Journal of Innovative Engineering and Research (JIER)
Vol.- 2,Issue - 1, April 2019, pp. 1-4 (4 pages)

A Brief Study of Various Data Mining Techniques and its Applications in Internet Banking

NIDHI NIGAM VERMA[1], DEEPIKA PATHAK[2]

[1,2]*Department of Computer Application, Dr. A. P. J. Abdul Kalam University, Indore*
Corresponding Author Email: nidhinigam2007@gmail.com

Abstract— Data mining is a process of extracting usable information or knowledge from large amounts of data. For the banking industry, data mining, its importance, and its techniques are vital because it helps to extract useful information from a large amount of historical data which enable to make useful decisions. Nowadays Data mining is vital for many corporate firms, banking is one of them. In banking, industry customers are becoming more and more demanding in terms of the level of service, responsiveness, costs, quality of products etc. Banks could only reach to these goals furthermore if they take initiative to invest in new technologies. In the recent era, a new technology that has achieved considerable attention, especially among banks, is Internet banking. Its large scope of applications, its advantages brings an immoderate change in a common human's life. In order to provide useful knowledge for the purpose of future research and development, a detailed and most up to date analysis on the current status of Data Mining in banking will be very beneficial. This paper is an attempt to review some of the most useful data mining techniques that can be greatly used banking areas. It as well provides an insight into how these data mining techniques can be utilized in banking areas and help in the decision-making process.

Index Terms—Data mining techniques, internet banking, clustering, Association, classification, prediction, Association Rule Mining, Data Mining applications in Internet banking.

I. INTRODUCTION

In today's world technology improves the functioning of the banking industry and the services offered by them. The Internet has appeared as the best medium for the delivery of banking products and services. In the early nineties, each bank of India has done some advancement in technology by the utilization of internet banking, telemarketing, ATM, mobile banking and much more [1]. This is a technology-driven advent to reach the maximum number of customers at low cost in an effective manner. Nowadays the banking industry has become highly competitive. To be capable to grow and survive in this changing environment of business and marketing banks are going to utilize updated and latest technologies, it is as well consider as a tool of effective communication with clients and cost reduction and institution matter with a banking transaction. With the advent of computers, its technology and a broad application of databases, the bank's associated huge amounts of data, which is collected in different forms but analyze usable knowledge for taking a crucial decision sometimes become very difficult. At this instance of time, a new emerging concept of Data mining with its various useful techniques came into existence that makes the capability to find out the true value of data and through which decision-making process became somewhat easier [1, 3].

In recent era several industries mostly banks, as well as financial industries, have realized the true importance of the information they have by their customers. Data mining minimizes the cost of purchasing and provides a reasonable solution to various other organizations problems. Data mining have plenty of usable applications across industries. It offers organizations a novel way of doing business. It opens new paths and horizons for understanding business needs and serves their customer in a better way. There are several other useful applications of data mining which includes database marketing, sales forecasting, behavior pattern analysis and many more[1]. It is important to apply the data mining tools and techniques in a profitable manner it is required to maintain the pace of business and to keep beside industry dynamics.

Thousands of decisions are taken by a bank on daily basis. They are relationship startup, credit decisions, relationship startup, default decisions, investment decisions, and Illegal financing [2]. For this process, one requires to depend on various reports provided by the bank to come at crucial decisions. But this manual process is very time consuming due to huge volumes of transactional data and is error-prone. Interesting patterns and useable data can be mined from this huge volume of data that in turn can be utilized for this decision making process.

II. DATA MINING

Data mining is the process of extracting predictive information from large data bases. It can as well be defined as

a process of analyzing data from various perspectives and optimize it into useable information. Data mining is fully associated with data science which comprises classification and manipulation of data by implement mathematical and statistical concepts. Data's are available everywhere around the world. It can be utilized for the prediction of future problems. For this mostly the statistical approach is used [4]. Sometimes, data might be in different formats as it receives from various different sources, inappropriate attributes and missing values in data. Therefore, data cleaning is necessary before applying any technique of data mining.

Data mining is as well known by some other names which include information discovery, knowledge extraction, information harvesting, and data pattern processing and data archeology [4]. Data mining with its applications will come out as emerging and assuring technological developments that offer valuable means to access different types of data and information available globally. Apart from this, these applications as well contribute in decision making. There has been a drastic boost in amount of data and information which is gathered in electronic format considering last few decades. Nowadays the size of database due to continuous increment and has reached till terabytes.

III. DATA MINING TECHNIQUES USED IN INTERNET BANKING

There are various techniques and algorithms of data mining developed and used to solve the critical problems related to internet banking and several other issues. They are:

3.1 CLUSTERING

Clustering is a process of distribution of data into the groups of similar objects, called clusters. It includes objects that are having similar properties between themselves and unlike to objects of another groups [3]. The main objective of cluster analysis is to determine a system of organizing observations, mostly people, into groups, groups members share common. This procedure aims to classify relatively homogeneous groups called clusters, which is based on some characteristics, using an algorithm that can be applicable on large numbers of cases [3].Although; the algorithm needs you to mention the total number of clusters.

Clustering in internet banking is used to analyze the characteristics of customer and their behaviors with particular criteria's like date of transaction, access time, transaction status etc. The advantages are profitable for the bank to improve their services and reach customer satisfaction level. Concept formation is a intently related process to clustering [7].For banking purposes, Concept formation and Clustering can be used for classifying the customers with same types of queries or transactions to same products or having similar risk aptitude. Having appropriate knowledge about these classes will definitely help the banking industry in developing products for each class of customers and can start working on effective and targeted marketing campaigns[7].

3.2 CLASSIFICATION

Classification is a most commonly used data mining technique to predict group membership for data elements in a dataset. Popular classification techniques which are used include decision trees and neural networks [8]. Classification is a data mining technique which is used mostly in order to defend from fraud and escape from risk of credit [1].In classification technique of data mining data accuracy is very important. Especially in classification method test data are utilized to anticipate the accuracy of classification. If that accuracy is acceptable then only that value can be applied to new data set.

Credit risk and fraud detection applications are mostly well suited to this kind of analysis. This approach often employs neural network or decision tree based classification algorithms.

3.3 PREDICTION

It is a most useful technique of data mining, as it name suggests it determine the relationship between a dependent and an independent variable. Most of the time regression analysis is used for prediction model, as it establishes the relationship between one or more variables in the data set. This technique is really very beneficial and helpful in discover the pattern through which one can able make a logical or reasonable prediction.

For prediction, Regression technique of data mining can be used. Regression analysis may be used to exemplary the relationship between one or more independent and dependent variables. Here in data mining, independent variables are the variables or attributes which are known in advance and the response variables are those which we want to predict [10].

This is the fact that many real-world problems are not just prediction. For example, stock prices, sales volumes, and product failure rates are all not very easy to predict since they might depend on various complex interactions of several predictor variables. Therefore, some more complex techniques like decision trees, logistic regression or neural networks will be required to forecast future values.

3.4 REGRESSION

Regression technique is broadly used to predict the relationship among two or more independent and dependent variables. Linear Regression is a mostly used technique to create a relationship among a dependent variable and the independent variable. Nowadays, Data Mining techniques are really very useful in the banking industry for more suitable targeting and bring in new customers, most precious customer detention, automatic credit approval which is utilized for fraud detection, prevention in real time, offering segment based products, customers analysis, risk management and marketing.

Regression is a most popular technique of data mining which is used to predict a range of numeric values called continuous values as well, for given a specific dataset. Regression technique is applied across different industries for business

and marketing planning, environmental modeling, financial forecasting and trends analysis.

3.5 ASSOCIATION RULE MINING

Association rule mining is a popular and most used method for identifying interesting relations among variables in large data sets. It is intended to determine strong association rules discovered in data sets using various measures of concern [8]. Association rule mining aims to find interesting correlation or association relationships between a large set of data items. The discovery of such interesting association relationships among large amounts of customer transaction records can support in decision making processes for many businesses such as cross marketing, catalog design and loss-leader analysis[

One of the most popular techniques of data mining used by banking industry is the discovery for association rules among products and services a bank offers. Generally results are very impressive as in many cases strong relations among data items are established, which are not at all easily observed at a first glance. These rules are now established as additional tools used for the continuous improvement of bank products and services and also help in attracting new customers towards internet banking.

Assurance of association rules regarding banking data is a challenging though demanding task because:

- The volume of banking data is immense. So the data must be appropriately prepared before the final step of the application of method.
- The objective must be clearly stated from the beginning. In several cases not having a clear objective results in inaccurate or no results at all.
- Good knowledge of the dataset is necessary not only for the data miner but also for the final analyst, if not false results will be produced by the data miner and inaccurate conclusions will be making by the manager.

Not all the rules discovered are interesting. The analyst must recognize strong association rules for decision making.

III. CONCLUSION

We Data mining is a useful process to extract information or knowledge from an existing database. It is a most useable tool in internet banking sector to facilitate better decision-making. It involves knowledge data discovery process, which includes various steps as data selection, data integration, data transformation, data mining, pattern evaluation, knowledge presentation. Banks utilize data mining tools and techniques in various fields like fraud detection, marketing, risk management, money laundering detection and in investment banking.

A comparative study of different data mining techniques is discussed in this paper. It is very necessary to decide which technique to use for a particular dataset and when. Clustering in internet banking is used to analyze the characteristics of customer and their behaviors with particular criteria's like date of transaction, access time, transaction status etc. Prediction can be applied on data to predict their future performance. For Credit risk and fraud detection applications, classification techniques are mostly well suited to this kind. Regression is generally used to predict a continuous or numeric value while classification allots dataset into discrete categories. One of the most popular techniques of data mining used by banking industry is the development for association rules among various products and services a bank offers. Generally results are very impressive as in many cases strong relations among data items are established, which are not at all easily observed at a first glance. Data Mining is very useful for banking sector for better acquiring and targeting new customers and helps to analyze customers and their transaction behaviors.

REFERENCES

1. Md. R. Farooqi and N. Iqbal, "Effectiveness of Data mining in Banking Industry: An empirical study", international Journal of Advanced Research in Computer Science, Volume 8, No. 5, pp. 827-830, May-June 2017.
2. S. Pulakkazh and R.V.S. Balan," Data Mining In Banking and its applications- A review", Journal of Computer Science, Vol. 9 Issue 10, pp. 252-1259, 2013.
3. M. Dash and S. K. Nayak," Data Mining Need In Profiling Internet Banking Customer" International Journal of Computer Science and Management Research, 2012.
4. K. Rangra and K. L. Bansal," Comparative Study of Data Mining Tools ",International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 6, pp. 216-223, June 2014.
5. S. Raj, "A Review Of Data Mining Applications In Banking", International Journal of Computer Science and Management Research, June 2015.
6. V. Jayasree and R.V. Siva Balan, "A Review on Data Mining in Banking Sector", American Journal of Applied Sciences, Vol. 10 Issue 10, pp. 1160-1165, 2013.
7. C. Rajabhushanam and S. cumaar, "An Study Of Data Mining Applications In Banking", International Journal of Pure and Applied Mathematics, Volume 116 No. 15, pp. 265-271, 2017.
8. N. K. Dokania and N. Kaur," Comparative Study Of Various Techniques In Data Mining", International Journal Of Engineering Sciences & Research Technology, Vol. 7 Issue 5, pp. 202-209, May 2018.
9. W. Niyagas, A Srivihok, and S Kitisin, Non-members," Clustering e-Banking Customer using Data Mining and Marketing Segmentation", ECTI Transactions On Computer And Information Technology Vol.2, No.1, pp. 63-69, May 2006.
10. K. I. Moin, Q. B. Ahmed, "Use of Data Mining in Banking", International Journal of Engineering Research and Applications,Vol. 2, Issue 2,Mar-Apr 2012, pp.738-742.
11. A. Raorane and R. V. Kulkarni," Data Mining Techniques: A Source for Consumer Behavior Analysis ", International journal of Core Engineering, September 2011.
12. V. Bhambri, "Application of Data Mining in Banking Sector", International journal of computer science and technology, Vol. 2, Issue 2, June 2011, pp. 199-202.

13. V. Aggelis, "Association rules model of e-banking services ",Journal of Internet banking and Commerce, 2004, pp. 167-173.
14. F. Halili and A Rustemi, " Predictive Modeling: Data Mining Regression Technique Applied in a Prototype", International Journal of Computer Science and Mobile Computing, Vol. 5, Issue. 8, August 2016, pp. 207– 215.
15. S. Jindal, "Study of E-Banking Scenario in India", International Journal of Science and Research, Volume.5 Issue.12, December 2016, pp. 680-683.
16. S. N. Hojjati,A. Monavvarian and F.S. Roozbahani," Evaluation of Factors Influencing Internet Banking Acceptance by Internet Users in Iran", International Journal of Advanced Networking and Applications, Volume 6 Issue.6, 2015, pp. 2503-2519.
17. Y. Dalvadi, "A Study of Factors Affecting Adoption of Internet Banking In India". Vidyasagar University Journal of Commerce, Vol. 20, 2015.
18. T. P. Singh Brar, D. Sharma and Dr. S. Khurmi, "Factors Influencing E-Banking Adoption in India: An Amalgamation of TAM and TPB", International Journal of Computer Science and Technology, Vol. 5 isasue.1, March 2014.
19. G. A. Weshah," The role of internet banking in continuous improvement areas: Quantitative evidence from Jordanian banks" International Journal of Business Performance Management, January 2013, Vol.14, No.2, pp.181- 196.
20. R. Farooqi, "Impact of Internet Banking Service Quality On Customer Satisfaction", Journal of Internet Banking and Com, April 2017, vol. 22 Issue No. 1, pp. 1-17.
21. P. Singh and D. Kumar," Impact of Internet Banking on Customer", International Journal of Research, Vol. 1 issue.4, May 2014, pp. 394-412.