Name: Shaikh Inamul Hasan

Roll No: 100

Lab 8: Supervised Learning – Classification

1. Implementation and analysis of Classification algorithms like:

ID3 , C4.5 using Fish.csv dataset - Visualize the output

- predict the test data

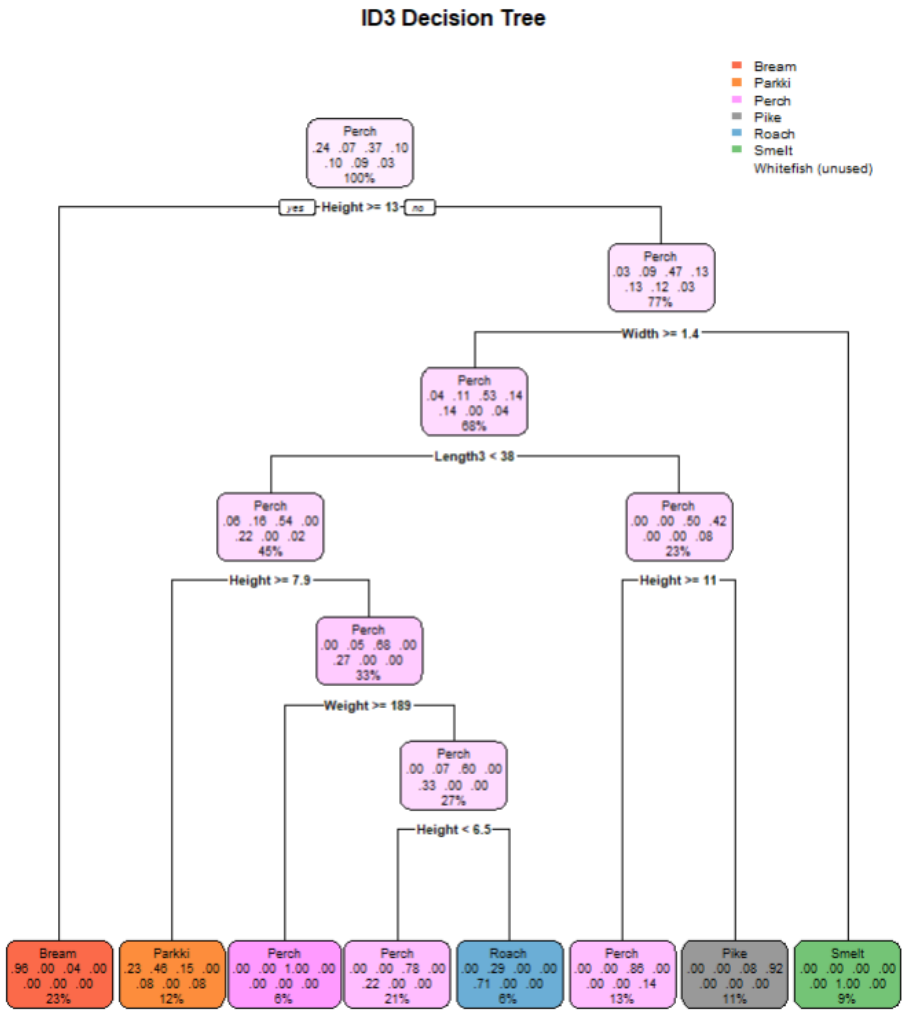
- Verify the result

Code & Output:

ID3:

```
> library(rpart)
> library(rpart.plot)
> fish_data <- read.csv("Fish.csv")
> sample_index <- sample(1:nrow(fish_data), 0.7 * nrow(fish_data))
> train_data <- fish_data[sample_index, ]
> test_data <- fish_data[-sample_index, ]
> id3_model <- rpart(Species ~ ., data = train_data, method = "class")
> rpart.plot(id3_model, main = "ID3 Decision Tree")
> predictions <- predict(id3_model, test_data, type = "class")
> predictions
     1      3      5      7     10     14     17     18     27     29     32     34     38     46     52
Parkki Parkki Parkki  Bream  Bream  Bream  Bream  Bream  Bream  Bream  Bream  Bream  Perch  Roach  Perch
    54     56     57     58     60     65     68     69     70     71     75     77     78     79     84
Parkki Parkki  Perch Parkki  Perch Parkki Parkki Parkki Parkki Parkki  Perch  Perch  Perch  Perch  Perch
    88     96    100    102    103    108    110    116    118    119    124    135    138    150    152
 Perch  Roach  Perch  Perch Parkki  Perch   Pike  Perch  Perch  Perch  Bream   Pike   Pike  Smelt  Smelt
   154    156    158
 Smelt  Smelt  Perch
Levels: Bream Parkki Perch Pike Roach Smelt Whitefish
> conf_matrix <- table(predictions, test_data$Species)
> print("Confusion Matrix:")
[1] "Confusion Matrix:"
> print(conf_matrix)

predictions Bream Parkki Perch Pike Roach Smelt Whitefish
  Bream         9      0     1    0     0     0         0
  Parkki        3      5     1    0     1     0         2
  Perch         0      0    12    0     2     1         2
  Pike          0      0     1    2     0     0         0
  Roach         0      0     1    0     1     0         0
  Smelt         0      0     0    0     0     4         0
  Whitefish     0      0     0    0     0     0         0
> accuracy <- sum(diag(conf_matrix)) / sum(conf_matrix)
> print(paste("Accuracy: ", round(accuracy, 4)))
[1] "Accuracy:  0.6875"
```

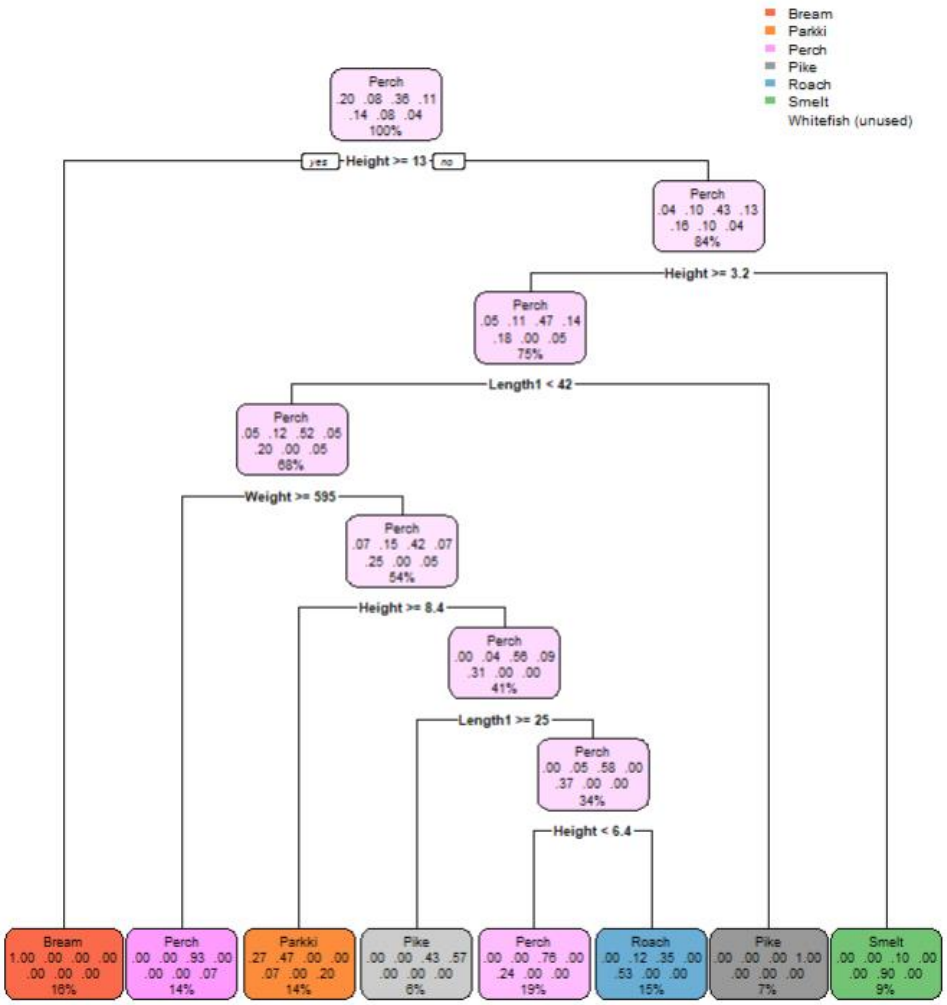


C4.5:

```
> library(C50)
> library(rpart.plot)
> fish_data <- read.csv("Fish.csv")
> train_indices <- sample(1:nrow(fish_data), 0.7 * nrow(fish_data))
> train_data$Species <- as.factor(train_data$Species)
> test_data$Species <- as.factor(test_data$Species)
> c45_model <- C5.0(Species ~ ., data = train_data)
> predictions <- predict(c45_model, test_data)
> confusion_matrix <- table(predictions, test_data$Species)
> print(confusion_matrix)

predictions Bream Parkki Perch Pike Roach Smelt Whitefish
  Bream        13      0     0    0     0     0         1
  Parkki        0      1     0    0     0     0         0
  Perch         0      0    15    0     3     0         1
  Pike          0      0     0    5     1     0         0
  Roach         0      1     1    0     0     0         0
  Smelt         0      0     0    0     0     5         0
  Whitefish     0      0     0    0     1     0         0
> c45_model <- rpart(Species ~ ., data = train_data, method = "class")
> rpart.plot(c45_model, main = "C4.5 Decision Tree")
> accuracy <- sum(diag(confusion_matrix)) / sum(confusion_matrix)
> print(paste("Accuracy:", round(accuracy, 2)))
[1] "Accuracy: 0.81"
```

C4.5 Decision Tree



2. Implementation and analysis of Classification algorithms like:

Naive Bayesian, K-Nearest Neighbour using Fish.csv dataset - Visualize the output

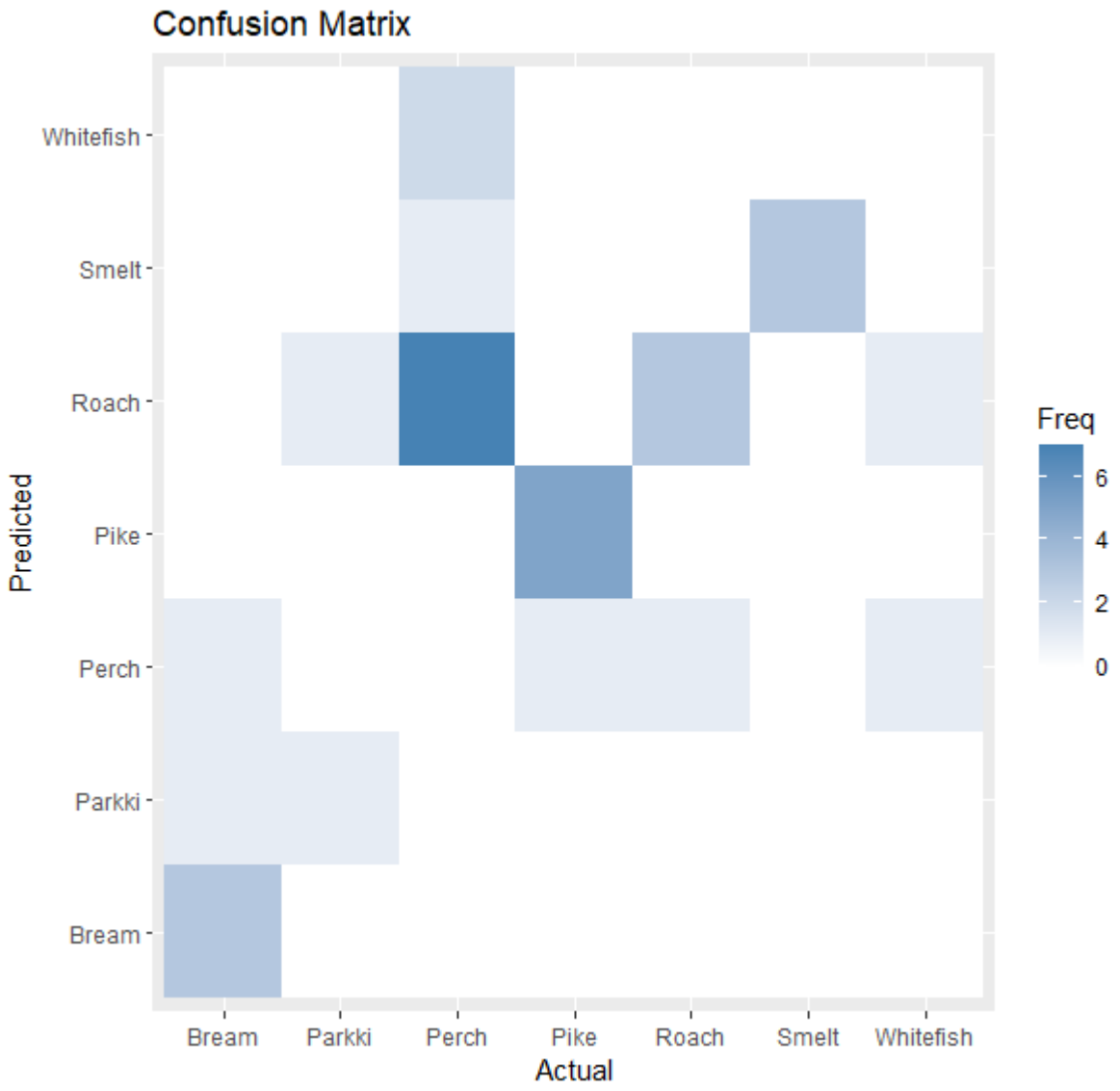
- Predict the test data

- Verify the result

Code & Output:

Naive Bayesian:

```
> library(e1071)
> library(ggplot2)
> fish_data <- read.csv("Fish.csv")
> str(fish_data)
'data.frame':   159 obs. of  7 variables:
 $ Species: chr  "Bream" "Bream" "Bream" "Bream" ...
 $ Weight : num  242 290 340 363 430 450 500 390 450 500 ...
 $ Length1: num  23.2 24 23.9 26.3 26.5 26.8 26.8 27.6 27.6 28.5 ...
 $ Length2: num  25.4 26.3 26.5 29 29 29.7 29.7 30 30 30.7 ...
 $ Length3: num  30 31.2 31.1 33.5 34 34.7 34.5 35 35.1 36.2 ...
 $ Height : num  11.5 12.5 12.4 12.7 12.4 ...
 $ Width  : num  4.02 4.31 4.7 4.46 5.13 ...
> head(fish_data)
  Species Weight Length1 Length2 Length3  Height  Width
1   Bream    242    23.2    25.4    30.0 11.5200 4.0200
2   Bream    290    24.0    26.3    31.2 12.4800 4.3056
3   Bream    340    23.9    26.5    31.1 12.3778 4.6961
4   Bream    363    26.3    29.0    33.5 12.7300 4.4555
5   Bream    430    26.5    29.0    34.0 12.4440 5.1340
6   Bream    450    26.8    29.7    34.7 13.6024 4.9274
> split_index <- sample(1:nrow(fish_data), 0.8 * nrow(fish_data))
> train_data <- fish_data[split_index, ]
> test_data <- fish_data[-split_index, ]
> nb_model <- naiveBayes(Species ~ ., data = train_data)
> predictions <- predict(nb_model, test_data)
> conf_matrix_df <- as.data.frame(as.table(conf_matrix))
> print(conf_matrix_df)
   predictions      Var2 Freq
1        Bream     Bream    3
2       Parkki     Bream    1
3        Perch     Bream    1
4         Pike     Bream    0
5        Roach     Bream    0
6        Smelt     Bream    0
7    Whitefish     Bream    0
8        Bream    Parkki    0
9       Parkki    Parkki    1
10       Perch    Parkki    0
11        Pike    Parkki    0
12       Roach    Parkki    1
> accuracy <- sum(diag(conf_matrix)) / sum(conf_matrix)
> print(paste("Accuracy:", accuracy))
[1] "Accuracy: 0.46875"
> names(conf_matrix_df) <- c("Predicted", "Actual", "Freq")
> ggplot(data = conf_matrix_df, aes(x = Actual, y = Predicted, fill = Freq)) +
+   geom_tile() +
+   scale_fill_gradient(low = "white", high = "steelblue") +
+   labs(title = "Confusion Matrix",
+        x = "Actual",
+        y = "Predicted")
```



K-Nearest Neighbour:

```
> library(class)
> library(ggplot2)
> fish_data <- read.csv("Fish.csv")
> str(fish_data)
'data.frame':   159 obs. of  7 variables:
 $ Species: chr  "Bream" "Bream" "Bream" "Bream" ...
 $ Weight : num  242 290 340 363 430 450 500 390 450 500 ...
 $ Length1: num  23.2 24 23.9 26.3 26.5 26.8 26.8 27.6 27.6 28.5 ...
 $ Length2: num  25.4 26.3 26.5 29 29 29.7 29.7 30 30 30.7 ...
 $ Length3: num  30 31.2 31.1 33.5 34 34.7 34.5 35 35.1 36.2 ...
 $ Height : num  11.5 12.5 12.4 12.7 12.4 ...
 $ Width  : num  4.02 4.31 4.7 4.46 5.13 ...
> head(fish_data)
  Species Weight Length1 Length2 Length3  Height  Width
1   Bream    242    23.2    25.4    30.0 11.5200 4.0200
2   Bream    290    24.0    26.3    31.2 12.4800 4.3056
3   Bream    340    23.9    26.5    31.1 12.3778 4.6961
4   Bream    363    26.3    29.0    33.5 12.7300 4.4555
5   Bream    430    26.5    29.0    34.0 12.4440 5.1340
6   Bream    450    26.8    29.7    34.7 13.6024 4.9274
> split_index <- sample(1:nrow(fish_data), 0.8 * nrow(fish_data))
> train_data <- fish_data[split_index, ]
> test_data <- fish_data[-split_index, ]
> knn_model <- knn(train = train_data[, -1], test = test_data[, -1], cl = train_data$Species, k = 3)
> ggplot(data = test_data, aes(x = Length1, y = Length2, color = factor(predictions))) +
+   geom_point() +
+   labs(title = "K-Nearest Neighbors Prediction",
+        x = "Length1",
+        y = "Length2",
+        color = "Predicted Species")
> predictions <- knn(train = train_data[, -1], test = test_data[, -1], cl = train_data$Species, k = 3)
> conf_matrix <- table(predictions, test_data$Species)
```

```
> print(conf_matrix)

predictions Bream Parkki Perch Pike Roach Smelt
  Bream         9      0     1    1     1     0
  Parkki        0      0     2    0     0     0
  Perch         0      0     5    0     2     0
  Pike          0      0     0    2     0     0
  Roach         0      0     3    0     1     0
  Smelt         0      0     0    0     0     2
  Whitefish     0      1     0    2     0     0
> accuracy <- sum(diag(conf_matrix)) / sum(conf_matrix)
> print(paste("Accuracy:", accuracy))
[1] "Accuracy: 0.59375"
```