



## Experiment 5

**Student Name:** Praduman Kumar

**Branch:** BE-CSE

**Semester:** 6

**Subject Name:** Data Mining Lab

**UID:** 20BCS9446

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

**Subject Code:** 20CSP-376

**Date of Performance:** 11-04-2023

**1. Aim/Overview of the practical:** To perform the classification by decision tree induction.

**2. Tools used:** RStudio

**3. Code:**

```
library(RWeka)
library(partykit)
library(caTools)
setwd("C:\\Users\\hp\\Documents\\DATA MINING
CODES\\EXPERIMENT 5")
getwd()
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)
summary(m1)
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()
```

## 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 5/
> library(Rweka)
> library(partykit)
> library(caTools)
> setwd("C:\\Users\\hp\\Documents\\DATA MINING CODES\\EXPERIMENT 5")
> getwd()
[1] "C:/Users/hp/Documents/DATA MINING CODES/EXPERIMENT 5"
> 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 1 ...
> summary(iris_data)
 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
> spl = sample.split(iris_data, splitRatio = 0.7)
> dataTrain = subset(iris_data, spl==TRUE)
> dataTest = subset(iris_data, spl==FALSE)
> m1 <- J48(Species~., dataTrain)
> summary(m1)

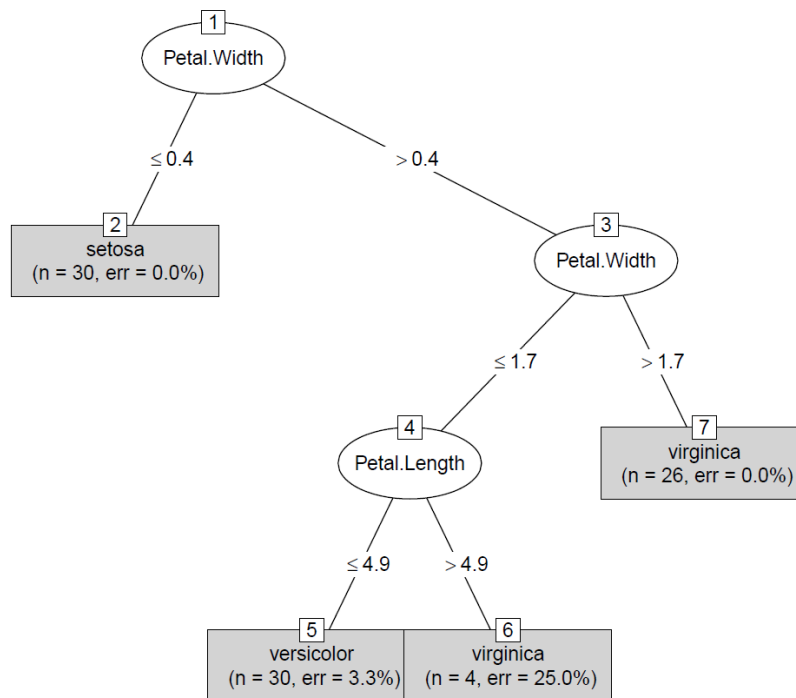
=== Summary ===

Correctly Classified Instances 88 97.7778 %
Incorrectly Classified Instances 2 2.2222 %
Kappa statistic 0.9667
Mean absolute error 0.0254
Root mean squared error 0.1128
Relative absolute error 5.7222 %
Root relative squared error 23.9212 %
Total Number of Instances 90

=== Confusion Matrix ===

 a b c <-- classified as
30 0 0 | a = setosa
0 29 1 | b = versicolor
0 1 29 | c = virginica
> dataTestPred <- predict(m1, newdata = dataTest)
> table_matrix <- table(dataTest$Species, dataTestPred)
> print(table_matrix)
dataTestPred
setosa versicolor virginica
setosa 18 2 0
versicolor 0 18 2
virginica 0 0 20
> 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
```

Iris\_decision\_plot file:



## 5. Observation:

- Learnt how to use R and create a file in Rstudio.
- Learnt how to load dataset iris in Rstudio.
- Learned How to Implement Decision Tree in R Studio.