



**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH
TECHNOLOGY**

Effective Use of Data Mining in Banking

Sandeep Sharma^{*1}, Sashi Tarun², Pankaj Sharma³

^{*1}M.Tech(CSE) Student, ²Assistant Professor, ³Head of Department Arni University, Indora, Kangra, India

Sanintel123@gmail.com

Abstract

Data mining is one of the tasks in the process of knowledge discovery from the database. In the corporate world every organization is competing the other organization in terms of their value towards the business and the financial growth. 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. In this paper we discuss about the basic details of data mining and the use of knowledge discovery process and the new techniques from the business point of view. In our approach we make an efficient system so that the organization will get the right information at the right time and right to access the necessary information for their growth. The growth of the organization depends on the quality of service, competing with the other organizations, provide required information to the customers, satisfaction of the employees working in the organization. In the banking sector all the financial work can be done in the computers and their connectivity through World Wide Web the softwares get automatically updated in time, use of internet banking and ATM makes the big change in the banking sector. The banks have realized that their biggest asset is the knowledge and the planning to implement the right knowledge at the right time, the financial resources and the techniques of datamining for customer segmentation and profitability, marketing, risk management and customer relationship management and the fraud detection.

Keywords: Data Mining, Banking Sector, Association, Classification, Risk Management, forecasting, CRM..

Introduction

In the banking sector facilities refer to credit line such as overdrafts, loans, import and export lines, etc. offered by the Bank to the customer. Information available under facility enquiry is the facility limit, utilised amount, available amount and hold amount of each facility. 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[2]. 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 softwares 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. Data mining can

help companies in better understanding of the vast volume of data collected by the CRM systems. In the past few years, many organisations have recognised the vital importance of the information they have on their customers[1]. Data mining, can identify products that are often purchased together, which can help build product bundles that are more likely to be successful. 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.

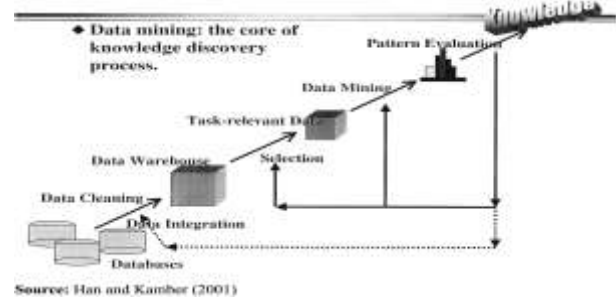
Data might be one of the most valuable resources of any bank but only if it knows how to expose valuable knowledge hidden in raw data. Data mining allows extracting knowledge from the historical data, and predicting outcomes of future situations. It helps optimize business decisions, increase the value of each customer and communication, and improve customer satisfaction.

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. The data may be spatial data, multimedia data, time series data, text data and web data. It is the set of activities used to find new, hidden or unexpected patterns in data or unusual patterns in data[2]. 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. A.Vasudevan, recommended the use of data mining techniques, data available at various computer systems can be accessed by a combination of techniques like classification, clustering, segmentation, association rules, sequencing, decision trees.

Data mining[5] also have one of the tasks in the process of knowledge discovery from the database. Fig. 1 shows the process of knowledge discovery. The steps involved in Knowledge discovery are :

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.



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. The industry needs to explore following data[3]:

1. What is the profile, taste and preferences, attitude of the customer? (Used to Cross sell the products).
2. What transactions does a customer do before shifting to a competitor? (To prevent shifting of customers)
3. What patterns in credit transactions lead to fraud?(To detect and deter fraud)
4. What is the profile of a high-risk borrower?(To prevent defaults, bad loans, and improve screening)
5. Identifying the customers who are getting all types of services from your company?(Identifying 'Loyal' Customers).

Data mining techniques

The various techniques of data mining are:

Association

Association rules are created by analyzing data for frequent patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database[3]. Confidence indicates the number of times the if/then statements have been found to be true. In data mining, association rules are useful for analyzing and predicting customer behavior. They play an important part in shopping basket analysis, product clustering, catalog design and store layout. Programmers use association rules to build programs capable of machine learning. Machine learning is a type of artificial intelligence that seeks to build programs with the ability to become more efficient without being explicitly programmed. 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.

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. 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. Decision tree models are used to solve classification and prediction problems where instances are classified into one of two classes, typically positive and negative, or churner and non-churner in the churn classification case. These models are represented and evaluated in a top-down manner[4].

Following are the examples of cases where the data analysis task is Classification :

A bank loan officer wants to analyse the data in order to know which customer (loan applicant) are risky or which are safe.

A marketing manager at a company needs to analyse to guess a customer with a given profile will buy a new computer.

In both of the above examples a model or classifier is constructed to predict categorical labels. These labels are risky or safe for loan application data and yes or no for marketing data.

Risk Management

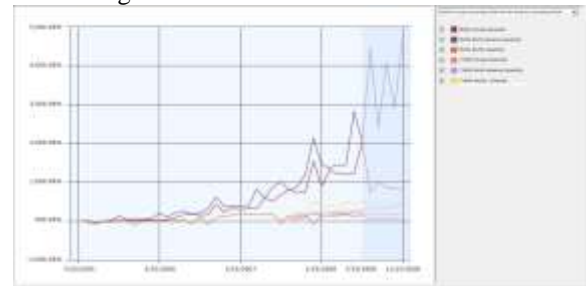
Risk Management is a logical and systematic method of identifying, analyzing, treating and monitoring the risks involved in any activity or process. The key to successful risk management lies in the ability to tailor a formal risk management process that addresses the complementary needs of the business and its customers. A formal risk management process is a continuous process for systematically addressing risk throughout the product/project life-cycle. Risks can be introduced (or latently reside) at the very earliest stages of the project life-cycle. The ability to identify risks earlier translates into earlier risk removal, at less cost, which promotes higher project success probability. Data

mining refers to discovery or “mining” of knowledge from large amounts of data. Data Mining has been described as a confluence of different disciplines primarily database systems, statistics, machine learning and information science.

Managing and measurement of risk is at the core of every financial institution. Today's major challenge in the banking and insurance world is therefore the implementation of risk management systems in order to identify, measure, and control business exposure. Here credit and market risk present the central challenge, one can observe a major change in the area of how to measure and deal with them, based on the advent of advanced database and data mining technology.

Forecasting

In the context of forecasting, the decision-maker needs to find ways to derive value from big data. Data mining for forecasting offers the opportunity to leverage the numerous sources of time-series data, both internal and external, now readily available to the business decision-maker, into actionable strategies that can directly impact profitability. Deciding what to make, when to make it and for whom is a complex process. Understanding what factors drive demand and how these factors interact with production processes or demand and change over time are keys to deriving value in this context. Regression technique can be adapted for prediction. 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. Below is an example of the forecasting model :



Customer relationship management

CRM is applicable only for managing relationships between businesses and consumers. A closer examination reveals that it is even more crucial for business customers. In business-to-business (B2B) environments, a tremendous amount of information is exchanged on a regular basis E-mail

alerts and new product information tailored to different roles in the buyer company can help increase the effectiveness of the sales pitch. Trust and authority are enhanced if targeted academic reports or industry news are delivered to the relevant individuals. All of these can be considered among the benefits of CRM[5].

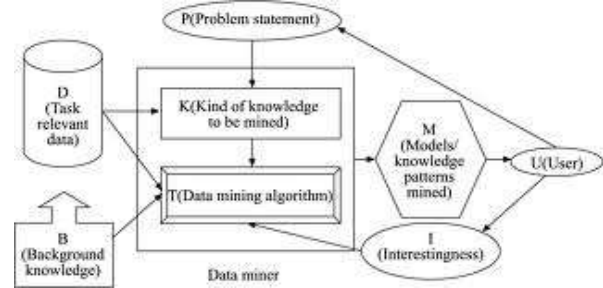
Data mining can be useful in all the three phases of a customer relationship cycle: Customer Acquisition, Increasing value of the customer and Customer retention. Data mining technique can be used to create customer profiling to group the like minded customers in to one group and hence they can be dealt accordingly. The information collected can be used for different purposes like making new marketing initiatives, market segmentation, risk analysis and revising company customer policies according to the need of the customers.

In the banking sector the banks do have their customer relationships software with the help of which the all necessary details of all the customers who have their account in the bank, can be saved and provide the required detail of the on the demand of the customer telephonically. The concept of data mining plays a major role in all these facilities provided by the bank to the customers. Similarly In the business process outsourcing(BPO) the organizations can deal with their customers troubleshoot the fault and to provide the necessary details about their problem. All these facilities helps to retain the old customers and make the new customers as well.

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 fault, whether providing loan to a particular customer will result in bad loans etc. The data mining techniques can help them to detect and hence prevent frauds.

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. The following figure explains the sequential pattern[3].



Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics[2].

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data preprocessing and model parameters until the result achieves the desired properties.

Fraud Detection

Data mining and statistics help to anticipate and quickly detect fraud and take immediate action to minimize costs. Through the use of sophisticated data mining tools, millions of transactions can be searched to spot patterns and detect fraudulent transactions. An authenticated person will be allowed to do the transactions for example the internet facility provided by the bank require the User name and the password if both the information is correct then only it will allow to do the transaction otherwise it will give error. This facility provided by the bank will

make the banking safe and secure and hence will be able to make more and more customers. An important early step in fraud detection is to identify factors that can lead to fraud. What specific phenomena typically occur before, during, or after a fraudulent incident? What other characteristics are generally seen with fraud? When these phenomena and characteristics are pinpointed, predicting and detecting fraud becomes a much more manageable task.

Prominent examples of fraud areas are

- Insurance fraud (e.g. fraudulent car or health insurance claims)
- Telecommunications fraud (phone cloning, subscription fraud)
- Investment fraud (pyramid schemes, insider trading)
- Employee fraud (falsification of balance sheets, embezzlement)
- Credit Card fraud (stolen or cloned credit cards)
- Retail fraud (forgeries, fake sales)
- Advance fee fraud (Nigerian money offer, lottery scam)
- Computer and internet fraud (Phishing, Spoofing, etc.)

Conclusion

Data Mining techniques makes the effective help to the banks and other institutions which is related to for better growth of the organization by providing them the new customers and retaining the old customers by using certain techniques, how to fulfill the targets within the certain deadline of the organization, how to combine the homogeneous task in the single module known as clustering so that they can be easily done, 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] Vivek Bhambri "Application of Data Mining in Banking Sector" *International Journal of Computer Science and Technology* Vol. 2, Issue 2, June 2011.
- [2] B.Muneendra nayak, Naveen kumar, R. Mahammad Shafi "A FOCUS ON DIFFERENT FRAUDS AND USING DATA MINING TO ENHANCE BUSINESS PROCESS IN BANKING SECTOR" *International Journal of Engineering Sciences Research* Vol 04, Special Issue 01, 2013.
- [3] Kazi Imran Moin, Dr. Qazi Baseer Ahmed "Use of Data Mining in Banking" *International Journal of Engineering Research and Applications* Vol. 2, Issue 2, Mar-Apr 2012, pp.738-742.
- [4] Dr. K. Chitra, B. Subashini "Data Mining Techniques and its Applications in Banking Sector" *International Journal of Emerging Technology and Advanced Engineering* Volume 3, Issue 8, August 2013.
- [5] Vivek Bhambri, Gagandeep "COEXISTENCE OF DATA MINING AND PRIVACY OF DATA" *International Journal of Research in IT & Management* Volume 2, Issue 2 (February 2012).