

Deliberation of Data Mining in Banking

Miss Rujuta Shinde*, Miss Priya Vaghurdekar**, Prof. Santaji Shinde***

*Department of IT, BVCOE, Kolhapur, M.S. India

**Department of IT, BVCOE, Kolhapur, M.S. India

***Department of IT, BVCOE, Kolhapur, M.S. India

Abstract

In recent years the ability to generate, capture and store data has increased enormously. The information contain in this data can be very important. 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. The wide availability of huge amounts of the data and the need for transforming such data into knowledge encourage IT industry to use Data Mining. 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 banking sector to enhance the performance of the core business process in banking sector.

Keywords

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

1. Introduction

In India, after the globalization the banking sector has undergone tremendous changes in the way the business is conducted. New products have been introduced to attract customers from different segments . Use of computers and information technology has made banks to ease their business transactions with the clients. With the recent implementation, greater user acceptance and usage of e-banking, the capturing of daily transactional data has become easier and, simultaneously, the volume of such data has grown considerably over a period of time. Thus huge electronic data repositories are being maintained by banks. It is beyond human capability

to analyze this huge amount of data to come up with interesting information that will help in the decision making process. Therefore the analyst is unable or failing to effectively transform the data into useful knowledge for the organization. A number of commercial software 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 rise more rapidly in times to come.

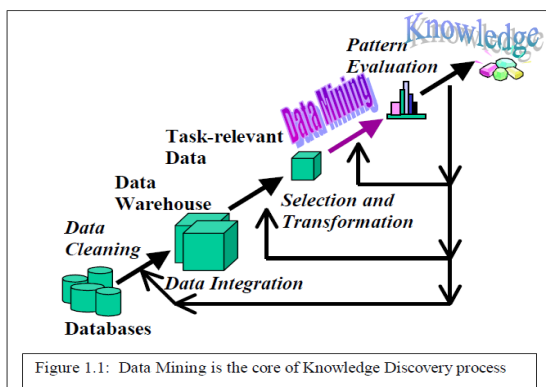
Data mining can contribute to solving business problems in banking sector by identifying patterns and trends like, how customers will react to adjustments in interest rates, which customers will be likely to accept new product offers, the risk profile of a customer segment for defaulting on loans, Know Your Customer (KYC) etc. Causalities and correlation structure among different variables in business data and market prices are not immediately perceptible to managers because the volume of data is too large or is generated too quickly to screen by analysts. Therefore the managers of the banks may go a step further to find the sequences and periodicity of the transaction behavior of their customers which may help them in actually better understanding, segmenting, retaining and maintaining a profitable customer base. In this process Business Intelligence (analytics) and data mining techniques help them in identifying various classes of customers and come up with a class based service/product and/or pricing approach that may garner better revenue management as well.

The banking industry needs to update its customer database as it knows the importance of the information it has about its customers, covering customer demographic profiles, types of transactions, credit /debit card usage pattern, and other details of customer. Since Customer profiling is a data mining

process that builds customer profiles of different groups from the bank'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 bank's customer policies. Thus Valuable business information can be extracted from these data bases. With the emergence of service sector , technological changes that took place over a period of time, the task of maintaining a strong and effective customer relationship management (CRM) becomes a critical issue, as today's technology is capable of generating immense amounts of data, and there may be real-time or near real time requirements placed on the data mining application. 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 customized products and services to those customers.[4]

2. Data Mining

With the enormous amount of data stored in files, databases, and other repositories, it is increasingly important, if not necessary, to develop powerful means for analysis and perhaps interpretation of such data and for the extraction of interesting knowledge that could help in decision-making. *Data Mining*, also popularly known as *Knowledge Discovery in Databases* (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. While data mining and knowledge discovery in databases (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. The following figure (Figure 1.1) shows data mining as a step in an iterative knowledge discovery process.



The Knowledge Discovery in Databases process comprises of a few steps leading from raw data

collections to some form of new knowledge. The iterative process consists of the following steps:

- 1 Data cleaning:** also known as data cleansing, it is a phase in which noise data and irrelevant data are removed from the collection.
- 2 Data integration:** at this stage, multiple data sources often heterogeneous, may be combined in a common source.
- 3 Data selection:** at this step, the data relevant to the analysis is decided on and retrieved from the data collection.
- 4 Data transformation:** also known as data Consolidation , it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.
- 5 Data mining:** it is the crucial step in which clever Techniques are applied to extract patterns potentially useful.
- 6 Pattern evaluation:** in this step, strictly interesting patterns representing knowledge are identified based on given measures.
- 7 Knowledge representation:** is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results.

It is common to combine some of these steps together. For instance, *data cleaning* and *data integration* can be performed together as a pre-processing phase to generate a data warehouse. *Data selection* and *data transformation* can also be combined where the consolidation of the data is the result of the selection, or, as for the case of data warehouses, the selection is done on transformed data.

What kind of Data can be mined?

In principle, data mining is not specific to one type of media or data. Data mining should be applicable to any kind of information repository. However, algorithms and approaches may differ when applied to different types of data. Indeed, the challenges presented by different types of data vary significantly. Data mining is being put into use and studied for databases, including relational databases, object-relational databases and object oriented databases, data warehouses, transactional databases, unstructured and semistructured repositories such as the World Wide Web, advanced databases such as spatial databases, multimedia databases, time-series databases and textual databases, and even flat files. Here are some examples in more detail:

1. Flat files: Flat files are actually the most common data source for data mining algorithms, especially at the research level. Flat files are simple data files in text or binary format with a structure known by the data mining algorithm to be applied. The data in these files can be transactions, time-series data, scientific measurements, etc.

2 Relational Databases: Briefly, a relational database consists of a set of tables containing either values of entity attributes, or values of attributes from entity relationships. Tables have columns and rows, where columns represent attributes and rows represent tuples. A tuple in a relational table corresponds to either an object or a relationship between objects and is identified by a set of attribute values representing a unique key.

3. Data Mining Techniques

The Various techniques of data mining are:

3.1 .Clustering

Similar to classification, clustering is the organization of data in classes. However, unlike classification, in clustering, class labels are unknown and it is up to the clustering algorithm to discover acceptable classes. Clustering is also called *unsupervised classification*, because the classification is not dictated by given class labels. There are many clustering approaches all based on the principle of maximizing the similarity between objects in a same class (*intra-class similarity*) and minimizing the similarity between objects of different classes (*inter-class similarity*).

There are several different approaches to the computation of clusters. Clustering algorithms may be characterized as:

1. **Hierarchical** — Groups data objects into a hierarchy of clusters. The hierarchy can be formed top-down or bottom-up. Hierarchical methods rely on a distance function to measure the similarity between clusters.
2. **Partitioning** — Partitions data objects into a given number of clusters. The clusters are formed in order to optimize an objective criterion such as distance.
3. **Locality-based** —Groups neighboring data objects into clusters based on local conditions.
4. **Grid-based** — Divides the input space into hyper-rectangular cells, discards the low-density cells, and then combines adjacent high-density cells to form clusters.

3.2. Association

Association rules are frequently used by retail stores to assist in marketing, advertising, floor placement and inventory control. Although they have direct applicability to retail business, they have been used for other purposes as well including predicting faults in telecommunication networks. Association rules are used to show the relationships between data items. These uncovered relationships are not inherent in the data, as with functional dependencies and they do not represent any sort of causality or correlation.[6]

Types of Association Rules:

1. Multilevel Association Rule
2. Multi Dimensional Association Rules
3. Quantitative Association Rules
4. Direct Association Rules
5. Indirect Association Rules

3.3. Prediction

Data mining can show how certain attributes within a data will behave in the future. Examples of predictive data mining include the analysis of buying transactions to predict what customers will buy under certain discounts, how much sales volumes a store would generate in a given period, whether deleting a product line would yield more profits. In such applications, Business Logic is used coupled with data mining.[7]

Many real world data mining applications can be seen as a predicating future data states based on past and current data. Prediction can be viewed as a type of classification. The difference is that predication is predicting a future state rather than a current state. Here we are referring to a type of application rather than to a type of data mining modeling approach. Prediction applications include flooding, speech recognition, machine learning and pattern recognition. Although future values may be predicated using time series analysis or regression Techniques.[6]

Types of Regression Methods:

1. Linear Regression
2. Multivariate Linear Regression
3. Non Linear Regression
4. Multivariate Non Linear Regression

3.4. Classification

Data mining can partition the data so that different classes or categories can be identified based on combinations of parameters. Sometimes classification based on common domain knowledge is used as an input to decompose the mining problem and make it simpler.

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 networks and statistics. In classification we make the software that can learn how to classify the data items into groups.[8]For a fraud detection application this would include complete records of both fraudulent and valid activities determined on a record by record basis.[9]

Types of Classification Modes:

1. Classification by Decision Tree Induction
2. Bayesian Classification
3. Neural Networks
4. Support Vector Machines
5. Classification Based on Associations.

4. Data Mining Applications in Banking

As banking competition becomes more and more global and intense, banks have to fight more creatively and proactively to gain or even maintain market shares. Banks which still rely on reactive customer service techniques and conventional mass marketing are doomed to failure or atrophy. The banks of the future will use one asset, knowledge and not financial resources, as their leverage for survival and excellence. Surprisingly, most of this knowledge are currently in the banking system and generated by daily transactions and operations. This valuable information need not be gathered by intrusive customer surveys or expensive market research programs. The only problem is that this storehouse of data has to be mined for useful information. Normally unmined and unappreciated, these terabytes of transaction data are collected, generated, printed, stored, only to be filed and discarded after they have served their short-lived purposes as audit trails and paper trails. Most data generated by the bank's information systems, manual or automated like ATM's and credit card processing, were designed to support or track transactions, satisfy internal and external audit requirements, and meet government or central bank regulations. Few are gathered

intentionally and originally to generate useful management reports. Current information systems are not designed as decision support systems (DSS) that would help management make effective decisions to manage resources, compete successfully, and enhance customer satisfaction and service. Consequently, adhoc or even the most basic management reports have to be extracted excruciatingly from scattered and autonomous data centers or islands of automation that use incompatible formats. The results are management reports that are perennially late, inaccurate, and incomplete. Executive decisions based on these misleading reports can lead to millions of dollars in short and long term losses and lost opportunities and markets.

In banking, the questions data mining can possibly answer are:

1. What transactions does a customer do before shifting to a competitor bank? (to prevent attrition)
2. What is the profile of an ATM customer and what type of products is he likely to buy? (to cross sell)
3. Which bank products are often availed of together by which groups of customers? (to cross sell and do 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)

Several data mining techniques such as distributed data mining has been researched, modeled and developed to help credit card fraud detection. 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. The hidden correlation's between different financial indicators can be discovered by using data mining.

From historical market data, data mining enable to identify stock trading rules. Banking data mining applications may, for example, need to track client spending habits in order to detect unusual transactions that might be fraudulent. Most banks and

financial institutions offer a wide variety of banking services (such as checking, saving, and business and individual customer transactions), credit (such as business, mortgage, and automobile loans), and investment services (such as mutual funds) It has also offer insurance services and stock services. For example it can also help in fraud detection by detecting a group of people who stage accidents to collect on insurance money. The following methods are used for financial data analysis.

1. Loan payment prediction and customer credit policy analysis
2. Classification and clustering of customers for targeted marketing
3. Detection of money laundering and other financial crimes [10]

4.1. CRM

Data mining is used in CRM implementation. For any organization to earn profit, it is necessary to understand its customers. Understanding and responding to customers needs and improving customer service have become important element of corporate strategy. CRM has gained momentum in recent years. The data mining techniques are really a boon for the industry. 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 . 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. The information collected can be used for different purposes like making new marketing initiatives, market segmentation, risk analysis and revising company policies according to the need of the customers. The profiling is usually done on the basis of demographic characteristics, life style and previous transactional behavior of a particular customer. 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. Data mining techniques can significantly improve the customer conversion rate by more focused marketing.[3]

4.2. Risk Management

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. It also helps to predict when the borrower is at default, whether providing loan to a particular customer will result in bad loans etc. 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.

4.3. Marketing

Know Your Customer (KYC) is the buzzword these days. Data mining techniques will help in making customer oriented strategies. Data mining techniques can be used to determine that how customers will react to adjustments in interest rates, reaction of the customer for the existing and new products can be recorded, according to which the future strategies can be designed.[2] 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 analyze 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.[1]

Data mining carry various analysis 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 behaviour 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. Another major area of development in banking is Cross selling i.e. banks makes an attractive offer to

its customer by asking them to buy additional product or service. For example, Home loan with insurance facilities and so on. With the help of data mining technique, banks are able to analyze which products and service are availed by most of the customers in cross selling and which type of consumers prefer to purchase cross selling products and so on.

4.4. Fraud Detection

Data Mining is now used in the banking sector for credit card fraud detection by identifying the patterns involved in fraudulent transactions. It is also used to reduce credit risk by classifying a potential client and predicting bad loans. 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. 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[1]

5. Conclusion

Data Mining can contribute to solving business problems in banking sectors. Data mining techniques can be very useful for the better targeting and acquiring new customers, fraud detection in credit cards fraudulent transactions.

Data Mining is used in CRM to understand and respond to customer's needs and to improve

customer's service in banking. Data Mining plays the role of decision support systems (DSS). Data mining technique helps to distinguish borrowers who repay loans promptly from those who don't.

Data Mining is one of the most promising interdisciplinary developments in Information Technology.

References

- [1] Dr. Madan Lal Bhasin, "Data Mining: A Competitive Tool in the Banking and Retail Industries", The Chartered Accountant October 2006
- [2] Vivek Bhambri "Role of Data Mining in Banking Sector" International Indexed & Referred Research Journal Vol.III ISSUE-33, June, 2012.
- [3] Babita Chopra, Vivek Bhambri, Balram Krishan "Implementation of Data Mining Techniques for Strategic CRM Issues"
- [4] I. Krishna Murthy, "Data Mining- Statistics Applications: A Key to Managerial Decision Making" socio – economic voices 2010
- [5] Data Mining Techniques, <http://www.dataminingtechniques.net/>
- [6] Margaret H. Dunham "Data Mining Introductory and advanced topic"
- [7] Ramez Elmasri, Shyam K. Gupta. "Fundamentals of Database Systems"
- [8] Kazi Imran Moin, Dr. Qazi Baseer Ahmed "Use of Data Mining in Banking" International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 2, Mar-Apr 2012
- [9] Vivek bhambri "Application of data mining in banking sector" International Journal of Computer Science and Technology Vol. 2, Issue 2, June 2011
- [10] Jiawei Han & Micheline Kamber. (2001) Data Mining: Concepts and Techniques, Morgan Kaufmann publishers, CA, USA.

IJERT