MIS40970 Data Mining - Assignment 3 Classification

```
#rm() is used to remove other objects from the environment
rm(list=ls())

#To check the working directory
getwd()
```

```
## [1] "G:/R_programs_git/R_Progams/Classification/Classification"
```

```
#A specific working drectory needs to be set for the loading of dataset
setwd("G:/R_programs_git/R_Progams/Classification/Classification")
```

Q4. Install R package "C50". Import customer churn dataset (churn) using data() function.
Examine the churnTrain dataset. Using R run a decision-tree classification algorithm of your choice constructing a full unpruned tree and a pruned tree. Compare classification results of the pruned and unpruned trees generated.

```
#install.packages('C50')

library(C50)
```

```
## Warning: package 'C50' was built under R version 3.3.3
```

```
#To examine datasets in package C50
data()
```

After examining data function, it has been found out that there are two datasets in 'Customer Churn Data' (C50 package), which are churnTest and churnTrain. To import datasets in C50 package, use command "data(churn)".

```
data(churn)

print("----------------------------Summary Table : churnTrain----------------------")
```

```
## [1] "----------------------------Summary Table : churnTrain----------------------"
```

```
#To understand dataset churnTrain
summary(churnTrain)
```

```
##      state         account_length          area_code    international_plan
##  WV     : 106   Min.   :  1.0   area_code_408: 838   no :3010
##  MN     :  84   1st Qu.: 74.0   area_code_415:1655   yes: 323
##  NY     :  83   Median :101.0   area_code_510: 840
##  AL     :  80   Mean   :101.1
##  OH     :  78   3rd Qu.:127.0
##  OR     :  78   Max.   :243.0
##  (Other):2824
##  voice_mail_plan number_vmail_messages total_day_minutes total_day_calls
##  no :2411        Min.   : 0.000        Min.   :  0.0     Min.   :  0.0
##  yes: 922        1st Qu.: 0.000        1st Qu.:143.7     1st Qu.: 87.0
##                  Median : 0.000        Median :179.4     Median :101.0
```

```
##                  Mean   : 8.099        Mean   :179.8     Mean   :100.4
##                  3rd Qu.:20.000        3rd Qu.:216.4     3rd Qu.:114.0
##                  Max.   :51.000        Max.   :350.8     Max.   :165.0
##
##  total_day_charge total_eve_minutes total_eve_calls total_eve_charge
##  Min.   : 0.00    Min.   :  0.0     Min.   :  0.0   Min.   : 0.00
##  1st Qu.:24.43    1st Qu.:166.6     1st Qu.: 87.0   1st Qu.:14.16
##  Median :30.50    Median :201.4     Median :100.0   Median :17.12
##  Mean   :30.56    Mean   :201.0     Mean   :100.1   Mean   :17.08
##  3rd Qu.:36.79    3rd Qu.:235.3     3rd Qu.:114.0   3rd Qu.:20.00
##  Max.   :59.64    Max.   :363.7     Max.   :170.0   Max.   :30.91
##
##  total_night_minutes total_night_calls total_night_charge
##  Min.   : 23.2       Min.   : 33.0     Min.   : 1.040
##  1st Qu.:167.0       1st Qu.: 87.0     1st Qu.: 7.520
##  Median :201.2       Median :100.0     Median : 9.050
##  Mean   :200.9       Mean   :100.1     Mean   : 9.039
##  3rd Qu.:235.3       3rd Qu.:113.0     3rd Qu.:10.590
##  Max.   :395.0       Max.   :175.0     Max.   :17.770
##
##  total_intl_minutes total_intl_calls total_intl_charge
##  Min.   : 0.00      Min.   : 0.000   Min.   :0.000
##  1st Qu.: 8.50      1st Qu.: 3.000   1st Qu.:2.300
##  Median :10.30      Median : 4.000   Median :2.780
##  Mean   :10.24      Mean   : 4.479   Mean   :2.765
##  3rd Qu.:12.10      3rd Qu.: 6.000   3rd Qu.:3.270
##  Max.   :20.00      Max.   :20.000   Max.   :5.400
##
##  number_customer_service_calls churn
##  Min.   :0.000                 yes: 483
##  1st Qu.:1.000                 no :2850
##  Median :1.000
##  Mean   :1.563
##  3rd Qu.:2.000
##  Max.   :9.000
##
```