

**Shri Ramdeobaba College of Engineering & Management, Nagpur**

**Department of Electronics and Communication**

**Academic Year 2021-2022**

**Machine Learning [ECT453-3] - VII<sup>th</sup> Semester**

**Teacher Assessment**



**Report On**

Loan Prediction Using Machine Learning.

**Submitted By**

Insiya Saher (27/A)

Chaitanya Sudan (41/A)

Neha Sahu (09/B)

Georgina Frank (11/B)

**Guided by**

Prof. A. Jaiswal

## **Abstract:**

In our banking system, banks have many products to sell but main source of income of any bank is on its credit line. So they can earn from interest of those loans which they credit. A bank's profit or a loss depends to a large extent on loans i.e. whether the customers are paying back the loan or defaulting. By predicting the loan defaulters, the bank can reduce its Non-performing Assets. This makes the study of this phenomenon very important. Previous research in this era has shown that there are so many methods to study the problem of controlling loan default. But as the right predictions are very important for the maximization of profits, it is essential to study the nature of the different methods and their comparison. A very important approach in predictive analytics is used to study the problem of predicting loan defaulters (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation.

**Key Words:** Loan Prediction, Machine Learning, Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Extra Trees Classifier.

## **Introduction:**

Loan Prediction is very helpful for employee of banks as well as for the applicant also. The aim of this Project is to provide quick, immediate and easy way to choose the deserving applicants. Customer first apply for loan after that company or bank validates the customer eligibility for loan. Company or bank wants to automate the loan eligibility process (real time) based on customer details provided while filling application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and other. This project has taken the data of previous customers of various banks to whom on a set of parameters loan were approved (Kaggle Dataset). So the machine learning model is trained on that record to get accurate results.

## **Objective:**

Our main objective of this project is to predict the best machine learning algorithm to predict with maximum accuracy that whether the applicant should be given the loan or not. To predict loan safety, the Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Extra Trees Classifier algorithm are used.

## **Problem Statement:**

To overcome the problems of existing system by improving the model selection before Loan Prediction analysis. Problem of existing is the less accuracy for predictive analysis. Only one prediction model is available to predict the loan approval.

## Existing Approach/ Algorithm:

Loan prediction model of existing system is a powerful tool for a range of possible circumstances. Convenient but still had some limitations to it. The history of prediction model is long and there had been many obstacles in its evolution. Although constantly increasing variations and evolution in technology managed to overcome many obstacles. The different prediction models use different approaches on dealing with these limitations. There are important observations to make about existing system of loan prediction. For the prediction of loan approval, mainly logistic regression approach had been used.

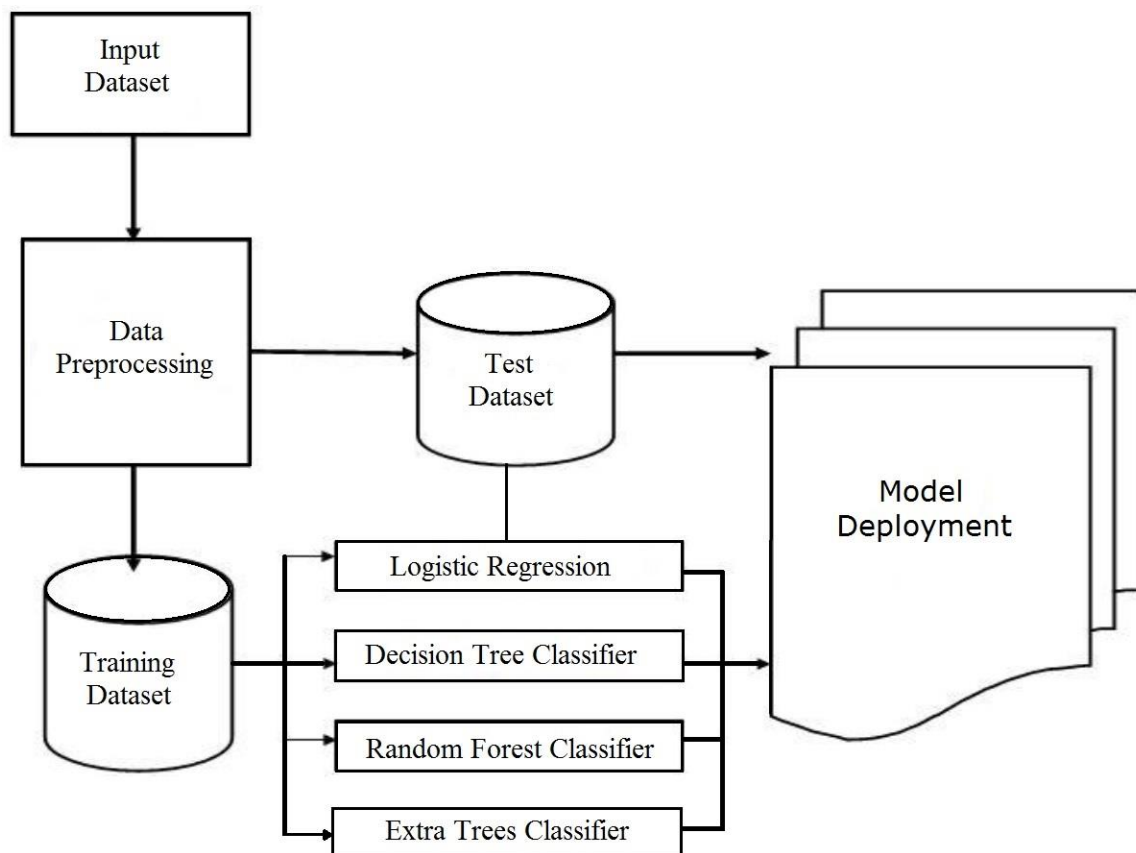
## Comparative Study:

Sr. No.	Paper Name	Author/ Publication	Technology	Advantages	Disadvantages
1.	Loan Approval Prediction with the help of Machine Learning Approach	Kumar Arun, IOSR	Decision Trees, Random Forest, Support Vector Machine, Linear Model, Neural network	Six ML based classification models have been used for prediction	The system is trained on old training dataset.
2.	Loan Prediction using Machine Learning Models	Pidiketi Supriya	Decision tree algorithm	The best accuracy on dataset test is 81.1%.	In this model only one method used for prediction
3.	Probabilistic and predictive approach using logistic regression: prediction of loan approval.	Ashlesha Vaidya/ Computer Science Engineering SRM University,	Logistic regression, decision trees, Artificial neural networks and Bayesian Networks.	Logistic regression is widely used in data analytics where analyzing of the pre - existing data within all kinds of organization is required.	Logistic regression requires a large sample for parameter estimation.

## Implemented Approach/ Algorithm:

### Algorithm of the Model Proposed:

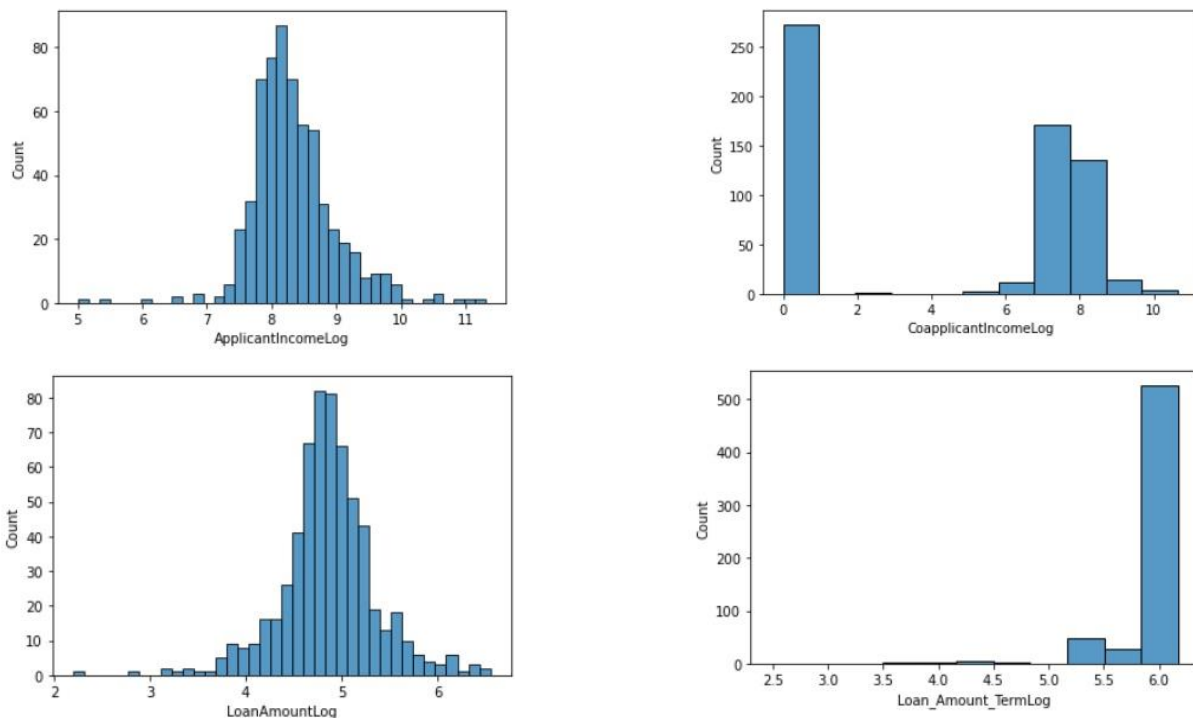
- Load the data
- Data cleaning and pre-processing
  1. Fill the missing values with mean values regarding numerical values.
  2. Fill the missing values with mode values regarding categorical variables.
  3. Outlier treatment
- Data Visualization using Countplot and Histplot
- Logarithm Conversion of Attributes and Visualization
- Correlation Matrix
- Apply the modelling for prediction
  1. Create the target variable. In this approach, target variable is loan-status
  2. Create a dummy variable for categorical variable (if required) and split the training and testing data for validation.
  3. Apply the model: Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Extra Trees Classifier
  4. Determine the accuracy and identify the best algorithm.



**Figure: System Implementation**

## Exploratory Data Analysis:

- The one whose salary is more can have a greater chance of loan approval.
- The one who is graduate has a better chance of loan approval.
- Married people would have an upper hand than unmarried people for loan approval.
- The applicant who has a smaller number of dependents have a high probability for loan approval.
- The lesser the loan amounts the higher the chance for getting loan.



Logarithmic Visualization Of Data

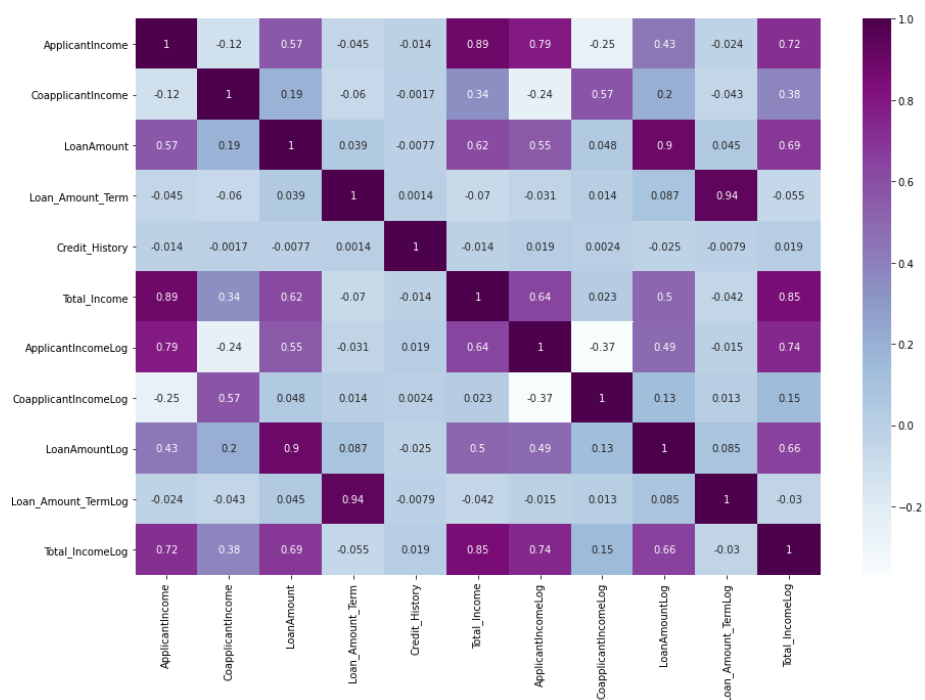


Figure: Correlation Matrix

## Result:

So here, it can be resulted with confidence that the Logistic Regression model is extremely efficient and gives a better result when compared to other models. It works correctly and fulfills all requirements of bankers. This system properly and accurately calculate the result. It predicts the loan is approve or reject to loan applicant or customer very accurately.

Sr. No.	Machine Learning Algorithm	Accuracy	Cross Validation
1.	Logistic Regression	81.82	80.95
2.	Decision Tree Classifier	76.62	70.36
3.	Random Forest Classifier	78.57	78.18
4.	Extra Trees Classifier	80.52	77.37

## Conclusion:

The main purpose of the project is to classify and analyze the nature of the loan applicants. From a proper analysis of available data and constraints of the banking sector, it can be concluded that by keeping safety in mind that this product is much effective or highly efficient. This application is operating efficiently and fulfilling all the major requirements of Banker. Although the application is flexible with various systems and can be plugged effectively.

This project can be extended to higher level in future so the software could have some better changes to make it more reliable, secure, and accurate. Thus, the system is trained with a present data sets which may be older in future so it can also take part in new testing to be made such as to pass new test cases.

There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system. So, in the near future the so – called software could be made more secure, reliable and dynamic weight adjustment. In near future this module of prediction can be integrated with the module of automated processing system.

## References:

- [1]. Kumar Arun, "Loan Approval Prediction based on Machine Learning Approach" IOSR-JCE (NCRTCSIT-2016).
- [2]. Pidikiti Supriya, "Loan Prediction using Machine Learning Models," International Journal of Engg. and Techniques, Mar-Apr 2019.
- [3]. Ashlesha Vaidya, "Predictive and probabilistic approach by logistic regression : prediction of loan approval," Computer Science Engineering SRM University, Chennai July 3-5, 2017, IIT Delhi
- [4]. Nikhil Madane, Siddharth Nanda, "Loan Prediction Analysis using Decision Tree", December 2019.
- [5]. Sivasree M.S., Rekha Sunny T., "Loan Credibility Prediction System using Decision Tree Algorithm", (IJERT), 09 September 2015.
- [6]. "Prediction of Loan Approval using Machine Learning" Rajiv Kumar, Vinod Jain, Premsagar Sharma, International Journal of Advanced Science and Technology, (2019)
- [7]. "Bank loan analysis using customer usage data: A big data approach using Hadoop," 2017 2nd International conference on Telecommunication & Networks.
- [8] J. R. Quinlan. Induction of Decision Tree. Machine Learning, Vol. 1, No. 1. pp. 81-106., 1986.
- [9] A. Goyal and R. Kaur, "A survey on Ensemble Model for Loan Prediction", International Journal of Engineering Trends and Applications (IJETA), vol. 3(1), pp. 32-37, 2016.
- [10] G. Shaath, "Credit Risk Analysis and Prediction Modelling of Bank Loans Using R".
- [11] A. Goyal and R. Kaur, "Accuracy Prediction for Loan Risk Using Machine Learning Models".
- [12] Hsieh, N. C., & Hung, L. P. (2010). A data driven ensemble classifier for credit scoring analysis. Expert systems with Applications, 37(1), 534-545.
- [13] [https://en.wikipedia.org/wiki/Exploratory\\_data\\_analysis](https://en.wikipedia.org/wiki/Exploratory_data_analysis)
- [14] <https://www.experian.com/blogs/ask-experian/credit-education/score-basics/what-is-a-good-credit-score>
- [15] Aboobyda Jafar Hamid and Tarig Mohammed Ahmed, —Developing Prediction Model of Loan Risk in Banks using Data Mining, Machine Learning and Applications: An International Journal (MLAIJ), Vol.3, No.1, pp. 1-9, March 2016.