## Predicting Customer Churn: Decision Tree Model in R

#Read the data from the working directory, create your own working directly to read the dataset.

data1 <- read.csv ("C:/Users/Deep/Desktop/data/

churn\_dataset.csv",header=TRUE,sep=",")

data2<-data.frame(data1)

#Perform exploratory data analysis to know about the data

#Display top 6 rows of dataset to see how data look like

head (data2)

#Display bottom 6 rows

tail(data2)

#Describe the structure of data

str(data2)

#Display the column name of the data

names(data2)

#Display the datatype

class(data2)

#Display the summary or descriptive statistics of the data

summary(data2$Monthly\_Charges)

#Set seed in order to reproduce the sample

set.seed(123)

#Splitting data set into training and testing dataset in 70:30

install.packages("caTools")

library(caTools)

sample <- sample.split(data2$Churn,SplitRatio=0.70)

#No of observations in train dataset

train\_data <- subset(data2,sample==TRUE)

# No of observations in test dataset

test\_data <- subset(data2,sample==FALSE)

#Growing full decision tree model by using rpart on training data

install.packages("rpart")

library(rpart)

**Program1:**

churn\_model <- rpart(Churn ~ ., data=train\_data,

method = "class", parms = list(split = 'information'),cp=-1)

churn\_model

#Installing rattle package and helper package like rpart.plot and RcolorBrewer package in order to create the decision tree

install.packages("rattle")

library(rattle)

install.packages("rpart.plot")

library(rpart.plot)

install.packages("RColorBrewer")

library(RColorBrewer)

#Creating the decision tree plot by using fancyRpartPlot function

fancyRpartPlot(churn\_model)

#List of options that control the rpart algorithm and are ways preventing a model from overfitting.

**Program1.1**

tree\_model <- rpart(Churn ~ ., data=train\_data,

method = "class",parms = list(split = 'information'), maxdepth = 3, minsplit = 2, minbucket = 2 )

tree\_model

#Creating the decision tree plot

fancyRpartPlot(tree\_model)

#Plotting it inorder to find cp value for pruning

plotcp(tree\_model)

#Finding the cross validation results

printcp(tree\_model)

**Program 1.2**

#Pruned model

prune\_model <- prune(tree\_model, cp=.02)

prune\_model

#Creating the pruned decision tree plot

fancyRpartPlot(prune\_model)

#Predicting the model using test data

test\_data$Churn\_Class <- predict(prune\_model,

newdata = test\_data, type="class")

#Display the confusion matrix or classification table

table(test\_data$Churn ,test\_data$Churn\_Class)

#Predicting the probability matrix using test data

pred <- predict(prune\_model, newdata = test\_data, type="prob")

pred1 <- data.frame(pred)