Rahul Aravind Mehalingam **rxm151730**

MACHINE LEARNING

6375.002

Assignment -1

Problem Statement:

Implementation of Decision tree using ID3 algorithm with two measures namely

- Entropy
- Variance Impurity

Implementation:

Following classes were used for implementation

Driver.java – Driver Program

TreeNode.java - It contains the signature of the tree and its methods

DecisionTreeBuilder.java – It contains the methods for constructing the decision tree.

InformationGain.java – It contains the methods for computing the information gain and entropy.

FileReader.java – Contains utilities for reading the contents of a file and getting data

DecisionTreePruner.java – contains methods which implements the pruning algorithm

Steps:

1) The decision tree is constructed using training data.

2) Once the tree is built, the accuracy of the tree is found using the test data.

3) Then the tree is pruned using validation data to make the accuracy of the tree better.

The pruning algorithm is implemented and the efficiency of the pruning depends on the two variable values ‘L’ and ‘K’.

Results:

Input: Data set 1

Information Gain Heuristic

Non-leaf nodes: 137

Accuracy: 75.95%

Variance Impurity Heuristic:

No leaf nodes: 133

Accuracy: 76.5%

Input: Data set 2

Information Gain Heuristic

Non-leaf nodes: 141

Accuracy: 72.83%

Variance Impurity Heuristic:

No leaf nodes: 141

Accuracy: 72.33%

Pruning:

Accuracy after pruning in entropy measure tree

Accuracy after pruning in variance impurity tree

Data Set	L - Value	K - Value	Accuracy after pruning in Information gain tree	Accuracy after pruning in Variance Impurity tree
Data Set 1	100	175	75.94	76.5
Data Set 1	25	40		
Data Set 2	100	200	72.83	73
Data Set 2	200	20	72.33	72
Data Set 2	500	20	72.67	72.5
Data Set 2	125	12	72.9	71.67
Data Set 2	177	8	73	73.16
Data Set 1	177	8	75	76.8
Data Set 1	375	15	76.1	76.55
Data Set 1	198	8	76	76.55

Result:

Thus the decision tree is constructed using different measures such as variance impurity and entropy and its accuracy has been increased by pruning the tree.