## **Assignment5 Name : Sasidhar Evuru ID: sxe140630**

**Part – II Question 4 – Programming Bagging part**

The code is an extension of the decision tree construction from the assignment1. The code is extended to support bagging.

The Bagging Procedure:

1. Initially the training records (N) are read from the text file and the corresponding objects are stores created for every record and these records are stored in the arraylist.
2. The number of bags (K) to be formed are read from the console input.
3. For every bag with the help of random number generator (generating the random number 0 to N-1) we generate N numbers and select that particular record number from the arraylist and form an arraylist with new set of records. (The new arraylist will also contain N records but some records may be duplicates as we are picking the records with replacement).
4. After generating K sets of Training records lists. We train K instances of decision tree with one bag each and now we will have K trees that are trained on the separate K bags.
5. Now we will compute the class value from each tree for each and every test record. The majority output from the trees is taken as the output class form the classifier. As we know the expected the class value we can calculate the accuracy percentage.

The Bagging procedure improves the accuracy of the unstable learners. Since Decision Tree Classifier is and unstable classifier it shall increase the performance when bagging is done.

Test cases were performed on the K trees that are generated from the train data (train-1.dat from assignemnt1). The K classifiers formed are tested on the test data (test-1.dat) and the accuracy is noted for different sizes of K.

|  |  |  |
| --- | --- | --- |
| **Train accuracy** | **Test Accuracy** | **Bagging accuracy** |
| 100 | 87.96 | 95.833(k=5) |
|  |  | 99.53(K=10) |
|  |  | 96.29(k=15) |
|  |  | 98.611(K=20) |
|  |  | 98.14(k=50) |
|  |  | 98.611(k=100) |