CREATION OF DECISTION TREE-IRIS DATA SET

1. Preliminary analysis

library(party)

data(iris)

str(iris)

summary(iris)

dim(iris)

```
> data(iris)
> str(iris)
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
> summary(iris)
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width          Species
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa    :50
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   versicolor:50
 Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
> dim(iris)
[1] 150   5
```

2. Checking missing values

df=data.frame(num_missing=colSums(is.na(iris)))

print(df)

```
             num_missing
Sepal.Length           0
Sepal.Width            0
Petal.Length           0
Petal.Width            0
Species                0
```

3. Partitioning of Data set into Training and Testing data

set.seed(555)

ind=sample(2,nrow(iris),replace=T,prob=c(0.8,0.2))

print(ind)

```
  [1] 1 2 1 1 1 1 2 1 1 2 2 1 1 1 2 1 1 1 1 2 1 1
1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
1 1 1 1 1 1
 [54] 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1
2 1 1 1 1 2 2 2 2 2 2 1 1 1 1 1 1 1 2 1 2 2 1 1 1
2 2 1 2 1 1
[107] 1 2 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1
2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 2
```

4. Creation of Training data set

train=iris[ind==1,]

print(head(train))

print(dim(train))

```
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
8          5.0         3.4          1.5         0.2  setosa
> print(dim(train))
[1] 118   5
```

5. Creation of Testing data set

test=iris[ind==2,]

print(head(test))

print(dim(test))

```
   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
2           4.9         3.0          1.4         0.2  setosa
7           4.6         3.4          1.4         0.3  setosa
10          4.9         3.1          1.5         0.1  setosa
11          5.4         3.7          1.5         0.2  setosa
15          5.8         4.0          1.2         0.2  setosa
20          5.1         3.8          1.5         0.3  setosa
> print(dim(test))
[1] 32  5
```

6. Creation of Decision Tree

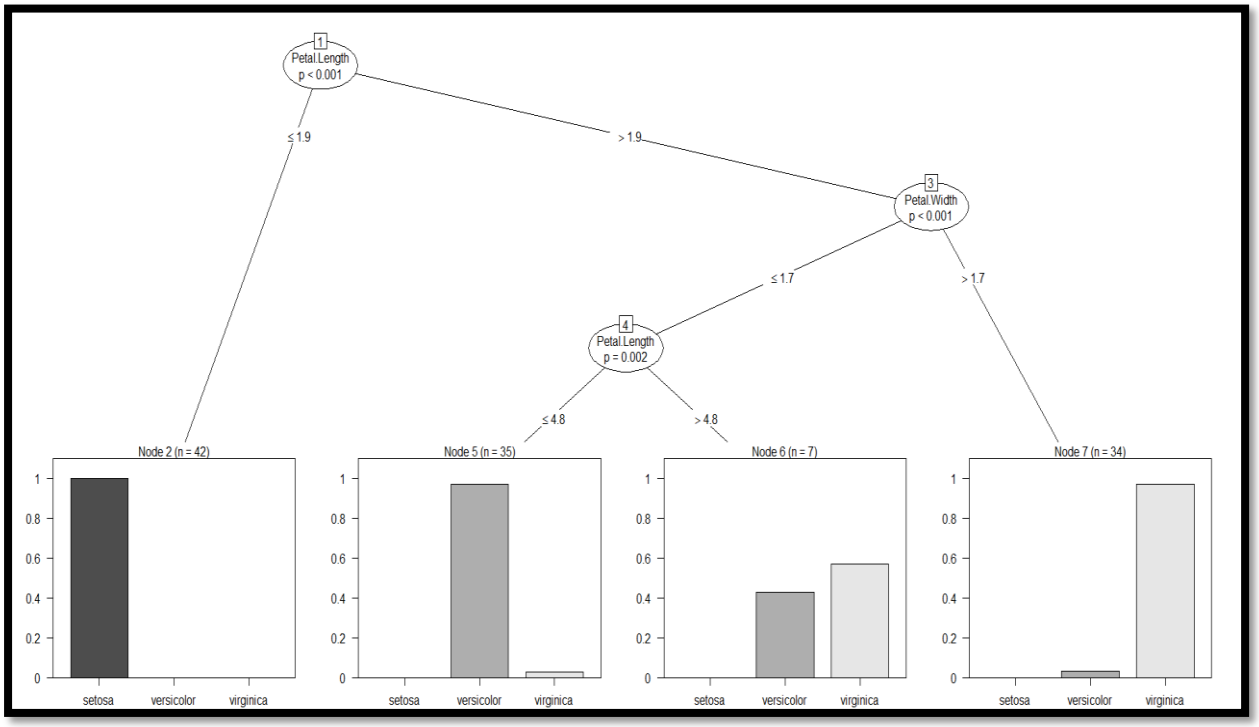
library(party)

```
tree=ctree(Species~.,train)
print(tree)
plot(tree)
```

```
        Conditional inference tree with 4 terminal nodes

Response:  Species
Inputs:  Sepal.Length, Sepal.Width, Petal.Length, Petal.Width
Number of observations:  118

1) Petal.Length ≤ 1.9; criterion = 1, statistic = 110.067
  2)*  weights = 42
1) Petal.Length > 1.9
  3) Petal.Width ≤ 1.7; criterion = 1, statistic = 50.039
    4) Petal.Length ≤ 4.8; criterion = 0.998, statistic = 12.36
      5)*  weights = 35
    4) Petal.Length > 4.8
      6)*  weights = 7
  3) Petal.Width > 1.7
    7)*  weights = 34
```



7. Create new data for testing the model

```
 new_data=list(
  Sepal.Length=c(5.1,5.5,6.3),
  Sepal.Width=c(3.5,2.8,3.3),
  Petal.Length=c(1.4,4.2,5.1),
  Petal.Width=c(0.2,1.3,1.8)
)
print(new_data)
```

```
$Sepal.Length
[1] 5.1 5.5 6.3

$Sepal.Width
[1] 3.5 2.8 3.3

$Petal.Length
[1] 1.4 4.2 5.1

$Petal.Width
[1] 0.2 1.3 1.8
```

8. Creating model Predictions on new data

predictions_newdata=predict(tree,newdata=new_data,type="response")

print(predictions_newdata)

```
[1] setosa     versicolor virginica
Levels: setosa versicolor virginica
```

9. Calculating Accuracy of the model

a. Get Model prediction on the test data

predictions_test=predict(tree,newdata=test,type="response")
print(predictions_test)

```
 [1] setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     versicolor
[10] versicolor versicolor versicolor versicolor versicolor virginica  versicolor versicolor versicolor
[19] versicolor versicolor virginica  virginica  virginica  virginica  virginica  virginica  virginica
[28] virginica  virginica  virginica  virginica  virginica
Levels: setosa versicolor virginica
```

b. Calculate Accuracy

accuracy=mean(predictions_test==test$Species)
print(paste("Accuracy:", accuracy))

```
[1] "Accuracy: 0.96875"
```