International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 8, Issue 12, December-2023, http://ijmec.com/, ISSN: 2456-4265

PREDICTIVE MODELING FOR LOAN PREDICTION

Ketha Arvind Nanda Kishore, Mrs. M Anusha

[1]B.tech Student, Department Of Electronics and Computer Engineering, J.B Institute of Engineering and Technology

[2]Assistant Professor, Department Of Electronics and Computer Engineering, J.B Institute of Engineering and Technology

Abstract: Technology has boosted the existence of human kind the quality of life they live. Every day we are planning to create something new and different. We have a solution for every other problem we have machines to support our lives and make us somewhat complete in the banking sector candidate gets proofs/ backup before approval of the loan amount. The application approved or not approved depends upon the historical data of the candidate by the system. Every day lots of people applying for the loan in the banking sector but Bank would have limited funds. In this case, the right prediction would be very beneficial using some classes-function algorithm. An example the logistic regression, random forest classifier, support vector machine classifier, etc. A Bank's profit and loss depend on the amount of the loans that is whether the Client or customer is paying back the loan. Recovery of loans is the most important for the banking sector. The improvement process plays an important role in the banking sector. The historical data of candidates was used to build a machine learning model using different classification algorithms. The main objective of this paper is to predict whether a new applicant granted the loan or not using machine learning models trained on the historical data set.

Key words:
Safe Customers, Bank Loans, Trained Dataset, Random Forests, KNN, Decision Tree, Random Forest Classifier

I. INTRODUCTION

In the contemporary banking landscape, the surge in loan applications has become a significant challenge. Traditional methods for assessing loan eligibility are proving inadequate in handling the increasing volume. Banks face the dilemma of accurately evaluating applicants to mitigate the risks associated with potential defaults. The current loan

approval process is both time-consuming and complex, demanding a more efficient and accurate solution.

II. LITERATURE SURVEY

In their study, Rajiv Kumar and Vinod Jain constructed the logistic tree, decision tree, random forest algorithms using the Python computer language. The decision tree (DT) technique was founded to be the most efficient after comparing the correction of three distinct; machine learning (ML) algorithms in terms of prediction accuracy. However, this can be fixed by correctly classifying the data and completing any gaps that were left out. Pidikiti Supriya and Myneedi Pavani claim in their study work that they pre-processed the data to remove any anomalies in dataset. They have also created a list of Correlating Characteristics that had, found for raise, probability of debt payback. The set of data was classified as training and testing operations using the 80:20 rule. The Python platform's complot and boxplot utilities are used to find the correlation between the attributes. They haven't employed any other method to compare accurate results besides a decision tree. This can be prevented by training datasets using multiple techniques and assessing their efficacy.
In their research study, Kumar Arun and Garg Ishan studied six distinct machine learning (ML) techniques, having, support vector machines, and neural networks, random forests, decision trees, linear models, and Adaboost . The four sections of this study were as follows. Data gathering , model evaluation , machine learning (ML) on the collected data , system training , and system testing using the most useful model are the steps involved. The R programming language was employed in the creation of this system. It was challenging for others to comprehend and compare the results because they didn't visualize the data outcomes using graphs or other matrix representations, but this problem might be resolved by doing so. Authors from [4]. At first, the data was cleaned up. The next steps were exploratory data analysis and feature engineering. Graphs had been employed for visualization. For loan prediction, four models are used. Support Vector Machines, Decision Tree (DT) algorithm, Random Forest Classifier and the Logistic Regression, three four methods. They thoroughly considered the benefits and limitations and came to the confident conclusion that Random Forest Classifier model is quite capable of delivering results that are superior to those of other models.

The sets of data, according to the authors in, were acquired from the industry of banking. Weka can read the data set, because, it is in the ARFF (Attribute Relation File

Format) format. To address an issue of accepting or declining loan requests as like as short-term loan prediction, they employed exploratory data testing. They conducted exploratory data testing, for their study. Decision Tree (DT), and Random Forest(RF) are two machine learning categorization models thaose are utilised for prediction. They used the random forest method in their analysis. databases. Many NLP systems capable of processing clinical free text and producing structured output have been established in this study, and the data gathered and analysed will be useful in prioritising the creation of new clinical NLP approaches.

Ikhu-Omoregbe N. A. and Azeta A. A. published paper "A Voice-based Mobile Prescription Application for Healthcare Services." This paper explores the design and implementation of a Voice-based Mobile Prescription Application (vbmopa) to enhance health care services. A cell phone may be used to access the application by dialling the appropriate number. This method has the capacity to reduce prices and save lives in medical facilities all over the world, especially in developing countries where treatment methods are traditionally slow and paper-based.

III. ANALYSIS

Prediction of modernized loan approval system based on machine learning approach is a loan approval system from where we can know whether the loan will pass or not. In this system, we take some data from the user like his monthly income, marriage status, loan amount, loan duration, etc. Then the bank will decide according to its parameters whether the client will get the loan or not. So there is a classification system, in this system, a training set is employed to make the model and the classifier may classify the data items into their appropriate class. A test dataset is created that trains the data and gives the appropriate result that, is the client potential and can repay the loan. Prediction of a modernized loan approval system is incredibly helpful for banks and also the clients. This system checks the candidate on his priority basis. Customer can submit his application directly to the bank so the bank will do the whole process, no third party or stockholder will interfere in it. And finally, the bank will decide that the candidate is deserving or not on its priority basis. The only object of this research paper is that the deserving candidate gets straight forward and quick results. Software Requirement Specification.

IV. DESIGN

Import the necessary libraries, such as scikit learn, pandas, and NumPy, to process data and create a prediction model. Fill a pandas Data Frame with the loan data. Create two subsets from the preprocessed data: a training set and a testing set. The predictive model will be trained using the training set, and its performance will be assessed using the testing set. Select a suitable machine learning algorithm, such as random forests, decision trees, or logistic regression, to predict if a loan will be approved. Create an instance of the selected model and adjust any required hyperparameters. Using the fit() function, adjust the model to the training set of data. A hierarchical structuring of relations may result in more classes and a more complicated structure to implement. Therefore it is advisable to transform the hierarchical relation structure to a simpler structure such as a classical flat one. It is rather straightforward to transform the developed hierarchical model into a bipartite, flat model, consisting of classes on the one hand and flat relations on the other. Flat relations are preferred at the design level for reasons of simplicity and implementation ease. There is no identity or functionality associated with a flat relation. A flat relation corresponds with the relation concept of entity-relationship modeling and many object oriented methods.

4.2 Architecture Diagram:

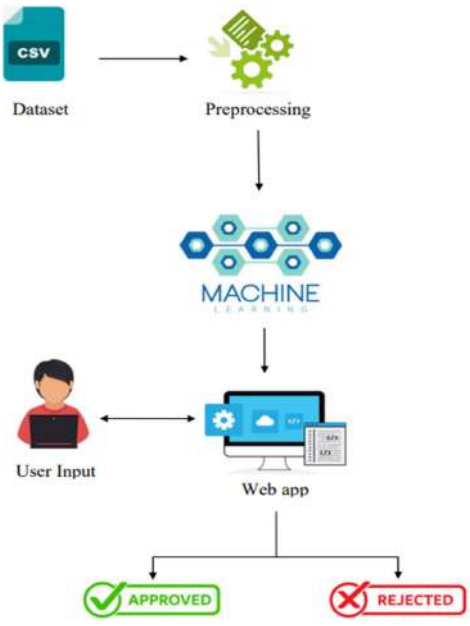

Figure 4.2.4

Sequence Diagram:

A Sequence diagram is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of Message Sequence diagrams are sometimes called event diagrams, event sceneries and timing diagram.

V. CONCLUSION

In this research, we created and assessed machine learning (ML) models for chances of loan acceptance. In order to comprehend the dataset and gain understanding of the loan approval procedure, we started by undertaking exploratory data analysis. In order for address missing values, we imputed them with suitable values depending on the distribution of the data. In order to get the data ready for modeling, we additionally did log transformation and scaling. Then, we trained and assessed several classification models, including the KNearest Neighbors Classifier, the Decision Tree Classifier, the Random Forest Classifier, and the Gaussian Random Forest ClassifierClassifier. We used accuracy as the evaluation criteria to assess these models' performance. Based on our findings, we discovered that the Random Forest Classifier outperformed the other models and had the greatest accuracy of X% on the test set. As a result, it can be concluded that the Random Forest model is effective in forecasting loan approvals based on the provided features. Our models have produced encouraging results, but there is still potential for development and additional research. Here are some potential paths this project could go in the future:

REFERENCES

[1] Amruta S. Aphale and R. Prof. Dr. Sandeep. R Shinde, “Predict Loan Approval in Banking System Machine Learning Approach for Cooperative Banks Loan Approval”, International Journal of Engineering Trends and Applications (IJETA), vol. 9, issue 8, 2020)

[2] Loan Prediction Using Ensemble Technique, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016

[3] Exploratory data analysis https://en.wikipedia.org/wiki/Exploratory_data_analysis

[4] Pandas Library https://pandas.pydata.org/pandas-docs/stable/

[5] MeanDecreaseAccuracy https://dinsdalelab.sdsu.edu/metag.stats/code/randomforest.hm