March-April 2020
ISSN: 0193-4120 Page No. 1482 - 1484

An Exploratory Data Analysis for Loan Prediction of Customers Based on Data Mining Approaches

[1]Mani Teja K, [2]R. Senthil Kumar
[1]Student, [2]Assistant Professor, Department of CSE, Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences, Chennai, India
[1]manitejakamineni667@gmail.com, [2]rsenthilmecse@gmail.com

Article Info
Volume 83
Page Number: 1482 - 1484
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 15 March 2020*

Abstract
The banking sector helpful for securing loan to customers by cross checking all the prerequisite documents required to process. DataMining classification algorithms helps to classify the loan applications and helpful automatic selection of customers who are eligible to get loan This paper is split into four sections (i)Data preprocessing (ii) Training the dataset and on collected information (iii) Applying the decision tree algorithms(iv) Testing the results. In this paper we introduce an effective prediction model for the bank that may help them predict the good customers who have applied for loan. Decision Tree data mining Algorithm is applied to predict the attributes relevant for credibility. A prototype of the model is described in this paper which can be used by the organizations in making the right decision to approve or reject the loan request of the customers.

Keywords: *Datasets, Classification, Data mining, Machine Learning, Prediction.*

1. Introduction

Distribution of the loans is that the core business a part of nearly each banks. The most portion the bank's assets is directly came from the profit attained from the loans distributed by the banks. The prime objective in banking atmosphere is to take a position their assets in safe hands wherever it's. Nowadays several banks/financial corporations approves loan once a regress method of verification and validation however still there's no surety whether or not the chosen applier is that the worthy right applier out of all candidates. Through this method we are able to predict whether or not that exact applier is safe or not and therefore the whole method of validation of options is automatic by machine learning technique. The disadvantage of this model is that it emphasize totally different weights to every issue however in reality someday loan will be approved on the idea of single robust issue solely, that isn't potential through this method. Loan Prediction is incredibly useful for worker of banks in addition as for the applier conjointly. The aim of this Paper is to supply fast, immediate and

straightforward thanks to select the worthy candidates.. The Loan Prediction System will mechanically calculate the load of every options collaborating in loan process. Loan Prediction System permits jumping to specific application in order that it will be check on priority basis. This Paper is completely for the managing authority of focus on the Acceptance (or) Deny of the and on new here it or not his/her loan will be sanctioned or check information same options are processed with relevance their associated weight.

2. Literature Survey

The Loan Prediction is the Process of applying loan process in various commercial banks. in banking sector as it is provides the most benefit plans to attract the loan applicants to their banks, the every applicant who applied for the loan has to be verified to know the his state that he can capable of repaying the Loan or not, as the Applicants increase the man force for the verification also required to overcome with this, Banking sector seek a algorithm to find the set of persons who are capable to get loan and repay them. As it requires the datasets with information of various attributes to predict the applicant eligible to get loan or not. We use the classification of decision tree algorithm with the help of data mining techniques, as classification and decision tree process

March-April 2020
ISSN: 0193-4120 Page No. 1482 - 1484

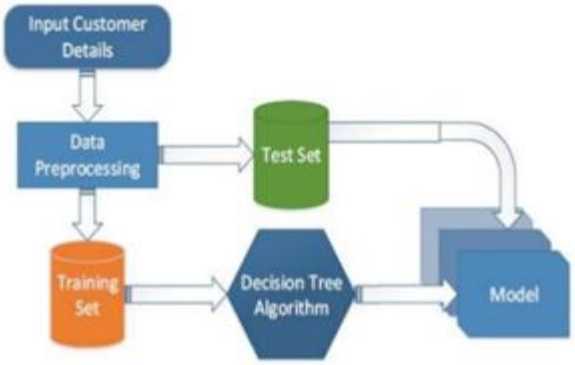
works very effectively than the matrices process and decision tree classification is also very secure process than the matrices classification because the classification of decision tree involves machine learning, it involves two types of learning the Supervised learning and the Unsupervised learning, the classification involves the supervised learning process, it means the data set itself has key in it and train the data set with help of the machine learning techniques.

As this involves the decision classification and machine learning process, the algorithm is so accurate as it uses tree classification techniques in prediction and the data is secure as data is trained with the machine learning techniques and it Data mining techniques are also involved in the Program tree classification algorithm involves the program or algorithm to make the code efficient.

3. Working Principle

The Loan Prediction System can automatically calculate the weight of each features taking part in loan processing and on new test data same features are processed with respect to their associated weight .A time limit can be set for the applicant to check whether his/her loan can be sanctioned or not. Loan Prediction System allows jumping to specific application so that it can be check on priority basis. This Paper is exclusively for the managing authority of Bank/finance Company, whole process of prediction is done privately no stakeholders would be able to alter the processing.

4. System Architecture



5. System Module

The data preprocessing involves Removal of unwanted attributes from our dataset which are not useful for association rule mining such as order id, hour of day etc. Then we have replaced the quantity of items in transactions which are greater than 1 by 1(since to generate association rules in weka we need the presence of item in the transaction but not the quantity). Next we have replaced all 1's in transaction by 't' and 0's by '?' .We created a dataset (items with price) with attributes item name and price for all 134 items and wrote its price in the data set.

Prediction based on generated decision tree:

We will calculate the class of the current transaction based on items of the transaction and then predict the missing items by checking the current transaction items with the antecedent of generated rules of same class.

Prediction based on training datasets:

All combo offers are loaded into database and every item of current transaction is crosschecked with the offer and if any offer is.

Prediction based on loans available:

If any item has a discount offer such as 50% off and if current transaction is having that item then we'll reduce the class of that item by 1 (if its original class is greater than 1 and discount is greater than 50%) and then calculate the current transaction class.

Prediction based on Customer history:

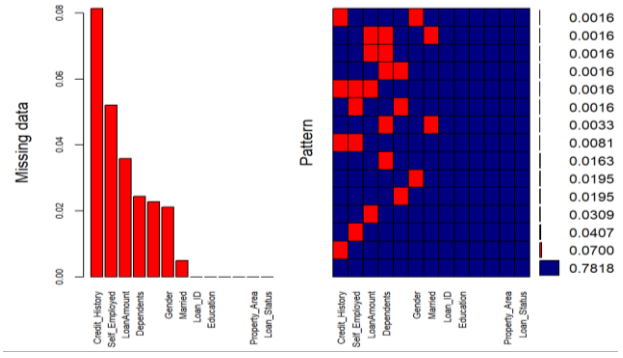
We will store all the previous transactions of customers and predict missing items from them. We will calculate probability of a particular customer buying an item and if the probability is greater than 0.5 and if that item is missing in his/her cart then we'll recommend that item to the customer(user based prediction).

Creation of interface using JFrames:

We have created an interface using Jframes which will take inputs such as customer id, item names along with its quantity and can also delete an item or can reduce its quantity and invoke the java code which would predict missing items.

6. Results

The prediction accuracy obtained through classification algorithm is more using the decision tree algorithm (91.80 %) is higher than the accuracy of the individual classification algorithms. The historical data of the customers like their age, income, loan amount, employment length etc. will be used in order to do the analysis. Prediction based on generated decision tree training datasets, loans available and Customer history.



7. Conclusion

The main purpose of the paper is to classify the acceptance or rejection of loan application based on proper analysis of data set and parameters of the banking sector, each data have been generated and processed through visualization. This paper work may be extended to future with Predictive model for loans that uses machine learning algorithms across multi users with different considerations.

References

[1] Z. Yang, Y. Zhang, B. Guo, B. Y. Zhao, and Y. Dai, "DeepCredit:Exploiting User Cickstream for Loan Risk Prediction in P2P Lending," in Twelfth International AAAI Conference on Web and Social Media,2018.

[2] J. H. Aboobyda and M. A. Tarig, "Developing Prediction Model Of Loan Risk In Banks Using Data Mining Machine Learning andApplications," An Int. J., vol. 3, no. 1, pp. 1–9, 2016.

[3] G. Sudhamathy, "Credit risk analysis and prediction modelling of bankloans using R," Int. J. Eng. Technol, vol. 8, pp. 1954–1966, 2016

[4] J. R. Quinlan. Induction of Decision Tree. Machine Learning, Vol. 1, No. 1. pp. 81-106., 1086.

[5] Mean Decrease Accuracy https://dinsdalelab.sdsu.edu/metag.stats/code/rando mforest.hml

[6] J.R. Quinlan. Induction of decision trees. Machine learning Springer, 1(1):81–106, 1086.

[7] Andy Liaw and Matthew Wiener. Classification and Regression by random Forest. R News (http://CRAN.R-project.org/doc/Rnews/), 2(3):9–22, 2002.

[8] S.S. Keerthi and E.G. Gilbert. Convergence of a generalizeSMO algorithm for SVM classifier design. Machine Learning, Springer, 46(1):351–360, 2002.

[9] J.M. Chambers. Computational methods for data analysis. Applied Statistics, Wiley, 1(2):1–10.