Data Analytics Lab 2

24/09/19

Step 1 - Import the excel file and checking content of the data

Convert the dataset into csv file, Read.csv tells R to read csv file.

```
Data <- read.csv("DT-Credit.csv", header=TRUE, sep= " ; ")
```

Check distinct categories of Variables useing STR function

```
str(Data)
```

```
attach(Data)
```

Step 2 - Install package rpart, and click on the checkbox in front of rpart library.

Develop the DT model:

```
DT_Model <-rpart(RESPONSE~., data=Data, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))
```

minsplit: the minimum number of observations that must exist in a node for a new split
minbucket: the minimum number of observations in any terminal <leaf> node
Maxdepth: Maximum depth for any node, with the root node counted as depth 0.

Step 3 - Install packgae partykit:

```
install.packages ("partykit")
```

```
library("partykit")
```

```
plot(as.party(DT_Model))
```

```
print(DT_Model)
```

Describe the results at the end of your scripts.

Working on binary response variable

Step 4. Change the response to Y/N answers.

Target=ifelse(RESPONSE==1,'Y','N')

Add Target to main dataset using

```
Data <- data.frame(Data,Target)
```

Check the content of the file

```
str(Data)
```

Drop the original variable (RESPONSE)

```
Data1=Data[,-32]
```

Develop the DT model again

```
DT_Model1<-rpart(Target~., data=Data1, control=rpart.control(minsplit=60, minbucket=30, maxdepth=4 ))
```

```
plot(as.party(DT_Model1)   )
```

```
print(DT_Model1)
```

Change the control parameters and see the change in the output (minsplit=60, minbucket=30, maxdepth=8).

Step 5- Procedure of Pruning

```
DT_Model2<-rpart(Target~., data=Data1, control=rpart.control)

plot(as.party(DT_Model2))
```

The following line fitted tree's CP table (Matrix of Information on optimal pruning given Complexity Parameter). Look where do you see the least error.

```
print(DT_Model2$cptable)
```

The line below automatically picks up the least error tree

```
opt <- which.min(DT_Model2$cptable [, "xerror"])
```

Step 6 - Pruning the tree to the least xerror

```
cp <- DT_Model2$cptable [opt,"CP"]

DT_Model_pruned <- prune(DT_Model2, cp=cp)

plot(as.party(DT_Model_pruned))
```