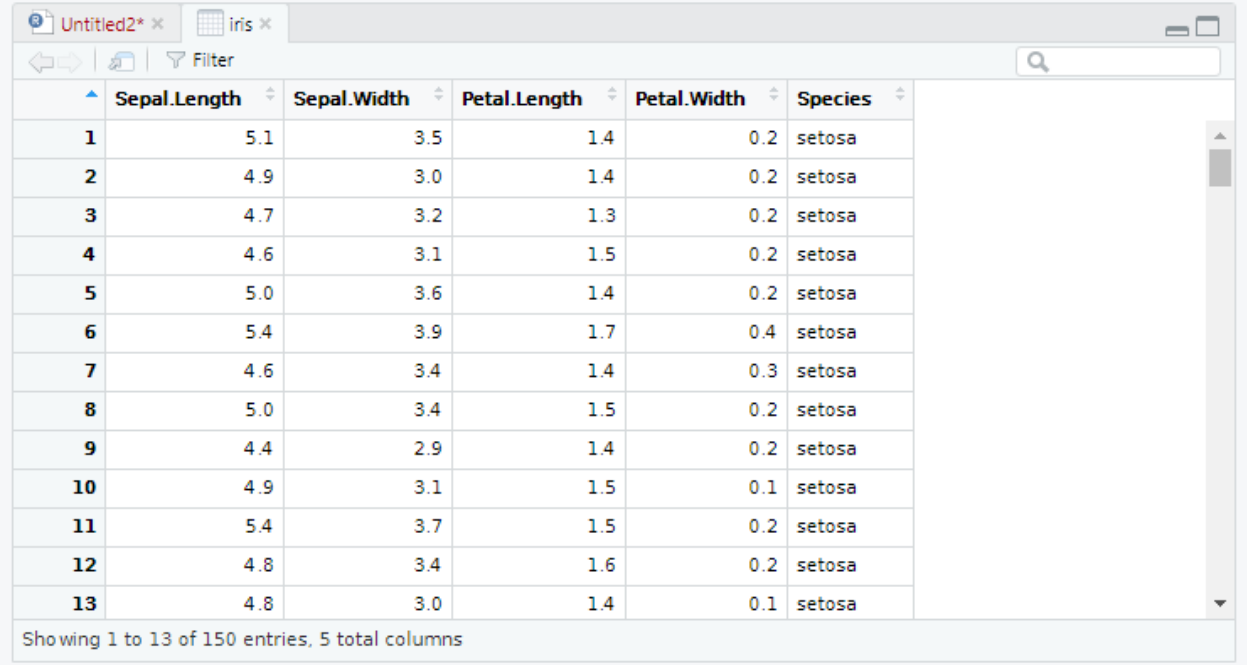
Name : Abin Anto Cheruvathoor
Roll No : TIT2425008

PRACTICAL 5

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 KNN model for this data with different ‘k’ values

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


Untitled2* × iris ×

Filter

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	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
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	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

Showing 1 to 13 of 150 entries, 5 total columns

#libraries to be included
library(caTools)

```
> setwd("D:/Aby/DM")
Warning message:
R graphics engine version 16 is not supported by this version of RStudio. The Plots tab
 will be disabled until a newer version of RStudio is installed.
> iris.df <- read.csv("Iris.csv")
> View(iris)
> library(caTools)
```

#split training and test data
split <- sample.split(iris$Species,SplitRatio=0.7)
split

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

```
package 'caTools' was built under R version 4.3.3
> split <- sample.split(iris$Species,SplitRatio=0.7)
> split
  [1] FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
 [14] FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE FALSE  TRUE
 [27]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
 [40]  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE
 [53]  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [66] FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
 [79] FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
 [92] FALSE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
[105]  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
[118]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[131]  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
[144] FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
> train_cl<-subset(iris,split=="TRUE")
```

#scaling training and test data
train_scale <- scale(train_cl[,1:4])
test_scale <- scale(test_cl[,1:4])

```
> train_scale
    Sepal.Length Sepal.Width Petal.Length  Petal.Width
2    -1.12716278 -0.06977776  -1.33213930 -1.320446985
5    -1.00986635  1.34828631  -1.33213930 -1.320446985
6    -0.54068066  2.05731835  -1.16402076 -1.058106525
7    -1.47905205  0.87559829  -1.33213930 -1.189276755
9    -1.71364489 -0.30612177  -1.33213930 -1.320446985
10   -1.12716278  0.16656626  -1.27609979 -1.451617216
11   -0.54068066  1.58463033  -1.27609979 -1.320446985
12   -1.24445920  0.87559829  -1.22006028 -1.320446985
13   -1.24445920 -0.06977776  -1.33213930 -1.451617216
15   -0.07149496  2.29366236  -1.44421833 -1.320446985
17   -0.54068066  2.05731835  -1.38817882 -1.058106525
18   -0.89256993  1.11194230  -1.33213930 -1.189276755
19   -0.18879139  1.82097434  -1.16402076 -1.189276755
21   -0.54068066  0.87559829  -1.16402076 -1.320446985
23   -1.47905205  1.34828631  -1.55629736 -1.320446985
24   -0.89256993  0.63925428  -1.16402076 -0.926936294
26   -1.00986635 -0.06977776  -1.22006028 -1.320446985
27   -1.00986635  0.87559829  -1.22006028 -1.058106525
28   -0.77527350  1.11194230  -1.27609979 -1.320446985
29   -0.77527350  0.87559829  -1.33213930 -1.320446985
30   -1.36175562  0.40291027  -1.22006028 -1.320446985
32   -0.54068066  0.87559829  -1.27609979 -1.058106525
33   -0.77527350  2.53000637  -1.27609979 -1.451617216
35   -1.12716278  0.16656626  -1.27609979 -1.320446985
36   -1.00986635  0.40291027  -1.44421833 -1.320446985
38   -1.12716278  1.34828631  -1.33213930 -1.451617216
39   -1.71364489 -0.06977776  -1.38817882 -1.320446985
40   -0.89256993  0.87559829  -1.27609979 -1.320446985
42   -1.59634847 -1.72418584  -1.38817882 -1.189276755
43   -1.71364489  0.40291027  -1.38817882 -1.320446985
44   -1.00986635  1.11194230  -1.22006028 -0.795766064
46   -1.24445920 -0.06977776  -1.33213930 -1.189276755
140   1.21876570  0.16656626   0.90944126  1.171787391
142   1.21876570  0.16656626   0.74132271  1.434127852
143  -0.07149496 -0.77880979   0.74132271  0.909446930
145   0.98417285  0.63925428   1.07755980  1.696468312
146   0.98417285 -0.06977776   0.79736223  1.434127852
147   0.51498716 -1.25149781   0.68528320  0.909446930
148   0.74958000 -0.06977776   0.79736223  1.040617161
150   0.04580146 -0.06977776   0.74132271  0.778276700
attr(,"scaled:center")
Sepal.Length  Sepal.Width Petal.Length  Petal.Width
    5.860952     3.029524     3.777143     1.206667
attr(,"scaled:scale")
Sepal.Length  Sepal.Width Petal.Length  Petal.Width
   0.8525409    0.4231121    1.7844552    0.7623681
> library(class)
```

#knn implementation (k values : 7,9,15,19)
library(class) #library for knn
library

knn_7 <- knn(train=train_scale,test=test_scale,cl=train_cl$Species,k=7)
knn_7

knn_9 <- knn(train=train_scale,test=test_scale,cl=train_cl$Species,k=9)
knn_9

knn_15 <- knn(train=train_scale,test=test_scale,cl=train_cl$Species,k=15)
knn_15

knn_19 <- knn(train=train_scale,test=test_scale,cl=train_cl$Species,k=19)
knn_19

```
> library(class)
> knn_7 <- knn(train=train_scale,test=test_scale,cl=train_cl$Species,k=7)
> knn_7
 [1] setosa     setosa     setosa     setosa     setosa     setosa     setosa
 [8] setosa     setosa     setosa     setosa     setosa     setosa     setosa
[15] setosa     versicolor versicolor versicolor versicolor versicolor versicolor
[22] versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[29] versicolor versicolor virginica  versicolor virginica  virginica  virginica
[36] virginica  virginica  virginica  virginica  virginica  virginica  virginica
[43] virginica  virginica  virginica
Levels: setosa versicolor virginica
>
> knn_9 <- knn(train=train_scale,test=test_scale,cl=train_cl$Species,k=9)
> knn_9
 [1] setosa     setosa     setosa     setosa     setosa     setosa     setosa
 [8] setosa     setosa     setosa     setosa     setosa     setosa     setosa
[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  virginica  virginica  virginica  virginica
[43] virginica  virginica  virginica
Levels: setosa versicolor virginica
>
```

```
> knn_15 <- knn(train=train_scale,test=test_scale,cl=train_cl$Species,k=15)
> knn_15
 [1] setosa     setosa     setosa     setosa     setosa     setosa     setosa
 [8] setosa     setosa     setosa     setosa     setosa     setosa     setosa
[15] setosa     versicolor versicolor versicolor versicolor versicolor versicolor
[22] versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[29] versicolor versicolor virginica  versicolor virginica  virginica  virginica
[36] virginica  virginica  virginica  virginica  virginica  virginica  virginica
[43] virginica  virginica  virginica
Levels: setosa versicolor virginica
>
> knn_19 <- knn(train=train_scale,test=test_scale,cl=train_cl$Species,k=19)
> knn_19
 [1] setosa     setosa     setosa     setosa     setosa     setosa     setosa
 [8] setosa     setosa     setosa     setosa     setosa     setosa     setosa
[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
>
```

B) Build the confusion matrix and calculate the accuracy for all ‘k’ values.

#confusion matrix

library(caret) #used for confusion matrix

cm1 <- table(test_cl$Species,knn_7)

cm1

confusionMatrix(cm1)

```
> confusionMatrix(cm1)
Confusion Matrix and Statistics

            knn_7
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          1        14

Overall Statistics

               Accuracy : 0.9778
                 95% CI : (0.8823, 0.9994)
    No Information Rate : 0.3556
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9667

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: setosa Class: versicolor Class: virginica
Sensitivity                 1.0000            0.9375           1.0000
Specificity                 1.0000            1.0000           0.9677
Pos Pred Value              1.0000            1.0000           0.9333
Neg Pred Value              1.0000            0.9667           1.0000
Prevalence                  0.3333            0.3556           0.3111
Detection Rate              0.3333            0.3333           0.3111
Detection Prevalence        0.3333            0.3333           0.3333
Balanced Accuracy           1.0000            0.9688           0.9839
> cm2 <- table(test_cl$Species,knn_9)
> cm2
            knn_9
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          1        14
```

cm2 <- table(test_cl$Species,knn_9)

cm2

confusionMatrix(cm2)

```
> cm2 <- table(test_cl$Species,knn_9)
> cm2
            knn_9
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          1        14
> confusionMatrix(cm2)
Confusion Matrix and Statistics

            knn_9
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          1        14

Overall Statistics

               Accuracy : 0.9778
                 95% CI : (0.8823, 0.9994)
    No Information Rate : 0.3556
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9667

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: setosa Class: versicolor Class: virginica
Sensitivity                 1.0000            0.9375           1.0000
Specificity                 1.0000            1.0000           0.9677
Pos Pred Value              1.0000            1.0000           0.9333
Neg Pred Value              1.0000            0.9667           1.0000
Prevalence                  0.3333            0.3556           0.3111
Detection Rate              0.3333            0.3333           0.3111
Detection Prevalence        0.3333            0.3333           0.3333
Balanced Accuracy           1.0000            0.9688           0.9839
```

cm3 <- table(test_cl$Species,knn_15)

cm3

confusionMatrix(cm3)

```
> cm3 <- table(test_cl$Species,knn_15)
> cm3
            knn_15
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          1        14
> confusionMatrix(cm3)
Confusion Matrix and Statistics

            knn_15
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          1        14

Overall Statistics

               Accuracy : 0.9778
                 95% CI : (0.8823, 0.9994)
    No Information Rate : 0.3556
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9667

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: setosa Class: versicolor Class: virginica
Sensitivity                 1.0000            0.9375           1.0000
Specificity                 1.0000            1.0000           0.9677
Pos Pred Value              1.0000            1.0000           0.9333
Neg Pred Value              1.0000            0.9667           1.0000
Prevalence                  0.3333            0.3556           0.3111
Detection Rate              0.3333            0.3333           0.3111
Detection Prevalence        0.3333            0.3333           0.3333
Balanced Accuracy           1.0000            0.9688           0.9839
```

cm4 <- table(test_cl$Species,knn_19)

cm4

confusionMatrix(cm4)

```
> cm4 <- table(test_cl$Species,knn_19)
> cm4
            knn_19
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          2        13
> confusionMatrix(cm4)
Confusion Matrix and Statistics

            knn_19
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          2        13

Overall Statistics

               Accuracy : 0.9556
                 95% CI : (0.8485, 0.9946)
    No Information Rate : 0.3778
    P-Value [Acc > NIR] : 2.61e-16

                  Kappa : 0.9333

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: setosa Class: versicolor Class: virginica
Sensitivity                 1.0000            0.8824           1.0000
Specificity                 1.0000            1.0000           0.9375
Pos Pred Value              1.0000            1.0000           0.8667
Neg Pred Value              1.0000            0.9333           1.0000
Prevalence                  0.3333            0.3778           0.2889
Detection Rate              0.3333            0.3333           0.2889
Detection Prevalence        0.3333            0.3333           0.3333
Balanced Accuracy           1.0000            0.9412           0.9688
> |
```

#accuracy for all k values

Error_1 <- mean(knn_7!=test_cl$Species)

print(paste("Accuracy of k=7 is",1-Error_1))

Error_2 <- mean(knn_9!=test_cl$Species)

print(paste("Accuracy of k=9 is",1-Error_2))

Error_3 <- mean(knn_15!=test_cl$Species)

print(paste("Accuracy of k=15 is",1-Error_3))

Error_4 <- mean(knn_19!=test_cl$Species)

print(paste("Accuracy of k=19 is",1-Error_4))

```
> Error_1 <- mean(knn_7!=test_cl$Species)
> print(paste("Accuracy of k=7 is",1-Error_1))
[1] "Accuracy of k=7 is 0.977777777777778"
> Error_2 <- mean(knn_9!=test_cl$Species)
> print(paste("Accuracy of k=9 is",1-Error_2))
[1] "Accuracy of k=9 is 0.977777777777778"
> Error_3 <- mean(knn_15!=test_cl$Species)
> print(paste("Accuracy of k=15 is",1-Error_3))
[1] "Accuracy of k=15 is 0.977777777777778"
> Error_4 <- mean(knn_19!=test_cl$Species)
> print(paste("Accuracy of k=19 is",1-Error_4))
[1] "Accuracy of k=19 is 0.955555555555556"
>
```