A Purpose of Data Mining in Banking Sector

Jimmy S Mehta[1]
Computer Science & Engineering
Sipna College of Engineering.
Amravati, India

Dr. A D Gawande[2]
Computer Science & Engineering
Sipna COET
Amravati, India

Abstract: Banking systems collect huge amounts of data on day to day basis, be it customer information, transaction details, risk profiles, credit card details, credit limit and collateral details, compliance and Anti Money Laundering (AML) related information, trade finance data, SWIFT and telex messages. Thousands of decisions are taken in a bank daily. These decisions include credit decisions, default decisions, relationship start up, investment decisions, AML and Illegal financing related. One needs to depend on various reports and drill down tools provided by the banking systems to arrive at these critical decisions. But this is a manual process and is error prone and time consuming due to large volume of transactional and historical data. Interesting patterns and knowledge can be mined from this huge volume of data that in turn can be used for this decision making process. This article explores and reviews various data mining techniques that can be applied in banking areas. It provides an overview of data mining techniques and procedures. It also provides an insight into how these techniques can be used in banking areas to make the decision making process easier and productive.

Keywords: data mining, banking, unstructured data, default detection, customer classification, AML.

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. 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 within the organization. However these reports available in the summarized form can be used by the governing authorities.

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. 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.

Jimmy et al.,

International Journal of Advance Research in Computer Science and Management Studies
Volume 3, Issue 3, March 2015 pg. 407-411

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.

➢ 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

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. 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 AND TECHNIQUES

The various techniques of data mining are:

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. 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:

• Multilevel association rule.

• Multidimensional association rule

• Quantitative association rule

• Direct association rule.

• Indirect association rule

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. For Example: The customer of a given geographic location and of a particular job profile demand 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, likewise 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.

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. 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. This technique of data mining will help in discovering patterns from which one can make reasonable predictions

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. 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.
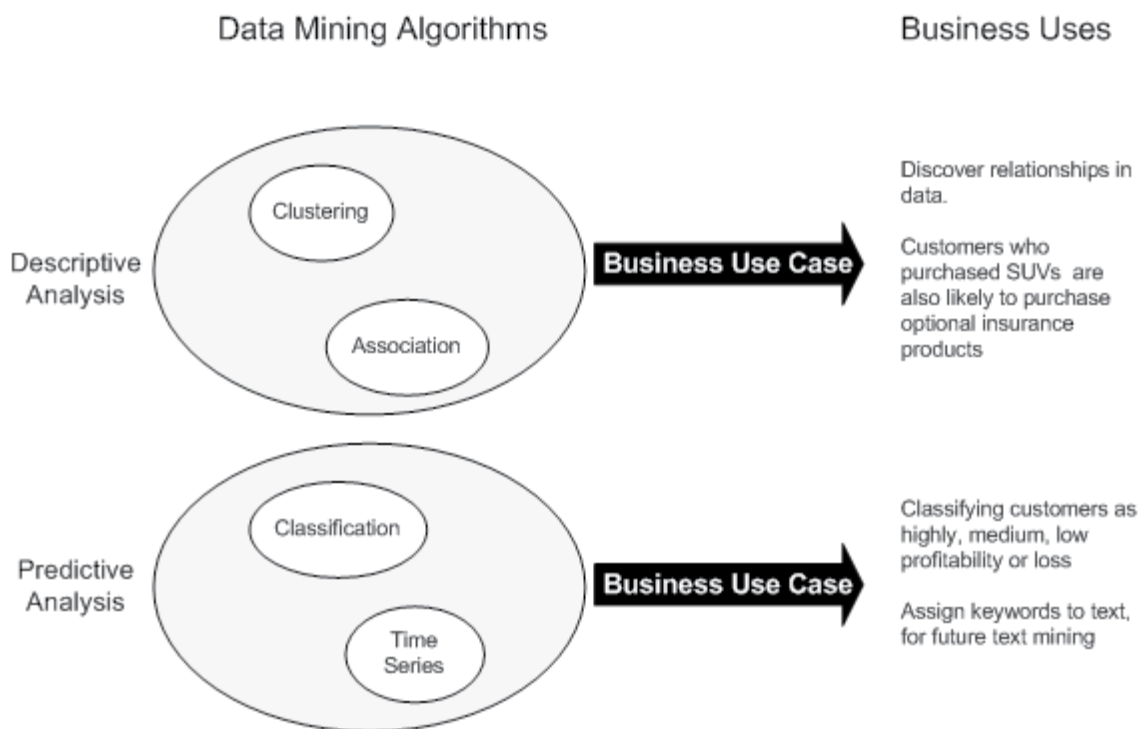

Fig. 1(Data Mining Techniques)

IV. DATA MINING APPLICATIONS IN BANKING

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 analyses 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. 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.

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. 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, and 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. Bank executives by using Data mining technique can also analyze the behavior 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. 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 behavior 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.

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. 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.

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. Customer

acquisition and retention are very important concerns for any industry, especially the banking industry. 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.

V. CONCLUSION

Data mining is a process to extract knowledge from existing data. It is used as a tool in banking and finance in general to discover useful information from the operational and historical data to enable better decision-making. It is an interdisciplinary field, confluence of Statistics, Database technology, Information science, Machine learning and Visualization. It involves steps that include data selection, data integration, data transformation, data mining, pattern evaluation, knowledge presentation. Banks use data mining in various application areas like marketing, fraud detection, risk management, money laundering detection and investment banking. The patterns detected help the bank to forecast future events that can help in its decision-making processes. More and more banks are investing in data mining technologies to be more competitive.

References

1. Sudhakar M & Dr. C. V. K Reddy "Application Areas of Data Mining in Indian Retail Banking Sector" Double Blind Peer Reviewed International Research Journal Volume 14 Issue 5 Version 1.0 Year 2014
2. Vivek Bhambri "Application of Data Mining in Banking Sector" www. i j c s t. c o m IJCST Vol. 2, Iss ue 2, June 2011
3. Aburrous, M., M. A. Hossain and K. Dahal and F.Thabtah, 2010. Intelligent phishing detection system for e-bankingusing fuzzy data mining. Expert Syst. Appli., 37:7913-7921.DOI:0.1016/j.eswa. 2010.04.044.

AUTHOR(S) PROFILE

Mr. Jimmy S Mehta, received the B.E.degree in Computer Science & Engineering from H.V.P.M's College Of Engineering andTechnology, Amravati in 2014. He is currently perusing Master's Degree in Computer Science and Engineering from Sipna College of Engineering and Technology, Amravati.

Dr. A D Gawande, received the PHD Degree from Amravati University, Amravati in year 2011. Currently working as professor & Head Dept of CSE at Sipna COET Amravati. His area of interest are Information storage, Data recovery, Image processing & Signal processing.