Data Analysis of Customer Complaints of Loan Approval and Financing in Banking Industry using SVM Classification

[1]A.P. Gopu, [2]S. Manikandan, [3]M. Nandhagopal, [4]V. Boopathi

[1]Assistant Professor, [2,3,4]UG Students – Final Year, Department of Computer Science and Engineering, Nandha College of Technology, Perundurai - 638052, Tamilnadu, India.

Abstract--This paper focus on discovering and analysing Customer Finance Complaints data, to find how many similar complaints are there in relation to the same bank or service or product. These datasets fall under the complaints of Credit reporting, Hypothecation, Dues Collection, Customer Loan and Finance Accounting. By using data mining techniques, cluster analysis as well as predictive modelling is applied to obtain valuable information about complaints in certain areas of the Country. The banks that are receiving customer complaints filed against them will analyse the complaint data to provide results on where the most complaints are being filed, what products services are producing the most complaints and other useful data. Our model will contribution banks in identifying the location and types of mistakes for resolution, important to increased customer satisfaction to drive income and profitability.

Keywords: **Customer, Complaints, analysis, clustering, predictive.**

I. INTRODUCTION

As we are awake that in today's modern world peoples are more into business, so receiving a complaint from a customer happens almost every day. A customer's complaints present bank or reporting agency with a prospect to detect and resolve specific problems with their current product and service. Service complaints management is a critical part of business management. A good complaint- management strategy will result in best customer relationship with least human-resource investment and so hope to find a connection between bank, and customers to refine bank applications to better accommodate customer needs. Progressively banks are identifying the value of a customer complaint in that it is feedback on their experience, and an opportunity to not only resolve a problem for that particular customer but feasibly also for a much larger number of customers and that leads to unavoidable amounts of data that has to be analysed and specific functions are used to aggregate the analysis results.

A customer's complaints present bank or reporting agency with an opportunity to identify and rectify specific problems with their current product or service. The banks that are receiving customer complaints filed against them will analyse the complaint data to provide results on where the most complaints are being filed, what products services are producing the most complaints and other useful data. This project assists banks in identifying the location and types of errors for resolution, leading to increased customer satisfaction to drive income and profitability. This project finds a connection between complaints, banks and customers to refine bank applications to better accommodate customer needs using a hybrid approach of hierarchal and k-means clustering.

In addition, using SVM classification, the complaints sentiment values are analysed and classified into positive or negative reviews. The project is designed using R Studio 1.0 as front end. The coding language used is R version 3.4.4.

II. LITERATURE REVIEW

The number of studies has been conducted regarding the services to customers and their awareness. As such, we have reviewed some of them.

Venkata Rao Maddumala(2020) determined accuracy of loan approval based on machine learningtechnique successfully.

Soni P M, Varghese Paul(2019) predicted a loan credibility system using data mining concept successfully.

Kamakodi (2007) concluded that modern day generation is influenced by the computation features used by banks and so the banks study about factors influencing their preferences. Residence relocation, salary fluctuation and unavailability banking based services are reasons enough to change bank.

Uppal and Kaur(2007) determined how consumer's awareness of web domains used by banks and gave some measures to make these applications more successful. They concluded that the limitation about today's web domain application is spreading the awareness about the varied features offered.

Mishra and Jain (2007) took up dimensions of consumer satisfaction in national and private banks. The study talks about how satisfaction is the foremost asset to the organization, which provides unmatched competitive edge that helps achieving loyalty of a customer.

Singh (2006) discusses CRM approaches in various banks. He emphasized on how the management targets customers in order to gain insight and gives out value added services and products. Web as provided a smooth user experience, giving access to the various features used by the customers thereby achieving customer satisfaction. Management has to strive to ensure end to end delivery and ensure customer satisfaction which is essential to the banks in terms of maintaining high regards and loyalty obtained from customers.

Jain and Jain (2006) demonstrated that the banking sector, both private and public have suffered radical as well as revolutionary changed due to the liberalization act of 1991. Retail banking is the consumer preferred choice which articulates itself responses received from 200 customers of HDFC bank, ICICI bank and some other banks in the city of Varanasi, Uttar Pradesh and he looked upon the schemes offered by the banks, quantized satisfaction in different types of services, expectations about theseschemes and the height of segmentation among the services offered.

Bhaskar (2004) computed that expansion of banking is directly proportional to the quality of services provided by the banks and satisfaction is regarded highly as customer's feedback is the only thing to lean on, when it comes to the highly competitive banking industry. Arguably, India's banking industry is highly thriving and depends heavily on customer morale and loyalty.

Although Singh (2004) spurred about the reality of banks in terms of providing customer support and found out that the customers are influenced by the banks location and the minutest

detail of the banking details including the banking interest rates as well as attitudes and customer supportprovided by the personnel.

Furthermore, Hasanbanu (2004) stated how the rural India is unaware about various schemes and benefits offered by the banks in order to ensure financial welfare. The majority of rural population is inaccessible to the web domain services of the banks and continue to prefer local moneylenders charging light interest rates. The study was conclusive and based on the data provided by the RBI, however, it isbased on the questionnaire and surveys.

III. METHODOLOGY AND PROCEDURE

The following modules are present in the project.

1. K-MEANS CLUSTERING

2. SUPPORT VECTOR MACHINE(SVM)

3. NEURAL NETWORK BASEDCLASSIFICATION

4. SIMILARITY PERCENT FINDING

A. K-MEANS CLUSTERING:

In this module, K-Means clustering of bank customer data are given as input to analyse and group the customers. The customer data set is taken with Customer Id, Annual Income, Loan Type, Interest Rate, Duration in Months, Loan Amount Requested and Loan Status columns of which Interest Rate and Duration in Months are taken as X and Y Axis for K- Means clustering. By default 3 is given for K, but wecan give any number to cluster the data.

1. Initialize cluster centroids $\mu_1, \mu_2, \quad \mu_k \in R^n$ randomly.

2. Repeat until convergence: {

For every i, set $C^{(i)} := {}^{\text{arg min}} ||x^{(i)} - \mu||^2$

For each j, set $\mu_j := \underline{\sum_{i=1}^m 1\{c^{(i)}=j\}x^{(i)}}$

$\sum_{i=1}^m 1\{c^{(i)}=j\}$

SVM Classification:

"Support Vector Machine" (SVM) is a supervised which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.
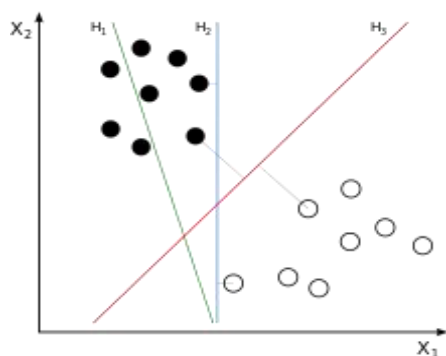

Figure1: H_1 does not separate the classes. H_2 does, butonly with a small margin. H_3 separates them with themaximal margin.

B. Neural network algorithm:

A neural network is a series of algorithms that endeavours to recognize underlying

relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature.
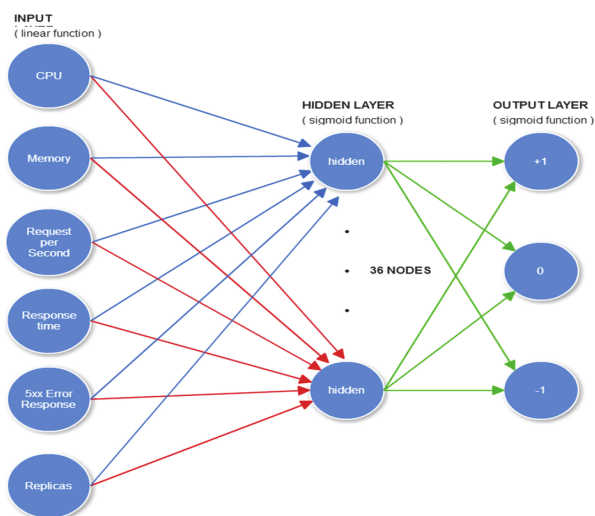

Figure2: The artificial neural network that is used in our scaling algorithm.

C. SIMILARITY PERCENT FINDING:

To find the nearest matching record values among the given training record set, hot encoding vector comparison is also made so that similarity percent of the test record in the overall training dataset records are also found out. The test record is classified into maximum similarity matching record.

IV. RESULTS

By performing K- Means, support vector machine(SVM) for positive negative analysis and neural network algorithm to find similarity percentage, we got a better insight by having 3 clusters. Figure1 shows clusters in having the highest density while the rest of the have a lower density in comparison. Figure2 plots below show us the results for our clustering model.
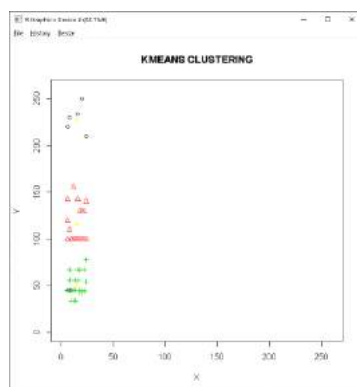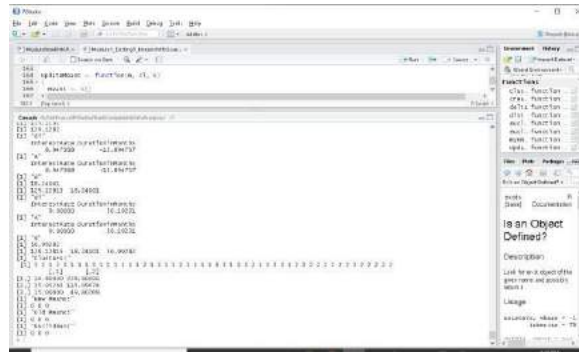

Figure3: K-MEANS CLUSTENIG

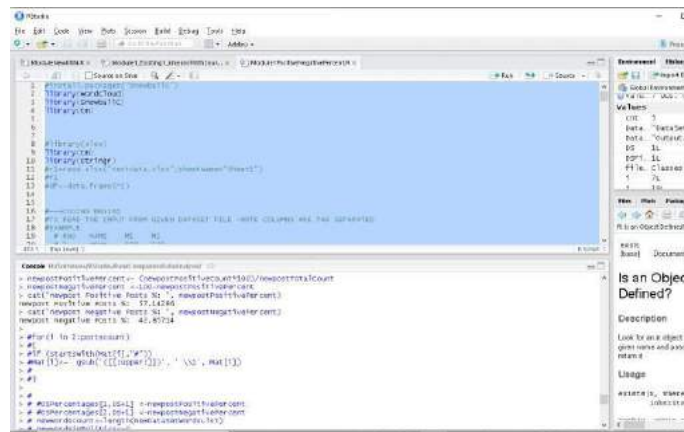**Figure4:** RECORDS GROUPED IN CLUSTERS

Figure5: POSITIVE/NEGATIVE POSTS PERCENTAGE
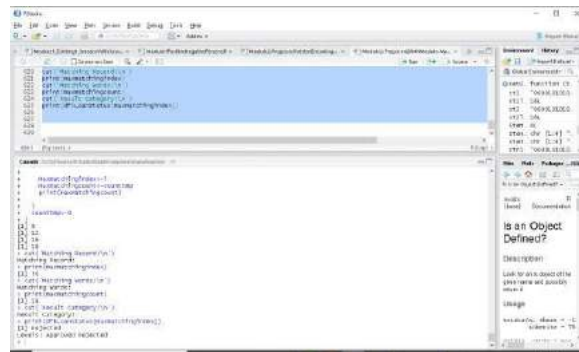
Figure8: MATCHING RECORD IN THE TRAINING DATA FOR THE   GIVEN TEST DATA.CONCLUSION:
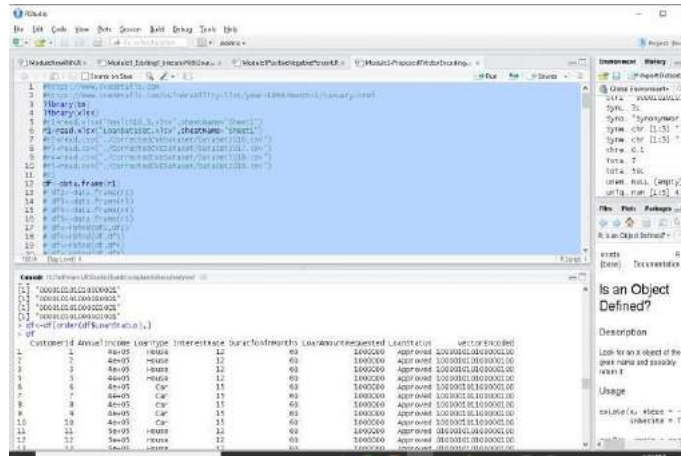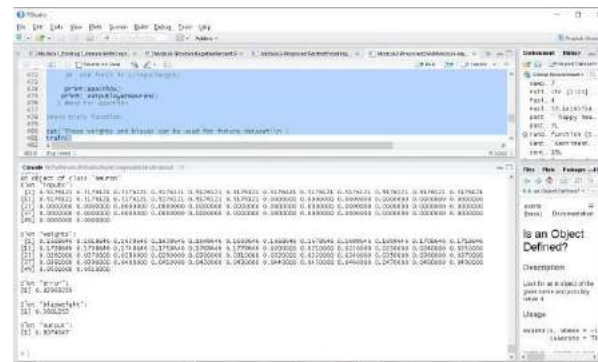
Figure6: VECTOR ENCODED VALUES

Figure7: NEURAL NETWORK MODEL[WEIGHTS AND BIASES]

The project has explored various concepts of data mining of bank customers' data set and the results show what problems customers are having with specific problems in particular loan provisions. This valuable information will show where companies will need to invest in to improve their overall performance in the view of their customers. This will lead to improved customer satisfaction. By maximizing customer satisfaction, the opportunity for repeat sales to customers can be increased.

Customer satisfaction also helps to increase customer loyalty, reducing the need to allocate marketing budget to acquire new customers.

Satisfied customers may also recommend your products or services to other potential customers, increasing the potential for additional revenue and profit. Future research can be done to collect more complaint data so that we can perform analysis for other products, services and companies. Since neural network is applied, the weight values and bias values are calculated which can be used forfurther records from customers in classifying the data.

REFERENCES

[1] Goyal S, Thakur KS (2008). A Study of Customer Satisfaction Public and Private Sector Banks of IndiaPunjab, J. Bus. Stud., 3(2): 121- 127.

[2] Uppal RK (2007). Customer Service in Banks- An Empirical Study', Bankers Conference Proceedings,pp. 36-42.

[3] Kamakodi N (2007). Customer Preferences on e- Banking ServicesUnderstanding through a Sample Survey of Customers of Present Day Banks in IndiaContributors, Banknet Publications, 4: 30-43.

[4] Mishra JK, Jain M (2007). Constituent Dimensions of Customer Satisfaction: A Study of Nationalized and Private Banks Prajnan, 35(4): 390-398.

[5] Jain AK, Jain P (2006). Customer Satisfaction in Retail Banking Services NICE, J. Bus. Stud., 1(2):95-102.

[6] Singh SB (2006). Customer Management in BanksVinimaya, 37(3): 31- 35.

[7] Bhaskar PV (2004). Customer Service in Banks IBABulletin, 36(8): 9- 13.

[8] Hasanbanu S (2004). Customer Service in Rural Banks: An Analytical Study of Attitude of Different types of Customers towards Banking Services IBA Bulletin, 36(8): 21-25.

[9] Singh S (2004). An Appraisal of Customer Serviceof Public Sector Banks IBA Bulletin, 36(8): 30-33.

[10]Shankar AG (2004). Customer Service inBanks IBA Bulletin, 36(8): 5- 7.

[11]Ganesh C, Varghese ME (2003). Customer Service in Banks: An Empirical Study'.Vinimaya, 36(2): 14-26.