Future of Loan Approvals with Explainable AI

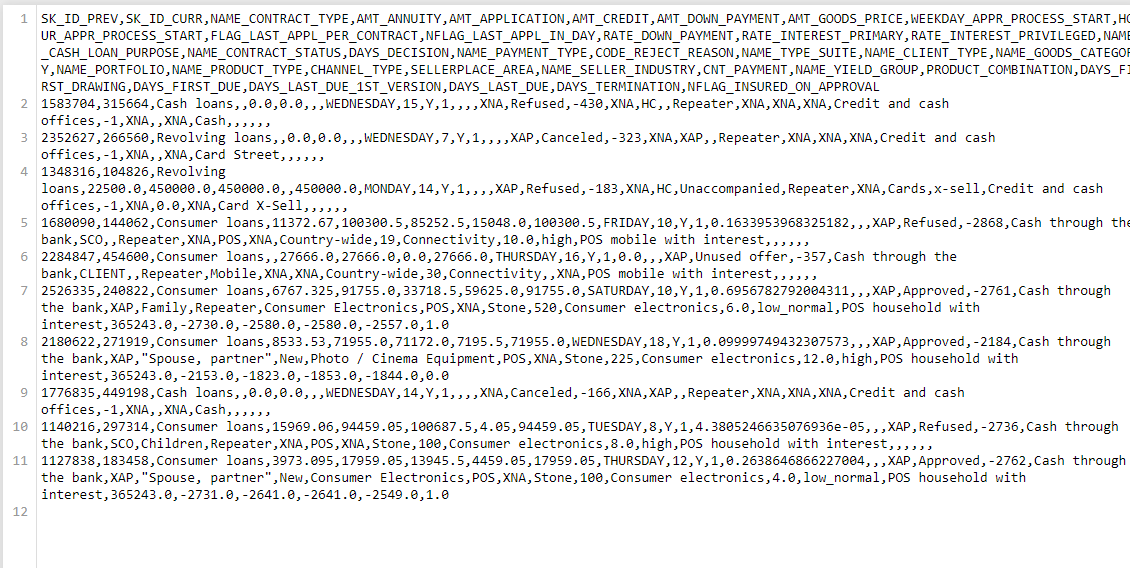
In this project you ask to implement AI algorithm which can explain about model who are using what all features to predict loan approval or rejection, in simple terms it will explain about model prediction and for explanation we are using SHAP module.

In propose AI algorithm Random Forest get trained on loan dataset and after training Random Forest will predict loan status as well as reason of rejection or approval.

We are using loan application dataset which contains both approval or rejection with reasons and this dataset can be downloaded from below KAGGLE website

<https://www.kaggle.com/code/idrisguclu/home-credit-risk-previous-application/input>

from above link only one dataset called ‘Previous\_Application.csv’ contains REASON and approval dataset and below is the dataset details



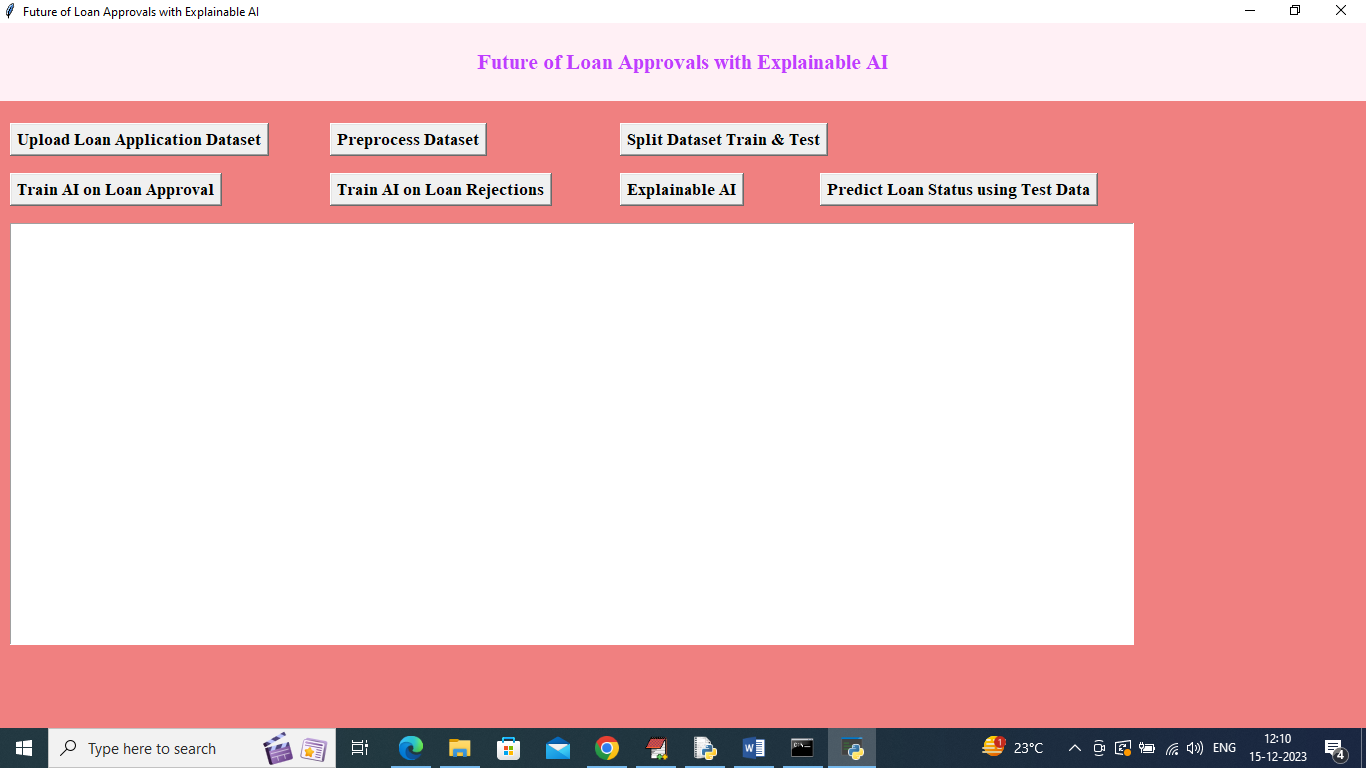
In above dataset screen first row represents dataset column names and remaining rows represents dataset values and by using above dataset we will train AI to predict loan approval status and reason.

To implement this project we have designed following modules

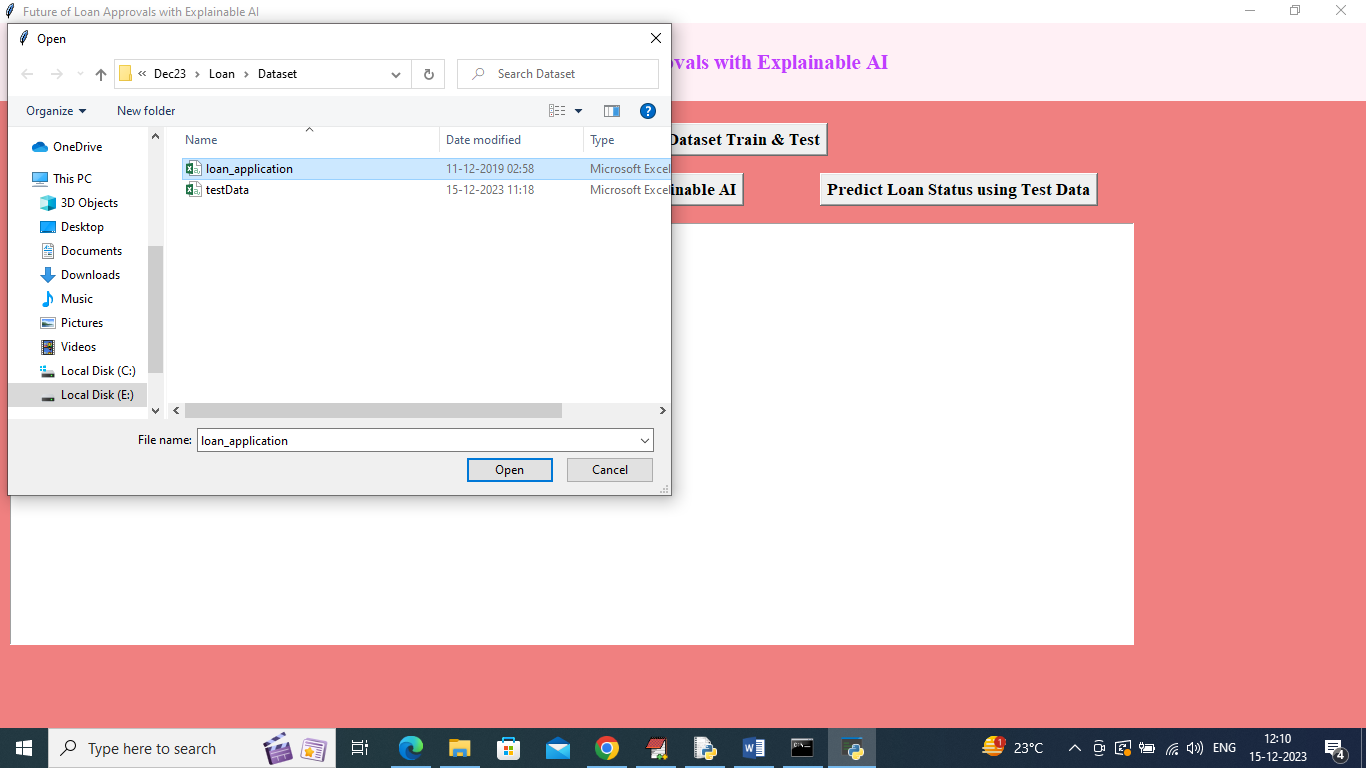
1. Upload Loan Application Dataset: using this module we will upload dataset to application and then application will read entire dataset and then find all class labels for loan and reject reason and plot them in a graph
2. Pre-process Dataset: dataset contains missing value and both numeric and non-numeric data so by employing label encoder class will convert all data into numeric format and then normalized all dataset values to make it clean.
3. Split Dataset Train & Test: using this module will split Dataset in to train and test where application using 80% dataset for training and 20% for testing
4. Train AI on Loan Approval: using this module will train AI on 80% training data to predict loan approval status and then perform prediction on 20% test data to calculate prediction accuracy
5. Train AI on Loan Rejections: using this module will train AI on 80% training data to predict REJECTION REASON and then perform prediction on 20% test data to calculate prediction accuracy
6. Explainable AI: using this module we will explain about features which are contributing most for label prediction
7. Predict Loan Status using Test Data: using this module we will upload test data and then AI model will predict loan status and then predict loan approval or rejection REASON.

SCREEN SHOTS

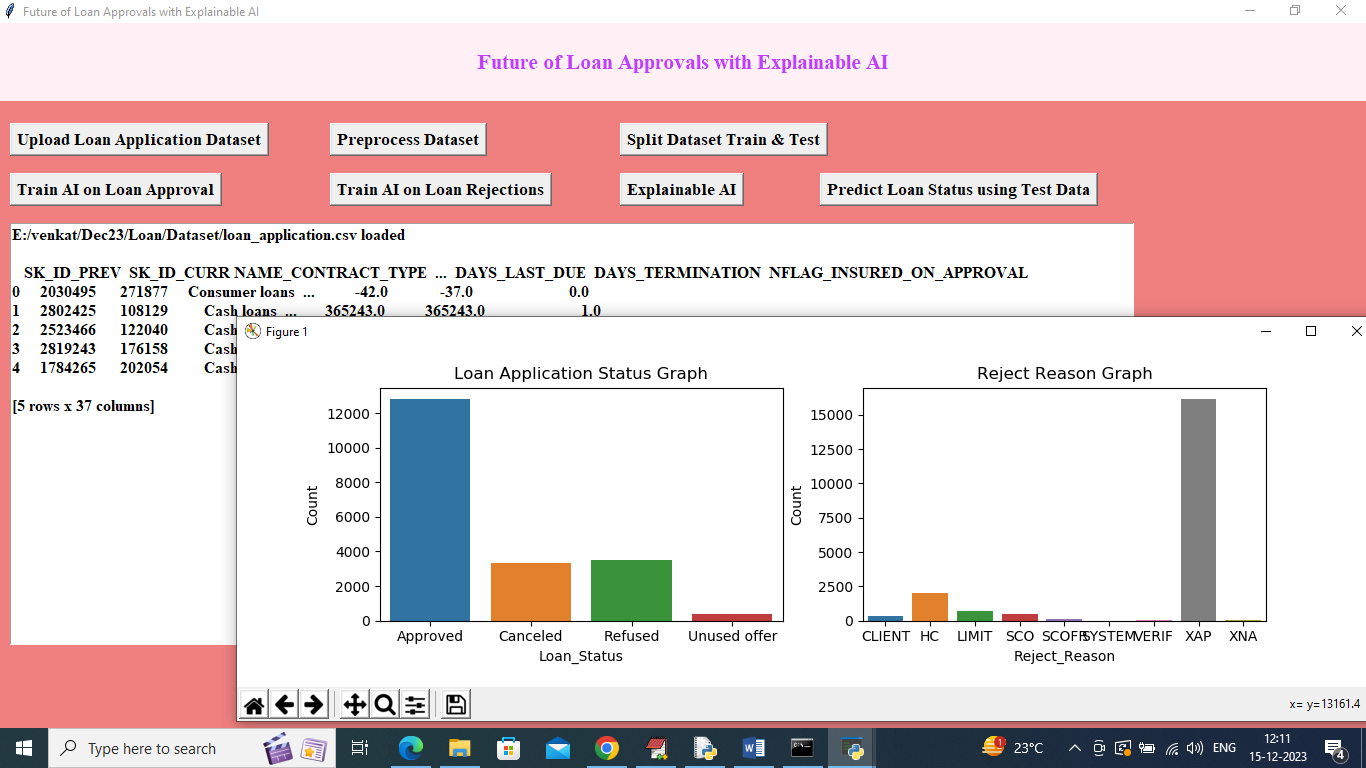
To run project double click on ‘run.bat’ file to get below screen



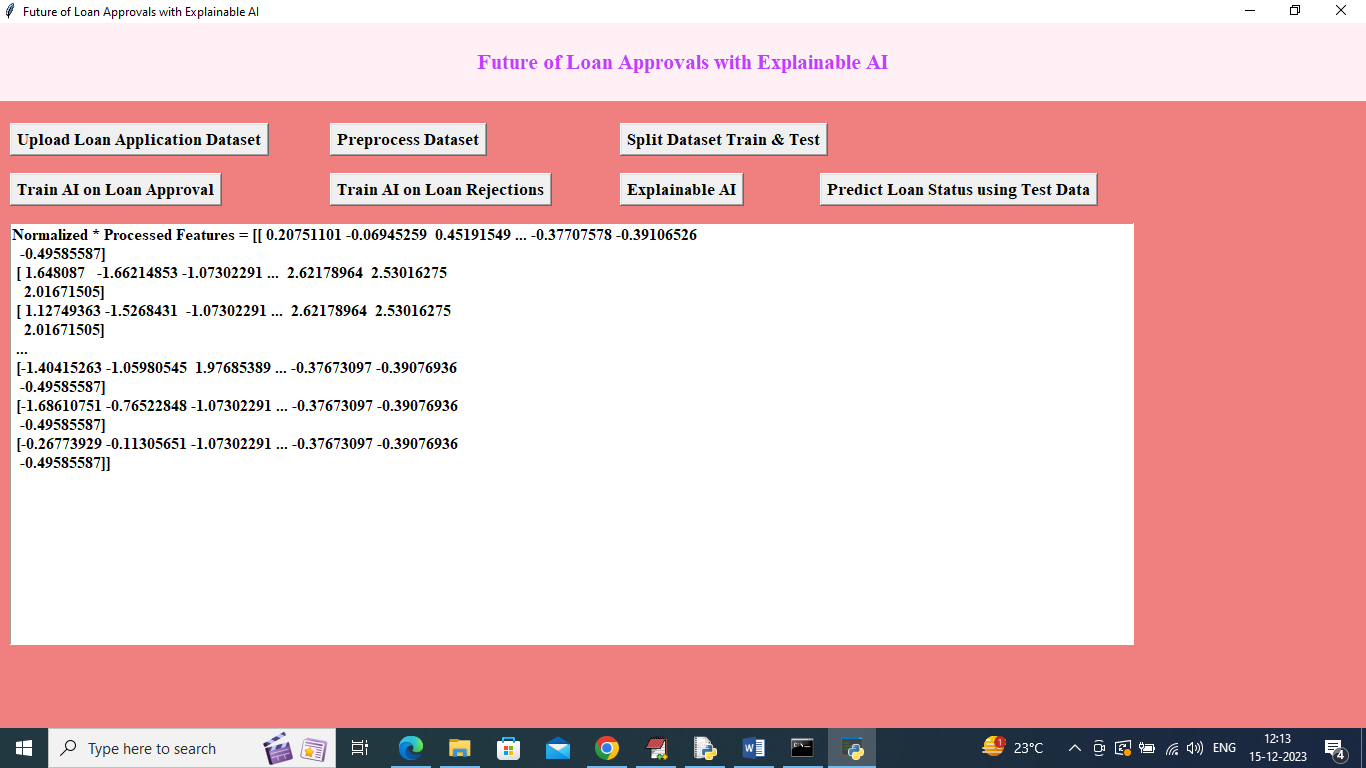
In above screen click on ‘Upload Loan Application Dataset’ button to upload dataset and then will get below output



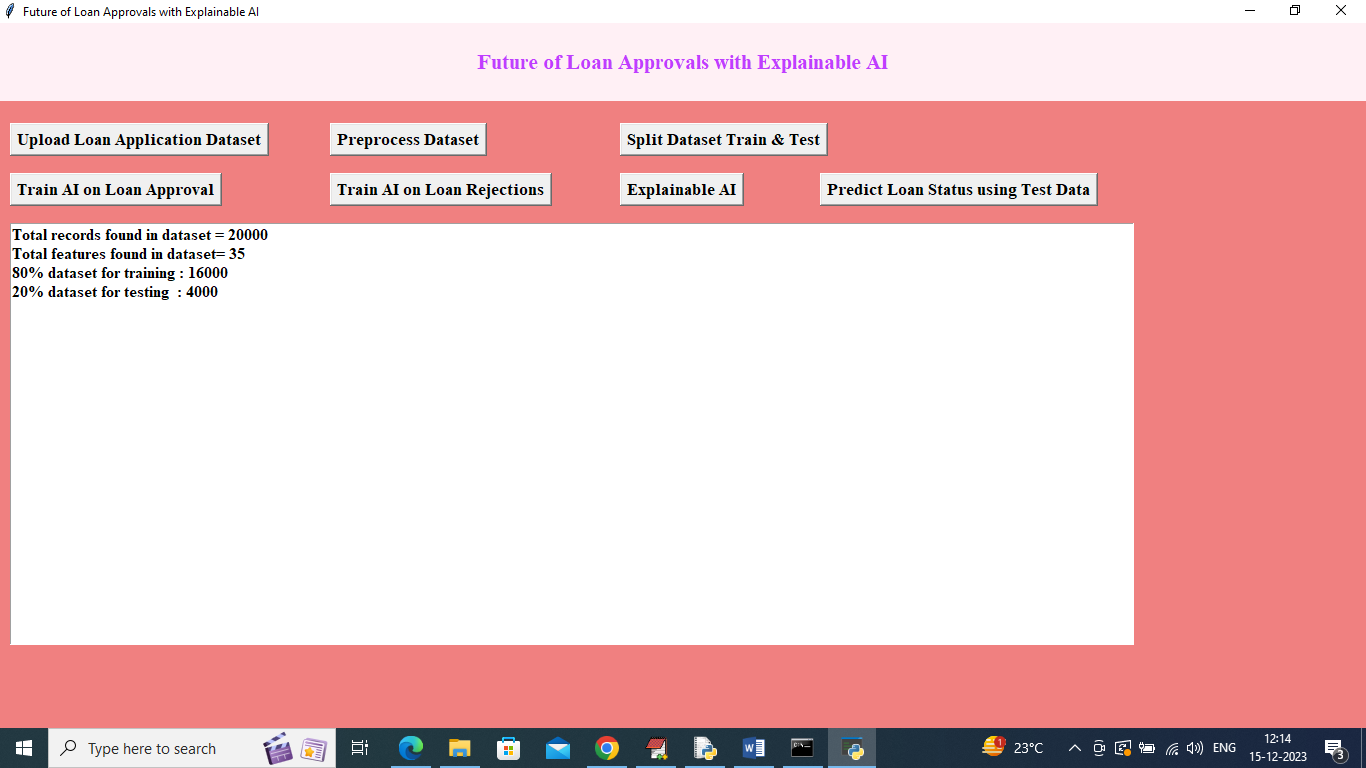
In above screen selecting and uploading ‘loan\_application.csv’ file and then click on ‘Open’ button to load dataset and get below output



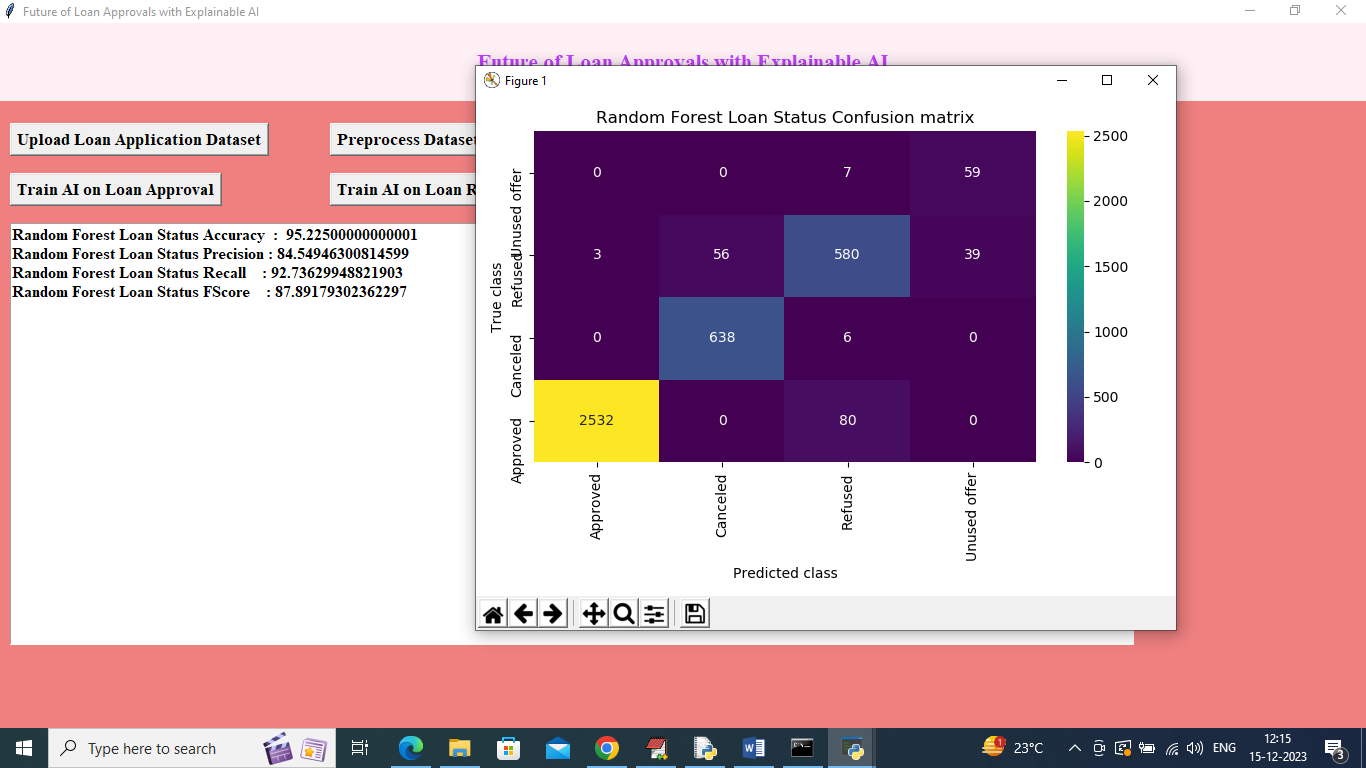
In above screen dataset loaded and in text area can see few records from dataset and in first graph x-axis represents LOAN STATUS and y-axis represents Number of Records available in that LOAN STATUS class label. In second graph x-axis represents REJECTION REASON and y-axis represents records size and in dataset we have both numeric and non-numeric values so to convert to numeric data then click on ‘Pre-process Dataset’ button to get below output



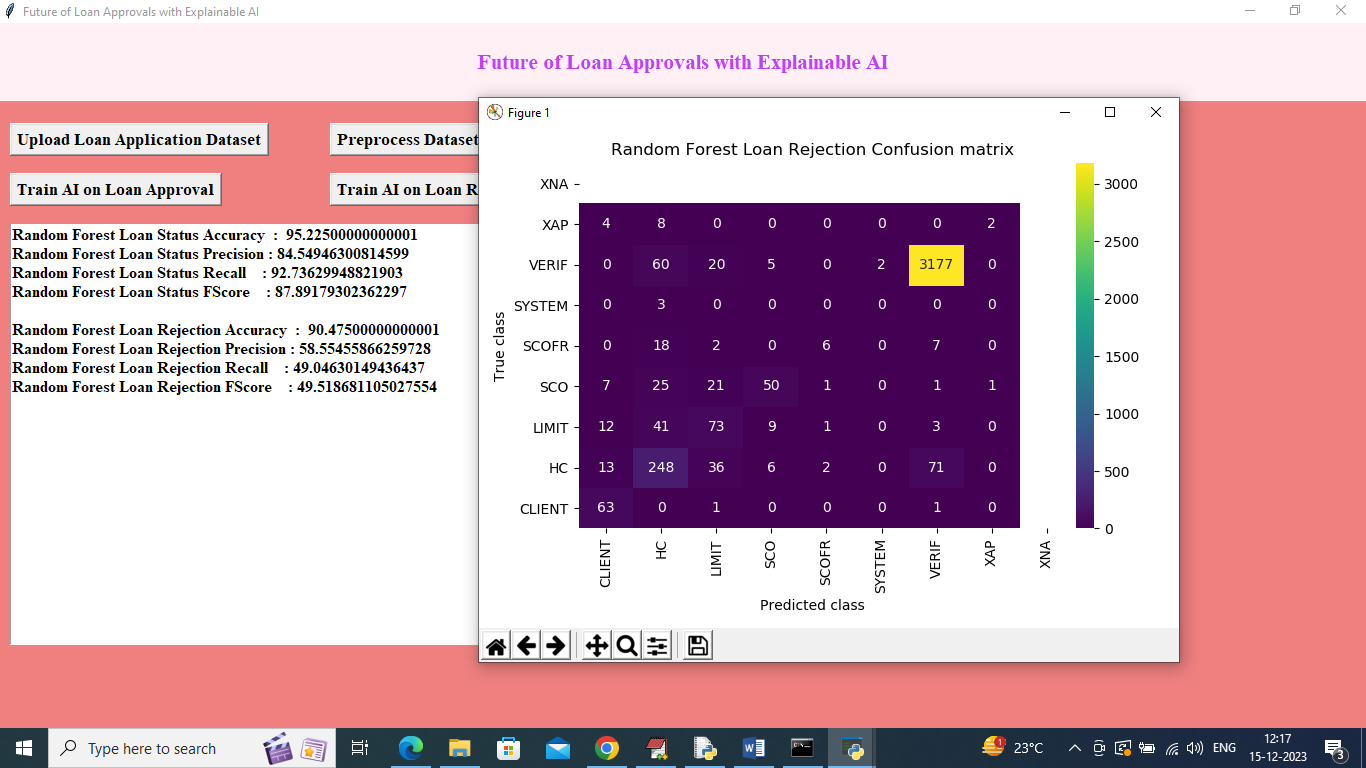
In above screen dataset converted to numeric format and then click on ‘Split Dataset Train & Test’ button to split dataset into train and test and then will get below output



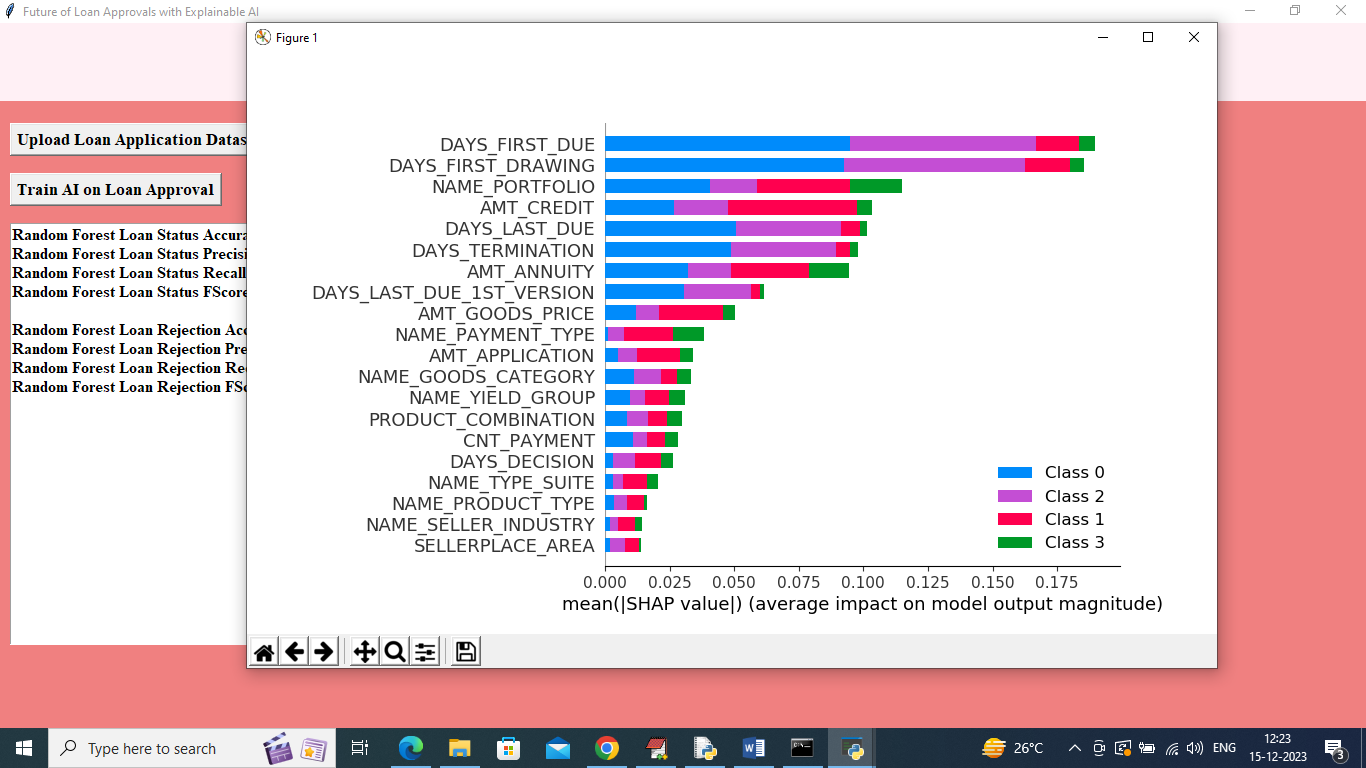
In above screen can see dataset size with total number of features and then can see TRAIN and TEST size and now click on ‘Train AO on Loan Status Approval’ button to train AI and get below output



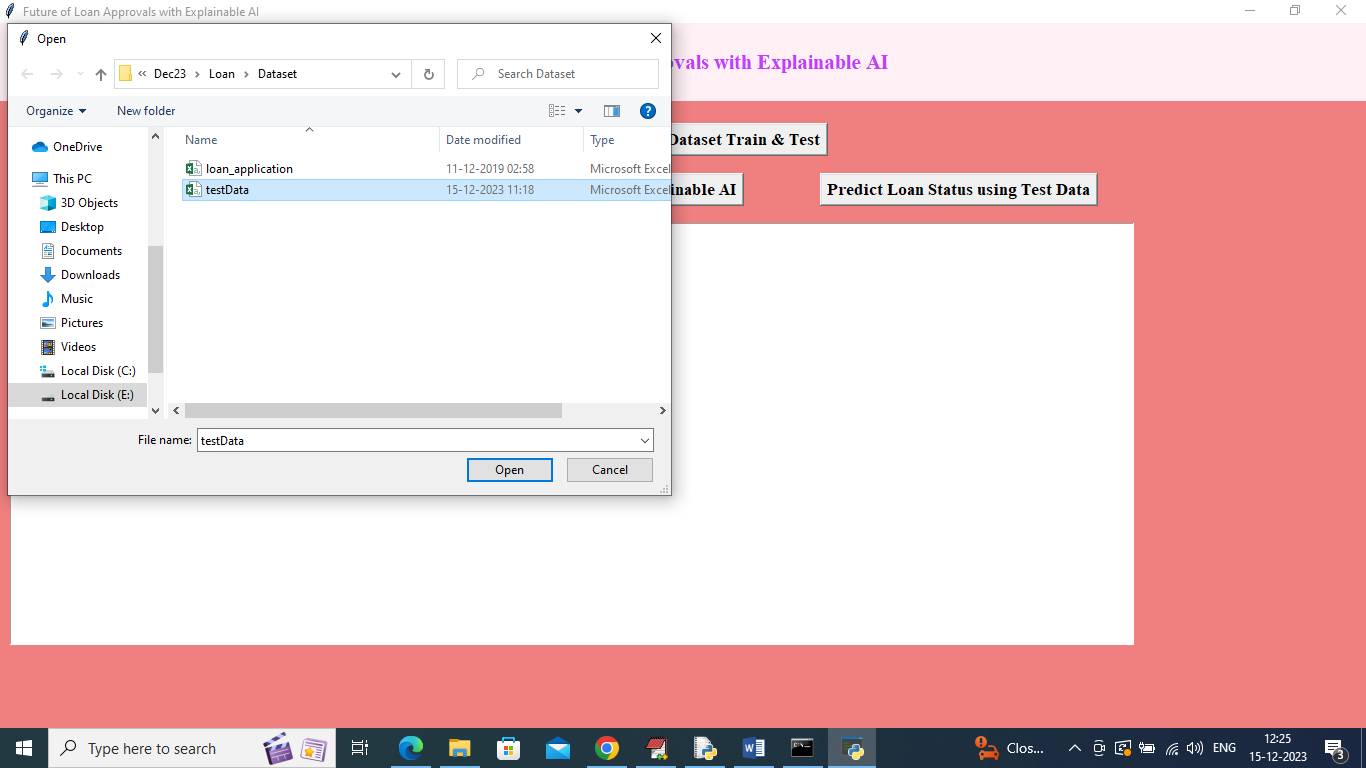
In above screen AI Random Forest got 95% accuracy on Loan STATUS and can see other metrics also. In above confusion matrix graph x-axis represents ‘LOAN STATUS Predicted Labels’ and y-axis represents TRUE labels and all boxes in diagnol contains correct prediction count and remaining blue boxes contains incorrect prediction count which are very few. Now click on ‘Train AI on Loan Rejections’ button to train AI on rejection reason and get below output



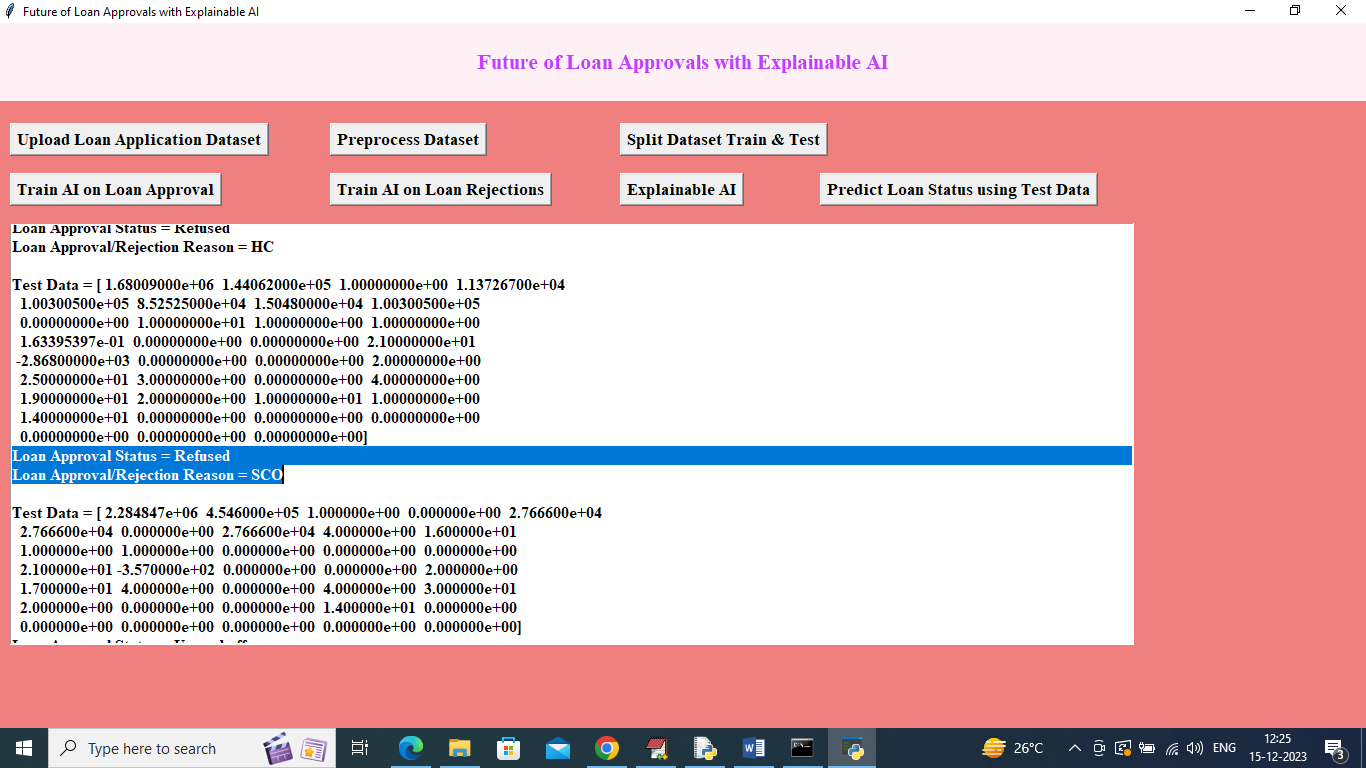
In above screen AI on REJECTION got 90% accuracy and in confusion matrix graph x-axis represents ‘Rejection Reason Predicted Labels’ and y-axis represents True label and in diagnol boxes we can see correct prediction count and remaining boxes contains incorrect prediction count. Now click on ‘Explainable AI’ button to get below features explanation on prediction



In above SHAP explanation screen in each bar we can see 4 different colours and each colour represents one class label and based on colour percentage we can say which feature names is contributing how much to predict that class label. Now close above graph and then click on ‘Predict Loan Status using Test Data’ button to upload test data and then will get below prediction



In above screen selecting and uploading testData.csv file and then click on ‘Open’ button to get below output

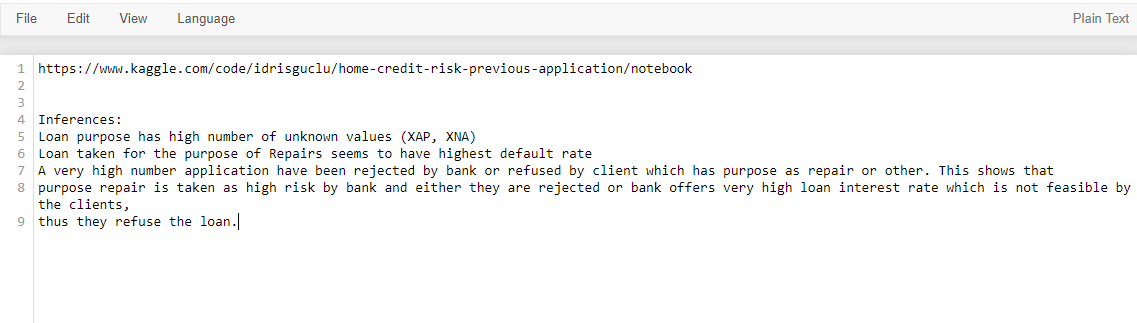


In above screen in square bracket we can see test data and then in blue colour selected line next to TEST data we can see LOAN STATUS prediction and REASON details. Scroll down above output to view all predictions



In above screen can see other prediction output.

For Reason Rejected code you can read below description



In above screen read blue colour selected text to know about REJECTED REASON codes