Assignment 8

Title: Supervised learning - Classification

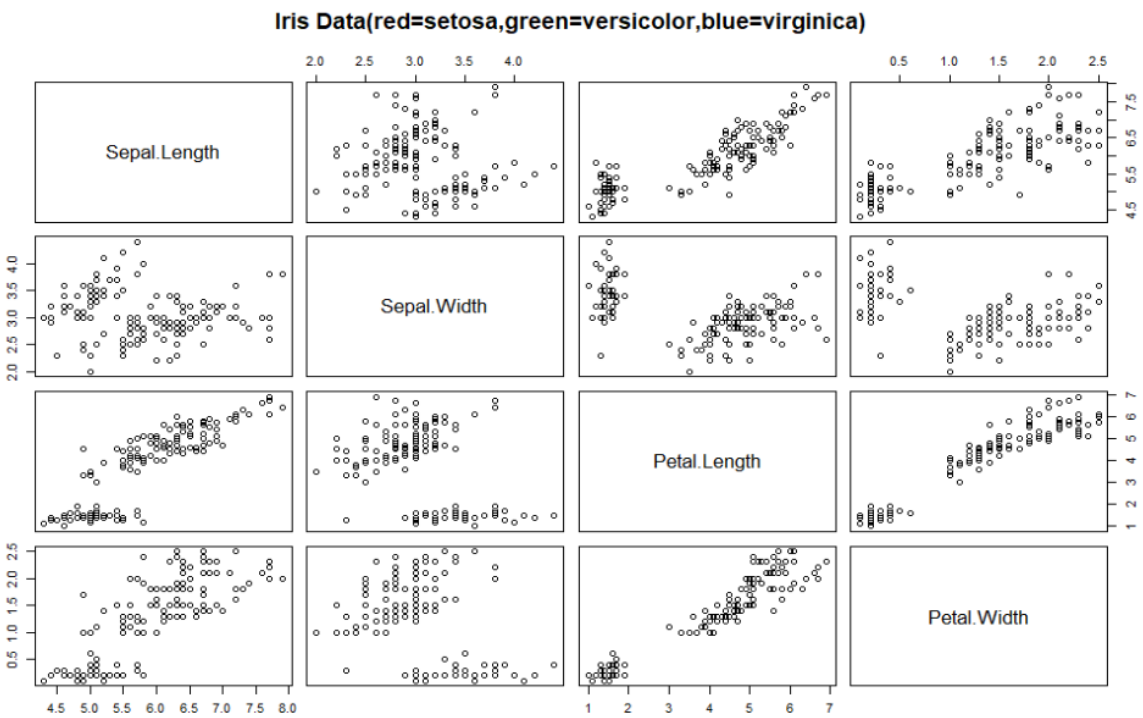
Questions:

1. **Implementation and analysis of Classification algorithms like:ID3 , C4.5 using Fish.csv dataset**
    - **Visualize the output**
    - **predict the test data**
    - **Verify the result**

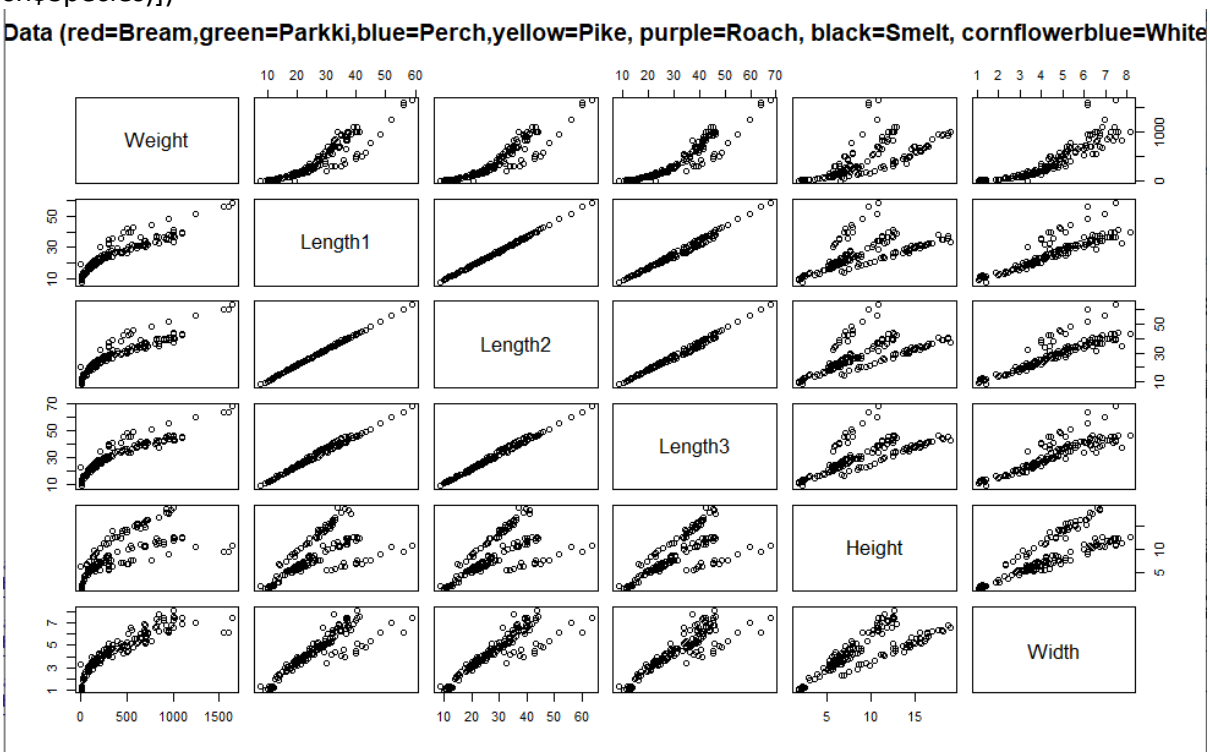
Code:

a) ID3 decision tree classification

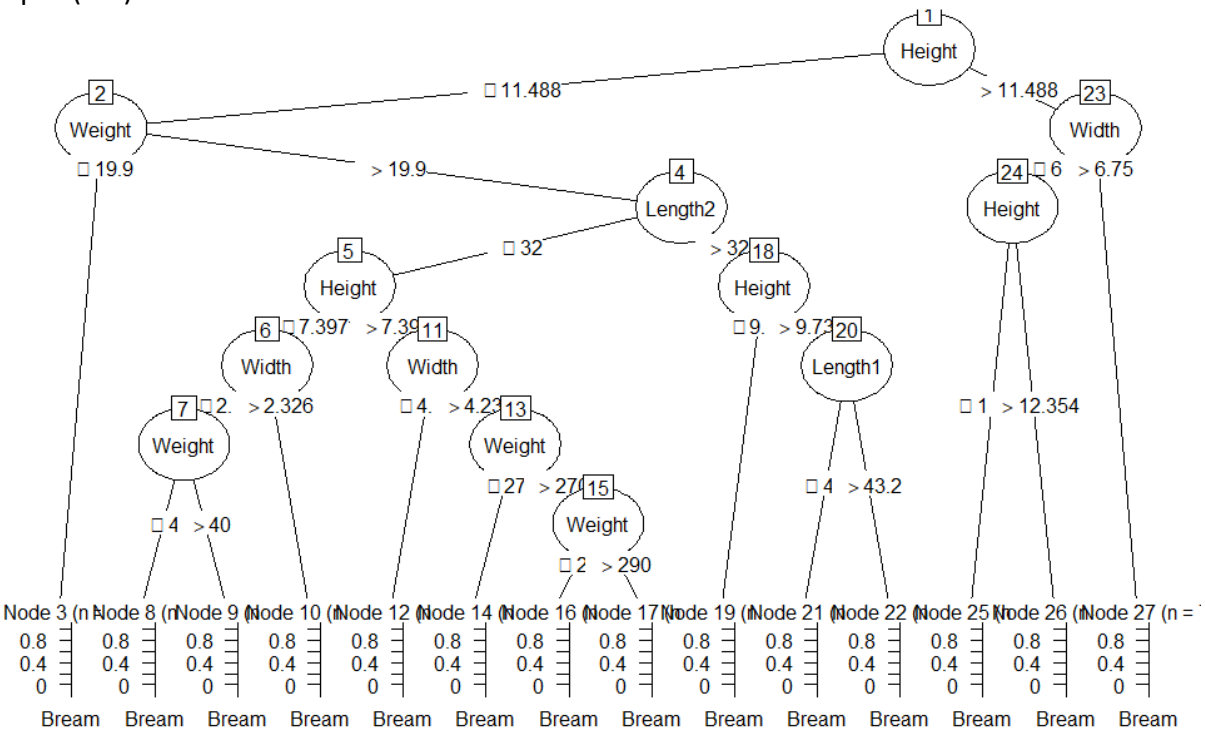
```
> library(RWeka)

> library(party)

> library(caret)

>mydata=read.csv('iris.csv')

>
```



```
> cores = c('blue','green','red')

> val = c('setosa','setosa','virginica','versicolor','virginica','setosa')

> val_fac=factor(val)

> unclass(val)

[1] "setosa"     "setosa"     "virginica"  "versicolor" "virginica"  "setosa"

> unclass(val_fac)

[1] 1 1 3 2 3 1

attr(,"levels")

[1] "setosa"     "versicolor" "virginica"

> cores[unclass(val_fac)]

[1] "blue"  "blue"  "red"   "green" "red"   "blue"

> myfish=read.csv('Fish.csv')

> pairs(myfish[1:6], main="Fish Data (red=Bream,green=Parkki,blue=Perch,yellow=Pike,
purple=Roach, black=Smelt, cornflowerblue=Whitefish )",
pch=21,bg=c("red","green","blue","yellow","purple","black","cornflowerblue")[unclass(myfi
sh$Species)])
```



```
> myfish$Species=as.factor(myfish$Species)
> myfish$Species=as.factor(myfish$Species)
> myfish=myfish[sample(nrow(myfish),),]
> head(myfish)
    Weight Length1 Length2 Length3  Height  Width Species
117    900    36.5    39.0    41.4 11.1366 7.4934   Perch
79      78    16.8    18.7    19.4  5.1992 3.1234   Perch
10     500    28.5    30.7    36.2 14.2266 4.9594   Bream
129    200    30.0    32.3    34.8  5.5680 3.3756    Pike
26     725    31.8    35.0    40.9 16.3600 6.0532   Bream
124   1100    39.0    42.0    44.6 12.8002 6.8684   Perch

> TrainData=myfish[1:112,]
> TestData=myfish[113:159,]
> m1=J48(Species~.,data=TrainData)
> plot(m1)
```




```
> summary(m1)
```

```
=== Summary ===

Correctly Classified Instances         100               89.2857 %
Incorrectly Classified Instances        12               10.7143 %
Kappa statistic                          0.8631
Mean absolute error                      0.0435
Root mean squared error                  0.1476
Relative absolute error                 19.0604 %
Root relative squared error             43.7508 %
Total Number of Instances              112

=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 25  0  0  0  0  0  1 |  a = Bream
  0  8  0  0  0  0  0 |  b = Parkki
  0  0 36  0  0  1  0 |  c = Perch
  0  0  0 13  0  0  0 |  d = Pike
  0  0  9  0  2  0  1 |  e = Roach
  0  0  0  0  0 10  0 |  f = Smelt
  0  0  0  0  0  0  6 |  g = Whitefish
```

```
> fishpred = predict(m1,TestData)
> df=data.frame(fishpred,TestData$Species)
> df
```

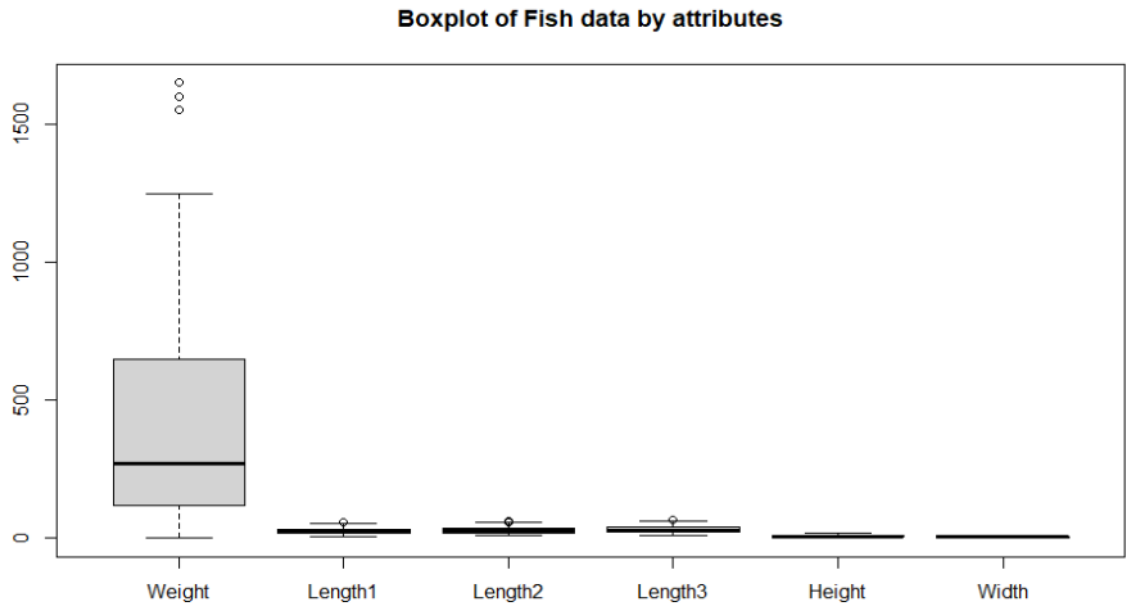
	fishpred	TestData.Species
1	Bream	Bream
2	Pike	Pike
3	Perch	Perch
4	Perch	Roach
5	Whitefish	Perch
6	Bream	Bream
7	Smelt	Roach
8	Parkki	Roach
9	Perch	Perch
10	Perch	Perch
11	Smelt	Smelt
12	Bream	Bream
13	Pike	Pike
14	Smelt	Smelt
15	Perch	Perch
16	Bream	Bream
17	Perch	Roach
18	Whitefish	Perch
19	Parkki	Parkki
20	Perch	Perch
21	Perch	Roach
22	Perch	Perch
23	Bream	Bream
24	Parkki	Perch
25	Perch	Perch
26	Smelt	Smelt
27	Perch	Perch
28	Bream	Bream
29	Perch	Roach
30	Perch	Perch
31	Bream	Bream
32	Perch	Perch
33	Pike	Pike
34	Pike	Pike
35	Perch	Roach
36	Perch	Roach

b) C5.0 decision tree classification

```
> library(C50)
> fs=read.csv('Fish.csv')
> summary(fs)
```

```
     Weight          Length1         Length2         Length3          Height
 Min.   :   0.0   Min.   : 7.50   Min.   : 8.40   Min.   : 8.80   Min.   : 1.728
 1st Qu.: 120.0   1st Qu.:19.05   1st Qu.:21.00   1st Qu.:23.15   1st Qu.: 5.945
 Median : 273.0   Median :25.20   Median :27.30   Median :29.40   Median : 7.786
 Mean   : 398.3   Mean   :26.25   Mean   :28.42   Mean   :31.23   Mean   : 8.971
 3rd Qu.: 650.0   3rd Qu.:32.70   3rd Qu.:35.50   3rd Qu.:39.65   3rd Qu.:12.366
 Max.   :1650.0   Max.   :59.00   Max.   :63.40   Max.   :68.00   Max.   :18.957

     Width           Species
 Min.   :1.048   Bream    :35
 1st Qu.:3.386   Parkki   :11
 Median :4.248   Perch    :56
 Mean   :4.417   Pike     :17
 3rd Qu.:5.585   Roach    :20
 Max.   :8.142   Smelt    :14
                 Whitefish: 6
```

> boxplot(fs[,-7], main = 'Boxplot of Fish data by attributes')



> fs$Species=as.factor(fs$Species)
> fs=fs[sample(nrow(fs),),]
> head(fs)

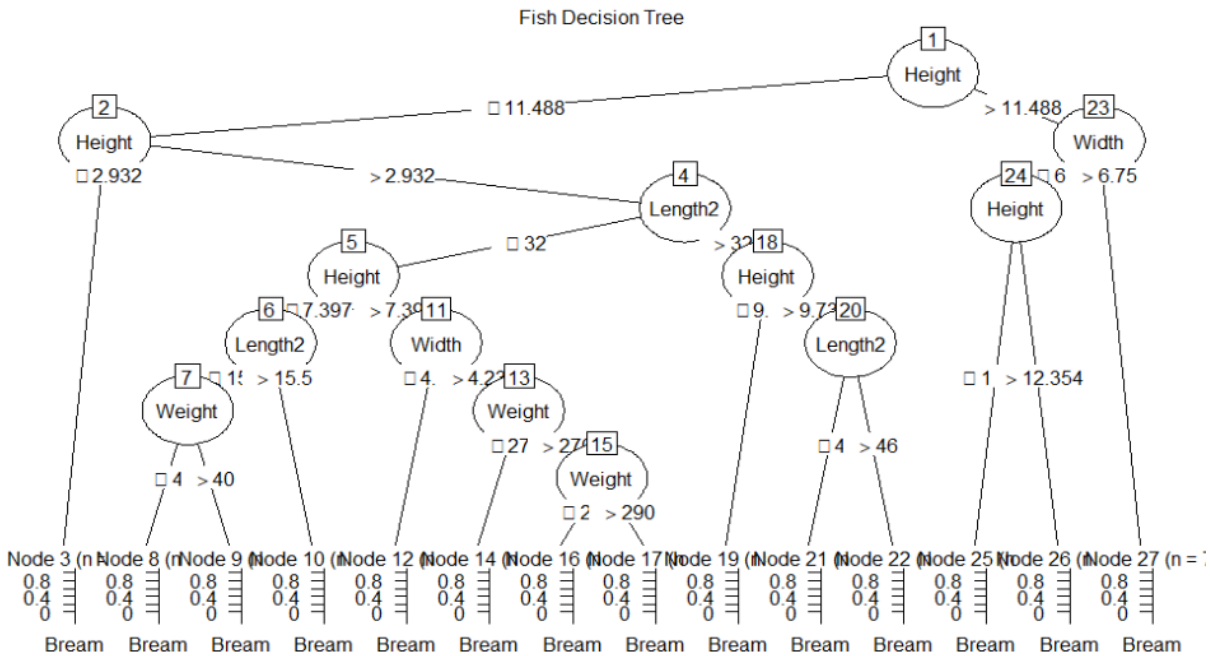
	Weight	Length1	Length2	Length3	Height	Width	Species
81	85	17.8	19.6	20.8	5.1376	3.0368	Perch
1	242	23.2	25.4	30.0	11.5200	4.0200	Bream
8	390	27.6	30.0	35.0	12.6700	4.6900	Bream
26	725	31.8	35.0	40.9	16.3600	6.0532	Bream
21	575	31.3	34.0	39.5	15.1285	5.5695	Bream
119	820	36.6	39.0	41.3	12.4313	7.3514	Perch

> Traindata=fs[1:112,]
> TestData=fs[113:159,]
> fstree=C5.0(TrainData[,-7],TrainData[,7])
> summary(fstree)

```
Call:
C5.0.default(x = TrainData[, -7], y = TrainData[, 7])


C5.0 [Release 2.07 GPL Edition]         Fri Feb 10 22:04:55 2023
-------------------------------

Class specified by attribute `outcome'

Read 112 cases (7 attributes) from undefined.data

Decision tree:

Height > 11.4884:
:...Width > 6.7497: Perch (7)
:   Width <= 6.7497:
:   :...Height <= 12.354: Whitefish (3/1)
:       Height > 12.354: Bream (25)
Height <= 11.4884:
:...Height <= 2.9322: Smelt (11/1)
    Height > 2.9322:
    :...Length2 > 32:
        :...Height <= 9.7364: Pike (11)
        :   Height > 9.7364:
        :   :...Length2 <= 46: Perch (7)
        :       Length2 > 46: Pike (2)
        Length2 <= 32:
        :...Height <= 7.3968:
            :...Length2 > 15.5: Perch (28/8)
            :   Length2 <= 15.5:
            :   :...Weight <= 40: Perch (3/1)
            :       Weight > 40: Parkki (2)
            Height > 7.3968:
            :...Width <= 4.239: Parkki (6)
                Width > 4.239:
```

```
:...Weight <= 270: Whitefish (2)
    Weight > 270:
    :...Weight <= 290: Roach (2)
        Weight > 290: Whitefish (3/1)

Evaluation on training data (112 cases):

          Decision Tree
        ----------------
        Size      Errors

          14   12(10.7%)   <<


         (a)   (b)   (c)   (d)   (e)   (f)   (g)    <-classified as
        ----  ----  ----  ----  ----  ----  ----
          25                                  1    (a): class Bream
                 8                                 (b): class Parkki
                      36                    1      (c): class Perch
                            13                      (d): class Pike
                       9           2          1    (e): class Roach
                                             10      (f): class Smelt
                                                 6   (g): class Whitefish


        Attribute usage:

        100.00% Height
         58.93% Length2
         42.86% Width
         10.71% Weight


Time: 0.0 secs
```

> plot(fstree,main="Fish Decision Tree")



```
> fsrules=C5.0(TrainData[,-7],TrainData[,7],rules=TRUE)
> pred=predict(fsrules, newdata = TestData)
> df = data.frame(TestData$Species,pred)
> df
   TestData.Species      pred
1             Perch     Perch
2              Pike      Pike
3             Smelt     Smelt
4             Perch     Perch
5             Bream     Bream
6             Bream     Bream
7             Perch     Perch
8             Perch     Perch
9             Perch     Perch
10        Whitefish Whitefish
11            Roach     Perch
12            Bream     Bream
13            Bream     Bream
14            Bream     Bream
15           Parkki    Parkki
16            Perch     Perch
17            Perch     Perch
18             Pike      Pike
19             Pike      Pike
20            Smelt     Smelt
21            Perch     Perch
22            Perch     Smelt
23            Perch     Perch

> confusion=table(TestData$Species,pred)
> confusion
            pred
             Bream Parkki Perch Pike Roach Smelt Whitefish
  Bream          9      0     0    0     0     0          0
  Parkki         0      3     0    0     0     0          0
  Perch          0      0    16    0     0     1          1
  Pike           0      0     0    8     0     0          0
  Roach          0      0     2    0     0     0          0
  Smelt          0      0     0    0     0     4          0
  Whitefish      0      0     0    0     0     0          3

> confusionMatrix(confusion)
Confusion Matrix and Statistics

          pred
           Bream Parkki Perch Pike Roach Smelt Whitefish
  Bream        9      0     0    0     0     0          0
  Parkki       0      3     0    0     0     0          0
  Perch        0      0    16    0     0     1          1
  Pike         0      0     0    8     0     0          0
  Roach        0      0     2    0     0     0          0
  Smelt        0      0     0    0     0     4          0
  Whitefish    0      0     0    0     0     0          3

Overall Statistics

               Accuracy : 0.9149
                 95% CI : (0.7962, 0.9763)
    No Information Rate : 0.383
    P-Value [Acc > NIR] : 3.265e-14

                  Kappa : 0.8893

 Mcnemar's Test P-Value : NA
```

```
Statistics by Class:

                     Class: Bream Class: Parkki Class: Perch Class: Pike Class: Roach
Sensitivity                1.0000       1.00000       0.8889      1.0000           NA
Specificity                1.0000       1.00000       0.9310      1.0000      0.95745
Pos Pred Value             1.0000       1.00000       0.8889      1.0000           NA
Neg Pred Value             1.0000       1.00000       0.9310      1.0000           NA
Prevalence                 0.1915       0.06383       0.3830      0.1702      0.00000
Detection Rate             0.1915       0.06383       0.3404      0.1702      0.00000
Detection Prevalence       0.1915       0.06383       0.3830      0.1702      0.04255
Balanced Accuracy          1.0000       1.00000       0.9100      1.0000           NA
                     Class: Smelt Class: Whitefish
Sensitivity               0.80000          0.75000
Specificity               1.00000          1.00000
Pos Pred Value            1.00000          1.00000
Neg Pred Value            0.97674          0.97727
Prevalence                0.10638          0.08511
Detection Rate            0.08511          0.06383
Detection Prevalence      0.08511          0.06383
Balanced Accuracy         0.90000          0.87500
> |
```

2. Implementation and analysis of Classification algorithm like: Naïve Bayesian, K-nearest neighbour using fish.csv dataset.

- **Visualize the output**
- **Predict the test data**
- **Verify the result**

Code:

a)Performing Nave Bayes Classification

```
> library(e1071)
> library(caTools)
> library(caret)
> fish=read.csv("Fish.csv")
> fish=fish[sample(nrow(fish),),]
> head(fish)
     Weight Length1 Length2 Length3  Height  Width Species
118  650.0    36.5    39.0    41.4 11.1366 6.0030   Perch
47   140.0    21.0    22.5    25.0  6.5500 3.3250   Roach
67   140.0    19.0    20.7    23.2  8.5376 3.2944  Parkki
22   685.0    31.4    34.0    39.2 15.9936 5.3704   Bream
154    9.8    11.4    12.0    13.2  2.2044 1.1484   Smelt
72   300.0    24.0    26.0    29.0 11.3680 4.2340  Parkki

> split=sample.split(fish,SplitRatio = 0.7)
> TrainData=subset(fish,split=="TRUE")
> TestData=subset(fish,split="FALSE")
> classifier_c1=naiveBayes(Species~.,data=TrainData)
> classifier_c1
```

```
Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
     Bream     Parkki      Perch       Pike      Roach      Smelt  Whitefish
0.18681319 0.07692308 0.35164835 0.10989011 0.15384615 0.06593407 0.05494505

Conditional probabilities:
           Weight
Y                [,1]       [,2]
  Bream     553.17647 182.907502
  Parkki    129.28571  45.132081
  Perch     421.03125 368.479043
  Pike      634.80000 458.395705
  Roach     144.42857  61.342311
  Smelt      12.86667   5.856506
  Whitefish 477.20000 313.223882

           Length1
Y                [,1]     [,2]
  Bream     29.29412 3.074587
  Parkki    18.07143 2.288897
  Perch     26.69687 9.081121
  Pike      40.85000 8.746333
  Roach     20.07143 2.963737
  Smelt     11.55000 1.751285
  Whitefish 27.82000 5.632673


           Length2
Y                [,1]     [,2]
  Bream     31.97647 3.358670
  Parkki    19.62857 2.493802
  Perch     28.93125 9.563993
  Pike      43.73000 9.391728
  Roach     21.65714 3.105755
  Smelt     12.30000 2.051341
  Whitefish 30.30000 5.761944


           Length2
Y                [,1]     [,2]
  Bream     31.97647 3.358670
  Parkki    19.62857 2.493802
  Perch     28.93125 9.563993
  Pike      43.73000 9.391728
  Roach     21.65714 3.105755
  Smelt     12.30000 2.051341
  Whitefish 30.30000 5.761944
```

```
           Length3
Y               [,1]      [,2]
  Bream     37.17059  3.690489
  Parkki    22.00000  2.797022
  Perch     30.70000 10.084770
  Pike      46.79000  9.813647
  Roach     24.33571  3.453116
  Smelt     13.40000  2.058155
  Whitefish 33.26000  6.081365

           Height
Y               [,1]      [,2]
  Bream     14.404859 1.7265524
  Parkki     8.555314 1.2895295
  Perch      8.122075 2.9718284
  Pike       7.576330 1.6858995
  Roach      6.513100 1.1398280
  Smelt      2.359950 0.4802198
  Whitefish  9.680360 1.8124523


           Width
Y               [,1]      [,2]
  Bream     5.191324 0.6501576
  Parkki    3.044743 0.4485892
  Perch     4.906678 1.8109504
  Pike      4.882590 1.1624993
  Roach     3.521893 0.5721338
  Smelt     1.451133 0.4175045
  Whitefish 5.252940 1.1914080
```

> y_pred=predict(classifier_c1,newdata = TestData)
> y_pred

```
  [1] Perch     Roach     Parkki    Bream     Smelt     Perch     Smelt     Roach
  [9] Bream     Bream     Pike      Perch     Roach     Smelt     Parkki    Bream
 [17] Bream     Roach     Perch     Smelt     Parkki    Perch     Roach     Roach
 [25] Roach     Perch     Perch     Perch     Bream     Perch     Perch     Bream
 [33] Roach     Perch     Bream     Roach     Perch     Pike      Perch     Smelt
 [41] Roach     Bream     Roach     Smelt     Bream     Perch     Perch     Smelt
 [49] Perch     Bream     Parkki    Perch     Roach     Smelt     Bream     Bream
 [57] Bream     Pike      Perch     Roach     Smelt     Whitefish Perch     Perch
 [65] Perch     Smelt     Roach     Pike      Whitefish Bream     Pike      Roach
 [73] Roach     Parkki    Bream     Bream     Perch     Pike      Roach     Perch
 [81] Smelt     Bream     Pike      Perch     Roach     Whitefish Roach     Bream
 [89] Perch     Pike      Roach     Perch     Perch     Perch     Perch     Perch
 [97] Perch     Parkki    Roach     Roach     Bream     Bream     Perch     Roach
[105] Roach     Pike      Roach     Roach     Bream     Roach     Roach     Perch
[113] Roach     Whitefish Pike      Roach     Roach     Bream     Roach     Smelt
[121] Smelt     Perch     Bream     Bream     Roach     Parkki    Roach     Bream
[129] Roach     Smelt     Whitefish Roach     Bream     Pike      Perch     Roach
[137] Perch     Bream     Pike      Bream     Perch     Bream     Roach     Bream
[145] Pike      Roach     Roach     Whitefish Bream     Smelt     Roach     Bream
[153] Smelt     Perch     Parkki    Roach     Perch     Roach     Parkki
Levels: Bream Parkki Perch Pike Roach Smelt Whitefish
```

> cm=table(TestData$Species,y_pred)
> cm

```
          y_pred
           Bream Parkki Perch Pike Roach Smelt Whitefish
  Bream       30      0     5    0     0     0          0
  Parkki       0      9     2    0     0     0          0
  Perch        2      0    22    0    27     2          3
  Pike         0      0     4   13     0     0          0
  Roach        0      0     3    0    16     0          1
  Smelt        0      0     0    0     0    14          0
  Whitefish    1      0     3    0     0     0          2
```

> confusionMatrix(cm)

```
Confusion Matrix and Statistics

          y_pred
           Bream Parkki Perch Pike Roach Smelt Whitefish
  Bream       30      0     5    0     0     0          0
  Parkki       0      9     2    0     0     0          0
  Perch        2      0    22    0    27     2          3
  Pike         0      0     4   13     0     0          0
  Roach        0      0     3    0    16     0          1
  Smelt        0      0     0    0     0    14          0
  Whitefish    1      0     3    0     0     0          2

Overall Statistics

               Accuracy : 0.6667
                 95% CI : (0.5877, 0.7393)
    No Information Rate : 0.2704
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.589

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: Bream Class: Parkki Class: Perch Class: Pike Class: Roach
Sensitivity                0.9091       1.00000       0.5641      1.00000       0.3721
Specificity                0.9603       0.98667       0.7167      0.97260       0.9655
Pos Pred Value             0.8571       0.81818       0.3929      0.76471       0.8000
Neg Pred Value             0.9758       1.00000       0.8350      1.00000       0.8058
Prevalence                 0.2075       0.05660       0.2453      0.08176       0.2704
Detection Rate             0.1887       0.05660       0.1384      0.08176       0.1006
Detection Prevalence       0.2201       0.06918       0.3522      0.10692       0.1258
Balanced Accuracy          0.9347       0.99333       0.6404      0.98630       0.6688
                     Class: Smelt Class: Whitefish
Sensitivity               0.87500          0.33333
Specificity               1.00000          0.97386
Pos Pred Value            1.00000          0.33333
Neg Pred Value            0.98621          0.97386
Prevalence                0.10063          0.03774
Detection Rate            0.08805          0.01258
Detection Prevalence      0.08805          0.03774
Balanced Accuracy         0.93750          0.65359
```

b) Performing K-Nearest Neighbour on Fish Dataset

```
> library(e1071)
> library(caTools)
> library(class)
> fish=read.csv('Fish.csv')
> myfish=fish[sample(nrow(fish),),]
> head(myfish)
    Weight Length1 Length2 Length3  Height  Width   Species
10     500    28.5    30.7    36.2 14.2266 4.9594     Bream
40     120    18.6    20.0    22.2  6.2160 3.5742     Roach
77      70    15.7    17.4    18.5  4.5880 2.9415     Perch
68     170    19.0    20.7    23.2  9.3960 3.4104    Parkki
61    1000    37.3    40.0    43.5 12.3540 6.5250 Whitefish
56     270    23.6    26.0    28.7  8.3804 4.2476 Whitefish

> split=sample.split(myfish,SplitRatio = 0.7)
> TrainData=subset(myfish,split=="TRUE")
> TestData=subset(myfish,split="FALSE")
> train_scale=scale(TrainData[,1:6])
> test_scale=scale(TestData[,1:6])
> classifier_knn=knn(train=train_scale,test=test_scale,cl=TrainData$Species,k=1)
> classifier_knn
  [1] Bream     Roach     Perch     Parkki    Perch     Whitefish Bream     Perch
  [9] Pike      Perch     Roach     Bream     Smelt     Pike      Whitefish Bream
 [17] Roach     Roach     Bream     Parkki    Roach     Parkki    Perch     Parkki
 [25] Smelt     Pike      Bream     Bream     Bream     Bream     Perch     Roach
 [33] Perch     Perch     Perch     Smelt     Perch     Perch     Perch     Perch
 [41] Parkki    Roach     Bream     Perch     Whitefish Bream     Roach     Perch
 [49] Roach     Perch     Bream     Bream     Smelt     Perch     Roach     Perch
 [57] Pike      Perch     Bream     Pike      Perch     Roach     Bream     Bream
 [65] Bream     Perch     Parkki    Pike      Roach     Perch     Perch     Perch
 [73] Bream     Bream     Perch     Pike      Bream     Roach     Whitefish Perch
 [81] Bream     Bream     Bream     Perch     Pike      Smelt     Bream     Pike
 [89] Parkki    Perch     Bream     Bream     Perch     Roach     Bream     Perch
 [97] Bream     Perch     Whitefish Perch     Whitefish Perch     Perch     Pike
[105] Parkki    Pike      Smelt     Perch     Perch     Bream     Perch     Pike
[113] Smelt     Parkki    Parkki    Bream     Perch     Roach     Perch     Perch
[121] Perch     Perch     Perch     Perch     Perch     Bream     Parkki    Bream
[129] Whitefish Perch     Smelt     Perch     Perch     Smelt     Smelt     Roach
[137] Bream     Perch     Perch     Smelt     Smelt     Perch     Perch     Perch
[145] Parkki    Perch     Smelt     Perch     Perch     Bream     Perch     Perch
[153] Perch     Smelt     Pike      Perch     Perch     Smelt     Roach
Levels: Bream Parkki Perch Pike Roach Smelt Whitefish

> cm=table(TestData$Species,classifier_knn)
            classifier_knn
             Bream Parkki Perch Pike Roach Smelt Whitefish
  Bream         34      1     0    0     0     0          0
  Parkki         0     11     0    0     0     0          0
  Perch          0      0    53    0     1     1          1
  Pike           0      0     4   13     0     0          0
  Roach          0      0     2    0    16     0          2
  Smelt          0      0     0    0     0    14          0
  Whitefish      0      0     2    0     0     0          4
```

#For k=1

```
> misClassError=mean(classifier_knn!=TestData$Species)
> print(paste('Accuracy=',1-misClassError))
[1] "Accuracy= 0.911949685534591"
```

#For k=3

```
> classifier_knn=knn(train=train_scale,test=test_scale,cl=TrainData$Species,k=3)
> misClassError=mean(classifier_knn!=TestData$Species)
> print(paste('Accuracy=',1-misClassError))
[1] "Accuracy= 0.786163522012579"
```

#For k=7

```
> classifier_knn=knn(train=train_scale,test=test_scale,cl=TrainData$Species,k=7)
> misClassError=mean(classifier_knn!=TestData$Species)
> print(paste('Accuracy=',1-misClassError))
[1] "Accuracy= 0.729559748427673"
```

#For k=15

```
> classifier_knn=knn(train=train_scale,test=test_scale,cl=TrainData$Species,k=15)
> misClassError=mean(classifier_knn!=TestData$Species)
> print(paste('Accuracy=',1-misClassError))
[1] "Accuracy= 0.635220125786164"
```

#For k=19

```
> classifier_knn=knn(train=train_scale,test=test_scale,cl=TrainData$Species,k=19)
> misClassError=mean(classifier_knn!=TestData$Species)
> print(paste('Accuracy=',1-misClassError))
[1] "Accuracy= 0.559748427672956"
```