Classifying Bank Customers Data using Data Mining and Clustering Techniques

Dr.K.Venkateswara Rao[1], U.V.Sravya[2], P.Bharadwaz[3], N.Durga Prasad[4]
[1]Assoc.Prof, [2,3,4]CSE students
Dept. of CSE, LBRCE, Mylavaram, Krishna (dt), AP

Abstract - The Main stay of this work is classification of bank customers data using k-means clustering in data mining. This helps banks to provide different marketing strategies to different groups of customers. It also enables the banks to identify the hidden patterns that are present in huge amount of data. Here we have used k-means clustering algorithm. In this paper we have taken an excel file that contains information of different customers that use services provided by a particular bank. The excel file consists of 19 fields like "Loan Id", "Customer", "Loan Status", "Current Loan Amount", "Term", "Credit Score", "Annual Income", "Years in current job", "Home Ownership", "Purpose", "Monthly Debt", "Years Of Credit History", "Months since last delinquent", "Number of open Accounts" etc., Based on the above specified fields, the customers were classified. Firstly the customers are classified based on the term i.e., short-term and long-term. Then clustering is performed based on the 5 selected fields. Based on these results, the banks will decide to provide the suitable marketing strategies to the companies. The above strategy improves profit.

I. INTRODUCTION

Not only banks but also different organisations should understand the behaviour of their customers in order to provide services more efficiently. Customer relationship management suits well for this purpose. Banks should concentrate more on the customers who have been receiving services from long time .This helps banks to increase their shares and also to find the loyal customers to the bank. Different groups of customers have different features. Based on these features bank should prepare marketing strategies. Now a days there is a severe competition among different banks and organisations. To survive in such a competitive world, classification should be definitely done on the data to find new techniques of marketing. Traditional statistical methods are not useful to discover hidden patterns in the data. Data mining is an advanced technique which is beneficial for the banks to understand customer's requirements and provide them with loans or their required essentials. 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) is known as clustering analysis.

II. RESEARCH THEORITICAL BASES

Classification - The banking system's production is effective when possessions are used for better customers and those who are loyal to the bank receive services for long-term. If the facilities and utilities of bank are good then the productivity will be maximum. Customer relationship management (CRM) helps in the improvement of services of the bank. There are two kinds of customer relationship management. They are operational Customer relationship management and analytical Customer relationship management.The services helps organisations to take care of their customers comes under the category of operational CRM. Sales, service, marketing and other business processes are supported by operational CRM. Web sites, data aggregation systems and call centres are a few examples of operational CRM. Middle database stores and organises the data. Organizational responses that are required by customers are expanded with the help of operational CRM. The second type is analytical CRM(ACRM).ACRM is more effective in promoting marketing strategies. Pattern discovery association rules, databases, sequential patterns, and prediction motors supports ACRM. Banks can find hidden patterns and can get necessary details about customers with the help of ACRM.

Clustering - Data mining helps to find information from large databases. Data mining helps in making decisions. The techniques in data mining consists of Association, Classification, Clustering, Succession models, Artificial neural networks, Regression, Decision trees. These help in discovering future trends depending on past trends. Clustering means grouping a set of similar data objects into a set of meaningful sub-classes/groups, called clusters. There are two types of data mining techniques. They are supervised learning and unsupervised learning. Clustering comes under unsupervised learning technique. Classification comes under supervised learning because here, the number of classes and their features are known previously and data is kept in pre-defined set of classes. we don't know about the number of classes and their features previously, in the case of clustering and related data is grouped into clusters. There are two groups of Clustering algorithms. They are non-hierarchical clustering and hierarchical clustering. Hierarchical clustering methods are of two types: top-down methods and bottom-up methods. Top-down method begins with a cluster and this large cluster is divided into a number of small clusters Smaller clusters are combined at each

stage to form a larger cluster in bottom-up clustering. In the data, the elimination of unnecessary variables is done with the help of Two-step clustering algorithm. The data is divided into 'k' number of clusters using K-means algorithm.

III. RESEARCH METHODOLOGY

The variables that fits best for the clustering are identified. Along with this, the determination of number of clusters for k-means algorithm also took place. The data Preprocessing techniques are applied on the dataset which is an excel file, to escape from the problem of incomplete data. Then the data is prepared for implementation of data mining. The dataset which is an excel file has 19 data fields which contains like Loan Id ,Customer, Loan Status, Current Loan Amount, Term, Credit Score, Annual Income, Years in current job, Home Ownership, Purpose, Monthly Debt ,Years Of Credit History, Months since last delinquent, Number of open Accounts, Number of Credit Problems, Current Credit Balance, Maximum Open Credit, Bankruptcies, Tax Liens.

Step-1: The customers are categorised as short term and long term.

Figure 1: GUI of Long-term and Short-term

Step-2: Among all the attributes, five sensitive attributes are identified for long term customers.

Figure 2: Sensitive Attributes

Step-3: Firstly the attributes "Annual Income" and "Monthly Debt " are considered and the customers who are paying less than or equal to 25 percentage of their total salary are considered. Then the result obtained after this step is taken as input for the second step.

Figure 3: Cluster 1 Result

In this step the attributes such as "CibilScore","Months Since last Delinquent" and "No of credit problems" are considered and then the result is obtained.

Figure 4: Identification of Best Customers

Now different types of services are provided to those customers who are selected after these two steps.These services depends on the type of the customers.

Figure 5: Provision of Services

IV. CLUSTERING ANALYSIS

If all the variables are inserted in the clustering algorithm the quality of resultant clusters will be poor.We first identify the variables that are well suitable for clustering .Only some variables are given as input to the k-means clustering algorithm. So finally we select 5 variables out of the 19 variables. They are
1) Annual Income
2) Credit Score
3) Monthly Debt
4) Months since last delinquent
5) Number of credit problems

Based on the 5 variables used in the k-means clustering algorithm, different services are provided to different customers based on the results of the k-means algorithm.

Advantages -
1. Cost of providing services is reduced.
2. Better conditions can be created.
3. Customer orientation culture can be expanded
4. Production sector customers were supported as bank loyal customers

V. CONCLUSION

This segmentation helps to find the loyal customers. This helps the organisation to communicate effectively with their loyal customers and provide best marketing strategies to them which improves the profitability of the organization. Customer satisfaction strategies helps to find the behaviour of the and trust worthiness of the customer i.e., whether the customer will remain in the organisation or whether the customer leaves the organisation. Locking strategy is applied on the new customers. It prevents the customers from leaving the organisation. In this study, we used methods in clustering along with variables to categorise the customers legally.

VI. REFERENCES

[1]. J.Han,M. Kamber,"Data Mining: Concepts and Techniques", Harcourt India / Morgan Kauffman, 2001.
[2]. PaulrajPonniah,"Data Warehousing Fundamentals", Wiley-Interscience Publication,2003
[3]. https://www.tutorialride.com/data- mining/clustering-in-datamining.html
[4]. https://www.statisticssolutions.com /directory-of-statistical-analyses- cluster-analysis/