Name : Cheruvathoor Abin Anto
Roll No : TIT2425008

PRACTICAL 6

The file Iris.csv contains 50 samples from each of 3 species of Iris (Iris setosa,Iris virginica,Iris versicolor).

A) Split the data to training and test data. Build the decision tree for this data

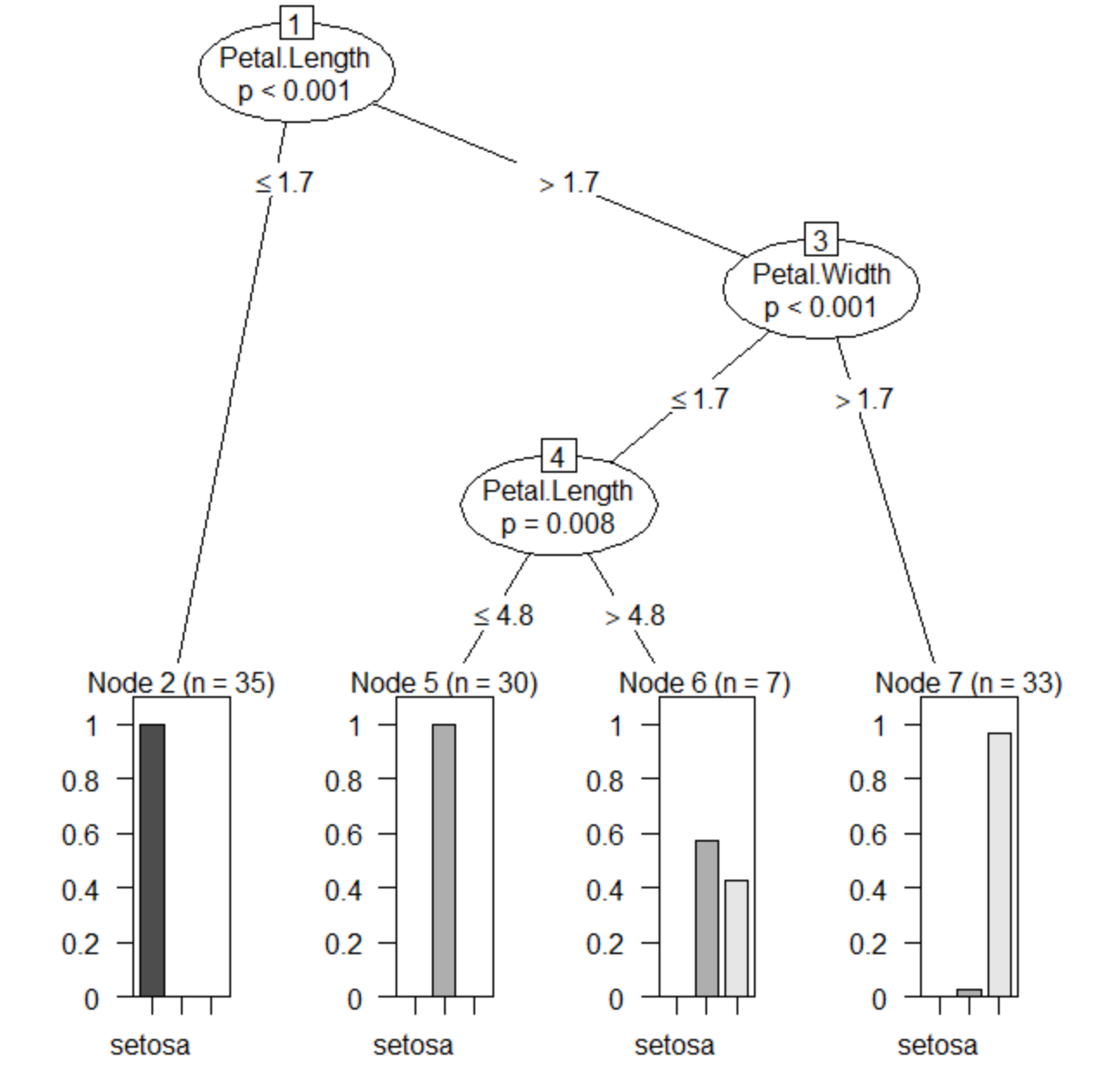
```
iris.df <- read.csv("Iris.csv")

#libraries to be Included

library(caTools) #for split fn

install.packages("party") #for ctree fn

library(party)


#split training and test data

split < sample.splt(iris,SplitRatio = 0.8)

train_cl <- subset(iris,split == "TRUE")

test_cl <- subset(iris,split == "FALSE")
```

```
> train_cl <- subset(iris,split == "TRUE")
> test_cl <- subset(iris,split == "FALSE")
> train_cl
    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
2            4.9         3.0          1.4         0.2     setosa
5            5.0         3.6          1.4         0.2     setosa
6            5.4         3.9          1.7         0.4     setosa
7            4.6         3.4          1.4         0.3     setosa
9            4.4         2.9          1.4         0.2     setosa
10           4.9         3.1          1.5         0.1     setosa
11           5.4         3.7          1.5         0.2     setosa
12           4.8         3.4          1.6         0.2     setosa
13           4.8         3.0          1.4         0.1     setosa
15           5.8         4.0          1.2         0.2     setosa
17           5.4         3.9          1.3         0.4     setosa
18           5.1         3.5          1.4         0.3     setosa
19           5.7         3.8          1.7         0.3     setosa
21           5.4         3.4          1.7         0.2     setosa
23           4.6         3.6          1.0         0.2     setosa
24           5.1         3.3          1.7         0.5     setosa
26           5.0         3.0          1.6         0.2     setosa
27           5.0         3.4          1.6         0.4     setosa
28           5.2         3.5          1.5         0.2     setosa
29           5.2         3.4          1.4         0.2     setosa
30           4.7         3.2          1.6         0.2     setosa
32           5.4         3.4          1.5         0.4     setosa
33           5.2         4.1          1.5         0.1     setosa
```

#Create the Decision Tree

dt <- ctree(Species~.,train_cl)

plot(dt)


1
Petal.Length
p < 0.001
≤ 1.7
> 1.7
3
Petal.Width
p < 0.001
≤ 1.7
> 1.7
4
Petal.Length
p = 0.008
≤ 4.8
> 4.8
Node 2 (n = 35)
Node 5 (n = 30)
Node 6 (n = 7)
Node 7 (n = 33)
1
0.8
0.6
0.4
0.2
0
setosa

B) Predict the species for the test data and determine the accuracy of the model

#predict the test data

p <- predict(dt,test_cl)

p

```
> dt <- ctree(Species~.,train_cl)
> plot(dt)
> p <- predict(dt,test_cl)
> p
 [1] setosa     setosa     setosa     setosa     setosa     setosa     setosa
 [8] setosa     versicolor setosa     setosa     setosa     setosa     versicolor
[15] setosa     versicolor versicolor versicolor versicolor versicolor versicolor
[22] versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[29] versicolor versicolor virginica  versicolor virginica  virginica  virginica
[36] virginica  virginica  virginica  versicolor virginica  virginica  virginica
[43] virginica  virginica  virginica
Levels: setosa versicolor virginica
```

#confusion matrix
cm <- table(test_cl$Species,p)
cm
confusionMatrix(cm)

```
> cm <- table(test_cl$Species,p)
> cm
            p
             setosa versicolor virginica
  setosa         13          2         0
  versicolor      0         15         0
  virginica       0          2        13
> confusionMatrix(cm)
Confusion Matrix and Statistics

            p
             setosa versicolor virginica
  setosa         13          2         0
  versicolor      0         15         0
  virginica       0          2        13

Overall Statistics

               Accuracy : 0.9111
                 95% CI : (0.7878, 0.9752)
    No Information Rate : 0.4222
    P-Value [Acc > NIR] : 7.909e-12

                  Kappa : 0.8667

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: setosa Class: versicolor Class: virginica
Sensitivity                 1.0000            0.7895           1.0000
Specificity                 0.9375            1.0000           0.9375
Pos Pred Value              0.8667            1.0000           0.8667
Neg Pred Value              1.0000            0.8667           1.0000
Prevalence                  0.2889            0.4222           0.2889
Detection Rate              0.2889            0.3333           0.2889
Detection Prevalence        0.3333            0.3333           0.3333
Balanced Accuracy           0.9688            0.8947           0.9688
```

#accuracy
acc <- sum(diag(cm))/sum(cm)
print(paste("Accuracy is",acc))

```
> acc <- sum(diag(cm))/sum(cm)
> print(paste("Accuracy is",acc))
[1] "Accuracy is 0.911111111111111"
>
```