

# A CASE STUDY ON DATA MINING APPLICATIONS ON BANKING SECTOR

<sup>1\*</sup>M.V. Jisha, <sup>2</sup>D. Vimal Kumar

<sup>1</sup>Nehru Arts and Science College, T.M.Palayam, Coimbatore, India

<sup>2</sup>Dept. Computer Science Nehru Arts and Science College, T.M.Palayam, Coimbatore, India

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

**Abstract-** Data mining is becoming strategically important area for many business organizations including banking sector. It is a process of analyzing the data from various perspectives and summarizing it into valuable information. Data mining assists the banks to look for hidden pattern in a group and discover unknown relationship in the data. Today, customers have so many opinions with regard to where they can choose to do their business. Early data analysis techniques were oriented toward extracting quantitative and statistical data characteristics. These techniques facilitate useful data interpretations for the banking sector to avoid customer attrition. Customer retention is the most important factor to be analyzed in today's competitive business environment. Fraud is another significant problem in banking sector. Detecting and preventing fraud is difficult, because fraudsters develop new schemes all the time, and the schemes grow more and more sophisticated to elude easy detection. This paper analyzes the data mining techniques and its applications in banking sector like fraud prevention and detection, customer retention, marketing and risk management.

**Keywords**— Banking Sector, Customer Retention, Credit Approval, Data mining, Fraud Detection.

## I. INTRODUCTION

Technological innovations have enabled the banking industry to open up efficient delivery channels. IT has helped the banking industry to deal with the challenges the new economy poses. Now a days, Banks have realized that customer relationships are a very important factor for their success. Customer relationship management (CRM) is a strategy that can help them to build long-lasting relationships with their customers and increase their revenues and profits. CRM is of greater importance in the banking sector. The CRM focus is shifting from customer acquisition to customer retention and ensuring the appropriate amounts of time, money and managerial resources. The challenge the bank face is how to retain the most profitable customers at the lowest cost. At the same time, they need to find and implement this solution quickly and the solution to be flexible. To detect fraud, they require complex and time-consuming investigations that deal with different domains of knowledge like financial, economics, business practices and law. Fraud instances can be similar in content and appearance but usually are not identical. In developing countries like India, Bankers face more problems with the fraudsters. Using data mining technique, it is simple to build a successful predictive model and visualize the report into meaningful information to the user.

This paper discusses about the application of DM in various activities on banking sector in the Section 2. Section 3 specifies the fraud analysis, Section 4 highlights the determinants of customer retention.

## II. APPLICATION AREAS OF DATA MINING TECHNIQUES IN BANKING

**Security and fraud detection:** Big secondary data like transaction records are monitored and analysed to enhance banking security and distinguish the unusual behaviour and patterns indicating fraud, phishing, or money laundering (among others).

**Risk management and investment banking:** Analysis of in-house credit card data freely accessible for banks enables credit scoring and credit granting which form part of the most popular tools for risk management and investment evaluation.

**CRM:** DM techniques have been widely applied in banking for marketing and customer relationship management related purposes such as customer profiling, customer segmentation, and cross/up selling. These help the banking sector to have a better understanding of their customers, predict customer behaviour, accurately target potential customers and further improve customer satisfaction with a strategic service design.

**Other advanced supports:** A few less mainstream applications focus on branching strategy, and efficiency and performance evaluation, which can significantly assist in achieving strategic branch locating and expansion plans.

### III. FRAUD ANALYSIS

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

Cyber crimes that are committed in banks include hacking, Credit card fraud, money laundering, DoS attacks, phishing, salami attacks, ATM card cloning etc. Cyber threats such as pharming , phishing, tempted reveal of private details like identity theft are the security qualms that subsist in the brains of clients in banking and financial sectors. Crime prediction uses past data and after analyzing data, predict the future crime with location and time. At present, serial criminal cases rapidly occur so it is an challenging task to predict future crime accurately with better performance [5]. Credit card and web based crime are increasingly as more technologies are rising high. To deal and overcome fraud, clustering and classification techniques are implemented [6]. Fraud detection is one of the difficult process not only technically, but also in crime investigations. The method of fraud detection is based on simple comparisons, and also based on association, clustering, prediction and outlier detection [7]. Association Rule mining as generates “n” best association rules based on n selected and Classification and Regression Tree (CART) that predict categorical class labels [8]. Clustering technique is devised as a multi purposive optimization crisis. The suitable clustering algorithm and parameter locations depend on the entity dataset and projected use of consequences.

### IV. DETERMINANTS OF CUSTOMER RETENTION

Following factors are based on thorough review of literature and interaction with experts and customers –

**Physical Appearance:** It included modern infrastructure, appealing environment, physical facilities are matching with the services expected by our customers, customers feel relaxed when they enter into the premises of our bank etc.

**Responsiveness:** This consists of Proper information about usage and benefits of the product/services are communicated to customers, employees have the knowledge to answer the

customers' questions, all true and meaningful information to customers are provided, convenient opening and closing time of the banks is satisfactory for our customers, Customers are informed regularly about latest and forthcoming schemes offered, Various problems which occur during electronic transactions are timely responded by the bank officials etc.

**Grievance Redressal System:** It denotes an effective of grievance redressal system, able to handle the customers' complaints easily and immediately, Personal attention to customers is provided whenever they required etc.

**Reliability:** comprised of trust and the Bank is reliable because it is mainly concerned with the investor's interest, cash transaction system of this bank is trustworthy, and has accuracy in performing financial transactions, Bank has sound and attractive return policies for investors etc.

**Service Orientation:** reflects in the consistency of services, to train the employees for bundling of services, Bank officials are more committed towards their customers, and Computerized information system provides best and quick services to the customers.

**Convenience:** is comprised of location, speedy documentation, clear departmentation according to the needs of customers.

**Customer Orientation:** reveals the Long term relationships with the customers, Bank adapted new technologies to improve communication with customers, Bank officials build mutual trust between the bank and its client, Bank encourages customers to purchase their products and services, Adaptability of different measures to meet customers' urgent requirements, care for customers etc.

**Security in Transactions:** This factor mentions that Banks ensure proper security system to protect customer's transaction, implementation of core banking solutions, and Customers get instant alerts regarding their transactions on their communication devices.

**Loyalty Programs:** Through this programme customers are easily benefited with promotional offers provided by the bank, promotional schemes that offer good value for money, benefitted by receiving extra rewards points on higher spending, earn rewards points etc.

**PRICE:** includes flexible rate of return for various products or services that meet the needs of customers, fair pricing on SMS updates, ATMs transactions, locker services, issuing of check book etc. and best interest rate on FDs to its customers.

**CRM Strategies:** reflect that Product features are developed to allow customization ,Technology enabled customers'

interactions, promoting a culture of diversity, Measurable objectives, NRI banking services are provided to customers by our banks etc.

**Corporate Image:** is denoted by proactive, good reputation, financially affordable for its customers and high creditability.

In corporate sector, particularly banking sector, the retention of customer is an important factor. We propose a classification based model to assess the customer behaviour in banking sector with respect to RFM. Three classification models like decision tree, ANNs and k-NN are studied and evaluated.

In any case, certain concepts are to be further defined and elaborated:

#### A. Customer Life Time Value (CLV) or Customer Loyalty

Customer happiness and attraction are the core objectives of any company. Therefore, customer loyalty or Customer lifetime value (CLV) is calculated in terms of recency, frequency and monetary variables (RFM).

#### B. RFM Analysis

Customer relationship management (CRM) system is used to manage company relations with the existing and prospect customers. RFM stands for recency, frequency and monetary. This technique is commonly used in database marketing and direct marketing to identify customer behaviour or purchasing behaviour.

We represent the customer behaviour in three variables:

- Recency: How recently did the customer purchase? This is the interval between the time of most recent and current customer purchase.
- Frequency: How frequently do they purchase? This is the number of transactions in a specific period.

- Monetary (value): How much do they spend? This is the amount of money that customer spent during a specific period.

#### C. Decision Tress

Decision tree is a data mining and classification technique. This technique is mostly used for prediction and classification. A tree consists of paths, branches and nodes. A collection of branches is called path and represents the attribute value. Class value is represented by leaves. Each path in decision tree represents a rule which is used for classification or prediction. Decision tree divides data into subsets or nodes. Root node is the top node. Root node represents the complete dataset. Tree pruning is performed when tree is completely built. Pruning is started from the leaf node.

#### D. Weka Tool

It is a machine learning toolkit extensively used for education, projects and research. It is an open source tool and developed by Waikato University, New Zealand. All the major machine learning algorithms are implemented in Weka. Data mining tasks like classification, clustering, association rules, data preparation or datapreprocessing etc. can be performed using this tool.

#### E. ANN (Artificial Neural Network)

ANN is a data mining classification technique and it is used as a classifier. This structure consists of layered approach. Collection of neurons is called a layer. We can define multiple or hidden layers in the ANN structure. Computation is performed in the hidden layer.

#### F. k-Nearest Neighbours (k-NN)

k-NN is a very simple data mining technique and is used for classification purposes. Here the symbol "k" is a static value and mostly it takes an odd value like 3, 5 and 7. Euclidean distance formula is used for measuring distances between the two entities and serves as a similarity index.

Table1. Summary table of Data Mining Application in Banking Sector

Sector	Key Techniques	Purposes
<b>Security and fraud detection</b>	classification (DT, NN, SVM, NB), k-mean clustering, ARM	Identifying phishing, fraud, money laundering, credit card fraud, security trend of mobile/online/traditional banking.
<b>Risk management and investment banking</b>	classification (DT, NN, SVM, NB, LR), k-mean clustering UCI Repository International Dataset	Credit scoring, credit granting, risk management for peer-to-peer lending.
<b>Customer profiling and</b>	classification (DT, NN),	Efficiently build accurate customer profiles.

<b>knowledge</b>	k-mean clustering	
<b>Customer segmentation</b>	k-mean clustering	Provide sufficient customer segmentation, conduct customer-centric business strategies.
<b>CRM Customer satisfaction</b>	k-mean clustering, classification (NN)	Make the most strategic investment on maintaining and enhancing customer satisfaction.
<b>Customer development and customization</b>	classification (DT, NN, NB, LR, SVM), k-mean clustering	Strategic banking via direct marketing, targeted marketing, product cross/up selling.
<b>Customer retention and acquisition</b>	classification (DT, NN, LR, SVM), ARM, k-mean clustering	Customer churn prediction and prevention, attracting potential customers and strategic future service design.
<b>Other advanced supports</b>	classification (NN, DT, SVM), k-mean clustering	Branch strategy, bank efficiency evaluation, deposit pricing, early warning of failing bank.

## V. CONCLUSIONS

This paper successfully captured and systematically reviewed few DM applications in banking sector. It fulfils the literature gap and serves as a quick reference guide for recent DM implementations in banking. Having reviewed these recent publications, it can be concluded that the banking sector has adopted DM mainly for fraud detection, risk management and CRM.

Apart from the comprehensive summary of recent developments of DM applications in banking, this study also aims to present insights into the challenges and directions for future research. Firstly, it is noted that although the big banking data consists of large volumes of unstructured data, there are many DM techniques which continue to be rarely exploited, e.g., text mining, entity extraction, and social network analysis. This unbalanced exploration status can be caused by the limited access of banking data, the shortage of researchers with relevant skill set, system constraints, and the lack of advanced data analytic tools [7]. Specifically, the confidentiality restrictions of banking related data have limited the progression of research. Therefore, seeking a proper solution for data availability will make a significant difference for future research.

## REFERENCES

- [1].T. F. Bahari, M. S. Elayidom, "An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour", *Procedia Computer Science*, Vol. 46, pp. 725-731, 2015
- [2].A. Shrivastava, B. Kumari, "Implementation of classifiers and their performance evaluation", *International Journal of Engineering Research Online*, Vol. 3, No. 2, pp. 71-78, 2015
- [3]. N. Khan, F. Khan, "Fuzzy based decision making for promotional marketing campaigns", *International Journal of Fuzzy Logic Systems*, Vol. 3, No. 1, pp. 64-77, 2013
- [4]. H. A. Elsalamony, "Bank Direct Marketing Analysis of Data Mining Techniques", *International Journal of Computer Applications*, Vol. 85, No. 7, pp. 12-22, 2014
- [5]. A. Nachev, "Application of Data Mining Techniques for Direct Marketing", in: *Computational Models for Business and Engineering Domains*, pp.86-95, ITHEA, Rzeszow – Sofia, 2014
- [6]. Vikas Jayasree and Rethnamoney Vijayalakshmi Siva Balan, A REVIEW ON DATA MINING IN BANKING SECTOR.: *American Journal of Applied Sciences*, 2013.
- [7]. Lagazio, M.; Sherif, N.; Cushman, M. A multi-level approach to understanding the impact of cyber crime on the financial sector. *Comput. Secur.* 2014, 45, 58–74.
- [8]. Kharote, M.; Kshirsagar, V.P. Data mining model for money laundering detection in financial domain. *Int. J. Comput. Appl.* 2014, 85, 61–64. [[CrossRef](#)]
- [9]. Jayasree, V.; Balan, R.V.S. A review on data mining in banking sector. *Am. J. Appl. Sci.* 2013, 10, 1160. Pulakkazhy, S.; Balan, R.V.S. Data mining in banking and its applications—A review. *J. Comput. Sci.* 2013, 1252–1259.
- [10]. Amani, F.A.; Fadlalla, A.M. Data mining applications in accounting: A review of the literature and organizing framework. *Int. J. Account. Inf. Syst.* 2017, 24, 32–58. [[CrossRef](#)]
- [11]. Wongchinsri, P.; Kuratach,W. A survey-data mining frameworks in credit card processing. In *Proceedings of the 2016 13th International Conference on Electrical Engineering ;Electronics,Computer, telecommunications and Information Technology (ECTI-CON)*, ChiangMai, Thailand, 28 June–1 July 2016; pp. 1–6.
- [12]. Bhasin, M.L. Menace of frauds in the Indian banking industry: An empirical study. *Aust. J. Bus. Manag. Res.* 2015, 4, 1–13. [[CrossRef](#)]
- [13].Hassani, H.; Huang, X.; Silva, E.S.; Ghodsi, M. A review of data mining applications in crime. *Stat. Anal. Data Min. ASA Data Sci. J.* 2016, 9, 139–154. [[CrossRef](#)]
- [14]. Hassani, H.; Huang, X.; Ghodsi, M. Big Data and Causality. *Ann. Data Sci.* 2018, 5, 133–156. [[CrossRef](#)]
- [15]. Hassani, H.; Silva, E.S. Big Data: a big opportunity for the petroleum and petrochemical industry. *OPEC Energy Rev.* 2018, 42, 74–89. [[CrossRef](#)]