Experiment-2.1

Student Name: Akshat Kumar **UID: 20BCS8283**

Branch: CSE **Section/Group: 20BCS-DM-705 (A)**

Semester: 6th **Date of Performance: 03 Apr 2023**

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

Aim – To perform the classification by decision tree induction using WEKA tools.

Objective-

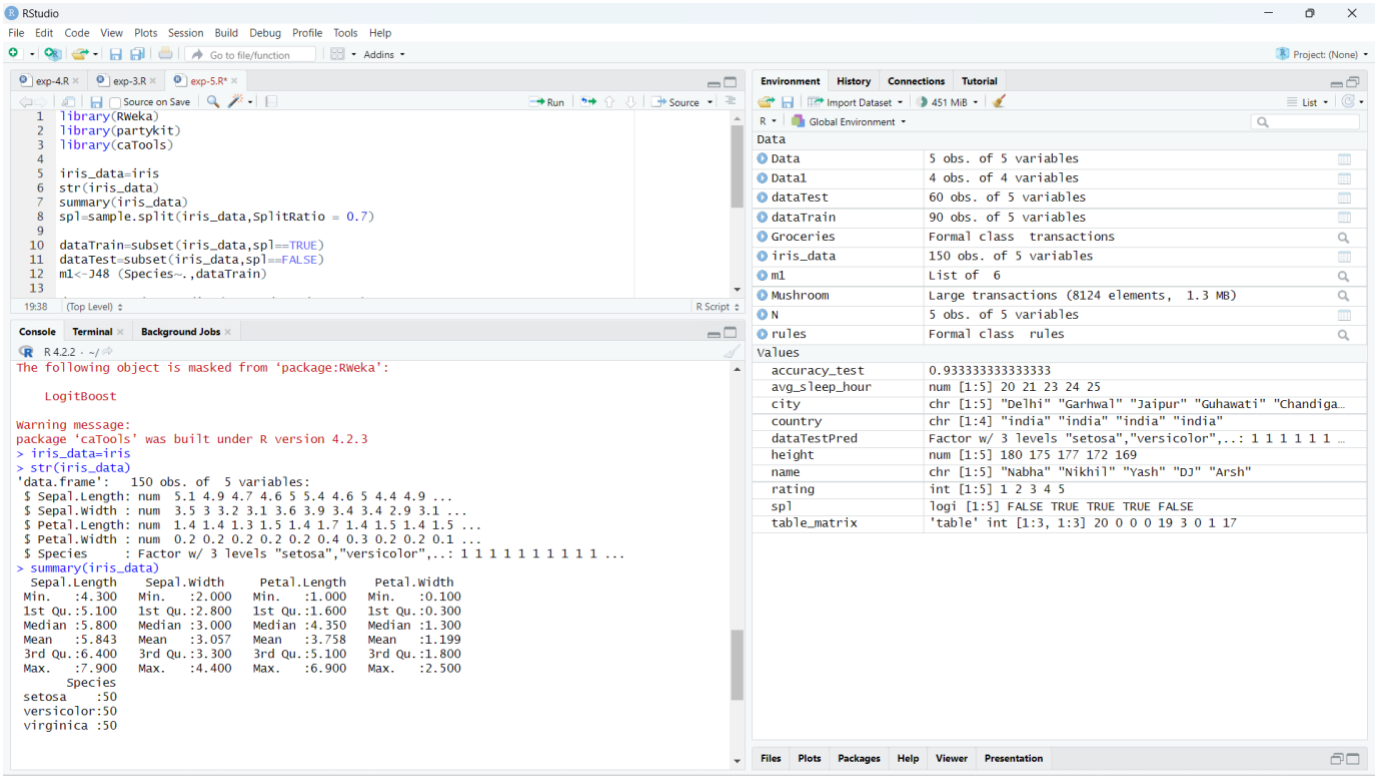
- Represent the reading of file using R studio
- Displaying the pattern on RWeka, partykit and caTools Tool.
- Demonstration of Decision Tree.

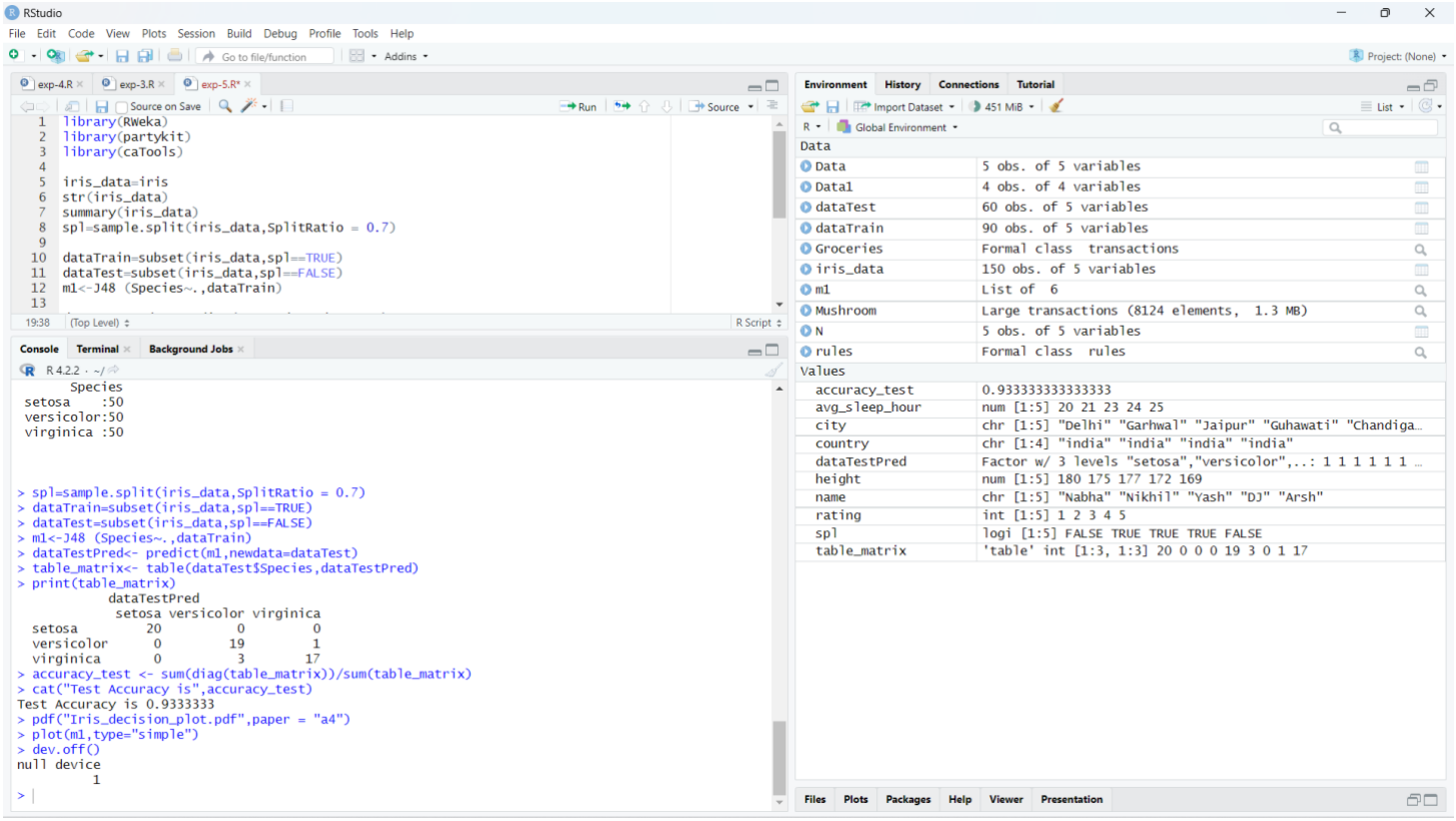
Script and Output-

```
library(RWeka)
library(partykit)
library(caTools)

iris_data=iris
str(iris_data)
summary(iris_data)
spl=sample.split(iris_data,SplitRatio = 0.7)

dataTrain=subset(iris_data,spl==TRUE)
dataTest=subset(iris_data,spl==FALSE)
m1<-J48 (Species~.,dataTrain)

dataTestPred<- predict(m1,newdata=dataTest)
table_matrix<- table(dataTest$Species,dataTestPred)
print(table_matrix)
```

```
accuracy_test <- sum(diag(table_matrix))/sum(table_matrix)
cat("Test Accuracy is",accuracy_test)
pdf("Iris_decision_plot.pdf",paper = "a4")

plot(m1,type="simple")

  dev.off()
```

Output-




RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
exp-4.R
exp-3.R
exp-5.R*
1 library(RWeka)
2 library(partykit)
3 library(caTools)
5 iris_data=iris
6 str(iris_data)
7 summary(iris_data)
8 spl=sample.split(iris_data,SplitRatio = 0.7)
10 dataTrain=subset(iris_data,spl==TRUE)
11 dataTest=subset(iris_data,spl==FALSE)
12 m1<-J48 (Species~.,dataTrain)
Console Terminal Background Jobs
R 4.2.2 · ~/
Species
setosa :50
versicolor:50
virginica :50
> spl=sample.split(iris_data,SplitRatio = 0.7)
> dataTrain=subset(iris_data,spl==TRUE)
> dataTest=subset(iris_data,spl==FALSE)
> m1<-J48 (Species~.,dataTrain)
> dataTestPred<- predict(m1,newdata=dataTest)
> table_matrix<- table(dataTest$Species,dataTestPred)
> print(table_matrix)
dataTestPred
setosa versicolor virginica
setosa 20 0 0
versicolor 0 19 1
virginica 0 3 17
> accuracy_test <- sum(diag(table_matrix))/sum(table_matrix)
> cat("Test Accuracy is",accuracy_test)
Test Accuracy is 0.9333333
> pdf("Iris_decision_plot.pdf",paper = "a4")
> plot(m1,type="simple")
> dev.off()
null device
1
Environment History Connections Tutorial
Import Dataset
Global Environment
Data
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
m1 List of 6
Mushroom Large transactions (8124 elements, 1.3 MB)
N 5 obs. of 5 variables
rules Formal class rules
Values
accuracy_test 0.933333333333333
avg_sleep_hour num [1:5] 20 21 23 24 25
city chr [1:5] "Delhi" "Garhwal" "Jaipur" "Guhawati" "Chandiga...
country chr [1:4] "india" "india" "india" "india"
dataTestPred Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 ...
height num [1:5] 180 175 177 172 169
name chr [1:5] "Nabha" "Nikhil" "Yash" "DJ" "Arsh"
rating int [1:5] 1 2 3 4 5
spl logi [1:5] FALSE TRUE TRUE TRUE FALSE
table_matrix 'table' int [1:3, 1:3] 20 0 0 0 19 3 0 1 17
Files Plots Packages Help Viewer Presentation