

In Banking Loan Approval Prediction Using Machine Learning

Wani Dhanashri¹, Tajane Adarsha², Doshi Tejal³

¹Wani Dhanashri, JCOE Kuran

²Tajane Adarsha, JCOE Kuran

³Doshi Tejal, JCOE Kuran

⁴Prof. Bhosale Sachin, Department of Computer Engineering, JCOE Kuran, Maharashtra, India

Abstract - Bank provides many services to the customer but the main source of income of bank is on its credit line. Hence they can earn from an interest of those loans which they credit. A bank profit or a loss depends to a large extent on loans and also on credits. The datasets are used from the Kaggle for prediction that is going to help to develop a prediction model. Loan approval is an important process for banking. On the basis of the income, dependents, property area and other factors the bank approved or reject the application of a loan. Recovery of loans is the most important contributing parameter in the financial system of a bank.

Many researchers did work on loan approval prediction systems. In this paper we use the machine learning algorithms Decision tree, Logistic Regression, and Random Forest. The results conclude that the accuracy of the Random Forest machine learning algorithm is better as compared to Decision Tree and Logistic Regression machine learning approaches.

Key Words- Loan Prediction, Machine Learning Cloud services, testing and training.

1.INTRODUCTION

People depend on bank loans to fulfill some needs. The rate of loan applications increases with very fast speed in recent years. The banking officers are very conscious about the payment of loan amounts by their customers. Even after taking a lot of precautions and analyzing loan applicants' whole data, the loan approval decisions are not always correct. Hence there is a need for automation of this process so that loan approval is less risky and suffers less loss for banks. Nowadays Artificial Intelligence is an emerging technology. It creates a model from training data. The machine learning algorithm trained the system using some of the data available and test the remaining data to build the prediction model.

The ML techniques can be applied to a sample test data first and then can be used in making prediction-related decisions. In this paper, the supervised machine learning approaches are used in solving the loan approval problem of the banking sector. The next section discusses the literature survey. Then proposed algorithms, results, and analysis are discussed.

After that conclusion and future scope are discussed which is followed by the references used in this paper.

2. LITERATURE SURVEY

In this paper, we tried to minimize the risk factor behind approving a loan to the safe person to save lots of bank efforts and money. The approval of a loan mostly depends on the credit history of the candidate. Those applicants who were applying for low amounts but having high income get easily approved. In this given report we have implemented Logistic regression models as Predictive analysis tools. It is used for the given problem of prediction of loan approval.

[1].H. Ramachandra, G. Balaraju, R. Divyashree and H. Patil, "Design and Simulation of Loan Approval Prediction Model using AWS Platform," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 53-56. The main goal of this paper is to implement cloud-based machine learning algorithms. This paper is divided into four sections (i)Data Prepossessing (ii)Data Filtering(iii)Decision Tree Algorithm(iv)Logistic Regression

[2]."Prediction for Loan Approval using Machine Learning Algorithm" Author-Ashwini S. Kadam, Shraddha R. Nikam, Ankita A. Aher, Gayatri V. Shelke, Amar S. Chandgude. Year-2021 . Aim of these paper is to predict loan safety. The SVM and Naïve bayes algorithms are used. The data cleaning is perform to avoid the missing values in the data set. The paper has following sections(i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation

[3]A. Gupta, V. Pant, S. Kumar and P. K. Bansal, "Bank Loan Prediction System using Machine Learning," 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), 2020. This paper allows jumping on particular applications that deserve to be approved on a priority basis.

[4]."Prediction of Loan Defaulter-A data Science Perspective" Author- P. Maheshwari, CH .V. Narayana, year-2020.The objective of these paper is to use data science and data analytical methods to achieve a business objective. This process involves several steps like data cleaning, preparation, modeling and model evaluation.

3. PROCEDURE

Step i: Start

Step ii: Recording the loan data(ID, married, loan amount, education, dependent Etc.)

Step iii: Data is trained.

Step iv: All Algorithms were processed on data.

Step v: Data is tested.

Algorithms that are used in this model are:

1. Logistic Regression
2. Decision Tree
3. Random forest

Logistic Regression:

Logistic Regression is a machine learning algorithm..

There are the following basic points:

1. In Logistic regression there are binary dependent variables.
2. In Binary regression dependent variables have level 1.
3. Every Included variable should have meaning. All included independent variables should be self-reliant.

Decision Tree:

The Decision tree is a supervised machine learning technique that is nonparametric in nature. In this algorithm predefined target variables are generally used in classification problems. It is useful for both classification and regression. It works categorical for input and output variables. The purest node in a given model.

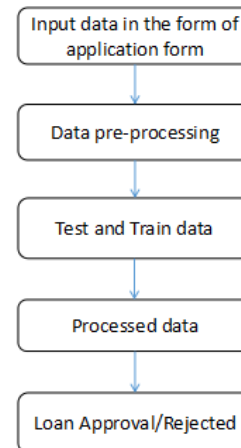
Random Forest

Random Forest is a very useful machine learning algorithm.

Random Forest is a approach which uses test data for the model for training. It creates random forests for the problem set using a decision tree and then finds the solution using these random forests. At the time of the training the data Random forest creates the decision trees to predict the outcome.

4. PROPOSED SYSTEM

This proposed system gives the predicted value to a customer with details then the person is eligible or not for a loan, According to information provided by the applicant as input in real-time or collected by bank employees. The Proposed system's prediction model includes the Decision tree model and Random forest model along with the Logistic regression model present in an existing system.



This Simple website is very helpful for users like customers or applicants and bank employees. The User interface is user-friendly there are no complications in using the interface, and it can be used just by entering necessary details into the UI in real-time it'll give the predicted value like if the customer is eligible or not and give some scenarios about eligibility.

5. MATHEMATICAL MODEL

Let s be System :

S=, I, P, O

S: is a System

P= Process which includes a whole procedure

I1: Loan datasets

I2: Trained datasets.

DC: Data Cleaning

DP: Data Processing

DV: Data Verification

DT: Decision Tree

RD: Random Forest

LG: Logistic Regression

6. RESULT AND ANALYSIS

There is a three machine learning approaches we have used to test the data to predict the loan defaulters from loan applications. We employed our 75% data from our available dataset for training and remaining 25% data from our dataset is employed for testing. Then prediction accuracy of the various Machine Learning approaches is calculated and compared.

Based on the train data set system analyze the rest of the 25% data and predict the results in terms of loan status either accepted or rejected.

Table-1: Comparison of prediction accuracy of machine learning algorithms

Sr.No.	Machine learning Algorithm	Accuracy Percentage
1	Logistic Regression	83.04
2	Decision Tree	85
3	Random Forest	88.53

7. CONCLUSION

We have used the three machine learning algorithms Decision tree, Random forest and Logistic regression in the prediction of loan approval. The results concludes that the prediction accuracy is 83.04%, 85%, and 88.53% for Logistic Regression, Decision Tree algorithm Random Forest algorithms respectively. From these three algorithms, the accuracy of the Random Forest algorithm is high so it is best for the prediction of loans defaulters.

8. FUTURE SCOPE

In the future the Decision Tree, Logistic regression are algorithms that can be applied to other data sets available for loan approvals to further investigate the accuracy. The analysis of other machine learning algorithms other than these three algorithms can also use in the future to investigate the power of machine learning algorithms for loan approval prediction.

9. REFERENCES

1) H. Ramachandra, G. Balaraju, R. Divyashree and H. Patil, "Design and Simulation of Loan Approval Prediction Model using AWS Platform," 2021 International Conference on

Emerging Smart Computing and Informatics (ESCI), 2021, pp. 53-56, doi: 10.1109/ESCI50559.2021.9397049.

2) A. Gupta, V. Pant, S. Kumar and P. K. Bansal, "Bank Loan Prediction System using Machine Learning," 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), 2020, pp. 423-426, doi: 10.1109/SMART50582.2020.9336801.

3) M. Alaradi and S. Hilal, "Tree-Based Methods for Loan Approval," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 2020, pp. 1-6, doi: 10.1109/ICDABI51230.2020.9325614.

4) 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.

5) P. Maheswari and C. V. Narayana, "Predictions of Loan Defaulter - A Data Science Perspective," 2020 5th International Conference on Computing, Communication and Security (ICCCS), 2020, pp. 1-4, doi: 10.1109/ICCCS49678.2020.9277458.