KNN, Random forest, Logistic Regression &SVM

S. Varatharajan

Tue Oct 16 21:35:29 2018

```
library(ggplot2)
library(reshape2)
library(corrplot)

## corrplot 0.84 loaded

library(e1071)
library(caret)

## Loading required package: lattice

library(rpart)
library(C50)
library(party)

## Loading required package: grid

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

#library(partykit)
library(randomForest)

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin

library(ROCR)

## Loading required package: gplots

##
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
##
##     lowess

library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:randomForest':
##
##     combine

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(car)

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:modeltools':
##
##     Predict

setwd("C:/Users/tsraj/Desktop/Acadgild project")
library(readr)
churn <- read_csv("churn.csv")

## Parsed with column specification:
## cols(
```

```
##   .default = col_integer(),
##   `Day Mins` = col_double(),
##   `Eve Mins` = col_double(),
##   `Night Mins` = col_double(),
##   `Intl Mins` = col_double(),
##   `Day Charge` = col_double(),
##   `Eve Charge` = col_double(),
##   `Night Charge` = col_double(),
##   `Intl Charge` = col_double(),
##   State = col_character(),
##   Phone = col_character()
## )

## See spec(...) for full column specifications.

View(churn)
library(doParallel)

## Loading required package: foreach

## Loading required package: iterators

## Loading required package: parallel

registerDoParallel()
set.seed(12345)
train<-churn[1:2500,]
test<-churn[2501:3033,]
mydata1 <- rbind(test,train)
#Removing unwanted variables for analysis
mydata2<-mydata1[,-21]
mydata3<-mydata2[,-20]
mydata<-mydata3[,-19]
names(mydata)

##  [1] "Account Length" "VMail Message"  "Day Mins"       "Eve Mins"
##  [5] "Night Mins"     "Intl Mins"      "CustServ Calls" "Churn"
##  [9] "Int'l Plan"     "VMail Plan"     "Day Calls"      "Day Charge"
## [13] "Eve Calls"      "Eve Charge"     "Night Calls"    "Night Charge"
## [17] "Intl Calls"     "Intl Charge"

#Remove observations that are missing from datasell
#na.omit(mydata)%>%
head(mydata)

## # A tibble: 6 x 18
##   `Account Length` `VMail Message` `Day Mins` `Eve Mins` `Night Mins`
##              <int>           <int>      <dbl>      <dbl>        <dbl>
## 1              121              41       216.       242.         147
## 2              101               0       125.       257.         193.
## 3              115               0       179.       271          246.
## 4              168               0       183.       179.         293.
```

```
## 5               90               0       168.       139.         138.
## 6               70               0       147.       200          235.
## # ... with 13 more variables: `Intl Mins` <dbl>, `CustServ Calls` <int>,
## #   Churn <int>, `Int'l Plan` <int>, `VMail Plan` <int>, `Day
## #   Calls` <int>, `Day Charge` <dbl>, `Eve Calls` <int>, `Eve
## #   Charge` <dbl>, `Night Calls` <int>, `Night Charge` <dbl>, `Intl
## #   Calls` <int>, `Intl Charge` <dbl>
```

```
summary(mydata)
```

```
##  Account Length VMail Message       Day Mins        Eve Mins
##  Min.   :  1    Min.   : 0.000   Min.   :  0.0   Min.   :  0.0
##  1st Qu.: 74    1st Qu.: 0.000   1st Qu.:144.0   1st Qu.:166.6
##  Median :101    Median : 0.000   Median :179.3   Median :201.4
##  Mean   :101    Mean   : 8.085   Mean   :179.8   Mean   :201.0
##  3rd Qu.:127    3rd Qu.:20.000   3rd Qu.:216.2   3rd Qu.:235.8
##  Max.   :243    Max.   :51.000   Max.   :350.8   Max.   :363.7
##    Night Mins      Intl Mins     CustServ Calls      Churn
##  Min.   : 23.2   Min.   : 0.00   Min.   :0.000   Min.   :0.0000
##  1st Qu.:167.1   1st Qu.: 8.50   1st Qu.:1.000   1st Qu.:0.0000
##  Median :201.4   Median :10.30   Median :1.000   Median :0.0000
##  Mean   :200.8   Mean   :10.24   Mean   :1.554   Mean   :0.1447
##  3rd Qu.:235.3   3rd Qu.:12.10   3rd Qu.:2.000   3rd Qu.:0.0000
##  Max.   :395.0   Max.   :20.00   Max.   :9.000   Max.   :1.0000
##    Int'l Plan        VMail Plan       Day Calls       Day Charge
##  Min.   :0.00000   Min.   :0.0000   Min.   :  0.0   Min.   : 0.00
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.: 87.0   1st Qu.:24.48
##  Median :0.00000   Median :0.0000   Median :101.0   Median :30.48
##  Mean   :0.09792   Mean   :0.2766   Mean   :100.4   Mean   :30.57
##  3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:114.0   3rd Qu.:36.75
##  Max.   :1.00000   Max.   :1.0000   Max.   :165.0   Max.   :59.64
##    Eve Calls        Eve Charge     Night Calls      Night Charge
##  Min.   :  0.00   Min.   : 0.00   Min.   : 33.00   Min.   : 1.040
##  1st Qu.: 87.00   1st Qu.:14.16   1st Qu.: 87.00   1st Qu.: 7.520
##  Median :100.00   Median :17.12   Median :100.00   Median : 9.060
##  Mean   : 99.97   Mean   :17.09   Mean   : 99.99   Mean   : 9.036
##  3rd Qu.:113.00   3rd Qu.:20.04   3rd Qu.:113.00   3rd Qu.:10.590
##  Max.   :168.00   Max.   :30.91   Max.   :175.00   Max.   :17.770
##    Intl Calls     Intl Charge
##  Min.   : 0.00   Min.   :0.000
##  1st Qu.: 3.00   1st Qu.:2.300
##  Median : 4.00   Median :2.780
##  Mean   : 4.47   Mean   :2.764
##  3rd Qu.: 6.00   3rd Qu.:3.270
##  Max.   :19.00   Max.   :5.400
```

```
View(mydata)
sapply(mydata, sd)
```

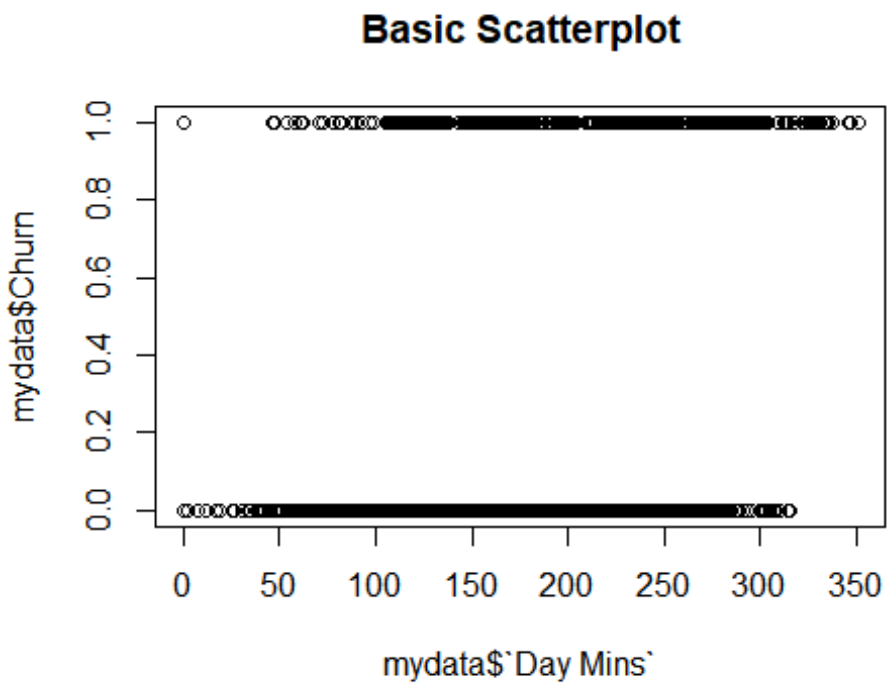
```
## Account Length  VMail Message       Day Mins       Eve Mins     Night Mins
##     39.9007277     13.6682673     54.5676061     51.0143598     50.5824663
```

```
##      Intl Mins CustServ Calls          Churn     Int'l Plan     VMail Plan
##      2.7865124      1.3117890      0.3518977      0.2972593      0.4474026
##      Day Calls     Day Charge      Eve Calls     Eve Charge    Night Calls
##     20.1037777      9.2764625     19.8697990      4.3361860     19.5717652
##   Night Charge     Intl Calls    Intl Charge
##      2.2762640      2.4435317      0.7523476
```

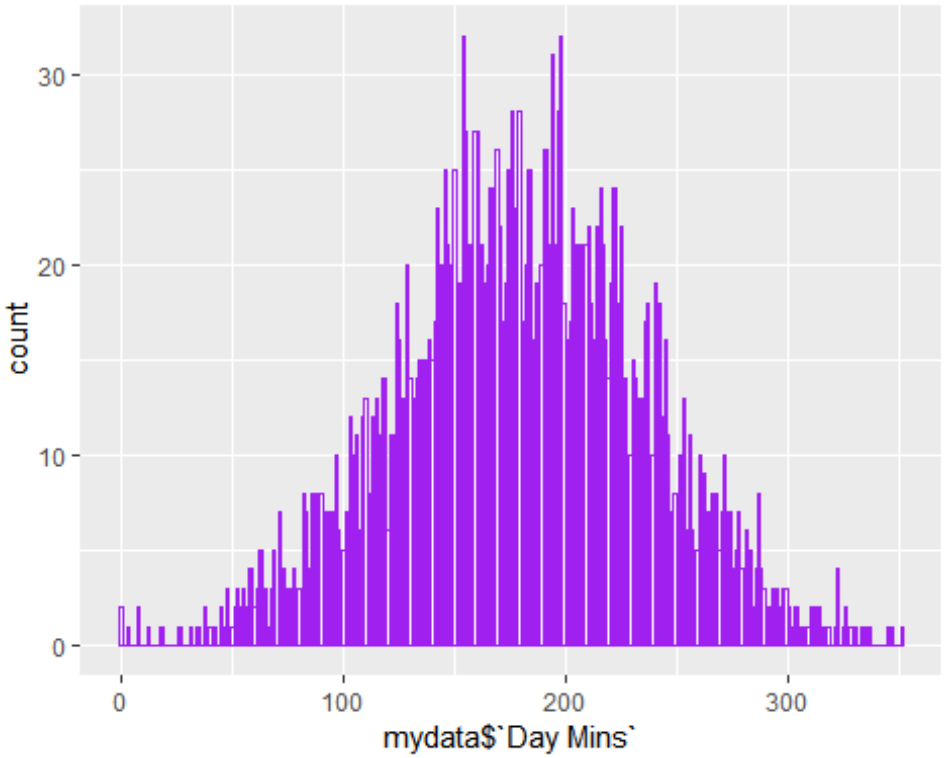
```
cormatrix <- round(cor(mydata), digits = 2 )
cormatrix
```

```
##                 Account Length VMail Message Day Mins Eve Mins Night Mins
## Account Length            1.00         -0.01     0.01    -0.01      -0.01
## VMail Message            -0.01          1.00     0.00     0.02       0.01
## Day Mins                  0.01          0.00     1.00     0.00       0.00
## Eve Mins                 -0.01          0.02     0.00     1.00      -0.01
## Night Mins               -0.01          0.01     0.00    -0.01       1.00
## Intl Mins                 0.02          0.02    -0.01    -0.01      -0.01
## CustServ Calls            0.00          0.00    -0.01    -0.01       0.00
## Churn                     0.02         -0.08     0.21     0.09       0.04
## Int'l Plan                0.02          0.02     0.06     0.02      -0.02
## VMail Plan                0.00          0.96     0.00     0.02       0.01
## Day Calls                 0.04          0.00     0.01    -0.02       0.02
## Day Charge                0.01          0.00     1.00     0.00       0.00
## Eve Calls                 0.02         -0.01     0.03    -0.01       0.00
## Eve Charge               -0.01          0.02     0.00     1.00      -0.01
## Night Calls               0.00          0.01     0.03     0.01       0.00
## Night Charge             -0.01          0.01     0.00    -0.01       1.00
## Intl Calls                0.02          0.02     0.00     0.01       0.00
## Intl Charge               0.02          0.02    -0.01    -0.01      -0.01
##                 Intl Mins CustServ Calls Churn Int'l Plan VMail Plan
## Account Length       0.02           0.00  0.02       0.02       0.00
## VMail Message        0.02           0.00 -0.08       0.02       0.96
## Day Mins            -0.01          -0.01  0.21       0.06       0.00
## Eve Mins            -0.01          -0.01  0.09       0.02       0.02
## Night Mins          -0.01           0.00  0.04      -0.02       0.01
## Intl Mins            1.00          -0.02  0.07       0.05       0.02
## CustServ Calls      -0.02           1.00  0.19      -0.03      -0.01
## Churn                0.07           0.19  1.00       0.26      -0.10
## Int'l Plan           0.05          -0.03  0.26       1.00       0.01
## VMail Plan           0.02          -0.01 -0.10       0.01       1.00
## Day Calls            0.01          -0.02  0.01       0.00       0.00
## Day Charge          -0.01          -0.02  0.21       0.06       0.00
## Eve Calls            0.01           0.01  0.01       0.00      -0.01
## Eve Charge          -0.01          -0.01  0.09       0.02       0.02
## Night Calls          0.00          -0.01  0.01       0.01       0.02
## Night Charge        -0.01           0.00  0.04      -0.02       0.01
## Intl Calls           0.03          -0.02 -0.05       0.00       0.01
## Intl Charge          1.00          -0.02  0.07       0.05       0.02
##                 Day Calls Day Charge Eve Calls Eve Charge Night Calls
## Account Length       0.04       0.01      0.02      -0.01        0.00
```

```
## VMail Message       0.00       0.00     -0.01       0.02        0.01
## Day Mins            0.01       1.00      0.03       0.00        0.03
## Eve Mins           -0.02       0.00     -0.01       1.00        0.01
## Night Mins          0.02       0.00      0.00      -0.01        0.00
## Intl Mins           0.01      -0.01      0.01      -0.01        0.00
## CustServ Calls     -0.02      -0.02      0.01      -0.01       -0.01
## Churn               0.01       0.21      0.01       0.09        0.01
## Int'l Plan          0.00       0.06      0.00       0.02        0.01
## VMail Plan          0.00       0.00     -0.01       0.02        0.02
## Day Calls           1.00       0.01      0.01      -0.02       -0.01
## Day Charge          0.01       1.00      0.03       0.00        0.03
## Eve Calls           0.01       0.03      1.00      -0.01        0.01
## Eve Charge         -0.02       0.00     -0.01       1.00        0.01
## Night Calls        -0.01       0.03      0.01       0.01        1.00
## Night Charge        0.02       0.00      0.00      -0.01        0.00
## Intl Calls          0.00       0.00      0.00       0.01        0.00
## Intl Charge         0.01      -0.01      0.01      -0.01        0.00
##                Night Charge Intl Calls Intl Charge
## Account Length        -0.01       0.02        0.02
## VMail Message          0.01       0.02        0.02
## Day Mins               0.00       0.00       -0.01
## Eve Mins              -0.01       0.01       -0.01
## Night Mins             1.00       0.00       -0.01
## Intl Mins             -0.01       0.03        1.00
## CustServ Calls         0.00      -0.02       -0.02
## Churn                  0.04      -0.05        0.07
## Int'l Plan            -0.02       0.00        0.05
## VMail Plan             0.01       0.01        0.02
## Day Calls              0.02       0.00        0.01
## Day Charge             0.00       0.00       -0.01
## Eve Calls              0.00       0.00        0.01
## Eve Charge            -0.01       0.01       -0.01
## Night Calls            0.00       0.00        0.00
## Night Charge           1.00       0.00       -0.01
## Intl Calls             0.00       1.00        0.03
## Intl Charge           -0.01       0.03        1.00

plot.new()
plot(mydata$Churn ~mydata$`Day Mins`)
title('Basic Scatterplot')
```



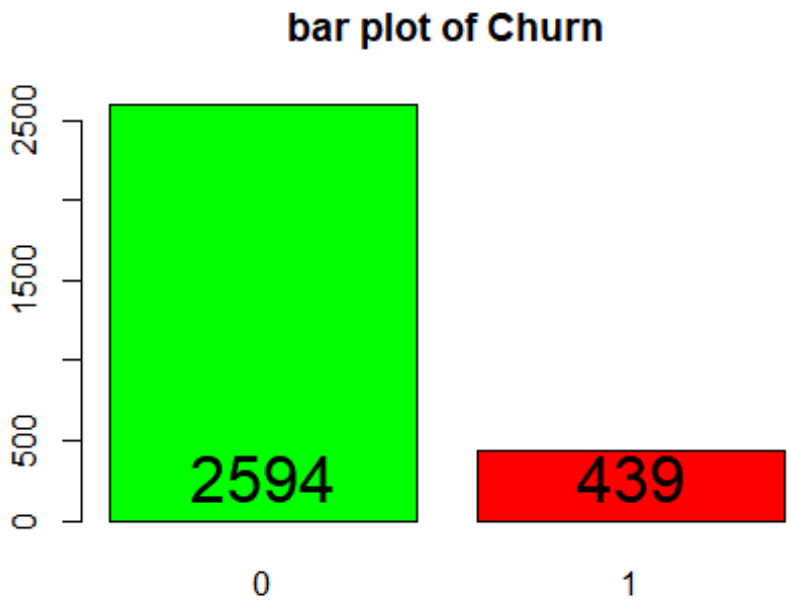

```
ggplot(mydata, aes(x=mydata$`Day Mins`)) + geom_histogram(binwidth = 1, fill
= "white", color = "purple")
```



```r
#Randomly split data into train and test set
#70% will be ssigned to train set, 30% will be assigned to tst set

set.seed(1234)
ind <- sample(2, nrow(mydata), replace = TRUE, prob = c(.7,.3))
traindata <- mydata[ind == 1,]
testdata <- mydata[ind == 2,]
names(mydata)
```

```
##  [1] "Account Length" "VMail Message"  "Day Mins"       "Eve Mins"
##  [5] "Night Mins"     "Intl Mins"      "CustServ Calls" "Churn"
##  [9] "Int'l Plan"     "VMail Plan"     "Day Calls"      "Day Charge"
## [13] "Eve Calls"      "Eve Charge"     "Night Calls"    "Night Charge"
## [17] "Intl Calls"     "Intl Charge"
```

```r
#Nearest Neighbhour classification

barplot(table(mydata$Churn), col= c("green", "red"), main='bar plot of
Churn')
text(barplot(table(mydata$Churn), col =c('green' , 'red'), main='bar plot of
Churn'), 0,table(mydata$Churn), cex =2 , pos =3)
```




```r
#proportion
round(prop.table(table(mydata$Churn))*100,digits = 2)
```

```
##
##     0     1
## 85.53 14.47
```

```
names(mydata)
```

```
##  [1] "Account Length" "VMail Message"  "Day Mins"       "Eve Mins"
##  [5] "Night Mins"     "Intl Mins"      "CustServ Calls" "Churn"
##  [9] "Int'l Plan"     "VMail Plan"     "Day Calls"      "Day Charge"
## [13] "Eve Calls"      "Eve Charge"     "Night Calls"    "Night Charge"
## [17] "Intl Calls"     "Intl Charge"
```

```
normalize<-function(x){return((x-min(x))/(max(x)-min(x)))}
mydata_n<-as.data.frame(lapply(mydata[1:18],normalize))
str(mydata)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    3033 obs. of  18 variables:
##  $ Account Length: int  121 101 115 168 90 70 138 43 117 108 ...
##  $ VMail Message : int  41 0 0 0 0 0 0 0 20 0 ...
##  $ Day Mins      : num  216 125 179 183 168 ...
##  $ Eve Mins      : num  242 257 271 179 139 ...
##  $ Night Mins    : num  147 193 246 293 138 ...
##  $ Intl Mins     : num  9.6 13.4 16.4 9.9 13 12.5 13.2 12.6 8.7 10.9 ...
##  $ CustServ Calls: int  1 0 2 2 1 3 1 0 2 2 ...
##  $ Churn         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Int'l Plan    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ VMail Plan    : int  1 0 0 0 0 0 0 0 1 0 ...
##  $ Day Calls     : int  95 66 114 131 96 105 107 125 98 105 ...
##  $ Day Charge    : num  36.6 21.2 30.4 31.1 28.5 ...
##  $ Eve Calls     : int  92 85 96 73 104 135 120 88 107 114 ...
##  $ Eve Charge    : num  20.6 21.9 23 15.2 11.8 ...
##  $ Night Calls   : int  108 115 94 100 87 65 119 106 147 98 ...
##  $ Night Charge  : num  6.61 8.69 11.07 13.18 6.23 ...
##  $ Intl Calls    : int  3 4 5 5 1 9 4 3 3 3 ...
##  $ Intl Charge   : num  2.59 3.62 4.43 2.67 3.51 3.38 3.56 3.4 2.35 2.94
...
```

```
str(mydata_n)
```

```
## 'data.frame':    3033 obs. of  18 variables:
##  $ Account.Length: num  0.496 0.413 0.471 0.69 0.368 ...
##  $ VMail.Message : num  0.804 0 0 0 0 ...
##  $ Day.Mins      : num  0.614 0.356 0.509 0.522 0.477 ...
##  $ Eve.Mins      : num  0.665 0.707 0.745 0.493 0.382 ...
##  $ Night.Mins    : num  0.333 0.457 0.599 0.725 0.31 ...
##  $ Intl.Mins     : num  0.48 0.67 0.82 0.495 0.65 0.625 0.66 0.63 0.435
0.545 ...
##  $ CustServ.Calls: num  0.111 0 0.222 0.222 0.111 ...
##  $ Churn         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Int.l.Plan    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ VMail.Plan    : num  1 0 0 0 0 0 0 0 1 0 ...
```

```
##  $ Day.Calls     : num  0.576 0.4 0.691 0.794 0.582 ...
##  $ Day.Charge    : num  0.614 0.356 0.509 0.522 0.478 ...
##  $ Eve.Calls     : num  0.548 0.506 0.571 0.435 0.619 ...
##  $ Eve.Charge    : num  0.665 0.707 0.745 0.493 0.382 ...
##  $ Night.Calls   : num  0.528 0.577 0.43 0.472 0.38 ...
##  $ Night.Charge  : num  0.333 0.457 0.6 0.726 0.31 ...
##  $ Intl.Calls    : num  0.1579 0.2105 0.2632 0.2632 0.0526 ...
##  $ Intl.Charge   : num  0.48 0.67 0.82 0.494 0.65 ...
```

```
mydata_train<-mydata_n[1:2800,]
mydata_test<-mydata_n[2801:3033,]
mydata_train_labels<-mydata_n[1:2800,7]
mydata_test_labels<-mydata_n[2801:3033,7]
str(mydata_train)
```

```
## 'data.frame':    2800 obs. of  18 variables:
##  $ Account.Length: num  0.496 0.413 0.471 0.69 0.368 ...
##  $ VMail.Message : num  0.804 0 0 0 0 ...
##  $ Day.Mins      : num  0.614 0.356 0.509 0.522 0.477 ...
##  $ Eve.Mins      : num  0.665 0.707 0.745 0.493 0.382 ...
##  $ Night.Mins    : num  0.333 0.457 0.599 0.725 0.31 ...
##  $ Intl.Mins     : num  0.48 0.67 0.82 0.495 0.65 0.625 0.66 0.63 0.435
0.545 ...
##  $ CustServ.Calls: num  0.111 0 0.222 0.222 0.111 ...
##  $ Churn         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Int.l.Plan    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ VMail.Plan    : num  1 0 0 0 0 0 0 0 1 0 ...
##  $ Day.Calls     : num  0.576 0.4 0.691 0.794 0.582 ...
##  $ Day.Charge    : num  0.614 0.356 0.509 0.522 0.478 ...
##  $ Eve.Calls     : num  0.548 0.506 0.571 0.435 0.619 ...
##  $ Eve.Charge    : num  0.665 0.707 0.745 0.493 0.382 ...
##  $ Night.Calls   : num  0.528 0.577 0.43 0.472 0.38 ...
##  $ Night.Charge  : num  0.333 0.457 0.6 0.726 0.31 ...
##  $ Intl.Calls    : num  0.1579 0.2105 0.2632 0.2632 0.0526 ...
##  $ Intl.Charge   : num  0.48 0.67 0.82 0.494 0.65 ...
```

```
str(mydata_train_labels)
```

```
##  num [1:2800] 0.111 0 0.222 0.222 0.111 ...
```

```
str(mydata_test)
```

```
## 'data.frame':    233 obs. of  18 variables:
##  $ Account.Length: num  0.116 0.351 0.566 0.599 0.719 ...
##  $ VMail.Message : num  0 0.314 0.725 0 0 ...
##  $ Day.Mins      : num  0.893 0.415 0.216 0.558 0.376 ...
##  $ Eve.Mins      : num  0.595 0.525 0.477 0.629 0.636 ...
##  $ Night.Mins    : num  0.525 0.517 0.375 0.754 0.781 ...
##  $ Intl.Mins     : num  0.64 0.565 0.41 0.61 0.435 0.585 0.38 0.495 0.385
0.295 ...
##  $ CustServ.Calls: num  0.222 0 0 0.333 0.111 ...
```

```
##  $ Churn         : num  1 0 0 0 0 0 0 0 0 0 ...
##  $ Int.l.Plan    : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ VMail.Plan    : num  0 1 1 0 0 0 0 1 0 0 ...
##  $ Day.Calls     : num  0.624 0.533 0.618 0.521 0.576 ...
##  $ Day.Charge    : num  0.893 0.415 0.216 0.558 0.376 ...
##  $ Eve.Calls     : num  0.899 0.768 0.875 0.488 0.44 ...
##  $ Eve.Charge    : num  0.595 0.525 0.478 0.629 0.636 ...
##  $ Night.Calls   : num  0.514 0.345 0.444 0.43 0.528 ...
##  $ Night.Charge  : num  0.525 0.518 0.375 0.754 0.781 ...
##  $ Intl.Calls    : num  0.211 0.368 0.684 0.211 0.526 ...
##  $ Intl.Charge   : num  0.641 0.565 0.409 0.609 0.435 ...
```

```
str(mydata_test_labels)
```

```
##  num [1:233] 0.222 0 0 0.333 0.111 ...
```

```
library(class)
#Apply knn
mydata_test_pred<-knn(train = mydata_train,test = mydata_test,
cl=mydata_train_labels,k=53)
summary(mydata_test_pred)
```

```
##                 0 0.111111111111111 0.222222222222222 0.333333333333333
##                 8               165                42                 5
## 0.444444444444444 0.555555555555556 0.666666666666667 0.777777777777778
##                11                 2                 0                 0
## 0.888888888888889                 1
##                 0                 0
```

```
#Evalulalte model
library(gmodels)
CrossTable(x=mydata_test_labels, y=mydata_test_pred,prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  233
##
##
##                   | mydata_test_pred
## mydata_test_labels |                 0 | 0.111111111111111 |
0.222222222222222 | 0.333333333333333 | 0.444444444444444 | 0.555555555555556
|         Row Total |
```

```
## -------------------|-------------------|--------------------|--------------
-----|-------------------|-------------------|-------------------|-----------
--------|
##                  0 |                 3 |                51 |
0 |                  0 |                 0 |                 0 |
54 |
##                    |             0.056 |             0.944 |
0.000 |             0.000 |             0.000 |             0.000 |
0.232 |
##                    |             0.375 |             0.309 |
0.000 |             0.000 |             0.000 |             0.000 |
|
##                    |             0.013 |             0.219 |
0.000 |             0.000 |             0.000 |             0.000 |
|
## -------------------|-------------------|--------------------|--------------
-----|-------------------|-------------------|-------------------|-----------
--------|
##  0.111111111111111 |                 3 |                71 |
4 |                  0 |                 1 |                 0 |
79 |
##                    |             0.038 |             0.899 |
0.051 |             0.000 |             0.013 |             0.000 |
0.339 |
##                    |             0.375 |             0.430 |
0.095 |             0.000 |             0.091 |             0.000 |
|
##                    |             0.013 |             0.305 |
0.017 |             0.000 |             0.004 |             0.000 |
|
## -------------------|-------------------|--------------------|--------------
-----|-------------------|-------------------|-------------------|-----------
--------|
##  0.222222222222222 |                 2 |                33 |
16 |                  0 |                 0 |                 0 |
51 |
##                    |             0.039 |             0.647 |
0.314 |             0.000 |             0.000 |             0.000 |
0.219 |
##                    |             0.250 |             0.200 |
0.381 |             0.000 |             0.000 |             0.000 |
|
##                    |             0.009 |             0.142 |
0.069 |             0.000 |             0.000 |             0.000 |
|
## -------------------|-------------------|--------------------|--------------
-----|-------------------|-------------------|-------------------|-----------
--------|
##  0.333333333333333 |                 0 |                10 |
15 |                  2 |                 0 |                 0 |
```

```
27 |
##                    |             0.000 |             0.370 |
0.556 |             0.074 |             0.000 |             0.000 |
0.116 |
##                    |             0.000 |             0.061 |
0.357 |             0.400 |             0.000 |             0.000 |
|
##                    |             0.000 |             0.043 |
0.064 |             0.009 |             0.000 |             0.000 |
|
## -------------------|-------------------|-------------------|-------------
-----|-------------------|-------------------|-------------------|-----------
--------|
##  0.444444444444444 |                 0 |                 0 |
4 |                 3 |                 4 |                 1 |
12 |
##                    |             0.000 |             0.000 |
0.333 |             0.250 |             0.333 |             0.083 |
0.052 |
##                    |             0.000 |             0.000 |
0.095 |             0.600 |             0.364 |             0.500 |
|
##                    |             0.000 |             0.000 |
0.017 |             0.013 |             0.017 |             0.004 |
|
## -------------------|-------------------|-------------------|-------------
-----|-------------------|-------------------|-------------------|-----------
--------|
##  0.555555555555556 |                 0 |                 0 |
3 |                 0 |                 3 |                 0 |
6 |
##                    |             0.000 |             0.000 |
0.500 |             0.000 |             0.500 |             0.000 |
0.026 |
##                    |             0.000 |             0.000 |
0.071 |             0.000 |             0.273 |             0.000 |
|
##                    |             0.000 |             0.000 |
0.013 |             0.000 |             0.013 |             0.000 |
|
## -------------------|-------------------|-------------------|-------------
-----|-------------------|-------------------|-------------------|-----------
--------|
##  0.666666666666667 |                 0 |                 0 |
0 |                 0 |                 2 |                 1 |
3 |
##                    |             0.000 |             0.000 |
0.000 |             0.000 |             0.667 |             0.333 |
0.013 |
##                    |             0.000 |             0.000 |
```

```
0.000 |             0.000 |             0.182 |             0.500 |
|
##                   |             0.000 |             0.000 |
0.000 |             0.000 |             0.009 |             0.004 |
|
## ------------------|-------------------|-------------------|-------------
-----|-------------------|-------------------|-------------------|-----------
--------|
##                 1 |                 0 |                 0 |
0 |                 0 |                 1 |                 0 |
1 |
##                   |             0.000 |             0.000 |
0.000 |             0.000 |             1.000 |             0.000 |
0.004 |
##                   |             0.000 |             0.000 |
0.000 |             0.000 |             0.091 |             0.000 |
|
##                   |             0.000 |             0.000 |
0.000 |             0.000 |             0.004 |             0.000 |
|
## ------------------|-------------------|-------------------|-------------
-----|-------------------|-------------------|-------------------|-----------
--------|
##       Column Total |                 8 |               165 |
42 |                 5 |                11 |                 2 |
233 |
##                   |             0.034 |             0.708 |
0.180 |             0.021 |             0.047 |             0.009 |
|
## ------------------|-------------------|-------------------|-------------
-----|-------------------|-------------------|-------------------|-----------
--------|
##
##
```

```
sapply(mydata_n, sd)
```

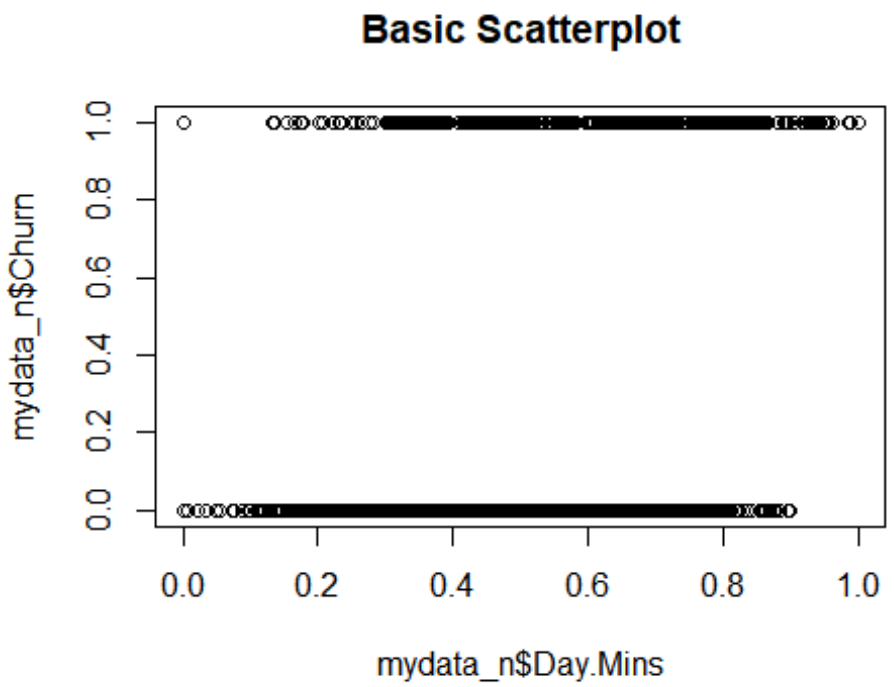
```
## Account.Length  VMail.Message       Day.Mins       Eve.Mins     Night.Mins
##      0.1648790      0.2680052      0.1555519      0.1402649      0.1360475
##      Intl.Mins CustServ.Calls          Churn     Int.l.Plan     VMail.Plan
##      0.1393256      0.1457543      0.3518977      0.2972593      0.4474026
##      Day.Calls     Day.Charge      Eve.Calls     Eve.Charge    Night.Calls
##      0.1218411      0.1555410      0.1182726      0.1402842      0.1378293
##   Night.Charge     Intl.Calls    Intl.Charge
##      0.1360588      0.1286069      0.1393236
```

```
cormatrix <- round(cor(mydata_n), digits = 2 )
cormatrix
```

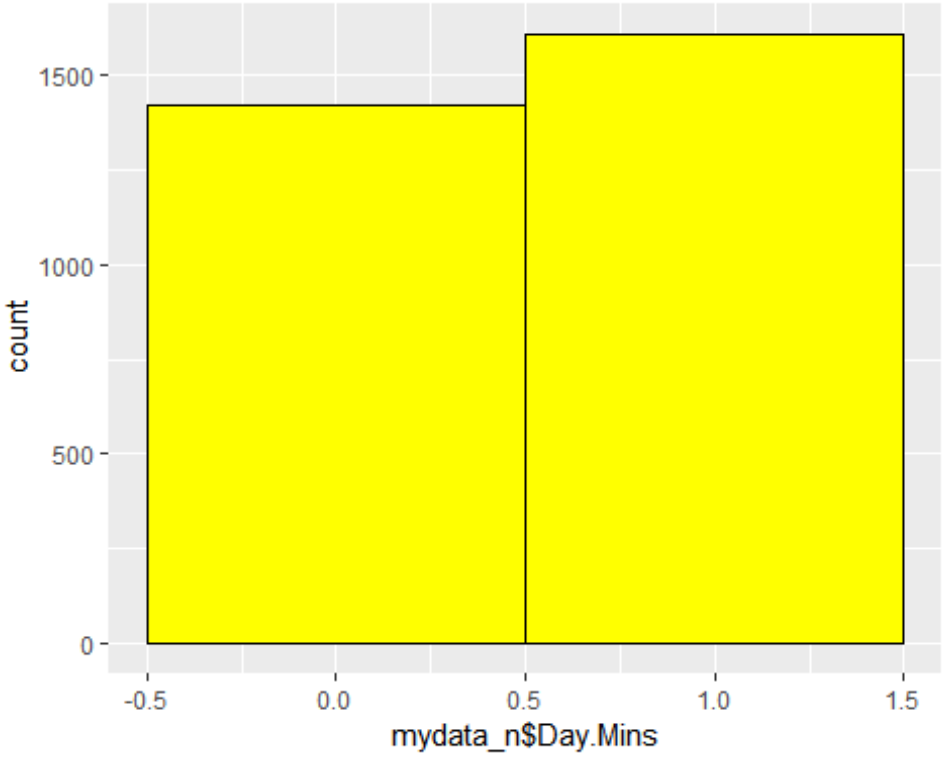
```
##                Account.Length VMail.Message Day.Mins Eve.Mins Night.Mins
## Account.Length           1.00         -0.01     0.01    -0.01       -0.01
```

```
## VMail.Message           -0.01          1.00     0.00     0.02       0.01
## Day.Mins                 0.01          0.00     1.00     0.00       0.00
## Eve.Mins                -0.01          0.02     0.00     1.00      -0.01
## Night.Mins              -0.01          0.01     0.00    -0.01       1.00
## Intl.Mins                0.02          0.02    -0.01    -0.01      -0.01
## CustServ.Calls           0.00          0.00    -0.01    -0.01       0.00
## Churn                    0.02         -0.08     0.21     0.09       0.04
## Int.l.Plan               0.02          0.02     0.06     0.02      -0.02
## VMail.Plan               0.00          0.96     0.00     0.02       0.01
## Day.Calls                0.04          0.00     0.01    -0.02       0.02
## Day.Charge               0.01          0.00     1.00     0.00       0.00
## Eve.Calls                0.02         -0.01     0.03    -0.01       0.00
## Eve.Charge              -0.01          0.02     0.00     1.00      -0.01
## Night.Calls              0.00          0.01     0.03     0.01       0.00
## Night.Charge            -0.01          0.01     0.00    -0.01       1.00
## Intl.Calls               0.02          0.02     0.00     0.01       0.00
## Intl.Charge              0.02          0.02    -0.01    -0.01      -0.01
##                 Intl.Mins CustServ.Calls Churn Int.l.Plan VMail.Plan
## Account.Length       0.02           0.00  0.02       0.02       0.00
## VMail.Message        0.02           0.00 -0.08       0.02       0.96
## Day.Mins            -0.01          -0.01  0.21       0.06       0.00
## Eve.Mins            -0.01          -0.01  0.09       0.02       0.02
## Night.Mins          -0.01           0.00  0.04      -0.02       0.01
## Intl.Mins            1.00          -0.02  0.07       0.05       0.02
## CustServ.Calls      -0.02           1.00  0.19      -0.03      -0.01
## Churn                0.07           0.19  1.00       0.26      -0.10
## Int.l.Plan           0.05          -0.03  0.26       1.00       0.01
## VMail.Plan           0.02          -0.01 -0.10       0.01       1.00
## Day.Calls            0.01          -0.02  0.01       0.00       0.00
## Day.Charge          -0.01          -0.02  0.21       0.06       0.00
## Eve.Calls            0.01           0.01  0.01       0.00      -0.01
## Eve.Charge          -0.01          -0.01  0.09       0.02       0.02
## Night.Calls          0.00          -0.01  0.01       0.01       0.02
## Night.Charge        -0.01           0.00  0.04      -0.02       0.01
## Intl.Calls           0.03          -0.02 -0.05       0.00       0.01
## Intl.Charge          1.00          -0.02  0.07       0.05       0.02
##                 Day.Calls Day.Charge Eve.Calls Eve.Charge Night.Calls
## Account.Length       0.04       0.01      0.02      -0.01        0.00
## VMail.Message        0.00       0.00     -0.01       0.02        0.01
## Day.Mins             0.01       1.00      0.03       0.00        0.03
## Eve.Mins            -0.02       0.00     -0.01       1.00        0.01
## Night.Mins           0.02       0.00      0.00      -0.01        0.00
## Intl.Mins            0.01      -0.01      0.01      -0.01        0.00
## CustServ.Calls      -0.02      -0.02      0.01      -0.01       -0.01
## Churn                0.01       0.21      0.01       0.09        0.01
## Int.l.Plan           0.00       0.06      0.00       0.02        0.01
## VMail.Plan           0.00       0.00     -0.01       0.02        0.02
## Day.Calls            1.00       0.01      0.01      -0.02       -0.01
## Day.Charge           0.01       1.00      0.03       0.00        0.03
## Eve.Calls            0.01       0.03      1.00      -0.01        0.01
```

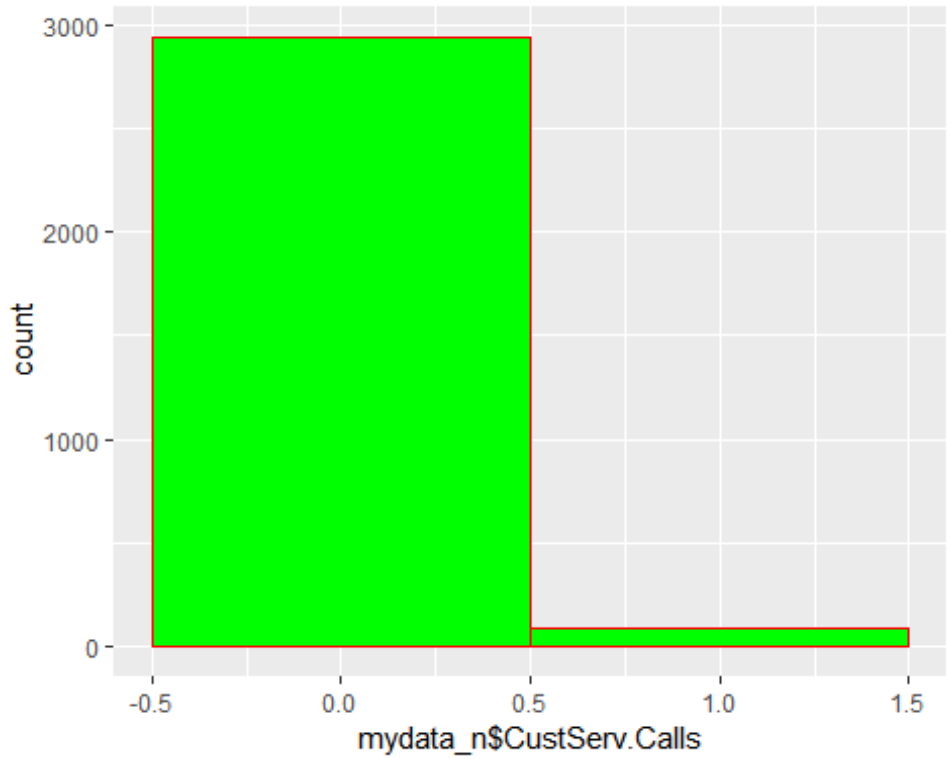
```
## Eve.Charge          -0.02       0.00     -0.01       1.00         0.01
## Night.Calls         -0.01       0.03      0.01       0.01         1.00
## Night.Charge         0.02       0.00      0.00      -0.01         0.00
## Intl.Calls           0.00       0.00      0.00       0.01         0.00
## Intl.Charge          0.01      -0.01      0.01      -0.01         0.00
##                 Night.Charge Intl.Calls Intl.Charge
## Account.Length         -0.01       0.02        0.02
## VMail.Message           0.01       0.02        0.02
## Day.Mins                0.00       0.00       -0.01
## Eve.Mins               -0.01       0.01       -0.01
## Night.Mins              1.00       0.00       -0.01
## Intl.Mins              -0.01       0.03        1.00
## CustServ.Calls          0.00      -0.02       -0.02
## Churn                   0.04      -0.05        0.07
## Int.l.Plan             -0.02       0.00        0.05
## VMail.Plan              0.01       0.01        0.02
## Day.Calls               0.02       0.00        0.01
## Day.Charge              0.00       0.00       -0.01
## Eve.Calls               0.00       0.00        0.01
## Eve.Charge             -0.01       0.01       -0.01
## Night.Calls             0.00       0.00        0.00
## Night.Charge            1.00       0.00       -0.01
## Intl.Calls              0.00       1.00        0.03
## Intl.Charge            -0.01       0.03        1.00

plot.new()
plot(mydata_n$Churn ~mydata_n$Day.Mins)
title('Basic Scatterplot')
```




```
ggplot(mydata_n, aes(x=mydata_n$Day.Mins)) + geom_histogram(binwidth = 1,
fill = "yellow", color = "black")
```



```
ggplot(mydata_n, aes(x=mydata_n$CustServ.Calls)) + geom_histogram(binwidth =
1, fill = "green", color = "red")
```




```
names(mydata_n)
```

```
##  [1] "Account.Length" "VMail.Message"  "Day.Mins"       "Eve.Mins"
##  [5] "Night.Mins"     "Intl.Mins"      "CustServ.Calls" "Churn"
##  [9] "Int.l.Plan"     "VMail.Plan"     "Day.Calls"      "Day.Charge"
## [13] "Eve.Calls"      "Eve.Charge"     "Night.Calls"    "Night.Charge"
## [17] "Intl.Calls"     "Intl.Charge"
```

```
#Forward elimination
#Lower AIC indicates a better model
forward <- step(glm(Churn ~ 1, data = mydata_train), direction = 'forward',
scope = ~Account.Length+VMail.Message+Day.Mins + Eve.Mins +
                  Night.Mins + Intl.Mins + CustServ.Calls + Int.l.Plan +
VMail.Plan +
                  Day.Calls + Day.Charge + Eve.Calls + Eve.Charge +
Night.Calls +
                  Night.Charge + Intl.Calls + Intl.Charge)
```

```
## Start:  AIC=2069.89
## Churn ~ 1
##
##                 Df Deviance    AIC
## + Int.l.Plan     1   318.50 1865.6
## + Day.Mins       1   328.10 1948.7
```

```
## + Day.Charge      1   328.10 1948.7
## + CustServ.Calls  1   332.16 1983.1
## + VMail.Plan      1   339.45 2043.9
## + Eve.Mins        1   339.52 2044.5
## + Eve.Charge      1   339.52 2044.5
## + VMail.Message   1   340.16 2049.8
## + Intl.Charge     1   340.90 2055.8
## + Intl.Mins       1   340.90 2055.8
## + Intl.Calls      1   342.20 2066.5
## + Night.Charge    1   342.24 2066.8
## + Night.Mins      1   342.24 2066.8
## <none>                342.86 2069.9
## + Account.Length  1   342.74 2070.9
## + Eve.Calls       1   342.83 2071.6
## + Day.Calls       1   342.85 2071.8
## + Night.Calls     1   342.86 2071.9
##
## Step:  AIC=1865.58
## Churn ~ Int.l.Plan
##
##                  Df Deviance    AIC
## + Day.Mins        1   305.81 1753.7
## + Day.Charge      1   305.81 1753.7
## + CustServ.Calls  1   306.92 1763.9
## + VMail.Plan      1   314.85 1835.3
## + Eve.Mins        1   315.49 1841.0
## + Eve.Charge      1   315.49 1841.0
## + VMail.Message   1   315.56 1841.6
## + Intl.Charge     1   317.21 1856.2
## + Intl.Mins       1   317.21 1856.2
## + Night.Charge    1   317.67 1860.2
## + Night.Mins      1   317.67 1860.2
## + Intl.Calls      1   317.83 1861.6
## <none>                318.50 1865.6
## + Account.Length  1   318.45 1867.1
## + Eve.Calls       1   318.48 1867.4
## + Night.Calls     1   318.48 1867.4
## + Day.Calls       1   318.50 1867.6
##
## Step:  AIC=1753.73
## Churn ~ Int.l.Plan + Day.Mins
##
##                  Df Deviance    AIC
## + CustServ.Calls  1   293.63 1641.9
## + VMail.Plan      1   302.18 1722.3
## + VMail.Message   1   302.81 1728.1
## + Eve.Mins        1   302.94 1729.3
## + Eve.Charge      1   302.94 1729.3
## + Intl.Charge     1   304.40 1742.8
## + Intl.Mins       1   304.40 1742.8
```

```
## + Night.Charge    1   304.97 1748.0
## + Night.Mins      1   304.97 1748.0
## + Intl.Calls      1   305.07 1748.9
## <none>                305.81 1753.7
## + Night.Calls     1   305.74 1755.0
## + Account.Length  1   305.76 1755.3
## + Day.Charge      1   305.78 1755.4
## + Eve.Calls       1   305.81 1755.7
## + Day.Calls       1   305.81 1755.7
##
## Step:  AIC=1641.91
## Churn ~ Int.l.Plan + Day.Mins + CustServ.Calls
##
##                  Df Deviance    AIC
## + VMail.Plan      1   290.02 1609.3
## + VMail.Message   1   290.61 1614.9
## + Eve.Mins        1   290.71 1615.9
## + Eve.Charge      1   290.71 1615.9
## + Intl.Charge     1   292.01 1628.4
## + Intl.Mins       1   292.01 1628.4
## + Night.Charge    1   292.75 1635.5
## + Night.Mins      1   292.75 1635.5
## + Intl.Calls      1   292.92 1637.2
## <none>                293.63 1641.9
## + Night.Calls     1   293.57 1643.4
## + Account.Length  1   293.57 1643.4
## + Day.Charge      1   293.62 1643.8
## + Day.Calls       1   293.62 1643.8
## + Eve.Calls       1   293.63 1643.9
##
## Step:  AIC=1609.27
## Churn ~ Int.l.Plan + Day.Mins + CustServ.Calls + VMail.Plan
##
##                  Df Deviance    AIC
## + Eve.Mins        1   287.04 1582.4
## + Eve.Charge      1   287.05 1582.4
## + Intl.Charge     1   288.33 1594.9
## + Intl.Mins       1   288.33 1594.9
## + Night.Charge    1   289.13 1602.6
## + Night.Mins      1   289.13 1602.6
## + Intl.Calls      1   289.32 1604.5
## <none>                290.02 1609.3
## + VMail.Message   1   289.94 1610.5
## + Account.Length  1   289.97 1610.7
## + Night.Calls     1   289.98 1610.9
## + Day.Charge      1   290.00 1611.1
## + Day.Calls       1   290.01 1611.2
## + Eve.Calls       1   290.02 1611.3
##
## Step:  AIC=1582.39
```

```
## Churn ~ Int.l.Plan + Day.Mins + CustServ.Calls + VMail.Plan +
##     Eve.Mins
##
##                   Df Deviance    AIC
## + Intl.Charge      1   285.26 1567.0
## + Intl.Mins        1   285.27 1567.0
## + Night.Charge     1   286.10 1575.2
## + Night.Mins       1   286.10 1575.2
## + Intl.Calls       1   286.33 1577.5
## <none>                 287.04 1582.4
## + Eve.Charge       1   286.90 1583.0
## + VMail.Message    1   286.95 1583.5
## + Account.Length   1   286.98 1583.8
## + Night.Calls      1   287.00 1584.0
## + Day.Charge       1   287.01 1584.1
## + Day.Calls        1   287.03 1584.2
## + Eve.Calls        1   287.04 1584.3
##
## Step:  AIC=1566.97
## Churn ~ Int.l.Plan + Day.Mins + CustServ.Calls + VMail.Plan +
##     Eve.Mins + Intl.Charge
##
##                   Df Deviance    AIC
## + Night.Charge     1   284.26 1559.1
## + Night.Mins       1   284.26 1559.1
## + Intl.Calls       1   284.49 1561.3
## <none>                 285.26 1567.0
## + Eve.Charge       1   285.14 1567.7
## + VMail.Message    1   285.19 1568.2
## + Account.Length   1   285.21 1568.5
## + Night.Calls      1   285.22 1568.5
## + Intl.Mins        1   285.23 1568.7
## + Day.Charge       1   285.24 1568.7
## + Day.Calls        1   285.25 1568.8
## + Eve.Calls        1   285.26 1569.0
##
## Step:  AIC=1559.06
## Churn ~ Int.l.Plan + Day.Mins + CustServ.Calls + VMail.Plan +
##     Eve.Mins + Intl.Charge + Night.Charge
##
##                   Df Deviance    AIC
## + Intl.Calls       1   283.48 1553.4
## <none>                 284.26 1559.1
## + Eve.Charge       1   284.13 1559.8
## + VMail.Message    1   284.18 1560.3
## + Account.Length   1   284.20 1560.5
## + Night.Calls      1   284.20 1560.5
## + Day.Charge       1   284.23 1560.8
## + Intl.Mins        1   284.23 1560.8
## + Night.Mins       1   284.23 1560.8
```

```
## + Day.Calls       1   284.25 1561.0
## + Eve.Calls       1   284.25 1561.0
##
## Step:  AIC=1553.4
## Churn ~ Int.l.Plan + Day.Mins + CustServ.Calls + VMail.Plan +
##     Eve.Mins + Intl.Charge + Night.Charge + Intl.Calls
##
##                  Df Deviance    AIC
## <none>                283.48 1553.4
## + Eve.Charge      1   283.35 1554.1
## + VMail.Message   1   283.39 1554.5
## + Account.Length  1   283.41 1554.8
## + Night.Calls     1   283.42 1554.8
## + Intl.Mins       1   283.44 1555.0
## + Night.Mins      1   283.45 1555.1
## + Day.Charge      1   283.45 1555.1
## + Day.Calls       1   283.47 1555.3
## + Eve.Calls       1   283.48 1555.4
```

```
logit<- glm(Churn ~Account.Length+Day.Mins+ Day.Charge
+CustServ.Calls+VMail.Plan +Eve.Mins+ Eve.Charge+VMail.Message+Day.Calls
+Eve.Calls+ Intl.Mins + Night.Calls+Intl.Calls, data = mydata_train, family =
"binomial")
summary(logit)
```

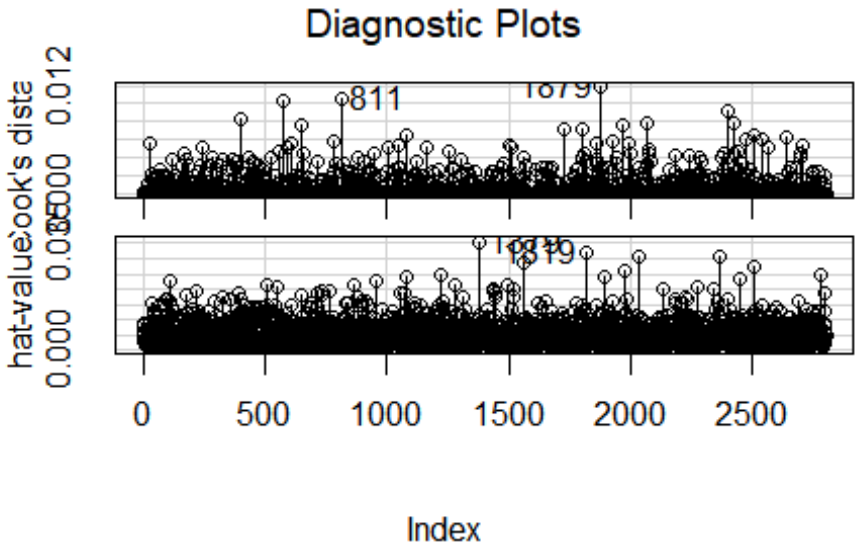
```
##
## Call:
## glm(formula = Churn ~ Account.Length + Day.Mins + Day.Charge +
##     CustServ.Calls + VMail.Plan + Eve.Mins + Eve.Charge + VMail.Message +
##     Day.Calls + Eve.Calls + Intl.Mins + Night.Calls + Intl.Calls,
##     family = "binomial", data = mydata_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6222  -0.5655  -0.4011  -0.2479   3.0151
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -7.25383    0.66439 -10.918  < 2e-16 ***
## Account.Length    0.37156    0.35201   1.056  0.29117
## Day.Mins        269.45486 1194.53382   0.226  0.82153
## Day.Charge     -264.95008 1194.61886  -0.222  0.82448
## CustServ.Calls    3.63509    0.36714   9.901  < 2e-16 ***
## VMail.Plan       -1.65418    0.58581  -2.824  0.00475 **
## Eve.Mins        663.66121  626.47216   1.059  0.28943
## Eve.Charge     -661.14858  626.37712  -1.056  0.29119
## VMail.Message     1.38872    0.94640   1.467  0.14228
## Day.Calls         0.18253    0.47698   0.383  0.70196
## Eve.Calls         0.12810    0.48716   0.263  0.79259
## Intl.Mins         2.10973    0.42728   4.938 7.91e-07 ***
```

```
## Night.Calls      -0.07677    0.42299  -0.181  0.85599
## Intl.Calls       -1.35828    0.48166  -2.820  0.00480 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2296.7  on 2799  degrees of freedom
## Residual deviance: 1981.5  on 2786  degrees of freedom
## AIC: 2009.5
##
## Number of Fisher Scoring iterations: 5
```

```
#evaluate model's fit and performance
influenceIndexPlot(logit, vars = c('Cook', "hat"), id.n =4)
```

```
## Warning in plot.window(...): "id.n" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "id.n" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "id.n" is not
## a graphical parameter


## Warning in axis(side = side, at = at, labels = labels, ...): "id.n" is not
## a graphical parameter

## Warning in box(...): "id.n" is not a graphical parameter

## Warning in title(...): "id.n" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "id.n" is not a
## graphical parameter


## Warning in plot.xy(xy.coords(x, y), type = type, ...): "id.n" is not a
## graphical parameter

## Warning in plot.window(...): "id.n" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "id.n" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "id.n" is not
## a graphical parameter


## Warning in axis(side = side, at = at, labels = labels, ...): "id.n" is not
## a graphical parameter

## Warning in box(...): "id.n" is not a graphical parameter

## Warning in title(...): "id.n" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "id.n" is not a
## graphical parameter
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "id.n" is not a
## graphical parameter
```




```
# Confidence interval using log-likelihood
confint(logit)

## Waiting for profiling to be done...

##                         2.5 %       97.5 %
## (Intercept)       -8.5721555   -5.9665847
## Account.Length    -0.3185885    1.0619748
## Day.Mins       -2073.6482603 2611.3934164
## Day.Charge     -2607.0467447 2078.3283908
## CustServ.Calls     2.9178587    4.3582284
## VMail.Plan        -2.8316748   -0.5326631
## Eve.Mins        -562.9378802 1894.1545109
## Eve.Charge     -1891.4503111  565.2693795
## VMail.Message     -0.4613650    3.2537413
## Day.Calls         -0.7518555    1.1187826
## Eve.Calls         -0.8261958    1.0844000
## Intl.Mins          1.2773818    2.9530184
## Night.Calls       -0.9064145    0.7524913
## Intl.Calls        -2.3182303   -0.4293632

exp(logit$coefficients)
```

```
##    (Intercept) Account.Length       Day.Mins     Day.Charge CustServ.Calls
##   7.074582e-04   1.450000e+00  1.053803e+117  8.583087e-116   3.790512e+01
##     VMail.Plan       Eve.Mins     Eve.Charge  VMail.Message      Day.Calls
##   1.912497e-01  1.676499e+288  7.359052e-288   4.009711e+00   1.200252e+00
##      Eve.Calls      Intl.Mins    Night.Calls     Intl.Calls
##   1.136669e+00   8.245980e+00   9.261057e-01   2.571039e-01
```

```
# Making a support vector machine (another prediction model)
svm_model <- svm(Churn ~., data= mydata_train, gamma = .1, cost =1)
print(svm_model)
```

```
##
## Call:
## svm(formula = Churn ~ ., data = mydata_train, gamma = 0.1, cost = 1)
##
##
## Parameters:
##    SVM-Type:  eps-regression
##  SVM-Kernel:  radial
##        cost:  1
##       gamma:  0.1
##     epsilon:  0.1
##
##
## Number of Support Vectors:  1650
```

```
summary(svm_model)
```

```
##
## Call:
## svm(formula = Churn ~ ., data = mydata_train, gamma = 0.1, cost = 1)
##
##
## Parameters:
##    SVM-Type:  eps-regression
##  SVM-Kernel:  radial
##        cost:  1
##       gamma:  0.1
##     epsilon:  0.1
##
##
## Number of Support Vectors:  1650
```

```
#Random forest model- takes decision trees and averages them
rf <- randomForest(Churn ~., data= mydata_train, ntree = 500, mtry = 5,
importance = TRUE)
```

```
## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?
```

```
print(rf)
```

```
##
## Call:
##  randomForest(formula = Churn ~ ., data = mydata_train, ntree = 500,
mtry = 5, importance = TRUE)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 5
##
##           Mean of squared residuals: 0.04381018
##                     % Var explained: 64.22
```

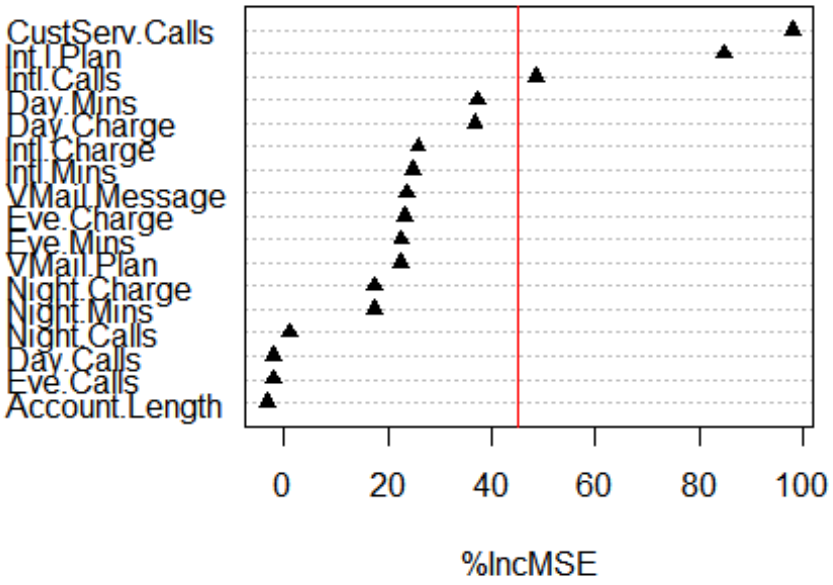
```
importance(rf)
```

```
##                   %IncMSE IncNodePurity
## Account.Length -3.3667171      8.958597
## VMail.Message  23.5093488     10.984626
## Day.Mins       37.3007328     46.899902
## Eve.Mins       22.6072179     22.906599
## Night.Mins     17.2330540     12.429405
## Intl.Mins      24.8707290     13.953814
## CustServ.Calls 97.9485792     33.708149
## Int.l.Plan     84.7632991     27.756730
## VMail.Plan     22.4330568      7.813428
## Day.Calls      -2.0564407      9.410889
## Day.Charge     36.6905800     44.824857
## Eve.Calls      -2.2526506      7.513130
## Eve.Charge     23.1060040     23.041325
## Night.Calls     0.9573658      8.763461
## Night.Charge   17.2960451     11.731147
## Intl.Calls     48.4776494     18.678216
## Intl.Charge    25.7659394     14.681465
```

%IncMSE is the most robust and informative measure. It is the increase in mse of predictions (estimated with out-of-bag-CV) as a result of variable j being permuted (values randomly shuffled). the higher number, the more important

IncNodePurity relates to the loss function which by best splits are chosen. The loss function is mse for regression and gini-impurity for classification. More useful variables achieve higher increases in node purities, that is to find a split which has a high inter node 'variance' and a small intra node 'variance'. IncNodePurity is biased and should only be used if the extra computation time of calculating %IncMSE is unacceptable

```
plot.new()
varImpPlot(rf, type = 1, pch = 17, col = 1, cex = 1.0, main = "")
abline(v= 45, col= "red")
```




```
mydata_n$Churn <- as.factor(mydata_n$Churn)
mydata$Churn <- as.factor(mydata$Churn)
summary(mydata_n)
```

```
##  Account.Length   VMail.Message       Day.Mins         Eve.Mins
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.3017   1st Qu.:0.0000   1st Qu.:0.4105   1st Qu.:0.4581
##  Median :0.4132   Median :0.0000   Median :0.5111   Median :0.5538
##  Mean   :0.4133   Mean   :0.1585   Mean   :0.5126   Mean   :0.5527
##  3rd Qu.:0.5207   3rd Qu.:0.3922   3rd Qu.:0.6163   3rd Qu.:0.6483
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##    Night.Mins       Intl.Mins      CustServ.Calls   Churn
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   0:2594
##  1st Qu.:0.3870   1st Qu.:0.4250   1st Qu.:0.1111   1: 439
##  Median :0.4793   Median :0.5150   Median :0.1111
##  Mean   :0.4776   Mean   :0.5118   Mean   :0.1727
##  3rd Qu.:0.5705   3rd Qu.:0.6050   3rd Qu.:0.2222
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##    Int.l.Plan        VMail.Plan       Day.Calls        Day.Charge
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.5273   1st Qu.:0.4105
##  Median :0.00000   Median :0.0000   Median :0.6121   Median :0.5111
##  Mean   :0.09792   Mean   :0.2766   Mean   :0.6084   Mean   :0.5125
##  3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:0.6909   3rd Qu.:0.6162
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##    Eve.Calls        Eve.Charge      Night.Calls      Night.Charge
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
```

```
##  1st Qu.:0.5179   1st Qu.:0.4581   1st Qu.:0.3803   1st Qu.:0.3873
##  Median :0.5952   Median :0.5539   Median :0.4718   Median :0.4794
##  Mean   :0.5951   Mean   :0.5528   Mean   :0.4718   Mean   :0.4779
##  3rd Qu.:0.6726   3rd Qu.:0.6483   3rd Qu.:0.5634   3rd Qu.:0.5708
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##    Intl.Calls      Intl.Charge
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.1579   1st Qu.:0.4259
##  Median :0.2105   Median :0.5148
##  Mean   :0.2353   Mean   :0.5119
##  3rd Qu.:0.3158   3rd Qu.:0.6056
##  Max.   :1.0000   Max.   :1.0000
```

```
str(mydata_n)
```

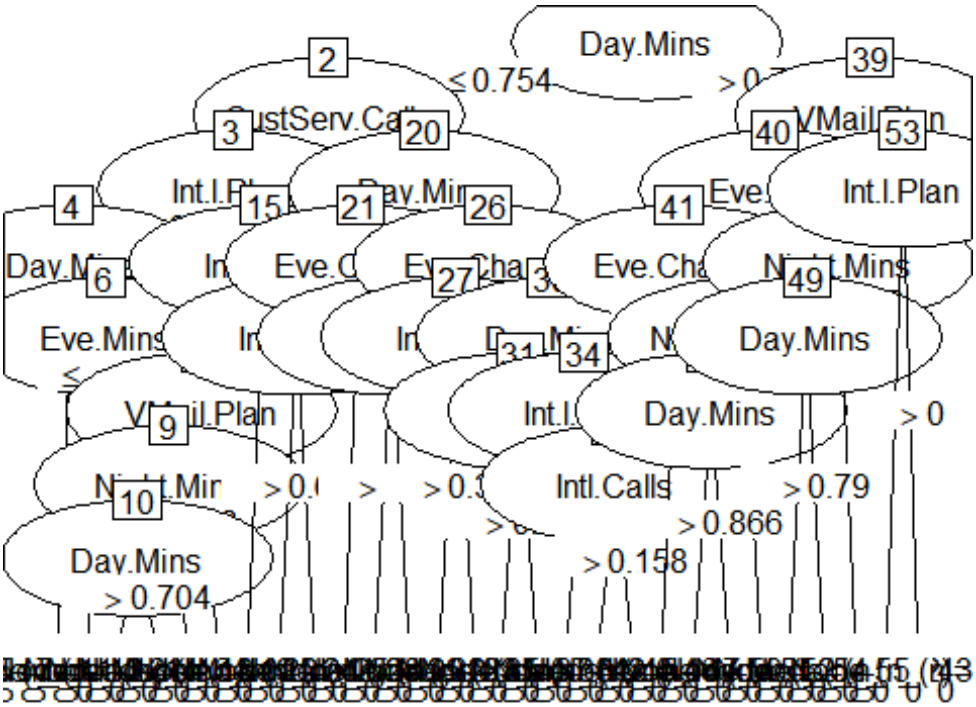
```
## 'data.frame':    3033 obs. of  18 variables:
##  $ Account.Length: num  0.496 0.413 0.471 0.69 0.368 ...
##  $ VMail.Message : num  0.804 0 0 0 0 ...
##  $ Day.Mins      : num  0.614 0.356 0.509 0.522 0.477 ...
##  $ Eve.Mins      : num  0.665 0.707 0.745 0.493 0.382 ...
##  $ Night.Mins    : num  0.333 0.457 0.599 0.725 0.31 ...
##  $ Intl.Mins     : num  0.48 0.67 0.82 0.495 0.65 0.625 0.66 0.63 0.435
0.545 ...
##  $ CustServ.Calls: num  0.111 0 0.222 0.222 0.111 ...
##  $ Churn         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 ...
##  $ Int.l.Plan    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ VMail.Plan    : num  1 0 0 0 0 0 0 0 1 0 ...
##  $ Day.Calls     : num  0.576 0.4 0.691 0.794 0.582 ...
##  $ Day.Charge    : num  0.614 0.356 0.509 0.522 0.478 ...
##  $ Eve.Calls     : num  0.548 0.506 0.571 0.435 0.619 ...
##  $ Eve.Charge    : num  0.665 0.707 0.745 0.493 0.382 ...
##  $ Night.Calls   : num  0.528 0.577 0.43 0.472 0.38 ...
##  $ Night.Charge  : num  0.333 0.457 0.6 0.726 0.31 ...
##  $ Intl.Calls    : num  0.1579 0.2105 0.2632 0.2632 0.0526 ...
##  $ Intl.Charge   : num  0.48 0.67 0.82 0.494 0.65 ...
```

```
#algorithm for decision tree
tree <- C5.0(Churn ~., data = mydata_n)
summary(tree)
```

```
##
## Call:
## C5.0.formula(formula = Churn ~ ., data = mydata_n)
##
##
## C5.0 [Release 2.07 GPL Edition]      Tue Oct 16 21:36:07 2018
## -------------------------------
##
## Class specified by attribute `outcome'
##
## Read 3033 cases (18 attributes) from undefined.data
```

```
##
## Decision tree:
##
## Day.Mins > 0.7537058:
## :...VMail.Plan > 0:
## :   :...Int.l.Plan <= 0: 0 (42/1)
## :   :   Int.l.Plan > 0: 1 (8/3)
## :   VMail.Plan <= 0:
## :   :...Eve.Mins > 0.5169095:
## :       :...Night.Mins > 0.2789134: 1 (82/1)
## :       :   Night.Mins <= 0.2789134:
## :       :   :...Day.Mins <= 0.7896237: 0 (4)
## :       :       Day.Mins > 0.7896237: 1 (3)
## :       Eve.Mins <= 0.5169095:
## :       :...Eve.Charge <= 0.3966354: 0 (15/1)
## :           Eve.Charge > 0.3966354:
## :           :...Night.Charge > 0.5098625: 1 (17/1)
## :               Night.Charge <= 0.5098625:
## :               :...Day.Mins <= 0.8663056: 0 (15/1)
## :                   Day.Mins > 0.8663056: 1 (6)
## Day.Mins <= 0.7537058:
## :...CustServ.Calls > 0.3333333:
##     :...Day.Mins <= 0.4566705:
##     :   :...Eve.Charge <= 0.6415399: 1 (68/3)
##     :   :   Eve.Charge > 0.6415399:
##     :   :   :...Day.Mins <= 0.3429304: 1 (8)
##     :   :       Day.Mins > 0.3429304: 0 (13/3)
##     :   Day.Mins > 0.4566705:
##     :   :...Eve.Charge <= 0.3710773:
##     :       :...Intl.Calls <= 0.3157895: 1 (11)
##     :       :   Intl.Calls > 0.3157895: 0 (2)
##     :       Eve.Charge > 0.3710773:
##     :       :...Day.Mins <= 0.4968643:
##     :           :...Eve.Mins <= 0.5820732: 1 (14/1)
##     :           :   Eve.Mins > 0.5820732: 0 (14)
##     :           Day.Mins > 0.4968643:
##     :           :...Int.l.Plan <= 0: 0 (85/6)
##     :               Int.l.Plan > 0:
##     :               :...Intl.Calls <= 0.1578947: 1 (4)
##     :                   Intl.Calls > 0.1578947: 0 (7)
##     CustServ.Calls <= 0.3333333:
##     :...Int.l.Plan > 0:
##         :...Intl.Calls <= 0.1052632: 1 (48)
##         :   Intl.Calls > 0.1052632:
##         :   :...Intl.Mins <= 0.655: 0 (159/7)
##         :       Intl.Mins > 0.655: 1 (40)
##         Int.l.Plan <= 0:
##         :...Day.Mins <= 0.63626: 0 (2027/58)
##             Day.Mins > 0.63626:
##             :...Eve.Mins <= 0.6662084: 0 (266/20)
```

```
##                 Eve.Mins > 0.6662084:
##                 :...VMail.Plan > 0: 0 (16)
##                     VMail.Plan <= 0:
##                     :...Night.Mins > 0.4061323: 1 (44/6)
##                         Night.Mins <= 0.4061323:
##                         :...Day.Mins <= 0.7035348: 0 (10)
##                             Day.Mins > 0.7035348: 1 (5/1)
##
##
## Evaluation on training data (3033 cases):
##
##      Decision Tree
##    ----------------
##    Size      Errors
##
##      28  113( 3.7%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##    2578    16    (a): class 0
##      97   342    (b): class 1
##
##
##  Attribute usage:
##
##  100.00% Day.Mins
##   93.67% CustServ.Calls
##   91.03% Int.l.Plan
##   16.85% Eve.Mins
##    9.20% Eve.Charge
##    8.94% Intl.Calls
##    8.80% VMail.Plan
##    6.56% Intl.Mins
##    4.88% Night.Mins
##    1.25% Night.Charge
##
##
## Time: 0.1 secs

plot(tree)
```




```
results <- C5.0(Churn ~., data = mydata_n, rules = TRUE)
summary(results)

##
## Call:
## C5.0.formula(formula = Churn ~ ., data = mydata_n, rules = TRUE)
##
##
## C5.0 [Release 2.07 GPL Edition]      Tue Oct 16 21:36:10 2018
## -------------------------------
##
## Class specified by attribute `outcome'
##
## Read 3033 cases (18 attributes) from undefined.data
##
## Rules:
##
## Rule 1: (2027/58, lift 1.1)
##  Day.Mins <= 0.63626
##  CustServ.Calls <= 0.3333333
##  Int.l.Plan <= 0
##  ->  class 0  [0.971]
##
## Rule 2: (42/1, lift 1.1)
##  Day.Mins > 0.7537058
##  Int.l.Plan <= 0
##  VMail.Plan > 0
```

```
##  ->  class 0  [0.955]
##
## Rule 3: (1798/82, lift 1.1)
##  Day.Mins <= 0.7537058
##  Intl.Mins <= 0.655
##  CustServ.Calls <= 0.3333333
##  Intl.Calls > 0.1052632
##  ->  class 0  [0.954]
##
## Rule 4: (180/8, lift 1.1)
##  Day.Mins > 0.3429304
##  Day.Mins <= 0.4566705
##  Eve.Charge > 0.6415399
##  ->  class 0  [0.951]
##
## Rule 5: (400/44, lift 1.0)
##  Eve.Charge <= 0.3966354
##  ->  class 0  [0.888]
##
## Rule 6: (1570/176, lift 1.0)
##  Day.Mins > 0.4566705
##  Day.Mins <= 0.7537058
##  Eve.Charge > 0.3710773
##  ->  class 0  [0.887]
##
## Rule 7: (57, lift 6.8)
##  Int.l.Plan > 0
##  Intl.Calls <= 0.1052632
##  ->  class 1  [0.983]
##
## Rule 8: (53, lift 6.8)
##  Intl.Mins > 0.655
##  Int.l.Plan > 0
##  ->  class 1  [0.982]
##
## Rule 9: (27, lift 6.7)
##  Day.Mins <= 0.3429304
##  CustServ.Calls > 0.3333333
##  ->  class 1  [0.966]
##
## Rule 10: (39/1, lift 6.6)
##  Day.Mins > 0.7035348
##  Eve.Mins > 0.6662084
##  Int.l.Plan <= 0
##  VMail.Plan <= 0
##  ->  class 1  [0.951]
##
## Rule 11: (18, lift 6.6)
##  CustServ.Calls > 0.3333333
##  Eve.Charge <= 0.3710773
```

```
##  Intl.Calls <= 0.3157895
##  ->  class 1  [0.950]
##
## Rule 12: (68/3, lift 6.5)
##  Day.Mins <= 0.4566705
##  CustServ.Calls > 0.3333333
##  Eve.Charge <= 0.6415399
##  ->  class 1  [0.943]
##
## Rule 13: (70/4, lift 6.4)
##  Day.Mins <= 0.4968643
##  Eve.Mins <= 0.5820732
##  CustServ.Calls > 0.3333333
##  ->  class 1  [0.931]
##
## Rule 14: (69/6, lift 6.2)
##  Day.Mins > 0.63626
##  Eve.Mins > 0.6662084
##  Night.Mins > 0.4061323
##  VMail.Plan <= 0
##  ->  class 1  [0.901]
##
## Rule 15: (192/78, lift 4.1)
##  Day.Mins > 0.7537058
##  ->  class 1  [0.593]
##
## Default class: 0
##
##
## Evaluation on training data (3033 cases):
##
##          Rules
##    ----------------
##      No      Errors
##
##      15  144( 4.7%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##    2564    30    (a): class 0
##     114   325    (b): class 1
##
##
##  Attribute usage:
##
##   97.99% Day.Mins
##   82.53% CustServ.Calls
##   72.93% Int.l.Plan
##   70.76% Eve.Charge
```

```
##   61.75% Intl.Calls
##   61.03% Intl.Mins
##    4.98% Eve.Mins
##    4.06% VMail.Plan
##    2.27% Night.Mins
##
##
## Time: 0.1 secs
```

Logistic_regression_model_churn_L.R

S.Varatharajan

Fri Oct 19 20:58:09 2018

```
setwd("C:/Users/tsraj/Desktop/Acadgild project")
library(readr)
churn <- read_csv("churn.csv")

## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   `Day Mins` = col_double(),
##   `Eve Mins` = col_double(),
##   `Night Mins` = col_double(),
##   `Intl Mins` = col_double(),
##   `Day Charge` = col_double(),
##   `Eve Charge` = col_double(),
##   `Night Charge` = col_double(),
##   `Intl Charge` = col_double(),
##   State = col_character(),
##   Phone = col_character()
## )

## See spec(...) for full column specifications.

View(churn)
library(doParallel)

## Loading required package: foreach

## Loading required package: iterators

## Loading required package: parallel

registerDoParallel()
set.seed(12345)
train<-churn[1:2800,]
test<-churn[2501:3333,]
mydata1 <- rbind(test,train)
```

```r
#Removing unwanted variables for analysis
mydata2<-mydata1[,-21]
mydata3<-mydata2[,-20]
mydata<-mydata3[,-19]
names(mydata)
```

```
##  [1] "Account Length" "VMail Message"  "Day Mins"       "Eve Mins"
##  [5] "Night Mins"     "Intl Mins"      "CustServ Calls" "Churn"
##  [9] "Int'l Plan"     "VMail Plan"     "Day Calls"      "Day Charge"
## [13] "Eve Calls"      "Eve Charge"     "Night Calls"    "Night Charge"
## [17] "Intl Calls"     "Intl Charge"
```

```r
datatrain<-mydata[1:2800,]
datatest<-mydata[2801:3333,]
# logistic regression model:
fit <- glm(Churn~.,data = datatrain,family = binomial(link='logit'))
summary(fit)
```

```
##
## Call:
## glm(formula = Churn ~ ., family = binomial(link = "logit"), data =
datatrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1238  -0.5193  -0.3447  -0.2066   3.2057
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -8.183e+00  7.835e-01 -10.445  < 2e-16 ***
## `Account Length`  6.085e-04  1.506e-03   0.404  0.68610
## `VMail Message`   3.074e-02  1.910e-02   1.610  0.10746
## `Day Mins`        1.733e-01  3.588e+00   0.048  0.96148
## `Eve Mins`        1.779e+00  1.799e+00   0.989  0.32278
## `Night Mins`     -4.138e-01  9.461e-01  -0.437  0.66180
## `Intl Mins`      -1.445e+00  5.743e+00  -0.252  0.80139
## `CustServ Calls`  5.018e-01  4.250e-02  11.808  < 2e-16 ***
## `Int'l Plan`      1.957e+00  1.569e-01  12.476  < 2e-16 ***
## `VMail Plan`     -1.812e+00  6.118e-01  -2.961  0.00307 **
## `Day Calls`       1.374e-03  3.002e-03   0.458  0.64706
## `Day Charge`     -9.443e-01  2.110e+01  -0.045  0.96431
## `Eve Calls`       1.091e-03  3.016e-03   0.362  0.71754
## `Eve Charge`     -2.085e+01  2.116e+01  -0.985  0.32466
## `Night Calls`    -5.059e-04  3.077e-03  -0.164  0.86939
## `Night Charge`    9.276e+00  2.102e+01   0.441  0.65904
## `Intl Calls`     -8.531e-02  2.667e-02  -3.198  0.00138 **
## `Intl Charge`     5.674e+00  2.127e+01   0.267  0.78965
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2310.9  on 2799  degrees of freedom
## Residual deviance: 1831.9  on 2782  degrees of freedom
## AIC: 1867.9
##
## Number of Fisher Scoring iterations: 6
```

```
library(MASS)
step_fit <- stepAIC(fit,method='backward')
```

```
## Start:  AIC=1867.93
## Churn ~ `Account Length` + `VMail Message` + `Day Mins` + `Eve Mins` +
##     `Night Mins` + `Intl Mins` + `CustServ Calls` + `Int'l Plan` +
##     `VMail Plan` + `Day Calls` + `Day Charge` + `Eve Calls` +
##     `Eve Charge` + `Night Calls` + `Night Charge` + `Intl Calls` +
##     `Intl Charge`
##
##                    Df Deviance    AIC
## - `Day Charge`      1   1831.9 1865.9
## - `Day Mins`        1   1831.9 1865.9
## - `Night Calls`     1   1832.0 1866.0
## - `Intl Mins`       1   1832.0 1866.0
## - `Intl Charge`     1   1832.0 1866.0
## - `Eve Calls`       1   1832.1 1866.1
## - `Account Length`  1   1832.1 1866.1
## - `Night Mins`      1   1832.1 1866.1
## - `Night Charge`    1   1832.1 1866.1
## - `Day Calls`       1   1832.1 1866.1
## - `Eve Charge`      1   1832.9 1866.9
## - `Eve Mins`        1   1832.9 1866.9
## <none>                  1831.9 1867.9
## - `VMail Message`   1   1834.5 1868.5
## - `VMail Plan`      1   1841.4 1875.4
## - `Intl Calls`      1   1842.8 1876.8
## - `CustServ Calls`  1   1974.9 2008.9
## - `Int'l Plan`      1   1980.3 2014.3
##
## Step:  AIC=1865.93
## Churn ~ `Account Length` + `VMail Message` + `Day Mins` + `Eve Mins` +
##     `Night Mins` + `Intl Mins` + `CustServ Calls` + `Int'l Plan` +
##     `VMail Plan` + `Day Calls` + `Eve Calls` + `Eve Charge` +
##     `Night Calls` + `Night Charge` + `Intl Calls` + `Intl Charge`
##
##                    Df Deviance    AIC
## - `Night Calls`     1   1832.0 1864.0
## - `Intl Mins`       1   1832.0 1864.0
## - `Intl Charge`     1   1832.0 1864.0
## - `Eve Calls`       1   1832.1 1864.1
## - `Account Length`  1   1832.1 1864.1
```

```
## - `Night Mins`      1   1832.1 1864.1
## - `Night Charge`    1   1832.1 1864.1
## - `Day Calls`       1   1832.1 1864.1
## - `Eve Charge`      1   1832.9 1864.9
## - `Eve Mins`        1   1832.9 1864.9
## <none>                  1831.9 1865.9
## - `VMail Message`   1   1834.5 1866.5
## - `VMail Plan`      1   1841.4 1873.4
## - `Intl Calls`      1   1842.8 1874.8
## - `Day Mins`        1   1960.7 1992.7
## - `CustServ Calls`  1   1975.0 2007.0
## - `Int'l Plan`      1   1980.4 2012.4
##
## Step:  AIC=1863.96
## Churn ~ `Account Length` + `VMail Message` + `Day Mins` + `Eve Mins` +
##     `Night Mins` + `Intl Mins` + `CustServ Calls` + `Int'l Plan` +
##     `VMail Plan` + `Day Calls` + `Eve Calls` + `Eve Charge` +
##     `Night Charge` + `Intl Calls` + `Intl Charge`
##
##                     Df Deviance    AIC
## - `Intl Mins`       1   1832.0 1862.0
## - `Intl Charge`     1   1832.0 1862.0
## - `Eve Calls`       1   1832.1 1862.1
## - `Account Length`  1   1832.1 1862.1
## - `Night Mins`      1   1832.2 1862.2
## - `Night Charge`    1   1832.2 1862.2
## - `Day Calls`       1   1832.2 1862.2
## - `Eve Charge`      1   1832.9 1862.9
## - `Eve Mins`        1   1832.9 1862.9
## <none>                  1832.0 1864.0
## - `VMail Message`   1   1834.6 1864.6
## - `VMail Plan`      1   1841.5 1871.5
## - `Intl Calls`      1   1842.8 1872.8
## - `Day Mins`        1   1960.7 1990.7
## - `CustServ Calls`  1   1975.1 2005.1
## - `Int'l Plan`      1   1980.4 2010.4
##
## Step:  AIC=1862.02
## Churn ~ `Account Length` + `VMail Message` + `Day Mins` + `Eve Mins` +
##     `Night Mins` + `CustServ Calls` + `Int'l Plan` + `VMail Plan` +
##     `Day Calls` + `Eve Calls` + `Eve Charge` + `Night Charge` +
##     `Intl Calls` + `Intl Charge`
##
##                     Df Deviance    AIC
## - `Eve Calls`       1   1832.2 1860.2
## - `Account Length`  1   1832.2 1860.2
## - `Night Mins`      1   1832.2 1860.2
## - `Night Charge`    1   1832.2 1860.2
## - `Day Calls`       1   1832.2 1860.2
## - `Eve Charge`      1   1833.0 1861.0
```

```
## - `Eve Mins`         1   1833.0 1861.0
## <none>                   1832.0 1862.0
## - `VMail Message`    1   1834.7 1862.7
## - `VMail Plan`       1   1841.5 1869.5
## - `Intl Calls`       1   1842.8 1870.8
## - `Intl Charge`      1   1847.9 1875.9
## - `Day Mins`         1   1960.8 1988.8
## - `CustServ Calls`   1   1975.2 2003.2
## - `Int'l Plan`       1   1980.4 2008.4
##
## Step:  AIC=1860.15
## Churn ~ `Account Length` + `VMail Message` + `Day Mins` + `Eve Mins` +
##     `Night Mins` + `CustServ Calls` + `Int'l Plan` + `VMail Plan` +
##     `Day Calls` + `Eve Charge` + `Night Charge` + `Intl Calls` +
##     `Intl Charge`
##
##                    Df Deviance    AIC
## - `Account Length`  1   1832.3 1858.3
## - `Night Mins`      1   1832.4 1858.4
## - `Night Charge`    1   1832.4 1858.4
## - `Day Calls`       1   1832.4 1858.4
## - `Eve Charge`      1   1833.1 1859.1
## - `Eve Mins`        1   1833.1 1859.1
## <none>                  1832.2 1860.2
## - `VMail Message`   1   1834.8 1860.8
## - `VMail Plan`      1   1841.6 1867.6
## - `Intl Calls`      1   1842.9 1868.9
## - `Intl Charge`     1   1848.0 1874.0
## - `Day Mins`        1   1961.2 1987.2
## - `CustServ Calls`  1   1975.3 2001.3
## - `Int'l Plan`      1   1980.6 2006.6
##
## Step:  AIC=1858.33
## Churn ~ `VMail Message` + `Day Mins` + `Eve Mins` + `Night Mins` +
##     `CustServ Calls` + `Int'l Plan` + `VMail Plan` + `Day Calls` +
##     `Eve Charge` + `Night Charge` + `Intl Calls` + `Intl Charge`
##
##                    Df Deviance    AIC
## - `Night Mins`      1   1832.5 1856.5
## - `Night Charge`    1   1832.5 1856.5
## - `Day Calls`       1   1832.5 1856.5
## - `Eve Charge`      1   1833.3 1857.3
## - `Eve Mins`        1   1833.3 1857.3
## <none>                  1832.3 1858.3
## - `VMail Message`   1   1835.0 1859.0
## - `VMail Plan`      1   1841.8 1865.8
## - `Intl Calls`      1   1843.1 1867.1
## - `Intl Charge`     1   1848.2 1872.2
## - `Day Mins`        1   1961.5 1985.5
## - `CustServ Calls`  1   1975.6 1999.6
```

```
## - `Int'l Plan`      1   1981.5 2005.5
##
## Step:  AIC=1856.54
## Churn ~ `VMail Message` + `Day Mins` + `Eve Mins` + `CustServ Calls` +
##     `Int'l Plan` + `VMail Plan` + `Day Calls` + `Eve Charge` +
##     `Night Charge` + `Intl Calls` + `Intl Charge`
##
##                    Df Deviance    AIC
## - `Day Calls`       1   1832.8 1854.8
## - `Eve Charge`      1   1833.5 1855.5
## - `Eve Mins`        1   1833.5 1855.5
## <none>                  1832.5 1856.5
## - `VMail Message`   1   1835.2 1857.2
## - `Night Charge`    1   1841.7 1863.7
## - `VMail Plan`      1   1842.0 1864.0
## - `Intl Calls`      1   1843.3 1865.3
## - `Intl Charge`     1   1848.4 1870.4
## - `Day Mins`        1   1961.6 1983.6
## - `CustServ Calls`  1   1976.1 1998.1
## - `Int'l Plan`      1   1981.5 2003.5
##
## Step:  AIC=1854.76
## Churn ~ `VMail Message` + `Day Mins` + `Eve Mins` + `CustServ Calls` +
##     `Int'l Plan` + `VMail Plan` + `Eve Charge` + `Night Charge` +
##     `Intl Calls` + `Intl Charge`
##
##                    Df Deviance    AIC
## - `Eve Charge`      1   1833.7 1853.7
## - `Eve Mins`        1   1833.7 1853.7
## <none>                  1832.8 1854.8
## - `VMail Message`   1   1835.4 1855.4
## - `Night Charge`    1   1842.0 1862.0
## - `VMail Plan`      1   1842.2 1862.2
## - `Intl Calls`      1   1843.6 1863.6
## - `Intl Charge`     1   1848.8 1868.8
## - `Day Mins`        1   1961.9 1981.9
## - `CustServ Calls`  1   1976.1 1996.1
## - `Int'l Plan`      1   1981.7 2001.7
##
## Step:  AIC=1853.68
## Churn ~ `VMail Message` + `Day Mins` + `Eve Mins` + `CustServ Calls` +
##     `Int'l Plan` + `VMail Plan` + `Night Charge` + `Intl Calls` +
##     `Intl Charge`
##
##                    Df Deviance    AIC
## <none>                  1833.7 1853.7
## - `VMail Message`   1   1836.2 1854.2
## - `VMail Plan`      1   1843.0 1861.0
## - `Night Charge`    1   1843.0 1861.0
## - `Intl Calls`      1   1844.4 1862.4
```

```
## - `Intl Charge`     1   1849.8 1867.8
## - `Eve Mins`        1   1865.1 1883.1
## - `Day Mins`        1   1963.2 1981.2
## - `CustServ Calls`  1   1977.2 1995.2
## - `Int'l Plan`      1   1982.0 2000.0
```

```
summary(step_fit)
```

```
##
## Call:
## glm(formula = Churn ~ `VMail Message` + `Day Mins` + `Eve Mins` +
##     `CustServ Calls` + `Int'l Plan` + `VMail Plan` + `Night Charge` +
##     `Intl Calls` + `Intl Charge`, family = binomial(link = "logit"),
##     data = datatrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0697  -0.5183  -0.3438  -0.2067   3.1823
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -7.934964   0.558044 -14.219  < 2e-16 ***
## `VMail Message`   0.030137   0.019022   1.584  0.11312
## `Day Mins`        0.012753   0.001177  10.832  < 2e-16 ***
## `Eve Mins`        0.006890   0.001246   5.530 3.21e-08 ***
## `CustServ Calls`  0.502024   0.042394  11.842  < 2e-16 ***
## `Int'l Plan`      1.950763   0.156264  12.484  < 2e-16 ***
## `VMail Plan`     -1.794702   0.609260  -2.946  0.00322 **
## `Night Charge`    0.080332   0.026367   3.047  0.00231 **
## `Intl Calls`     -0.084671   0.026609  -3.182  0.00146 **
## `Intl Charge`     0.325860   0.081975   3.975 7.03e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2310.9  on 2799  degrees of freedom
## Residual deviance: 1833.7  on 2790  degrees of freedom
## AIC: 1853.7
##
## Number of Fisher Scoring iterations: 6
```

```
confint(step_fit)
```

```
## Waiting for profiling to be done...
```

```
##                       2.5 %       97.5 %
## (Intercept)      -9.046956802 -6.858483637
## `VMail Message`  -0.006994403  0.067678500
## `Day Mins`        0.010469969  0.015087589
## `Eve Mins`        0.004461103  0.009347615
```

```
## `CustServ Calls`  0.419446811  0.585772595
## `Int'l Plan`      1.644797742  2.257881599
## `VMail Plan`     -3.020152155 -0.628670873
## `Night Charge`    0.028779331  0.132186595
## `Intl Calls`     -0.137765570 -0.033429008
## `Intl Charge`     0.166131510  0.487631516

#ANOVA on base model
anova(fit,test = 'Chisq')

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Churn
##
## Terms added sequentially (first to last)
##
##
##                  Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                              2799     2310.9
## `Account Length`  1    0.716      2798     2310.2  0.397574
## `VMail Message`   1   23.230      2797     2287.0 1.438e-06 ***
## `Day Mins`        1  116.475      2796     2170.5 < 2.2e-16 ***
## `Eve Mins`        1   23.927      2795     2146.6 1.000e-06 ***
## `Night Mins`      1    4.016      2794     2142.6  0.045074 *
## `Intl Mins`       1   17.262      2793     2125.3 3.256e-05 ***
## `CustServ Calls`  1  126.467      2792     1998.8 < 2.2e-16 ***
## `Int'l Plan`      1  145.302      2791     1853.5 < 2.2e-16 ***
## `VMail Plan`      1    9.241      2790     1844.3  0.002366 **
## `Day Calls`       1    0.294      2789     1844.0  0.587490
## `Day Charge`      1    0.015      2788     1844.0  0.901166
## `Eve Calls`       1    0.091      2787     1843.9  0.763442
## `Eve Charge`      1    0.867      2786     1843.0  0.351779
## `Night Calls`     1    0.035      2785     1843.0  0.850852
## `Night Charge`    1    0.197      2784     1842.8  0.657042
## `Intl Calls`      1   10.803      2783     1832.0  0.001013 **
## `Intl Charge`     1    0.071      2782     1831.9  0.789638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#ANOVA from reduced model after applying the Step AIC
anova(step_fit,test = 'Chisq')

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Churn
##
## Terms added sequentially (first to last)
```

```
##
##
##                  Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                              2799     2310.9
## `VMail Message`   1   23.311      2798     2287.6 1.378e-06 ***
## `Day Mins`        1  116.493      2797     2171.1 < 2.2e-16 ***
## `Eve Mins`        1   23.844      2796     2147.3 1.044e-06 ***
## `CustServ Calls`  1  123.315      2795     2024.0 < 2.2e-16 ***
## `Int'l Plan`      1  146.934      2794     1877.0 < 2.2e-16 ***
## `VMail Plan`      1    8.819      2793     1868.2  0.002981 **
## `Night Charge`    1    8.705      2792     1859.5  0.003173 **
## `Intl Calls`      1    9.683      2791     1849.8  0.001860 **
## `Intl Charge`     1   16.150      2790     1833.7 5.853e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#plot the fitted model
plot(fit$fitted.values)
```



```r
pred_link <- predict(fit,newdata = datatest,type = 'link')
```

```r
#check for multicollinearity
library(car)
```

```
## Loading required package: carData
```

```r
vif(fit)
```

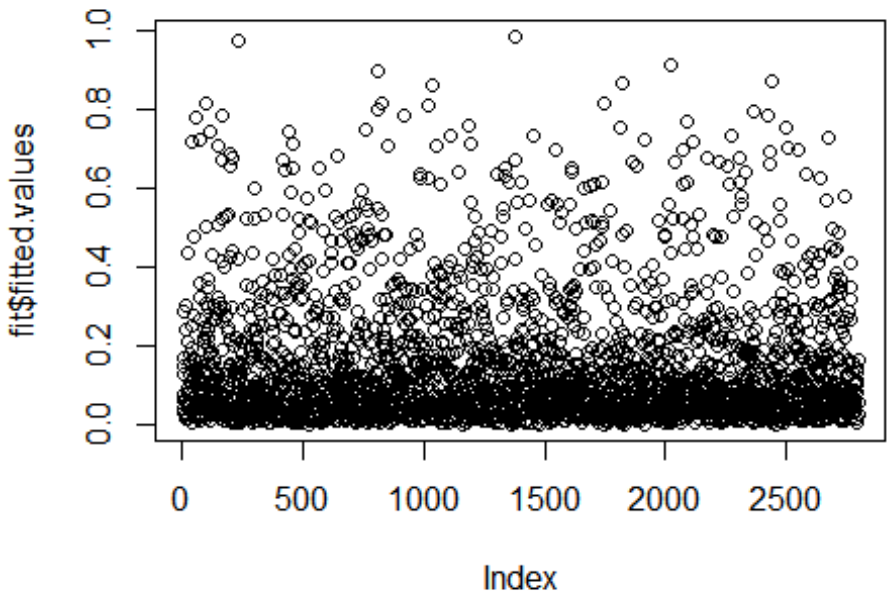
```
## `Account Length`  `VMail Message`       `Day Mins`       `Eve Mins`
##     1.006587e+00     1.554716e+01     9.719736e+06     2.131214e+06
##     `Night Mins`      `Intl Mins` `CustServ Calls`     `Int'l Plan`
##     6.445626e+05     6.823175e+04     1.085420e+00     1.065205e+00
##     `VMail Plan`      `Day Calls`     `Day Charge`      `Eve Calls`
##     1.556532e+01     1.007791e+00     9.719766e+06     1.004732e+00
##     `Eve Charge`    `Night Calls`   `Night Charge`     `Intl Calls`
##     2.131205e+06     1.006817e+00     6.445605e+05     1.013840e+00
##    `Intl Charge`
##     6.823230e+04
```

```r
vif(step_fit)
```

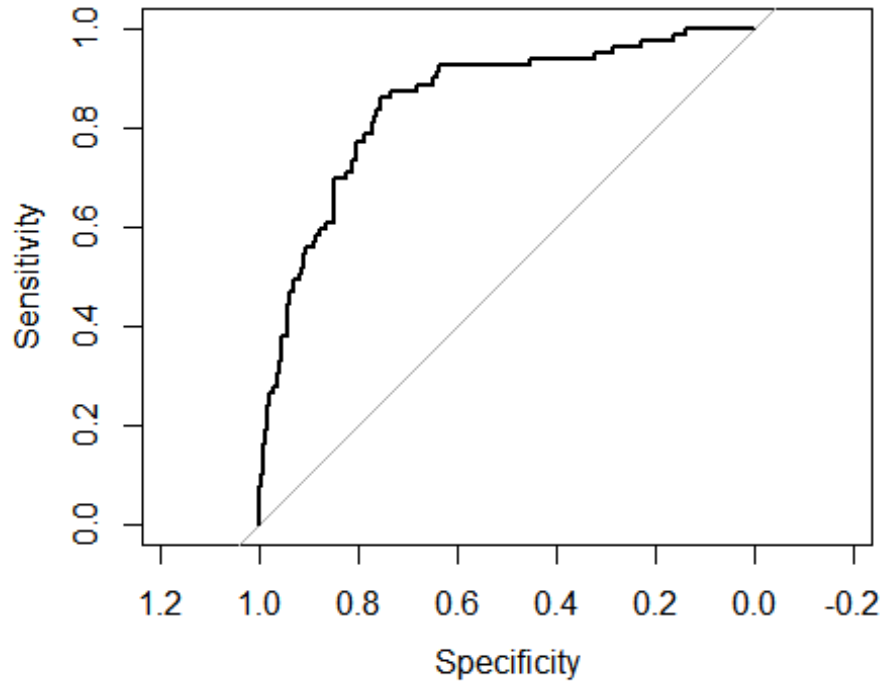
```
##  `VMail Message`       `Day Mins`       `Eve Mins` `CustServ Calls`
##        15.444162         1.046201         1.024934         1.082159
##     `Int'l Plan`     `VMail Plan`   `Night Charge`     `Intl Calls`
##         1.057428        15.467684         1.015481         1.010426
##    `Intl Charge`
##         1.014511
```

```r
pred <- predict(fit,newdata = datatest,type = 'response')
```

```r
#check the AUC curve
library(pROC)

## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```




```r
g <- roc(Churn ~ pred, data = datatest)
g

##
## Call:
## roc.formula(formula = Churn ~ pred, data = datatest)
##
## Data: pred in 454 controls (Churn 0) < 79 cases (Churn 1).
## Area under the curve: 0.8498

plot(g)
```




```
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
#with default prob cut 0.50
datatest$pred_Churn <- ifelse(pred<0.7,'yes','no')

table(datatest$pred_Churn,datatest$Churn)
```

```
##
##         0   1
##   no    1   7
##   yes 453  72
```

```
#training split of churn classes
round(table(datatrain$Churn)/nrow(datatrain),2)*100
```

```
##
##  0  1
## 86 14
```

```
# test split of churn classes
round(table(datatest$Churn)/nrow(datatest),2)*100
```

```
##
##  0  1
## 85 15
```

```
#predicted split of churn classes
round(table(datatest$pred_Churn)/nrow(datatest),2)*100
```

```
##
##  no yes
##   2  98
```

```
#create confusion matrix
#confusionMatrix(datatest$Churn,datatest$pred_churn)
```

```
#how do we create a cross validation scheme
control <- trainControl(method = 'repeatedcv',
                        number = 10,
                        repeats = 3)
seed <-7
metric <- 'Accuracy'
set.seed(seed)
fit_default <- train(Churn~.,
                     data = datatrain,
                     method = 'glm',
                     metric = NaN,
                     trControl = control)
```

```
## Warning in train.default(x, y, weights = w, ...): You are trying to do
## regression and your outcome only has two possible values Are you trying to
## do classification? If so, use a 2 level factor as your outcome column.
```

```
## Warning in train.default(x, y, weights = w, ...): The metric "NaN" was not
## in the result set. RMSE will be used instead.
```

```
print(fit_default)
```

```
## Generalized Linear Model
##
## 2800 samples
##   17 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 2520, 2520, 2520, 2520, 2520, 2520, ...
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.3214677  0.1639757  0.2214014
```

```
library(caret)
varImp(step_fit)
```

```
##                     Overall
## `VMail Message`   1.584309
## `Day Mins`       10.831501
## `Eve Mins`        5.529505
## `CustServ Calls` 11.841926
## `Int'l Plan`     12.483795
## `VMail Plan`      2.945711
## `Night Charge`    3.046687
## `Intl Calls`      3.181969
## `Intl Charge`     3.975115
```

```
varImp(fit_default)
```

```
## glm variable importance
##
##                           Overall
## `\\`Int'l Plan\\``     100.00000
## `\\`CustServ Calls\\``  86.85313
## `\\`Intl Calls\\``      21.44255
## `\\`VMail Plan\\``      17.93683
## `\\`VMail Message\\``    6.74396
## `\\`Eve Mins\\``         6.12577
## `\\`Eve Charge\\``       6.10076
## `\\`Night Charge\\``     4.01050
## `\\`Night Mins\\``       3.98598
## `\\`Account Length\\``   2.82527
## `\\`Day Calls\\``        1.64410
## `\\`Intl Charge\\``      0.69957
## `\\`Intl Mins\\``        0.59750
## `\\`Day Mins\\``         0.45869
## `\\`Day Charge\\``       0.43391
## `\\`Eve Calls\\``        0.03577
## `\\`Night Calls\\``      0.00000
```

```
library(devtools)
```

```
install_github("riv","tomasgreif")
```

```
## Warning: Username parameter is deprecated. Please use tomasgreif/riv
```

```
## Skipping install of 'woe' from a github remote, the SHA1 (43fcf268) has
not changed since last install.
##   Use `force = TRUE` to force installation
```

```
install_github("woe","tomasgreif")
```

```
## Warning: Username parameter is deprecated. Please use tomasgreif/woe
```

```
## Skipping install of 'woe' from a github remote, the SHA1 (43fcf268) has
not changed since last install.
##   Use `force = TRUE` to force installation
```

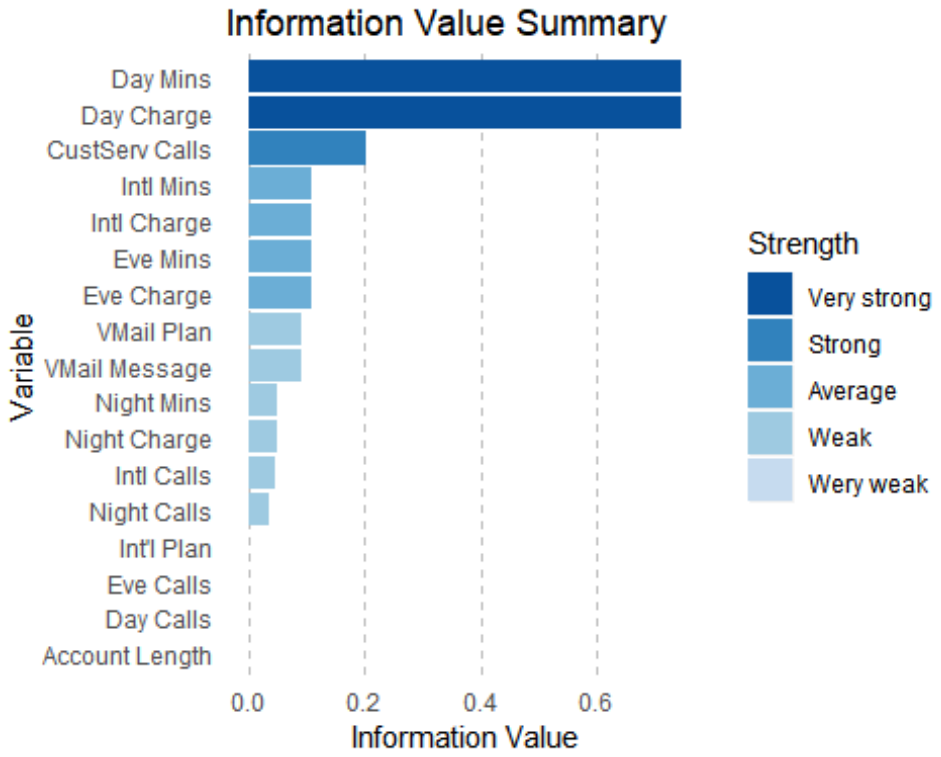
```r
library(woe)

library(riv)

## Loading required package: rrcov

## Loading required package: robustbase

## Scalable Robust Estimators with High Breakdown Point (version 1.4-4)

## Loading required package: quantreg

## Loading required package: SparseM

##
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':
##
##     backsolve

datatrain<-as.data.frame(datatrain)

iv_df <- iv.mult(datatrain, y="Churn", summary=TRUE, verbose=TRUE)

## Started processing of data frame: datatrain
## Calling iv.num for variable: Account Length
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 1
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0
##   Formatting output
## Calling iv.num for variable: VMail Message
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 2
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.09
##   Formatting output
## Calling iv.num for variable: Day Mins
##   Building rpart model
##   Model finished
##   Sending model to tree parser
```

```
##   Rules parsed: 5
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.74
##   Formatting output
## Calling iv.num for variable: Eve Mins
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 5
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.11
##   Formatting output
## Calling iv.num for variable: Night Mins
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 4
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.05
##   Formatting output
## Calling iv.num for variable: Intl Mins
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 4
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.11
##   Formatting output
## Calling iv.num for variable: CustServ Calls
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 3
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.2
```

```
##   Formatting output
## Calling iv.num for variable: Int'l Plan
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 1
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0
##   Formatting output
## Calling iv.num for variable: VMail Plan
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 2
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.09
##   Formatting output
## Calling iv.num for variable: Day Calls
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 1
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0
##   Formatting output
## Calling iv.num for variable: Day Charge
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 5
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.74
##   Formatting output
## Calling iv.num for variable: Eve Calls
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 1
```

```
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0
##   Formatting output
## Calling iv.num for variable: Eve Charge
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 5
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.11
##   Formatting output
## Calling iv.num for variable: Night Calls
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 4
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.04
##   Formatting output
## Calling iv.num for variable: Night Charge
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 4
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.05
##   Formatting output
## Calling iv.num for variable: Intl Calls
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 2
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.04
##   Formatting output
```

```
## Calling iv.num for variable: Intl Charge
##   Building rpart model
##   Model finished
##   Sending model to tree parser
##   Rules parsed: 4
##   Mapping nodes to data
##     SQL Merge
##     DF Merge
##   Calling iv.str for nodes
## Information Value 0.11
##   Formatting output
## Preparing summary

iv_df

##          Variable InformationValue Bins ZeroBins    Strength
## 1      Day Charge       0.74480154    5        0 Very strong
## 2        Day Mins       0.74480154    5        0 Very strong
## 3  CustServ Calls       0.20352690    3        0      Strong
## 4     Intl Charge       0.11004327    4        0     Average
## 5       Intl Mins       0.11004327    4        0     Average
## 6      Eve Charge       0.10851920    5        0     Average
## 7        Eve Mins       0.10851920    5        0     Average
## 8   VMail Message       0.09143542    2        0        Weak
## 9      VMail Plan       0.09143542    2        0        Weak
## 10     Night Mins       0.05097587    4        0        Weak
## 11   Night Charge       0.04953913    4        0        Weak
## 12     Intl Calls       0.04487300    2        0        Weak
## 13    Night Calls       0.03625925    4        0        Weak
## 14 Account Length       0.00000000    1        0   very weak
## 15      Day Calls       0.00000000    1        0   very weak
## 16      Eve Calls       0.00000000    1        0   very weak
## 17     Int'l Plan       0.00000000    1        0   very weak

# Plot information value summary

iv.plot.summary(iv_df)
```




```
library(ISLR)

#Check what models are better then others
#logistic_model <- predict(logit, mydata_test, type = "response")
#svm_predict <- predict(svm_model, mydata_test, type = "response")
#rf_predict <- predict(rf, mydata_test, type = "response")
#mydata_test$Yhat1 <- logistic_model
#mydata_test$Yhat2 <- svm_predict
#mydata_test$Yhat3 <- rf_predict

#setting threshold parameters
#predict1 <- function(x) ifelse(logistic_model > x, 1, 0)
#predict2 <- function(x) ifelse(svm_predict > x, 1, 0)
#predict3 <- function(x) ifelse(rf_predict > x, 1, 0)


#confusionMatrix(predict1(.5),mydata_test$Churn)
#confusionMatrix(predict2(.5),mydata_test$Churn)
```