Name: Sumit Singh
Roll No: 380
Class: TYBSC CS A
Subject: Data Science
Practical No : 6

Aim: Demonstration of Decision Tree.
Code:
Steps:
Step1: click on packages and set cran mirror.
Step2: click on packages and select install packages and install 3 packages (rpart,tree,rattle)
Step3:(OPTIONAL Application for version 4.2)
install.packages("rpart") install.packages("tree")
install.packages("rattle")

```
[Previously saved workspace restored]

> chooseCRANmirror()
> utils:::menuInstallPkgs()

  There is a binary version available but the source version is later:
      binary source needs_compilation
rpart 4.1.16 4.1.19              TRUE

  Binaries will be installed
trying URL 'http://ftp.ussg.iu.edu/CRAN/bin/windows/contrib/4.0/rpart_4.1.16.zi$
Content type 'application/zip' length 982973 bytes (959 KB)
downloaded 959 KB

package 'rpart' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\admin\AppData\Local\Temp\Rtmp2BLQPi\downloaded_packages
> x=read.csv("C:/Users/admin/Desktop/weather1.csv")
> x
    outlook temp humidity windy play.golf
1     rainy  hot     high FALSE        no
2     rainy  hot     high  TRUE        no
3  overcast  hot     high FALSE       yes
```

Step4: **Create an excel data save it with .csv extension.**

Code:

Read excel data in rstudio

> x=read.csv("C:/weather1.csv")
> x

```
> x=read.csv("C:/Users/admin/Desktop/weather1.csv")
> x
    outlook temp humidity windy play.golf
1     rainy  hot     high FALSE        no
2     rainy  hot     high  TRUE        no
3  overcast  hot     high FALSE       yes
4     sunny mild     high FALSE       yes
5     sunny cool   normal FALSE       yes
6     sunny cool   normal  TRUE        no
7  overcast cool   normal  TRUE       yes
8     rainy mild     high FALSE       yes
9     rainy cool   normal FALSE       yes
10    sunny mild   normal FALSE       yes
11    rainy mild   normal  TRUE       yes
12 overcast mild     high  TRUE       yes
13 overcast  hot   normal FALSE       yes
14    sunny mild     high  TRUE        no
> sample_weather=sample(nrow(x),.7*nrow(x))
> weather_tr=x[sample_weather,]
> weather_test=x[-sample_weather,]
> weather_test
    outlook temp humidity windy play.golf
2     rainy  hot     high  TRUE        no
3  overcast  hot     high FALSE       yes
```

Create sample partition of the excel data

> sample_weather=sample(nrow(x),.7*nrow(x))

Create a weather partition for training

> weather_tr=x[sample_weather,]

Create a weather partition for testing

> weather_test=x[-sample_weather,]

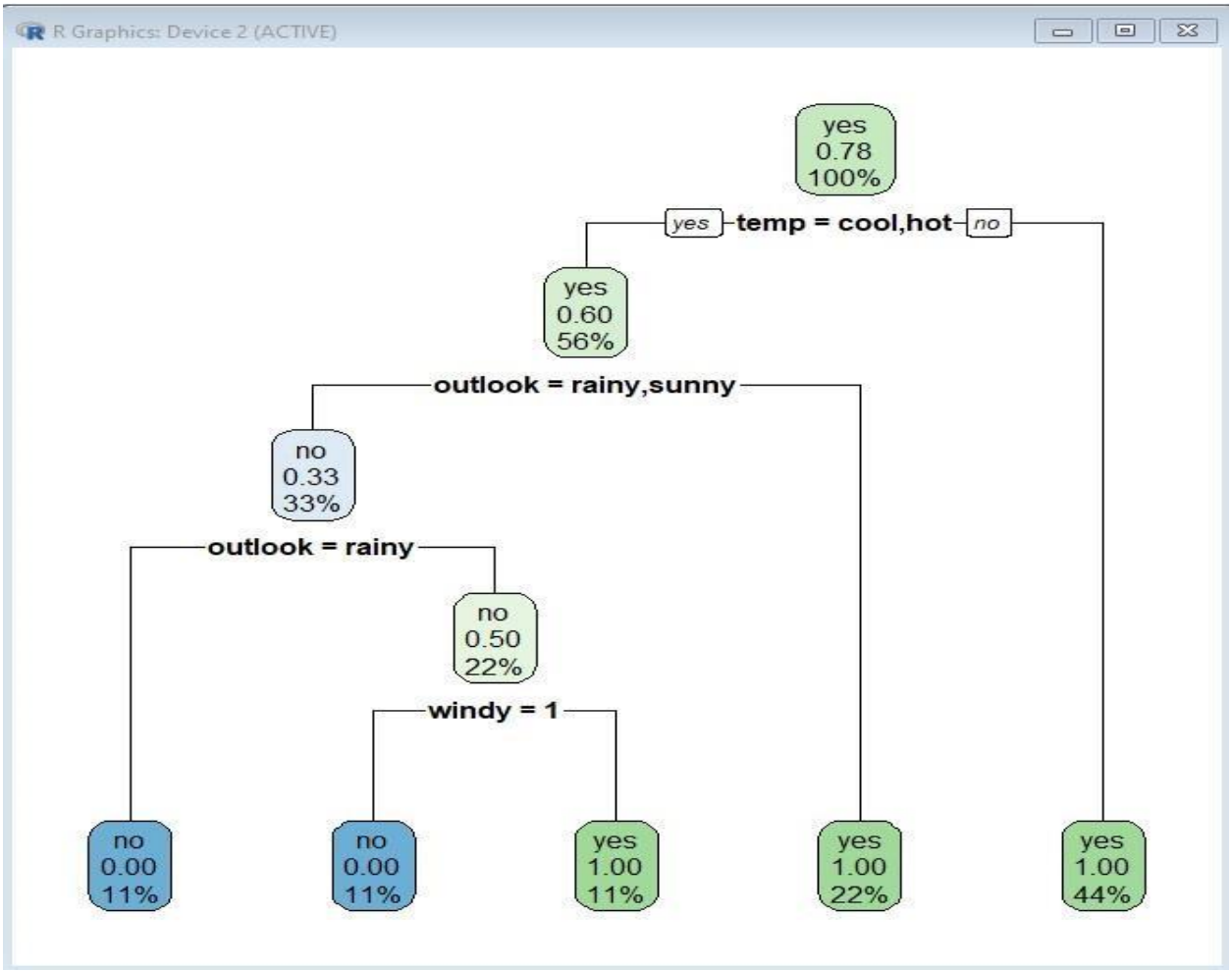
> weather_test

Call rpart packages

> library(rpart)

> library(rpart.plot) **Plot tree**

dtreemod=rpart(play.golf~.,data=weather_tr,method="class",control=rpart.control(minsplit=1,min bucket=1)) rpart.plot(dtreemod)



Predict Tree:

> p=predict(dtreemod,weather_test,type="class")
> weather_test
> table(weather_test$play.golf,p)

```
> p=predict(dtreemod,weather_test,type="class")
> weather_test
    outlook temp humidity windy play.golf
2     rainy  hot     high  TRUE        no
3  overcast  hot     high FALSE       yes
6     sunny cool   normal  TRUE        no
13 overcast  hot   normal FALSE       yes
14    sunny mild     high  TRUE        no
> table(weather_test$play.golf,p)
     p
      no yes
  no   1   2
  yes  2   0
>
```

Printing rules with rpart.rules

rpart.rules(dtreemod) play.golf

0.00 when temp is hot
1.00 when temp is cool or mild
>

Regression Tree:

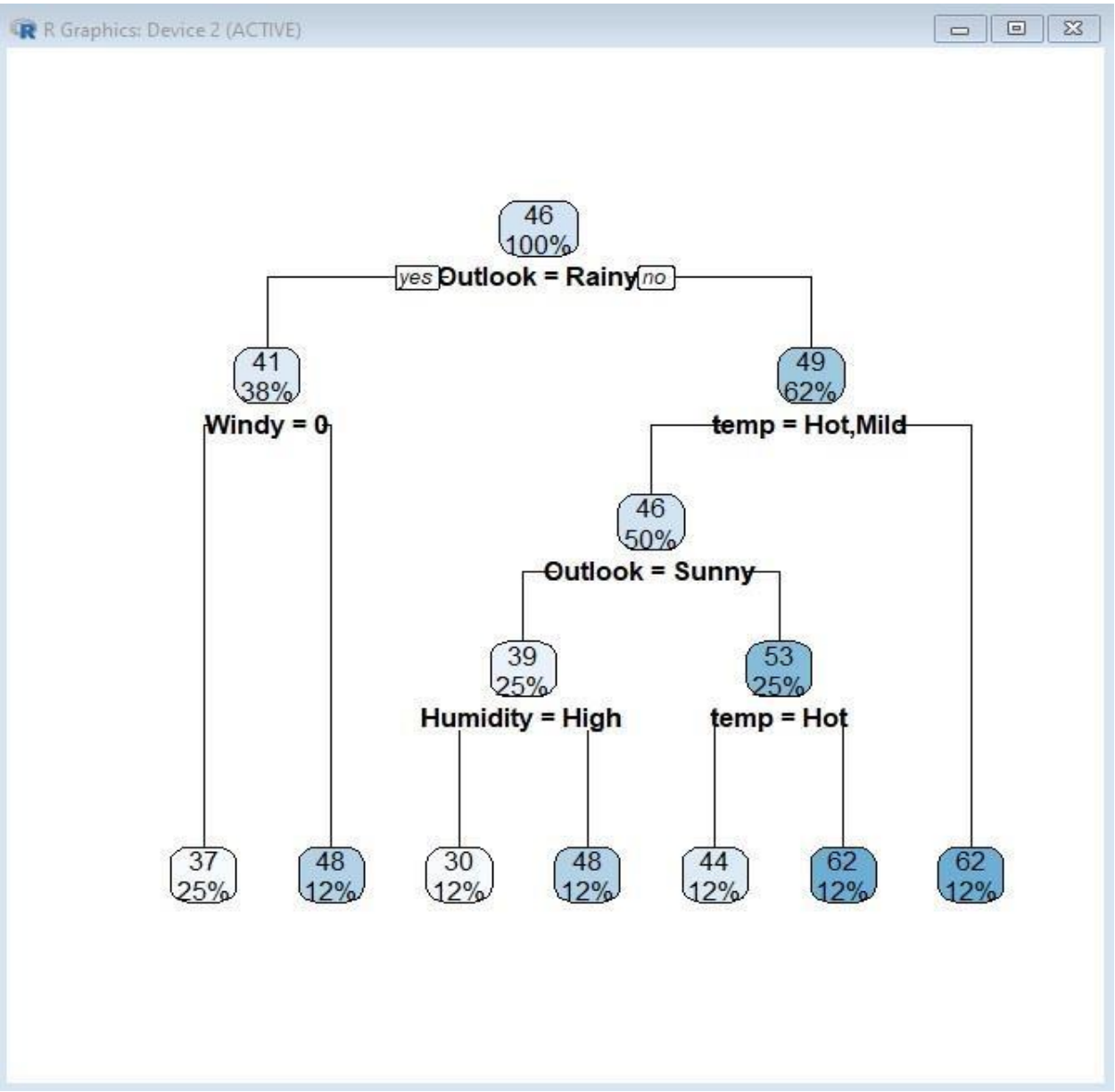
> x2=read.csv("C:/Users/admin/Desktop/weather2.csv") >
x2

```
> x2=read.csv("C:/Users/admin/Desktop/weather2.csv")
> x2
    Outlook  temp Humidity Windy Hours.Played
1     Rainy  Hot      High FALSE           26
2     Rainy  Hot      High  TRUE           30
3  Overcast  Hot      High FALSE           48
4     Sunny  Mild     High FALSE           46
5     Sunny  Cool   Normal FALSE           62
6  Overcast  Cool   Normal  TRUE           43
7     Rainy  Mild     High FALSE           36
8     Rainy  Cool   Normal FALSE           38
9     Sunny  Mild   Normal FALSE           48
10    Rainy  Mild   Normal  TRUE           48
11 Overcast  Mild     High  TRUE           62
12 Overcast   Hot   Normal FALSE           44
13    Sunny  Mild     High  TRUE           30
```

weather_tr2=x2[S2,] >
s2=sample(nrow(x),.7*nrow(x)) > weather_tr2=x2[s2,]
> weather_test2=x2[-s2,]
> weather_test2

```
> weather_tr2=x2[S2,]
> s2=sample(nrow(x),.7*nrow(x))
> weather_tr2=x2[s2,]
> weather_test2=x2[-s2,]
> weather_test2
   Outlook  temp Humidity Windy Hours.Played
1    Rainy  Hot      High FALSE           26
2    Rainy  Hot      High  TRUE           30
3 Overcast  Hot      High FALSE           48
4    Sunny  Mild     High FALSE           46
6 Overcast  Cool   Normal  TRUE           43
```

dtreemod2=rpart(Hours.Played~.,data=weather_tr2,method="anova",control=rpart.control(minsplit=1,minbucket=1)) > rpart.rules(dtreemod2)



Prediction:

> actuals_preds<- data.frame(cbind(actuals=weather_test2$Hours.played,predicts=p)) >

actuals_preds

```
> actuals_preds<- data.frame(cbind(actuals=weather_test2$Hours.played,predicts=p))
> actuals_preds
   predicts
2         1
7         2
9         1
12        2
14        2
> 
```

Conclusion: Hence we successfully implemented decision tree.