



Experiment-2.1

Student Name: Himanshu

UID: 20BCS7944

Branch: CSE

Section: 905/A

Semester: 6

Date of Performance: 30/03/2023

Subject Name: Data Mining Lab

Subject Code: 20CSP-376

1) **Aim:**

To perform the classification by decision tree induction.

2) **Objective:**

Classifying the data using decision tree algorithm.

3) **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)
```

```
pdf("Iris_decision_plot.pdf", paper="a4")
```

```
plot(m1, type="simple")
```

```
dev.off()
```

4) Output:

```

Console Terminal Background Jobs
R 4.2.3 ~

[Workspace loaded from ~/.RData]

> source("~/Documents/DMClassWork/exp5.R", echo=TRUE)

> library(RWeka)

> library(partykit)
Loading required package: grid
Loading required package: libcoin
Loading required package: mvtnorm

> library(caTools)

Attaching package: 'caTools'

The following object is masked from 'package:RWeka':

  LogitBoost

> 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

> summary(m1)

=== Summary ===

Correctly Classified Instances          90           100    %
Incorrectly Classified Instances         0            0    %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                   0    %
Root relative squared error              0    %
Total Number of Instances               90

```

```
Total Number of Instances          90

=== Confusion Matrix ===
  a  b  c   <-- classified as
30  0  0 | a = setosa
 0 30  0 | 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      18         2         0
versicolor   0        19         1
virginica    0         5        15

> accuracy_Test <- sum(diag(table_matrix)) / sum(table_matrix)

> cat("Test Accuracy is: ", accuracy_Test)
Test Accuracy is: 0.8666667
> # Initiate PDF File
> pdf("Iris_decision_plot.pdf", paper="a4")

> plot(m1, type="simple")

> #Close PDF file
> dev.off()
null device
      1
> |
```