#Iris Decision tree code

```
library(readxl)#to read excel
library(plyr)#split and put data
library(DescTools) #For Mode
library(rpart.plot) #For DT Graph
library(psych) #descriptive stats
library(dplyr)#data manipulation
library(rpart)#For DT model
library(moments)#skewness kurtosis stuff
library(caret)#data splitting. pre-processing. feature selection
library(purrr)#package for vector

#getting the data set
iris<- read.csv("C:\\Users\\sujoydutta\\Desktop\\Data analysis\\Datasets\\Clustering\\Iris.csv")



#seeing the data set
View(iris)
str(iris)

str(iris)
#checking null values
map(iris, ~sum(is.na(.)))

#summary statistics
describe(iris)

#viewing outliers
boxplot(iris)$out
```

```
#dropping uneccesary outliers

iris<- iris[-c(1)]

iris

# Eliminating outliers using Quartile method

iqr <- IQR(iris$SepalWidthCm)

Q <- quantile(irisk$SepalWidthCm, probs=c(.25, .75), na.rm = FALSE)

eliminated <- subset(iris, irisk$SepalWidthCm > (Q[1] - 1.5*iqr) & iris$SepalWidthCm <
(Q[2]+1.5*iqr))

iqr <- IQR(eliminated$SepalWidthCm)

Q <- quantile(eliminated$SepalWidthCm, probs=c(.25, .75), na.rm = FALSE)

irisdt <- subset(eliminated, eliminated$SepalWidthCm > (Q[1] - 1.5*iqr) & eliminated$SepalWidthCm
< (Q[2]+1.5*iqr))

remove(eliminated)


# For reproducibility

set.seed(100)


# Create index for testing and training data

indexes = sample(80, 70)


#splitting dataset into train and test


iris_train = iris[indexes,]

iris_test = iris[-indexes,]


#building the model

target = Species ~ SepalLengthCm + SepalWidthCm + PetalLengthCm + PetalWidthCm


target = Species ~ .
```

```
#Build and plot model
tree = rpart(target, data = iris_train, method = "class")
rpart.plot(tree)



#predictions
predictions = predict(tree, iris_test)

#confusion matrix
table(predictions[0:150], iris$Species[0:150])
```