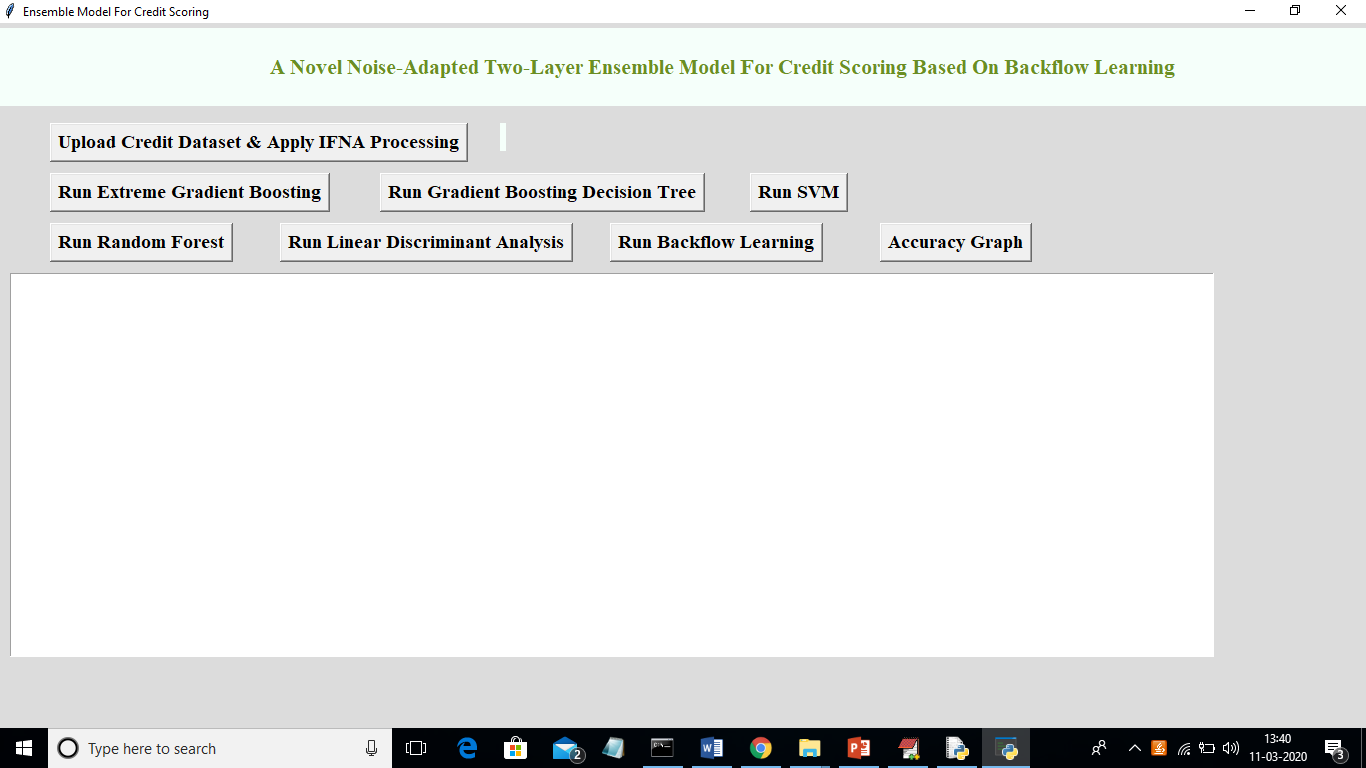
A novel noise-adapted two-layer ensemble model for credit scoring based on backflow learning

In this project using credit scoring dataset we are evaluating performance of various machine learning algorithms as base classifiers and later this base classifiers will be pass to backflow ensemble algorithm by using Voting classifier. Voting classifier will choose best performing base classifier. Here credit scoring dataset may contains outlier/class imbalance and to remove outlier we are applying SMOTE algorithm.

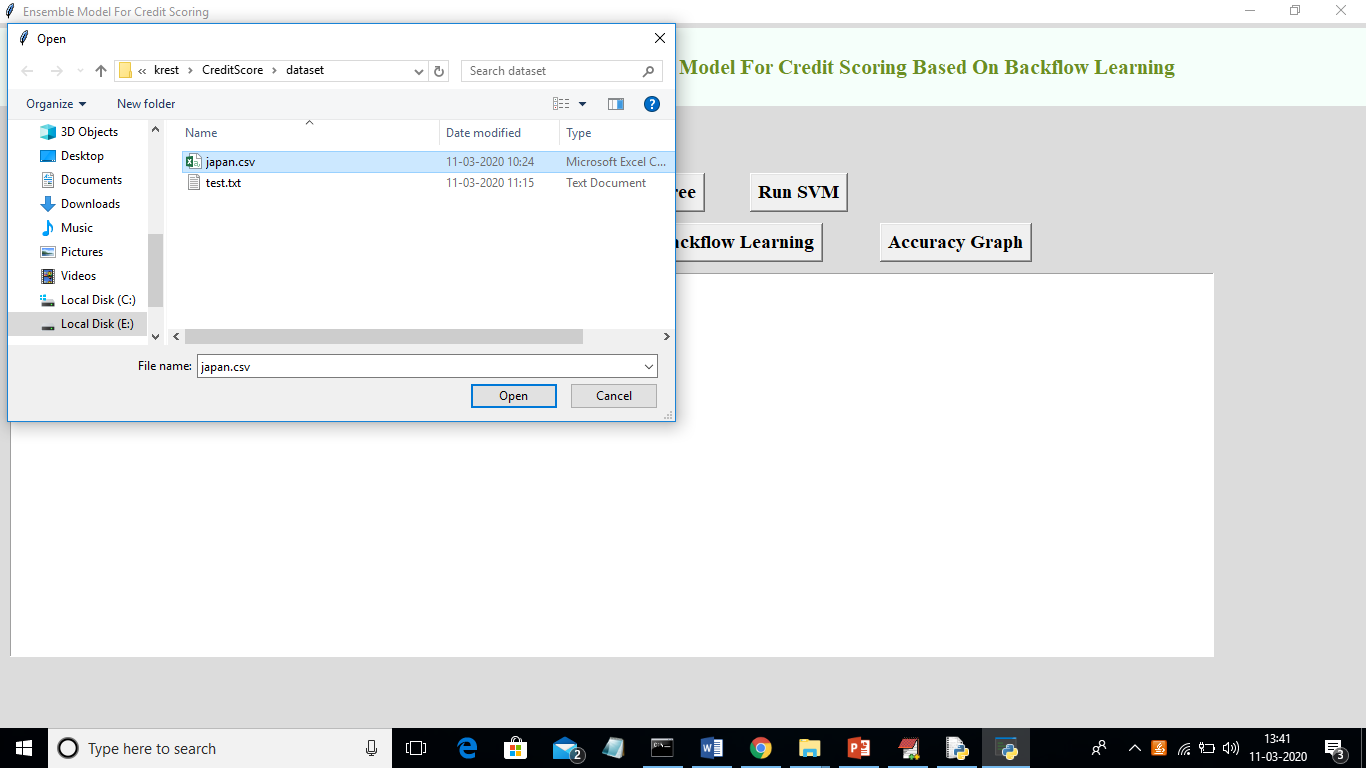
As base classifiers we are using SVM, Random Forest, Extreme Gradient Boosting, Gradient Boosting Decision Tree and LDA (linear Deterministic Analysis). As ensemble model for backflow we are using Voting classifier.

Screenshots

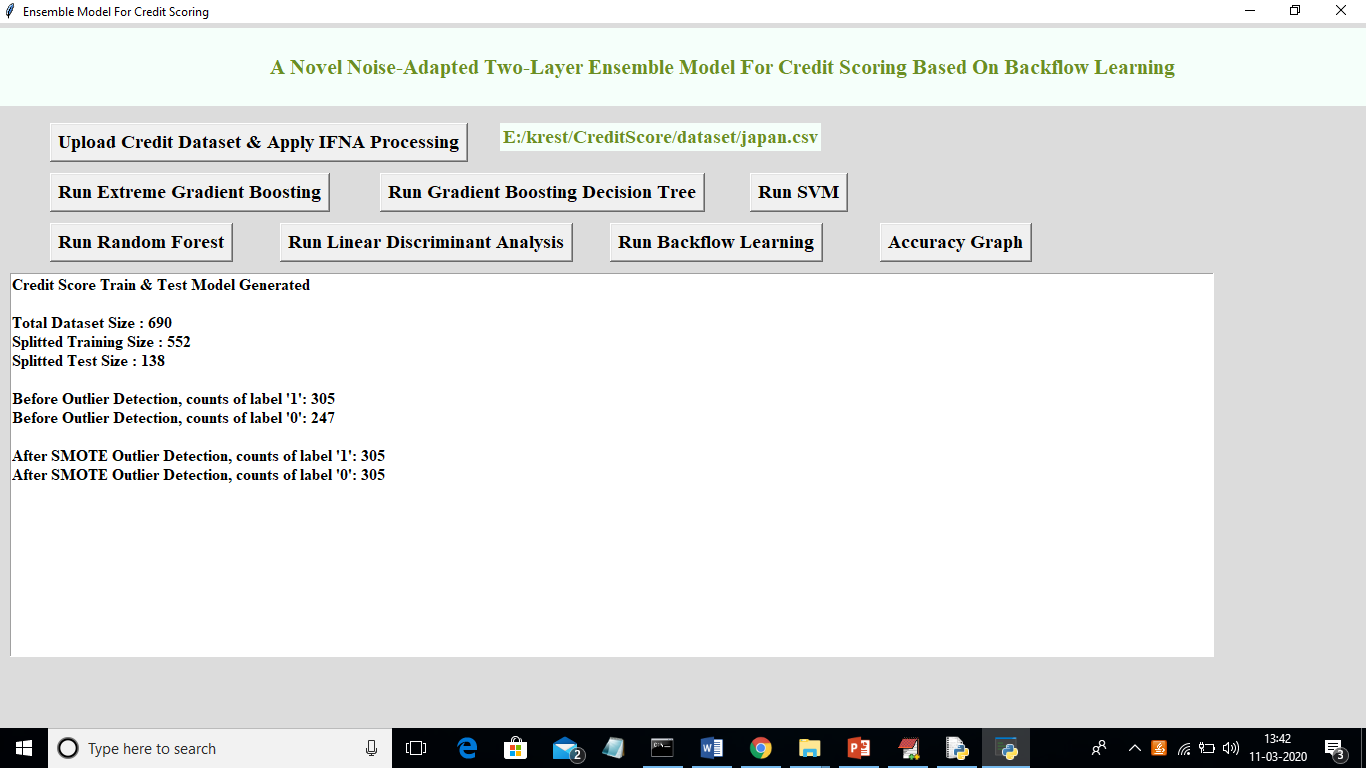
To run this project double click on ‘run.bat’ file to get below screen.



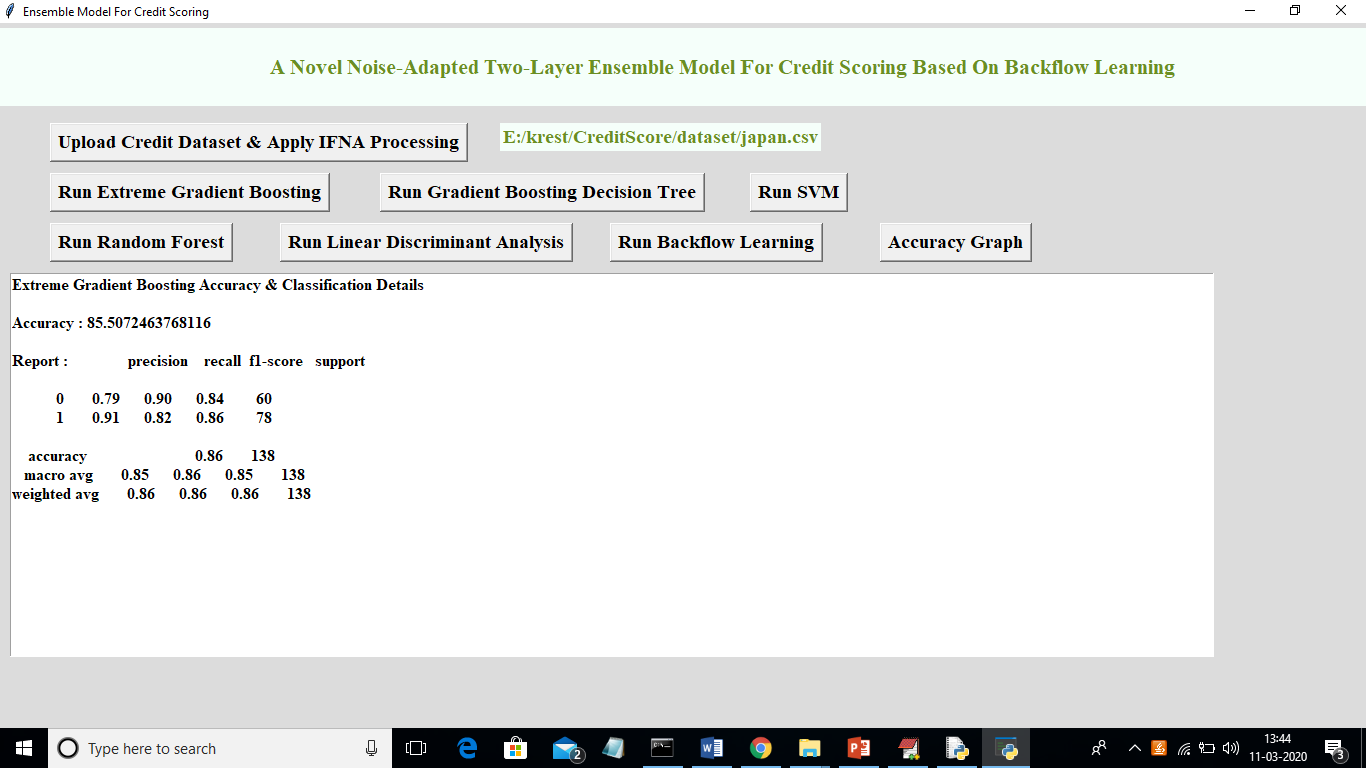
In above screen click on ‘Upload Credit Dataset & Apply IFNA Processing’ button and upload dataset and apply SMOTE algorithm to remove outliers and to solve class imbalance problem



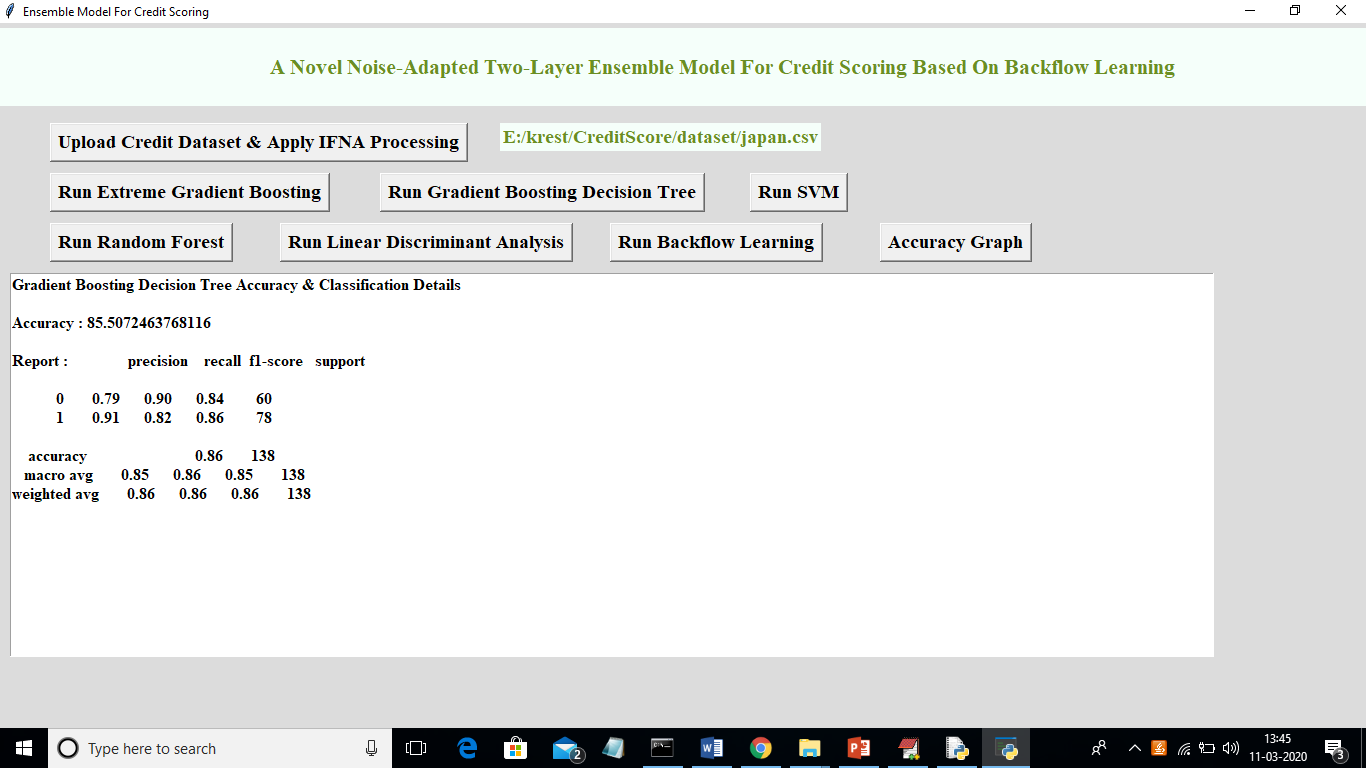
In above screen I am uploading ‘japan.csv’ dataset and after uploading will get below screen



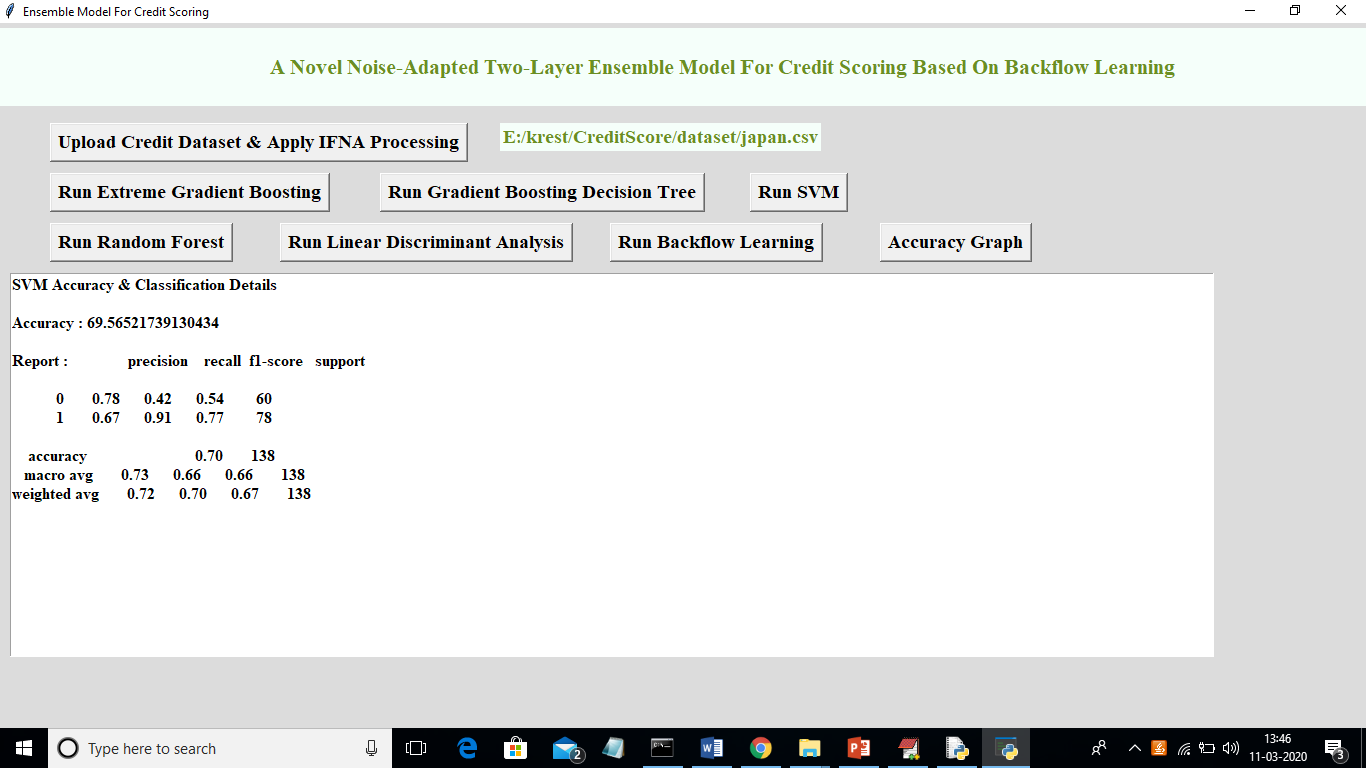
In above screen we can see dataset contains 690 records and application using 552 (80% dataset records) for training model and 138 for testing. Before applying smote we can see there is class imbalance as class 1 contains 305 records and class 0 contains 247 records. After applying SMOTE we got equal number of classes. Now click on ‘Run Extreme Gradient Boosting’ button to build boosting model and to calculate accuracy on test data



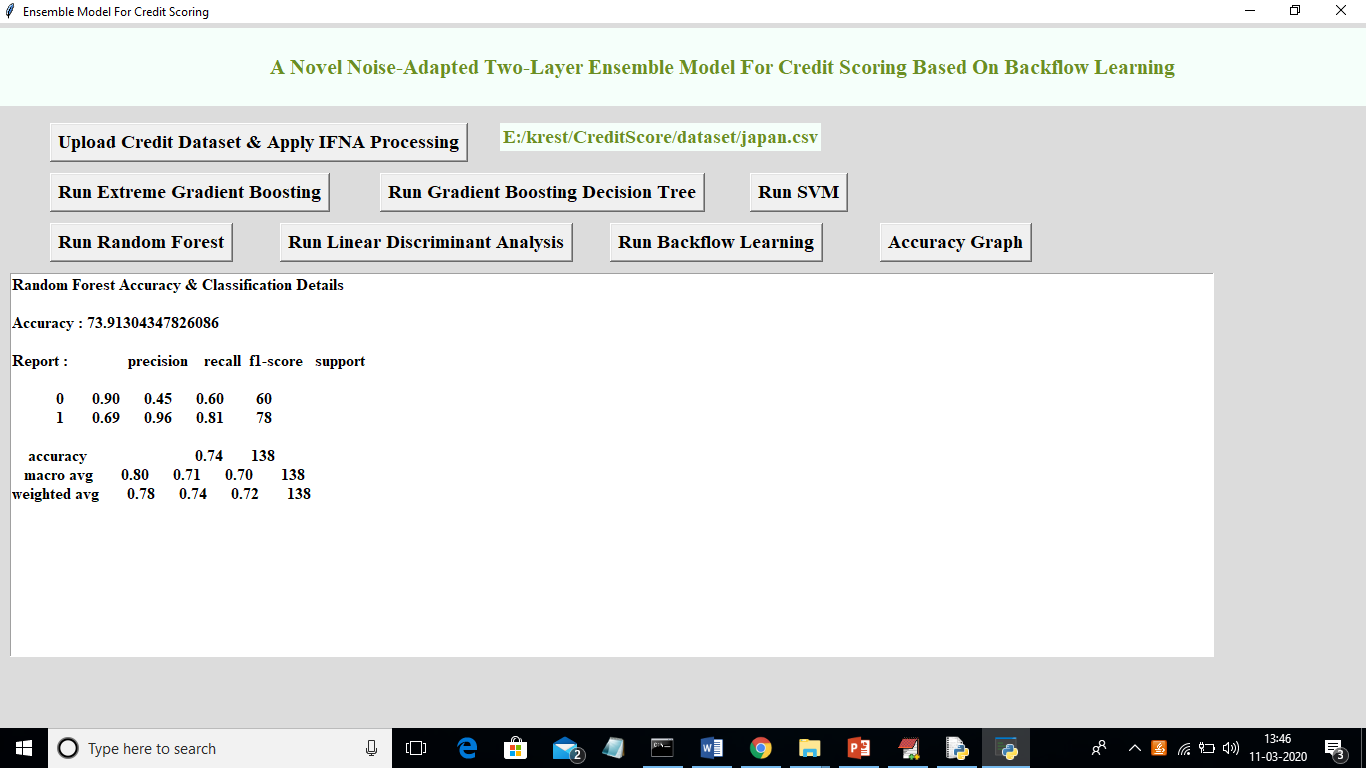
In above screen we can see Extreme Gradient boosting got 85% accuracy and we can see precision and recall values also. Now click on ‘Run Gradient Boosting Decision Tree’ button to calculate its accuracy



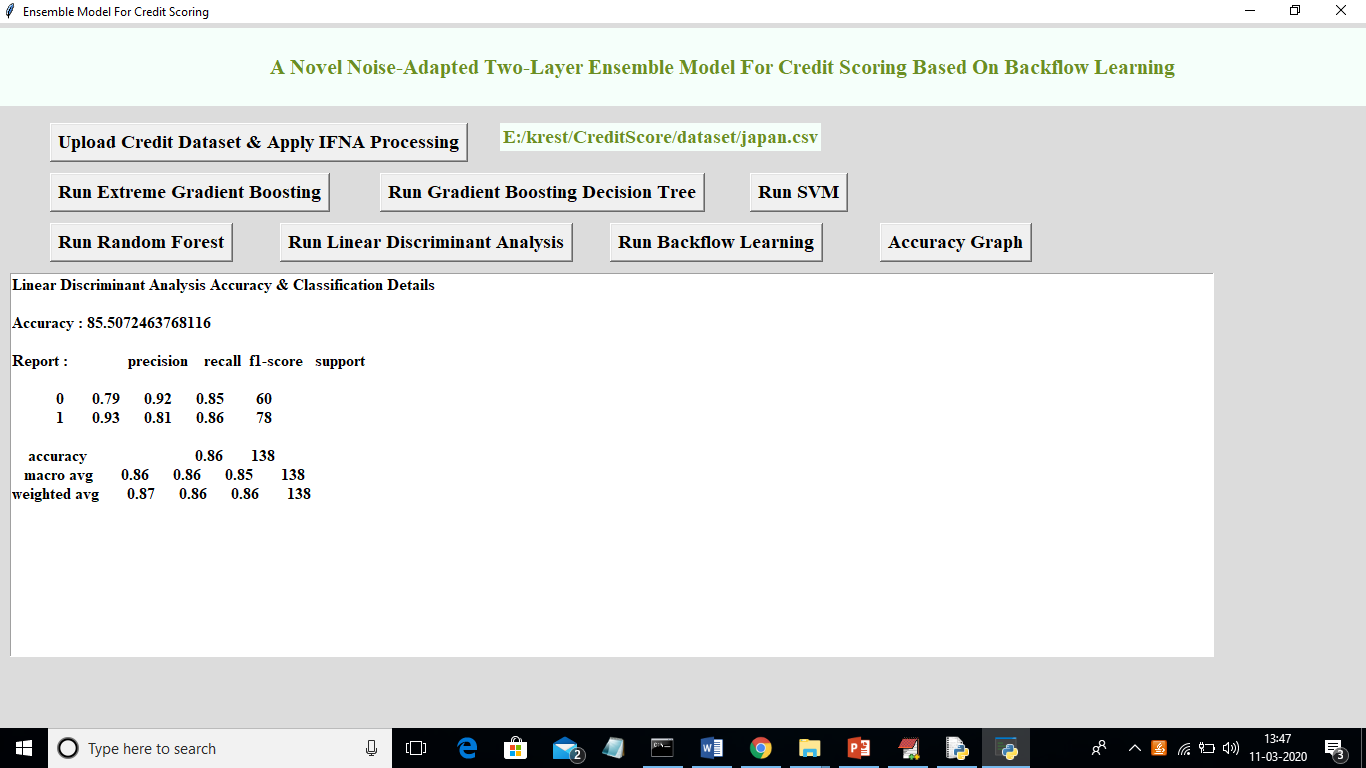
In above screen decision tree also got 85% accuracy and now click on ‘Run SVM’ to get its accuracy



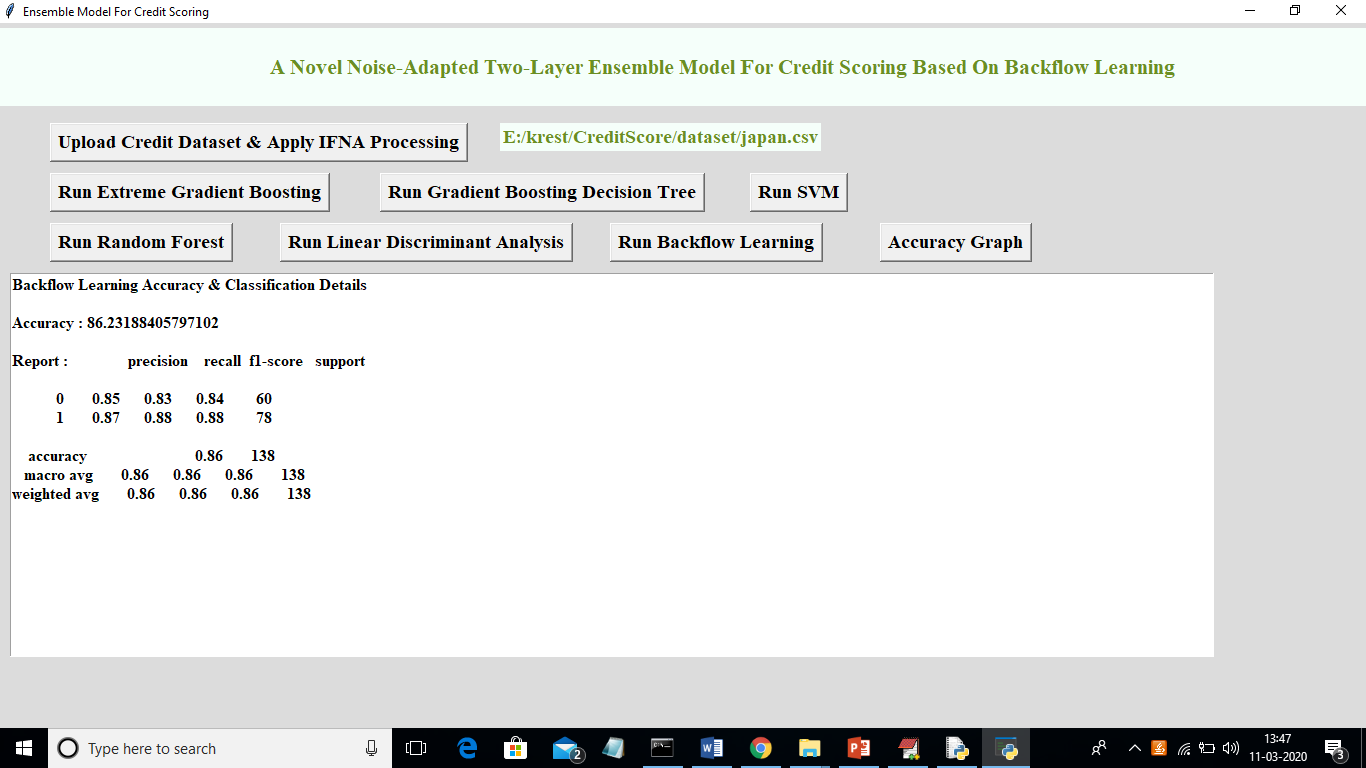
In above screen SVM got 69% accuracy and now click on ‘Run Random Forest’ button to get its accuracy



With random forest we got 79% accuracy and now click on ‘Run Linear Discriminant Analysis’ button to get its accuracy



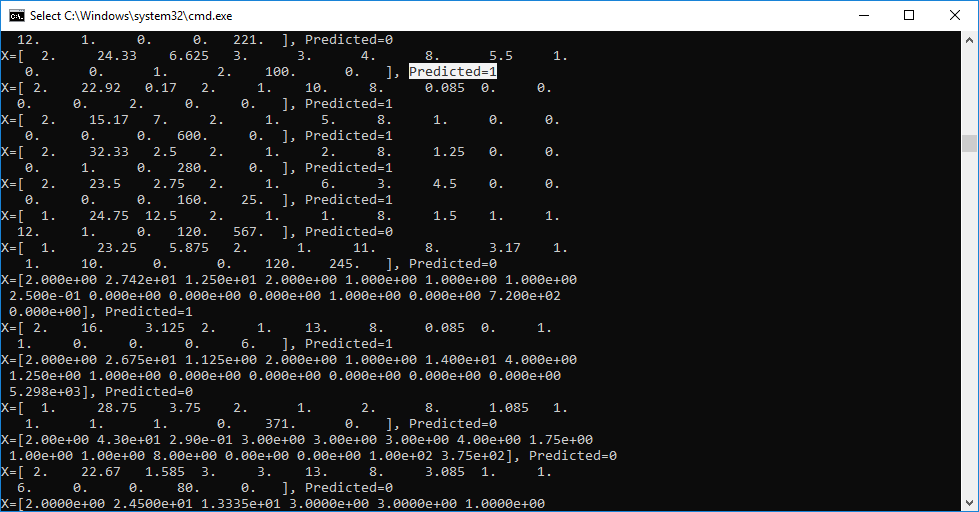
In above screen LDA got 85.50% accuracy and now click on ‘Run Backflow Learning’ button to get backflow accuracy



In above screen Backflow got 86.23% accuracy which is more than all other bas classifiers. Now click on ‘Accuracy Graph’ button to get below graph



In above graph x-axis represents algorithm names and y-axis represents accuracy values. From all algorithms backflow ensemble algorithm got better accuracy. In below black console u can see predicted values for each classifier



In above screen for each test record we can see the predicted class