Experiment 6

Student Name: Anshuman Singh
UID: 20BCS2665
Branch: CSE
Section/Group: 902-A
Semester: 6th
Date of Performance: 7/04/23
Subject Name: Data Mining Lab
Subject Code: 20CSP-376

Aim: To perform classification using the Bayesian classification algorithm using R

Script and Output:

```
library(e1071)

library(gmodels)

library(dplyr)  iris <-

read.csv('iris_1.csv')

str(iris)

summary(iris)
```

```
index = sample(2,nrow(iris),prob=c(0.8,0.2),replace=TRUE)



set.seed(1234)

#training set  train=

iris[index==1,]

#testing set  test=

iris[index==2,]

#test_data will be given as an input to the model to predict species  test_data

= test[1:4]

#test_labels are the actual values of species of the test data

test_label=test[,5]

print(test_label)
```

```
model=naiveBayes(train$class~.,train)

model  test_result=predict(model,test_data)

test_result

#compare the predicted and actual values


CrossTable(x=test_label, y=test_result)
```

OUTPUT:

R 4.1.2 · ~/

```
> library(e1071)
> library(gmodels)
> library(dplyr)
> iris <-read.csv('iris_1.csv')
> str(iris)
'data.frame':   150 obs. of  5 variables:
 $ sepal_length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ sepal_width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ petal_length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ petal_width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ class       : chr  "Iris-setosa" "Iris-setosa" "Iris-setosa" "Iris-setosa" ...
> summary(iris)
  sepal_length    sepal_width     petal_length    petal_width       class
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   Length:150
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   Class :character
 Median :5.800   Median :3.000   Median :4.350   Median :1.300   Mode  :character
 Mean   :5.843   Mean   :3.054   Mean   :3.759   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
>
> index = sample(2,nrow(iris),prob=c(0.8,0.2),replace=TRUE)
>
> set.seed(1234)
> #training set
> train= iris[index==1,]
> #testing set
> test= iris[index==2,]
> #test_data will be given as an input to the model to predict species
> test_data = test[1:4]
> #test_labels are the actual values of species of the test data
> test_label=test[,5]
> print(test_label)
 [1] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"
 [6] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-versicolor" "Iris-versicolor"
[11] "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor"
[16] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[21] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[26] "Iris-virginica"  "Iris-virginica"
```

```
> model=naiveBayes(train$class~.,train)
> model

Naive Bayes Classifier for Discrete Predictors

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

A-priori probabilities:
Y
    Iris-setosa Iris-versicolor  Iris-virginica
      0.3414634       0.3495935       0.3089431

Conditional probabilities:
                 sepal_length
Y                     [,1]      [,2]
  Iris-setosa     5.009524 0.3369950
  Iris-versicolor 5.986047 0.5138934
  Iris-virginica  6.586842 0.6597178

                 sepal_width
Y                     [,1]      [,2]
  Iris-setosa     3.419048 0.3690714
  Iris-versicolor 2.783721 0.2910819
  Iris-virginica  2.968421 0.3425700

                 petal_length
Y                     [,1]      [,2]
  Iris-setosa     1.473810 0.1767726
  Iris-versicolor 4.295349 0.4765718
  Iris-virginica  5.568421 0.5723849

                 petal_width
Y                     [,1]      [,2]
  Iris-setosa     0.250000 0.1109823
  Iris-versicolor 1.330233 0.1994455
  Iris-virginica  2.023684 0.2755184


> test_result=predict(model,test_data)
> test_result
 [1] Iris-setosa     Iris-setosa     Iris-setosa     Iris-setosa     Iris-setosa
 [6] Iris-setosa     Iris-setosa     Iris-setosa     Iris-versicolor Iris-versicolor
[11] Iris-versicolor Iris-versicolor Iris-versicolor Iris-versicolor Iris-versicolor
[16] Iris-virginica  Iris-virginica  Iris-virginica  Iris-virginica  Iris-virginica
[21] Iris-virginica  Iris-virginica  Iris-versicolor Iris-virginica  Iris-virginica
[26] Iris-virginica  Iris-virginica
Levels: Iris-setosa Iris-versicolor Iris-virginica
> #compare the predicted and actual values
> CrossTable(x=test_label, y=test_result)
```

```
Levels: Iris-setosa Iris-versicolor Iris-virginica
> #compare the predicted and actual values
> CrossTable(x=test_label, y=test_result)


   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  27


                | test_result
     test_label |     Iris-setosa | Iris-versicolor |  Iris-virginica |        Row Total |
----------------|-----------------|-----------------|-----------------|-----------------|
    Iris-setosa |               8 |               0 |               0 |               8 |
                |          13.370 |           2.370 |           3.259 |                 |
                |           1.000 |           0.000 |           0.000 |           0.296 |
                |           1.000 |           0.000 |           0.000 |                 |
                |           0.296 |           0.000 |           0.000 |                 |
----------------|-----------------|-----------------|-----------------|-----------------|
Iris-versicolor |               0 |               7 |               0 |               7 |
                |           2.074 |          11.699 |           2.852 |                 |
                |           0.000 |           1.000 |           0.000 |           0.259 |
                |           0.000 |           0.875 |           0.000 |                 |
                |           0.000 |           0.259 |           0.000 |                 |
----------------|-----------------|-----------------|-----------------|-----------------|
 Iris-virginica |               0 |               1 |              11 |              12 |
                |           3.556 |           1.837 |           7.639 |                 |
                |           0.000 |           0.083 |           0.917 |           0.444 |
                |           0.000 |           0.125 |           1.000 |                 |
                |           0.000 |           0.037 |           0.407 |                 |
----------------|-----------------|-----------------|-----------------|-----------------|
   Column Total |               8 |               8 |              11 |              27 |
                |           0.296 |           0.296 |           0.407 |                 |
----------------|-----------------|-----------------|-----------------|-----------------|

> |
```

4. Learning Outcome:

- Setting and getting directory path for R project.
- Creating and working with FP growth algorithm.
- Saving data frame in .csv file format in R.
- Loading .csv file in data frame and print it.
- Basic of the data frame and r studio uses in the data frame