

Smart Lender-Applicant Credibility Prediction for Loan Approval

Team Leader :

Mr. Ragul V

Team Members :

Mr. Aswath P

Mr. Murali Krishnan N

Mr. Nivas Saravana N

Guide :

Dr. Ramya J

1. INTRODUCTION:

Despite the fact that our banking system offers a wide range of goods, a bank's credit line is its primary source of income. As a result, they are able to profit from interest on the loans they credit [1]. Lenders always seek to lower their credit risk since commercial loans have historically represented a significant portion of the banking sector [5]. The function that banks play in the modern market economy is substantial. Loans, or whether clients repay or don't return them, significantly impact a bank's profitability [1]. Before granting loans to borrowers, banks must determine if they are good (non-defaulters) or bad (defaulters). The creditworthiness of the borrowers is one of the most critical issues in commercial loan financing.

The probability that borrowers may default on their loan commitments is referred to as credit risk [5]. For any bank or institution, determining whether a borrower will be good or bad is a very difficult process. The banking system employs a manual procedure to determine whether or not a borrower has defaulted. The manual method will undoubtedly be more precise and efficient, but it will not be able to handle a high volume of loan applications at once. When a situation like this arises, it will take a very long time to make decisions and a lot of labour will be needed.

Therefore, the objective is to determine if the borrower is good or bad, that is, whether or not the borrower will be able to repay the loans.

With the use of machine learning algorithms, this is possible.

2. ABSTRACT:

So, - Many people and businesses are requesting bank loans right now. The distribution of loans is every bank's primary activity. Giving their assets to trustworthy hands is the primary goal of the banking industry. The verification and validation processes, however, are quite

time-consuming for banks or financial institutions, and even after going through such an arduous process, there is no guarantee if the applicant selected is deserving or not. To address this issue, we created a technique that enables us to forecast if the selected candidate would be a deserving applicant for loan approval or not.

. Based on the model that has been developed using machine learning techniques, the system makes predictions. We even evaluated the precision of several machine learning techniques. We obtained accuracy rates ranging from 75% to 85%, but the greatest accuracy we obtained was via logistic regression or 88.70%.

The system comes with a user interface web application where the user may input the information needed for the model to forecast. The disadvantage of this model is that it considers several factors, but in real life, a loan application may occasionally be accepted based on a single, significant factor, which is not conceivable with this approach.

3. LITERATURE SURVEY:

There is no comparison of several methods in

[1], just one algorithm is employed. The greatest accuracy they could get with the algorithm, logistic regression, was 81.11%. The ultimate finding was that only people with strong credit, high income, and modest loan amount requirements would have their loan requests granted.

[2]. The accuracy of the two algorithms—two-class decision jungle and two-class decision—was 77.00% and 81.00%, respectively. They also estimated measures including Precision, Recall, F1 Score, and AUC in addition to these.

[3]. Gradient Boosting, Logistic Regression, Random Forest, and CatBoost Classifier were the methods employed.

[4] include Logistic Regression, Support Vector Machine, Random Forest, and Extreme Gradient Boosting. The accuracy % of each algorithm was rather consistent. But the lowest variance was provided by the support vector machine. The less variation, the less the scores will fluctuate, and the more accurate and stable the model will be.

[5,] there is just the use of the K Nearest Neighbour Classifier. The Min-Max Normalisation procedure is applied. It is a method of breaking down traits and values. The dataset was divided 50–50, with k set at 30, and the greatest accuracy they obtained was 75.08%. Only Logistic Regression is employed in [6]. The algorithm's accuracy wasn't calculated.

4. METHODOLOGY:

The suggested system contains a web application with a model installed that was learned using machine learning methods. The user must fill out a total of 11 fields on the form. There are 11 characteristics total in the dataset we used to train the model. Before utilising this dataset to train the model, preprocessing is performed on it. The dataset's null values are replaced with mean and mode methods as part of the pre-processing, and string values are replaced with 1 and 0 using a label encoder.

The dataset was then split into train and test halves. 10% of the dataset is utilised to assess the accuracy that the model will provide for various techniques, while 90% of the dataset is used for training purposes.

Different methods were used once the dataset was divided, and each of them produced results with varying degrees of accuracy. The greatest result we had was 88% from the Logistic Regression. A pickle file of the model is produced once it has been trained. To anticipate whether or not their loan will be approved, clients must first fill out a form on our website.

The user only needs to complete the form and click the MAKE PREDICTION button to find out if the customer's loan will be accepted or denied based on the pickle file or model that we have trained. In the same way, we compared the accuracy of several machine learning techniques.

5. CONCLUSION:

We have selected the machine learning technique to analyse the bank dataset in order to forecast the loan approval status of the applied client. We tested a number of machine learning algorithms to see which one would work the best on the dataset to get the most accurate results. Using this strategy, we discovered that, aside from logistic regression, the other algorithms delivered accurate results in a suitable manner.

The remaining algorithms' accuracy ranged from 75% to 85%. The logistic regression, however, provided us with the highest level of accuracy (88.70%) after comparison of all the methods.

We also identified the key elements that have the greatest impact on the loan approval status. The performance accuracy of these most crucial characteristics applied to a few chosen algorithms is then compared to the case where all features were employed. The banks may use this model to determine what elements are crucial for the loan approval process. Based on their accuracy, the comparison research clearly identifies which algorithm will be the best and disregards the others.

6. REFERENCES:

- [1] M. A. Sheikh, A. K. Goel, and T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 490-494, DOI: 10.1109/ICESC48915.2020.9155614
- [2] K. Alshouli, A. AlGhamdi, and D. P. Agrawal, "AzureML Based Analysis and Prediction Loan Borrowers Creditworthy," 2020 3rd International Conference on Information and Computer Technologies (ICICT), 2020, pp. 302-306, DOI: 10.1109/ICICT50521.2020.00053.
- [3] B. Patel, H. Patil, J. Hembram and S. Jaswal, "Loan Default Forecasting using Data Mining," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-4, DOI: 10.1109/INCET49848.2020.9154100
- [4] S. Z. H. Shoumo, M. I. M. Dhruva, S. Hossain, N. H. Ghani, H. Arif, and S. Islam, "Application of Machine Learning in Credit Risk Assessment: A Prelude to Smart Banking," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), 2019, pp. 2023-2028, DOI: 10.1109/TENCON.2019.8929527.
- [5] G. Arutjothi, C. Senthamarai, "Prediction of loan status in commercial banks using machine learning classifier" 2018 International Conference Sustainable Systems (ICISS).
- [6] Ashlesha Vaidya, "Predictive and Probabilistic approach using Logistic Regression" 2017 8th International Conference on Computing, Communication and Networking Technologies.