

Data Mining: A Tool for Enhancing Business Process in Banking Sector

Dr.R.Mahammad Shafi, Porandla Srinivas

Abstract— Significant shifts in the business environment, economic volatility, changing customer and staff expectations, and the adoption of new technology make it increasingly challenging for banks to navigate technology strategy alternatives and prioritize technology investments. The banking industry around the world has undergone a tremendous change in the way business is conducted. 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 Data Mining.

Data might be one of the most valuable assets of any corporation, but only if it knows how to reveal valuable knowledge hidden in raw data. Data mining allows extracting diamonds of 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. Data mining is the process of extracting previously unknown information, typically in the form of patterns and associations, from large databases. Today, organizations are realizing the numerous advantages that come with data mining. It is a valuable tool, by identifying potentially useful information from the large amounts of data collected. An organization can gain a clear advantage over its competitors.

The banking sector consists of public sector, private sector and foreign banks, apart from smaller regional and cooperative banks. In the market, various IT-based banking products, services and solutions are available. The most common of them are Phone Banking; ATM facility; Credit, Debit and Smart Cards; Internet Banking & Mobile Banking; SWIFT Network & INFINET Network; connectivity of bank branches to facilitate anywhere banking.

Index Terms— Credit Card, Debit Card, INFINET Network Internet Banking, Phone Banking, Smart Card, SWIFT Network.

1 INTRODUCTION

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” [3]. Data mining is one of the tasks in the process of knowledge discovery from the database [5]. Fig. 1 shows the process of knowledge discovery. The steps involved in Knowledge discovery are [4, 5]:

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.
3. Data Cleaning: It is also known as data cleansing; in this phase noise data and irrelevant data are removed from the

collected data.

4. Data Integration: In this stage, multiple data sources, often heterogeneous, are combined in a common source.
5. Data Transformation: In this phase the selected data is transformed into forms appropriate for the mining procedure.
6. 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.
7. 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.

-
- Dr.R.Mahammad Shafi is currently working as a professor in Sree Vidyaniketan Engineering College, Rangampet, India.
E-mail: rmdshafi@gmail.com
 - Porandla Srinivas is currently pursuing Ph.D. in Computer Science engineering in University of Allahabad, India
E-mail: porandla.srinivas@gmail.com

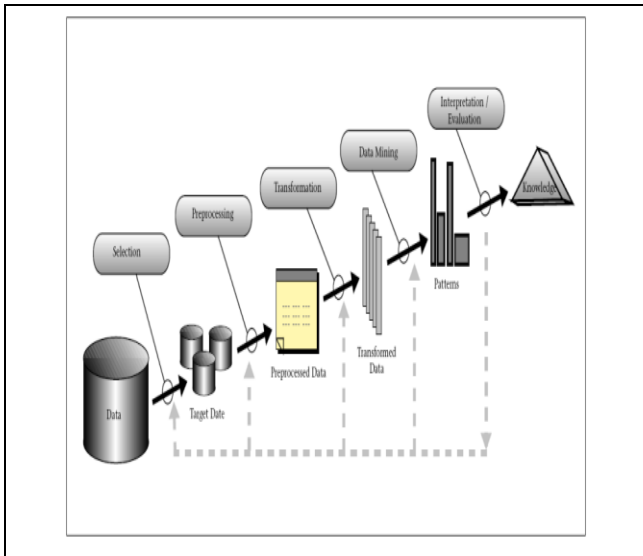


FIG: KNOWLEDGE DISCOVERY PROCESS

2 DATA MINING OPERATIONS

Data mining operations are classified in different ways. A few categorizations of data mining operations are: clustering/segmentation, visualization, predictive modelling, link analysis, deviation detection, dependency modelling, and summarization. With most analysis tools, users know what questions they want to ask before using the tool. With data mining, users do not necessarily need to know what questions to ask; the software uncovers patterns for the users. Data visualization tools allow users to 'see' patterns in the data. These tools allow users to build quite complex, three-dimensional representations of the data.

Many of the methodologies that are commonly described as forming part of a standard data mining tool kit have been developed either by the machine-learning community or by statistical data analysts. In fact, it is really difficult to specify any particular technique as being unique to the data mining community. Therefore, any listing of data analysis methodologies as data mining techniques would appear 'subjective' and somewhat arbitrary. Some widely used techniques in data mining include artificial neural networks, genetic algorithms, K-nearest neighbour method, decision trees, and data reduction.

3 DATA MINING IN THE BANKING INDUSTRY

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 organisation. The enormous amount of data that banks have been collecting over the years can greatly influence the success of data mining efforts. 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. 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.

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, optimising 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. There is, therefore, a need to build an analytical capability to address the above-stated issues and data mining attempts to provide the answer.

Following are some examples of how the banking industry has been effectively utilizing data mining in these areas.

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 and develop statistically sound profiles of individual customer preferences for products and services. By offering only those products and services that customers really want, banks can save substantial money on promotions and offerings that would otherwise be unprofitable. Bank marketers, therefore, need to focus on their customers by learning more about them. Bank of America, for instance, uses database marketing to improve customer service and increase profits. By consolidating five years of customer history records, the bank was able to market and sell targeted services to customers.

Risk Management: Data mining is widely used for risk man-

agement 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. Data mining, however, can be used to reduce the risk of banks that issue credit cards by determining those customers who are likely to default on their accounts. An example was reported in the press of a bank discovering that cardholders who withdrew money at casinos had higher rates of delinquency and bankruptcy. It is a common practice on the part of banks to analyse customers' transaction behaviours in their deposit accounts to determine their probability of default in their loan accounts. 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. Lenders would not have expanded the number of loans they give out without having an accurate, objective, and controllable risk assessment tool. The examples of both a 'good' and 'bad' loan applicant's histories can be used to develop a profile for a good and bad new loan applicant.

Data mining can also derive the credit behaviour of individual borrowers with instalment, 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. Customers who have been with the bank for longer periods of time, remained in good standing, and have higher salaries/wages, are more likely to receive a loan than a new customer who has no history with the bank, or who earns low salaries/wages. 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 (potentially containing transaction information from many companies) and uses 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 Fal-

con's 'fraud assessment system'. It is used by nine of the top ten credit card issuing banks, where it examines the transactions of 80 per cent of cards held in the US. Mellon Bank also uses data mining for fraud detection and is able to better protect itself and its customers' funds from potential credit card fraud.

Customer Acquisition and Retention: Not only can data mining help the banking industry to gain new customers, it can also help retain existing customers. Customer acquisition and retention are very important concerns for any industry, especially the banking industry. Today, customers have so many opinions with regard to where they can choose to do their business. Executives in the banking industry, therefore, must be aware that if they are not giving each customer their full attention, the customer can simply find another bank that will. Data mining can also help in targeting 'new' customers for products and services and in discovering a customer's previous purchasing patterns so that the bank will be able to retain existing customers by offering incentives that are individually tailored to each customer's needs. When Chase Manhattan Bank in New York began to lose customers to competitors, it began using data mining to analyse customer accounts and make changes in its account requirements, thereby allowing the bank to retain its profitable customers. Data mining is also being used by Fleet Bank, Boston, to identify the best candidates for mutual fund offerings. The bank mines customer demographics and account data along different product lines to determine which customers may be likely to invest in a mutual fund, and this information is used to target those customers. Bank of America's West Coast customer service call centre has its representatives ready with customer profiles gathered from data mining to pitch new products and services that are the most relevant to each individual caller. Mortgage bankers are also concerned with retaining customers. The program uses leading-edge Internet technologies, predictive models, and customer-direct marketing to enable lenders to identify new customers and retain those that they already have.

4 APPLICATIONS OF DATA MINING IN THE 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?

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).

What transactions does a customer do before shifting to a competitor? (To prevent shifting of customers)

Which products are often purchased together by the customers of which particular profile? (For target marketing)

What patterns in credit transactions lead to fraud? (To detect and deter fraud)

What is the profile of a high-risk borrower? (To prevent defaults, bad loans, and improve screening)

What services and benefits would current customers likely desire? (To increase loyalty and customer retention)

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.

5 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 software 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

6 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 analysed and the information that is captured is used throughout the organisation to support decision-making. It is universally accepted that many industries (including banking, retail and telecom) are using data mining effectively. Undoubtedly, data mining has many uses in industries. Its practical applications in such areas as analysing medical outcomes, detecting credit card fraud, predicting customer purchase behaviour, predicting the personal interests of Web users, optimizing manufacturing processes etc. have been very successful. It has also led to a set of fascinating scientific questions

about how computers might automatically learn from past experience. The retail industry is also realising that data mining could give them a competitive advantage. A majority of the banks in developing countries (particularly in the public sector) are not usually known to exploit their information 'asset' for deriving business value through data mining and gain competitive advantage. But with progressive liberalisation of rules on entry for private and foreign multinational banks, under the GATS framework of WTO, competitive pressure on domestic banks is increasing. Thus, customer retention and acquisition will be an important determinant of the banks' bottom lines. Those banks and retailers that have realized the utility of data mining and are in the process of building a data mining environment for their decision-making process will reap immense benefit and derive considerable competitive advantage to withstand competition in future.

7 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]. D. Muraleedharan, "Modern Banking: Theory and Practice", PHI Learning private Limited, 2009.
- [4]. Bharati M. Ramager, "Data Mining Techniques and Applications", International Journal of Computer Science and Engineering.
- [5]. S.P. Deshpande, Dr. V.M. Thakare, "Data Mining System and Applications: A Review".