1 INTRODUCTION

1.1 Overview

The immense increase in capitalism, the fast-paced development and instantaneous changes in the lifestyle has us in awe. Emi, loans at nominal rate, housing loans, vehicle loans, these are some of the few words which have skyrocketed from the past few years. The needs, wants and demands have never been increased this before. People gets loan from banks; however, it may be baffling for the bankers to judge who can pay back the loan nevertheless the bank shouldn't be in loss. Banks earn most of their profits through the loan sanctioning. Generally, bankspass loan after completing the numerous verification processes despite all these, it is still not confirmed that the borrower will pay back the loan or not. To get over the dilemma, I have built up a prediction model which says if the loan has been assigned in the safe hands or not. Government agencies like keep under surveillance why one person got a loan and the other person could not. In Machine Learning techniques which include classification and prediction can be applied to conquer this to a brilliant extent. Machine learning has eased today's world by developing these prediction models.

1.2 Purpose

By using this we can find whether a person is eligible for the loan or not.Here we will be using the fine techniques of machine learning – Decision tree algorithm to build this prediction model for loan assessment. It is as so because decision tree gives accuracy in the prediction and is often used in the industry for these models .

2 LITERATURE SURVEY

2.1 Existing Problem

Machine learning helps to learn data from its own experience and in prediction and decision of data. It has brought the revolution in the field of computer science because data is the most important thing in the world. To analyze data, Machine Learning has given many solutions by its algorithms.I have gone through many papers before developing this model to collect information. In paper (1), authors have tried to reduce the efforts of banks by generating a model by various machine learning models and explained which of the methods can be accurate. The paper was divided into 4 sections- data collection, comparison of various machine learning models on data, training of data and testing. They have done by the mining the previous data. Author has also mentioned that this system can be integrated with the Automaticprocessing

system in future. In paper (2), authors basically want to know about the nature of client by analyzing various variables.

2.2 Proposed Solution

In this paper, author compares various attributes by exploratory data analysis technique, seven graphs were generated and short-term loan was preferred in the paper. In paper (3), the authors have explained about whether it is safe or not to give loans to the person. They have used map function to predict the records. In paper (4), authors have used decision tree induction algorithm to implement a model and tried to review credit scoring of mortgage loans and criteria for the applicants. Credit score helps in sanctioning of the loan. They have developed a model to predict it is safe or not for loan sanctioning and it was concluded that mostly low-income applicants receive loan approval as they are likely to repay their loan back. Dataset is collected from the Kaggle. In paper (6), authors have tried to evaluate the credit risks and to identify the loan repayment prediction. They have made use of decision tree algorithm in the paper. A test set is utilized to approve the form. In paper (7), authors have used data mining to develop a model and the system has 3 components- preprocessing, classification and database updation. KNN, Binning and naïve-bayesian algorithm was used for prediction model. A hybrid of naïvebayesian and K-means was used for improving the efficiency of the system. In paper (8), authors have used three types of machine algorithm which are Logistic Regression, Decision tree and Random Forest algorithm. After analyzing the data sets on these models, Random Forest algorithm come out with the highest accuracy of all the model. The main aim of this model was to sanction loan to applicants in a short span of time. In paper (9), an ensemble model is used by the authors which gives better accuracy and efficiency than the individual models for prediction. In paper (10), authors have used classifier technique in which combination of min-max normalization and KNN algorithm is used which gives 75.08% accuracy result. R programming language is used to build KNN soring credit model. This paper focuses on predicting the loan status in commercial banks. It cannot be clear which model is best because each model has its own specification. Therefore, an efficient model is important to be developed which can give higher accuracy.

3 THEORITICAL ANALYSIS

3.1

1 INTRODUCTION

1.1 Overview

The immense increase in capitalism, the fast-paced development and

instantaneous changes in the lifestyle has us in awe. Emi, loans at nominal rate, housing loans, vehicle loans, these are some of the few words which have skyrocketed from the past few years. The needs, wants and demands have never been increased this before. People gets loan from banks; however, it may be baffling for the bankers to judge who can pay back the loan nevertheless the bank shouldn't be in loss. Banks earn most of their profits through the loan sanctioning. Generally, bankspass loan after completing the numerous verification processes despite all these, it is still not confirmed that the borrower will pay back the loan or not. To get over the dilemma, I have built up a prediction model which says if the loan has been assigned in the safe hands or not. Government agencies like keep under surveillance why one person got a loan and the other person could not. In Machine Learning techniques which include classification and prediction can be applied to conquer this to a brilliant extent. Machine learning has eased today's world by developing these prediction models.

1.2 Purpose

By using this we can find whether a person is eligible for the loan or not.Here we will be using the fine techniques of machine learning – Decision tree algorithm to build this prediction model for loan assessment. It is as so because decision tree gives accuracy in the prediction and is often used in the industry for these models .

2 LITERATURE SURVEY

2.1 Existing Problem

Machine learning helps to learn data from its own experience and in prediction and decision of data. It has brought the revolution in the field of computer science because data is the most important thing in the world. To analyze data, Machine Learning has given many solutions by its algorithms.I have gone through many papers before developing this model to collect information. In paper (1), authors have tried to reduce the efforts of banks by generating a model by various machine learning models and explained which of the methods can be accurate. The paper was divided into 4 sections- data collection, comparison of various machine learning models on data, training of data and testing. They have done by the mining the previous data. Author has also mentioned that this system can be integrated with the Automaticprocessing system in future. In paper (2), authors basically want to know about the nature of client by analyzing various variables.

2.2 Proposed Solution

In this paper, author compares various attributes by exploratory data analysis technique, seven graphs were generated and short-term loan was preferred in the paper. In paper (3), the authors have explained about whether it is safe or not to give loans to the person. They have used map function to predict the records. In paper (4), authors have used decision tree induction algorithm to implement a model and tried to review credit scoring of mortgage loans and criteria for the applicants. Credit score helps in sanctioning of the loan. They have developed a model to predict it is safe or not for loan sanctioning and it was concluded that mostly low-income applicants receive loan approval as they are likely to repay their loan back. Dataset is collected from the Kaggle. In paper (6), authors have tried to evaluate the credit risks and to identify the loan repayment prediction. They have made use of decision tree algorithm in the paper. A test set is utilized to approve the form. In paper (7), authors have used data mining to develop a model and the system has 3 components- preprocessing, classification and database updation. KNN, Binning and naïve-bayesian algorithm was used for prediction model. A hybrid of naïvebayesian and K-means was used for improving the efficiency of the system. In paper (8), authors have used three types of machine algorithm which are Logistic Regression, Decision tree and Random Forest algorithm. After analyzing the data sets on these models, Random Forest algorithm come out with the highest accuracy of all the model. The main aim of this model was to sanction loan to applicants in a short span of time. In paper (9), an ensemble model is used by the authors which gives better accuracy and efficiency than the individual models for prediction. In paper (10), authors have used classifier technique in which combination of min-max normalization and KNN algorithm is used which gives 75.08% accuracy result. R programming language is used to build KNN soring credit model. This paper focuses on predicting the loan status in commercial banks. It cannot be clear which model is best because each model has its own specification. Therefore, an efficient model is important to be developed which can give higher accuracy.

3 THEORITICAL ANALYSIS

3.1

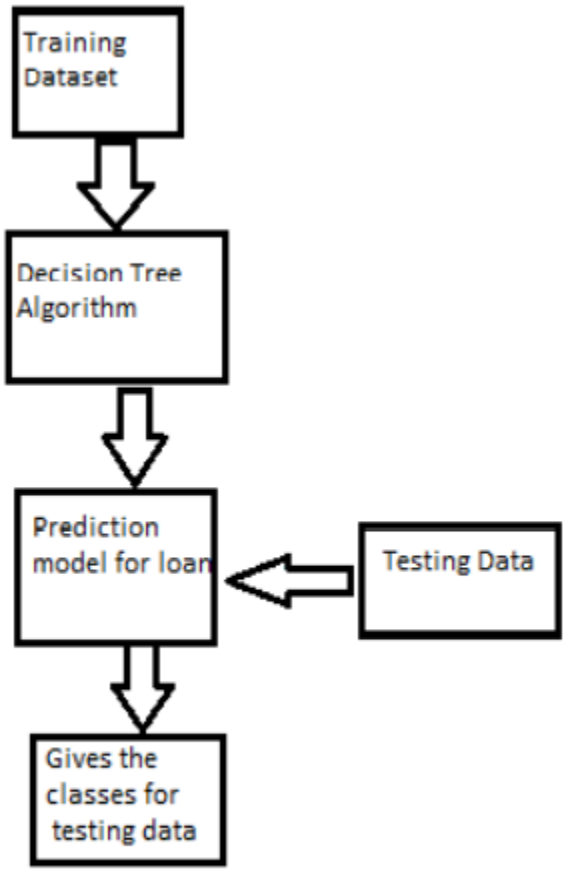
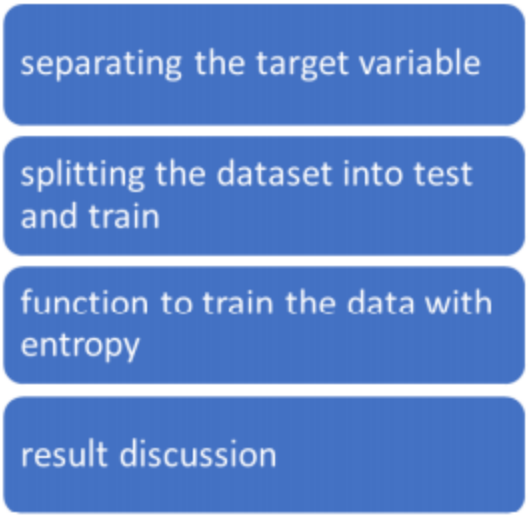


Fig 1: Architecture of proposed model

3.2 Hardware/Software Designing

In the proposed model for loan prediction, Dataset is split into training and testing data. After then training datasets are trained using the decision tree algorithm and a prediction model is developed using the algorithm. Testing datasets are then given to model for the prediction of loan. The motive of this paper is to predict the defaults who will repay the loan or not. Various libraries like pandas, numpy have been used. After the loading of datasets, Data Preprocessing like missing value treatment of numerical and categorical is done by checking the values. Numerical and categorical values are segregated. Outliers and frequency analysis are done, outliers are checked by getting the boxplot diagram of attributes.

4 FLOWCHART



6 RESULT

Expected Result:

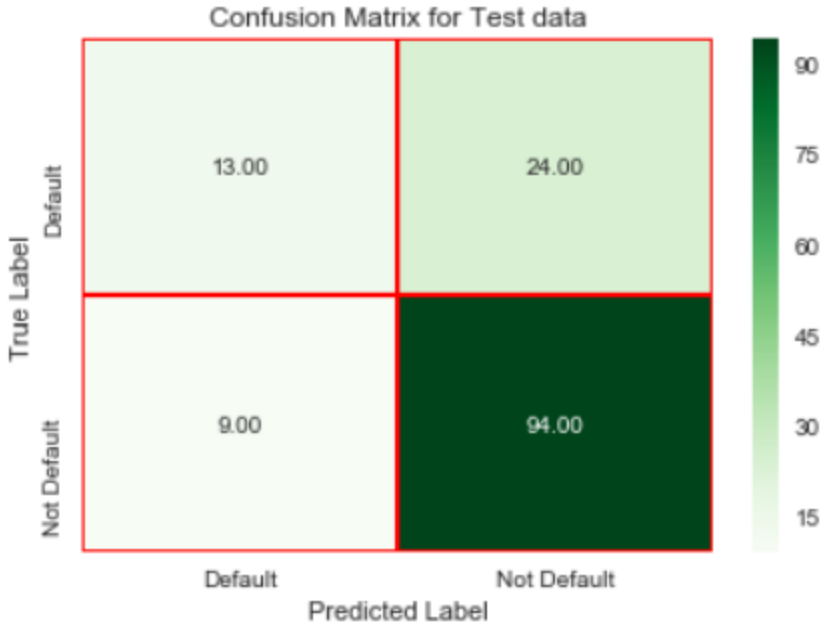
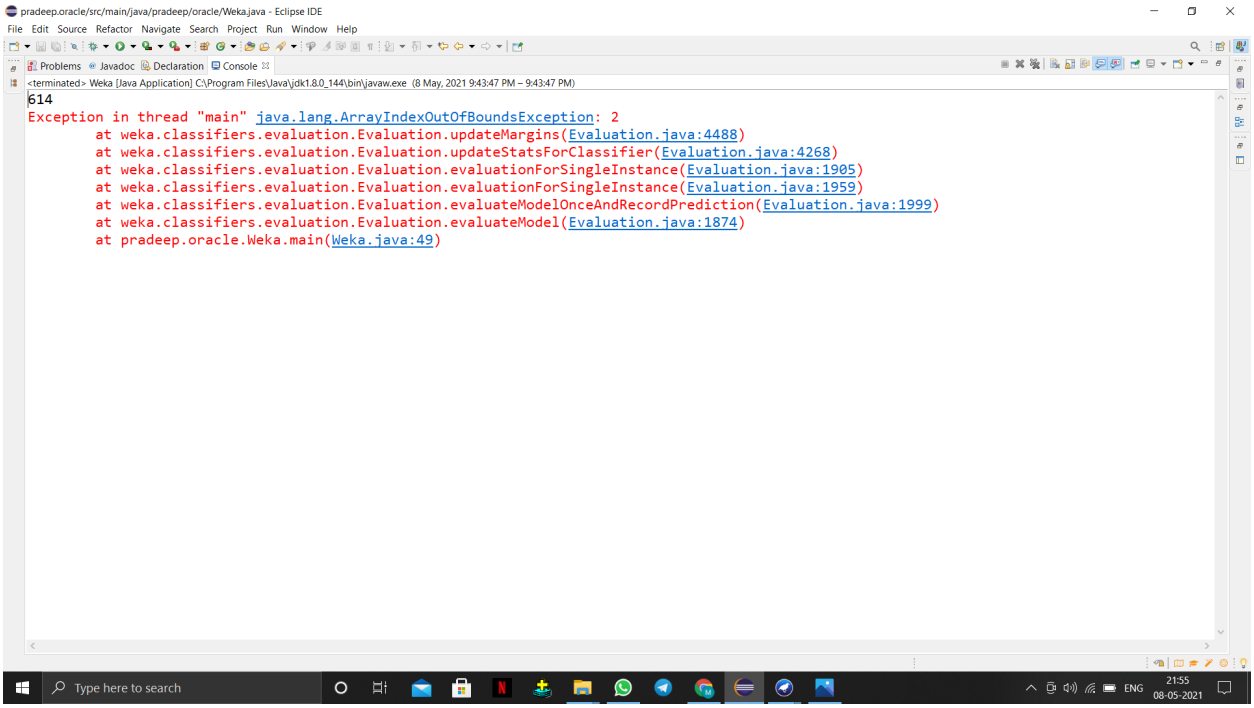
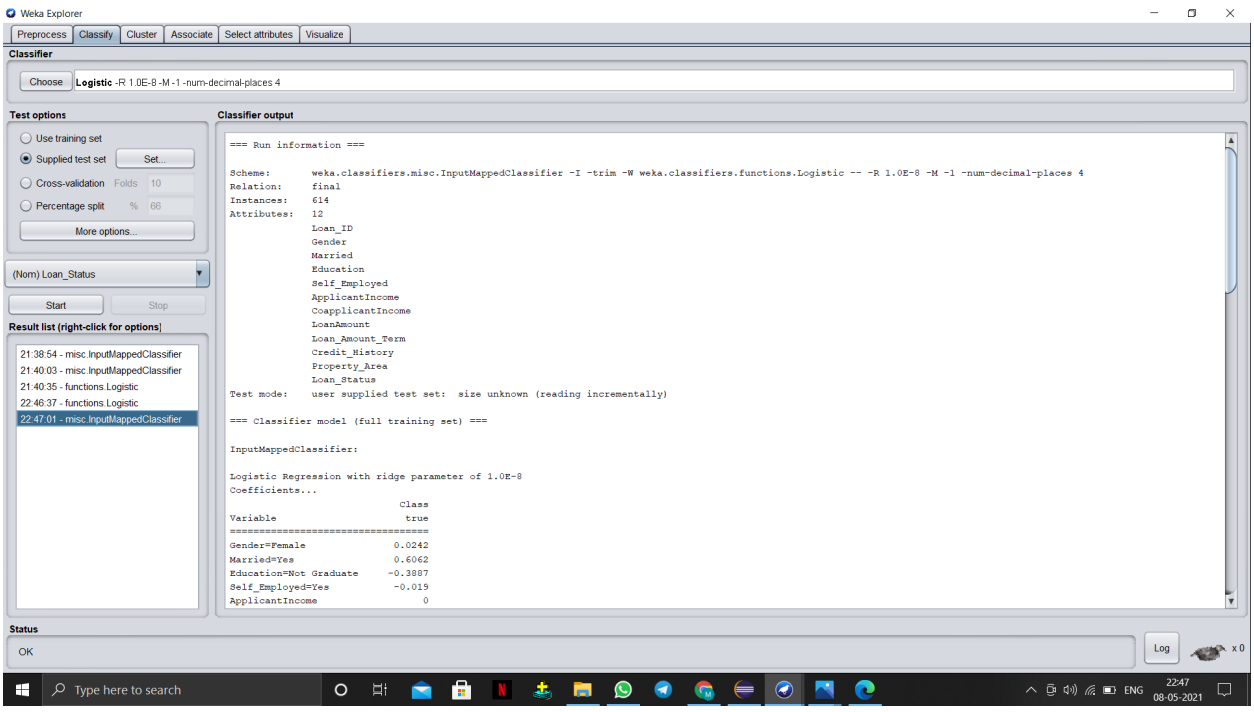
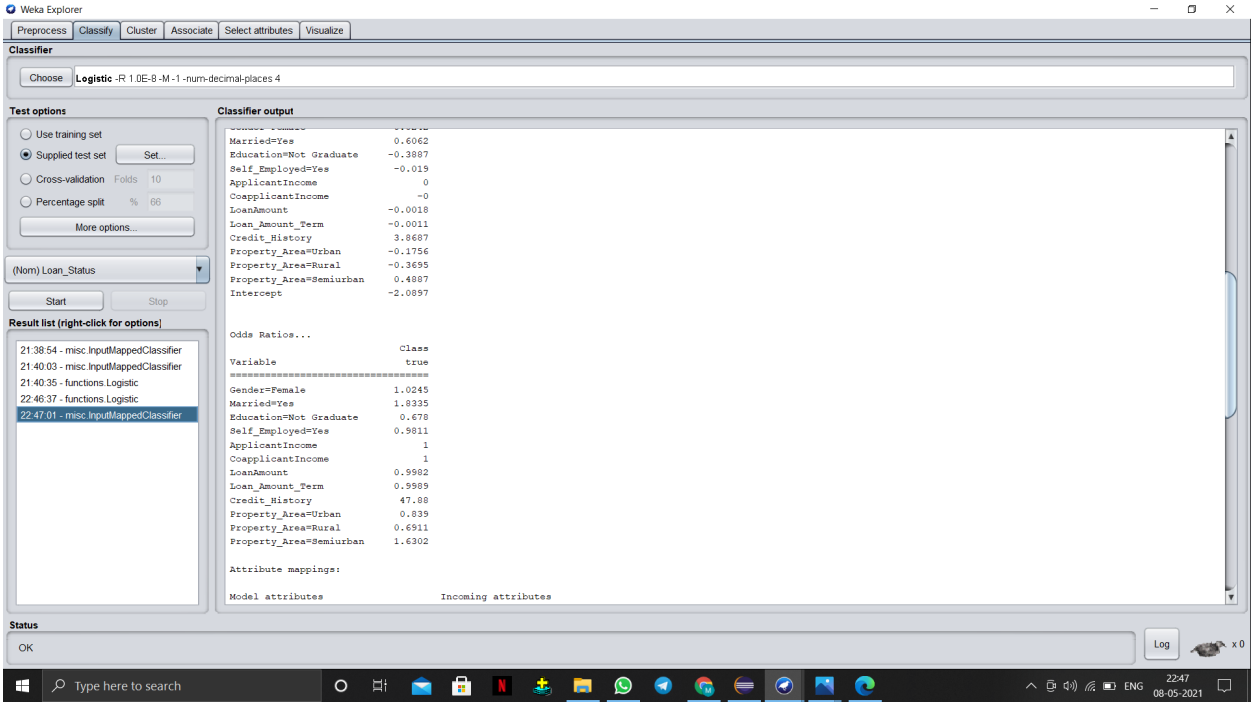


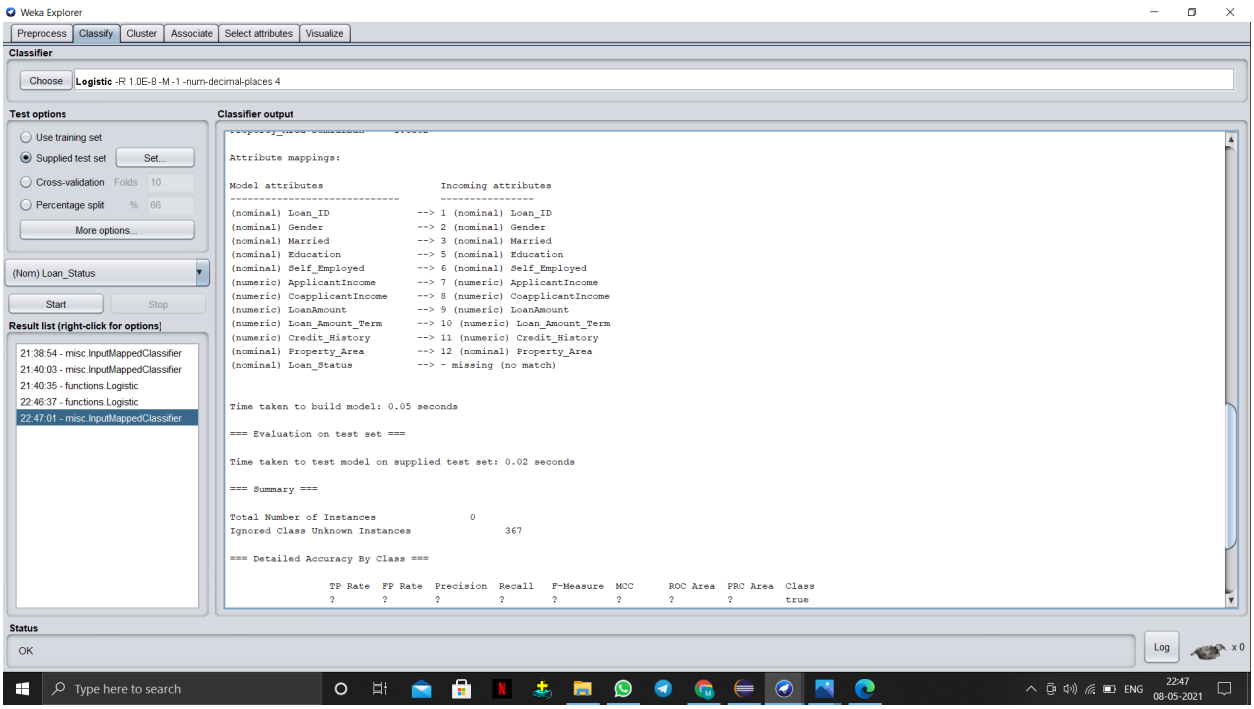
Fig 10: Confusion Matrix for the Test data

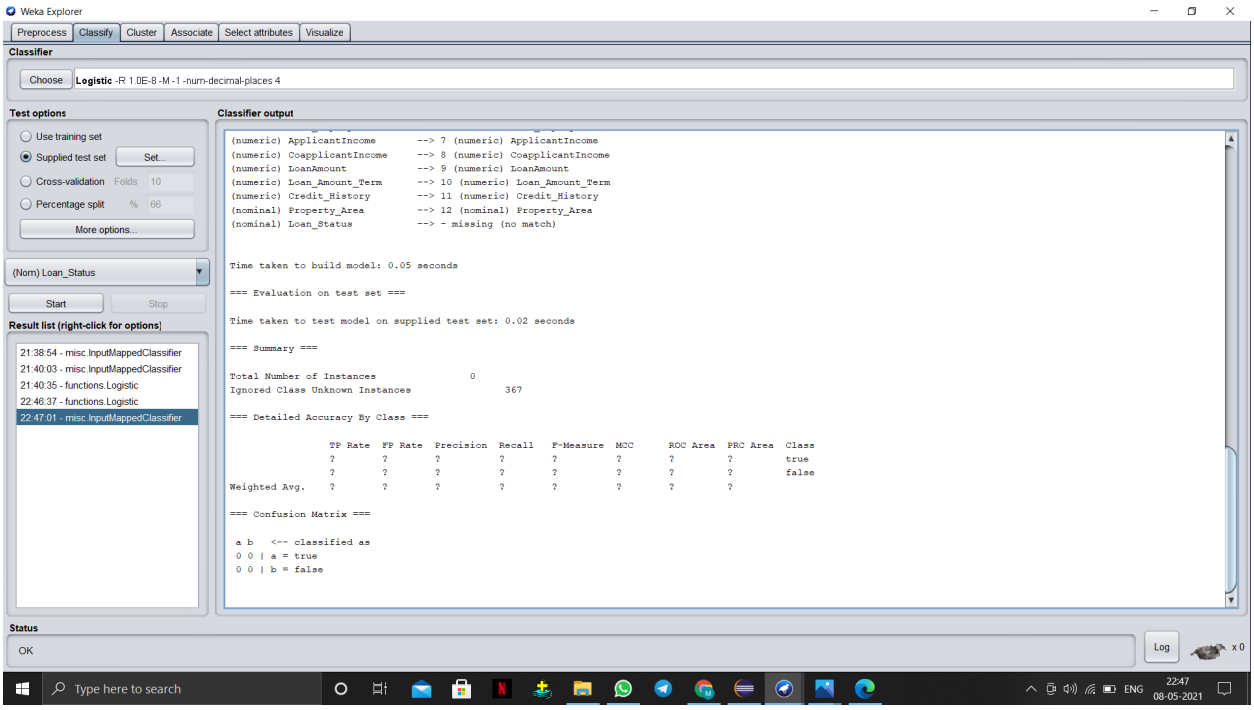
Obtained Result:






Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Loan_Status
Start
Stop
Result list (right-click for options)
21:38:54 - misc.InputMappedClassifier
21:40:03 - misc.InputMappedClassifier
21:40:35 - functions.Logistic
22:46:37 - functions.Logistic
22:47:01 - misc.InputMappedClassifier
Classifier output
Married=Yes 0.6062
Education=Not Graduate -0.3887
Self_Employed=Yes -0.019
ApplicantIncome 0
CoapplicantIncome -0
LoanAmount -0.0018
Loan_Amount_Term -0.0011
Credit_History 3.8687
Property_Area=Urban -0.1756
Property_Area=Rural -0.3695
Property_Area=Semiurban 0.4887
Intercept -2.0897
Odds Ratios...
Class
Variable true
Gender=Female 1.0245
Married=Yes 1.8335
Education=Not Graduate 0.678
Self_Employed=Yes 0.9811
ApplicantIncome 1
CoapplicantIncome 1
LoanAmount 0.9982
Loan_Amount_Term 0.9989
Credit_History 47.88
Property_Area=Urban 0.839
Property_Area=Rural 0.6911
Property_Area=Semiurban 1.6302
Attribute mappings:
Model attributes Incoming attributes
Status
OK
Log
x 0
Type here to search
ENG
22:47
08-05-2021


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Loan_Status
Start
Stop
Result list (right-click for options)
21:38:54 - misc.InputMappedClassifier
21:40:03 - misc.InputMappedClassifier
21:40:35 - functions.Logistic
22:46:37 - functions.Logistic
22:47:01 - misc.InputMappedClassifier
Classifier output
Attribute mappings:
Model attributes Incoming attributes
(nominal) Loan_ID --> 1 (nominal) Loan_ID
(nominal) Gender --> 2 (nominal) Gender
(nominal) Married --> 3 (nominal) Married
(nominal) Education --> 5 (nominal) Education
(nominal) Self_Employed --> 6 (nominal) Self_Employed
(numeric) ApplicantIncome --> 7 (numeric) ApplicantIncome
(numeric) CoapplicantIncome --> 8 (numeric) CoapplicantIncome
(numeric) LoanAmount --> 9 (numeric) LoanAmount
(numeric) Loan_Amount_Term --> 10 (numeric) Loan_Amount_Term
(numeric) Credit_History --> 11 (numeric) Credit_History
(nominal) Property_Area --> 12 (nominal) Property_Area
(nominal) Loan_Status --> - missing (no match)
Time taken to build model: 0.05 seconds
=== Evaluation on test set ===
Time taken to test model on supplied test set: 0.02 seconds
=== Summary ===
Total Number of Instances 0
Ignored Class Unknown Instances 367
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
? ? ? ? ? ? ? ? true
Status
OK
Log
x 0
Type here to search
ENG
22:47
08-05-2021



7 ADVANTAGES

- Fast
- More Reliable
- Good Accuracy
- Work can be done easily with machine learning

DISADVANTAGES

- Accuracy is not 100 percent
- Errors may occur
- Sometimes the obtained output can be wrong

8 APPLICATIONS

- Used in banks for providing loans.
- Used for financers to decide whether the applicant can repay or not within time.

9 CONCLUSION

After this work, we are able to conclude that Decision tree version is extraordinary efficient and gives a higher end result. We have developed a model which can easily predict that the person will repay its loan or not. we can see our model has reduced the efforts of bankers. Machine learning has helped a lot in developing this model which gives precise results.

10 FUTURE SCOPE

1. Kumar Arun,GaurIshan,KaurSanmit, “Loan Prediction based on machine Learning approach”, IOSR Journal of computer engineering, pp.18-21 2016.

2. X.FrencisJensy, V.P.Sumathi,Janani Shiva Shri, “An exploratory Data Analysis for Loan

Prediction based on nature of clients", International Journal of Recent Technology and Engineering (IJRTE),Volume-7 Issue-4S, November 2018.
3. PidikitiSupriya, MyneediPavani, NagarapuSaisushma,NamburiVimalaKumari, k Vikash,"Loan Prediction by using Machine Learning Models", International Journal of Engineering and Techniques.Volume 5 Issue 2, Mar-Apr 2019
4. Nikhil Madane, SiddharthNanda,"Loan Prediction using Decision tree", Journal of the Gujrat Research History,Volume 21 Issue 14s, December 2019.
5. Dataset Source, https://www.kaggle.com
6. Oracle Academy.