An Approach for Prediction of Loan Approval using ML algorithm

P.S.A.Bhaskar Reddy[1],M.Gnaneswar Reddy [2], Ponmagal R.S[3]

Department of Computing Technologies, SRM Institute of Science of Technology, Kattankulathur, India, 603203

Corresponding author Email : sp4781@srmist.edu.in

Co author's Email : mr1409@srmist.edu.in , ponmagas@srmist.edu.in

Abstract

Several individuals are looking for banking advancements resulting from the financial sector's improvement, the bank can only offer certain people access to its limited resources. So, determining who can be granted credit is a common interaction and will be a more secure option for the bank. To conserve a tonne of bank resources and labour, we therefore try to lessen the element of risk involved in choosing the protected person in this assignment. To do this, it is necessary to examine the prior records of those who have previously received advances, base the machine's preparation on these records and encounters, and use the model that yields the most accurate results.This project's main goal is to determine whether giving someone a cash advance will protect them. Four pieces make up the project. Data gathering, ML model comparison using the information obtained, framework training using the most promising model, and testing are the first three steps.

Keywords - Machine Learning,Support Vector Machine,K-Nearest neighbor,Logistic Regression

I. INTRODUCTION

In a market economy, banks are essential. The industry capacity to assess risk has a significant impact on whether a business succeeds or fails. Before approving a loan, The bank decides if a borrower is trustworthy or not. Any business or bank will find it difficult to foresee a borrower's condition, such as whether they would default or not in the future. The prediction of loan defaulters is essentially a binary classification problem. Sum borrowed the customer's credit history determines his eligibility to receive a loan. The data used for this work came from previous customers of various banks who received loan approvals across a variety of boundaries. The AI model is therefore set up on that data to produce accurate results. Identifying the security of credit is our main goal in conducting this investigation. The strategic relapse calculation is applied to forecast future health. In order to prevent the informational collection from having any missing components, it must first be cleaned. A data set of 1000 examples, 10 mathematical formulas, and 8 explicit traits were used to create our model. Several criteria, such as CIBIL Score (Credit History), Company Value, and others must be met in order to credit an advance to a client.

II. EXISTING SYSTEM

Abhilash had suggested a method for approving using a decision tree to forecast loans. One of the most used models is the decision tree and very useful Regression algorithm.To decide whether to divide a bump into two or more sub-knots, decision trees employ a variety of strategies. Future sub-knot formation becomes simpler. As a result, we can state that the bump's chasteness grows in connection to the desired variable. Current System's drawbacks include (1).Less efficiency (2).Less accuracy (3).Issue With Prediction (4).It is not a web-based programme.

III. PROPOSED SYSTEM

We suggest a machine learning-based logistic regression analysis model for loan approval in this paper. among the most effective and well-liked classification-based algorithms is logistic regression.. The purpose or the importance of using Logistic Regression was that it uses the concept of predictive analysis which was suitable enough for describing the data.The advantages of Proposed System is (1).Great accuracy (2).Time efficient (3).Simple UI (4).It's a web based application (5).It gives perfect result.

IV. LITERATURE SURVEY

By creating and evaluating the efficacy of several models using data from the Bank of Kigali, Ndayisenga et al. (2021) helped to the effort with commercial banks to anticipate the behaviours of borrowers. The data was split into two groups, training and test, with training dataset making up 70% of the overall total and test dataset making up 30%. To find the best machine learning techniques to use for forecasting bank loan default, ensembles were used. Gradient Boosting (Accuracy 80.40%) and XGBoosting were shown to be the two most accurate models for forecasting bank loan default, with decision trees, random forest, and logistic regression doing poorly.

According to the clients' history financial and credit scores, a reliable predictive modelling strategy was developed by Tejaswini et al. (2020) to approve or reject loan applications. The goal of this study was to provide a rapid, simple, and effective procedure for choosing competent applications. A number of financial institutions provided the information. The machine learning model received the training data set, and the model was trained using that data

set. The data submitted by each new applicant on the application form acts as a test data set. In this study, they employed three machine learning techniques—Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF)—to forecast customer loan approval.When compared to the machine learning algorithms for Logistic Regression and Random Forest, the testing results demonstrate that the Decision Tree method has a greater accuracy of 82.00%.

A credit scoring model for credit information. A range of machine learning techniques are employed in the construction of the financial credit score model. We suggest a classifier-based machine learning analysis methodology for credit data. With the software package tool, the goal is accomplished. With the best precision, this suggested model offers the crucial data. In commercial banks, a machine learning classifier is used to forecast the status of loans. They can be grouped using a range of machine learning (ML) algorithms, such as SVM, neural networks, RF (random forest), decision trees, Xg boost, and so forth. The suggested methodology uses both the Random Forest(RF) and Decision-Tree classifiers with 72% accuracy for Random Forest classifiers and a 76% accuracy for decision tree classifiers.

V. DATASETS

By examining a wide range of attributes, data sets have been made rich, which will increase the model's efficacy and precision. We outline the essential traits that should be looked at in order to identify if it's a legitimate website or not.

Using Loan_Id

A loan id is a special set of digits that your bank assigns to your loan account once it has been approved and opened. Each loan account is given a different loan account number. If you've obtained many loans from the same bank, each loan will have a unique loan account number. With the help of the loan account number, which is specific to each loan account, banks keep a record of all the loans they have authorised. We have used this attribute.

Using Gender

Whether they are male or female. We have used this attribute.

Using Marriage Status

Whether they are married or not.We used these attributes.

Using Dependencies

The state of being dependent on somebody/something. The state of being unable to live without someone. We have used this attribute.

Using Education

The imparting or obtaining specific knowledge or abilities, such as those needed for a profession. a qualification, level, or type of instruction at a university. We have included whether they are graduation or not.We have used this attribute.

Using Self_Employment

the practise of working independently for oneself as a freelancer or business owner as opposed to an employer. We have used this attribute.

Using Applicant Income

Income from employment is the money you earn from a job.Applicant Income refers to the income of the employee. We have used this attribute

Using Co- Applicant Income

Co-Applicant Income refers to the income of family or relation or friends or the nearest relatives. We have used this attribute.

Using Loan_Amount

The loan amount is the sum of money that you are currently obligated to pay us under this mortgage. UN-paid principal, interest on underpaid principal, missed payments, interest on missed payments, other fees, and interest on other fees could all be included in the loan amount..We have used this attribute.

Using Loan_Amount_Term

The length of time you take to repay the loan is known as the tenure. The amount of interest you must pay depends on how long it takes you to return your personal loan; the longer it takes, the more interest you must pay. We have used this attribute.

Using Credit_history
Your payback history for obligations like credit card and loan balances is documented in your credit history. Your credit reports, which often include extra financial information, include a record of your credit history.We have used this attribute.

Using Property Area
a region of land that is used for a certain purpose (such as building or farming). We have classified into three categories.They are rural,semi-rural,urban.We have used this attributes.

Using Loan_Status
The loan status indicates where your loan is in the process. We have used this attribute to display whether the loan is granted or not.

VI.   MACHINE LEARNING METHODOLOGY

LOGISTIC REGRESSION
Statistically, the connection between a binary dependent variable and one or more independent variables can be modelled using logistic regression. It is a kind of regression analysis that models the data using a logistic function.Probability of the dependent variable taking a certain value (typically 0 or 1) based on the values of the independent variables.In other words, logistic regression is a technique based on the values of the input variables, is employed to forecast the likelihood that an event will occur. When attempting to forecast whether an observation belongs to one of two classes in binary classification tasks, it is frequently employed in machine learning and statistical analysis..Logistic regression produces a logistic curve, which is an S-shaped curve that demonstrates the connection between the dependent and independent variables. It is ideal for binary classification jobs because the logistic function employed in logistic regression makes sure that the predicted probabilities are always between 0 and 1.
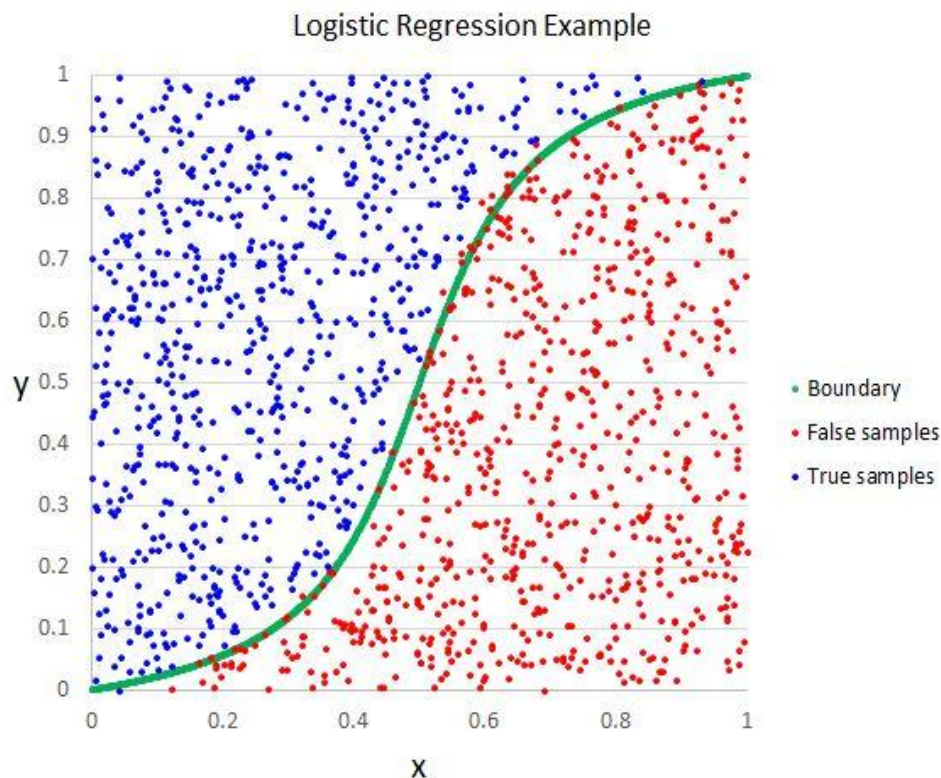
Fig. 1. Logistic Regression

VII. IMPLEMENTATION

The dataset is split into two parts: test and train. To improve accuracy, we used training and testing ratios: 70:30, and 80:20. Then in all of ratio the data goes into data preprocessing It addresses the data must be altered or encoded in order for a device to quickly parse it. The programme should immediately identify and comprehend the model must take into account the properties of the data in order to produce precise and accurate predictions. Unprocessed data are converted into manageable numerical features using the feature extraction procedure, which keeps the integrity of the original data set intact. It produces better outcomes than usingML on only raw data. The data is subsequently passed on to the ML algorithms that have been developed. They are trained first, then tested to determine the correctness of each method.
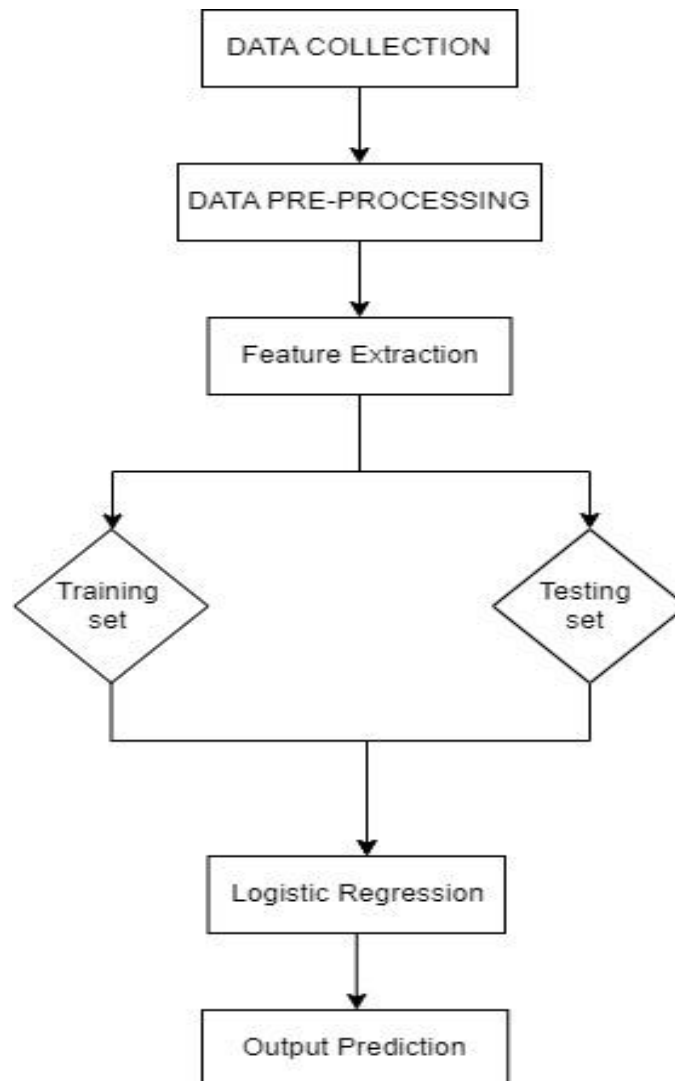
Fig.2 . System Architecture

VIII. RESULT

Different tools were used to import machine learning models. Especially Sklearn tool is used to import the methodologies. The datasets are separated into training and testing sets in 70:30, and 80:20 ratio.

Using the training set the classifier is taught and trained, and the testing set is used to assess the classifier's efficiency. By calculating scores for accuracy, false-negative, and false-positive rates for each classifier, classification performance was evaluated. Confusion matrix is used further to determine the accuracy.
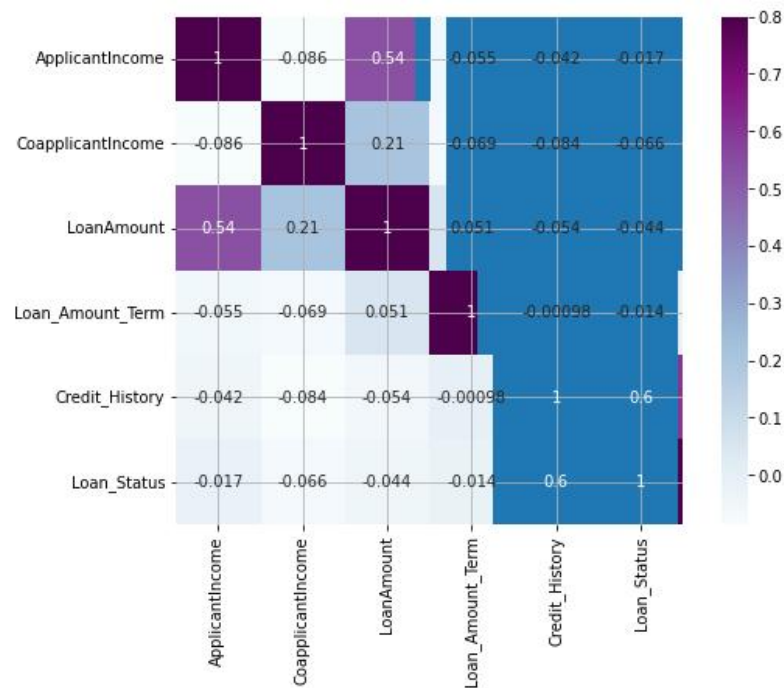
Fig. 3. Confusion Matrix

IX. CONCLUSION

This paper tries to make the detection techniques for loan prediction more effective. With logistic regression method, we had the lowest false positive rate and 82% detection accuracy. Additionally, the findings demonstrable classifiers function more effectively when a large amount of data is used as training data. Future loan prediction websites will be more successfully discovered by combining machine learning's logistic regression algorithm .

REFERENCES

[1] An Method for Predicting Loan Approval Using Machine Learning Algorithm, International Conference on Electronics and Sustainable Communication Systems (ICESC), M. Sheikh, A. Goel, and T. Kumar (2020).

[2] S. M. S. and R. Sunny T, "Loan Credibility Prediction System Based on Decision Tree Algorithm," International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue 09, (2015).

[3] "Loan Approval Prediction Based on Machine Learning Method," IOSR Journal of Computer Engineering, by A. Kumar, I. Garg, and S. Kaur (2016).

[4] "Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Approaches," by Dr. K. Kavitha Issue 2 of Volume 6 of IJARCSSE (2016). 65–80 in Library, vol. 38, no. 1, 2020.

[5] "Prediction of Loan Status in Commercial Bank Using Machine Learning Classifier," Proceedings of the International Conference on Intelligent Sustainable Systems, G. Arutjothi and Dr. C. Senthamarai (2017).

[6] "Loan Prediction by Using Machine Learning Models," International Journal of Engineering and Technology, P. Supriya, M. Pavani, N. Saisushma, N. Kumari, and K. Vikas (2019).

[7] "HOME LOAN DATA ANALYSIS AND VISUALIZATION," International Journal of Creative Research Thinking (IJCRT), R. Salvi, R. Ghule, T. Sanadi, and M. Bhajibhakare (2021).

[8] "Domain-specific modification of a partial least squares regression model for loan defaults prediction," 11th IEEE International Conference on Data Mining Workshops, B. Srinivasan, N. Gnanasambandam, S. Zhao, and R. Minhas (2011).

[9] "Neural Networks for Prediction of Loan Default Using Attribute Relevance Analysis," International Conference on Signal Acquisition and Processing, M. V. Reddy and Dr. B. Kavitha (2010).

[10] Business-Oriented Feature Selection for Hybrid Classification Model of Credit Scoring, IEEE Second International Conference on Data Stream Mining & Processing, August G. Chornous, I. Nikolskyi (2018).