International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 01 | Jan 2021 www.irjet.net p-ISSN: 2395-0072

LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING

Deepak Ishwar Gouda[1], Ashok Kumar A[2], Anil Manjunatha Madivala[3], Dilip Kumar R [4],

Dr.Ravikumar[5]

[1,2,3,4]B.E. Student, Department of CSE, Sir M Visvesvaraya Institute of Technology, VTU, Bengaluru, India
[5]Associate Professor, Department of CSE, Sir M Visvesvaraya Institute of Technology, VTU, Bengaluru, India

---

Abstract - *The enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this paper we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this paper is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing.*

***Key Words*: Loan Prediction, Big data, Machine Learning, Logistic Regression, SVM, Decision Tree, Naïve Bayes.**

1. INTRODUCTION

Distribution of the loans is the core business part of almost every banks. The main portion the bank's assets is directly came from the profit earned from the loans distributed by the banks. The prime objective in banking environment is to invest their assets in safe hands where it is. Today many banks/financial companies approves loan after a regress process of verification and validation but still there is no surety whether the chosen applicant is the deserving right applicant out of all applicants. Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique. The disadvantage of this model is that it emphasize different weights to each factor but in real life sometime loan can be approved on the basis of single strong factor only, which is not possible through this system.

Loan Prediction is very helpful for employee of banks as well as for the applicant also. The aim of this Paper is to provide quick, immediate and easy way to choose the deserving applicants. It can provide special advantages to the bank. 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. Result against particular Loan Id can be send to various department of banks so that they can take appropriate action on application. This helps all others department to carried out other formalities.

2. LITERATURE SURVEY

In 2019, Vimala and Sharmili [1] proposed a loan prediction model using NB and Support Vector Machines (SVM) methods. Naïve Bayes, an independent speculation approach, encompasses probability theory regarding the data classification. On the other hand, SVM uses statistical learning model for classification of predictions. Dataset from UCI repository with 21 attributes was adopted to evaluate the proposed method. Experimentations concluded that, rather than individual performances of classifiers (NB and SVM), the integration of NB and SVM resulted in an efficient classification of loan prediction.

In 2019, Jency, Sumathi and Shiva Sri [2] proposed a Exploratory Data Analysis (EDA) regarding the loan prediction procedure based on the client's nature and their requirements. The major factors concentrated during the data analysis were annual income versus loan purpose, customer's trust, loan tenure versus delinquent months, loan tenure versus credit category, loan tenure versus number of years in the current job, and chances for loan repayment versus the house ownership. Finally, the outcome of the present work was to infer the constraints on the customer who are applying for the loan followed by the prediction regarding the repayment. Further, results showed that, the customers were interested more on availing short-tenure loans rather than long-tenure loans.

In 2019, Supriya, Pavani, Saisushma, Vimala Kumari and Vikas [3] presented a ML based loan prediction model. The modules in the present approach were data collection and pre-processing, applying the ML models, training followed by testing the data. During the pre-processing stage, the detection and removal of outliers and imputation removal processing were carried out. In the present method, SVM, DT, KNN and gradient boosting models were employed to predict the possibilities of current status regarding the loan approval process. The conventional 80:20 rule was adopted to split the dataset into training and *testing processes. Experimentation concluded that, DT has significantly higher loan prediction accuracy than the other models.*

In 2017, Goyal and Kaur [4] presented a loan prediction model using several Machine Learning (ML) algorithms. The dataset with features, namely, gender, marital status, education, number of dependents, employment status, income, co-applicant's income, loan amount, loan tenure, credit history, existing loan status, and property area, are used for determining the loan eligibility regarding the loan sanctioning process. Various ML models adopted in the present method includes, Linear model, Decision Tree (DT), Neural Network (NN), Random Forest (RF), SVM, Extreme learning machines, Model tree, Multivariate Adaptive Regression Splines, Bagged Cart Model, NB and TGA. When evaluated these models using R Environment in five runs, TGA resulted in better loan forecasting performance than the other methods.

In 2016, Aboobyda Jafar Hamid and Tarig Mohammed Ahmed [5] presented a loan risk prediction model based on the data mining techniques, such as Decision Tree (J48), Naïve Bayes (NB) and BayseNet approaches. The procedure followed was training set preparation, building the model, Applying the model and finally. Evaluating the accuracy. This approach was implemented using Weka Tool and considered a dataset with eight attributes, namely, gender, job, age, credit amount, credit history, purpose, housing, and class. Evaluating these models on the dataset, experimental results concluded that, J48 based loan prediction approach resulted in better accuracy than the other methods.

In 2016, Kacheria, Shivakumar, Sawkar and Gupta [6] suggested a loan sanctioning prediction procedure basded on NB approach integrated with K-Nearest Neighbor (KNN) and binning algorithms. The seven parameters considered were income, age, profession, existing loan with its tenure, amount and approval status. The sub-processes include, Pre-processing (handling the missing values with KNN and data refinement using binning algorithm), Classification using NB approach and Updating the dataset frequently results in appropriate improvement in the loan prediction process. Experimentation put-forth the conclusion that, integration of KNN and binning algorithm with NB resulted in improved prediction of loan sanctioning process.

In 2016, Goyal and Kaur [7] suggested an ensemble technique based loan prediction procedure for the customers. The sub-processes in the present method includes, data collection, filtering the data, feature extraction, applying the model, and finally analysis the results. The various loan prediction procedures implemented in the present method were Random Forest (RF), SVM and Tree model with Genetic Algorithm (TGA). The parameters considered for evaluating the models were accuracy, Gini Coefficient, Area Under Curve (AUC), Receiver Operating Curve (ROC), Kolmogorov - Smirnov (KS) Chart, Minimum Cost - Weighted Error Rate, Minimum Error Rate, and K-Fold Cross Validation parameters. Experimentation outcome concluded that the integration of three methods (RF, SVM and TGA) resulted in improved loan - prediction results rather than individual method's prediction.

In 2006, Sudhamathy [8] suggested a risk analysis method in sanctioning a loan for the customers using R package. The various modules include data selection, pre-processing, feature extraction and selection, building the model, prediction followed by the evaluation. The dataset used for evaluation in this method was adopted from UCI repository. To fine tune the prediction accuracy, the pre-processing operation includes the following sub-processes: detection, ranking and removal of outliers, removal of imputation, and balancing of dataset by proportional bifurcation regarding testing and training process. Further, feature selection process improves the prediction accuracy. When evaluated, the DT model resulted in 94.3% prediction accuracy.

The process of analyzing data from different perspectives and extracting useful knowledge from it. It is the core of knowledge discovery process. The various steps involved in extracting knowledge from raw data. Different data mining techniques include classification, clustering, association rule mining, prediction and sequential patterns, neural networks, regression etc. Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to classification technique. This approach frequently employs Decision tree based classification Algorithm. In classification, a training set is used to build the model as the classifier which can classify the data items into its appropriate classes. A test set is used to validate the model.

3. CONCLUSION

This application is working properly and meeting to all Banker requirements. This component can be easily plugged in many other systems. It works correctly and fulfills all requirements of bankers and can be connected to many other systems. There were multiple malfunctions in the computers, content errors and fixing of weight in computerized prediction systems. In the near term, the banking software could be more reliable, accurate, and dynamic in nature and can be fit in with an automated processing unit. There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system more secure, reliable and dynamic weight adjustment. The system is trained on old training dataset in future software can be made such that new testing date should also take part in training data after some fix time. Machine learning helps to understand the factors which affect the specific outcomes most. Other models like neutral network and discriminate analysis can be used individually or combined for enhancing reliability and accuracy prediction.

REFERENCES

[1] S. Vimala, K.C. Sharmili, -Prediction of Loan Risk using NB and Support Vector Machine||, International Conference on Advancements in Computing Technologies (ICACT 2018), vol. 4, no. 2, pp. 110-113, 2018.

[2] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, K. Vikas, -Loan Prediction by using Machine Learning.

[3] X. Francis Jency, V.P.Sumathi, Janani Shiva Sri, -An Exploratory Data Analysis for Loan Prediction Based on Nature of the Clients||, International Journal of Recent Technology and Engineering (IJRTE), Vol. 7, No. 48, pp. 176-179, 2018.

[4] Anchal Goyal, Ranpreet Kaur, -Loan Prediction Using Ensemble Technique||, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 3, pp. 523 – 526, March 2016.

[5] Aboobyda Jafar Hamid and Tarig Mohammed Ahmed, -Developing Prediction Model of Loan Risk in Banks using Data Mining||, Machine Learning and Applications: An International Journal (MLAIJ), Vol.3, No.1, pp. 1-9, March 2016.

[6] Aditi Kacheria, Nidhi Shivakumar, Shreya Sawkar, Archana Gupta,Loan Sanctioning Prediction System, International Journal of Soft Computing and Engineering (IJSCE), vol. 6, no. 4, pp. 50-53, 2016.

[7] Anchal Goyal, Ranpreet Kaur, -Accuracy Prediction for Loan Risk using Machine Learning Models||, International Journal of Computer Science Trends and Technology (I JCST), Vol. 4, Issue 1, pp. 52-57, Jan- Feb 2016.

[8] Sudhamathy, Credit Risk Analysis and Prediction Modelling of Bank Loans using R||, International Journal of Engineering and Technology (IJET), Vol. 8, No. 5, pp. 1954-1966, Oct-Nov 2016.

[9] Rattle data mining tool: available from http://rattle.togaware.com/rattle-download.html

[10] Aafer Y, Du W &Yin H 2013, DroidAPIMiner: 'Mining API-Level Features for Robust Malware Detection in Android', in: Security and privacy in Communication Networks Springer, pp 86-103.

[11] Ekta Gandotra, Divya Bansal, Sanjeev Sofat 2014, 'Malware Analysis and Classification: A Survey'available from http:// www.scirp.org/journal/jis

[12] K. Hanumantha Rao, G. Srinivas, A. Damodhar, M. Vikas Krishna: Implementation of Anomaly Detection Technique Using Machine Learning Algorithms: Internatinal Journal of Computer Science and Telecommunications (Volume2, Issue3, June 2011).