Application of Association Rule Mining in Banking Sector

Vivek Bhambri

Assistant Prof., Department of Computer Applications, A.S.College, Khanna

Abstract

In the present age of globalization, data is considered as most important asset of an organization. But the irony of the situation is that, we are becoming data rich and information poor i.e we have collected a huge amount of data, but are not able to get the required information out of it. The organizations which will be able to convert their data into knowledge and hence will be able to use it for the decision making will rule the world. So the techniques of data mining are proving to be a boon in such context. Data mining is an emerging field and it helps to find out interesting patterns and knowledge from the large amount of data in the transactional and interpersonal database. One of the most important data mining technique is association rule mining whose main purpose is to find frequent patterns, associations and relationship between various database items using different Algorithms. The present paper explores the use of this technique in banking sector. The data collected from the employees of public and private sector banks has been analyzed by applying association rule mining to find the association between the educational qualifications of employee and their readiness to adopt the data mining techniques.

Keywords: Data Mining, Association Rule, Frequent Pattern mining.

I.   Introduction

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. 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. 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. In the present day environment, the huge amount of electronic data is being maintained by banks around the globe. The huge size of these data bases makes it impossible for the organizations to analyze these data bases to retrieve useful information as per the need of the decision makers. Due to the vast expansion of the horizons of the data and its multivariate uses, the organizations and the individuals are feeling a need for some centralized data management and retrieval system. In such a situation, data mining is a boon for one and all.

What is Data Mining?

Data mining methods are expected to provide knowledge previously hidden in the database so as to produce more valuable information which can be used to make the future decisions by taking into considerations the various aspects which can affect the decision.

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.

It is "the science of extracting useful information from large databases".

Data Mining is a process to explore the added value of information that has not been known manually from a database by extracting patterns of data in order to manipulate data into information obtained by extracting and recognizing the important or interesting patterns of data that are present in the database.

A data mining approach comprises of a variety of

techniques and tools to explore, summarize, compare, analyze, forecast and estimate the data for various decisions to be taken by the organization. It provides a methodology for problem solving, analysis, planning, diagnosis, detection, integration, prevention, learning and innovations.

The objective of data mining is to identify valid, novel, potentially useful, and understandable correlations and patterns in existing data.

Data mining is a process that uses statistics, artificial intelligence and machine learning techniques to extract and identify useful information, and subsequent knowledge, from large databases.

The various data mining techniques are: Association Rule Mining, Classification, Clustering, Sequence Path Analysis etc.

II.   Data Mining in Banking Sector

The banks who have realized the importance of data mining are in the process of reaping huge profits and considerable competitive advantage. According to the regulations given by Reserve Bank of India, the banks have to Provide Off-site Monitoring Surveillance (OSMOS) reports on regular basis in electronic format only and Regulatory requirement of filing of statutory returns such as the one under Section 42 of the Reserve Bank

Corresponding author: Vivek Bhambri.

of India Act, 1934 for working out Cash Reserve Ratio (CRR) and Statutory Liquidity Ratio (SLR) obligations in electronic format. According to the Committee formed by Reserve Bank of India Headed by Dr. A. Vasudevan to go through the details of this topic, gave his report on 17th July,1999, the committee highlighted that by the use of data mining techniques, data available at various computer systems can be accessed and by a combination of techniques like classification, clustering, segmentation, association rules, sequencing, decision tree various ALM reports such as Statement of Structural Liquidity, Statement of Interest Rate Sensitivity etc. or accounting reports like Balance Sheet and Profit & Loss Account can be generated instantaneously for any desired period/ date. Trends can be analyzed and predicted with the availability of historical data and the data warehouse assures that everyone is using the same data at the same level of extraction, which eliminates conflicting analytical results and arguments over the source and quality of data used for analysis. In short, data warehouse enables information processing to be done in a credible, efficient manner. The Committee recognizes the need for data warehouses and data mining both at the individual bank level and at industry level. As RBI has instructed the banks to use these technologies. The banks face many problems in implementing this technology; the problems can be managerial, technical or operational. It is to be checked whether, the management is ready to invest on this new technology or not, whether they have the required hardware and softwares and the skilled manpower to run this technology or not. It is to be checked whether the employees are ready to adopt this new technology or not, factors such as size, educational qualification of the employees, their age, sex and the internal needs of an organization are the key factors affecting adoption of the new information technology. Individual differences such as age, gender, education, job function are argued to influence innovativeness which then contributes to the readiness to implement technology.

The readiness towards a particular technology is dependent on:

1. Optimism: A positive view about technology and a belief that it offers people increased control, flexibility and efficiency in their lives.

2. Innovativeness: A tendency to be a technology pioneer and thought leader.

3. Discomfort: A perceived lack of control over technology and a feeling of being overwhelming by it.

4. Insecurity: Distrust of technology and skepticism about its ability to work properly.

Readiness to use the technology can also measured on the basis of *perceived usefulness* and *perceived ease-of-use*, where, Perceived Usefulness is defined as the extent to which a potential adopter views and believes a particular technology can offer value over alternative ways of performing the same task. It is 'the degree to which a person believes that using a particular system would enhance his or her job performance.

Perceived ease-of-use refers to the degree to which a potential adopter views and expects that the usage of a particular technology will be simple and relatively free of effort.

III. Association Rule Mining

Association rule mining is a data-mining technique used to find out the association between various item combinations. The concept of association rule is originated from the market basket analysis which aims to identify frequent item sets in a supermarket shopping. The main idea is to find products that are frequently bought together for improving marketing strategies. Mining association rules is particularly useful for discovering relationships among items from large databases. Association rule mining finds frequent patterns, interesting associations, causal structures or correlations among a large set of data items or objects in transactional database, relational data base or other information repositories. If such associations are made available then it will help the decision makers in making more viable decisions in catalog design, cross-marketing, show room layout and carrying out the analysis of customer shopping behavior etc. The first and the foremost task to find associations among the data items is to find the frequent patterns i.e the patterns which are repeating in the data base again and again. However, the association may be direct or in direct. Direct such as purchasing a pen and paper, that means when the customer buys paper then he/she will buy a pen also, this association will help the organization in designing the layout of store, by placing these two products adjacent to each other, which will lead to convenience to the customer and organization. As these results can be used for designing the pricing decision and offers can be given based on this study. The organization can find that which customer buys which product most of the times together and hence can provide discounts based on the results. It will increase the satisfaction level of the customers and hence churning of customers will reduce.

The various types of associations include:

 - Multilevel association rule.

- Multidimensional association rule

- Quantitative association rule

- Direct association rule.

- Indirect association rule.

An Association rule is an implication of the form A=> B, where A is subset of I, B is subset of I and A != Ø and B != Ø and A ∩ B = Ø. Where Ø means that the set is empty.

The association rules are normally measured with the help of following two measures:

The rule A => B, holds in the transaction set DB with:

Support: It is denoted by s, support of a rule indicates how frequently its items appear in the data base.

It is the percentage of transactions in DB that contains A U B (i.e both A and B), it is taken to be the P(A U B)

Support (A => B) = P (A U B)

Confidence: It is denoted by c, confidence of a rule indicates the probability that if the left hand side appears in T, the right hand side will also appear.

It is the percentage of transaction in DB containing A that also contain B. It is taken to the conditional probability, P (B| A).

Confidence (A => B) = P(A | B)

 The values of support and confidence occur between 0 % to 100% rather than 0 to 1.

 A standard association rule is a rule of the form X→ Y which says that if X is true of an instance in a database, so is Y true of the same instance, with a certain level of significance as measured by two indicators, support and confidence.

The goal of standard association rule mining is to output all rules whose support and confidence are respectively above some given support and coverage thresholds.

The mining process of association rules can be divided into two steps:

Frequent Itemset Generation: In this phase, generate all sets of items that have support greater than a certain threshold, called minsupport.

Association Rule Generation: From the frequent item sets, generate all association rules that have confidence greater than a certain threshold called minconfidence.

The various association rule mining algorithms were used in different applications to determine interesting frequent patterns. One of the association rule mining algorithm such as Apriori algorithm used the property of support and confidence to generate frequent patterns. Apriori Algorithm for Association Rules Association rules are useful for the analysis of customer data.

An association rule can be represented in the form X => Y, indicating that when product X is purchased, product Y is also purchased. X is known as the antecedent and Y is the consequent, such that X triggers the purchase of Y. The algorithms which can be used for market basket analysis are for a generic problem of association: the Apriori algorithm, Frequent-pattern growth algorithm and Eclat algorithm. These algorithms involve two stages. The first stage concerns the discovery of the products which are purchased more frequently. The Apriori algorithm is the most well known association rule mining algorithm.

Corresponding author: Vivek Bhambri.

IV. Methodology Used

In the present study, the data from various public and private sector banks of Punjab, Haryana and Union Territory of Chandigarh has been collected through a questionnaire. The questionnaire tries to measure the extent of usage of IT and data mining in the banking sector, the problems faced by the banks which are using data mining and who are not using data mining, the study tries to explore the feasibility of implementation and problems faced by the banks in implementing such techniques. Further, through the application of association rule mining, the study tries to find out the association between the problems faced by the banks and the type of banks, further it has been analyzed that whether the employees in the banks are ready to use the data mining techniques or not, for this purpose, association rule mining has been used to analyze the association between the educational qualification with the readiness to adopt this technology.

The data from the public and private sector banks has been collected, it is being stored in the excel tables, and then it is being preprocessed to build the classification models, identification of association rules and sequential patterns. After that the strategy is drawn to access the need of data mining in the banking sector. For that purpose the technique of association rule mining is used to test the hypothesis and results are interpreted accordingly. The various steps followed are:

- Data Collection.

- Recording of data in files.

- Preprocessing of data (Filtering, Removing Missing Values and removing the outliers)

- Classification of data (Based on various criterias)

- Applying Association Rules.

 - Drawing Conclusions and

- Testing of Hypothesis

V. Analysis of Data and Interpretation of Results

The customer data comprising of the above mentioned attributes is observed and the techniques of association rule mining have been applied to interpret the various results and hence to find out the associations with in the various variables and how the attributes are associated with each other and how they impact each other.

Association between the Education of the Employee and Readiness towards technology. The association between the educational qualification of the employees and their readiness towards the adoption of a data mining technology is a major area of concern. The educational qualification of the customer has a great impact on the attitudes, psychology and perceptions of the individuals. The educational qualification of the customers has been divided broadly into two categories: technical and non-technical. In technical category, we have considered the employees have some sort of IT related qualification and in non- technical category, we have considered in the employees who do not posses any IT related degree and hence are working in marketing or operations sections of the banks.

Table I

Analysis of Association Rules between "Category of employees (Based on their Educational Qualification) and Readiness towards the technology

Category of employees (Based on their Educational Qualification)	Support (%)	Confidence (%)
Technical	0.84	70.7
Non- Technical	0.33	11

H02: There is no significant difference in average response of respondents having different educational qualification regarding readiness towards the implementation of data mining techniques in banks.

Ha2: There is significant difference in average response of respondents having different educational qualification regarding readiness towards the implementation of data mining techniques in banks.

The analysis of the table I show the association between the educational qualification of the employees and the readiness towards the adoption of technology. It has been observed that customers with technical background are more ready to implement these techniques as compared to the employees with non-technical background. The employees belonging to technical background show a high confidence value, but employees from non-technical background feel disagreement on the issue and show a low confidence value.

The support of first rule is more as compared to second rule, so these two rules imply that there is close association between readiness towards the technology and educational qualification of employees. Based on the analysis, the Null Hypothesis is rejected, and is concluded that the employees having technical qualifications face less problems as compared to the employees belonging to non- technical background.

VI. Findings of the Study: Association rule mining has been used to find the association between the various findings drawn from the data obtained from the bank employees. The important factor which has been analyzed is the readiness towards the implementation of the data mining techniques and their usage, the data has been analyzed w.r.t to educational qualification of employees and it has been observed that the employees having technical qualification are more ready to implement data mining techniques as compared to the employees having non- technical qualification.

VII. Conclusion

Data mining techniques can be of immense help to the organization in solving business problems by finding patterns, associations and correlations which are hidden in the business information stored in the data bases. Organizations can use these techniques for acquiring new customers, fraud detection, providing segment based products for better targeting the customers, analysis of the customers' purchase patterns over time for better

retention and relationship. The implementation and use of association rule mining to comfort the process of finding trends in large data sets will be helpful in effective and efficient decision making. The main purpose is to find the frequent patterns (FP), association, and relationship between various databases using different methods. Upon applying the association rule mining algorithm it has been concluded that the employees having technical qualification are more ready to implement data mining techniques as compared to the employees having non-technical qualification. The results obtained from the present research will help the decision makers in the banks to plan future strategies.

References

C. Angeli, S. K. Howard, J. Ma, J. Yang, and P. A. Kirschner, "Data mining in educational technology classroom research: Can it make a contribution?," *Comput. Educ.*, vol. 113, pp. 226–242, 2017.

M. Zechner and W. Kienreich, "Scalable realtime glyph rendering on consumer level graphics hardware: A hybrid approach," in *Proceedings of the 7th IASTED International Conference on Visualization, Imaging, and Image Processing, VIIP 2007*, 2007.

A. Houari, W. Ayadi, and S. Ben Yahia, "Mining Negative Correlation Biclusters from Gene Expression Data using Generic Association Rules," *Procedia Comput. Sci.*, vol. 112, pp. 278–287, 2017.

Behrouz Minaei-Bidgoli, Elham Akhondzadeh- A New Approach of Using Association Rule Mining in Customer Complaint Management- IJCSI, International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010

A.Vasudevan, "Report of the Committee on Technology Up gradation in the Banking Sector", Constituted by Reserve Bank of India, Chairman of Committee, 1999

P.Isakki alias Devi, S.P.Rajagopalan- Analysis of Customer Behavior using Clustering and Association Rules.

R. Agrawal, and R. Srikant, "Fast algorithm for mining association rules", In Proceedings of the Twentieth International Conference on Very Large Databases, 1994, pp. 487–499.

Bharati M. Ramageri [2010]-Data Mining Techniques and Applications- Indian Journal of Computer Science and Engineering- Vol. 1 No. 4 301-305.

Madan Lal Bhasin, "Data Mining: A Competitive Tool in the Banking and Retail Industries", The Chartered Accountant October 2006

J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Elsevier Science, 2011.

Vivek Bhambri. (2021). Application of Association Rule Mining in Banking Sector.
International Journal of Economic Perspectives,*15*(1),335–343.
Retrieved from https://ijeponline.org/index.php/journal/article/view/61

P. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining", Addison Wesley, ISBN: 0-321-32136-7, 2006.

J. Han and M. Kamber, Data Mining: "Concepts and Techniques, Second Edition", Morgan Kaufman Publisher,2006.

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.

B. Desai and Anita Desai, "The Role of Data mining in Banking Sector", IBA Bulletin, 2004.

S.S.Kaptan, "New Concepts in Banking", Sarup and Sons, Edition, 2002

Rajanish Dass, "Data Mining in Banking and Finance: A Note for Bankers", Indian Institute of Management Ahmadabad.

Corresponding author: Vivek Bhambri.