International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Fraud Detection in Banking Sector using Data mining

B. Rajdeepa[1], D. Nandhitha[2]

[1]Assistant Professor, Department of Computer Science, PSG College of Arts & Science, Coimbatore, Tamilnadu

[2]Research scholar, Department of Computer Science, PSG College of Arts & Science, Coimbatore, Tamilnadu

Abstract: ***In present scenario when the term fraud comes into a discussion, banking fraud clicks to mind until now. In recent years banking fraud has increasing extremely. Every year fraud in banking is rising. Fraud presents significant cost to our economy. For customer segmentation and productivity, most of the banks are using data mining, and also for credit scores and approval, predicting payment default, marketing, detecting fraudulent transactions, etc. This paper provides an overview of the concept of Data Mining and different frauds in Banking. The banking sector consists of public sector, private sector and foreign banks, apart from smaller regional and cooperative banks. IT-based banking products, services and solutions are available in market. Phone Banking; ATM facility; Credit, Debit and Smart Cards; Internet Banking & Mobile Banking; SWIFT Network & INFINET Network are most common in banking.***

Keywords: Fraud, Banking, Data Mining, Fraud Detection

1. Data Mining

Data mining is a process to extract the implicit information and knowledge which is potentially useful. The data is extracted from the mass, incomplete, noisy, fuzzy and random data by which the data mining process is done.

In business, scientific domain and in all sectors the need of the database applications has been increased, so the need of data and the storage device are also increased. This explosion in the amount of electronically stored data was accelerated by the success of the relational model for storing data and the development and maturing of data retrieval and manipulation technologies. This technology is used for storing the data as the demand increased; little stress was paid to developing software for analyzing the data until recently. when companies realized that hidden within these masses of data was a resource that was being ignored. The large amount of stored data contains knowledge about a number of aspects of their business waiting to be harnessed and used for more effective business decision support. Database Management Systems used to manage these data sets at present only allow the user to access information explicitly present in the databases. In database the stored data is only a small part of the 'iceberg of information' available from it. Contained implicitly within this data is knowledge about a number of aspects of their business waiting to be harnessed and used for more effective business decision support. This extraction of knowledge from large data sets is called Data Mining or Knowledge Discovery in Databases. It is defined as the non-trivial extraction of implicit, previously unknown and potentially useful information from data. The benefits of Data Mining have resulted in a lot of resources being directed towards its development.

Developments in the database field, machine learning research were developed based on different models of human learning. Learning by example, cased-based reasoning, learning by observation and neural networks are some of the most popular learning techniques that were being used to create the ultimate thinking machine.

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.

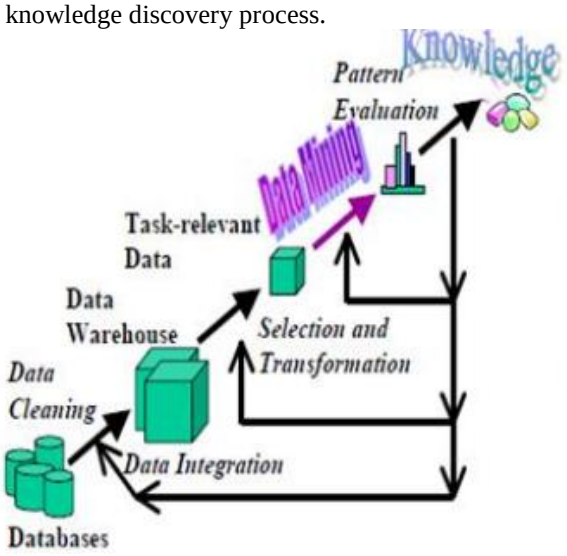

Figure 1: KDD Process

2. Data Mining Algorithms and Techniques

Various algorithms and techniques are used for databases like Classification, Clustering, Regression, Neural Networks, Association Rules etc.,

A. 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. For this type of analysis Fraud detection and credit

risk applications is well suited. This approach implies decision tree or neural network-based classification algorithms. Learning and classification are the process that involves in data classification. In Learning the training data are analyzed by classification algorithm. To estimate the accuracy of the classification rules data test is used. 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. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier. Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

B. Clustering

Identification of similar classes of objects is known as clustering. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly, so clustering can be used as preprocessing approach for attribute subset selection and classification. For example, to form group of customers based on purchasing patterns, to categories genes with similar functionality. Types of clustering methods are

- Partitioning Methods
- Hierarchical methods
- Density based methods
- Grid-based methods
- Model-based methods

C. Predication

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. The same model types can often be used for both regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models. Types of regression methods

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

D. Association rule

Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. 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 (if any) value. Types of association rule

- Multilevel association rule
- Multidimensional association rule
- Quantitative association rule

E. Neural networks

Neural network is a set of connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights so as to be able to predict the correct class labels of the input tuples. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. These are well suited for continuous valued inputs and outputs. For example handwritten character reorganization, for training a computer to pronounce English text and many real world business problems and have already been successfully applied in many industries.

3. Top 10 Frauds in Indian Banking Sector

The Reserve Bank of India – RBI maintains data on frauds on the basis of area of operation under which the frauds have been perpetrated. According to such data pertaining, top 10 categories under which frauds have been reported by banks are as follows

1) Credit Cards
2) Deposits – Savings A/C
3) Internet Banking
4) Housing Loans
5) Term Loans
6) Cheque / Demand Drafts
7) Cash Transactions
8) Cash Credit A/c (Types of Overdraft A/C)
9) Advances
10) ATM / Debit Cards

4. Data Mining Applications in Banking

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 analyses this huge amount of raw data and to effectively transform the data into useful knowledge for the organization. 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. 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.

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 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. 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 '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 Falcon'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.

5. Conclusion

Data mining is a technique used to extract vital information from existing huge amount of data and enable better decision-making for 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 analyzed and the information that is captured is used throughout the organization to support decision-making. Data Mining techniques are very useful to the banking sector for better targeting and acquiring new customers, most valuable customer retention, automatic credit approval which is used for fraud prevention, fraud detection in real time, providing segment based products, analysis of the customers, transaction patterns over time for better retention and relationship, risk management and marketing.

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] Phua, C., Lee, V., Smith, K. & Gayler, R. (2005) A comprehensive survey of data mining-based fraud detection research, Artificial Intelligence Review (2005) 1–14.
[5] B. Desai and Anita Desai, "The Role of Data mining in Banking Sector", IBA Bulletin, 2004.
[6] S.S.Kaptan, “New Concepts in Banking”, Sarup and Sons, Edition, 2002.
[7] 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.
[8] 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.
[9] 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.
[10] K. Chitra, B.Subashini, Fraud Detection in the Banking Sector, Proceedings of National Level Seminar on Globalization and its Emerging Trends, December 2012.

Author Profile

B. Rajdeepa , Assistant Professor, Department of Computer Science, PSG College of Arts & Science, Coimbatore, Tamilnadu, India

D. Nandhitha , Research scholar, Department of Computer Science, PSG College of Arts & Science, Coimbatore, Tamilnadu, India