

Importance of Data Mining in Banking Sectors

SK. GHOUSE BASHA

Associate Professor, Dept of CSE, SSN Engineering College, Ongole, AP, India, E-mail: smile2ghouse@gmail.com.

Abstract: In today's globalization and cut throat competition the banks are struggling to gain a competitive edge over each other. Apart from execution of business processes, the creation of knowledge base and its utilization for the benefit of the bank is becoming a strategy tool to compete. In recent years the ability to generate, capture and store data has increased enormously. The information contained in this data can be very important. The wide availability of huge amounts of data and the need for transforming such data into knowledge encourage IT industry to use data mining. The banking industry around the world has undergone a tremendous change in the way business is conducted. The banking industry has started realizing the need of the techniques like data mining which can help them to compete in the market. Leading banks are using Data Mining (DM) tools for customer segmentation and profitability, credit scoring and approval, predicting payment default, marketing, detecting fraudulent transactions, etc. This paper provides an overview of the concept of DM and highlights the applications of data mining to enhance the performance of some of the core business processes in banking industry.

Keywords: Banking industry, Data Mining, Fraud Detection, MIS, TBC.

I. INTRODUCTION

In the financial services industry throughout the world, the traditional face-to-face customer contacts are being replaced by electronic points of contact to reduce the time and cost of processing an application for various products and ultimately improve the financial performance. The computerization of financial operations, use of internet and automated software's has completely changed the basic concept of business and the way the business operations are being carried out. The banking sector is not an exception to it. It has also witnessed a tremendous change in the way the banking operations are carried out [1]. Since 1990's the whole concept of banking has been shifted to centralized databases, online transactions and ATM's all over the world, which has made banking system technically strong and more customer oriented[1]. Data might be one of the most valuable resources of any bank but only if it knows how to expose valuable knowledge hidden in raw data. Data mining allows extracting knowledge from the historical data, and predicting outcomes of future situations. It helps optimize business decisions, increase the value of each customer and communication, and improve customer satisfaction [2]. The amount of data collected by banks has grown rapidly in recent years. Existing statistical data analysis techniques find it difficult to manage with the large volumes of data now available.

This explosive growth has lead to the need for new data analysis techniques and tools in order to find the information hidden in this data. Banking is an area where vast amounts of data are collected. This data can be generated from bank account transactions, loan applications, loan repayments, credit card repayments, etc. It is assumed that valuable

information on the financial profile of customers is hidden within these massive operational databases and this information can be used to improve the performance of the bank [3]. In the beginning Total Branch Computerization (TBC) software packages being used at various branch for daily transactions, Designing the new MIS or restructuring the existing ones would not be possible by just replacing the existing Total Branch Computerization packages. The solution to this problem is to implement the concept of data warehousing and data mining [1].

II. DATA MINING

Data mining refers to extracting knowledge from large amounts of data. The data may be spatial data, multimedia data, time series data, text data and web data. Data mining is the process of extraction of interesting, nontrivial, implicit, previously unknown and potentially useful patterns or knowledge from huge amounts of data. It is the set of activities used to find new, hidden or unexpected patterns in data or unusual patterns in data. Using information contained within data warehouse, data mining can often provide answers to questions about an organization that a decision maker has previously not thought to ask [4].

- Which products should be promoted to a particular customer? – Targeted Marketing
- What is the probability that a certain customer will leave for a competitor? – Customer Relationship Management
- What is the appropriate medical diagnosis for this patient? – Bio medical
- What is the likelihood that a certain customer will default or pay back a loan? – Banking

- Which products are bought most often together? – Market Basket Analysis
- How to identify fraudulent users in telecommunication industry? – Fraudulent pattern analysis.

These types of questions can be answered quickly and easily if the information hidden among the huge amount of data in the databases can be located and utilized. Data mining is often referred to as „analytical intelligence“. Several recent trends have increased the interest in data mining because of decreasing cost of data storage and the increasing ease of collecting data. With greater data storage capabilities and decreasing costs, data mining has offered organizations a new way of doing business. Data mining can help organizations better understand their business, be able to better serve their customers, and increase the effectiveness of the organization in the long run [4]. Today, banks are realizing the various advantages of data mining. It is a valuable tool by which banks can identify potentially useful information from the large amounts of data. This can help banks to gain a clear advantage over its competitors. Data mining can help banks in better understanding of the vast volume of data collected by the CRM systems.

III. DATA MINING ALGORITHM AND TECHNIQUES

There are several data mining techniques and algorithms have been developed and used in data mining like association, classification, clustering, prediction and sequential patterns, Regression, Neural Networks etc. [5]. We will briefly examine those data mining techniques.

A. Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large [6]. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we make the software that can learn how to classify the data items into groups [5]. Fraud detection and credit risk applications are particularly well suited to this type of analysis. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. For a fraud detection application, this would include complete records of both fake and valid activities determined on a record-by-record basis. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper judgment. The algorithm then encodes these parameters into a model called a classifier [6]. Types of classification models:

- Classification by decision tree induction
- Bayesian Classification

- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

B. Association

Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction [5]. Association and correlation is usually used to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis [6]. For example, the association technique is used in market basket analysis to identify what products that customers frequently purchase together. Based on this data businesses can have corresponding marketing campaign to sell more products to make more profit [5]. The various types of associations include:

- Multilevel association rule.
- Multidimensional association rule
- Quantitative association rule
- Direct association rule.
- Indirect association rule.

C. Clustering

Clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique [5]. Clustering technique also defines the classes and put objects in them, while in classification objects are assigned into predefined classes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification [7]. For Example: The customer of a given geographic location and of a particular job profile demand a particular set of services, like in banking sector the customers from the service class always demand for the policy which ensures more security as they are not intending to take risks, similarly the same set of service class people in rural areas have a the preferences for some particular brands which may vary from their counterparts in urban areas. This information will help the organization in cross-selling their products, The bank's customer service representatives can be equipped with customer profiles enriched by data mining that help them to identify which products and services are most relevant to callers. This technique will help the management in finding the solution of 80/20 principle of marketing, which says: Twenty per cent of your customers will provide you with 80 per cent of your profits, then problem is to identify those 20 % and the techniques of clustering will help in achieving the same [1]. Types of clustering methods

- Partitioning Methods
- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods [6]

D. Prediction

The prediction as its name implied is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables [5]. Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict. Unfortunately, many real-world problems are not simply prediction. For instance, sales volumes, stock prices, and product failure rates are all very difficult to predict because they may depend on complex interactions of multiple predictor variables. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values [6]. Types of regression methods

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

E. Sequential Patterns

Sequential patterns analysis is one of data mining technique that seeks to discover similar patterns in data transaction over a business period. The uncover patterns are used for further business analysis to recognize relationships among data [5].

IV. DATA MINING APPLICATIONS IN BANKING

The banking industry across the world has undergone tremendous changes in the way the business is conducted. With the recent implementation, greater acceptance and usage of „electronic“ banking, the capturing of transactional data has become easier and, simultaneously, the volume of such data has grown considerably. It is beyond human capability to analyse this huge amount of raw data and to effectively transform the data into useful knowledge for the organization [2]. Data Mining can help by contributing in solving business problems by finding patterns, associations and correlations which are hidden in the business information stored in the data bases [1]. By using data mining to analyse patterns and trends, bank executives can predict, with increased accuracy, how customers will react to adjustments in interest rates, which customers will be likely to accept new product offers, which customers will be at a higher risk for defaulting on a loan, and how to make customer relationships more profitable [2]. The banking industry is widely recognizing the importance of the information it has about its customers. Undoubtedly, it has among the richest and largest pool of customer information, covering customer demographics, transactional data, credit cards usage pattern, and so on. As banking is in the service industry, the task of maintaining a strong and effective CRM is a critical issue. To do this, banks need to invest their resources to better understand their existing and prospective customers.

By using suitable data mining tools, banks can subsequently offer „tailor-made“ products and services to those customers [2]. There are numerous areas in which data mining can be used in the banking industry, which include customer segmentation and profitability, credit scoring and approval, predicting payment default, marketing, detecting fraudulent transactions, cash management and forecasting operations, optimizing stock portfolios, and ranking investments. In addition, banks may use data mining to identify their most profitable credit card customers or high-risk loan applicants. To help bank to retain credit card customers, data mining is used. By analyzing the past data, data mining can help banks to predict customers that likely to change their credit card affiliation so they can plan and launch different special offers to retain those customers. Credit card spending by customer groups can be identified by using data mining. Following are some examples of how the banking industry has been effectively utilizing data mining in these areas.

A. Marketing

One of the most widely used areas of data mining for the banking industry is marketing. The bank's marketing department can use data mining to analyse customer databases. Data mining carry various analyses on collected data to determine the consumer behavior with reference to product, price and distribution channel. The reaction of the customers for the existing and new products can also be known based on which banks will try to promote the product, improve quality of products and service and gain competitive advantage. Bank analysts can also analyze the past trends, determine the present demand and forecast the customer behavior of various products and services in order to grab more business opportunities and anticipate behavior patterns. Data mining technique also helps to identify profitable customers from non-profitable ones [8]. The data mining techniques can be used to determine that how customers will react to adjustments in interest rates, the risk profile of a customer segment for defaulting on loans [9].

B. Risk Management

Data mining is widely used for risk management in the banking industry. Bank executives need to know whether the customers they are dealing with are reliable or not. Offering new customers credit cards, extending existing customers lines of credit, and approving loans can be risky decisions for banks if they do not know anything about their customers [2]. Banks provide loan to its customers by verifying the various details relating to the loan such as amount of loan, lending rate, repayment period, type of property mortgaged, demography, income and credit history of the borrower. Customers with bank for longer periods, with high income groups are likely to get loans very easily. Even though, banks are cautious while providing loan, there are chances for loan defaults by customers. Data mining technique helps to distinguish borrowers who repay loans promptly from those who don't [8]. Bank executives by using Data mining technique can also analyze the behaviour and reliability of the customers while selling credit cards too. It also helps to

analyze whether the customer will make prompt or delay payment if the credit cards are sold to them [8]. Credit scoring, in fact, was one of the earliest financial risk management tools developed. Credit scoring can be valuable to lenders in the banking industry when making lending decisions. Data mining can also derive the credit behaviour of individual borrowers with installment, mortgage and credit card loans, using characteristics such as credit history, length of employment and length of residency. A score is thus produced that allows a lender to evaluate the customer and decide whether the person is a good candidate for a loan, or if there is a high risk of default. By knowing what the chances of default are for a customer, the bank is in a better position to reduce the risks [2].

C. Fraud Detection

Another popular area where data mining can be used in the banking industry is in fraud detection. Being able to detect fraudulent actions is an increasing concern for many businesses; and with the help of data mining more fraudulent actions are being detected and reported. Two different approaches have been developed by financial institutions to detect fraud patterns. In the first approach, a bank taps the data warehouse of a third party and use data mining programs to identify fraud patterns. The bank can then cross-reference those patterns with its own database for signs of internal trouble. In the second approach, fraud pattern identification is based strictly on the bank's own internal information. Most of the banks are using a „hybrid“ approach [2]. One system that has been successful in detecting fraud is Falcon's „fraud assessment system“. It is used by nine of the top ten credit card issuing banks. The data mining techniques will help the organization to focus on the ways and means of analyzing the customer data in order to identify the patterns that can lead to frauds [10].

D. Customer Relationship Management

In the era of cut throat competition the customer is considered as the king. 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 [11]. Customer acquisition and retention are very important concerns for any industry, especially the banking industry [2]. Today customers have wide range of products and services provided by different banks. Hence, banks have to cater the needs of the customer by providing such products and services which they prefer. This will result in customer loyalty and customer retention. Data mining techniques helps to analyze the customers who are loyal from those who shift to other banks for better services. If the customer is shifting from his bank to another, reasons for such shifting and the last transaction performed before shifting can be known which will help the banks to perform better and retain its customers [8].

V. CONCLUSION

Data mining is a tool used to extract important information from existing data and enable better decision-making throughout the banking and retail industries. They use data

warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analyzed and the information that is captured is used throughout the organization to support decision-making. Data Mining techniques can be very helpful to the banks for better targeting and acquiring new customers, fraud detection in real time, providing segment based products, analysis of the customers' purchase patterns over time for better retention and relationship. Those banks that have realized the usefulness of data mining and are in the process of building a data mining environment for their decision-making process will obtain huge benefit and derive considerable competitive advantage in future.

VI. REFERENCES

- [1]Vivek Bhambri "Application of Data Mining in Banking Sector", International Journal of ComputerScience and Technology Vol. 2, Issue 2, June 2011
- [2]Dr. Madan Lal Bhasin, "Data Mining: A Competitive Tool in the Banking and Retail Industries", The Chartered Accountant October 2006
- [3]R. I. Scott S. Svinterikou C. Tjortjis J. A. Keane "Experiences of using Data Mining in a Banking Application."
- [4]P.Sundari, Dr.K.Thangadurai "An Empirical Study on Data Mining Applications", Global Journal ofComputer Science and Technology, Vol. 10 Issue 5 Ver. 1.0 July 2010.
- [5]DataMiningTechniques,<http://www.dataminingtechniques.net/>
- [6]Bharati M. Ramageri, "DATA MINING TECHNIQUES AND APPLICATIONS", IndianJournal of Computer Science and Engineering Vol. 1 No. 4
- [7]Hillol Kargupta, Anupam Joshi, Krishnamoorthy Siva Kumar, Yelena Yesha, "Data Mining: Next Generation Challenges and Future Directions", Publishers: Prentice-Hall of India, Private Limited, 2005.
- [8]B. Desai and Anita Desai, "The Role of Data mining in Banking Sector", IBA Bulletin, 2004.
- [9]S.S.Kaptan, "New Concepts in Banking", Sarup andSons, Edition, 2002
- [10]S. S. Kaptan, N S Chobey, "Indian Banking in Electronic Era", Sarup and Sons, Edition 2002.
- [11]Rajanish Dass, "Data Mining in Banking and Finance: A Note for Bankers", Indian Institute of Management Ahmadabad.