Application of Data Mining in Banking Sector

Vivek Bhambri
Dept. of Computer Sciences, Desh Bhagat Institute of Management and Computer Sciences, Mandi Gobindgarh, Punjab, India

Abstract
In the era of globalization and cut throat competition the organizations today are striving 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 organization is becoming a strategy tool to compete. The organizations and individuals having right access to the right information at the right moment of time will be the one to rule the system. In spite of having ever growing data bases the problem is that the banks fail to fully capitalize the true benefits which can be gained from this great wealth of information. The banking sector has started realizing the need of the techniques like data mining which can help them to compete in the market. This paper highlights the perspective applications of data mining to enhance the performance of some of the core business processes in banking sector.

Keywords
Data Mining, Banking Sector, Risk Management, CRM.

I. Introduction

The computerization of financial operations, connectivity through World Wide Web and the support of 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. 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. In the present day environment, the huge amount of electronic data is being maintained by banks around the globe. The huge size of these data bases makes it impossible for the organizations to analyze these data bases and to retrieve useful information as per the need of the decision makers [3,5]. Since 1980's the banking sector is incorporating the concept of Management Information System, through which banks are generating various kinds of reports, which are then presented and analyzed for the decision making with in the organization. However these reports available in the summarized form can be used by the governing authorities [2]. While dealing with banking sector, which itself is an information intensive industry, is quite cumbersome task. The banks at present generate reports from the periodic paper reports and the statements submitted by various constitute units. Such reports have a high degree of error, due to data being recorded and interpreted by various parties at various levels [2]. Moreover the Total Branch Computerization (TBC) software packages being used at various branch levels are transaction oriented, as these were designed keeping day to day transactions in mind. Designing the new MIS or restructuring the existing ones would not be possible by just replacing the existing Total Branch Computerization packages. The solution seems to be in incorporating the concept of data warehousing and data mining. Due to the vast expansion of the horizons of the data and its multivariate uses, the organizations and the individuals are feeling a need for some centralized data management and retrieval system. The centralization of the data is required basically for better processing and in turn facilitating the user access and analysis.

II. Data Mining

Data Mining is the process of extracting knowledge hidden from large volumes of raw data. The knowledge must be new, not obvious, and one must be able to use it. Data mining has been defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data [1]. It is "the science of extracting useful information from large databases" [6]. Data mining is one of the tasks in the process of knowledge discovery from the database [10]. Fig. 1 shows the process of knowledge discovery. The steps involved in Knowledge discovery are [7,10]:

1. Data Selection The data relevant to the analysis is decided and retrieved from the various data locations.
2. Data Preprocessing: In this stage the process of data cleaning and data integration is done.
- Data Cleaning: It is also known as data cleansing; in this phase noise data and irrelevant data are removed from the collected data.
- Data Integration: In this stage, multiple data sources, often heterogeneous, are combined in a common source.
3. Data Transformation: In this phase the selected data is transformed into forms appropriate for the mining procedure.
4. Data Mining: It is the crucial step in which clever techniques are applied to extract potentially useful patterns. The decision is made about the data mining technique to be used.
5. Interpretation and Evaluation: In this step, interesting patterns representing knowledge are identified based on given measures. The discovered knowledge is visually presented to the user. This essential step uses visualization techniques to help users understand.

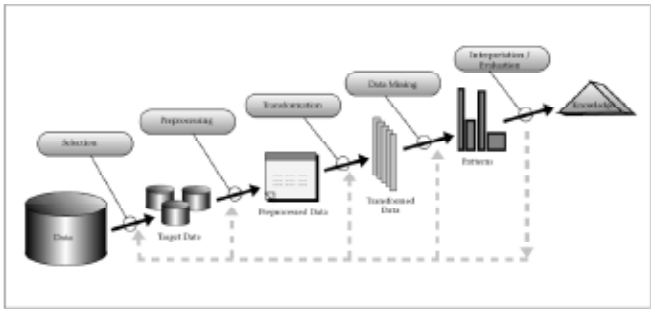

Fig. 1: Knowledge Discovery Process

III. Data Mining Techniques:

The various techniques of data mining are:

A. Association

Association and correlation is usually to find frequently used data items in the large data sets. It is the technique of finding patterns where one event is connected to another event. This type of findings help businesses to make certain decisions regarding pricing, selling and to design the strategies for marketing, such as catalogue design, cross marketing and customer shopping behavior analysis [8]. However the number of possible Association Rules

for a given dataset is generally very large and a high proportion of the rules are usually of little value. The various types of associations include [7]:
- Multilevel association rule.
- Multidimensional association rule
- Quantitative association rule
- Direct association rule.
- Indirect association rule.

B. Clustering

Clustering can be said as identification of similar classes of objects. This is the technique of combining the transactions with similar behavior into one group, or the customers with same set of queries or transactions into one group. Classification approach can also be used as effective mean of distinguishing groups. So clustering can be used as preprocessing approach for attribute subset selection and classification [1]. 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, like wise the same set of service class people in rural areas have a the preferences for some particular brands which may differ from their counterparts in urban areas. This information will help the organization in cross-selling their products, Instead of mass pitching a certain "hot" product, 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.

C. Forecasting

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 [8]. 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 [1,8]. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values. This technique of data mining will help in discovering patterns from which one can make reasonable predictions

D. 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. 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 [8,9]. 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 fraudulent and valid activities determined on a record-by-record basis.

IV. Applications of Data Mining in Banking Sector:

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.

What Customer Data the industry needs to explore & Why ?

1. What is the profile, taste and preferences, attitude of the customer and what is the purchasing behavior of the customer since the time he/she is with the bank? (Used to Cross sell the products).
2. What transactions does a customer do before shifting to a competitor? (To prevent shifting of customers)
3. Which products are often purchased together by the customers of which particular profile? (For target marketing)
4. What patterns in credit transactions lead to fraud? (To detect and deter fraud)
5. What is the profile of a high-risk borrower? (To prevent defaults, bad loans, and improve screening)
6. What services and benefits would current customers likely desire? (To increase loyalty and customer retention)
7. Identifying the customers who are getting all types of services from your company? (Identifying 'Loyal' Customers)

The banks who have realized the importance of data mining are in the process of reaping huge profits and considerable competitive advantage. According to the regulations given by Reserve Bank of India, the banks have to Provide Off-site Monitoring Surveillance (OSMOS) reports on regular basis in electronic format only and Regulatory requirement of filing of statutory returns such as the one under Section 42 of the Reserve Bank of India Act, 1934 for working out Cash Reserve Ratio (CRR) and Statutory Liquidity Ratio (SLR) obligations in electronic format [2]. According to the Committee formed by Reserve Bank of India Headed by Dr. A. Vasudevan to go through the details of this topic, gave his report on 17th July,1999, the committee highlighted that by the use of data mining techniques, data available at various computer systems can be accessed and by a combination of techniques like classification, clustering, segmentation, association rules, sequencing, decision tree various ALM reports such as Statement of Structural Liquidity, Statement of Interest Rate Sensitivity etc. or accounting reports like Balance Sheet and Profit & Loss Account can be generated instantaneously for any desired period/ date [2]. Trends can be analyzed and predicted with the availability of historical data and the data warehouse assures that everyone is using the same data at the same level of extraction, which eliminates conflicting analytical results and arguments over the source and quality of data used for analysis. In short, data warehouse enables information processing to be done in a credible, efficient manner. The Committee recognizes the need for data warehouses and data mining both at the individual bank level and at industry level [1]. The implication of adopting such technology in a bank would be as under [2]:

- All transactions captured at the branch level would get consolidated at a central location. Such a central location could be called the Data Warehouse of the concerned bank. For this to happen, one of the requirements would be to establish connectivity between the branches on the one hand and the Data Warehouse platform on the other.
- For banks with large number of branches, it may not be desirable to consolidate the transaction details at one place only. It can be decentralized
- By way of data mining techniques, data available at various computer systems can be accessed and by a combination of

data mining techniques various decisions can be made.

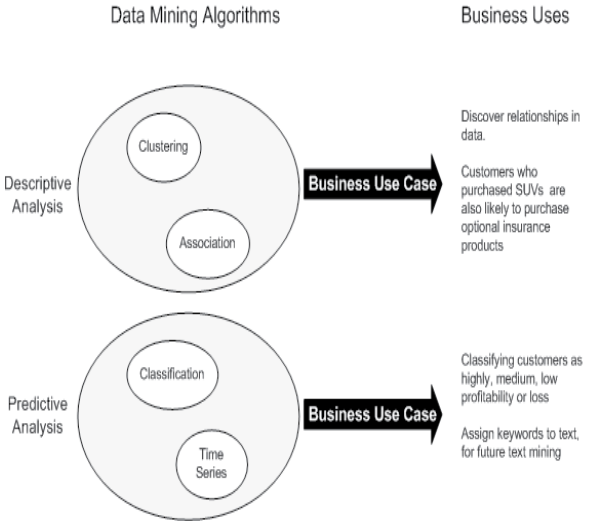

Fig. 2: Usage of Data Mining Algorithms in Analysis.

The broad categories of application of Data Mining and Business Intelligence Techniques in the banking and financial industry vertical may be viewed as follows:

A. Customer Relationship Management

In the era of cut throat competition the customer is considered as the king and it's the customer only who is ruling the whole show. The concept of selling a product to the customer is outdated and obsolete, now the objective is to reach to the heart of the customer and hence to develop a sense of belongingness for the organization. The huge data bases of various organizations are storing billions of data items about the customers. 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 [5]. Data mining technique can be used to create customer profiling to group the like minded customers in to one group and hence they can be dealt accordingly [8]. The information collected can be used for different purposes like making new marketing initiatives, market segmentation, risk analysis and revising company customer policies according to the need of the customers [9]. The profiling is usually done on the basis of demographic characteristics, life style and previous transactional behavior of a particular customer. Customer profiling is to characterize features of special customer groups [10].

B. Marketing

As we are already aware of cut throat competition prevailing in the market in almost all areas, and banking sector is not an exception to it. The marketing and customer care goes hand in hand. Know Your Customer (KYC) is the buzzword these days. 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. Moreover the uncertain behavior of the customer is making this task more tedious. 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 [6,8]. Data mining techniques will help in making customer oriented strategies for their customers in various categories. The data mining techniques can be used to determine that how customers will react to adjustments in interest rates, which customers will be likely to accept new product offers, the risk profile of a customer segment for defaulting on loans, etc. [4,9]. The reaction of the customer for the existing and new products can be recorded, according to which the future strategies can be designed. They can also use the data mining techniques for cross selling. 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 [5].

C. Risk Management

The customer and the banks while dealing with each other will always try to cover the risk factor. To identify, quantify and control the risk factor is always an area of concern for every business organization. In commercial lending, risk assessment is usually an attempt to quantify the risk of loss to the lender while making a particular lending decision. Data mining technique helps to distinguish borrowers who repay loans promptly from those who don't [8]. It also helps to predict when the borrower is at fault, whether providing loan to a particular customer will result in bad loans etc. Such techniques come under the category of credit risk, where we wish to check the behavior of the prospective customers [10]. Bank executives by using Data mining technique can also analyze the behavior and reliability of the customers while selling credit cards too.

D. Fraud Detection

While dealing with banks, the customers and the banks have the chances of falling an easy prey to the frauds. So both the parties wish to be secure while dealing with each other. The data mining techniques can help them to detect and hence prevent frauds. 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 [4,8].

V. Software Support

Keeping in mind the usefulness and applicability of data mining techniques in various sectors, the software development companies have come up with various applications, which can automate the task of data mining. Some such softwares are:

STATISTICA Data Miner, A venture of StatSoft worldwide, is a revolutionary product in the data mining applications. It enables financial institutions to Detect patterns of fraud; Identify causes of risk; create sophisticated and automated models of risk, Segment and predict behavior of homogeneous groups of customers, Uncover hidden correlations between different indicators,.

11Ants Analytics Ltd is a venture backed company located in Hamilton, New Zealand. 11Ants Analytics is committed to making advanced data mining accessible to non-technical users. They have built incredibly powerful data mining software which is deceptively simple to use.

Data Mining with SAS® Enterprise Miner: SAS data mining software helps customers to: detect fraud; anticipate resource demands, increase acquisitions, curb customer attrition

VI. Present Industry Status

The banks in Indian and abroad have started using the techniques of data mining. Chase Manhattan Bank in New York, was facing a financial crunch mainly due to constant decrease in the customer base, then the bank used the techniques of data mining to analyze customer profiles to use them for their benefits and hence chalked out the strategy for the survival and succeed in its attempt [4].

Data mining is also being used by Fleet Bank, Boston, to identify the best candidates for mutual fund offerings [4].

In India banks like ICICI, IDBI, Citi bank, HDFC and PNB has started reaping the benefits of data mining techniques [11]. Citibank is a leader in this area. It has built a data-profile of its customers for a decade and over. Citibank is using data mining results in association with Hutchison Max in Mumbai and Airtel in Delhi. Likewise, HDFC Bank is working with all of its mobile service providers - Hutchison Max, BPL Mobile, Tata Cellular, RPG, Airtel, Aircell, Cellphone and Command. ICICI Bank has followed suit with Hutchison Max, Airtel and Command in Calcutta [11]. `Data mining helps to increase sales by targeting the right customers and to make the right offers to customers. Banks, who have their ears to the ground regarding their customer's tastes and preferences, gather a lot of data.what data mining does is that it sifts through all the voluminous data and ekes out a pattern, which enables the bank to personalize its communication towards the customer as much as possible [11]. With data mining, banks can get a better understanding what drives the customer relationship.

VII. Conclusion

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 approach in a highly competitive market adding a lot more value to existing products and services and launching of new product and service bundles. Data mining has wide application domain almost in every industry where the data is generated that's why data mining is considered one of the most important frontiers in database and information systems and one of the most promising interdisciplinary developments in Information Technology.

References

[1] 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.

[2] Dr. A.Vasudevan, "Report of the Committee on Technology Up gradation in the Banking Sector", Constituted by Reserve Bank of India, Chairman of Committee, 1999

[3] S.R. Mittal, Report of Committee on Internet Banking (2001), Constituted by Reserve bank of India, Chairman of the Committee

[4] Dr.Madan Lal Bhasin, "Data Mining:A Competitive Tool in the Banking and Retail Industries", The Chartered Accountant October ,2006.

[5] Rajanish Dass, "Data Mining in Banking and Finance: A Note for Bankers", Indian Institute of Management Ahmadabad.

[6] D. Muraleedharan, "Modern Banking: Theory and Practice", PHI Learning private Limited, 2009.

[7] Bharati M. Ramager, "Data Mining Techniques And Applications", International Journal of Computer Science and Engineering.

[8] S. S. Kaptan, N S Chobey, "Indian Banking in Electronic Era", Sarup and Sons, Edition 2002.

[9] S.S.Kaptan, "New Concepts in Banking", Sarup and Sons, Edition, 2002

[10] S.P. Deshpande, Dr. V.M. Thakare, "Data Mining System And Applications: A Review".

[11] Mudit Saxena, Vice-President (Retail Marketing), HDFC Bank, Press Statement, Reported by Leena Baliga in India Express Newspaper on 9th Nov., 2000.

Vivek Bhambri received his MCA degree from Punjab Technical University in 2003 and has secured University Position and did his M.Phil from Periyar University, Selam. He is working as Assistant Professor and Head of Computer Science Department at Desh Bhagat Institute of Management and Computer Sciences, Mandi Gobindgrah. He has 4 research publications in International and National Conferences to his credit. His areas of interest include Data Mining and Computer Architecture.