ASSIGNMENT-2

Obtained Accuracies for BNA dataset in
- Single – DT : 94.16
- Bagged – DT : 96.89

- The accuracy of single – DT varies Bagged one by 2.8 percent(approx)

Obtained Accuracy of SDD data set in
- Single : 86.32
- Bagged : 84
- In this case we observe that bagging has reduce the accuracy metric but this can be increased because this is executed in a random manner.

The metrics can be varied because the execution is done in a random manner.

Structure of single – DT for BNA data set and also for SDD data set
- Maximum tree depth is 5
- Minimum number of data points for splitting at each node is 10

- The class of each terminal node is determined by calculating the dominent number of class members in the split with maximum information gain.

- Splitting feature, By considering each and every data vector of data set and splitting the data set and calculating the information gain of every such split and considering the splitting point at which we observe maximum information gain.

- Splitting value for the feature at each node is the value of attribute at which there is highest information gain.

- The number of data sets at each terminal node if ther are less than 10 node points at each split then the splitting is stopped this.