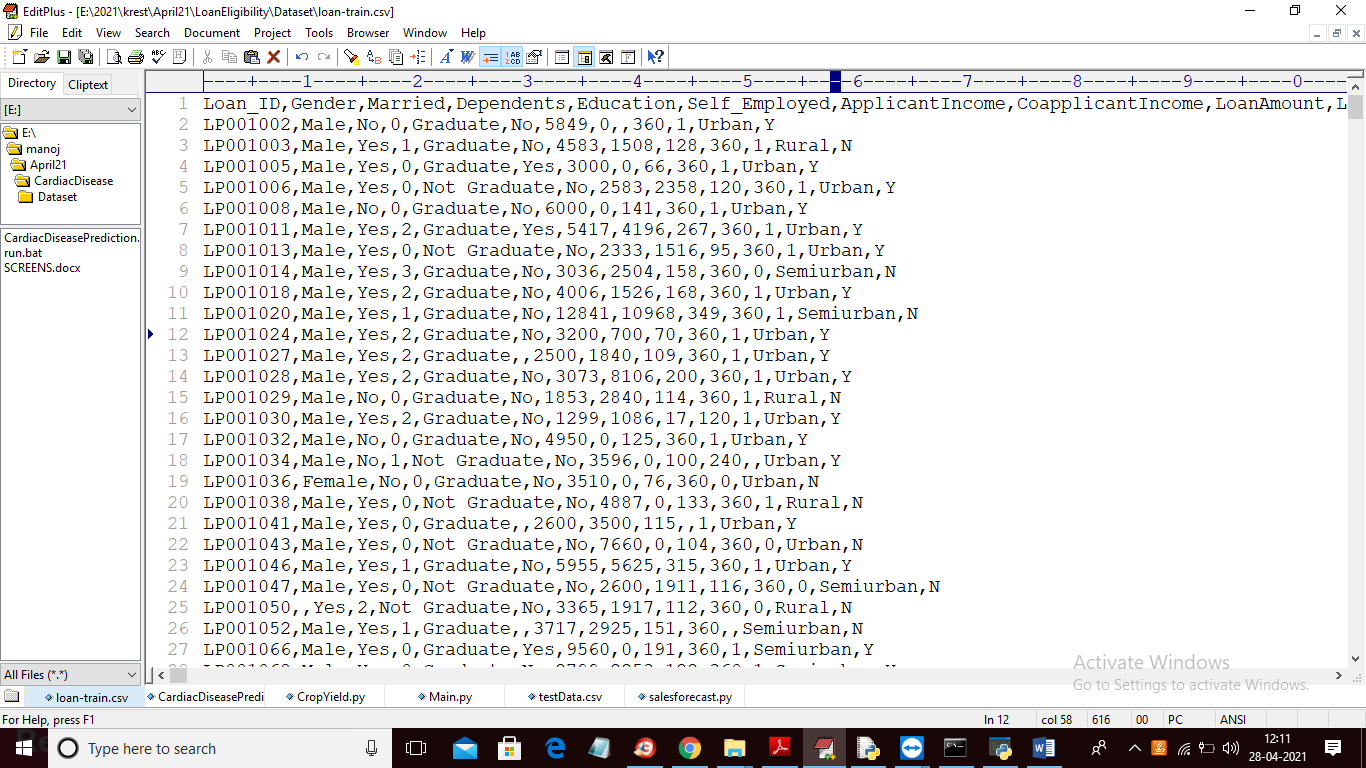
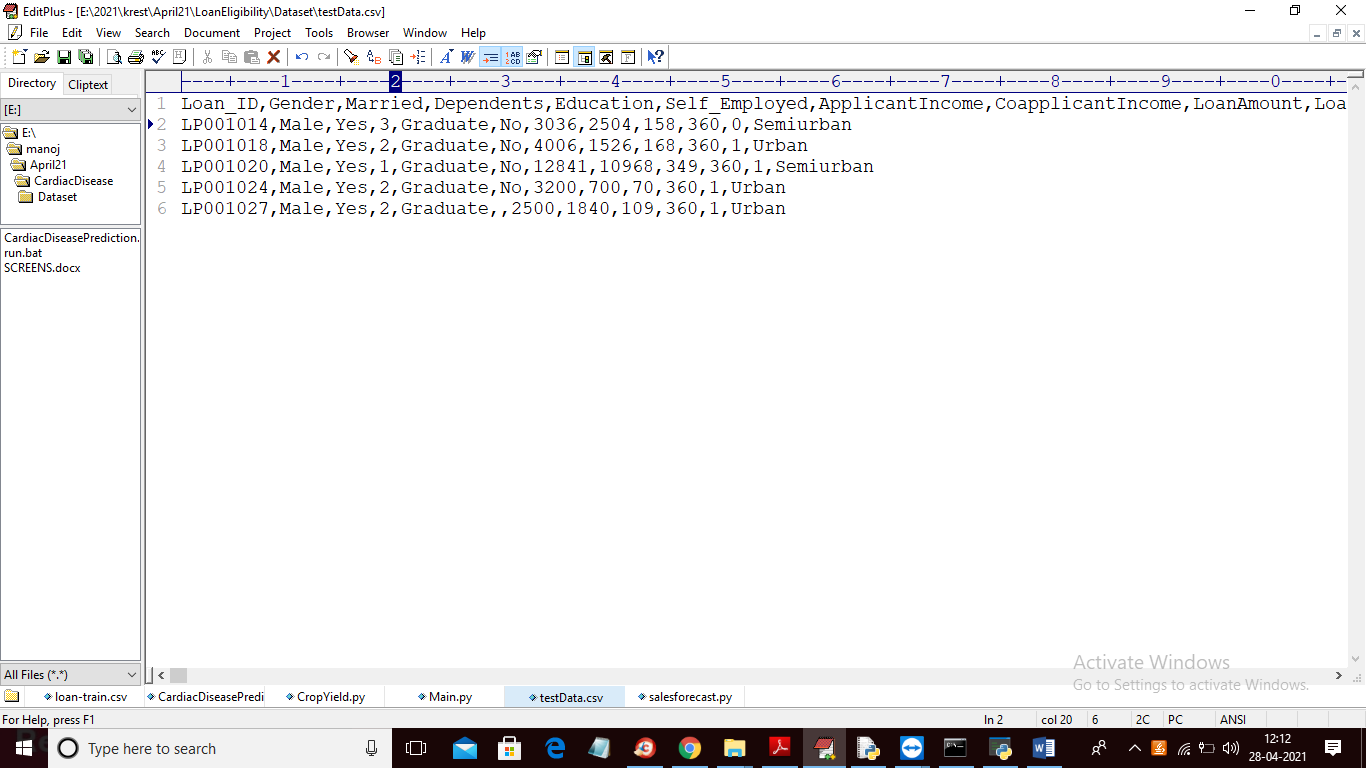
PREDICTION OF LOAN ELIGIBILITY OF THE CUSTOMER

In this project we are using machine learning algorithm called Random Forest to predict loan eligibility and to train this random forest we are using below dataset



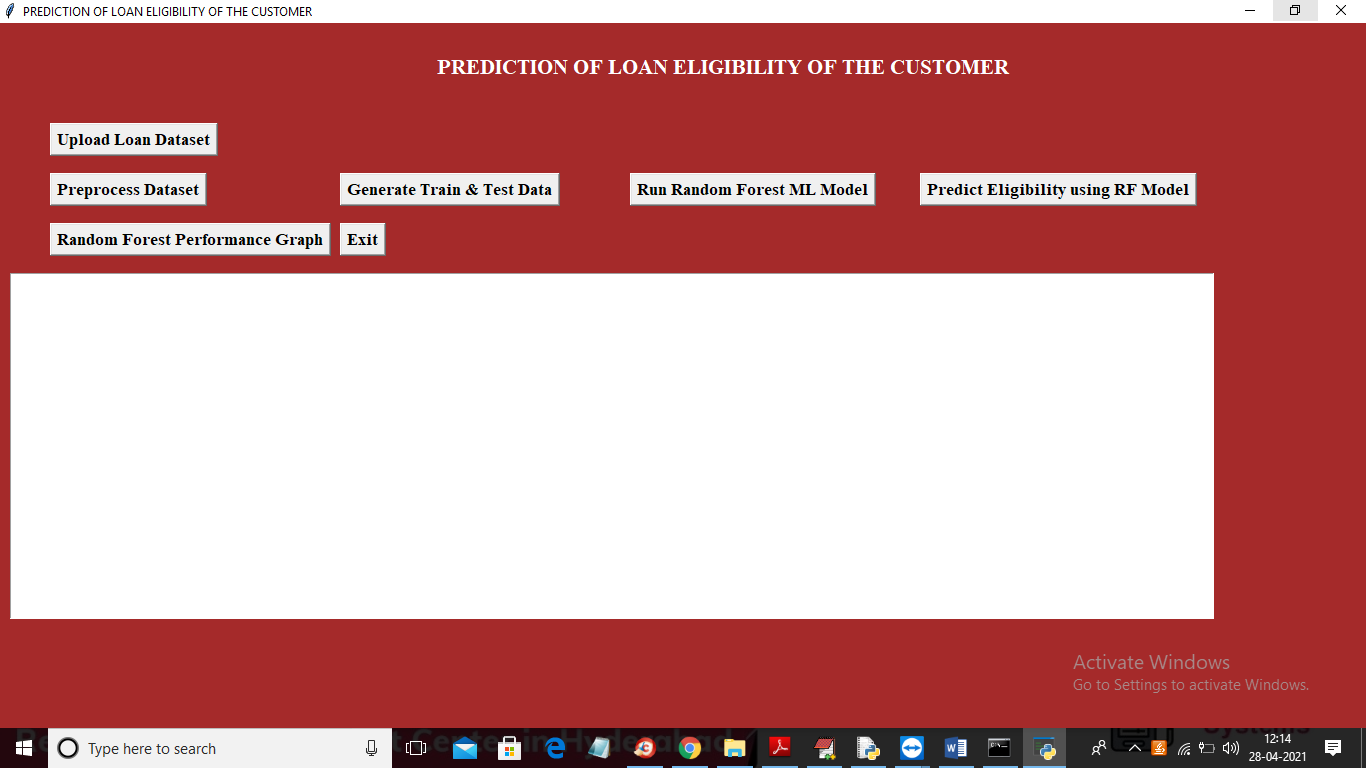
In above dataset in first row we can see dataset column names and in other rows we have dataset values and in last column we have class label as Y or N where Y means eligible and N means not eligible and now we used above dataset to train machine learning model and after training we will upload test dataset and then application will predict class label Y or N and below is test dataset screen shots



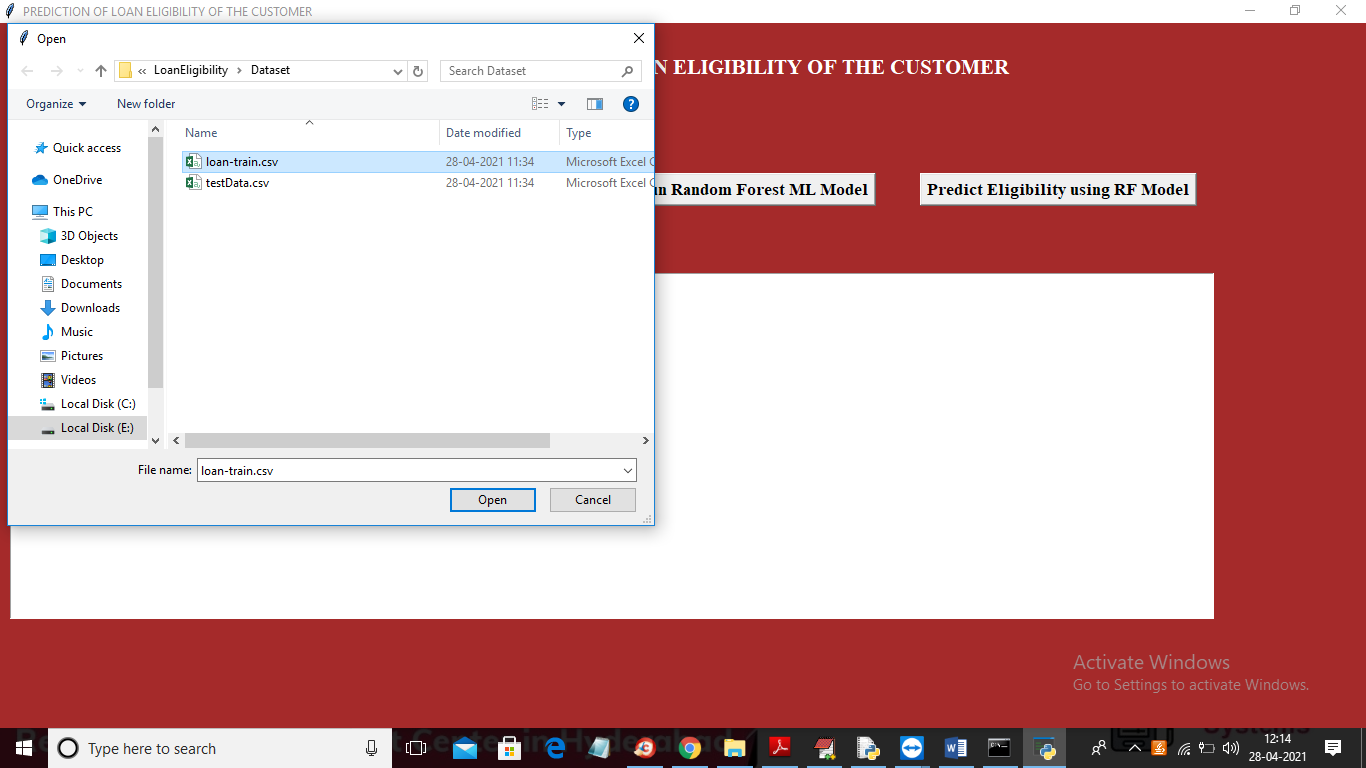
In above test data we don’t have any N or Y class label and by analysing above records machine learning will predict eligibility.

SCREEN SHOTS

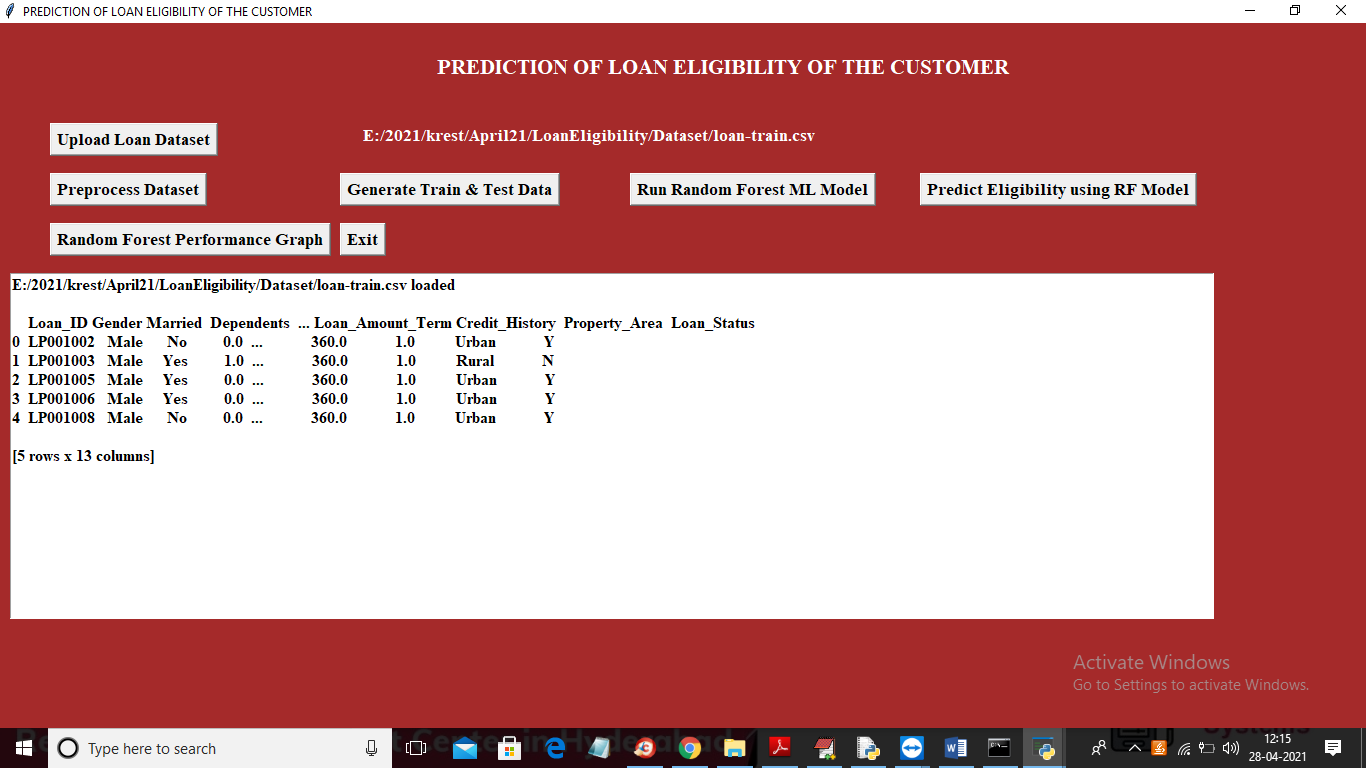
To run project double click on ‘run.bat’ file to get below screen



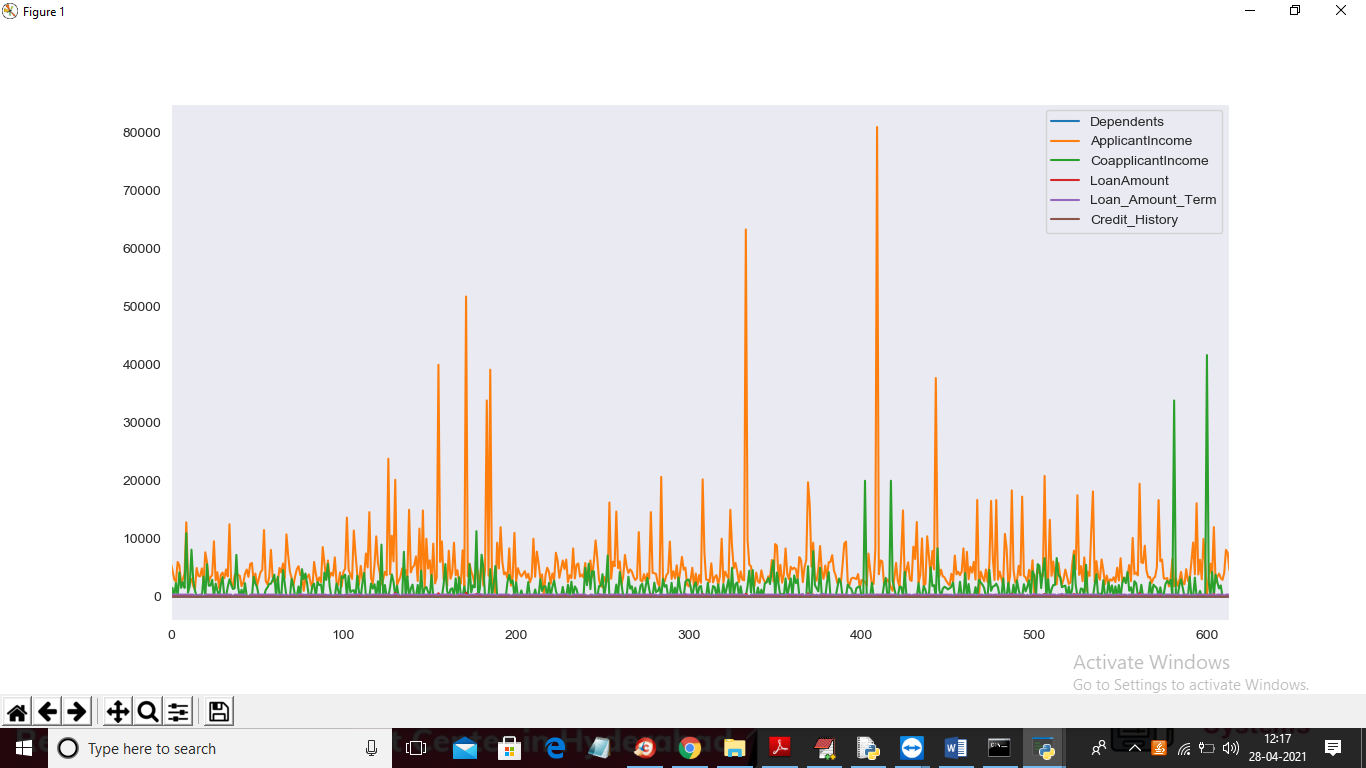
In above screen click on ‘Upload Loan Dataset’ button to load dataset



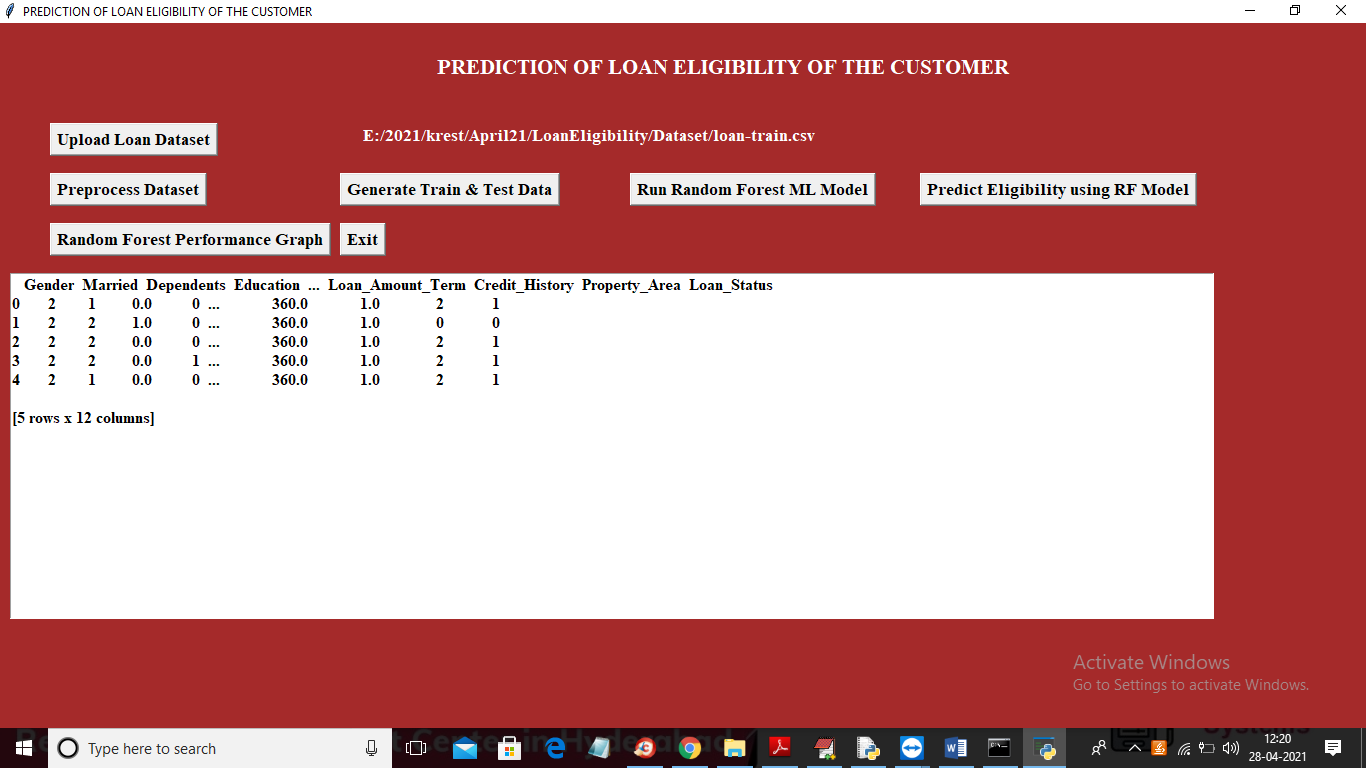
In above screen selecting and uploading ‘loan-train.csv’ file and then click on ‘Open’ button to load dataset and to get below screen



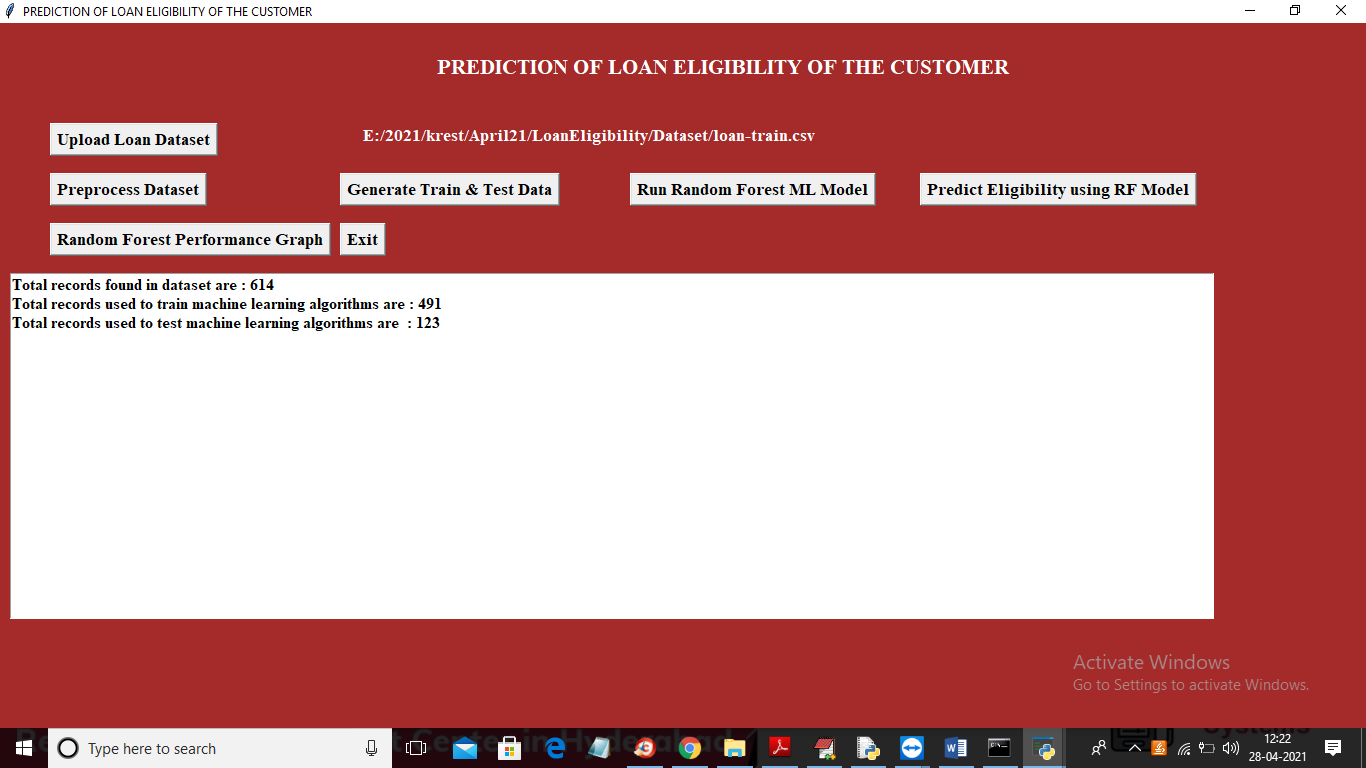
In above screen dataset loaded and all columns contains non-numeric values and machine learning will not accept non-numeric values so we need to convert all those values to numeric by assigning ID’s to them where MALE will replace with 0 and FEMALE will replace with 1 and below graph showing number of different values in dataset



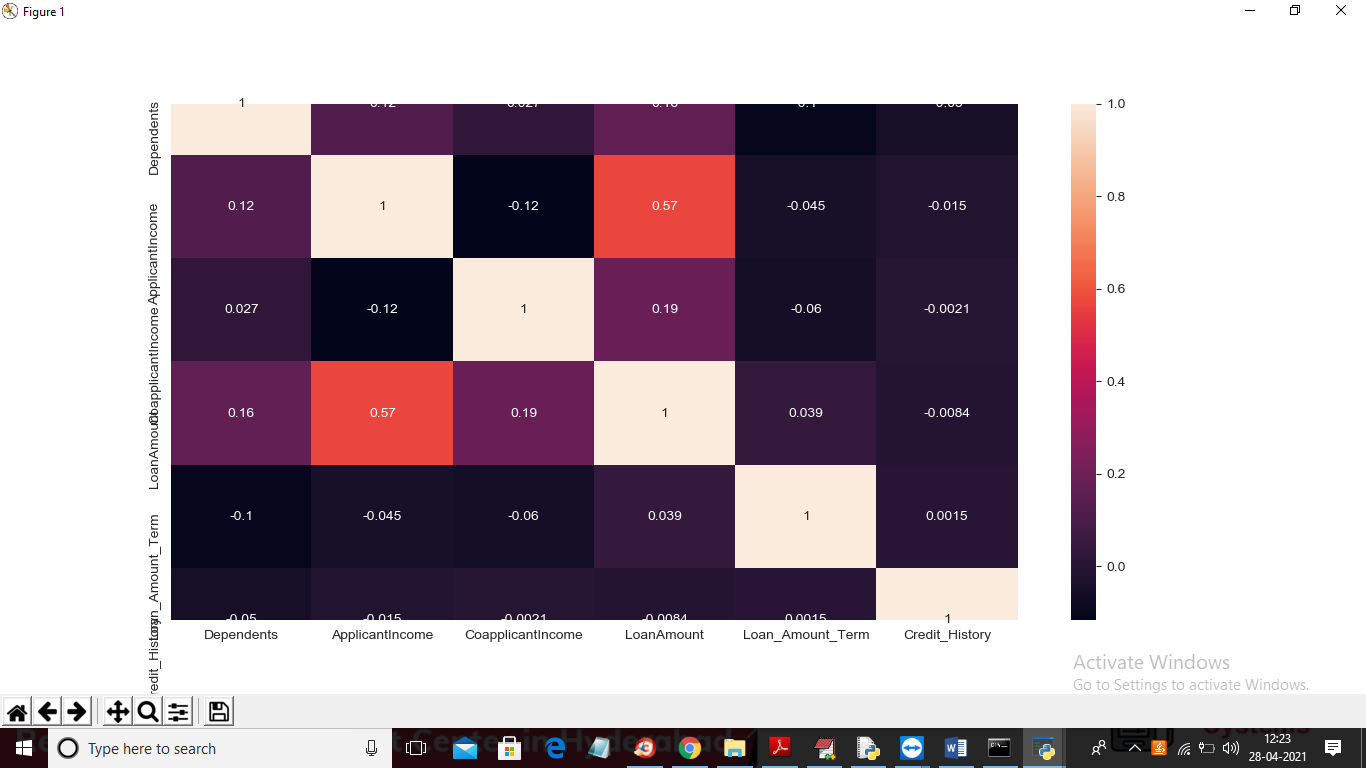
In above graph different colour lines represents counts of that column and you can see column names with colour in graph top right side. Now click on ‘Preprocess Dataset’ button to clean dataset



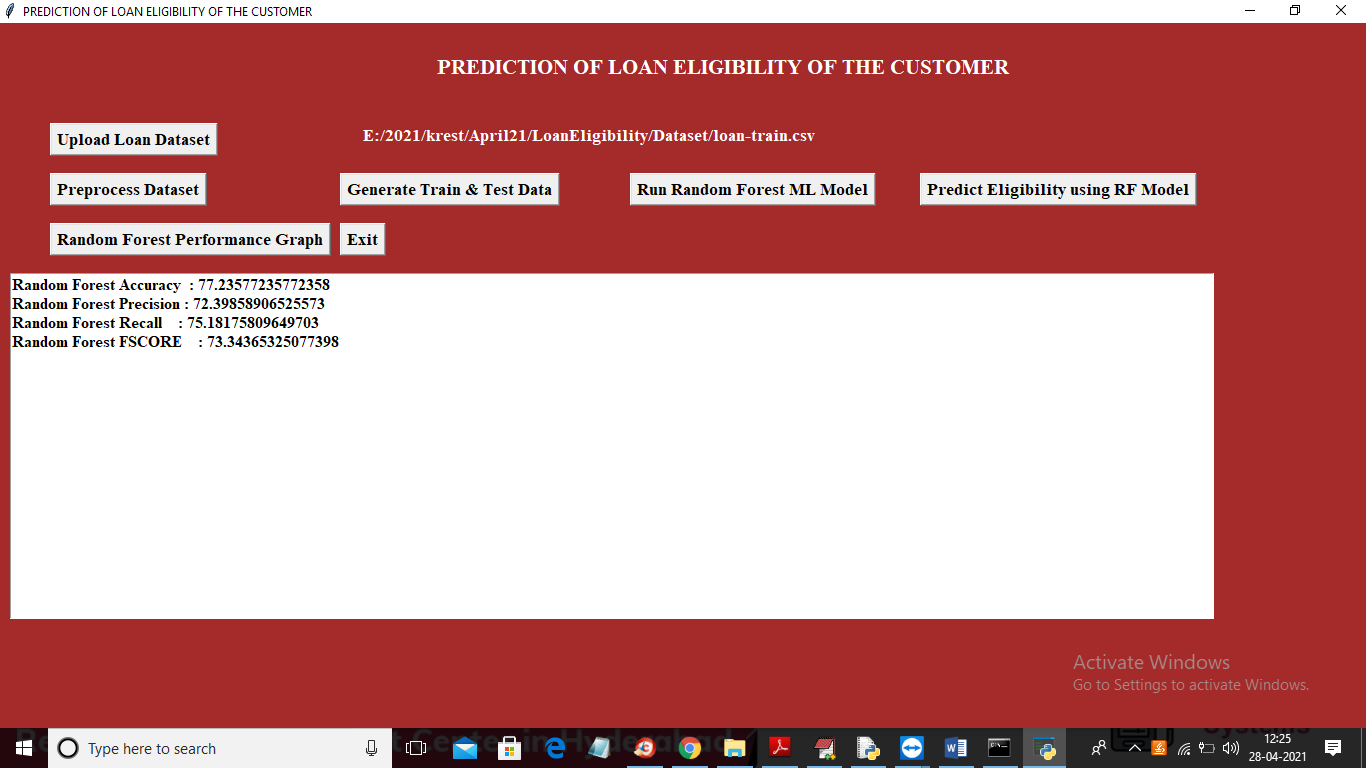
In above screen all non-numeric data is replace with numeric values and now click on ‘Generate Train & Test Data’ button to split dataset into train and test part



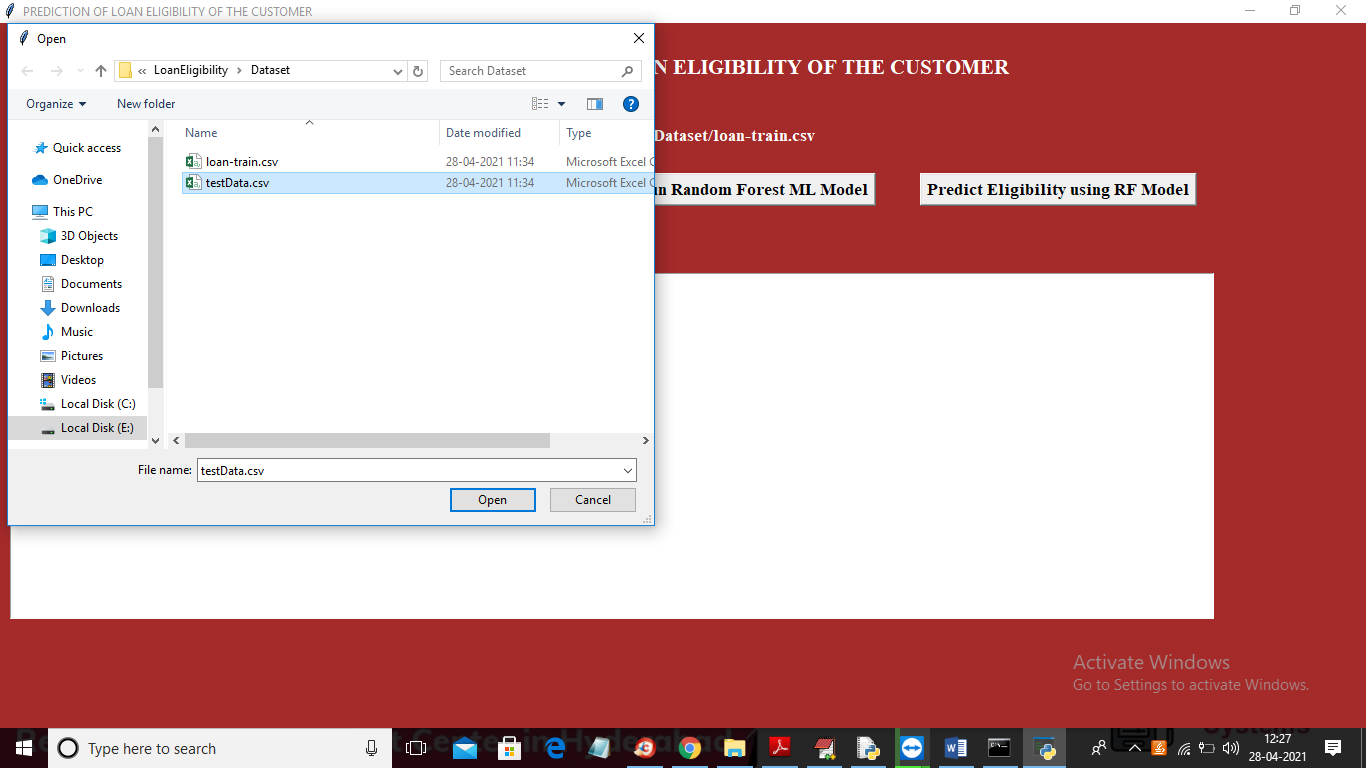
In above screen dataset contains 614 records and using 491 records to train ML and 123 records to test ML accuracy. In below graph we can see importance of each attribute with other attribute by using graph correlation metric



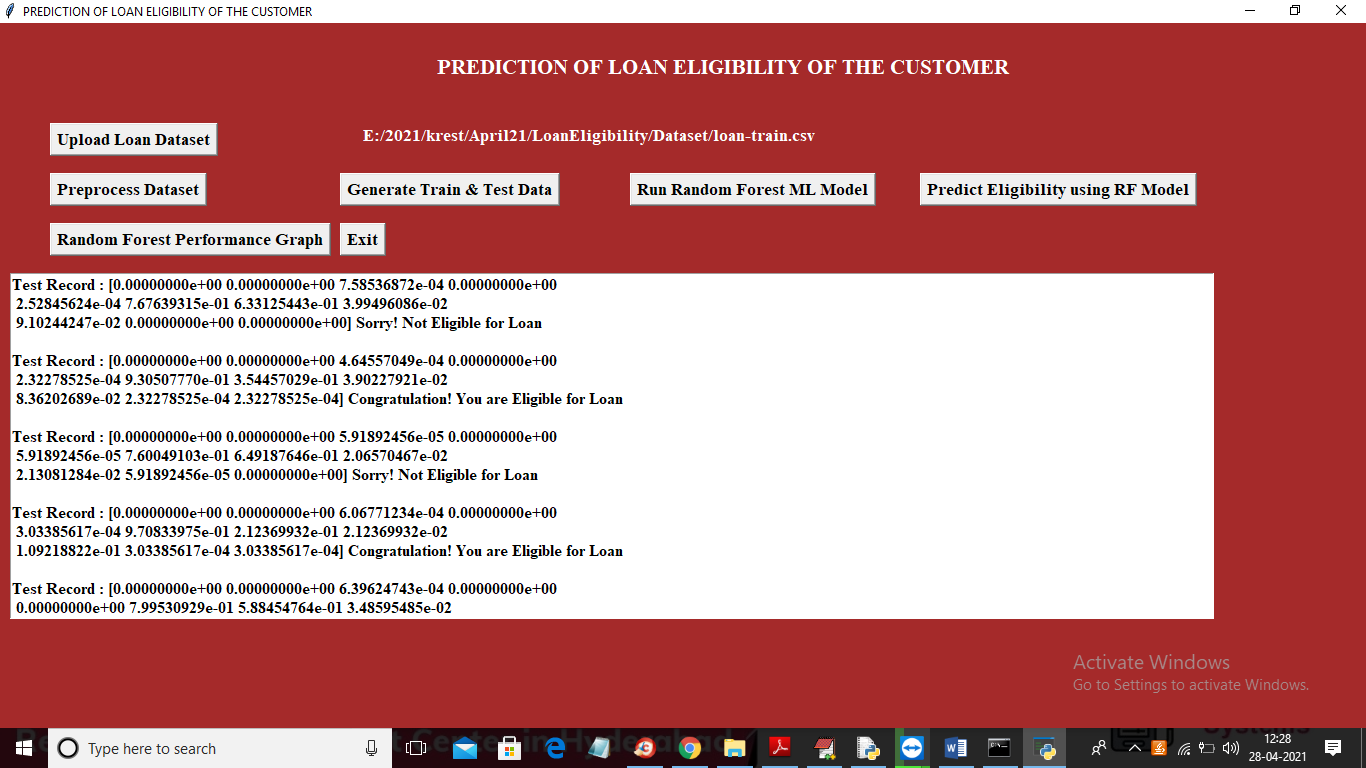
In above graph whatever column in x-axis and y-axis having value >0 will be consider as important features or column. Now click on ‘Run Random Forest Ml Model’ to build random forest model on above dataset



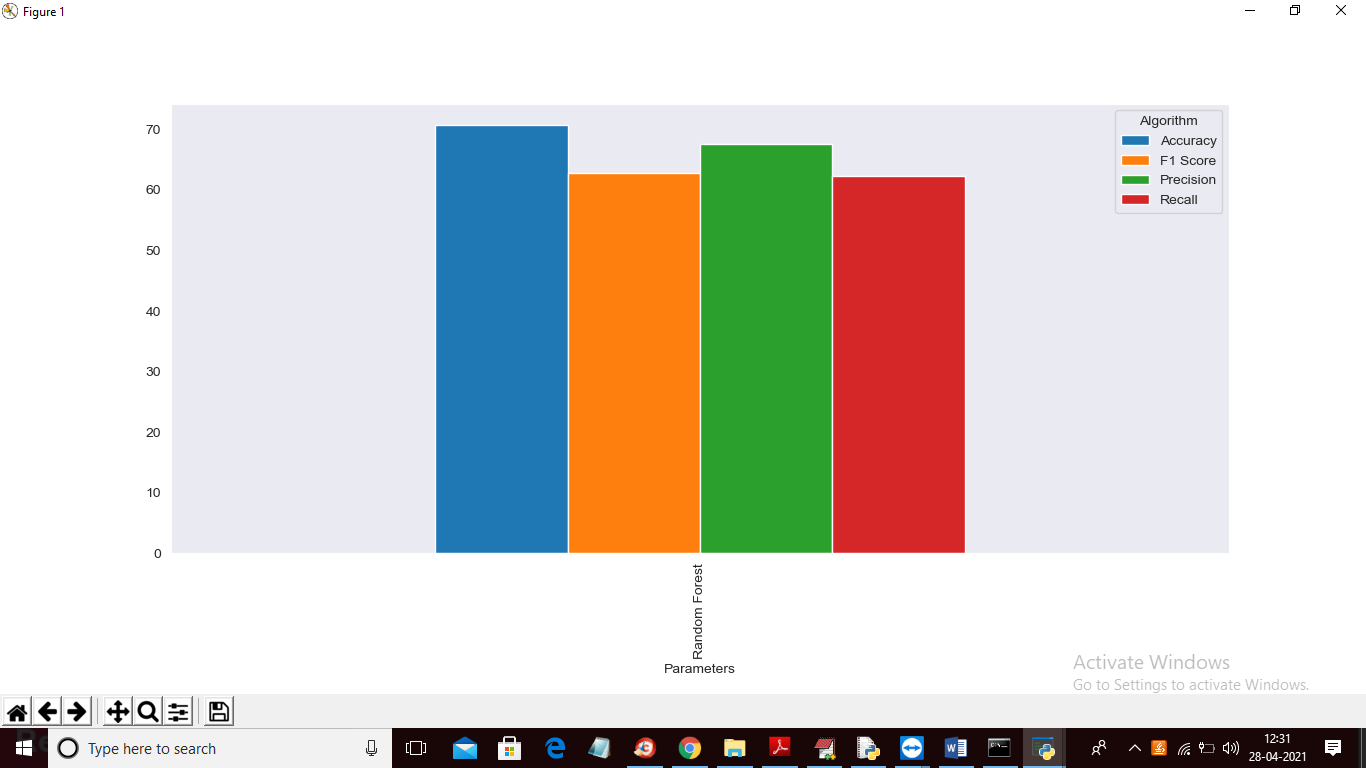
In above screen random forest model generated with 77% accuracy and we can see its precision, recall and FSCORE value and now click on ‘Predict Eligibility using RF Model’ button to upload test data and perform eligibility prediction



In above screen selecting and uploading ‘testData.csv’ file and then click on ‘Open’ button to load test data and then will get below prediction result



In above screen in square bracket we can see normalized test values and after square bracket we can see the prediction result as eligible or not eligible. You can scroll down above text area to view all predicted records and now click on ‘Random Forest Performance Graph’ button to get below graph



In above graph we can see accuracy, precision, recall and FSCORE values of random forest and graph y-axis represents %value where accuracy got 80% and Precision got 65%. Each metric bar colour name you can see from top right side