* ID3 implementation is done in Java
* Compilation:  
   javac Node.java.  
   java Node trainingDataPath validationDataPath testingDataPath prunningFactor.
* Assumptions:

1. If a leaf node has no data Instances then we labeled the node as 0 (negative class label).
2. After splitting with all the attributes, if the data is still not pure then we labelled the node as 1 if no of positive data instances greater than equal to negative data instances else as 0.
3. Root node of the decision tree is not considered for pruning.
4. The first line will contain the attribute names
5. For every data instance, it is assumed that data value for each entry is non empty

* Accomplished:

Implemented ID3 for decision tree model and did post pruning for improving accuracy in case of overfitting. Based on random number selected pruned nodes, post pruned accuracies vary. To get the maximum accuracy, we are pruning the decision tree for 5 times and outputting the best result obtained based on the validation accuaracy.

* Learned:

Learnt how to construct decision tree training model and how pruning avoids over fitting and improves accuracy provided the nodes chosen to prune are good potential ones. Also learnt how to implement tree data structure in java.

* One of the best result using second dataset and the screenshot of output of Program is copied below:

E:\MS\ML\PA\2\src>java Node training\_set.csv validation\_set.csv test\_set.csv 0.05

Decision Tree Before Pruning

XI = 0 :

| XU = 0 :

| | XQ = 0 :

| | | XF = 0 :

| | | | XO = 0 :

| | | | | XJ = 0 : 0

| | | | | XJ = 1 :

| | | | | | XG = 0 : 0

| | | | | | XG = 1 : 1

| | | | XO = 1 :

| | | | | XB = 0 :

| | | | | | XD = 0 :

| | | | | | | XS = 0 : 0

| | | | | | | XS = 1 : 1

| | | | | | XD = 1 : 0

| | | | | XB = 1 :

| | | | | | XL = 0 : 1

| | | | | | XL = 1 :

| | | | | | | XH = 0 :

| | | | | | | | XS = 0 :

| | | | | | | | | XK = 0 :

| | | | | | | | | | XN = 0 : 1

| | | | | | | | | | XN = 1 : 0

| | | | | | | | | XK = 1 : 1

| | | | | | | | XS = 1 : 0

| | | | | | | XH = 1 : 1

| | | XF = 1 : 0

| | XQ = 1 :

| | | XB = 0 :

| | | | XT = 0 :

| | | | | XL = 0 :

| | | | | | XR = 0 :

| | | | | | | XN = 0 :

| | | | | | | | XP = 0 : 1

| | | | | | | | XP = 1 :

| | | | | | | | | XG = 0 : 1

| | | | | | | | | XG = 1 :

| | | | | | | | | | XS = 0 :

| | | | | | | | | | | XO = 0 : 1

| | | | | | | | | | | XO = 1 : 0

| | | | | | | | | | XS = 1 : 0

| | | | | | | XN = 1 : 0

| | | | | | XR = 1 : 1

| | | | | XL = 1 :

| | | | | | XF = 0 : 0

| | | | | | XF = 1 :

| | | | | | | XC = 0 :

| | | | | | | | XH = 0 : 1

| | | | | | | | XH = 1 :

| | | | | | | | | XS = 0 : 1

| | | | | | | | | XS = 1 : 0

| | | | | | | XC = 1 :

| | | | | | | | XJ = 0 : 0

| | | | | | | | XJ = 1 :

| | | | | | | | | XK = 0 : 1

| | | | | | | | | XK = 1 :

| | | | | | | | | | XM = 0 : 1

| | | | | | | | | | XM = 1 : 0

| | | | XT = 1 :

| | | | | XO = 0 : 1

| | | | | XO = 1 :

| | | | | | XM = 0 :

| | | | | | | XH = 0 :

| | | | | | | | XL = 0 : 0

| | | | | | | | XL = 1 :

| | | | | | | | | XS = 0 : 1

| | | | | | | | | XS = 1 :

| | | | | | | | | | XR = 0 : 1

| | | | | | | | | | XR = 1 : 0

| | | | | | | XH = 1 : 1

| | | | | | XM = 1 : 0

| | | XB = 1 :

| | | | XO = 0 : 0

| | | | XO = 1 :

| | | | | XM = 0 :

| | | | | | XF = 0 :

| | | | | | | XP = 0 :

| | | | | | | | XL = 0 :

| | | | | | | | | XC = 0 : 1

| | | | | | | | | XC = 1 :

| | | | | | | | | | XK = 0 :

| | | | | | | | | | | XT = 0 :

| | | | | | | | | | | | XS = 0 :

| | | | | | | | | | | | | XR = 0 :

| | | | | | | | | | | | | | XN = 0 :

| | | | | | | | | | | | | | | XJ = 0 : 0

| | | | | | | | | | | | | | | XJ = 1 :

| | | | | | | | | | | | | | | | XH = 0 : 0

| | | | | | | | | | | | | | | | XH = 1 :

| | | | | | | | | | | | | | | | | XG = 0 :

| | | | | | | | | | | | | | | | | | XE = 0 :

| | | | | | | | | | | | | | | | | | | XD = 0 : 1

| | | | | | | | | | | | | | | | | | | XD = 1 : 0

| | | | | | | | | | | | | | | | | | XE = 1 : 0

| | | | | | | | | | | | | | | | | XG = 1 : 0

| | | | | | | | | | | | | | XN = 1 : 0

| | | | | | | | | | | | | XR = 1 : 0

| | | | | | | | | | | | XS = 1 : 0

| | | | | | | | | | | XT = 1 : 0

| | | | | | | | | | XK = 1 : 1

| | | | | | | | XL = 1 : 0

| | | | | | | XP = 1 : 1

| | | | | | XF = 1 : 0

| | | | | XM = 1 : 0

| XU = 1 :

| | XG = 0 :

| | | XS = 0 :

| | | | XJ = 0 :

| | | | | XN = 0 :

| | | | | | XB = 0 :

| | | | | | | XF = 0 : 0

| | | | | | | XF = 1 : 1

| | | | | | XB = 1 :

| | | | | | | XK = 0 :

| | | | | | | | XT = 0 : 0

| | | | | | | | XT = 1 : 1

| | | | | | | XK = 1 : 0

| | | | | XN = 1 : 0

| | | | XJ = 1 :

| | | | | XN = 0 :

| | | | | | XQ = 0 : 0

| | | | | | XQ = 1 :

| | | | | | | XO = 0 :

| | | | | | | | XR = 0 :

| | | | | | | | | XM = 0 : 0

| | | | | | | | | XM = 1 : 1

| | | | | | | | XR = 1 : 0

| | | | | | | XO = 1 :

| | | | | | | | XM = 0 : 1

| | | | | | | | XM = 1 : 0

| | | | | XN = 1 :

| | | | | | XE = 0 :

| | | | | | | XD = 0 :

| | | | | | | | XT = 0 :

| | | | | | | | | XR = 0 : 0

| | | | | | | | | XR = 1 : 1

| | | | | | | | XT = 1 : 0

| | | | | | | XD = 1 : 1

| | | | | | XE = 1 : 1

| | | XS = 1 :

| | | | XO = 0 : 0

| | | | XO = 1 :

| | | | | XL = 0 :

| | | | | | XN = 0 :

| | | | | | | XR = 0 : 0

| | | | | | | XR = 1 : 1

| | | | | | XN = 1 : 1

| | | | | XL = 1 : 0

| | XG = 1 : 0

XI = 1 :

| XK = 0 :

| | XC = 0 :

| | | XS = 0 :

| | | | XG = 0 :

| | | | | XM = 0 : 1

| | | | | XM = 1 :

| | | | | | XT = 0 : 1

| | | | | | XT = 1 :

| | | | | | | XR = 0 :

| | | | | | | | XQ = 0 :

| | | | | | | | | XO = 0 : 0

| | | | | | | | | XO = 1 :

| | | | | | | | | | XU = 0 : 1

| | | | | | | | | | XU = 1 : 0

| | | | | | | | XQ = 1 : 1

| | | | | | | XR = 1 :

| | | | | | | | XQ = 0 : 1

| | | | | | | | XQ = 1 : 0

| | | | XG = 1 :

| | | | | XO = 0 : 1

| | | | | XO = 1 :

| | | | | | XL = 0 :

| | | | | | | XH = 0 : 0

| | | | | | | XH = 1 : 1

| | | | | | XL = 1 : 0

| | | XS = 1 :

| | | | XR = 0 :

| | | | | XN = 0 : 1

| | | | | XN = 1 :

| | | | | | XH = 0 :

| | | | | | | XU = 0 : 1

| | | | | | | XU = 1 :

| | | | | | | | XJ = 0 :

| | | | | | | | | XD = 0 : 0

| | | | | | | | | XD = 1 : 1

| | | | | | | | XJ = 1 : 1

| | | | | | XH = 1 :

| | | | | | | XT = 0 : 0

| | | | | | | XT = 1 :

| | | | | | | | XF = 0 : 0

| | | | | | | | XF = 1 : 1

| | | | XR = 1 :

| | | | | XG = 0 :

| | | | | | XM = 0 :

| | | | | | | XH = 0 : 0

| | | | | | | XH = 1 :

| | | | | | | | XB = 0 : 1

| | | | | | | | XB = 1 :

| | | | | | | | | XF = 0 : 1

| | | | | | | | | XF = 1 : 0

| | | | | | XM = 1 :

| | | | | | | XD = 0 : 0

| | | | | | | XD = 1 : 1

| | | | | XG = 1 : 0

| | XC = 1 :

| | | XD = 0 :

| | | | XS = 0 :

| | | | | XF = 0 :

| | | | | | XP = 0 : 0

| | | | | | XP = 1 :

| | | | | | | XR = 0 : 0

| | | | | | | XR = 1 : 1

| | | | | XF = 1 :

| | | | | | XP = 0 : 1

| | | | | | XP = 1 :

| | | | | | | XH = 0 : 1

| | | | | | | XH = 1 :

| | | | | | | | XR = 0 : 0

| | | | | | | | XR = 1 :

| | | | | | | | | XU = 0 :

| | | | | | | | | | XQ = 0 : 1

| | | | | | | | | | XQ = 1 : 0

| | | | | | | | | XU = 1 : 1

| | | | XS = 1 :

| | | | | XJ = 0 :

| | | | | | XQ = 0 :

| | | | | | | XM = 0 : 0

| | | | | | | XM = 1 :

| | | | | | | | XL = 0 : 1

| | | | | | | | XL = 1 :

| | | | | | | | | XT = 0 : 1

| | | | | | | | | XT = 1 :

| | | | | | | | | | XP = 0 : 0

| | | | | | | | | | XP = 1 : 1

| | | | | | XQ = 1 : 0

| | | | | XJ = 1 :

| | | | | | XO = 0 :

| | | | | | | XH = 0 : 1

| | | | | | | XH = 1 :

| | | | | | | | XN = 0 : 0

| | | | | | | | XN = 1 : 1

| | | | | | XO = 1 :

| | | | | | | XR = 0 :

| | | | | | | | XP = 0 :

| | | | | | | | | XT = 0 : 0

| | | | | | | | | XT = 1 : 1

| | | | | | | | XP = 1 : 1

| | | | | | | XR = 1 : 1

| | | XD = 1 :

| | | | XF = 0 : 0

| | | | XF = 1 :

| | | | | XS = 0 :

| | | | | | XQ = 0 : 1

| | | | | | XQ = 1 : 0

| | | | | XS = 1 : 0

| XK = 1 :

| | XD = 0 :

| | | XT = 0 :

| | | | XF = 0 :

| | | | | XB = 0 :

| | | | | | XS = 0 : 0

| | | | | | XS = 1 :

| | | | | | | XU = 0 : 1

| | | | | | | XU = 1 :

| | | | | | | | XO = 0 : 1

| | | | | | | | XO = 1 : 0

| | | | | XB = 1 :

| | | | | | XO = 0 : 0

| | | | | | XO = 1 : 1

| | | | XF = 1 :

| | | | | XO = 0 : 1

| | | | | XO = 1 :

| | | | | | XH = 0 :

| | | | | | | XL = 0 :

| | | | | | | | XS = 0 : 0

| | | | | | | | XS = 1 :

| | | | | | | | | XU = 0 : 0

| | | | | | | | | XU = 1 : 1

| | | | | | | XL = 1 : 1

| | | | | | XH = 1 :

| | | | | | | XN = 0 :

| | | | | | | | XM = 0 :

| | | | | | | | | XS = 0 : 1

| | | | | | | | | XS = 1 : 0

| | | | | | | | XM = 1 : 1

| | | | | | | XN = 1 : 1

| | | XT = 1 :

| | | | XB = 0 :

| | | | | XF = 0 :

| | | | | | XM = 0 : 1

| | | | | | XM = 1 :

| | | | | | | XE = 0 : 0

| | | | | | | XE = 1 :

| | | | | | | | XR = 0 : 1

| | | | | | | | XR = 1 : 0

| | | | | XF = 1 :

| | | | | | XH = 0 : 1

| | | | | | XH = 1 :

| | | | | | | XU = 0 :

| | | | | | | | XR = 0 : 0

| | | | | | | | XR = 1 :

| | | | | | | | | XQ = 0 : 1

| | | | | | | | | XQ = 1 : 0

| | | | | | | XU = 1 : 1

| | | | XB = 1 : 0

| | XD = 1 : 0

Pre-Pruned Accuracy

--------------------------------

Number of training instances = 600

Number of training attributes = 20

Total number of nodes in the tree = 307

Number of leaf nodes in the tree = 154

Accuracy of the model on traianing dataset = 0.9983333333333333

Number of validation instances = 600

Number of validation attributes = 20

Accuracy of the model on validation dataset before pruning = 0.7666666666666667

Number of testing instances = 600

Number of testing attributes = 20

Accuracy of the model on testing dataset = 0.745

Decision Tree After Pruning

XI = 0 :

| XU = 0 :

| | XQ = 0 :

| | | XF = 0 :

| | | | XO = 0 : 0

| | | | XO = 1 :

| | | | | XB = 0 :

| | | | | | XD = 0 :

| | | | | | | XS = 0 : 0

| | | | | | | XS = 1 : 1

| | | | | | XD = 1 : 0

| | | | | XB = 1 : 1

| | | XF = 1 : 0

| | XQ = 1 :

| | | XB = 0 :

| | | | XT = 0 :

| | | | | XL = 0 :

| | | | | | XR = 0 : 1

| | | | | | XR = 1 : 1

| | | | | XL = 1 :

| | | | | | XF = 0 : 0

| | | | | | XF = 1 :

| | | | | | | XC = 0 :

| | | | | | | | XH = 0 : 1

| | | | | | | | XH = 1 :

| | | | | | | | | XS = 0 : 1

| | | | | | | | | XS = 1 : 0

| | | | | | | XC = 1 :

| | | | | | | | XJ = 0 : 0

| | | | | | | | XJ = 1 :

| | | | | | | | | XK = 0 : 1

| | | | | | | | | XK = 1 :

| | | | | | | | | | XM = 0 : 1

| | | | | | | | | | XM = 1 : 0

| | | | XT = 1 :

| | | | | XO = 0 : 1

| | | | | XO = 1 :

| | | | | | XM = 0 :

| | | | | | | XH = 0 :

| | | | | | | | XL = 0 : 0

| | | | | | | | XL = 1 :

| | | | | | | | | XS = 0 : 1

| | | | | | | | | XS = 1 : 1

| | | | | | | XH = 1 : 1

| | | | | | XM = 1 : 0

| | | XB = 1 :

| | | | XO = 0 : 0

| | | | XO = 1 :

| | | | | XM = 0 : 1

| | | | | XM = 1 : 0

| XU = 1 :

| | XG = 0 :

| | | XS = 0 :

| | | | XJ = 0 :

| | | | | XN = 0 :

| | | | | | XB = 0 :

| | | | | | | XF = 0 : 0

| | | | | | | XF = 1 : 1

| | | | | | XB = 1 :

| | | | | | | XK = 0 :

| | | | | | | | XT = 0 : 0

| | | | | | | | XT = 1 : 1

| | | | | | | XK = 1 : 0

| | | | | XN = 1 : 0

| | | | XJ = 1 :

| | | | | XN = 0 :

| | | | | | XQ = 0 : 0

| | | | | | XQ = 1 :

| | | | | | | XO = 0 :

| | | | | | | | XR = 0 :

| | | | | | | | | XM = 0 : 0

| | | | | | | | | XM = 1 : 1

| | | | | | | | XR = 1 : 0

| | | | | | | XO = 1 :

| | | | | | | | XM = 0 : 1

| | | | | | | | XM = 1 : 0

| | | | | XN = 1 :

| | | | | | XE = 0 : 1

| | | | | | XE = 1 : 1

| | | XS = 1 :

| | | | XO = 0 : 0

| | | | XO = 1 :

| | | | | XL = 0 :

| | | | | | XN = 0 :

| | | | | | | XR = 0 : 0

| | | | | | | XR = 1 : 1

| | | | | | XN = 1 : 1

| | | | | XL = 1 : 0

| | XG = 1 : 0

XI = 1 :

| XK = 0 :

| | XC = 0 :

| | | XS = 0 :

| | | | XG = 0 :

| | | | | XM = 0 : 1

| | | | | XM = 1 :

| | | | | | XT = 0 : 1

| | | | | | XT = 1 :

| | | | | | | XR = 0 :

| | | | | | | | XQ = 0 :

| | | | | | | | | XO = 0 : 0

| | | | | | | | | XO = 1 :

| | | | | | | | | | XU = 0 : 1

| | | | | | | | | | XU = 1 : 0

| | | | | | | | XQ = 1 : 1

| | | | | | | XR = 1 :

| | | | | | | | XQ = 0 : 1

| | | | | | | | XQ = 1 : 0

| | | | XG = 1 :

| | | | | XO = 0 : 1

| | | | | XO = 1 :

| | | | | | XL = 0 :

| | | | | | | XH = 0 : 0

| | | | | | | XH = 1 : 1

| | | | | | XL = 1 : 0

| | | XS = 1 :

| | | | XR = 0 :

| | | | | XN = 0 : 1

| | | | | XN = 1 :

| | | | | | XH = 0 :

| | | | | | | XU = 0 : 1

| | | | | | | XU = 1 :

| | | | | | | | XJ = 0 :

| | | | | | | | | XD = 0 : 0

| | | | | | | | | XD = 1 : 1

| | | | | | | | XJ = 1 : 1

| | | | | | XH = 1 :

| | | | | | | XT = 0 : 0

| | | | | | | XT = 1 :

| | | | | | | | XF = 0 : 0

| | | | | | | | XF = 1 : 1

| | | | XR = 1 :

| | | | | XG = 0 :

| | | | | | XM = 0 :

| | | | | | | XH = 0 : 0

| | | | | | | XH = 1 :

| | | | | | | | XB = 0 : 1

| | | | | | | | XB = 1 :

| | | | | | | | | XF = 0 : 1

| | | | | | | | | XF = 1 : 0

| | | | | | XM = 1 :

| | | | | | | XD = 0 : 0

| | | | | | | XD = 1 : 1

| | | | | XG = 1 : 0

| | XC = 1 :

| | | XD = 0 :

| | | | XS = 0 :

| | | | | XF = 0 :

| | | | | | XP = 0 : 0

| | | | | | XP = 1 :

| | | | | | | XR = 0 : 0

| | | | | | | XR = 1 : 1

| | | | | XF = 1 :

| | | | | | XP = 0 : 1

| | | | | | XP = 1 :

| | | | | | | XH = 0 : 1

| | | | | | | XH = 1 :

| | | | | | | | XR = 0 : 0

| | | | | | | | XR = 1 :

| | | | | | | | | XU = 0 :

| | | | | | | | | | XQ = 0 : 1

| | | | | | | | | | XQ = 1 : 0

| | | | | | | | | XU = 1 : 1

| | | | XS = 1 :

| | | | | XJ = 0 :

| | | | | | XQ = 0 :

| | | | | | | XM = 0 : 0

| | | | | | | XM = 1 :

| | | | | | | | XL = 0 : 1

| | | | | | | | XL = 1 :

| | | | | | | | | XT = 0 : 1

| | | | | | | | | XT = 1 :

| | | | | | | | | | XP = 0 : 0

| | | | | | | | | | XP = 1 : 1

| | | | | | XQ = 1 : 0

| | | | | XJ = 1 :

| | | | | | XO = 0 :

| | | | | | | XH = 0 : 1

| | | | | | | XH = 1 :

| | | | | | | | XN = 0 : 0

| | | | | | | | XN = 1 : 1

| | | | | | XO = 1 :

| | | | | | | XR = 0 :

| | | | | | | | XP = 0 :

| | | | | | | | | XT = 0 : 0

| | | | | | | | | XT = 1 : 1

| | | | | | | | XP = 1 : 1

| | | | | | | XR = 1 : 1

| | | XD = 1 :

| | | | XF = 0 : 0

| | | | XF = 1 :

| | | | | XS = 0 :

| | | | | | XQ = 0 : 1

| | | | | | XQ = 1 : 0

| | | | | XS = 1 : 0

| XK = 1 :

| | XD = 0 :

| | | XT = 0 :

| | | | XF = 0 :

| | | | | XB = 0 :

| | | | | | XS = 0 : 0

| | | | | | XS = 1 :

| | | | | | | XU = 0 : 1

| | | | | | | XU = 1 :

| | | | | | | | XO = 0 : 1

| | | | | | | | XO = 1 : 0

| | | | | XB = 1 :

| | | | | | XO = 0 : 0

| | | | | | XO = 1 : 1

| | | | XF = 1 :

| | | | | XO = 0 : 1

| | | | | XO = 1 :

| | | | | | XH = 0 :

| | | | | | | XL = 0 :

| | | | | | | | XS = 0 : 0

| | | | | | | | XS = 1 :

| | | | | | | | | XU = 0 : 0

| | | | | | | | | XU = 1 : 1

| | | | | | | XL = 1 : 1

| | | | | | XH = 1 : 1

| | | XT = 1 :

| | | | XB = 0 :

| | | | | XF = 0 :

| | | | | | XM = 0 : 1

| | | | | | XM = 1 :

| | | | | | | XE = 0 : 0

| | | | | | | XE = 1 :

| | | | | | | | XR = 0 : 1

| | | | | | | | XR = 1 : 0

| | | | | XF = 1 :

| | | | | | XH = 0 : 1

| | | | | | XH = 1 : 1

| | | | XB = 1 : 0

| | XD = 1 : 0

Post-Pruned Accuracy

--------------------------------

Number of training instances = 600

Number of training attributes = 20

Total number of nodes in the tree = 235

Number of leaf nodes in the tree = 118

Accuracy of the model on traianing dataset = 0.9583333333333334

Number of validation instances = 600

Number of validation attributes = 20

Accuracy of the model on validation dataset before pruning = 0.7716666666666666

Number of testing instances = 600

Number of testing attributes = 20

Accuracy of the model on testing dataset = 0.76