



# DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

## Experiment-2.1

Student Name: Akshat Kumar

UID: 20BCS8283

Branch: CSE

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

Semester: 6<sup>th</sup>

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)
```



# DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

```
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-

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for loading libraries, splitting the iris dataset, training a model, and generating a PDF plot.
- Console:** Shows the execution output, including a warning message about the 'caTools' package and a detailed summary of the iris dataset.
- Environment:** Lists the objects in the global environment, including 'Data', 'dataTrain', 'dataTest', 'iris\_data', 'm1', 'Mushroom', 'N', 'rules', and 'values'.
- Values:** Displays the values for the 'accuracy\_test' variable, which is 0.933333333333333.

**Console Output:**

```
The following object is masked from 'package:Rwaka':
  LogitBoost

Warning message:
package 'caTools' 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 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 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:**

Object	Description
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:**

Variable	Value
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 ...
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
split	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



# DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

RStudio interface showing R code execution and environment details.

```
1 library(Rweka)
2 library(partykit)
3 library(caTools)
4
5 iris_data=iris
6 str(iris_data)
7 summary(iris_data)
8 spl=sample.split(iris_data,SplitRatio = 0.7)
9
10 dataTrain=subset(iris_data,spl==TRUE)
11 dataTest=subset(iris_data,spl==FALSE)
12 m1<-J48 (Species~,dataTrain)
13
```

Console output:

```
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 details:

Object	Class	Attributes
Data	data.frame	5 obs. of 5 variables
Data1	data.frame	4 obs. of 4 variables
dataTest	data.frame	60 obs. of 5 variables
dataTrain	data.frame	90 obs. of 5 variables
Groceries	transactions	Formal class transactions
iris_data	data.frame	150 obs. of 5 variables
m1	list	List of 6
Mushroom	transactions	Large transactions (8124 elements, 1.3 MB)
N	data.frame	5 obs. of 5 variables
rules	rules	Formal class rules

Values:

Variable	Value
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 ...
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