

## **DATA MINING IN BANKING AND FINANCE: A NOTE FOR BANKERS**

**Rajanish Dass**

**Indian Institute of Management Ahmedabad**

[rajanish@iimahd.ernet.in](mailto:rajanish@iimahd.ernet.in)

### **Abstract**

*Currently, huge electronic data repositories are being maintained by banks and other financial institutions. 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 (or patterns) that will help in the decision making process. A number 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. This note is intended for bankers, who would like to get aware of the possible applications of data mining to enhance the performance of some of their core business processes. In this note, the author 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.*

*Keywords: Data Mining, Banks, Financial Institutions, Risk Management, Portfolio Management, Trading, CRM, Customer Profiling*

## **DATA MINING IN BANKING AND FINANCE: A NOTE FOR BANKERS**

**Rajanish Dass**

**Indian Institute of Management Ahmedabad**

[rajanish@iimahd.ernet.in](mailto:rajanish@iimahd.ernet.in)

As knowledge is becoming more and more synonymous to wealth creation and as a strategy plan for competing in the market place can be no better than the information on which it is based, the importance of knowledge and information in today's business can never be seen as an exogenous factor to the business. Organizations and individuals having access to the right information at the right moment, have greater chances of being successful in the epoch of globalization and cut-throat competition.

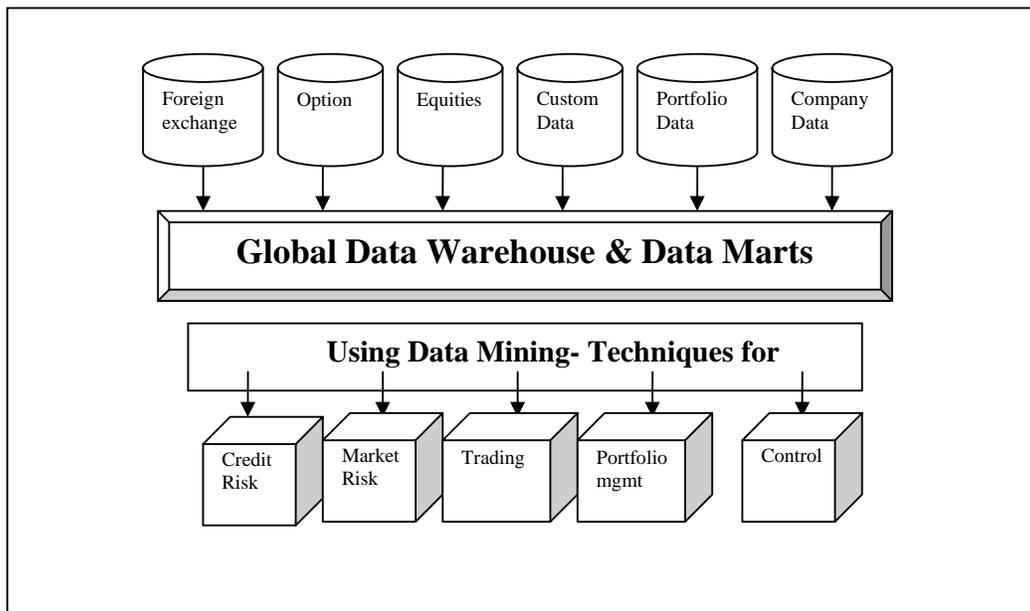
Currently, huge electronic data repositories are being maintained by banks and other financial institutions across the globe. Valuable bits of information are embedded in these data repositories. The huge size of these data sources make it impossible for a human analyst to come up with interesting information that will help in the decision making process. A number 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.

Business Intelligence focuses on discovering knowledge from various electronic data repositories, both internal and external, to support better decision making. Data mining techniques become important for this knowledge discovery from databases. In recent years, business intelligence systems have played pivotal roles in helping organizations to fine tune business goals such as improving customer retention, market penetration, profitability and efficiency. In most cases, these insights are driven by analyses of historical data.

Global competitions, dynamic markets, and rapidly decreasing cycles of technological innovation provide important challenges for the banking and finance industry. Worldwide just-in-time availability of information allows enterprises to improve their

flexibility. In financial institutions considerable developments in information technology have led to huge demand for continuous analysis of resulting data.

Data mining can contribute to solving business problems in banking and finance by finding patterns, causalities, and correlations in business information and market prices that are not immediately apparent to managers because the volume data is too large or is generated too quickly to screen by experts. The managers of the banks may go a step further to find the sequences, episodes and periodicity of the transaction behaviour of their customers which may help them in actually better segmenting, targeting, acquiring, retaining and maintaining a profitable customer base. Business Intelligence and data mining techniques can also help them in identifying various classes of customers and come up with a class based product and/or pricing approach that may garner better revenue management as well.



**Figure 1.** *The use of Data Mining Technique is a Global and Firm wide challenge for financial business. Firm-wide data source can be used through data mining for different business areas.*

The broad categories of application of Data Mining and Business Intelligence techniques in the banking and financial industry vertical may be viewed as follows<sup>1</sup>:

- **Risk Management**

Managing and measurement of risk is at the core of every financial institution. Today's major challenge in the banking and insurance world is therefore the implementation of risk management systems in order to identify, measure, and control business exposure. Here credit and market risk present the central challenge, one can observe a major change in the area of how to measure and deal with them, based on the advent of advanced database and data mining technology. ( Other types of risk is also available in the banking and finance i.e., liquidity risk, operational risk, or concentration risk. )

Today, integrated measurement of different kinds of risk (i.e., market and credit risk) is moving into focus. These all are based on models representing single financial instruments or risk factors, their behaviour, and their interaction with overall market, making this field highly important topic of research.

- **Financial Market Risk**

For single financial instruments, that is, stock indices, interest rates, or currencies, market risk measurement is based on models depending on a set of underlying risk factor, such as interest rates, stock indices, or economic development. One is interested in a functional form between instrument price or risk and underlying risk factors as well as in functional dependency of the risk factors itself.

Today different market risk measurement approaches exist. All of them rely on models representing single instrument, their behaviour and interaction with overall market. Many of this can only be built by using various data mining

---

<sup>1</sup> J. M. Zytow and W. Klösgen, *Handbook of Data Mining and Knowledge Discovery*. New York: Oxford, 2002.

techniques on the proprietary portfolio data, since data is not publicly available and needs consistent supervision.

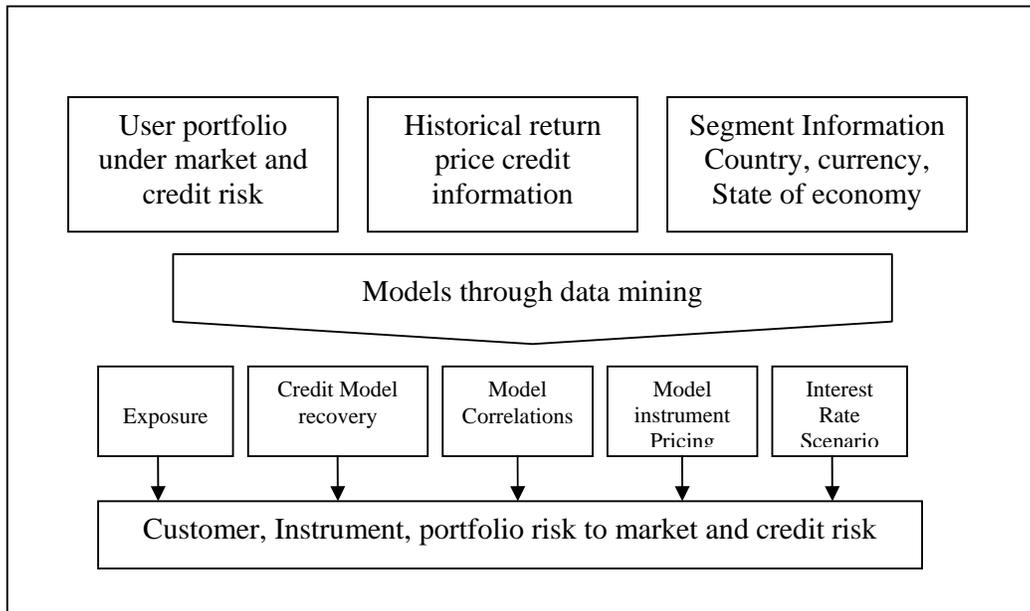
- **Credit Risk**

Credit risk assessment is key component in the process of commercial lending. Without it the lender would be unable to make an objective judgement of whether to lend to the prospective borrower, or if how much charge for the loan. Credit risk management can be classified into two basic groups:

- a. Credit scoring/credit rating. Assignment of a customer or a product to risk level.(i.e., credit approval)
- b. Behaviour scoring/credit rating migration analysis. Valuation of a customer's or product's probability of a change in risk level within a given time.(i.e., default rate volatility)

In commercial lending, risk assessment is usually an attempt to quantify the risk of loss to the lender when making a particular lending decision. Here credit risk can quantify by the changes of value of a credit product or of a whole credit customer portfolio, which is based on change in the instrument's rating, the default probability, and recovery rate of the instrument in case of default. Further diversification effects influence the result on a portfolio level. Thus a major part of implementation and care of credit risk management system will be a typical data mining problem: the modelling of the credit instrument's value through the default probabilities, rating migrations, and recovery rates.

Three major approaches exist to model credit risk on the transaction level: accounting analytic approaches, statistical prediction and option theoretic approaches. Since large amount of information about client exist in financial business, an adequate way to build such models is to use their own database and data mining techniques, fitting models to the business needs and the business current credit portfolio.



**Figure 2.** *Using Data Mining technique for customer, financial instrument, portfolio risk to market and credit risk measurement*

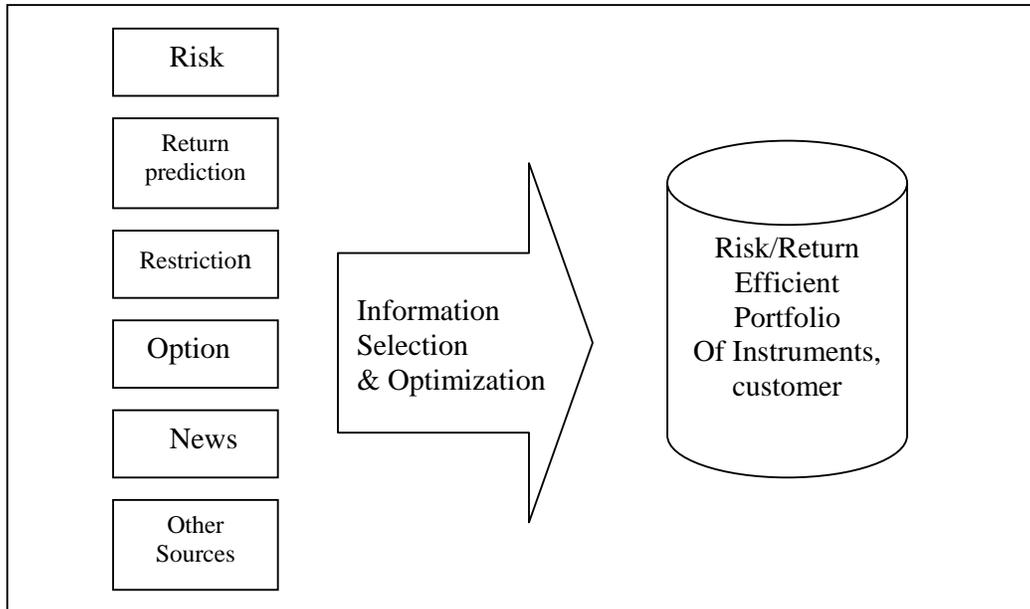
- **Portfolio Management**

Risk measurement approaches on an aggregated portfolio level quantify the risk of a set of instrument or customer including diversification effects. On the other hand, forecasting models give an induction of the expected return or price of a financial instrument. Both make it possible to manage firm wide portfolio actively in a risk/return efficient manner. The application of modern risk theory is therefore within portfolio theory, an important part of portfolio management.

With the data mining and optimization techniques investors are able to allocate capital across trading activities to maximise profit or minimise risk. This feature supports the ability to generate trade recommendations and portfolio structuring from user supplied profit and risk requirement.

With data mining techniques it is possible to provide extensive scenario analysis capabilities concerning expected asset prices or returns and the risk involved. With

this functionality, what if simulations of varying market conditions e.g. interest rate and exchange rate changes) can be run to assess impact on the value and/or risk associated with portfolio, business unit counterparty, or trading desk. Various scenario results can be regarded by considering actual market conditions. Profit and loss analyses allow users to access an asset class, region, counterparty, or custom sub portfolio can be benchmarked against common international benchmarks.



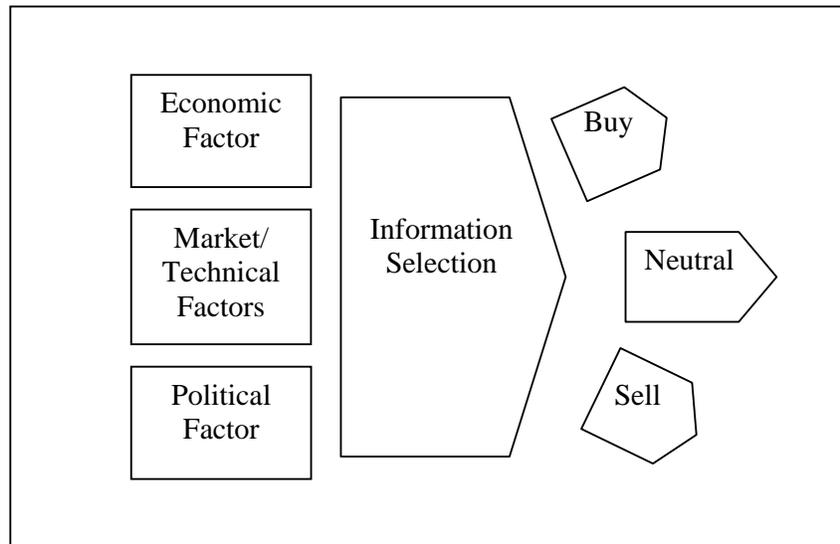
**Figure 3.** *The management of an instrumental portfolio is based on all reachable -information, that is risk, scenario and predicted credit ratings, but also on news and other information sources.*

### ▪ Trading

For the last few years a major topic of research has been the building of quantitative trading tools using data mining methods based on past data as input to predict short-term movements of important currencies, interest rates, or equities.

The goal of this technique is to spot times when markets are cheap or expensive by identifying the factor that are important in determining market returns. The trading system examines the relationship between relevant information and piece of financial assets, and gives you buy or sell recommendations when they suspect an under or overvaluation. Thus, even if some traders find the data mining approach too

mechanical or too risky to be used systematically, they may want to use it selectively as further opinion.



**Figure 4.** Market participants examine the relationship between relevant information and the price of financial assets, and buy or sell securities when they suspect an under or over valuation

Trading is based on the idea of predicting short term movements in the price/value of a product (currency/equity/interest rate etc.). With a reasonable guesstimate in place one may trade the product if he/she thinks it is going to be overvalued or undervalued in the coming future. Trading traditionally is done based on the instinct of the trader. If he/she thinks the product is not priced properly he/she may sell/buy it. This instinct is usually based on past experience and some analysis based on market conditions. However, the number of factors that even the most expert of traders can account for are limited. Hence, quite often these predictions fail.

The price of a financial asset is influenced by a variety of factors which can be broadly classified as economic, political and market factors. Participants in a market observe the relation between these factors and the price of an asset, account for the current value of these factors and predict the future values to finally arrive at the future value of the asset and trade accordingly. Quite often by the time a trained eye detects these favourable factors, many others may have discovered the opportunity, decreasing the possible revenues otherwise. Also these factors in turn may be related to several other factors making prediction difficult.

Data mining techniques are used to discover hidden knowledge, unknown patterns and new rules from large data sets, which maybe useful for a variety of decision making activity. With the increasing economic globalization and improvements in information technology, large amounts of financial data are being generated and stored. These can be subjected to data mining techniques to discover hidden patterns and obtain predictions for trends in the future and the behaviour of the financial markets. With the immediacy offered by data mining, latest data can be mined to obtain crucial information at the earliest. This in turn would result in an improved market place responsiveness and awareness leading to reduced costs and increased revenue.

Advancements made in technology have enabled to create faster and better prediction systems. These systems are based on a combination of data mining techniques and artificial intelligence methods like Case Based Reasoning (CBR) and Neural Networks (NN). A combination of such a forecasting system together with a good trading strategy offers tremendous opportunities for massive returns.

The value of a financial asset is dependent on both macroeconomic and microeconomic variables and this data is available in a variety of disparate formats. Data mining comes in here since it helps discover information and hidden patterns from large data sets and data sources in different formats. NN and CBR techniques can be applied extensively for predicting these financial variables.

NN are characterized by learning capabilities and the ability to improve performance over time. Also NN can generalize i.e. recognize new objects which may be similar but not exactly identical to previous objects. NN with their ability to derive meaning from imprecise data can be used to detect patterns which are otherwise too complex to be detected by humans. NN act as experts in the area that they have been trained to work in. these can be used to provide predictions for new situations and work in real time. Thus, historic data available about financial markets and the various variables can be used to train NN to simulate the market. Based on entry of current values of market variables, the NN can predict the status in the coming day and may be used to give a buy/sell recommendation.

CBR methodology is based on reasoning from past performances. It uses a large repository of data stored as cases which would include all the market variables in this case. When a new case is fed in (in the form of a case containing the concerned variables), the CBR algorithm predicts the performance/result of this case based on the cases it has in its repository. Data mining techniques can be used to detect hidden patterns in these cases which may then be used for further decision making. CBR methods can be used in real time which makes analysis really quick and helps in real time decision making resulting in immediate profits.

Thus data mining and business intelligence (CBR and NN) techniques may be used in conjunction in financial markets to predict market behaviour and obtain patterned behaviour to influence decision making.

#### ▪ Customer Profiling and Customer Relationship Management

Banks have many and huge databases containing transactional and other details of its customers. Valuable business information can be extracted from these data stores. But it is unfeasible to support analysis and decision making using traditional query languages; because human analysis breaks down with volume and dimensionality. Traditional statistical methods do not have the capacity and scale to analyse these data, and hence modern data mining methodologies and tools are increasingly being used for decision making process not only in banking and financial institutions, but across the industries.

Customer profiling is a data mining process that builds customer profiles of different groups from the company's existing customer database. The information obtained from this process can be used for different purposes, such as understanding business performance, making new marketing initiatives, market segmentation, risk analysis and revising company customer policies. The advantage of data mining is that it can handle large amounts of data and learn inherent structures and patterns in data. It can generate rules and models that are useful in enabling decisions that can be applied to future cases.

Customer Behaviour Modeling (CBM) or customer profiling is a tool to predict the future value of an individual and the risk category to which he belongs to based on his demographic characteristics, life-style and previous behaviour. This helps to focus on customer retention. The two important facts that have important implication in selecting customer profiling methods are:

- Profiling information can consist of many variables (or dozens of them).
- Majority of them are categorical variables (or non-numeric variables or nominal variables).

Customer profiling is to characterize features of special customer groups. Many data mining techniques search profiles of special customer groups systematically using Artificial Intelligence techniques. They generate accurate profiles based on beam search and incremental learning techniques.

Customer profiling also uses many predictive modeling methods. Predictive modeling techniques applicable can be categorized into two broad approaches. They depend on the type of predicted information or variables, also called target variables. If the type of predicted values is categorical, classification techniques is preferred to be used.

#### ***Classification Methods:***

In this approach, risk levels are organized into two categories based on past default history. For example, customers with past default history can be classified into "risky" group, whereas the rest are placed as "safe" group. Using this categorization information as target of prediction, Decision Tree and Rule Induction techniques can be used to build models that can predict default risk levels of new loan applications.

#### ***Value Prediction Methods:***

In this method, for example, instead of classifying new loan applications, it attempts to predict expected default amounts for new loan applications. The predicted values are numeric and thus it requires modeling techniques that can take numerical data as

target (or predicted) variables. Neural Network and regression are used for this purpose. The most common data mining methods used for customer profiling are:

- Clustering (descriptive)
- Classification (predictive) and regression (predictive)
- Association rule discovery (descriptive) and sequential pattern discovery (predictive)

In CRM, data mining is frequently used to assign a score to a particular customer or prospect indicating the likelihood that the individual will behave in a particular way. For example, a score could measure the propensity to respond to a particular insurance or credit card offer or to switch to a competitor's product.

Data mining can be useful in all the three phases of a customer relationship-cycle: customer acquisition, increasing value of the customer and customer retention. For example, a typical banking firm let say sends 1 million direct mails for credit card customer acquisition. Past researches<sup>2</sup> have shown that typically 6% of such target customers respond to these direct mails. Banks use their credit risk models to classify these respondents in good credit risk and bad credit risk classes. The proportion of good credit risk respondents is only 16% out of the total respondents. So, as net result, roughly only 1% of the total targeted customers are converted into the credit card customers through direct mailing. Seeing the huge cost and effort involved in such marketing process, data mining techniques can significantly improve the customer conversion rate by more focused marketing. Using a predictive test model using decision tree techniques like CHAID (Chi-squared Automatic Interaction Detection), CART (Classification And Regression Trees), Quest and C5.0; it can be analyzed which customers are more probable to respond. And using this with the risk model using techniques like neural network can help build a test model.

The way data mining can actually be built into the CRM application is determined by the nature of customer interaction. The customer interaction could be inbound (when the customer contacts the firm) or outbound (when the firm contacts customers). The deployment requirements are quite different. Outbound interactions such as direct

---

<sup>2</sup> "Building Profitable Customer Relations with Data Mining", Herb Edelstein

mail campaign involve the firm selecting the people whom to be mailed by applying the test model to the customer database. In other outbound campaigns like advertising, the profile of good prospects shown by the test model needs to be matched to the profile of the people the advertisement would reach.

For inbound transactions such as telephone or internet order, the application must respond in real time. Therefore the data mining model is embedded in the application and actively recommends an action. In either case, one of the key issues in applying a model to new data set is the transformations that are made in building the model. The ease with which these changes are embedded in the model determines the productivity of deploying these tools.

- **Marketing and customer care**

Because high competitions in the finance industry, intelligent business decisions in marketing are more important than ever for better customer targeting, acquisition, retention and customer relationship. There is a need for customer care and marketing strategies to be in place for the success and survival of the business. It is possible with the help of data mining and predictive analytics to make such strategies.

Financial institutions are finding it more difficult to locate new previously unsolicited buyers, and as a result they are implementing aggressive marketing program to acquire new customer from their competitors. The uncertainties of the buyer make planning of new services and media usage almost impossible. The classical solution is to apply subjective human expert knowledge as rules of thumb. Until recently, replacing the human expert by computer technology has been difficult.

An interesting tool available in marketing and financial institution is analysis of client's data. This allows analysis and calculation of key indicators that help bank to identify factors that affected customer's demand in the past and customer' need in the future.

Information about the customer's personal data can also give indications that affect future demand. In case of analysis of retail debtors and small corporations, marketing tasks will typically include factors about the customer himself, his credit record and rating made by external rating agencies.

With the advent of data mining and business intelligence tools it has become possible for banks to strengthen their customer acquisition by direct marketing and establish multi-channel contacts, to improve customer development by cross selling and up selling of products, and to increase customer retention by behaviour management. It is possible for the banks to use the data available to retain its best customers and to identify opportunities to sell them additional services. The profiling of all the valuable accounts can be done and the top most say 5-10 % can be assigned to *Relationship Managers*, whose job will be to identify new selling opportunities with these customers.

It is also possible to bundle various offers to meet the need of the valued customers. Data mining can also help the banks in customizing the various promotional offers. For example the direct mails can be customized as per the segment of the account holders in the bank. It is also possible for the banks to find out the *problem* customers who can be defaulters in the future, from their past payment records and the profile and the data patterns that are available. This can also help the banks in adjusting the relationship with these customers so that the loss in future is kept to its minimum.

Data mining can improve the response rates in the direct mail campaigns as the time required to classify the customers will be reduced, this in turn will increase the revenues, improve the sales force efficiency from the *target group*. Data mining helps the banks to optimize their portfolio of services, delivery channels. A record of past transactions can give useful insight to the bank and different locations /branches of same branch can also follow some patterns that when noticed can be used as past records to learn from and base the future actions upon.

Data Mining techniques can be of immense help to the banks and financial institutions in this arena for better targeting and acquiring new customers, fraud detection in real-

time, providing segment based products for better targeting the customers, analysis of the customers' purchase patterns over time for better retention and relationship, detection of emerging trends to take proactive stance in a highly competitive market adding a lot more value to existing products and services and launching of new product and service bundles.

## Reference:

- *Marketing Buzz*, retrieved 4th January, 2006 from, <http://www.cindiaifoline/fmcg/stma/st26.html>
- *Banking Software: Data Mining & Banking Intelligence*, retrieved 3rd January, 2006 from , [http://www.stratinfotech.com/banking\\_software/banking\\_software\\_business\\_intelligence\\_data\\_mining.htm](http://www.stratinfotech.com/banking_software/banking_software_business_intelligence_data_mining.htm)
- R. Savitha, *From Mine to Shine* , retrieved 6th January, 2006 from, <http://www.blonnet.com/ew/2003/07/30/stories/2003073000250100.htm>
- Petra Hunziker, Andreas Maier, Alex Nippe, Markus Tresch, Douglas Weers, and Peter Zemp, *Data Mining at a major bank: Lessons from a large marketing application* retrieved 5th January, 2006 from <http://homepage.sunrise.ch/homepage/pzemp/info/pkdd98.pdf>
- Michal Meltzer, *Using Data Mining on the road to be successful part III*, published in October 2004, retrieved 2nd January, 2006 from [http://www.dmreview.com/editorial/newsletter\\_article.cfm?nl=bireport&articleId=1011392&issue=20082](http://www.dmreview.com/editorial/newsletter_article.cfm?nl=bireport&articleId=1011392&issue=20082)
- Fuchs, Gabriel and Zwahlen, Martin, *What's so special about insurance anyway?*, published in DM Review Magazine, August 2003 issue, retrieved 5<sup>th</sup> January, 2006 from [http://www.dmreview.com/article\\_sub.cfm?articleId=7157](http://www.dmreview.com/article_sub.cfm?articleId=7157)
- Mehta, Radhika, *Future Perfect*, retrieved 4th January, 2006 from [http://www.tata.com/tata\\_aig\\_general/articles/20030920.htm](http://www.tata.com/tata_aig_general/articles/20030920.htm)
- Insurance Information Warehouse (IIW) General Information Manual *Transforming Insurance Information into Business Intelligence*, retrieved 6th January, 2006 from [http://www-03.ibm.com/industries/financialservices/doc/content/bin/fss\\_IIW\\_GIM\\_29June2004.pdf](http://www-03.ibm.com/industries/financialservices/doc/content/bin/fss_IIW_GIM_29June2004.pdf)
- Christos, Stergiou and Dimitrios, Siganos, *Neural Networks*, retrieved January 6, 2006 from [http://www.doc.ic.ac.uk/~nd/surprise\\_96/journal/vol4/cs11/report.html#Introduction%20to%20neural%20networks](http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Introduction%20to%20neural%20networks)
- Chun, Se-Hak and Kim, Steven, *Data mining or financial prediction and trading: application to single and multiple markets* (2003)
- J. M. Zytkow and W. Klösgen, *Handbook of Data Mining and Knowledge Discovery*. New York: Oxford, 2002.
- Herb Edelstein, *Building Profitable Customer Relationships With Data Mining*, SPSS, [www.spss.fi/PDF/Building\\_profitable\\_cust\\_relations\\_DM.pdf](http://www.spss.fi/PDF/Building_profitable_cust_relations_DM.pdf)