Experiment 5

Student Name: Anshuman Singh	**UID:** 20BCS2665
Branch: CSE	**Section/Group: 902**-A
Semester: 6th	**Date of Performance:** 31/03/23
Subject Name: Data Mining Lab	**Subject Code:** 20CSP-376

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

Code -

```
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) 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)
#initate PDF File pdf("Iris_decision_plot.pdf",paper
="a4") plot(m1, type="simple")
```

#Close PDF File dev.off()

Output -

```
Correctly Classified Instances          89              98.8889 %
Incorrectly Classified Instances        1                1.1111 %
Kappa statistic                        0.9833
Mean absolute error                    0.0119
Root mean squared error                0.077
Relative absolute error                 2.6667 %
Root relative squared error           16.3299 %
Total Number of Instances              90

=== Confusion Matrix ===

  a  b  c   <-- classified as
 30  0  0 |  a = setosa
  0 29  1 |  b = versicolor
  0  0 30 |  c = virginica
>
> dataTestPred <- predict(m1, newdata = dataTest)
> table_matrix <-table(dataTest$Species, dataTestPred)
>
> print(table_matrix)
            dataTestPred
            setosa versicolor virginica
  setosa        20          0         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:   0.9666667>
> #initate PDF File
> pdf("Iris_decision_plot.pdf",paper ="a4")
```

R ▾ | 🔷 Global Environment ▾ | 🔍

Data

▶ dataTest	150 obs. of 5 variables	▦
▶ dataTrain	150 obs. of 5 variables	▦
▶ Groceries	Formal class transactions	🔍
▶ iris_data	150 obs. of 5 variables	▦
▶ m1	List of 6	🔍
▶ rules	Formal class rules	🔍

Values

accuracy_Test	0.98
dataTestPred	Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 …
spl	logi [1:5] FALSE TRUE FALSE TRUE TRUE
table_matrix	'table' int [1:3, 1:3] 50 0 0 0 49 2 0 1 48