



Experiment 6

Student Name: Husanpreet Kaur

Branch: BE-CSE

Semester: 6

Subject Name: Data Mining Lab

UID: 20BCS9512

Section/Group: 20BCS-DM-714/A

Subject Code: 20CSP-376

Date of Performance: 25-04-2023

1. Aim/Overview of the practical: To perform the classification using Bayesian classification.

2. Tools used: RStudio

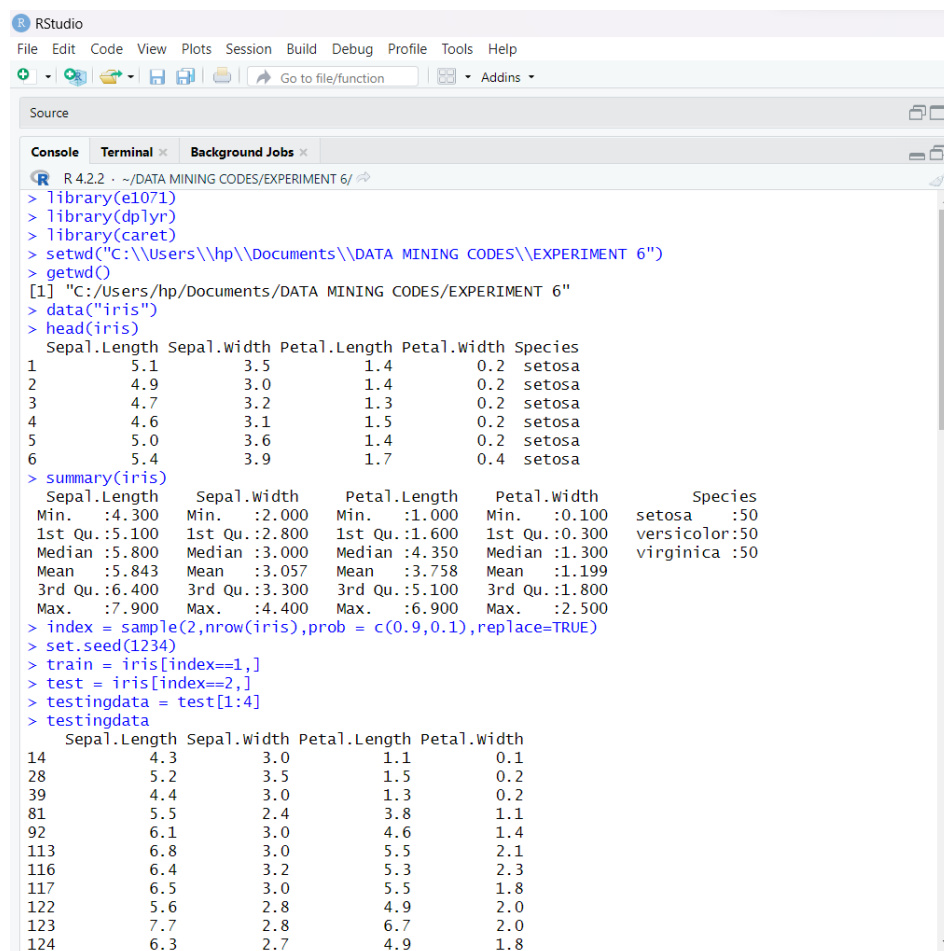
3. Code:

```
install.packages("naivebayes")
install.packages("e1071")
install.packages("caret")
library(e1071)
library(dplyr)
library(caret)
setwd("C:\\Users\\hp\\Documents\\DATA MINING
CODES\\EXPERIMENT 6")
getwd()
data("iris")
head(iris)
summary(iris)
index = sample(2,nrow(iris),prob = c(0.9,0.1),replace=TRUE)
set.seed(1234)
train = iris[index==1,]
test = iris[index==2,]
testingdata = test[1:4]
testingdata
label = test[,5]
```

```
label
model=naiveBayes(train$Species~.,train)
model
result=predict(model, testingdata)
result
table1=table(x=label, y=result)
table1
matrix=confusionMatrix(table1)
matrix
```

4. Output:

RStudio:



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source

Console Terminal Background Jobs

R 4.2.2 ~ /DATA MINING CODES/EXPERIMENT 6/
> library(e1071)
> library(dplyr)
> library(caret)
> setwd("C:\\Users\\hp\\Documents\\DATA MINING CODES\\EXPERIMENT 6")
> getwd()
[1] "C:/Users/hp/Documents/DATA MINING CODES/EXPERIMENT 6"
> data("iris")
> head(iris)
  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
> summary(iris)
  Sepal.Length      Sepal.Width      Petal.Length      Petal.Width      Species
Min.   :4.300      Min.   :2.000      Min.   :1.000      Min.   :0.100      setosa :50
1st Qu.:5.100      1st Qu.:2.800      1st Qu.:1.600      1st Qu.:0.300      versicolor:50
Median :5.800      Median :3.000      Median :4.350      Median :1.300      virginica :50
Mean   :5.843      Mean   :3.057      Mean   :3.758      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(0.9,0.1),replace=TRUE)
> set.seed(1234)
> train = iris[index==1,]
> test = iris[index==2,]
> testingdata = test[1:4]
> testingdata
  Sepal.Length Sepal.Width Petal.Length Petal.Width
14          4.3         3.0          1.1          0.1
28          5.2         3.5          1.5          0.2
39          4.4         3.0          1.3          0.2
81          5.5         2.4          3.8          1.1
92          6.1         3.0          4.6          1.4
113         6.8         3.0          5.5          2.1
116         6.4         3.2          5.3          2.3
117         6.5         3.0          5.5          1.8
122         5.6         2.8          4.9          2.0
123         7.7         2.8          6.7          2.0
124         6.3         2.7          4.9          1.8
```



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

```
131      7.4      2.8      6.1      1.9
135      6.1      2.6      5.6      1.4
137      6.3      3.4      5.6      2.4
142      6.9      3.1      5.1      2.3
149      6.2      3.4      5.4      2.3
> label = test[,5]
> label
[1] setosa      setosa      setosa      versicolor versicolor virginica  virginica  virginica
[9] virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica
Levels: setosa versicolor virginica
> model=naiveBayes(train$Species~.,train)
> model

Naive Bayes Classifier for Discrete Predictors

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

A-priori probabilities:
Y
      setosa versicolor  virginica
0.3507463  0.3582090  0.2910448

Conditional probabilities:
      Sepal.Length
Y      [,1]      [,2]
setosa  5.029787  0.3348685
versicolor 5.941667 0.5226094
virginica  6.594872 0.6529080

      Sepal.Width
Y      [,1]      [,2]
setosa  3.444681 0.3803937
versicolor 2.772917 0.3140264
virginica  2.971795 0.3386884

      Petal.Length
Y      [,1]      [,2]
setosa  1.472340 0.1690152
versicolor 4.262500 0.4724878
virginica  5.564103 0.5654468

      Petal.Width
Y      [,1]      [,2]
setosa  0.2510638 0.1060606
versicolor 1.3291667 0.1988888
virginica  2.0256410 0.2721435

> result=predict(model, testingdata)
> result
[1] setosa      setosa      setosa      versicolor versicolor virginica  virginica  virginica
[9] virginica  virginica  virginica  virginica  versicolor virginica  virginica  virginica
Levels: setosa versicolor virginica
> table1=table(x=label, y=result)
> table1
      y
x      setosa versicolor virginica
setosa      3          0          0
versicolor  0          2          0
virginica   0          1         10
> matrix=confusionMatrix(table1)
> matrix
Confusion Matrix and Statistics

      y
x      setosa versicolor virginica
setosa      3          0          0
versicolor  0          2          0
virginica   0          1         10

Overall Statistics

          Accuracy : 0.9375
          95% CI   : (0.6977, 0.9984)
    No Information Rate : 0.625
    P-Value [Acc > NIR] : 0.005746

          Kappa : 0.8779

    McNemar's Test P-Value : NA

Statistics by Class:
```

```
Statistics by Class:

              Class: setosa Class: versicolor Class: virginica
Sensitivity          1.0000          0.6667          1.0000
Specificity          1.0000          1.0000          0.8333
Pos Pred Value       1.0000          1.0000          0.9091
Neg Pred Value       1.0000          0.9286          1.0000
Prevalence           0.1875          0.1875          0.6250
Detection Rate       0.1875          0.1250          0.6250
Detection Prevalence 0.1875          0.1250          0.6875
Balanced Accuracy     1.0000          0.8333          0.9167
> |
```

5. Observation:

- Learnt how to use R and create a file in Rstudio.
- Learnt how to install packages in Rstudio.
- Learnt how to split data into training and test set.
- Learnt how to load dataset iris in Rstudio.
- Learnt the use of naivebayes, e1071 and caret libraries.
- Learnt how to model data, create table and confusion matrix.