Applicant Credibility Prediction For Loan Approval

Introduction

As the data are increasing daily due to digitization in the banking sector, people want to apply for loans through the internet. Artificial intelligence (AI), as a typical method for information investigation, has gotten more consideration increasingly. Individuals of various businesses are utilizing AI calculations to take care of the issues dependent on their industry information. Banks are facing a significant problem in the approval of the loan. Daily there are so many applications that are challenging to manage by the bank employees, and also the chances of some mistakes are high.Most banks earn profit from the loan, but it is risky to choose deserving customers from the number of applications.One mistake can make a massive loss to a bank. Loan distribution is the primary business of almost every bank. This project aims to provide a loan [1, 8] to a deserving applicant out of all applicants. An efficient and non-biased system that reduces the bank's time employs checking every applicant on a priority basis. The bank authorities complete all other customer's other formalities on time, which positively impacts the customers. The best part is that it is efficient for both banks and applicants. This system allows jumping on particular applications that deserve to be approved on a priority basis. There are some features for the prediction like,Gender, Married, Dependents, Education, Self-employed, ApplicantIncome, CoapplicantIncome, LoanAmount, Loan_Amount_Term, Credit_History, Property Area, Loan_Status.Loans have made our life easier, providing us the financial leverage that extends beyond our earnings. Be it Credit Card, Home Loan, Personal Loan or Auto Loan etc. loans are the credit extended to us by lenders on fulfilling certain key parameters. However, getting a loan in India can often be a tedious process for the un-initiated, but not for individuals with a good credit score. Whenever you apply for a loan, banks check your CIBIL Score and Report to evaluate your credit history and credit worthiness. The higher your score the better are the chances of your loan application getting approved.

Literature Review

[1] In 2019, Vimala and Sharmili proposed a loan prediction model using 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.

Advantages: 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.

Disadvantages: SVM algorithm is not suitable for large data sets. In cases where the number of features for each data point exceeds the number of training data samples, the SVM will underperform.

[2] In 2019, Jency, Sumathi and Shiva Sri proposed an 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

Advantages: Libraries helps to analyse the data. Statistical and prediction is very easy comparing to existing technologies. Results will be accurate compared to other methodologies.

Disadvantages: Complexity in analyzing the data. Prediction is challenging task working in the model. Coding is complex maintaining multiple methods. Libraries support was not that much familiar

[3] In 2019, Supriya, Pavani, Saisushma, Vimala Kumari and Vikas presented a ML based loan prediction model. Themodulesin 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.

Advantages: Experimentation concluded that, DT has significantly higher loan prediction accuracy than the other models.

Disadvantages: It requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality. There can also be times where they must wait for new data to be generated.

[4] In 2017, Goyal and Kaur 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.

Advantages: When evaluated these models using Environment in five runs, TGA resulted in better loan forecasting performance than the other methods.

Disadvantages: Another major challenge is the ability to accurately interpret results generated by the algorithms. You must also carefully choose the algorithms for your purpose.

[5] In 2016, Aboobyda Jafar Hamid and Tarig Mohammed Ahmed 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.

Advantages: Evaluating these models on the dataset, experimental results concluded that, J48 based loan prediction approach resulted in better accuracy than the other methods.

Disadvantages: The minimum age of the customer satisfies the rules predicted for the loan he applies.

[6] In 2016, Kacheria, Shivakumar, Sawkar and Gupta suggested a loan sanctioning prediction procedure based 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.

Advantages: Experimentation put-forth the conclusion that, integration of KNN and binning algorithm with NB resulted in improved prediction of loan sanctioning process

Disadvantages: If you want to work on datasets with many features this can be problematic with kNN. This can be a tedious extra task and it can also introduce wrong bias to the data.

[7] In 2016, Goyal and Kaur 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.

Advantages: 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.

Disadvantages: The art of ensembling is hard to learn and any wrong selection can lead to lower predictive accuracy than an individual model. Ensembling is expensive in terms of both time and space.

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), 2018

[2]. 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), 2018

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

[4]. Anchal Goyal, Ranpreet Kaur, ―Accuracy Prediction for Loan Risk using Machine Learning Models‖,International Journal of Computer Science Trends and Technology (I JCST) , Jan-Feb 2017

[5]. Aboobyda Jafar Hamid and Tarig Mohammed Ahmed, ―Developing Prediction Model of Loan Risk in Banks using Data Mining‖, Machine Learning andApplications: An International Journal (MLAIJ), March 2016

[6]. Aditi Kacheria, Nidhi Shivakumar, Shreya Sawkar, Archana Gupta, Loan Sanctioning Prediction System, International Journal of Soft Computing and Engineering (IJSCE), 2016

[7]. Anchal Goyal, Ranpreet Kaur, Loan Prediction Using Ensemble Technique, International Journal of Advanced Research in Computer and Communication Engineering, March 2016