

Name – Kanika Kapoor

NetID – kxk140230

Title – Machine Learning Assignment 02 Decision Trees

## **ACCURACY RESULTS**

Accuracy results have been tested for all cp values against minsplit = 2 and minsplit = 100. These 2 minsplit values have been chosen randomly whereas the default minsplit value is 20.

### **1) DATASET – 1**

DATASET	MINSPLIT	VALUE OF CP	ACCURACY(%)
Dataset - 1	2	0.172308	50
		0.042308	57.1
		0.015385	61.25
		0.014615	67.75
		0.013846	69.75
		0.013462	69.75
		0.011538	71.9
		0.010000	71.9

DATASET	MINSPLIT	VALUE OF CP	ACCURACY(%)
Dataset - 1	100	0.172308	50
		0.042308	57.1
		0.015385	61.25
		0.014615	67.65
		0.013846	68.75
		0.013462	68.75
		0.011538	70.9
		0.010000	70.65

## 2) DATASET - 2

DATASET	MINSPLIT	VALUE OF CP	ACCURACY(%)
Dataset - 2	2	0.170000	50
		0.050833	61
		0.033333	62
		0.031667	62
		0.030000	70.5
		0.029167	70.5
		0.016667	70.833
		0.013333	76
		0.010000	76

DATASET	MINSPLIT	VALUE OF CP	ACCURACY(%)
Dataset - 2	100	0.170000	50
		0.050833	61
		0.033333	62
		0.031667	62
		0.030000	70.5
		0.029167	70.5
		0.016667	69
		0.013333	72
		0.010000	72

## Workspace Screenshot

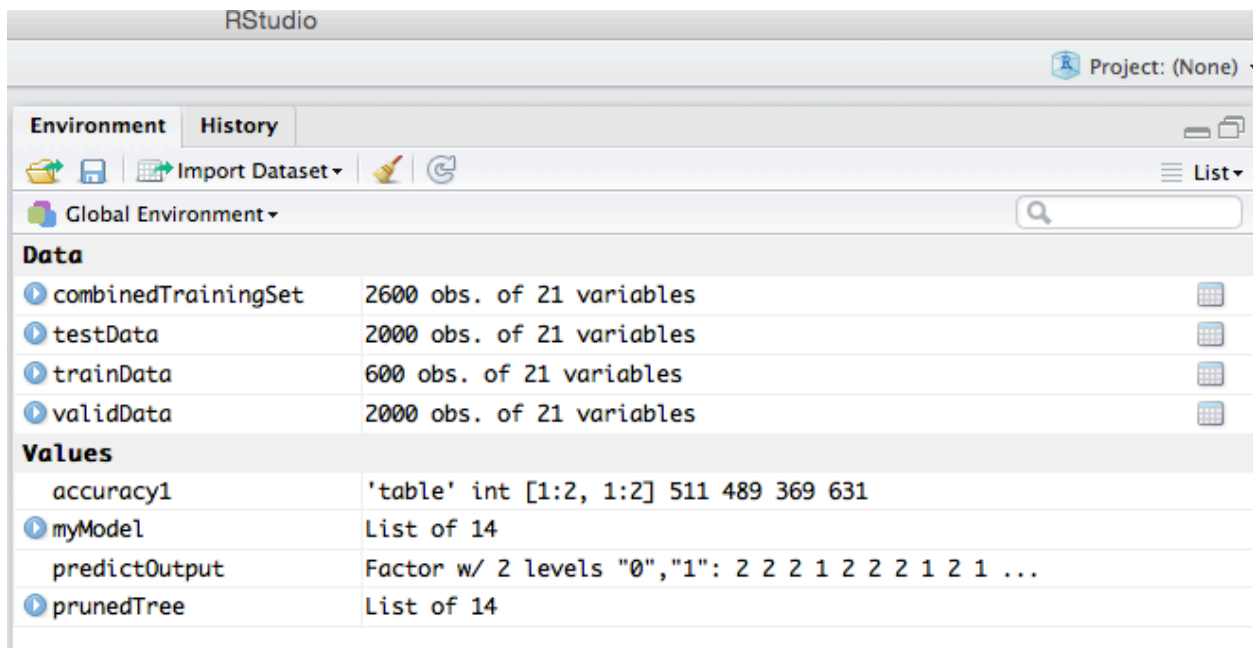
### 1) DataSet-1

It shows that for dataset – 1,

Training file contains 600 rows or observations

Validation file contains 2000 rows or observations

Combined Training Set contains 2600 rows



The screenshot shows the RStudio interface with the 'Environment' pane active. It displays the 'Global Environment' with a search bar. The 'Data' section lists four objects: 'combinedTrainingSet' (2600 obs. of 21 variables), 'testData' (2000 obs. of 21 variables), 'trainData' (600 obs. of 21 variables), and 'validData' (2000 obs. of 21 variables). The 'Values' section lists four objects: 'accuracy1' (a table of integers), 'myModel' (a list of 14), 'predictOutput' (a factor with 2 levels), and 'prunedTree' (a list of 14).

Data	
combinedTrainingSet	2600 obs. of 21 variables
testData	2000 obs. of 21 variables
trainData	600 obs. of 21 variables
validData	2000 obs. of 21 variables

Values	
accuracy1	'table' int [1:2, 1:2] 511 489 369 631
myModel	List of 14
predictOutput	Factor w/ 2 levels "0","1": 2 2 2 1 2 2 2 1 2 1 ...
prunedTree	List of 14

### 2) DataSet-2

It shows that for dataset – 2,

Training file contains 600 rows or observations

Validation file contains 600 rows or observations

Combined Training Set contains 1200 rows

