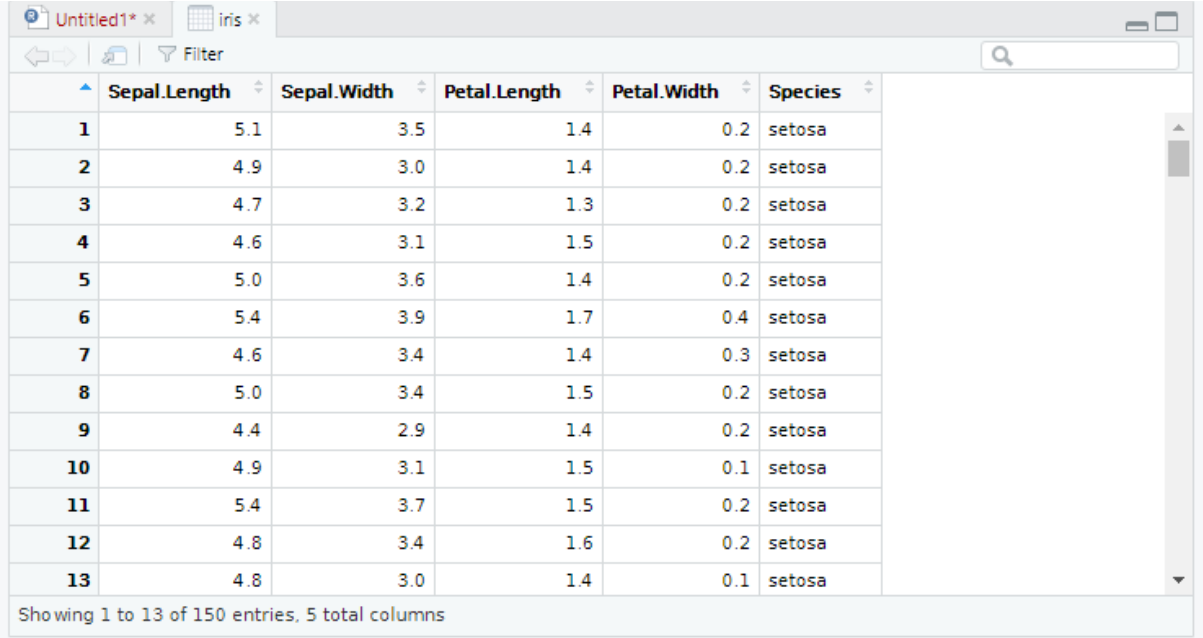
Name : Cheruvathoor Abin Anto
Roll No : TIT2425008

PRACTICAL 4 : Data Mining

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 Naïve Bayes Classifier model for this data

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



Untitled1* × 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

```
#split data into training and test data
install.packages("caTools")
library(caTools) #library used for split fn
split <- sample.split(iris$Species,SplitRatio = 0.7)
split
```

```
> 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")
train_cl #this is my training data

```
> train_cl <- subset(iris,split=="TRUE")
> train_cl
    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
2            4.9         3.0          1.4         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
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
15           5.8         4.0          1.2         0.2     setosa
17           5.4         3.9          1.3         0.4     setosa
18           5.1         3.5          1.4         0.3     setosa
19           5.7         3.8          1.7         0.3     setosa
21           5.4         3.4          1.7         0.2     setosa
23           4.6         3.6          1.0         0.2     setosa
24           5.1         3.3          1.7         0.5     setosa
26           5.0         3.0          1.6         0.2     setosa
27           5.0         3.4          1.6         0.4     setosa
28           5.2         3.5          1.5         0.2     setosa
29           5.2         3.4          1.4         0.2     setosa
30           4.7         3.2          1.6         0.2     setosa
32           5.4         3.4          1.5         0.4     setosa
33           5.2         4.1          1.5         0.1     setosa
35           4.9         3.1          1.5         0.2     setosa
36           5.0         3.2          1.2         0.2     setosa
38           4.9         3.6          1.4         0.1     setosa
39           4.4         3.0          1.3         0.2     setosa
40           5.1         3.4          1.5         0.2     setosa
42           4.5         2.3          1.3         0.3     setosa
43           4.4         3.2          1.3         0.2     setosa
44           5.0         3.5          1.6         0.6     setosa
46           4.8         3.0          1.4         0.3     setosa
48           4.6         3.2          1.4         0.2     setosa
```

test_cl <- subset(iris,split=="FALSE")
test_cl #this is my test data

```
> test_cl <- subset(iris,split=="FALSE")
> test_cl
    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
8            5.0         3.4          1.5         0.2     setosa
14           4.3         3.0          1.1         0.1     setosa
16           5.7         4.4          1.5         0.4     setosa
20           5.1         3.8          1.5         0.3     setosa
22           5.1         3.7          1.5         0.4     setosa
25           4.8         3.4          1.9         0.2     setosa
31           4.8         3.1          1.6         0.2     setosa
34           5.5         4.2          1.4         0.2     setosa
37           5.5         3.5          1.3         0.2     setosa
41           5.0         3.5          1.3         0.3     setosa
45           5.1         3.8          1.9         0.4     setosa
47           5.1         3.8          1.6         0.2     setosa
52           6.4         3.2          4.5         1.5 versicolor
55           6.5         2.8          4.6         1.5 versicolor
58           4.9         2.4          3.3         1.0 versicolor
59           6.6         2.9          4.6         1.3 versicolor
66           6.7         3.1          4.4         1.4 versicolor
68           5.8         2.7          4.1         1.0 versicolor
70           5.6         2.5          3.9         1.1 versicolor
75           6.4         2.9          4.3         1.3 versicolor
79           6.0         2.9          4.5         1.5 versicolor
80           5.7         2.6          3.5         1.0 versicolor
83           5.8         2.7          3.9         1.2 versicolor
86           6.0         3.4          4.5         1.6 versicolor
92           6.1         3.0          4.6         1.4 versicolor
94           5.0         2.3          3.3         1.0 versicolor
97           5.7         2.9          4.2         1.3 versicolor
104          6.3         2.9          5.6         1.8  virginica
107          4.9         2.5          4.5         1.7  virginica
110          7.2         3.6          6.1         2.5  virginica
112          6.4         2.7          5.3         1.9  virginica
```

#create naive bayes classifier model
install.packages("e1071")
library(e1071) #Library for naiveBayes
nbclassifier <- naiveBayes(Species~.,data=train_cl)
nbclassifier

```
> nbclassifier <- naiveBayes(Species~.,data=train_cl)
> nbclassifier

Naive Bayes Classifier for Discrete Predictors

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

A-priori probabilities:
Y
    setosa versicolor  virginica
 0.3333333  0.3333333  0.3333333

Conditional probabilities:
            Sepal.Length
Y               [,1]      [,2]
  setosa     4.997143 0.3527086
  versicolor 5.931429 0.5155074
  virginica  6.654286 0.6418461

            Sepal.Width
Y               [,1]      [,2]
  setosa     3.371429 0.3610210
  versicolor 2.748571 0.3220992
  virginica  2.968571 0.3287869

            Petal.Length
Y               [,1]      [,2]
  setosa     1.451429 0.1560004
  versicolor 4.308571 0.4692744
  virginica  5.571429 0.5680573
```

```
             Petal.Width
Y                [,1]      [,2]
  setosa     0.2457143 0.1120474
  versicolor 1.3485714 0.1900022
  virginica  2.0257143 0.2593463
```

#predicition on test data
pred <- predict(nbclassifier,newdata=test_cl)
pred

```
> pred <- predict(nbclassifier,newdata=test_cl)
> pred
 [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.

#confusion matrix
cm <- table(test_cl$Species,pred)
cm

```
> cm <- table(test_cl$Species,pred)
> cm
            pred
             setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         15         0
  virginica       0          2        13
> |
```

install.packages("caret")
library(caret) #Library for confusionMatrix

confusionMatrix(cm)

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

            pred
             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
> |
```