MACHINE LEARNING

6375.002

Assignment -1

Problem Statement:

Implementation of Decision tree using ID3 algorithm with two measures namely

- Entropy
- Variance Impurity

Implementation:

The implementation is done using two classes.

MachineLearning class – Driver Program

DecisionTree class - It contains the actual functions for decision tree construction and performance evaluation

The decision tree is constructed using training data. Once the tree is built, the accuracy of the tree is found using the test data. 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’.

The result of the implementation is follows

Input: Data set 1

Entropy Measure:

Non-leaf nodes: 137

Accuracy: 75.85%

Variance Impurity Measure:

No leaf nodes: 138

Accuracy: 76.65%

Pruning:

K value	L value	Accuracy after pruning in entropy measure tree	Accuracy after pruning in variance impurity tree
175	100	75.85%	76.65%
15	10	76.15%	76.85%
25	40	75.85%	77.1%
20	30	76.95%	76.95%
10	24	75.85%	76.8%
46	52	76.25%	76.65%
80	64	76.8%	76.65%
50	50	75.85%	76.9%
82	28	76.05%	76.65%
37	62	76.0%	76.85%

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.

Readme:

The code is implemented in a way that has been said in the question.

Implemented in java. Pass the specified arguments via command line.

The result gets printed in the console.

Sample output file is attached with the assignment zip.