Decision tree:

Dataset2:

Accuracy: 0.6079093432007401

Precision: 0.44196910545594764

Recall: 0.4623846913320597

F-1 Measure: 0.4450678981785526

Random Forests:

Code Implementation:

In this algorithm we generate T number of trees by taking N length dataset (different every time) and choose m variables which are to be considered while calculating the best split at every node in a tree. We calculate the predictions from all the trees and append them into a list and find the majority of the classifications. The class value with more instances will be assigned to the row.

Best split: we traverse through all the m random columns selected and considering all the unique values present in that certain, we split the node into its subtrees we generate the gain

and return the column number and the value which constitute to the low gini index

decision tree\_new : same steps as in decision\_tree

predicted\_new: this function returns the majority classes from the predictions trees in random forest

dataset1:

Accuracy: 0.9964912280701753

Precision: 0.9967741935483871

Recall: 0.9954545454545455

F-1 Measure: 0.996035074342356

dataset 2:

Accuracy: 0.9833333333333333

Precision: 0.9285714285714286

Recall: 1.0

F-1 Measure: 0.9629629629629629

Boosting:

Initially, we set up uniform weights on all the records . At each round we create a bootstrap sample based on the weights and train a decision tree classifier on the sample and apply it on the original training set and calculate the error using the existing and the predicted values. If the error rate is higher than 50%, start over. If the error is less than 0.5 we calculate the alpha and update the weights by checking whether the predictions are correct or not. The records that are mis-classified will have their weights increased whereas the records that are classified correctly will have their weights decreased. Final prediction is weighted average of all the classifiers with weight representing the training accuracy

Dataset1:

Accuracy: 95.77694235588972

Precision: 94.86029795203956

Recall: 93.54188311688311

F-1 Measure: 94.1115068882824

Dataset2:

Accuracy: 64.27382053654026

Precision: 49.59541262830737

Recall: 47.81043956043956

F-1 Measure: 48.05059105870464