# 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)
#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"
```

```
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 = datatra
in)
##
## 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
```

```r
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
```

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

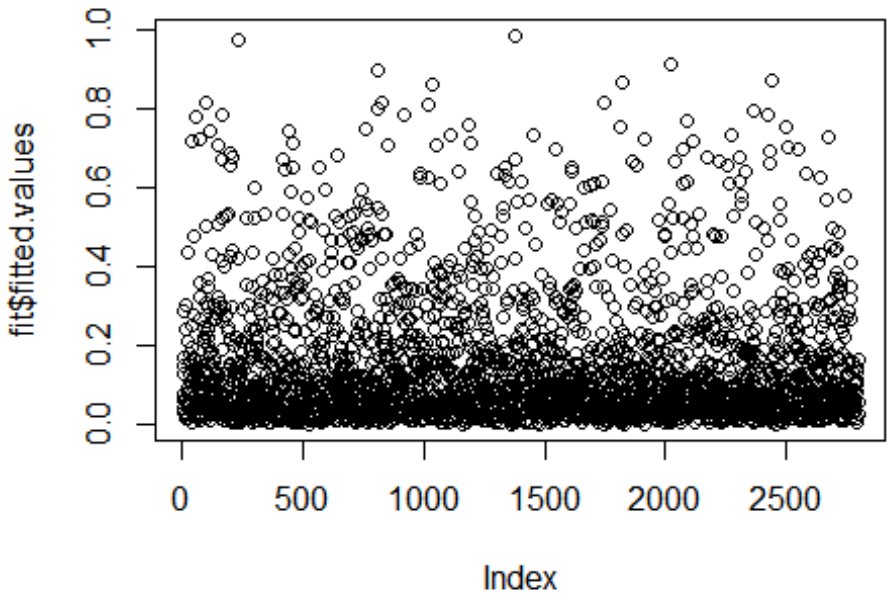
```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Churn
##
## Terms added sequentially (first to last)
##
##
##                  Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                              2799     2310.9
## `Account Length`  1    0.716      2798     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
```

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

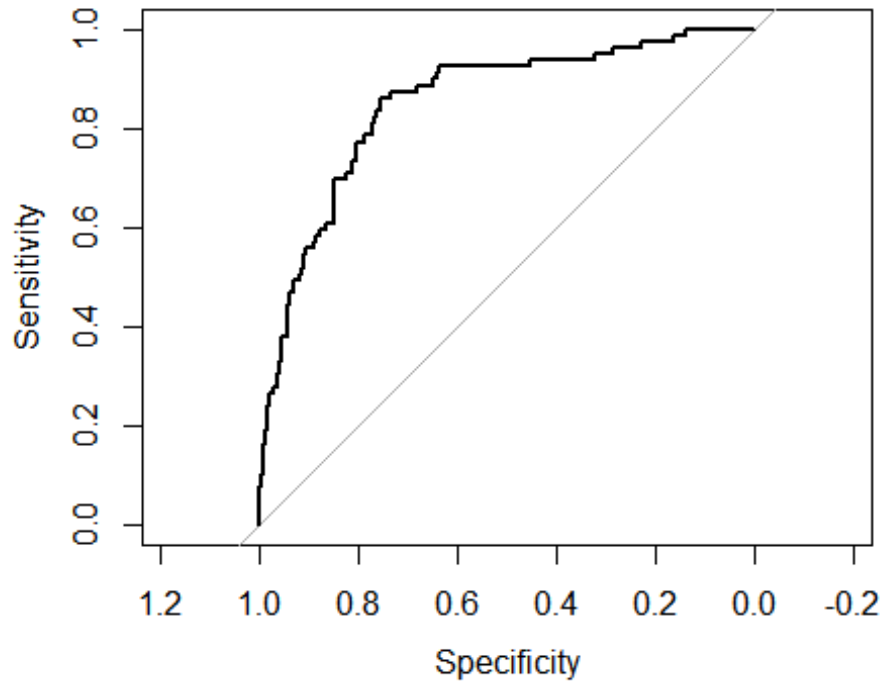
```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Churn
##
## Terms added sequentially (first to last)
##
##
##                  Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                              2799     2310.9
## `VMail Message`   1   23.311      2798     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

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

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

#check for multicollinearity
library(car)

## Loading required package: carData

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

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

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

#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
```




```
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 no
t 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 no
t changed since last install.
##   Use `force = TRUE` to force installation
```

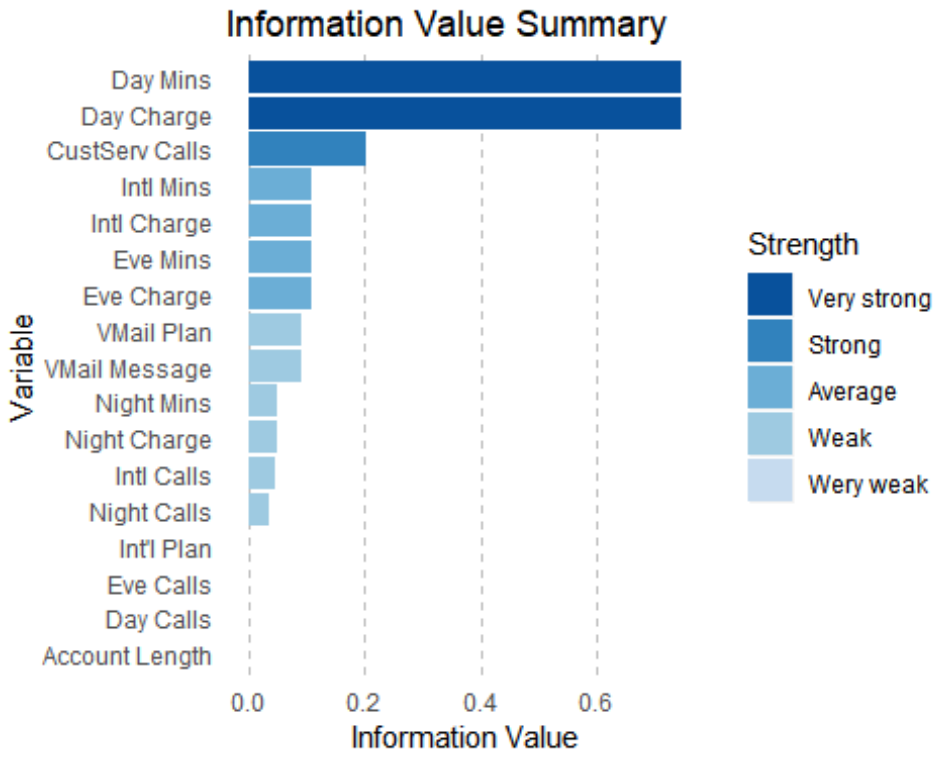
```r
library(woe)

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   Wery weak
## 15      Day Calls       0.00000000    1        0   Wery weak
## 16      Eve Calls       0.00000000    1        0   Wery weak
## 17     Int'l Plan       0.00000000    1        0   Wery weak

# Plot information value summary

iv.plot.summary(iv_df)
```


Information Value Summary
Day Mins
Day Charge
CustServ Calls
Intl Mins
Intl Charge
Eve Mins
Eve Charge
VMail Plan
VMail Message
Night Mins
Night Charge
Intl Calls
Night Calls
Int'l Plan
Eve Calls
Day Calls
Account Length
Variable
0.0
0.2
0.4
0.6
Information Value
Strength
Very strong
Strong
Average
Weak
Wery weak