CS 6375
ASSIGNMENT _____2_____

Names of students in your group:

NIKHIL KUMAR CHANDRA (nxc161330)
SHRUTI CHANDRASHEKHAR HARIHAR (sxh164530)

Number of free late days used: _____0_____

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

Tom Mitchell textbook
www.stackoverflow.com

REPORT

Assumptions:

1) If a noisy data is encountered then the class classification int the tree has no value i.e it mean a don't care value

2) we have only considered the internal nodes(excluding leaf nodes) for selecting the total number of prune nodes

3) Each Leaf node is the decision node i.e which stores the value of the class if the tree condition in that path followed.

Screenshots:

```
Enter the path of training data set
C:\Users\nxc161330\Downloads\ID3Implementation\data_sets1\training_set.csv
Enter the path for test data set
C:\Users\nxc161330\Downloads\ID3Implementation\data_sets1\validation_set.csv
Enter the path for validation data set
C:\Users\nxc161330\Downloads\ID3Implementation\data_sets1\test_set.csv
Enter the pruining factor for the decision tree
0.2

 XO = 0 :
|  XM = 0 :
| |  XF = 0 :
| | |  XB = 0 :
| | |  XG = 0 : 0
| | |  XG = 1 :
| | | |  XD = 0 :
| | | | |  XS = 0 : 0
| | | | |  XS = 1 :
| | | | | |  XC = 0 : 1
| | | | | |  XC = 1 :
| | | | | | |  XH = 0 : 0
| | | | | | |  XH = 1 : 1
| | | | |  XD = 1 :
| | | | |  XE = 0 : 0
| | | | |  XE = 1 :
| | | | | |  XK = 0 : 0
| | | | | |  XK = 1 : 1
| | |  XB = 1 :
| | |  XD = 0 : 0
| | |  XD = 1 :
| | | |  XI = 0 : 0
| | | |  XI = 1 :
| | | | |  XG = 0 : 1
| | | | |  XG = 1 : 0
| |  XF = 1 : 0
|  XM = 1 :
| |  XB = 0 :
| | |  XD = 0 :
| | |  XG = 0 :
| | | |  XF = 0 : 0
| | | |  XF = 1 :
| | | | |  XJ = 0 :
| | | | | |  XN = 0 : 1
| | | | | |  XN = 1 :
| | | | | | |  XE = 0 :
| | | | | | |  XK = 0 : 0
| | | | | | |  XK = 1 : 1
| | | | | | |  XE = 1 : 0
| | | | | |  XJ = 1 :
| | | | | |  XC = 0 :
| | | | | | |  XT = 0 :
| | | | | | | |  XL = 0 :
| | | | | | | | |  XI = 0 :
| | | | | | | | | |  XN = 0 : 1
| | | | | | | | | |  XN = 1 : 0
```

```
| | | | | | | | | |  XI = 1 : 1
| | | | | | | | |  XL = 1 : 0
| | | | | | | | |  XT = 1 : 1
| | | | | | | |  XC = 1 : 1
| | | |  XG = 1 :
| | | | |  XU = 0 : 1
| | | | |  XU = 1 :
| | | | | |  XI = 0 : 0
| | | | | |  XI = 1 : 1
| | |  XD = 1 :
| | | |  XC = 0 :
| | | | |  XF = 0 :
| | | | | |  XG = 0 : 0
| | | | | |  XG = 1 :
| | | | | | |  XP = 0 :
| | | | | | | |  XS = 0 : 0
| | | | | | | |  XS = 1 : 1
| | | | | | |  XP = 1 : 0
| | | | |  XF = 1 :
| | | | | |  XJ = 0 : 1
| | | | | |  XJ = 1 :
| | | | | | |  XE = 0 :
| | | | | | | |  XG = 0 :
| | | | | | | | |  XI = 0 : 1
| | | | | | | | |  XI = 1 : 0
| | | | | | | |  XG = 1 : 0
| | | | | | |  XE = 1 :
| | | | | | | |  XT = 0 :
| | | | | | | | |  XG = 0 : 1
| | | | | | | | |  XG = 1 : 0
| | | | | | | |  XT = 1 : 1
| | | |  XC = 1 : 0
| |  XB = 1 :
| | |  XI = 0 : 0
| | |  XI = 1 :
| | | |  XC = 0 :
| | | | |  XK = 0 :
| | | | | |  XP = 0 : 1
| | | | | |  XP = 1 :
| | | | | | |  XS = 0 :
| | | | | | | |  XG = 0 : 1
| | | | | | | |  XG = 1 :
| | | | | | | | |  XF = 0 : 0
| | | | | | | | |  XF = 1 : 1
| | | | | | | |  XS = 1 : 0
| | | | |  XK = 1 : 0
| | | |  XC = 1 : 0
 XO = 1 :
|  XI = 0 :
| |  XM = 0 :
| | |  XQ = 0 :
| | | |  XF = 0 :
| | | | |  XH = 0 :
| | | | | |  XB = 0 : 0
| | | | | |  XB = 1 :
```

```
| | | | | | | |  XC = 0 : 1
| | | | | | | |  XC = 1 : 0
| | | | |  XH = 1 : 1
| | | |  XF = 1 : 0
| | |  XQ = 1 :
| | | |  XJ = 0 :
| | | | |  XN = 0 :
| | | | | |  XP = 0 : 1
| | | | | |  XP = 1 :
| | | | | | |  XB = 0 :
| | | | | | | |  XF = 0 : 0
| | | | | | | |  XF = 1 : 1
| | | | | | |  XB = 1 : 0
| | | | |  XN = 1 : 0
| | | |  XJ = 1 :
| | | | |  XL = 0 :
| | | | | |  XH = 0 : 0
| | | | | |  XH = 1 :
| | | | | | |  XK = 0 :
| | | | | | | |  XU = 0 : 1
| | | | | | | |  XU = 1 : 0
| | | | | | |  XK = 1 : 1
| | | | |  XL = 1 : 1
| |  XM = 1 :
| | |  XQ = 0 :
| | | |  XF = 0 :
| | | | |  XL = 0 :
| | | | | |  XC = 0 : 1
| | | | | |  XC = 1 :
| | | | | | |  XH = 0 : 1
| | | | | | |  XH = 1 :
| | | | | | | |  XU = 0 :
| | | | | | | | |  XB = 0 :
| | | | | | | | | |  XJ = 0 : 1
| | | | | | | | | |  XJ = 1 : 0
| | | | | | | | |  XB = 1 : 0
| | | | | | | |  XU = 1 : 1
| | | | |  XL = 1 :
| | | | | |  XC = 0 :
| | | | | | |  XB = 0 : 0
| | | | | | |  XB = 1 :
| | | | | | | |  XP = 0 : 0
| | | | | | | |  XP = 1 : 1
| | | | | | |  XC = 1 : 1
| | | |  XF = 1 : 0
| | |  XQ = 1 : 0
|  XI = 1 :
| |  XT = 0 :
| | |  XH = 0 :
| | | |  XP = 0 :
| | | | |  XF = 0 : 0
| | | | |  XF = 1 :
| | | | | |  XQ = 0 :
| | | | | | |  XK = 0 : 1
| | | | | | |  XK = 1 :
```

```
| | | | | | | | |  XC = 0 : 0
| | | | | | | | |  XC = 1 : 1
| | | | | | |  XQ = 1 :
| | | | | | | |  XK = 0 : 0
| | | | | | | |  XK = 1 : 1
| | | |  XP = 1 :
| | | | |  XS = 0 :
| | | | | |  XD = 0 :
| | | | | | |  XC = 0 :
| | | | | | | |  XJ = 0 :
| | | | | | | | |  XN = 0 : 0
| | | | | | | | |  XN = 1 : 1
| | | | | | | |  XJ = 1 :
| | | | | | | | |  XB = 0 : 0
| | | | | | | | |  XB = 1 :
| | | | | | | | | |  XG = 0 : 1
| | | | | | | | | |  XG = 1 : 0
| | | | | | |  XC = 1 : 0
| | | | | |  XD = 1 :
| | | | | | |  XM = 0 :
| | | | | | | |  XC = 0 : 1
| | | | | | | |  XC = 1 : 0
| | | | | | |  XM = 1 : 1
| | | | |  XS = 1 : 1
| | |  XH = 1 :
| | | |  XJ = 0 :
| | | | |  XC = 0 :
| | | | | |  XN = 0 : 1
| | | | | |  XN = 1 :
| | | | | | |  XF = 0 :
| | | | | | | |  XG = 0 : 1
| | | | | | | |  XG = 1 : 0
| | | | | | |  XF = 1 : 0
| | | | |  XC = 1 :
| | | | | |  XM = 0 : 0
| | | | | |  XM = 1 :
| | | | | | |  XF = 0 :
| | | | | | | |  XR = 0 : 1
| | | | | | | |  XR = 1 : 0
| | | | | | |  XF = 1 : 1
| | | |  XJ = 1 :
| | | | |  XS = 0 : 1
| | | | |  XS = 1 :
| | | | | |  XG = 0 :
| | | | | | |  XB = 0 : 0
| | | | | | |  XB = 1 :
| | | | | | | |  XD = 0 : 1
| | | | | | | |  XD = 1 :
| | | | | | | | |  XE = 0 : 0
| | | | | | | | |  XE = 1 : 1
| | | | | |  XG = 1 :
| | | | | | |  XC = 0 : 1
| | | | | | |  XC = 1 :
| | | | | | | |  XD = 0 : 1
| | | | | | | |  XD = 1 : 0
```

```
| | XT = 1 :
| | | XS = 0 :
| | | | XQ = 0 :
| | | | | XK = 0 :
| | | | | | XC = 0 :
| | | | | | | XR = 0 :
| | | | | | | | XH = 0 :
| | | | | | | | | XE = 0 : 0
| | | | | | | | | XE = 1 : 1
| | | | | | | | XH = 1 : 1
| | | | | | | XR = 1 :
| | | | | | | | XB = 0 :
| | | | | | | | | XD = 0 : 0
| | | | | | | | | XD = 1 : 1
| | | | | | | | XB = 1 : 0
| | | | | | XC = 1 : 1
| | | | | XK = 1 :
| | | | | | XD = 0 :
| | | | | | | XF = 0 : 0
| | | | | | | XF = 1 : 1
| | | | | | XD = 1 : 0
| | | | XQ = 1 :
| | | | | XM = 0 :
| | | | | | XN = 0 :
| | | | | | | XU = 0 : 1
| | | | | | | XU = 1 : 0
| | | | | | XN = 1 :
| | | | | | | XP = 0 :
| | | | | | | | XB = 0 :
| | | | | | | | | XF = 0 : 0
| | | | | | | | | XF = 1 : 1
| | | | | | | | XB = 1 : 1
| | | | | | | XP = 1 : 1
| | | | | XM = 1 : 0
| | | XS = 1 :
| | | | XL = 0 :
| | | | | XD = 0 :
| | | | | | XU = 0 : 1
| | | | | | XU = 1 :
| | | | | | | XB = 0 :
| | | | | | | | XE = 0 : 1
| | | | | | | | XE = 1 :
| | | | | | | | | XC = 0 : 1
| | | | | | | | | XC = 1 : 0
| | | | | | | XB = 1 :
| | | | | | | | XG = 0 :
| | | | | | | | | XH = 0 : 0
| | | | | | | | | XH = 1 : 1
| | | | | | | | XG = 1 : 0
| | | | | XD = 1 :
| | | | | | XG = 0 : 0
| | | | | | XG = 1 : 1
| | | | XL = 1 :
| | | | | XH = 0 :
| | | | | | XD = 0 :
```

```
| | | | | | |  XQ = 0 : 0
| | | | | | |  XQ = 1 :
| | | | | | | |  XB = 0 : 0
| | | | | | | |  XB = 1 : 1
| | | | | |  XD = 1 :
| | | | | | |  XB = 0 : 1
| | | | | | |  XB = 1 : 0
| | | | |  XH = 1 : 0

Pre-Pruned Accuracy:
----------------------------------
Number of training instances = 600
Number of training attributes = 20
Number of non-leaf nodes in the tree = 137
Number of leaf nodes in the tree = 138
Number of nodes in the tree = 275
Accuracy of the model on the training dataset = 1.0

Number of validation instances = 2000
Number of validation attributes = 20
Accuracy of the model on the training dataset = 0.7585

Number of testing instances = 2000
Number of testing attributes = 20
Accuracy of the model on the testing dataset = 0.758


Post-Pruned Accuracy
----------------------------------
Pruned Tree:


 XO = 0 :
|  XM = 0 :
| |  XF = 0 :
| | |  XB = 0 :
| | | |  XG = 0 : 0
| | | |  XG = 1 : 1
| | |  XB = 1 :
| | | |  XD = 0 : 0
| | | |  XD = 1 :
| | | | |  XI = 0 : 0
| | | | |  XI = 1 :
| | | | | |  XG = 0 : 1
| | | | | |  XG = 1 : 0
| |  XF = 1 : 0
|  XM = 1 :
| |  XB = 0 : 1
| |  XB = 1 :
| | |  XI = 0 : 0
| | |  XI = 1 :
| | | |  XC = 0 :
| | | | |  XK = 0 :
| | | | | |  XP = 0 : 1
| | | | | |  XP = 1 :
```

```
| | | | | | | |  XS = 0 :
| | | | | | | | |  XG = 0 : 1
| | | | | | | | |  XG = 1 :
| | | | | | | | | |  XF = 0 : 0
| | | | | | | | | |  XF = 1 : 1
| | | | | | | |  XS = 1 : 0
| | | | |  XK = 1 : 0
| | | |  XC = 1 : 0
 XO = 1 :
|  XI = 0 : 0
|  XI = 1 :
| |  XT = 0 :
| | |  XH = 0 :
| | | |  XP = 0 :
| | | | |  XF = 0 : 0
| | | | |  XF = 1 :
| | | | | |  XQ = 0 :
| | | | | | |  XK = 0 : 1
| | | | | | |  XK = 1 :
| | | | | | | |  XC = 0 : 0
| | | | | | | |  XC = 1 : 1
| | | | | |  XQ = 1 :
| | | | | | |  XK = 0 : 0
| | | | | | |  XK = 1 : 1
| | | | |  XP = 1 : 1
| | | |  XH = 1 :
| | | | |  XJ = 0 :
| | | | | |  XC = 0 :
| | | | | |  XN = 0 : 1
| | | | | |  XN = 1 :
| | | | | | |  XF = 0 :
| | | | | | | |  XG = 0 : 1
| | | | | | | |  XG = 1 : 0
| | | | | | |  XF = 1 : 0
| | | | |  XC = 1 : 0
| | | |  XJ = 1 :
| | | | |  XS = 0 : 1
| | | | |  XS = 1 :
| | | | | |  XG = 0 :
| | | | | | | |  XB = 0 : 0
| | | | | | | |  XB = 1 :
| | | | | | | | |  XD = 0 : 1
| | | | | | | | |  XD = 1 :
| | | | | | | | | |  XE = 0 : 0
| | | | | | | | | |  XE = 1 : 1
| | | | | |  XG = 1 :
| | | | | | | |  XC = 0 : 1
| | | | | | | |  XC = 1 :
| | | | | | | | |  XD = 0 : 1
| | | | | | | | |  XD = 1 : 0
| |  XT = 1 : 1
Number of training instances = 600
Number of training attributes = 20
Number of non-leaf nodes in the tree = 37
Number of leaf nodes in the tree = 38
```

```
Number of leaf nodes in the tree = 38
Number of nodes in the tree = 75
Accuracy of the model on the training dataset = 0.735

Number of validation instances = 2000
Number of validation attributes = 20
Accuracy of the model on the training dataset = 0.666

Number of testing instances = 2000
Number of testing attributes = 20
Accuracy of the model on the testing dataset = 0.6855
```

What we Accomplished?

We have accomplished in creating Decision tree and pruned if for better results.

Observations:

1) We have observed that with the help of training data, we can built a decision tree which would work with 100% accuracy on training data and would be good enough for the test data.

2) If we encounter noisy data in the training data the accuracy of the decision tree decreases.

3) We have observed that if the decision tree is pruned its accuracy increases or decreases depending on the pruning factor and nodes pruned.

4) It is better to prune nodes which are close to the root nodes and to avoid pruning root nodes.

5) The highest accuracy we achieved was with pruning factor of 0.086
   The accuracy of Validation set increased from 0.7585 to 0.763
   The accuracy of Testing set increased from 0.758 to 0.765