Experiment-2.2

Student Name: Nabha Varshney **UID: 20BCS4995**

Branch: CSE **Section/Group: 20BCS-DM-704 (A)**

Semester: 6th **Date of Performance: thApr 2023**

Subject Name: Data Mining **Subject Code: 20CSP- 351**

Aim – To perform the classification using Bayesian classification algorithm using R.

Objective-

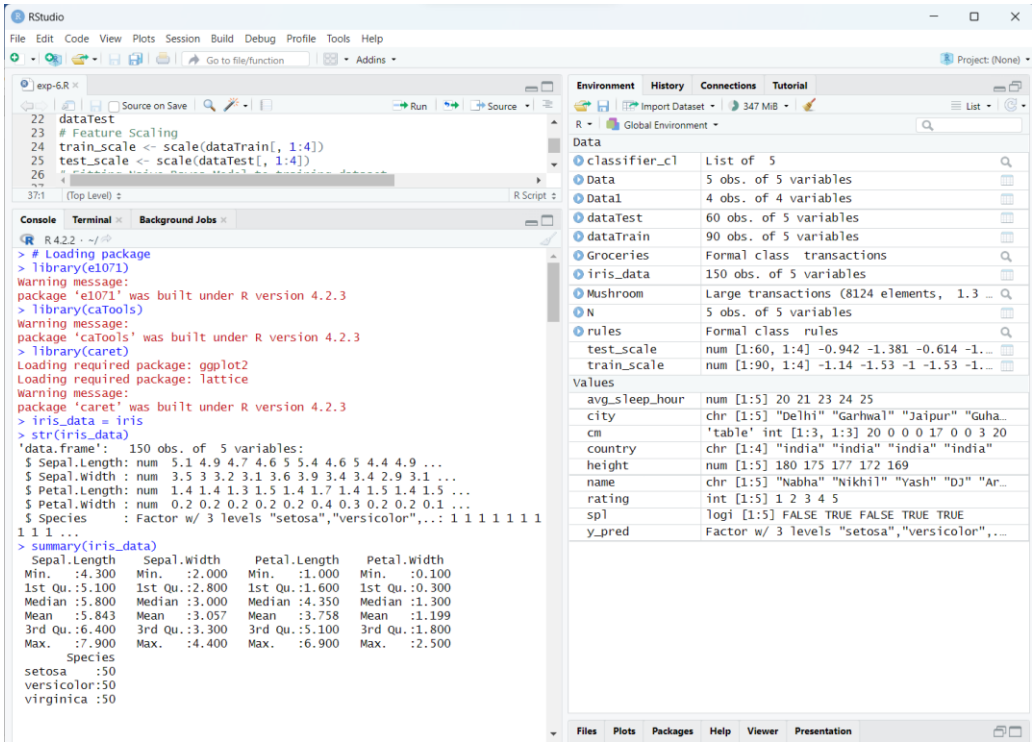
- Represent the reading of file using R studio
- Displaying the pattern on e1071, caret and caTools Tool.
- Demonstration of Non-linear classification algorithm.

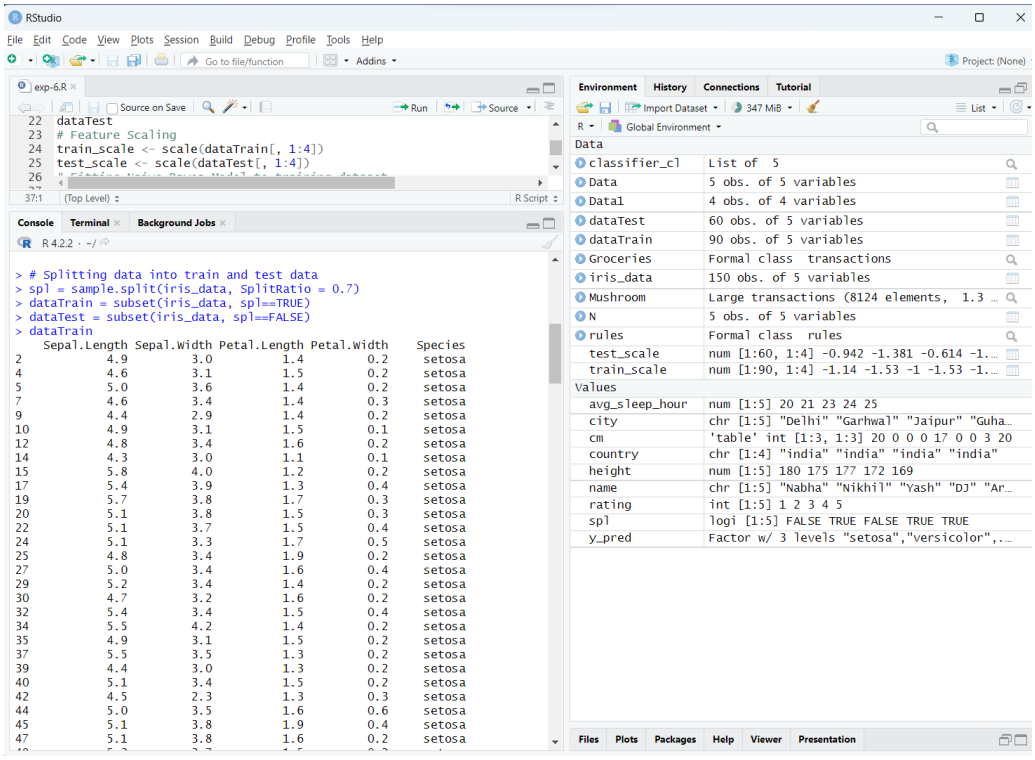
Script and Output-

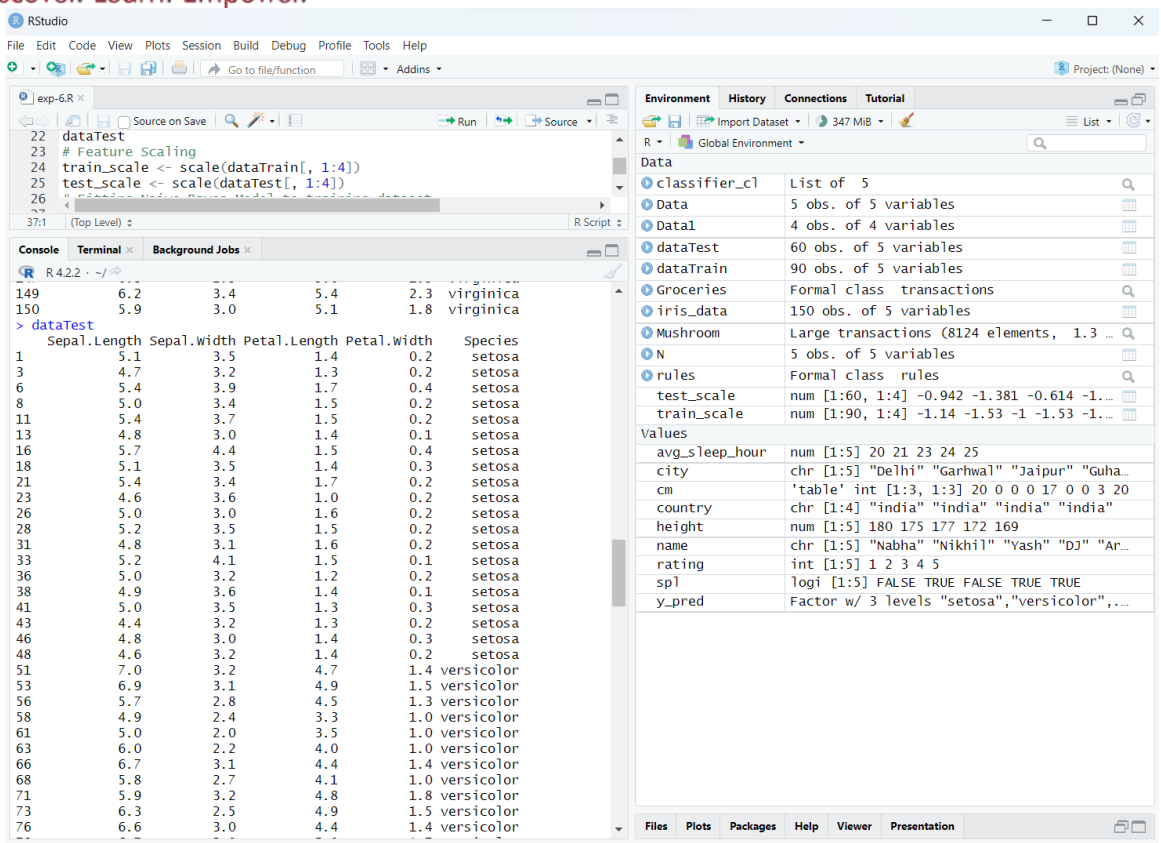
```
# Structure
str(iris)
#Performing Naive Bayes on Dataset

#Using Naive Bayes algorithm on the dataset which includes 11 persons and 6
variables or attributes

# Installing Packages
install.packages("e1071")
install.packages("caTools")
install.packages("caret")

# Loading package
library(e1071)
library(caTools)
library(caret)
```

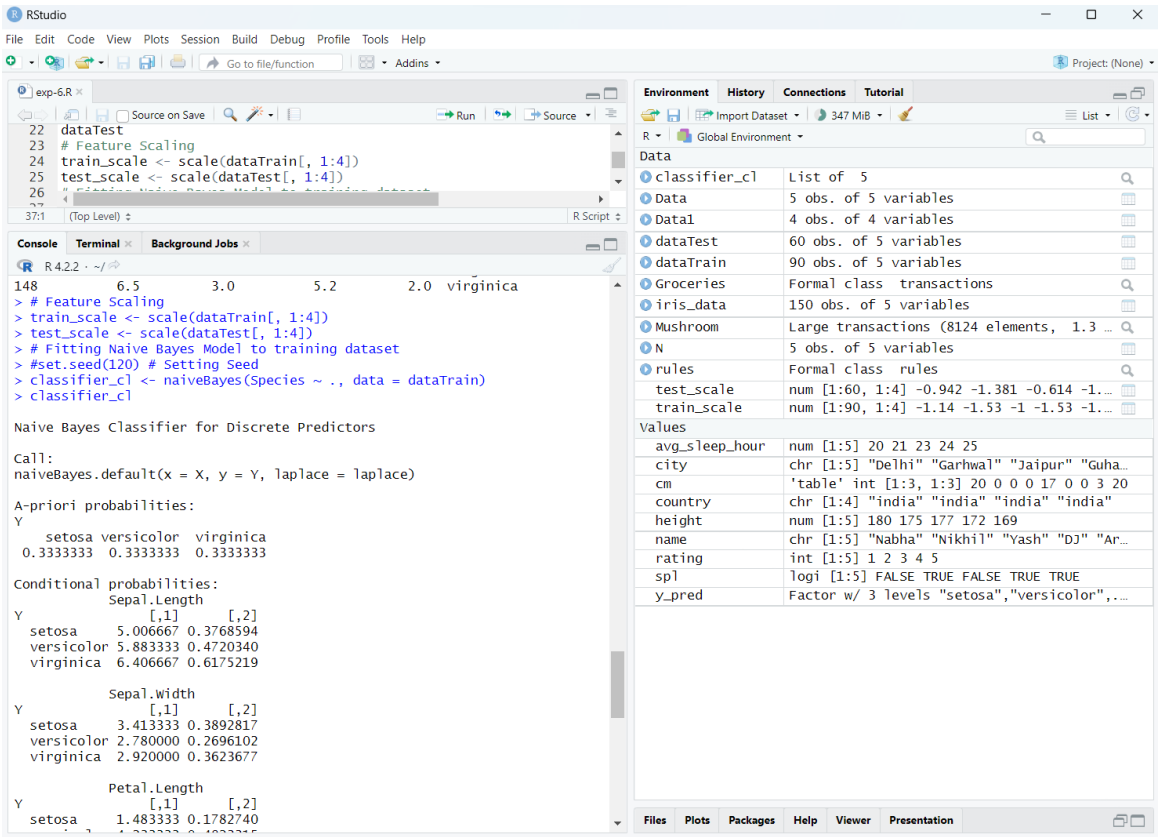
```
iris_data = iris
str(iris_data)
summary(iris_data)

# Splitting data into train and test data
spl = sample.split(iris_data, SplitRatio = 0.7)
dataTrain = subset(iris_data, spl==TRUE)
dataTest = subset(iris_data, spl==FALSE)
dataTrain

dataTest

# Feature Scaling
train_scale <- scale(dataTrain[, 1:4])
test_scale <- scale(dataTest[, 1:4])
# Fitting Naive Bayes Model to training dataset

#set.seed(120) # Setting Seed
classifier_cl <- naiveBayes(Species ~ ., data = dataTrain)
classifier_cl

# Predicting on test data'
y_pred <- predict(classifier_cl, newdata = dataTest)

# Confusion Matrix
cm <- table(dataTest$Species, y_pred)
cm

# Model Evaluation
confusionMatrix(cm)
```

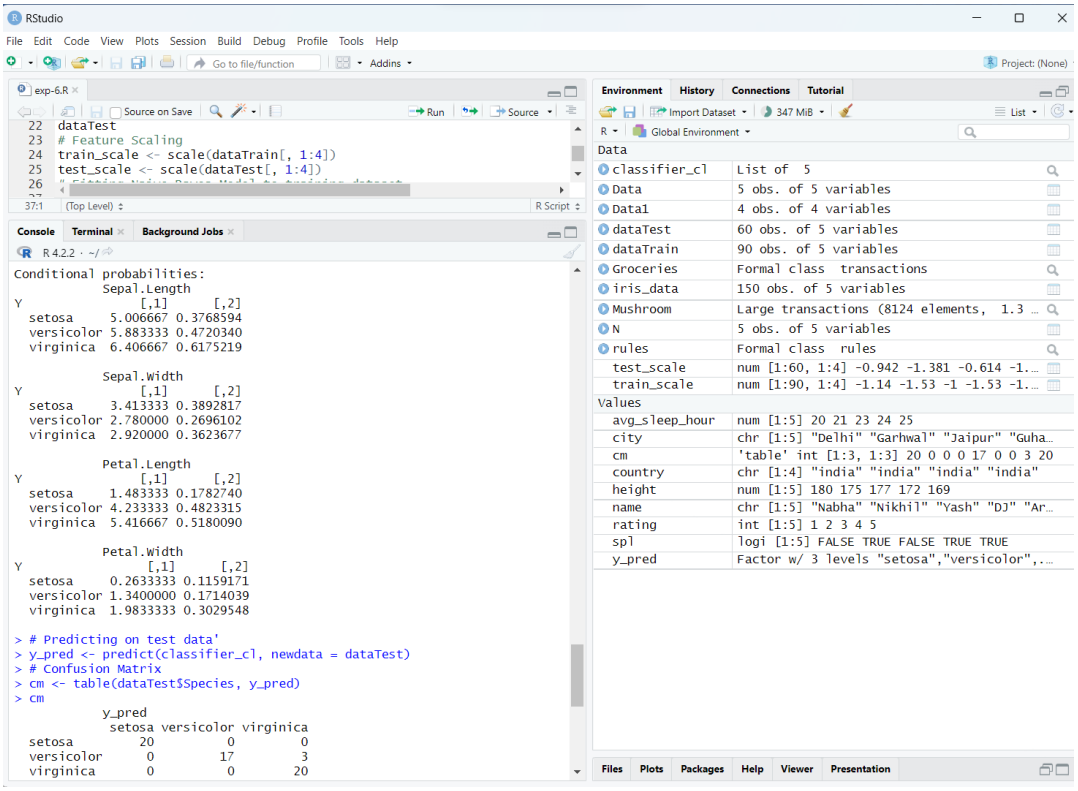
Output-

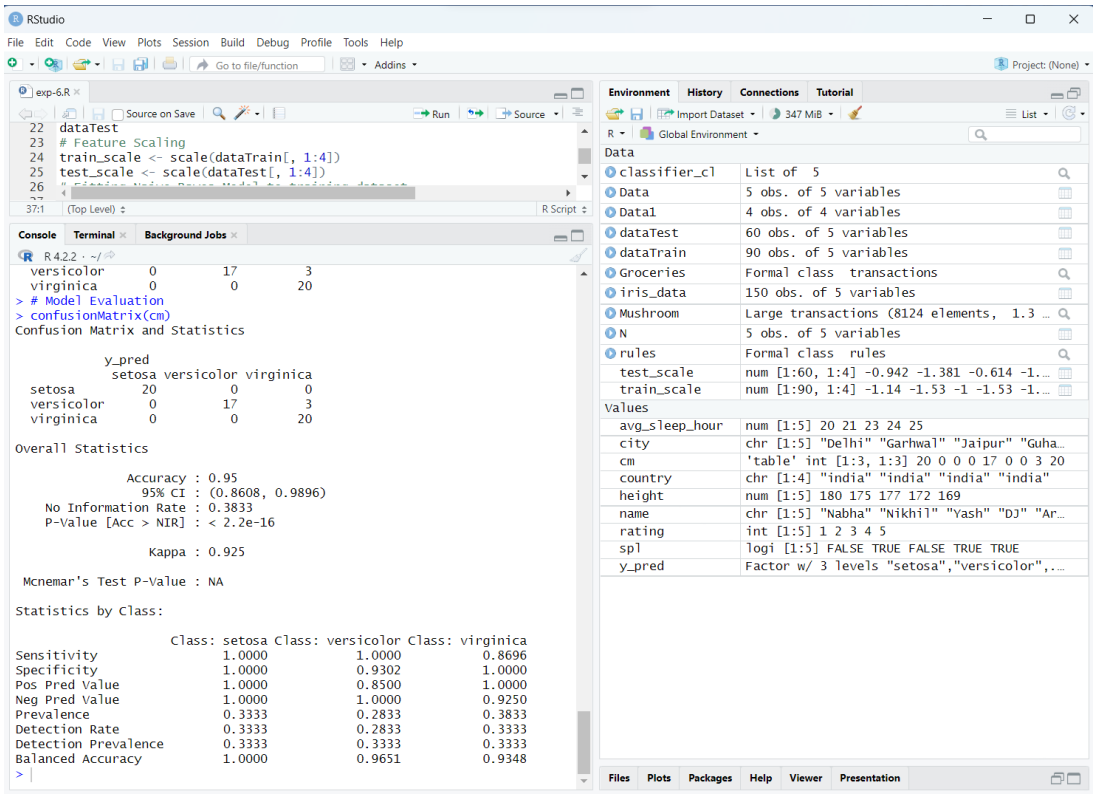

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
exp-6.R
22 dataTest
23 # Feature Scaling
24 train_scale <- scale(dataTrain[, 1:4])
25 test_scale <- scale(dataTest[, 1:4])
Console Terminal Background Jobs
> # Loading package
> library(e1071)
Warning message:
package 'e1071' was built under R version 4.2.3
> library(caTools)
Warning message:
package 'caTools' was built under R version 4.2.3
> library(caret)
Loading required package: ggplot2
Loading required package: lattice
Warning message:
package 'caret' was built under R version 4.2.3
> iris_data = iris
> str(iris_data)
'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 ...
$ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 ...
> summary(iris_data)
Sepal.Length Sepal.Width Petal.Length Petal.Width
Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100
1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300
Median :5.800 Median :3.000 Median :4.350 Median :1.300
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
Species
setosa :50
versicolor:50
virginica :50
Environment History Connections Tutorial
Import Dataset 347 MiB
Global Environment
Data
classifier_cl List of 5
Data 5 obs. of 5 variables
Data1 4 obs. of 4 variables
dataTest 60 obs. of 5 variables
dataTrain 90 obs. of 5 variables
Groceries Formal class transactions
iris_data 150 obs. of 5 variables
Mushroom Large transactions (8124 elements, 1.3 ...
N 5 obs. of 5 variables
rules Formal class rules
test_scale num [1:60, 1:4] -0.942 -1.381 -0.614 -1....
train_scale num [1:90, 1:4] -1.14 -1.53 -1 -1.53 -1....
Values
avg_sleep_hour num [1:5] 20 21 23 24 25
city chr [1:5] "Delhi" "Garhwal" "Jaipur" "Guha...
cm 'table' int [1:3, 1:3] 20 0 0 0 17 0 0 3 20
country chr [1:4] "india" "india" "india" "india"
height num [1:5] 180 175 177 172 169
name chr [1:5] "Nabha" "Nikhil" "Yash" "DJ" "Ar...
rating int [1:5] 1 2 3 4 5
spl logi [1:5] FALSE TRUE FALSE TRUE TRUE
y_pred Factor w/ 3 levels "setosa","versicolor",...
Files Plots Packages Help Viewer Presentation


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
exp-6.R
22 dataTest
23 # Feature Scaling
24 train_scale <- scale(dataTrain[, 1:4])
25 test_scale <- scale(dataTest[, 1:4])
Console Terminal Background Jobs
> # Splitting data into train and test data
> spl = sample.split(iris_data, SplitRatio = 0.7)
> dataTrain = subset(iris_data, spl==TRUE)
> dataTest = subset(iris_data, spl==FALSE)
> dataTrain
Sepal.Length Sepal.Width Petal.Length Petal.Width Species
2 4.9 3.0 1.4 0.2 setosa
4 4.6 3.1 1.5 0.2 setosa
5 5.0 3.6 1.4 0.2 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
12 4.8 3.4 1.6 0.2 setosa
14 4.3 3.0 1.1 0.1 setosa
15 5.8 4.0 1.2 0.2 setosa
17 5.4 3.9 1.3 0.4 setosa
19 5.7 3.8 1.7 0.3 setosa
20 5.1 3.8 1.5 0.3 setosa
22 5.1 3.7 1.5 0.4 setosa
24 5.1 3.3 1.7 0.5 setosa
25 4.8 3.4 1.9 0.2 setosa
27 5.0 3.4 1.6 0.4 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
34 5.5 4.2 1.4 0.2 setosa
35 4.9 3.1 1.5 0.2 setosa
37 5.5 3.5 1.3 0.2 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
44 5.0 3.5 1.6 0.6 setosa
45 5.1 3.8 1.9 0.4 setosa
47 5.1 3.8 1.6 0.2 setosa
Environment History Connections Tutorial
Import Dataset 347 MiB
Global Environment
Data
classifier_cl List of 5
Data 5 obs. of 5 variables
Data1 4 obs. of 4 variables
dataTest 60 obs. of 5 variables
dataTrain 90 obs. of 5 variables
Groceries Formal class transactions
iris_data 150 obs. of 5 variables
Mushroom Large transactions (8124 elements, 1.3 ...
N 5 obs. of 5 variables
rules Formal class rules
test_scale num [1:60, 1:4] -0.942 -1.381 -0.614 -1....
train_scale num [1:90, 1:4] -1.14 -1.53 -1 -1.53 -1....
Values
avg_sleep_hour num [1:5] 20 21 23 24 25
city chr [1:5] "Delhi" "Garhwal" "Jaipur" "Guha...
cm 'table' int [1:3, 1:3] 20 0 0 0 17 0 0 3 20
country chr [1:4] "india" "india" "india" "india"
height num [1:5] 180 175 177 172 169
name chr [1:5] "Nabha" "Nikhil" "Yash" "DJ" "Ar...
rating int [1:5] 1 2 3 4 5
spl logi [1:5] FALSE TRUE FALSE TRUE TRUE
y_pred Factor w/ 3 levels "setosa","versicolor",...
Files Plots Packages Help Viewer Presentation


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
exp-6.R
22 dataTest
23 # Feature Scaling
24 train_scale <- scale(dataTrain[, 1:4])
25 test_scale <- scale(dataTest[, 1:4])
Console Terminal Background Jobs
R 4.2.2 · ~/
149 6.2 3.4 5.4 2.3 virginica
150 5.9 3.0 5.1 1.8 virginica
> dataTest
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
6 5.4 3.9 1.7 0.4 setosa
8 5.0 3.4 1.5 0.2 setosa
11 5.4 3.7 1.5 0.2 setosa
13 4.8 3.0 1.4 0.1 setosa
16 5.7 4.4 1.5 0.4 setosa
18 5.1 3.5 1.4 0.3 setosa
21 5.4 3.4 1.7 0.2 setosa
23 4.6 3.6 1.0 0.2 setosa
26 5.0 3.0 1.6 0.2 setosa
28 5.2 3.5 1.5 0.2 setosa
31 4.8 3.1 1.6 0.2 setosa
33 5.2 4.1 1.5 0.1 setosa
36 5.0 3.2 1.2 0.2 setosa
38 4.9 3.6 1.4 0.1 setosa
41 5.0 3.5 1.3 0.3 setosa
43 4.4 3.2 1.3 0.2 setosa
46 4.8 3.0 1.4 0.3 setosa
48 4.6 3.2 1.4 0.2 setosa
51 7.0 3.2 4.7 1.4 versicolor
53 6.9 3.1 4.9 1.5 versicolor
56 5.7 2.8 4.5 1.3 versicolor
58 4.9 2.4 3.3 1.0 versicolor
61 5.0 2.0 3.5 1.0 versicolor
63 6.0 2.2 4.0 1.0 versicolor
66 6.7 3.1 4.4 1.4 versicolor
68 5.8 2.7 4.1 1.0 versicolor
71 5.9 3.2 4.8 1.8 versicolor
73 6.3 2.5 4.9 1.5 versicolor
76 6.6 3.0 4.4 1.4 versicolor
Environment History Connections Tutorial
Import Dataset 347 MiB
Global Environment
Data
classifier_cl List of 5
Data 5 obs. of 5 variables
Data1 4 obs. of 4 variables
dataTest 60 obs. of 5 variables
dataTrain 90 obs. of 5 variables
Groceries Formal class transactions
iris_data 150 obs. of 5 variables
Mushroom Large transactions (8124 elements, 1.3 …
N 5 obs. of 5 variables
rules Formal class rules
test_scale num [1:60, 1:4] -0.942 -1.381 -0.614 -1.…
train_scale num [1:90, 1:4] -1.14 -1.53 -1 -1.53 -1.…
Values
avg_sleep_hour num [1:5] 20 21 23 24 25
city chr [1:5] "Delhi" "Garhwal" "Jaipur" "Guha…
cm 'table' int [1:3, 1:3] 20 0 0 0 17 0 0 3 20
country chr [1:4] "india" "india" "india" "india"
height num [1:5] 180 175 177 172 169
name chr [1:5] "Nabha" "Nikhil" "Yash" "DJ" "Ar…
rating int [1:5] 1 2 3 4 5
spl logi [1:5] FALSE TRUE FALSE TRUE TRUE
y_pred Factor w/ 3 levels "setosa","versicolor",.…
Files Plots Packages Help Viewer Presentation


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
exp-6.R
22 dataTest
23 # Feature Scaling
24 train_scale <- scale(dataTrain[, 1:4])
25 test_scale <- scale(dataTest[, 1:4])
Console Terminal Background Jobs
R 4.2.2 · ~/
148 6.5 3.0 5.2 2.0 virginica
> # Feature Scaling
> train_scale <- scale(dataTrain[, 1:4])
> test_scale <- scale(dataTest[, 1:4])
> # Fitting Naive Bayes Model to training dataset
> #set.seed(120) # Setting Seed
> classifier_cl <- naiveBayes(Species ~ ., data = dataTrain)
> classifier_cl
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 5.006667 0.3768594
versicolor 5.883333 0.4720340
virginica 6.406667 0.6175219
Sepal.Width
Y [,1] [,2]
setosa 3.413333 0.3892817
versicolor 2.780000 0.2696102
virginica 2.920000 0.3623677
Petal.Length
Y [,1] [,2]
setosa 1.483333 0.1782740
Environment History Connections Tutorial
Import Dataset 347 MiB
Global Environment
Data
classifier_cl List of 5
Data 5 obs. of 5 variables
Data1 4 obs. of 4 variables
dataTest 60 obs. of 5 variables
dataTrain 90 obs. of 5 variables
Groceries Formal class transactions
iris_data 150 obs. of 5 variables
Mushroom Large transactions (8124 elements, 1.3 …
N 5 obs. of 5 variables
rules Formal class rules
test_scale num [1:60, 1:4] -0.942 -1.381 -0.614 -1.…
train_scale num [1:90, 1:4] -1.14 -1.53 -1 -1.53 -1.…
Values
avg_sleep_hour num [1:5] 20 21 23 24 25
city chr [1:5] "Delhi" "Garhwal" "Jaipur" "Guha…
cm 'table' int [1:3, 1:3] 20 0 0 0 17 0 0 3 20
country chr [1:4] "india" "india" "india" "india"
height num [1:5] 180 175 177 172 169
name chr [1:5] "Nabha" "Nikhil" "Yash" "DJ" "Ar…
rating int [1:5] 1 2 3 4 5
spl logi [1:5] FALSE TRUE FALSE TRUE TRUE
y_pred Factor w/ 3 levels "setosa","versicolor",.…
Files Plots Packages Help Viewer Presentation


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
exp-6.R
22 dataTest
23 # Feature Scaling
24 train_scale <- scale(dataTrain[, 1:4])
25 test_scale <- scale(dataTest[, 1:4])
Console Terminal Background Jobs
Conditional probabilities:
Sepal.Length
Y [,1] [,2]
setosa 5.006667 0.3768594
versicolor 5.883333 0.4720340
virginica 6.406667 0.6175219
Sepal.Width
Y [,1] [,2]
setosa 3.413333 0.3892817
versicolor 2.780000 0.2696102
virginica 2.920000 0.3623677
Petal.Length
Y [,1] [,2]
setosa 1.483333 0.1782740
versicolor 4.233333 0.4823315
virginica 5.416667 0.5180090
Petal.Width
Y [,1] [,2]
setosa 0.2633333 0.1159171
versicolor 1.3400000 0.1714039
virginica 1.9833333 0.3029548
> # Predicting on test data'
> y_pred <- predict(classifier_cl, newdata = dataTest)
> # Confusion Matrix
> cm <- table(dataTest$Species, y_pred)
> cm
y_pred
setosa versicolor virginica
setosa 20 0 0
versicolor 0 17 3
virginica 0 0 20
Environment History Connections Tutorial
Data
classifier_cl List of 5
Data 5 obs. of 5 variables
Data1 4 obs. of 4 variables
dataTest 60 obs. of 5 variables
dataTrain 90 obs. of 5 variables
Groceries Formal class transactions
iris_data 150 obs. of 5 variables
Mushroom Large transactions (8124 elements, 1.3 ...
N 5 obs. of 5 variables
rules Formal class rules
test_scale num [1:60, 1:4] -0.942 -1.381 -0.614 -1...
train_scale num [1:90, 1:4] -1.14 -1.53 -1 -1.53 -1...
Values
avg_sleep_hour num [1:5] 20 21 23 24 25
city chr [1:5] "Delhi" "Garhwal" "Jaipur" "Guha...
cm 'table' int [1:3, 1:3] 20 0 0 0 17 0 0 3 20
country chr [1:4] "india" "india" "india" "india"
height num [1:5] 180 175 177 172 169
name chr [1:5] "Nabha" "Nikhil" "Yash" "DJ" "Ar...
rating int [1:5] 1 2 3 4 5
spl logi [1:5] FALSE TRUE FALSE TRUE TRUE
y_pred Factor w/ 3 levels "setosa","versicolor",...
Files Plots Packages Help Viewer Presentation


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
exp-6.R
22 dataTest
23 # Feature Scaling
24 train_scale <- scale(dataTrain[, 1:4])
25 test_scale <- scale(dataTest[, 1:4])
Console Terminal Background Jobs
versicolor 0 17 3
virginica 0 0 20
> # Model Evaluation
> confusionMatrix(cm)
Confusion Matrix and Statistics
y_pred
setosa versicolor virginica
setosa 20 0 0
versicolor 0 17 3
virginica 0 0 20
Overall Statistics
Accuracy : 0.95
95% CI : (0.8608, 0.9896)
No Information Rate : 0.3833
P-Value [Acc > NIR] : < 2.2e-16
Kappa : 0.925
Mcnemar's Test P-Value : NA
Statistics by Class:
Class: setosa Class: versicolor Class: virginica
Sensitivity 1.0000 1.0000 0.8696
Specificity 1.0000 0.9302 1.0000
Pos Pred Value 1.0000 0.8500 1.0000
Neg Pred Value 1.0000 1.0000 0.9250
Prevalence 0.3333 0.2833 0.3833
Detection Rate 0.3333 0.2833 0.3333
Detection Prevalence 0.3333 0.3333 0.3333
Balanced Accuracy 1.0000 0.9651 0.9348
Environment History Connections Tutorial
Data
classifier_cl List of 5
Data 5 obs. of 5 variables
Data1 4 obs. of 4 variables
dataTest 60 obs. of 5 variables
dataTrain 90 obs. of 5 variables
Groceries Formal class transactions
iris_data 150 obs. of 5 variables
Mushroom Large transactions (8124 elements, 1.3 ...
N 5 obs. of 5 variables
rules Formal class rules
test_scale num [1:60, 1:4] -0.942 -1.381 -0.614 -1...
train_scale num [1:90, 1:4] -1.14 -1.53 -1 -1.53 -1...
Values
avg_sleep_hour num [1:5] 20 21 23 24 25
city chr [1:5] "Delhi" "Garhwal" "Jaipur" "Guha...
cm 'table' int [1:3, 1:3] 20 0 0 0 17 0 0 3 20
country chr [1:4] "india" "india" "india" "india"
height num [1:5] 180 175 177 172 169
name chr [1:5] "Nabha" "Nikhil" "Yash" "DJ" "Ar...
rating int [1:5] 1 2 3 4 5
spl logi [1:5] FALSE TRUE FALSE TRUE TRUE
y_pred Factor w/ 3 levels "setosa","versicolor",...
Files Plots Packages Help Viewer Presentation