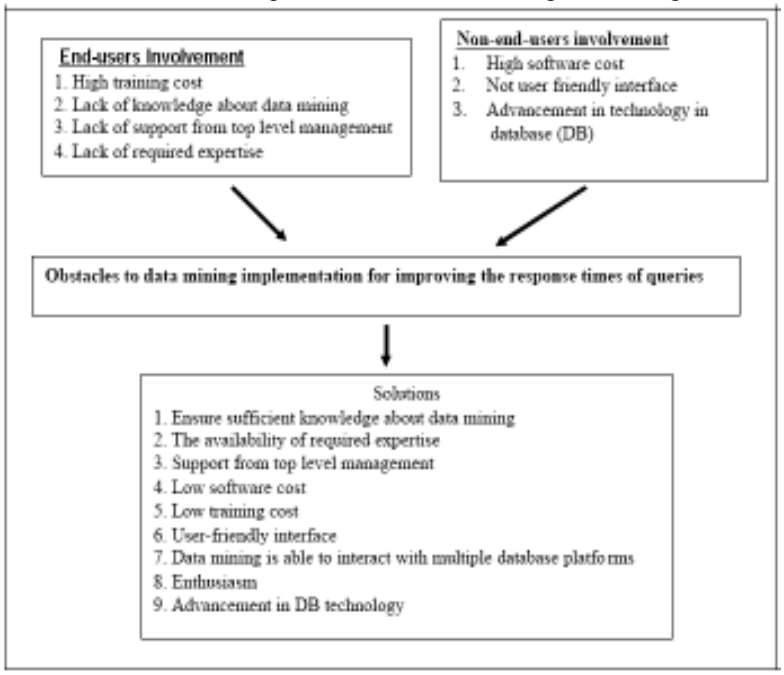

Figure 5. The Refined Conceptual Model

References:

[1] Web site: www.gao.gov/fraudnet/fraudnet.htm

□□□□□ODM literature and white papers (including this one) can be found at
http://www.oracle.com/technology/products/bi/odm/index.html
(or simply Google™ search on "Oracle Data Mining")

[3] Bank Negara Malaysia. (2002).Consolidation of the Banking Sector.
Available from World Wide Web:
http://www.bnm.gov.my/en/News/releases.asp?yr=2002&sid=0128a
Last modified : 22 March, 2002.

[4] Chaudhuri, S., Dayal U., Ganti, Venkatesh. (2001). Database Technology for Decision Support Systems. *Computer*. IEEE Computer Society.

[5] Frank I. and Richard T. (2001, August). SQL Optimization for the Data Warehousing Environment.
Lecco Technology.
Available from World Wide Web: http://www.hkcs.org.hk/dbwp1805.doc

[6] *Banking Software: Data Mining & Banking Intelligence*, retrieved 3rd January, 2006 from ,
http://www.stratinfotech.com/banking_software/banking_software_business_intelligence_data _mining.htm

[7] J. M. Zytkow and W. Klösgen, *Handbook of Data Mining and Knowledge Discovery*. New York: Oxford, 2002.

[8] Herb Edelstein, *Building Profitable Customer Relationships With Data Mining*, SPSS,
www.spss.fi/PDF/Building_proftable_cust_relations_DM.pdf