Hamid Muneerulhudhakalvathi N hxm151830@utdallas.edu

Assignment 2 Report

Approach 1 (using rpart):

source file: Assignment2.R
Packages used: rpart, rpart.plot
The Assignment2.R file has to be the same folder as the dataset csv files.

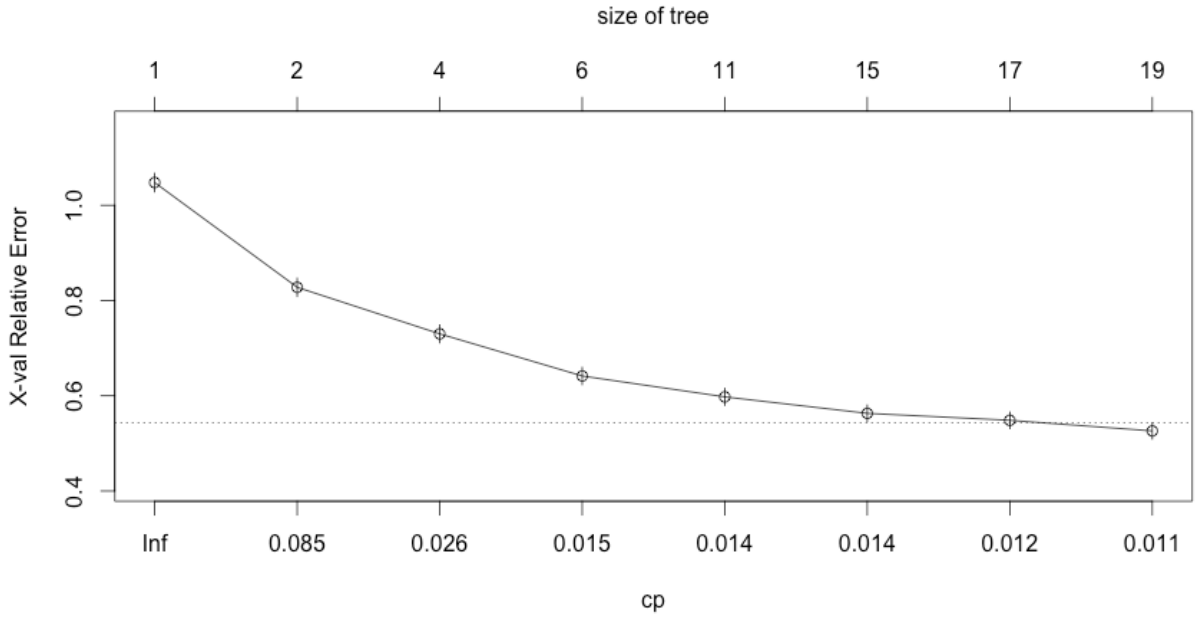
DataSet 1:

printcp on TrainingData before pruning:

```
Classification tree:
rpart(formula = Class ~ ., data = trainingData, method = "class",
    parms = list(split = "information"))

Variables actually used in tree construction:
 [1] XB XD XF XG XI XK XM XO XQ XS XU

Root node error: 1300/2600 = 0.5

n= 2600

        CP nsplit rel error  xerror     xstd
1 0.172308      0   1.00000 1.04769 0.019589
2 0.042308      1   0.82769 0.82769 0.019318
3 0.015385      3   0.74308 0.73000 0.018883
4 0.014615      5   0.71231 0.64154 0.018308
5 0.013846     10   0.60692 0.59769 0.017955
6 0.013462     14   0.55154 0.56308 0.017641
7 0.011538     16   0.52462 0.54846 0.017498
8 0.010000     18   0.50154 0.52615 0.017270
```

An appropriate cp value is chosen from the table for the prune function.

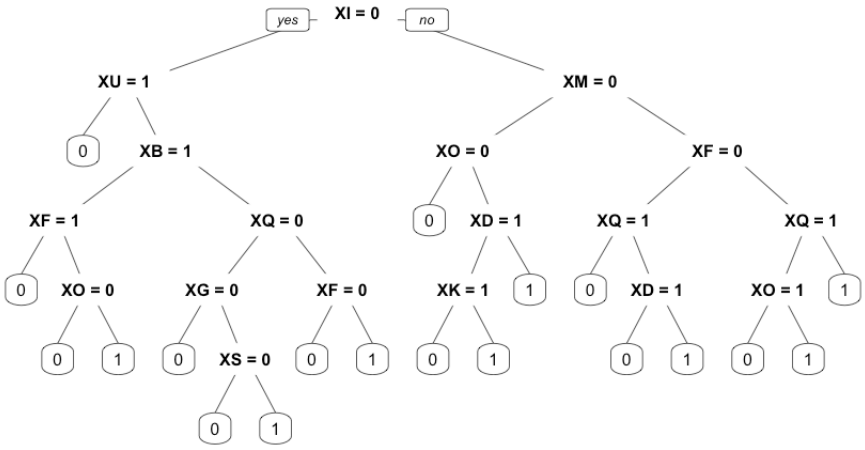
Plotting the cp values gets the following plot:



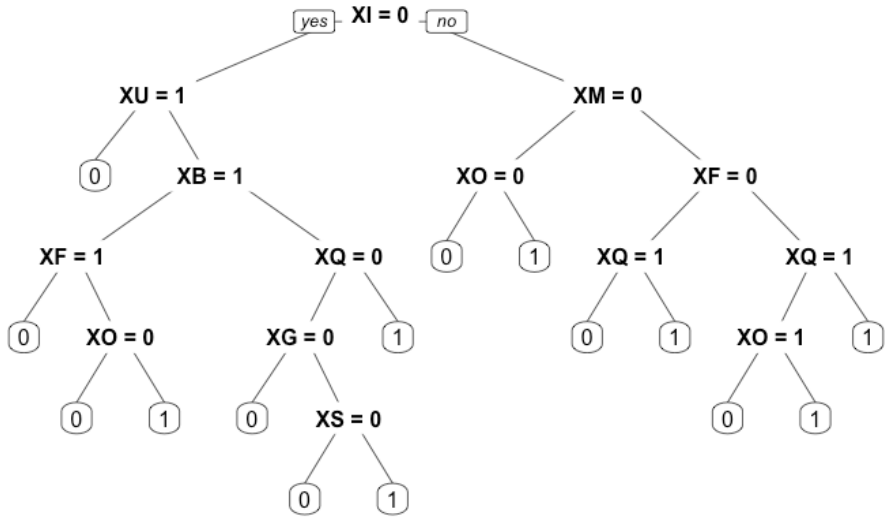
CP value is chosen from the plot as 0.013462

For plotting, prp function from the rpart.plot library is used.

Training Data plot before pruning:



Training Data plot after pruning:



Output:

Accuracy of the plot before pruning: 0.719
Accuracy of the plot after pruning: 0.6975

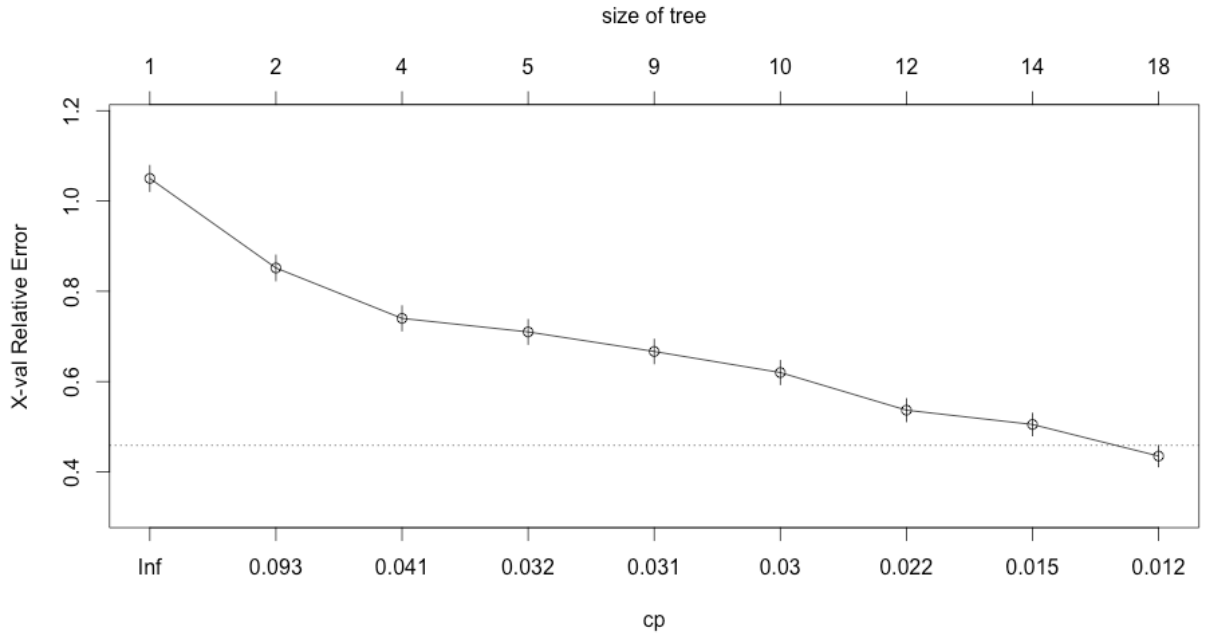
DataSet 2:

printcp on TrainingData before pruning:

```
Classification tree:
rpart(formula = Class ~ ., data = trainingData, method = "class",
    parms = list(split = "information"))

Variables actually used in tree construction:
 [1] XB XC XD XF XG XI XJ XK XO XQ XS XU

Root node error: 600/1200 = 0.5

n= 1200

        CP nsplit rel error  xerror     xstd
1 0.170000      0   1.00000 1.05000 0.028831
2 0.050833      1   0.83000 0.85167 0.028548
3 0.033333      3   0.72833 0.74000 0.027875
4 0.031667      4   0.69500 0.71000 0.027627
5 0.030000      8   0.56833 0.66667 0.027217
6 0.029167      9   0.53833 0.62000 0.026702
7 0.016667     11   0.48000 0.53667 0.025582
8 0.013333     13   0.44667 0.50500 0.025083
9 0.010000     17   0.39333 0.43500 0.023818
```

An appropriate cp value is chosen from the table for the prune function.

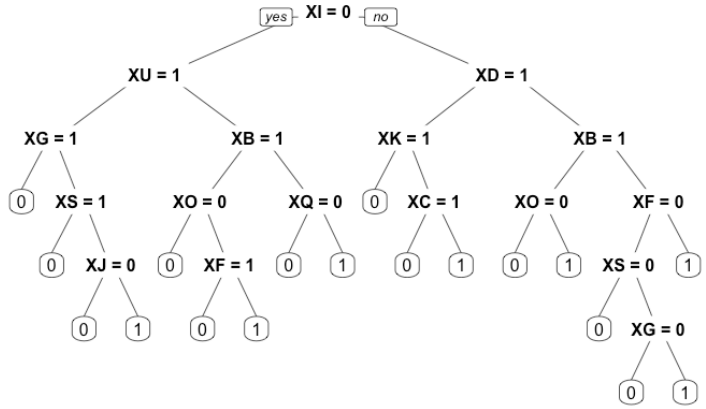
Plotting the cp values gets the following plot:



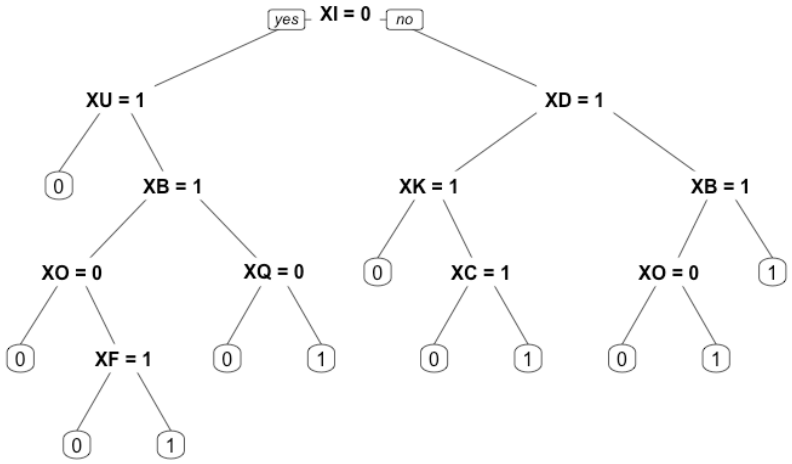
CP value is chosen from the plot as 0.016667

For plotting, prp function from the rpart.plot library is used.

Training Data plot before pruning:



Training Data plot after pruning:



Output:

Accuracy of the plot before pruning: 0.76
Accuracy of the plot after pruning: 0.708

Approach 2 (using Java):

Data Set 1:

Accuracy on data set 1: 0.427

Tree paths:

```
XC->XR->XG->XK->XI->XF->XB->0
XC->XR->XG->XK->XI->XF->XB->XL->XM->XH->0
XC->XR->XG->XK->XI->XF->XB->XL->XM->XH->1
XC->XR->XG->XK->XI->XF->XB->XL->XM->1
XC->XR->XG->XK->XI->XF->XB->XL->XD->XU->1
XC->XR->XG->XK->XI->XF->XB->XL->XD->XU->0
XC->XR->XG->XK->XI->XF->XB->XL->XD->0
XC->XR->XG->XK->XI->XF->XP->0
XC->XR->XG->XK->XI->XF->XP->XS->1
XC->XR->XG->XK->XI->XF->XP->XS->0
XC->XR->XG->XK->XI->XN->XO->XJ->1
XC->XR->XG->XK->XI->XN->XO->XJ->XE->
XC->XR->XG->XK->XI->XN->XO->XJ->XE->XT->XQ->0
XC->XR->XG->XK->XI->XN->XO->XJ->XE->XT->XQ->0
XC->XR->XG->XK->XI->XN->XO->XJ->XE->XT->0
XC->XR->XG->XK->XI->XN->XO->0
XC->XR->XG->XK->XI->XN->0
XC->XR->XG->XK->1
XC->XR->XG->0
XC->XR->1
XC->1
```

Tree printed in the requested format:

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

Data Set 2:

Accuracy on data set 2: 0.425

Tree printed in requested format:

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

Tree paths:

```
XI->XU->XJ->XN->XC->XO->XQ->0
XI->XU->XJ->XN->XC->XO->XQ->XR->XB->XF->XL->1
XI->XU->XJ->XN->XC->XO->XQ->XR->XB->XF->XL->0
XI->XU->XJ->XN->XC->XO->XQ->XR->XB->XF->1
XI->XU->XJ->XN->XC->XO->XQ->XR->XB->0
XI->XU->XJ->XN->XC->XO->XQ->XR->1
XI->XU->XJ->XN->XC->XO->XS->1
XI->XU->XJ->XN->XC->XO->XS->XG->XK->XP->XT->
XI->XU->XJ->XN->XC->XO->XS->XG->XK->XP->XT->XD->XE->XH->
XI->XU->XJ->XN->XC->XO->XS->XG->XK->XP->XT->XD->XE->XH->XM->0
XI->XU->XJ->XN->XC->XO->XS->XG->XK->XP->XT->XD->XE->XH->XM->1
XI->XU->XJ->XN->XC->XO->XS->XG->XK->XP->XT->XD->XE->0
XI->XU->XJ->XN->XC->XO->XS->XG->XK->XP->XT->XD->0
XI->XU->XJ->XN->XC->XO->XS->XG->XK->XP->1
XI->XU->XJ->XN->XC->XO->XS->XG->XK->0
XI->XU->XJ->XN->XC->XO->XS->XG->0
XI->XU->XJ->XN->XC->0
XI->XU->XJ->XN->1
XI->XU->XJ->0
XI->XU->1
XI->0
```

Pruning not implemented in Java.