Practical no 6

Aim: Demonstration of Decision Tree

Theory:

A decision tree is a flowchart-like tree structure where an internal node represents feature (or attribute), the

branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision

tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in

recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's

visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees

are easy to understand and interpret.

Decision Tree is a white box type of ML algorithm. It shares internal decision-making logic, which is not

available in the black box type of algorithms such as Neural Network. Its training time is faster compared to the

neural network algorithm. The time complexity of decision trees is a function of the number of records and

number of attributes in the given data. The decision tree is a distribution-free or non-parametric method,

which does not depend upon probability distribution assumptions. Decision trees can handle high dimensional

data with good accuracy.

How does the Decision Tree Algorithm Work?

The basic idea behind any decision tree algorithm is as follows:

1. Select the best attribute using Attribute Selection Measures (ASM) to split the records.

2. Make that attribute a decision node and breaks the dataset into smaller subsets.

Steps:

- Open Excel create a data
- Save it as .CSV(MS-DOS)
- Keep the dataset and R code in a same folder.

Datasets:

	A	B	C	D	E	F	G
1	outlook	temp	humidity	windy	play golf		
2	rainy	hot	high	FALSE	no		
3	rainy	hot	high	TRUE	no		
4	overcast	hot	high	FALSE	yes		
5	sunny	mild	high	FALSE	yes		
6	sunny	cool	normal	FALSE	yes		
7	sunny	cool	normal	TRUE	no		
8	overcast	cool	normal	TRUE	yes		
9	rainy	mild	high	FALSE	yes		
10	rainy	cool	normal	FALSE	yes		
11	sunny	mild	normal	FALSE	yes		
12	rainy	mild	normal	TRUE	yes		
13	overcast	mild	high	TRUE	yes		
14	overcast	hot	normal	FALSE	yes		
15	sunny	mild	high	TRUE	no		
16							
17							

	A	B	C	D	E	F
1	Outlook	temp	Humidity	Windy	Hours Played	
2	Rainy	Hot	High	FALSE	26	
3	Rainy	Hot	High	TRUE	30	
4	Overcast	Hot	High	FALSE	48	
5	Sunny	Mild	High	FALSE	46	
6	Sunny	Cool	Normal	FALSE	62	
7	Overcast	Cool	Normal	TRUE	43	
8	Rainy	Mild	High	FALSE	36	
9	Rainy	Cool	Normal	FALSE	38	
10	Sunny	Mild	Normal	FALSE	48	
11	Rainy	Mild	Normal	TRUE	48	
12	Overcast	Mild	High	TRUE	62	
13	Overcast	Hot	Normal	FALSE	44	
14	Sunny	Mild	High	TRUE	30	
15						
16						

Code:

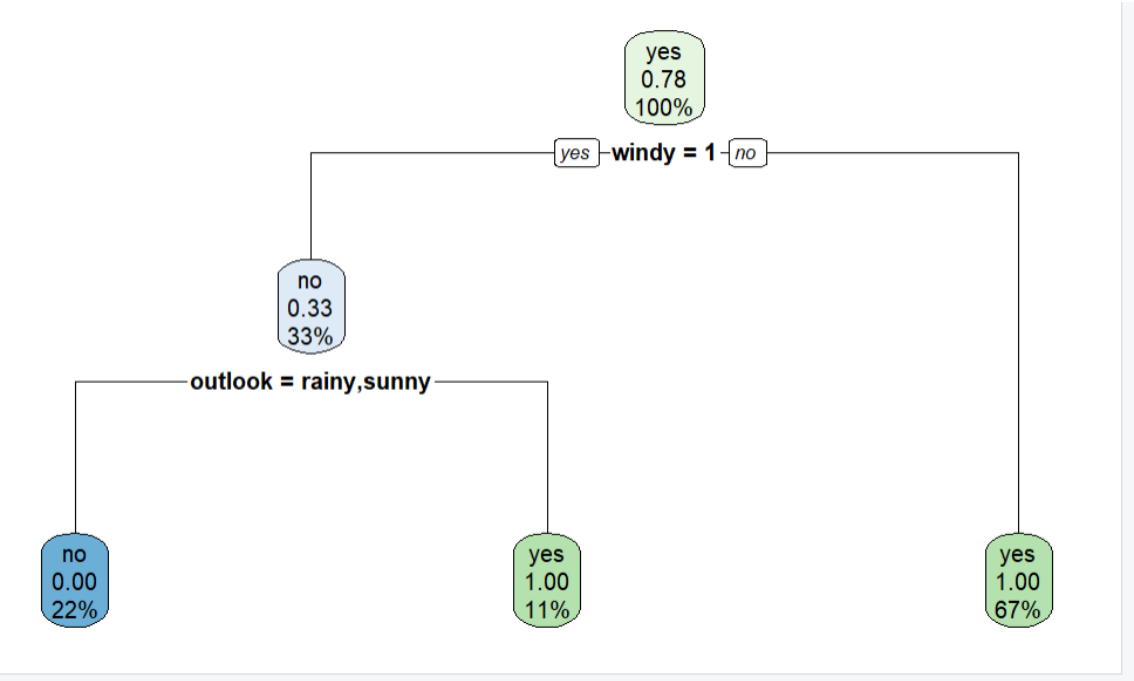
```
x=read.csv("C:/Users/admin/Downloads/Desktop/Materials/COMPUTER SCIENCE/Sem 6/Data Science/All Pracs/weather1.csv")
x
sample_weather=sample(nrow(x),.7*nrow(x))
weather_tr=x[sample_weather,]
weather_test=x[-sample_weather,]
weather_test
library(rpart)
library(rpart.plot)
dtreemod=rpart(play.golf~.,data=weather_tr,method="class",control=rpart.control(minsplit=1,minbucket=1))
rpart.plot(dtreemod)
p=predict(dtreemod,weather_test,type="class")
weather_test
table(weather_test$play.golf,p)
x2=read.csv("C:/Users/admin/Downloads/Desktop/Materials/COMPUTER SCIENCE/Sem 6/Data Science/All Pracs/weather2.csv")
x2
weather_tr2=x2[s2,]
s2=sample(nrow(x),.7*nrow(x))
weather_tr2=x2[s2,]
weather_test2=x2[-s2,]
weather_test2
dtreemod2=rpart(Hours.Played~.,data=weather_tr2,method="anova",control=rpart.control(minsplit=1,minbucket=1))
rpart.rules(dtreemod2)
actuals_preds<- data.frame(cbind(actuals=weather_test2$Hours.played,predicts=p))
actuals_preds

```

Output:

```
> 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
1     rainy  hot     high FALSE        no
4     sunny mild     high FALSE       yes
11    rainy mild   normal  TRUE       yes
12 overcast mild     high  TRUE       yes
14    sunny mild     high  TRUE        no
> library(rpart)
> library(rpart.plot)
> dtreemod=rpart(play.golf~.,data=weather_tr,method="class",control=rpart.control(minsp
lit=1,minbucket=1))
> rpart.plot(dtreemod)
> p=predict(dtreemod,weather_test,type="class")
> weather_test
    outlook temp humidity windy play.golf
1     rainy  hot     high FALSE        no
4     sunny mild     high FALSE       yes
11    rainy mild   normal  TRUE       yes
12 overcast mild     high  TRUE       yes
14    sunny mild     high  TRUE        no
> table(weather_test$play.golf,p)
     p
      no yes
```

```
> 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
   Outlook  temp Humidity Windy Hours.Played
1    Rainy  Hot      High FALSE           26
3 Overcast  Hot      High FALSE           48
5    Sunny  Cool   Normal FALSE           62
8    Rainy  Cool   Normal FALSE           38
9    Sunny  Mild   Normal FALSE           48
> dtreemod2=rpart(Hours.Played~.,data=weather_tr2,method="anova",control=rpart.control
(minsplit=1,minbucket=1))
> rpart.rules(dtreemod2)
 Hours.Played
           30 when Outlook is Rainy or Sunny & Humidity is   High & Windy is 1
           36 when Outlook is          Rainy & Humidity is   High & Windy is 0
           44 when Outlook is       Overcast                                  & temp
s Cool or Hot
           46 when Outlook is          Sunny & Humidity is   High & Windy is 0
           48 when Outlook is Rainy or Sunny & Humidity is Normal
           62 when Outlook is       Overcast                                  & temp
s        Mild
> actuals_preds<- data.frame(cbind(actuals=weather_test2$Hours.played,predicts=p))
> actuals_preds
   predicts
1         2
4         2
11        1
```

```
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
   Outlook  temp Humidity Windy Hours.Played
1    Rainy  Hot      High FALSE           26
3 Overcast  Hot      High FALSE           48
5    Sunny  Cool   Normal FALSE           62
8    Rainy  Cool   Normal FALSE           38
9    Sunny  Mild   Normal FALSE           48
> dtreemod2=rpart(Hours.Played~.,data=weather_tr2,method="anova",control=rpart.control
(minsplit=1,minbucket=1))
> rpart.rules(dtreemod2)
 Hours.Played
           30 when Outlook is Rainy or Sunny & Humidity is   High & Windy is 1
           36 when Outlook is          Rainy & Humidity is   High & Windy is 0
           44 when Outlook is       Overcast                                  & temp
s Cool or Hot
           46 when Outlook is          Sunny & Humidity is   High & Windy is 0
           48 when Outlook is Rainy or Sunny & Humidity is Normal
           62 when Outlook is       Overcast                                  & temp
s        Mild
> actuals_preds<- data.frame(cbind(actuals=weather_test2$Hours.played,predicts=p))
> actuals_preds
   predicts
1         2
4         2
11        1
12        2
14        1
>
```



Conclusion: Hence we have successfully learnt and performed Decision Tree