Credit Card Default Case Study

Dheeraj Kumar

2/13/2022

R Markdown

Credit card Default Case Study

Importing libraries

```
#install.packages("ggplot2")
#install.packages("dplyr")
#install.packages("psych")
#install.packages("MASS")
#install.packages("tidyverse")
#install.packages("corrplot")
#install.packages("leaps")
#install.packages("rpart")
#install.packages("mgcv")
#install.packages("glmnet")
#install.packages("boot")
#install.packages("caret")
#install.packages("rpart.plot")
#install.packages("tree")
#install.packages("plotmo")
#install.packages("ROCR")
#install.packages("PRROC")
#install.packages("pROC")
```

```
library(ggplot2)
library(dplyr)
library(psych)
library(MASS)
library(tidyverse)
library(corrplot)
library(leaps)
library(rpart)
library(mgcv)
library(glmnet)
library(boot)
library(caret)
library(rpart.plot)
library(tree)
library(plotmo)
library(ROCR)
```

```r
library(PRROC)
library(pROC)
```

Importing the dataset

```r
credit_card_data <- read.csv("D:/UC/Classes/Spring/Data Mining 1/Credit Card
Data/default_of_credit_card_clients.csv")
```

Checking head

```r
head(credit_card_data)
```

```
##   ID LIMIT_BAL SEX EDUCATION MARRIAGE AGE PAY_0 PAY_2 PAY_3 PAY_4 PAY_5 PA
Y_6
## 1  1     20000   2         2        1  24     2     2    -1    -1    -2
-2
## 2  2    120000   2         2        2  26    -1     2     0     0     0
2
## 3  3     90000   2         2        2  34     0     0     0     0     0
0
## 4  4     50000   2         2        1  37     0     0     0     0     0
0
## 5  5     50000   1         2        1  57    -1     0    -1     0     0
0
## 6  6     50000   1         1        2  37     0     0     0     0     0
0
##   BILL_AMT1 BILL_AMT2 BILL_AMT3 BILL_AMT4 BILL_AMT5 BILL_AMT6 PAY_AMT1 PAY
_AMT2
## 1      3913      3102       689         0         0         0        0
689
## 2      2682      1725      2682      3272      3455      3261        0
1000
## 3     29239     14027     13559     14331     14948     15549     1518
1500
## 4     46990     48233     49291     28314     28959     29547     2000
2019
## 5      8617      5670     35835     20940     19146     19131     2000
36681
## 6     64400     57069     57608     19394     19619     20024     2500
1815
##   PAY_AMT3 PAY_AMT4 PAY_AMT5 PAY_AMT6 default.payment.next.month
## 1        0        0        0        0                          1
## 2     1000     1000        0     2000                          1
## 3     1000     1000     1000     5000                          0
## 4     1200     1100     1069     1000                          0
## 5    10000     9000      689      679                          0
## 6      657     1000     1000      800                          0
```

Renaming target variable

```r
credit_card_data <- rename(credit_card_data, default_payment_next_month = def
ault.payment.next.month)
```

Checking structure of dataset

```
str(credit_card_data)

## 'data.frame':	30000 obs. of  25 variables:
##  $ ID                        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ LIMIT_BAL                 : int  20000 120000 90000 50000 50000 50000 5
00000 100000 140000 20000 ...
##  $ SEX                       : int  2 2 2 2 1 1 1 2 2 1 ...
##  $ EDUCATION                 : int  2 2 2 2 2 1 1 2 3 3 ...
##  $ MARRIAGE                  : int  1 2 2 1 1 2 2 2 1 2 ...
##  $ AGE                       : int  24 26 34 37 57 37 29 23 28 35 ...
##  $ PAY_0                     : int  2 -1 0 0 -1 0 0 0 0 -2 ...
##  $ PAY_2                     : int  2 2 0 0 0 0 0 -1 0 -2 ...
##  $ PAY_3                     : int  -1 0 0 0 -1 0 0 -1 2 -2 ...
##  $ PAY_4                     : int  -1 0 0 0 0 0 0 0 0 -2 ...
##  $ PAY_5                     : int  -2 0 0 0 0 0 0 0 0 -1 ...
##  $ PAY_6                     : int  -2 2 0 0 0 0 0 -1 0 -1 ...
##  $ BILL_AMT1                 : int  3913 2682 29239 46990 8617 64400 36796
5 11876 11285 0 ...
##  $ BILL_AMT2                 : int  3102 1725 14027 48233 5670 57069 41202
3 380 14096 0 ...
##  $ BILL_AMT3                 : int  689 2682 13559 49291 35835 57608 44500
7 601 12108 0 ...
##  $ BILL_AMT4                 : int  0 3272 14331 28314 20940 19394 542653
221 12211 0 ...
##  $ BILL_AMT5                 : int  0 3455 14948 28959 19146 19619 483003
-159 11793 13007 ...
##  $ BILL_AMT6                 : int  0 3261 15549 29547 19131 20024 473944
567 3719 13912 ...
##  $ PAY_AMT1                  : int  0 0 1518 2000 2000 2500 55000 380 3329
0 ...
##  $ PAY_AMT2                  : int  689 1000 1500 2019 36681 1815 40000 60
1 0 0 ...
##  $ PAY_AMT3                  : int  0 1000 1000 1200 10000 657 38000 0 432
0 ...
##  $ PAY_AMT4                  : int  0 1000 1000 1100 9000 1000 20239 581 1
000 13007 ...
##  $ PAY_AMT5                  : int  0 0 1000 1069 689 1000 13750 1687 1000
1122 ...
##  $ PAY_AMT6                  : int  0 2000 5000 1000 679 800 13770 1542 10
00 0 ...
##  $ default_payment_next_month: int  1 1 0 0 0 0 0 0 0 0 ...
```

Checking dimensions of dataset

```
dim(credit_card_data)

## [1] 30000    25
```

Checking for any missing values

```
sum(is.na(credit_card_data))
```

```
## [1] 0
```

Checking statistics on variables

```
summary(credit_card_data)

##        ID          LIMIT_BAL            SEX          EDUCATION
##  Min.   :    1   Min.   :  10000   Min.   :1.000   Min.   :0.000
##  1st Qu.: 7501   1st Qu.:  50000   1st Qu.:1.000   1st Qu.:1.000
##  Median :15000   Median : 140000   Median :2.000   Median :2.000
##  Mean   :15000   Mean   : 167484   Mean   :1.604   Mean   :1.853
##  3rd Qu.:22500   3rd Qu.: 240000   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :30000   Max.   :1000000   Max.   :2.000   Max.   :6.000
##     MARRIAGE          AGE            PAY_0              PAY_2
##  Min.   :0.000   Min.   :21.00   Min.   :-2.0000   Min.   :-2.0000
##  1st Qu.:1.000   1st Qu.:28.00   1st Qu.:-1.0000   1st Qu.:-1.0000
##  Median :2.000   Median :34.00   Median : 0.0000   Median : 0.0000
##  Mean   :1.552   Mean   :35.49   Mean   :-0.0167   Mean   :-0.1338
##  3rd Qu.:2.000   3rd Qu.:41.00   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :3.000   Max.   :79.00   Max.   : 8.0000   Max.   : 8.0000
##      PAY_3              PAY_4              PAY_5              PAY_6
##  Min.   :-2.0000   Min.   :-2.0000   Min.   :-2.0000   Min.   :-2.0000
##  1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   :-0.1662   Mean   :-0.2207   Mean   :-0.2662   Mean   :-0.2911
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   : 8.0000   Max.   : 8.0000   Max.   : 8.0000   Max.   : 8.0000
##    BILL_AMT1         BILL_AMT2        BILL_AMT3         BILL_AMT4
##  Min.   :-165580   Min.   :-69777   Min.   :-157264   Min.   :-170000
##  1st Qu.:   3559   1st Qu.:  2985   1st Qu.:   2666   1st Qu.:   2327
##  Median :  22382   Median : 21200   Median :  20089   Median :  19052
##  Mean   :  51223   Mean   : 49179   Mean   :  47013   Mean   :  43263
##  3rd Qu.:  67091   3rd Qu.: 64006   3rd Qu.:  60165   3rd Qu.:  54506
##  Max.   : 964511   Max.   :983931   Max.   :1664089   Max.   : 891586
##    BILL_AMT5        BILL_AMT6          PAY_AMT1         PAY_AMT2
##  Min.   :-81334   Min.   :-339603   Min.   :     0   Min.   :      0
##  1st Qu.:  1763   1st Qu.:   1256   1st Qu.:  1000   1st Qu.:    833
##  Median : 18105   Median :  17071   Median :  2100   Median :   2009
##  Mean   : 40311   Mean   :  38872   Mean   :  5664   Mean   :   5921
##  3rd Qu.: 50191   3rd Qu.:  49198   3rd Qu.:  5006   3rd Qu.:   5000
##  Max.   :927171   Max.   : 961664   Max.   :873552   Max.   :1684259
##     PAY_AMT3         PAY_AMT4         PAY_AMT5           PAY_AMT6
##  Min.   :     0   Min.   :     0   Min.   :     0.0   Min.   :     0.0
##  1st Qu.:   390   1st Qu.:   296   1st Qu.:   252.5   1st Qu.:   117.8
##  Median :  1800   Median :  1500   Median :  1500.0   Median :  1500.0
##  Mean   :  5226   Mean   :  4826   Mean   :  4799.4   Mean   :  5215.5
##  3rd Qu.:  4505   3rd Qu.:  4013   3rd Qu.:  4031.5   3rd Qu.:  4000.0
##  Max.   :896040   Max.   :621000   Max.   :426529.0   Max.   :528666.0
##  default_payment_next_month
##  Min.   :0.0000
##  1st Qu.:0.0000
```

```
##   Median :0.0000
##   Mean   :0.2212
##   3rd Qu.:0.0000
##   Max.   :1.0000
```

Converting required variables to factors

```
credit_card_data$SEX <- as.factor(credit_card_data$SEX)
credit_card_data$MARRIAGE <- as.factor(credit_card_data$MARRIAGE)
credit_card_data$EDUCATION <- as.factor(credit_card_data$EDUCATION)
credit_card_data$default_payment_next_month <- as.factor(credit_card_data$def
ault_payment_next_month)
credit_card_data$PAY_0 <- as.factor(credit_card_data$PAY_0)
credit_card_data$PAY_2 <- as.factor(credit_card_data$PAY_2)
credit_card_data$PAY_3 <- as.factor(credit_card_data$PAY_3)
credit_card_data$PAY_4 <- as.factor(credit_card_data$PAY_4)
credit_card_data$PAY_5 <- as.factor(credit_card_data$PAY_5)
credit_card_data$PAY_6 <- as.factor(credit_card_data$PAY_6)
```

Dropping ID column from dataset

```
credit_card_data = subset(credit_card_data, select = -c(ID))
```

Setting unique seed

```
set.seed(14283873)
```

Splitting dataset into Train and Test

```
index <- sample(nrow(credit_card_data), nrow(credit_card_data) * 0.80)
```

Creating Train and Test datasets

```
credit_card_data_train <- credit_card_data[index, ]
credit_card_data_test <- credit_card_data[-index, ]
```

Checking dimensions of Train and Test dataset

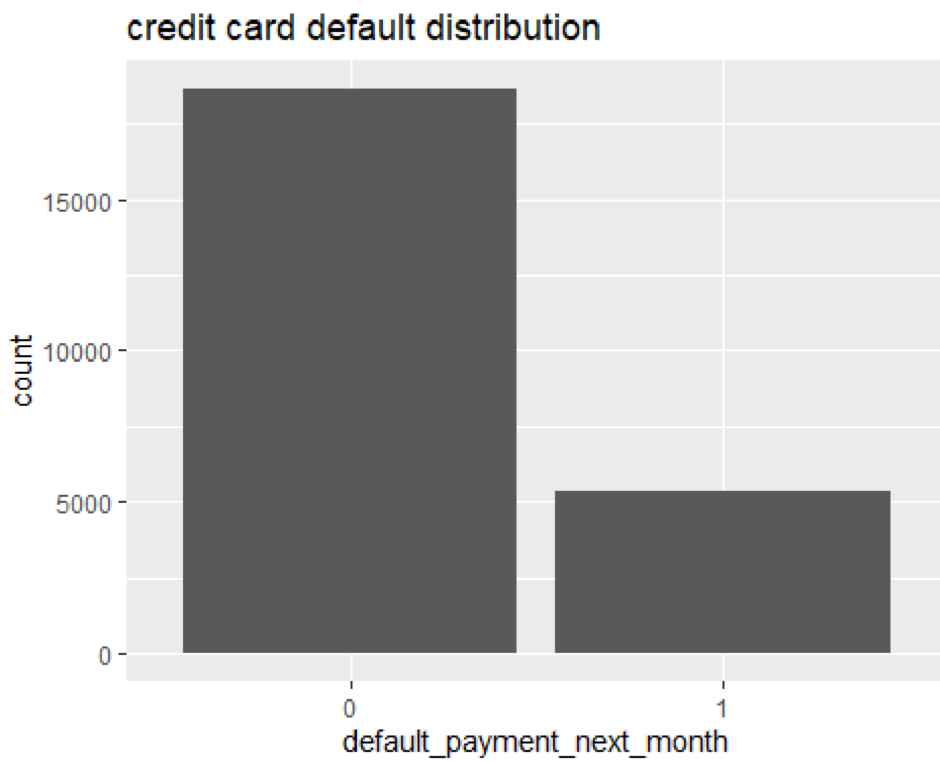
```
dim(credit_card_data_train)  ## 24000,24

## [1] 24000    24

dim(credit_card_data_test)   ## 6000,24

## [1] 6000   24
```

#———— A-1. Analysis with 80% Training Data ————————#

Performing EDA

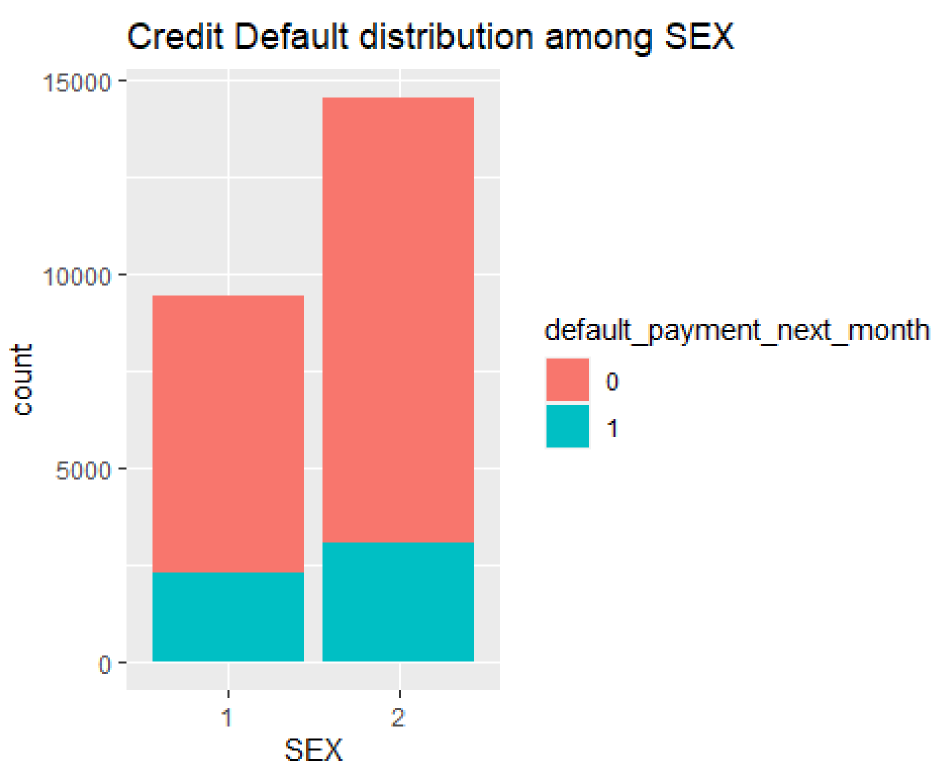
Countplot of default_payment_next_month

```
ggplot(credit_card_data_train,aes(default_payment_next_month))+geom_bar()+ggt
itle("credit card default distribution")
```



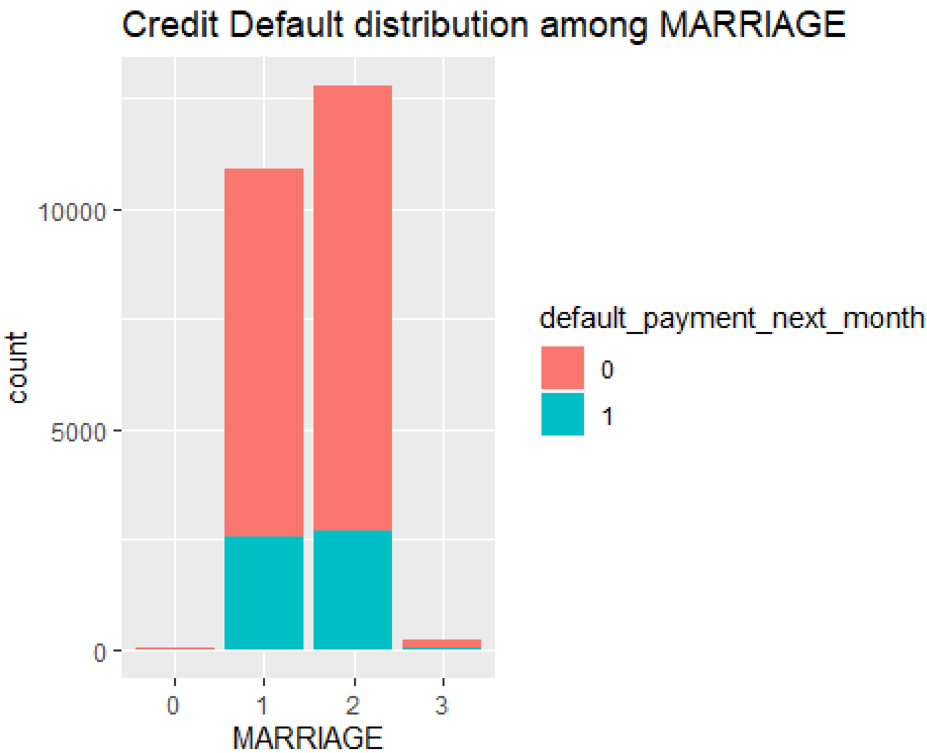
Distribution of default_payment_next_month vs SEX

```
ggplot(credit_card_data_train,aes(SEX))+geom_bar(aes(fill=default_payment_nex
t_month))+ggtitle("Credit Default distribution among SEX")
```



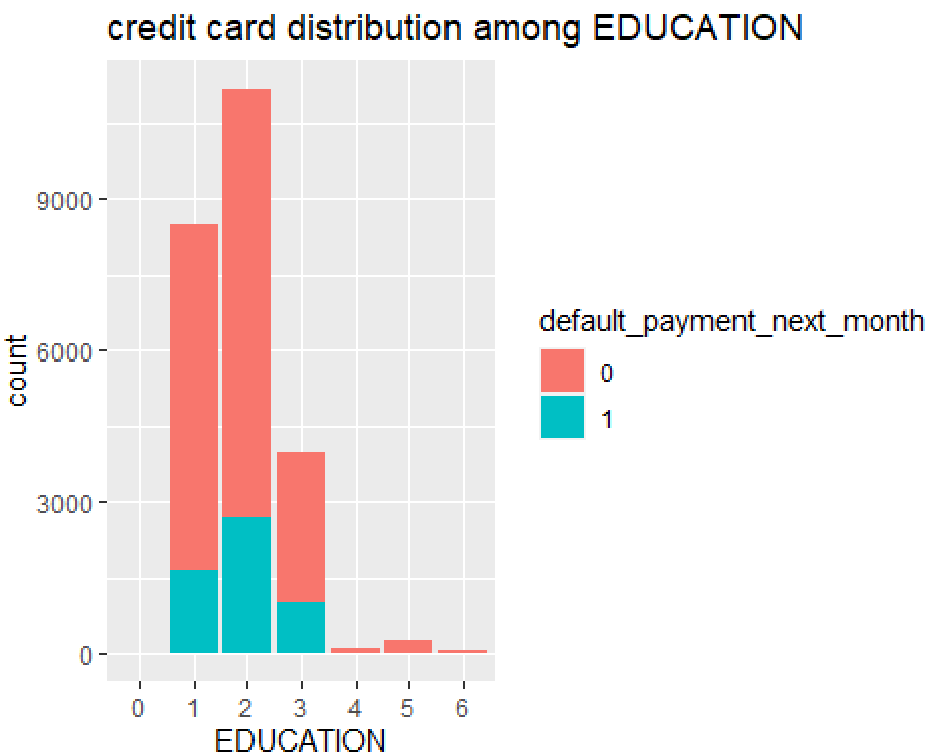
Distribution of default_payment_next_month vs MARRIAGE

```
ggplot(credit_card_data_train,aes(MARRIAGE))+geom_bar(aes(fill=default_paymen
t_next_month))+ggtitle("Credit Default distribution among MARRIAGE")
```



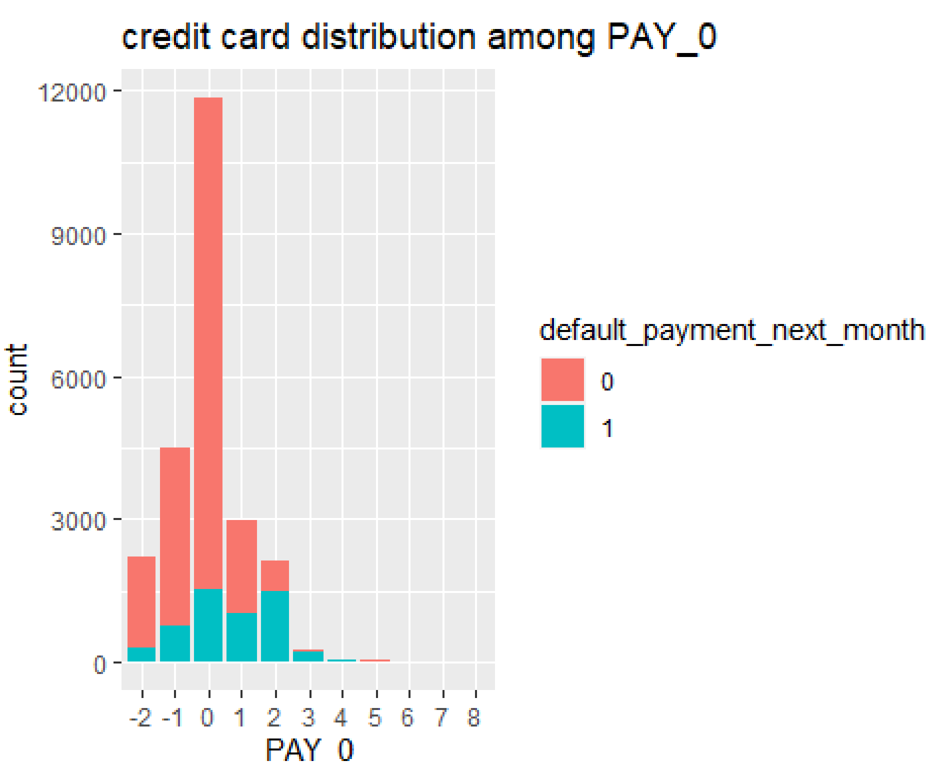
Distribution of default_payment_next_month vs EDUCATION

```
ggplot(credit_card_data_train,aes(EDUCATION))+geom_bar(aes(fill=default_payme
nt_next_month))+ggtitle("credit card distribution among EDUCATION")
```



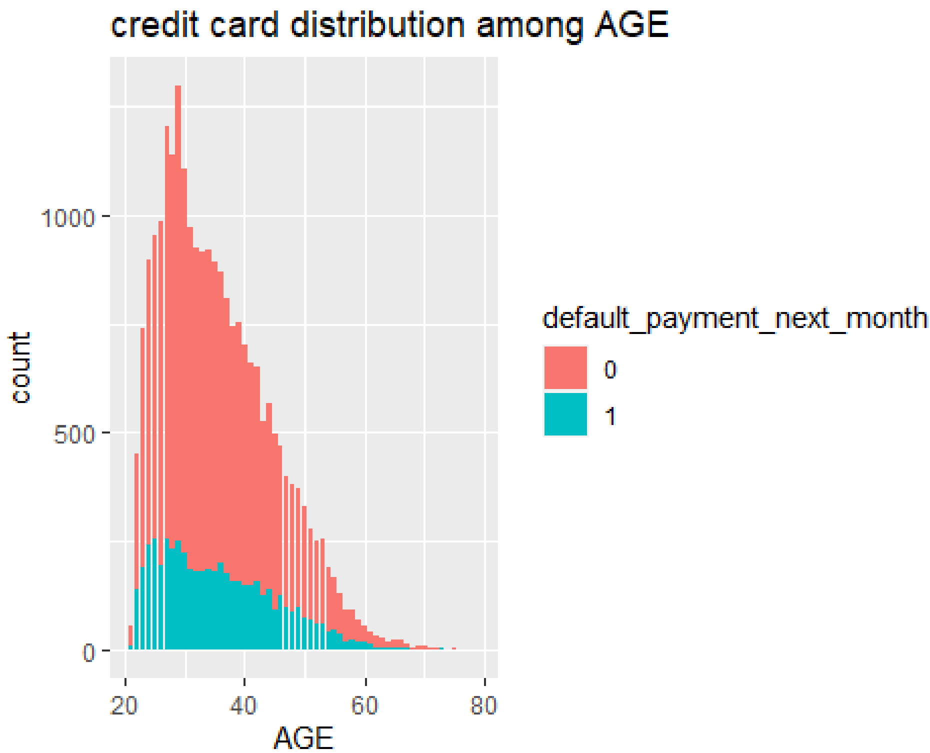
Distribution of default_payment_next_month vs PAY_0

```
ggplot(credit_card_data_train,aes(PAY_0))+geom_bar(aes(fill=default_payment_n
ext_month))+ggtitle("credit card distribution among PAY_0")
```



Distribution of default_payment_next_month vs AGE

```
ggplot(credit_card_data_train,aes(AGE))+geom_bar(aes(fill=default_payment_nex
t_month))+ggtitle("credit card distribution among AGE")
```



#———— A-2. Logistic Regression and Variable Selection ————————#

##——————- Part 1 ——————-##

##——————- Full Model ——————-##

Logistic Regression with full model

```
full.glm <- glm(default_payment_next_month ~ ., family=binomial, data = credi
t_card_data_train)
full.glm.summary <- summary(full.glm)
full.glm.summary

##
## Call:
## glm(formula = default_payment_next_month ~ ., family = binomial,
##     data = credit_card_data_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5178  -0.5993  -0.5098  -0.2951   3.5160
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept) -1.608e+01  2.604e+02  -0.062 0.950753    
## LIMIT_BAL   -1.757e-06  1.949e-07  -9.017  < 2e-16 ***
## SEX2        -1.314e-01  3.618e-02  -3.631 0.000282 ***
## EDUCATION1   1.283e+01  2.604e+02   0.049 0.960703    
## EDUCATION2   1.287e+01  2.604e+02   0.049 0.960595    
## EDUCATION3   1.279e+01  2.604e+02   0.049 0.960819    
## EDUCATION4   1.156e+01  2.604e+02   0.044 0.964598    
## EDUCATION5   1.161e+01  2.604e+02   0.045 0.964429    
## EDUCATION6   1.261e+01  2.604e+02   0.048 0.961392    
## MARRIAGE1    2.051e+00  6.532e-01   3.141 0.001686 ** 
## MARRIAGE2    1.912e+00  6.533e-01   2.927 0.003426 ** 
## MARRIAGE3    2.199e+00  6.723e-01   3.271 0.001071 ** 
## AGE          3.902e-03  2.205e-03   1.770 0.076796 .  
## PAY_0-1      6.146e-01  1.215e-01   5.060 4.19e-07 ***
## PAY_00      -1.853e-01  1.311e-01  -1.414 0.157489    
## PAY_01       8.259e-01  9.426e-02   8.762  < 2e-16 ***
## PAY_02       2.109e+00  1.190e-01  17.720  < 2e-16 ***
## PAY_03       1.956e+00  1.860e-01  10.511  < 2e-16 ***
## PAY_04       1.985e+00  3.493e-01   5.682 1.33e-08 ***
## PAY_05       1.438e+00  5.456e-01   2.636 0.008377 ** 
## PAY_06      -7.054e-01  1.375e+00  -0.513 0.608038    
## PAY_07      -1.262e+01  6.187e+02  -0.020 0.983731    
## PAY_08      -1.309e+01  8.827e+02  -0.015 0.988169    
## PAY_2-1     -3.672e-01  1.279e-01  -2.871 0.004094 ** 
## PAY_20      -1.044e-01  1.550e-01  -0.673 0.500779    
## PAY_21      -6.855e-01  5.795e-01  -1.183 0.236824    
## PAY_22      -4.859e-02  1.307e-01  -0.372 0.710042    
## PAY_23       1.405e-02  2.016e-01   0.070 0.944440    
## PAY_24      -5.142e-01  3.633e-01  -1.416 0.156899    
## PAY_25       2.323e+00  1.180e+00   1.969 0.049005 *  
## PAY_26       1.507e+01  6.187e+02   0.024 0.980574    
## PAY_27       9.470e-01  9.899e+02   0.001 0.999237    
## PAY_28       1.447e+01  1.134e+03   0.013 0.989826    
## PAY_3-1      7.783e-02  1.224e-01   0.636 0.524836    
## PAY_30       1.532e-01  1.412e-01   1.085 0.278116    
## PAY_31      -1.265e+01  8.827e+02  -0.014 0.988565    
## PAY_32       4.772e-01  1.428e-01   3.341 0.000834 ***
## PAY_33       4.969e-01  2.540e-01   1.956 0.050419 .  
## PAY_34      -4.239e-01  4.627e-01  -0.916 0.359540    
## PAY_35      -9.385e-01  8.686e-01  -1.080 0.279937    
## PAY_36       1.443e+01  4.479e+02   0.032 0.974288    
## PAY_37       1.389e-01  1.005e+00   0.138 0.890012    
## PAY_38      -2.578e+01  4.974e+02  -0.052 0.958669    
## PAY_4-1     -2.209e-01  1.229e-01  -1.796 0.072438 .  
## PAY_40      -2.460e-01  1.371e-01  -1.795 0.072700 .  
## PAY_41       2.852e+01  1.248e+03   0.023 0.981773    
## PAY_42       4.012e-02  1.463e-01   0.274 0.783887    
## PAY_43      -1.537e-01  2.794e-01  -0.550 0.582157    
## PAY_44       3.930e-01  5.098e-01   0.771 0.440855    
## PAY_45      -1.338e+00  8.441e-01  -1.585 0.112860    
```

```
## PAY_46      -2.933e+01  7.126e+02  -0.041 0.967170
## PAY_47      -1.951e+00  6.370e+02  -0.003 0.997557
## PAY_48      -3.130e+01  1.052e+03  -0.030 0.976253
## PAY_5-1     -5.877e-02  1.202e-01  -0.489 0.624934
## PAY_50       1.510e-01  1.328e-01   1.136 0.255810
## PAY_52       4.600e-01  1.486e-01   3.095 0.001970 **
## PAY_53       1.542e-01  2.733e-01   0.564 0.572644
## PAY_54       1.306e-01  5.291e-01   0.247 0.804956
## PAY_55       9.381e-01  9.618e-01   0.975 0.329380
## PAY_56       3.930e+01  8.145e+02   0.048 0.961519
## PAY_57       1.556e+01  5.716e+02   0.027 0.978289
## PAY_58       4.186e+01  2.098e+03   0.020 0.984077
## PAY_6-1     -1.054e-01  9.243e-02  -1.141 0.254059
## PAY_60      -3.156e-01  9.907e-02  -3.186 0.001444 **
## PAY_62      -1.543e-02  1.154e-01  -0.134 0.893675
## PAY_63       5.765e-01  2.621e-01   2.199 0.027848 *
## PAY_64      -3.145e-01  5.239e-01  -0.600 0.548240
## PAY_65       8.037e-01  1.019e+00   0.789 0.430157
## PAY_66      -2.789e-01  1.129e+00  -0.247 0.804860
## PAY_67      -1.269e+01  2.813e+02  -0.045 0.964022
## PAY_68       2.879e+01  1.270e+03   0.023 0.981917
## BILL_AMT1   -1.888e-06  1.242e-06  -1.520 0.128527
## BILL_AMT2    2.472e-06  1.676e-06   1.475 0.140315
## BILL_AMT3    2.636e-06  1.458e-06   1.809 0.070482 .
## BILL_AMT4   -3.516e-07  1.467e-06  -0.240 0.810642
## BILL_AMT5   -1.209e-06  1.719e-06  -0.703 0.481785
## BILL_AMT6    4.343e-07  1.368e-06   0.317 0.750872
## PAY_AMT1    -1.159e-05  2.595e-06  -4.467 7.94e-06 ***
## PAY_AMT2    -8.394e-06  2.287e-06  -3.671 0.000242 ***
## PAY_AMT3    -3.247e-06  2.130e-06  -1.524 0.127406
## PAY_AMT4    -9.531e-07  1.912e-06  -0.499 0.618086
## PAY_AMT5    -4.421e-06  2.054e-06  -2.153 0.031328 *
## PAY_AMT6    -3.114e-06  1.512e-06  -2.060 0.039392 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 25495  on 23999  degrees of freedom
## Residual deviance: 20832  on 23917  degrees of freedom
## AIC: 20998
##
## Number of Fisher Scoring iterations: 13
```

Displaying coefficients of the full model

```
data.frame(coef = round(full.glm$coefficients,2))
```

```
##                coef
## (Intercept) -16.08
## LIMIT_BAL     0.00
```

```
## SEX2          -0.13
## EDUCATION1   12.83
## EDUCATION2   12.87
## EDUCATION3   12.79
## EDUCATION4   11.56
## EDUCATION5   11.61
## EDUCATION6   12.61
## MARRIAGE1     2.05
## MARRIAGE2     1.91
## MARRIAGE3     2.20
## AGE           0.00
## PAY_0-1       0.61
## PAY_00       -0.19
## PAY_01        0.83
## PAY_02        2.11
## PAY_03        1.96
## PAY_04        1.99
## PAY_05        1.44
## PAY_06       -0.71
## PAY_07      -12.62
## PAY_08      -13.09
## PAY_2-1      -0.37
## PAY_20       -0.10
## PAY_21       -0.69
## PAY_22       -0.05
## PAY_23        0.01
## PAY_24       -0.51
## PAY_25        2.32
## PAY_26       15.07
## PAY_27        0.95
## PAY_28       14.47
## PAY_3-1       0.08
## PAY_30        0.15
## PAY_31      -12.65
## PAY_32        0.48
## PAY_33        0.50
## PAY_34       -0.42
## PAY_35       -0.94
## PAY_36       14.43
## PAY_37        0.14
## PAY_38      -25.78
## PAY_4-1      -0.22
## PAY_40       -0.25
## PAY_41       28.52
## PAY_42        0.04
## PAY_43       -0.15
## PAY_44        0.39
## PAY_45       -1.34
## PAY_46      -29.33
## PAY_47       -1.95
```

```
## PAY_48      -31.30
## PAY_5-1      -0.06
## PAY_50        0.15
## PAY_52        0.46
## PAY_53        0.15
## PAY_54        0.13
## PAY_55        0.94
## PAY_56       39.30
## PAY_57       15.56
## PAY_58       41.86
## PAY_6-1      -0.11
## PAY_60       -0.32
## PAY_62       -0.02
## PAY_63        0.58
## PAY_64       -0.31
## PAY_65        0.80
## PAY_66       -0.28
## PAY_67      -12.69
## PAY_68       28.79
## BILL_AMT1     0.00
## BILL_AMT2     0.00
## BILL_AMT3     0.00
## BILL_AMT4     0.00
## BILL_AMT5     0.00
## BILL_AMT6     0.00
## PAY_AMT1      0.00
## PAY_AMT2      0.00
## PAY_AMT3      0.00
## PAY_AMT4      0.00
## PAY_AMT5      0.00
## PAY_AMT6      0.00
```

```
AIC(full.glm) # 20997.55
```

```
## [1] 20997.55
```

```
BIC(full.glm) # 21668.67
```

```
## [1] 21668.67
```

Calculating in-sample residual deviance

```
full.glm$deviance  # 20831.55
```

```
## [1] 20831.55
```

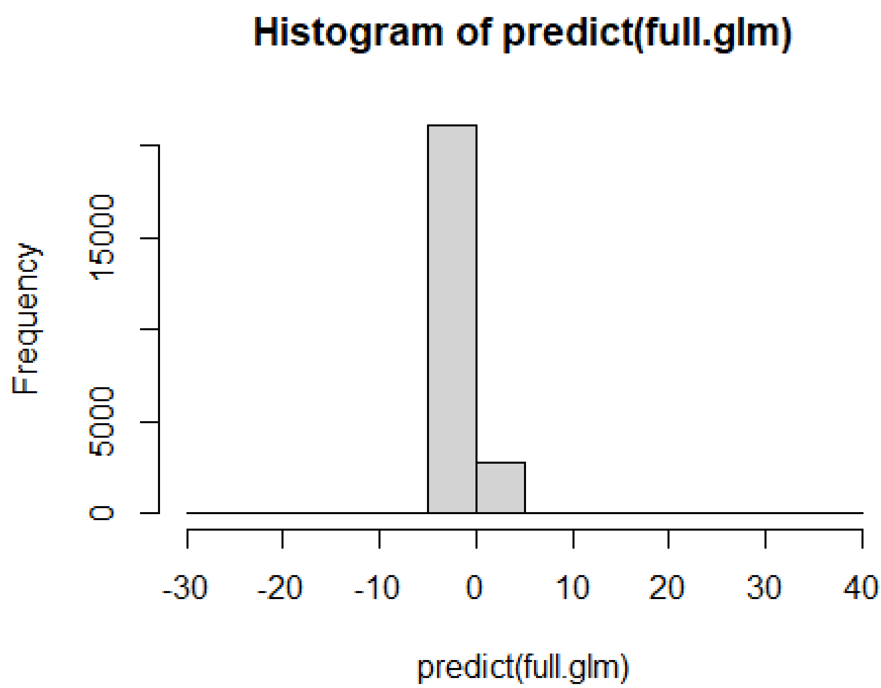
Calculating Model mean residual deviance (in-sample)

```
full.glm$dev/full.glm$df.residual # 0.8709933
```

```
## [1] 0.8709933
```

Plotting histogram for full model

```
hist(predict(full.glm))
```



##———————- Null Model ———————-##

Logistic Regression with Null model

```
null.glm=glm(default_payment_next_month ~ 1, family=binomial, data=credit_car
d_data_train)
null.glm.summary <- summary(null.glm)
null.glm.summary

##
## Call:
## glm(formula = default_payment_next_month ~ 1, family = binomial,
##     data = credit_card_data_train)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -0.711  -0.711  -0.711  -0.711   1.731
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.2461     0.0155  -80.41   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 25495  on 23999  degrees of freedom
## Residual deviance: 25495  on 23999  degrees of freedom
## AIC: 25497
##
## Number of Fisher Scoring iterations: 4
```

Displaying coefficients of the null model

```
data.frame(coef = round(null.glm$coefficients,2))

##               coef
## (Intercept) -1.25

AIC(null.glm) # 25497.03

## [1] 25497.03

BIC(null.glm) # 25505.12

## [1] 25505.12
```

Calculating in-sample residual deviance

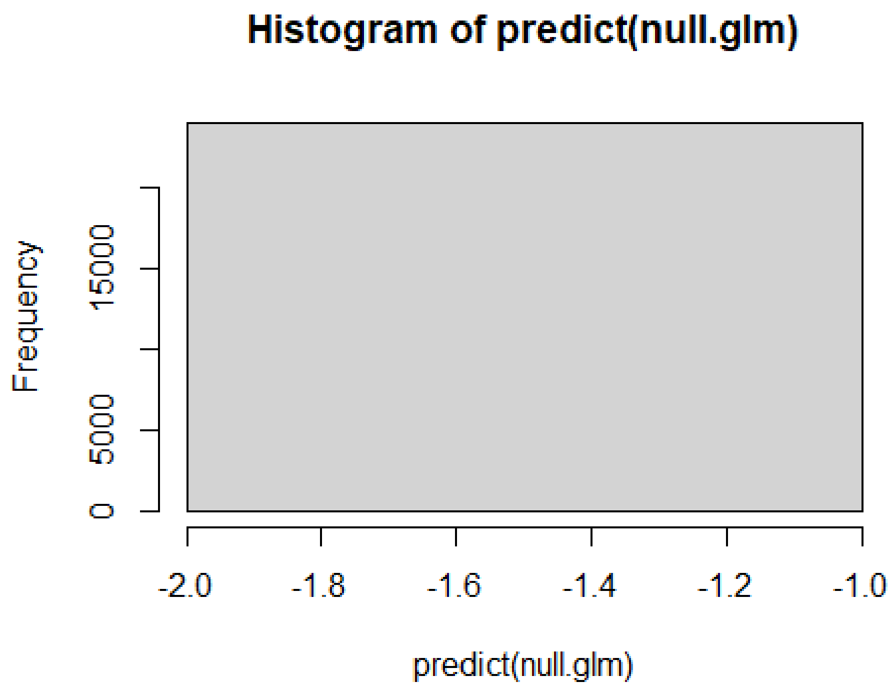
```
null.glm$deviance  # 25495.03

## [1] 25495.03
```

Calculating Model mean residual deviance (in-sample)

```
null.glm$dev/null.glm$df.residual # 1.062337

## [1] 1.062337
```

Plotting histogram for full model

```
hist(predict(null.glm))
```



##——————- Two Variable Model with EDUCATION and PAY_0 ——————-##

```
two_var.glm <- glm(default_payment_next_month ~ credit_card_data_train$EDUCAT
ION + credit_card_data_train$PAY_0, family=binomial, data=credit_card_data_tr
ain)
two_var.glm.summary <- summary(two_var.glm)
two_var.glm.summary

##
## Call:
## glm(formula = default_payment_next_month ~ credit_card_data_train$EDUCATIO
N +
##     credit_card_data_train$PAY_0, family = binomial, data = credit_card_da
ta_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6874  -0.5883  -0.5420  -0.4940   2.6494
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -13.00316   94.73660  -0.137 0.890829
## credit_card_data_train$EDUCATION1  11.07489   94.73658   0.117 0.906938
## credit_card_data_train$EDUCATION2  11.27311   94.73658   0.119 0.905280
## credit_card_data_train$EDUCATION3  11.26048   94.73659   0.119 0.905386
## credit_card_data_train$EDUCATION4   9.63751   94.73776   0.102 0.918972
## credit_card_data_train$EDUCATION5   9.82628   94.73698   0.104 0.917390
```

```
## credit_card_data_train$EDUCATION6  10.89622   94.73773   0.115 0.908434    
## credit_card_data_train$PAY_0-1      0.26171    0.07385   3.544 0.000394 **
*
## credit_card_data_train$PAY_00      -0.11377    0.06856  -1.659 0.097020 .  
## credit_card_data_train$PAY_01       1.16491    0.07356  15.835  < 2e-16 **
*
## credit_card_data_train$PAY_02       2.63277    0.07862  33.487  < 2e-16 **
*
## credit_card_data_train$PAY_03       2.87809    0.15409  18.678  < 2e-16 **
*
## credit_card_data_train$PAY_04       2.78840    0.29885   9.330  < 2e-16 **
*
## credit_card_data_train$PAY_05       1.98568    0.43716   4.542 5.57e-06 **
*
## credit_card_data_train$PAY_06       1.73215    0.81908   2.115 0.034451 *  
## credit_card_data_train$PAY_07       2.67923    0.83949   3.192 0.001415 **
## credit_card_data_train$PAY_08       2.28109    0.52061   4.382 1.18e-05 **
*
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 25495  on 23999  degrees of freedom
## Residual deviance: 21707  on 23983  degrees of freedom
## AIC: 21741
## 
## Number of Fisher Scoring iterations: 11
```

Displaying coefficients of the two variable model

```
data.frame(coef = round(two_var.glm$coefficients,2))
```

```
##                                      coef
## (Intercept)                        -13.00
## credit_card_data_train$EDUCATION1  11.07
## credit_card_data_train$EDUCATION2  11.27
## credit_card_data_train$EDUCATION3  11.26
## credit_card_data_train$EDUCATION4   9.64
## credit_card_data_train$EDUCATION5   9.83
## credit_card_data_train$EDUCATION6  10.90
## credit_card_data_train$PAY_0-1      0.26
## credit_card_data_train$PAY_00      -0.11
## credit_card_data_train$PAY_01       1.16
## credit_card_data_train$PAY_02       2.63
## credit_card_data_train$PAY_03       2.88
## credit_card_data_train$PAY_04       2.79
## credit_card_data_train$PAY_05       1.99
## credit_card_data_train$PAY_06       1.73
## credit_card_data_train$PAY_07       2.68
## credit_card_data_train$PAY_08       2.28
```

```r
AIC(two_var.glm) # 21741.37
```

```
## [1] 21741.37
```

```r
BIC(two_var.glm) # 21878.82
```

```
## [1] 21878.82
```

Calculating in-sample residual deviance

```r
two_var.glm$deviance  # 21707.37
```

```
## [1] 21707.37
```

Calculating Model mean residual deviance (in-sample)

```r
two_var.glm$dev/two_var.glm$df.residual # 0.9051147
```

```
## [1] 0.9051147
```

##———————- Part 2 ———————-##

Performing stepwise variable selection with AIC and BIC

```r
credit_default_glm_back <- step(full.glm)
```

```
## Start:  AIC=20997.55
## default_payment_next_month ~ LIMIT_BAL + SEX + EDUCATION + MARRIAGE +
##     AGE + PAY_0 + PAY_2 + PAY_3 + PAY_4 + PAY_5 + PAY_6 + BILL_AMT1 +
##     BILL_AMT2 + BILL_AMT3 + BILL_AMT4 + BILL_AMT5 + BILL_AMT6 +
##     PAY_AMT1 + PAY_AMT2 + PAY_AMT3 + PAY_AMT4 + PAY_AMT5 + PAY_AMT6
##
##               Df Deviance   AIC
## - BILL_AMT4   1    20832 20996
## - BILL_AMT6   1    20832 20996
## - PAY_AMT4    1    20832 20996
## - BILL_AMT5   1    20832 20996
## <none>             20832 20998
## - BILL_AMT2   1    20834 20998
## - BILL_AMT1   1    20834 20998
## - PAY_AMT3    1    20834 20998
## - AGE         1    20835 20999
## - BILL_AMT3   1    20835 20999
## - PAY_AMT6    1    20836 21000
## - PAY_AMT5    1    20837 21001
## - PAY_2      10    20861 21007
## - SEX         1    20845 21009
## - PAY_5       9    20863 21011
## - PAY_AMT2    1    20848 21012
## - PAY_4      10    20867 21013
## - PAY_6       9    20870 21018
## - MARRIAGE    3    20859 21019
## - PAY_3      10    20873 21019
```

```
## - PAY_AMT1   1    20858 21022
## - EDUCATION  6    20875 21029
## - LIMIT_BAL  1    20917 21081
## - PAY_0     10    22019 22165
##
## Step:  AIC=20995.6
## default_payment_next_month ~ LIMIT_BAL + SEX + EDUCATION + MARRIAGE +
##     AGE + PAY_0 + PAY_2 + PAY_3 + PAY_4 + PAY_5 + PAY_6 + BILL_AMT1 +
##     BILL_AMT2 + BILL_AMT3 + BILL_AMT5 + BILL_AMT6 + PAY_AMT1 +
##     PAY_AMT2 + PAY_AMT3 + PAY_AMT4 + PAY_AMT5 + PAY_AMT6
##
##             Df Deviance   AIC
## - BILL_AMT6  1    20832 20994
## - PAY_AMT4   1    20832 20994
## - BILL_AMT5  1    20832 20994
## <none>            20832 20996
## - BILL_AMT2  1    20834 20996
## - BILL_AMT1  1    20834 20996
## - AGE        1    20835 20997
## - PAY_AMT3   1    20835 20997
## - BILL_AMT3  1    20835 20997
## - PAY_AMT6   1    20836 20998
## - PAY_AMT5   1    20837 20999
## - PAY_2     10    20862 21006
## - SEX        1    20845 21007
## - PAY_5      9    20863 21009
## - PAY_AMT2   1    20848 21010
## - PAY_4     10    20867 21011
## - PAY_6      9    20870 21016
## - MARRIAGE   3    20859 21017
## - PAY_3     10    20873 21017
## - PAY_AMT1   1    20858 21020
## - EDUCATION  6    20875 21027
## - LIMIT_BAL  1    20917 21079
## - PAY_0     10    22019 22163
##
## Step:  AIC=20993.7
## default_payment_next_month ~ LIMIT_BAL + SEX + EDUCATION + MARRIAGE +
##     AGE + PAY_0 + PAY_2 + PAY_3 + PAY_4 + PAY_5 + PAY_6 + BILL_AMT1 +
##     BILL_AMT2 + BILL_AMT3 + BILL_AMT5 + PAY_AMT1 + PAY_AMT2 +
##     PAY_AMT3 + PAY_AMT4 + PAY_AMT5 + PAY_AMT6
##
##             Df Deviance   AIC
## - PAY_AMT4   1    20832 20992
## - BILL_AMT5  1    20833 20993
## <none>            20832 20994
## - BILL_AMT2  1    20834 20994
## - BILL_AMT1  1    20834 20994
## - AGE        1    20835 20995
## - PAY_AMT3   1    20835 20995
```

```
## - BILL_AMT3  1    20835 20995
## - PAY_AMT6   1    20837 20997
## - PAY_AMT5   1    20838 20998
## - PAY_2     10    20862 21004
## - SEX        1    20845 21005
## - PAY_5      9    20863 21007
## - PAY_AMT2   1    20848 21008
## - PAY_4     10    20867 21009
## - PAY_6      9    20870 21014
## - MARRIAGE   3    20859 21015
## - PAY_3     10    20873 21015
## - PAY_AMT1   1    20858 21018
## - EDUCATION  6    20875 21025
## - LIMIT_BAL  1    20917 21077
## - PAY_0     10    22019 22161
##
## Step:  AIC=20991.94
## default_payment_next_month ~ LIMIT_BAL + SEX + EDUCATION + MARRIAGE +
##     AGE + PAY_0 + PAY_2 + PAY_3 + PAY_4 + PAY_5 + PAY_6 + BILL_AMT1 +
##     BILL_AMT2 + BILL_AMT3 + BILL_AMT5 + PAY_AMT1 + PAY_AMT2 +
##     PAY_AMT3 + PAY_AMT5 + PAY_AMT6
##
##             Df Deviance   AIC
## - BILL_AMT5  1    20834 20992
## <none>            20832 20992
## - BILL_AMT2  1    20834 20992
## - BILL_AMT1  1    20835 20993
## - AGE        1    20835 20993
## - PAY_AMT3   1    20835 20993
## - BILL_AMT3  1    20836 20994
## - PAY_AMT6   1    20837 20995
## - PAY_AMT5   1    20838 20996
## - PAY_2     10    20862 21002
## - SEX        1    20845 21003
## - PAY_AMT2   1    20849 21007
## - PAY_5      9    20865 21007
## - PAY_4     10    20868 21008
## - PAY_6      9    20870 21012
## - MARRIAGE   3    20859 21013
## - PAY_3     10    20874 21014
## - PAY_AMT1   1    20859 21017
## - EDUCATION  6    20876 21024
## - LIMIT_BAL  1    20918 21076
## - PAY_0     10    22021 22161
##
## Step:  AIC=20991.88
## default_payment_next_month ~ LIMIT_BAL + SEX + EDUCATION + MARRIAGE +
##     AGE + PAY_0 + PAY_2 + PAY_3 + PAY_4 + PAY_5 + PAY_6 + BILL_AMT1 +
##     BILL_AMT2 + BILL_AMT3 + PAY_AMT1 + PAY_AMT2 + PAY_AMT3 +
##     PAY_AMT5 + PAY_AMT6
```

```
##
##              Df Deviance   AIC
## <none>            20834 20992
## - BILL_AMT2  1    20836 20992
## - BILL_AMT3  1    20836 20992
## - BILL_AMT1  1    20836 20992
## - AGE        1    20837 20993
## - PAY_AMT3   1    20839 20995
## - PAY_AMT6   1    20839 20995
## - PAY_AMT5   1    20840 20996
## - PAY_2     10    20864 21002
## - SEX        1    20847 21003
## - PAY_AMT2   1    20850 21006
## - PAY_5      9    20867 21007
## - PAY_4     10    20869 21007
## - PAY_6      9    20873 21013
## - MARRIAGE   3    20861 21013
## - PAY_3     10    20876 21014
## - PAY_AMT1   1    20860 21016
## - EDUCATION  6    20877 21023
## - LIMIT_BAL  1    20923 21079
## - PAY_0     10    22022 22160
```

```
summary(credit_default_glm_back)
```

```
##
## Call:
## glm(formula = default_payment_next_month ~ LIMIT_BAL + SEX +
##     EDUCATION + MARRIAGE + AGE + PAY_0 + PAY_2 + PAY_3 + PAY_4 +
##     PAY_5 + PAY_6 + BILL_AMT1 + BILL_AMT2 + BILL_AMT3 + PAY_AMT1 +
##     PAY_AMT2 + PAY_AMT3 + PAY_AMT5 + PAY_AMT6, family = binomial,
##     data = credit_card_data_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4765  -0.5989  -0.5099  -0.2951   3.4870
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.607e+01  2.603e+02  -0.062 0.950773
## LIMIT_BAL   -1.787e-06  1.938e-07  -9.223  < 2e-16 ***
## SEX2        -1.323e-01  3.616e-02  -3.659 0.000253 ***
## EDUCATION1   1.283e+01  2.603e+02   0.049 0.960710
## EDUCATION2   1.286e+01  2.603e+02   0.049 0.960601
## EDUCATION3   1.279e+01  2.603e+02   0.049 0.960823
## EDUCATION4   1.156e+01  2.603e+02   0.044 0.964578
## EDUCATION5   1.161e+01  2.603e+02   0.045 0.964433
## EDUCATION6   1.259e+01  2.603e+02   0.048 0.961427
## MARRIAGE1    2.051e+00  6.532e-01   3.140 0.001688 **
## MARRIAGE2    1.912e+00  6.533e-01   2.926 0.003431 **
```

```
## MARRIAGE3    2.202e+00  6.723e-01   3.276 0.001054 **
## AGE          3.911e-03  2.205e-03   1.773 0.076153 .
## PAY_0-1      6.151e-01  1.214e-01   5.067 4.04e-07 ***
## PAY_00      -1.845e-01  1.310e-01  -1.408 0.159156
## PAY_01       8.270e-01  9.420e-02   8.779  < 2e-16 ***
## PAY_02       2.109e+00  1.189e-01  17.739  < 2e-16 ***
## PAY_03       1.959e+00  1.860e-01  10.532  < 2e-16 ***
## PAY_04       1.982e+00  3.487e-01   5.684 1.32e-08 ***
## PAY_05       1.446e+00  5.463e-01   2.647 0.008117 **
## PAY_06      -7.208e-01  1.375e+00  -0.524 0.600050
## PAY_07      -1.262e+01  6.184e+02  -0.020 0.983714
## PAY_08      -1.310e+01  8.827e+02  -0.015 0.988164
## PAY_2-1     -3.656e-01  1.279e-01  -2.859 0.004248 **
## PAY_20      -1.013e-01  1.550e-01  -0.654 0.513261
## PAY_21      -6.862e-01  5.794e-01  -1.184 0.236333
## PAY_22      -4.595e-02  1.306e-01  -0.352 0.725062
## PAY_23       1.681e-02  2.015e-01   0.083 0.933507
## PAY_24      -5.108e-01  3.632e-01  -1.407 0.159563
## PAY_25       2.329e+00  1.179e+00   1.974 0.048354 *
## PAY_26       1.506e+01  6.184e+02   0.024 0.980575
## PAY_27       9.469e-01  9.899e+02   0.001 0.999237
## PAY_28       1.446e+01  1.135e+03   0.013 0.989830
## PAY_3-1      7.694e-02  1.223e-01   0.629 0.529391
## PAY_30       1.585e-01  1.411e-01   1.124 0.261223
## PAY_31      -1.264e+01  8.827e+02  -0.014 0.988572
## PAY_32       4.797e-01  1.427e-01   3.361 0.000776 ***
## PAY_33       4.993e-01  2.539e-01   1.966 0.049248 *
## PAY_34      -4.138e-01  4.627e-01  -0.894 0.371074
## PAY_35      -9.316e-01  8.682e-01  -1.073 0.283242
## PAY_36       1.444e+01  4.481e+02   0.032 0.974293
## PAY_37       1.450e-01  1.005e+00   0.144 0.885268
## PAY_38      -2.578e+01  4.975e+02  -0.052 0.958673
## PAY_4-1     -2.084e-01  1.225e-01  -1.702 0.088815 .
## PAY_40      -2.354e-01  1.368e-01  -1.721 0.085265 .
## PAY_41       2.854e+01  1.248e+03   0.023 0.981761
## PAY_42       4.815e-02  1.461e-01   0.330 0.741707
## PAY_43      -1.454e-01  2.793e-01  -0.520 0.602755
## PAY_44       4.021e-01  5.098e-01   0.789 0.430268
## PAY_45      -1.323e+00  8.438e-01  -1.568 0.116784
## PAY_46      -2.931e+01  7.128e+02  -0.041 0.967200
## PAY_47      -1.937e+00  6.371e+02  -0.003 0.997574
## PAY_48      -3.131e+01  1.052e+03  -0.030 0.976251
## PAY_5-1     -7.008e-02  1.193e-01  -0.587 0.556992
## PAY_50       1.475e-01  1.320e-01   1.117 0.264111
## PAY_52       4.564e-01  1.478e-01   3.088 0.002015 **
## PAY_53       1.544e-01  2.728e-01   0.566 0.571342
## PAY_54       1.256e-01  5.287e-01   0.238 0.812149
## PAY_55       9.432e-01  9.618e-01   0.981 0.326746
## PAY_56       3.930e+01  8.146e+02   0.048 0.961520
## PAY_57       1.559e+01  5.717e+02   0.027 0.978248
```

```
## PAY_58       4.195e+01  2.098e+03   0.020 0.984044    
## PAY_6-1     -1.102e-01  9.225e-02  -1.195 0.232238    
## PAY_60      -3.368e-01  9.757e-02  -3.452 0.000557 ***
## PAY_62      -4.135e-02  1.135e-01  -0.364 0.715491    
## PAY_63       5.523e-01  2.615e-01   2.112 0.034706 *  
## PAY_64      -3.448e-01  5.233e-01  -0.659 0.509995    
## PAY_65       7.694e-01  1.019e+00   0.755 0.450324    
## PAY_66      -3.397e-01  1.134e+00  -0.299 0.764584    
## PAY_67      -1.275e+01  2.812e+02  -0.045 0.963851    
## PAY_68       2.870e+01  1.271e+03   0.023 0.981980    
## BILL_AMT1   -1.914e-06  1.240e-06  -1.543 0.122756    
## BILL_AMT2    2.430e-06  1.680e-06   1.447 0.148034    
## BILL_AMT3    1.759e-06  1.171e-06   1.501 0.133272    
## PAY_AMT1    -1.157e-05  2.587e-06  -4.474 7.67e-06 ***
## PAY_AMT2    -8.201e-06  2.261e-06  -3.627 0.000287 ***
## PAY_AMT3    -4.075e-06  1.942e-06  -2.099 0.035838 *  
## PAY_AMT5    -4.157e-06  1.760e-06  -2.362 0.018174 *  
## PAY_AMT6    -3.205e-06  1.492e-06  -2.148 0.031719 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 25495  on 23999  degrees of freedom
## Residual deviance: 20834  on 23921  degrees of freedom
## AIC: 20992
## 
## Number of Fisher Scoring iterations: 13
```

```
credit_default_glm_back$deviance # 20833.88
```

```
## [1] 20833.88
```

```
AIC(credit_default_glm_back) # 20991.88
```

```
## [1] 20991.88
```

```
BIC(credit_default_glm_back) # 21630.66
```

```
## [1] 21630.66
```

Calculating Model mean residual deviance (in-sample)

```
credit_default_glm_back$dev/credit_default_glm_back$df.residual # 0.8709453
```

```
## [1] 0.8709453
```

#————— Stepwise variable selection using AIC —————#

```
model_stepwise_AIC <- step(null.glm, scope = list(lower= null.glm, upper= ful
l.glm), direction = "both", k = 2)
```

```
## Start:  AIC=25497.03
## default_payment_next_month ~ 1
##
##              Df Deviance   AIC
## + PAY_0      10    21792 21814
## + PAY_2      10    23061 23083
## + PAY_3      10    23666 23688
## + PAY_4      10    23883 23905
## + PAY_5       9    23991 24011
## + PAY_6       9    24234 24254
## + LIMIT_BAL   1    24889 24893
## + PAY_AMT1    1    25201 25205
## + PAY_AMT2    1    25238 25242
## + PAY_AMT3    1    25308 25312
## + EDUCATION   6    25348 25362
## + PAY_AMT4    1    25364 25368
## + PAY_AMT5    1    25366 25370
## + PAY_AMT6    1    25384 25388
## + SEX         1    25463 25467
## + MARRIAGE    3    25462 25470
## + BILL_AMT1   1    25481 25485
## + BILL_AMT2   1    25488 25492
## + BILL_AMT3   1    25489 25493
## + BILL_AMT4   1    25491 25495
## + AGE         1    25491 25495
## + BILL_AMT5   1    25493 25497
## <none>             25495 25497
## + BILL_AMT6   1    25493 25497
##
## Step:  AIC=21813.55
## default_payment_next_month ~ PAY_0
##
##              Df Deviance   AIC
## + PAY_3      10    21461 21503
## + PAY_4      10    21464 21506
## + PAY_5       9    21466 21506
## + LIMIT_BAL   1    21544 21568
## + PAY_6       9    21530 21570
## + PAY_2      10    21558 21600
## + PAY_AMT2    1    21686 21710
## + PAY_AMT1    1    21689 21713
## + PAY_AMT3    1    21715 21739
## + EDUCATION   6    21707 21741
## + PAY_AMT5    1    21735 21759
## + PAY_AMT6    1    21741 21765
## + PAY_AMT4    1    21742 21766
## + MARRIAGE    3    21764 21792
## + SEX         1    21772 21796
## + BILL_AMT4   1    21784 21808
## + BILL_AMT5   1    21784 21808
```

```
## + BILL_AMT3  1    21785 21809
## + BILL_AMT6  1    21785 21809
## + BILL_AMT1  1    21785 21809
## + BILL_AMT2  1    21787 21811
## + AGE        1    21787 21811
## <none>            21792 21814
## - PAY_0     10    25495 25497
##
## Step:  AIC=21502.86
## default_payment_next_month ~ PAY_0 + PAY_3
##
##             Df Deviance   AIC
## + LIMIT_BAL  1    21292 21336
## + PAY_5      9    21307 21367
## + PAY_6      9    21330 21390
## + PAY_AMT1   1    21356 21400
## + PAY_4     10    21349 21411
## + PAY_AMT2   1    21400 21444
## + PAY_AMT3   1    21404 21448
## + PAY_2     10    21390 21452
## + EDUCATION  6    21401 21455
## + PAY_AMT5   1    21415 21459
## + PAY_AMT6   1    21419 21463
## + PAY_AMT4   1    21421 21465
## + MARRIAGE   3    21431 21479
## + BILL_AMT5  1    21444 21488
## + BILL_AMT4  1    21444 21488
## + BILL_AMT6  1    21445 21489
## + SEX        1    21446 21490
## + BILL_AMT3  1    21447 21491
## + BILL_AMT2  1    21449 21493
## + BILL_AMT1  1    21451 21495
## + AGE        1    21454 21498
## <none>            21461 21503
## - PAY_3     10    21792 21814
## - PAY_0     10    23666 23688
##
## Step:  AIC=21336.11
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL
##
##             Df Deviance   AIC
## + PAY_5      9    21156 21218
## + PAY_6      9    21181 21243
## + PAY_4     10    21195 21259
## + PAY_AMT1   1    21239 21285
## + PAY_2     10    21233 21297
## + MARRIAGE   3    21251 21301
## + EDUCATION  6    21246 21302
## + PAY_AMT2   1    21266 21312
## + PAY_AMT3   1    21269 21315
```

```
## + AGE        1    21272 21318
## + PAY_AMT5   1    21277 21323
## + SEX        1    21277 21323
## + PAY_AMT6   1    21279 21325
## + PAY_AMT4   1    21279 21325
## + BILL_AMT1  1    21282 21328
## + BILL_AMT2  1    21284 21330
## + BILL_AMT3  1    21285 21331
## + BILL_AMT4  1    21288 21334
## + BILL_AMT6  1    21288 21334
## + BILL_AMT5  1    21288 21334
## <none>            21292 21336
## - LIMIT_BAL  1    21461 21503
## - PAY_3     10    21544 21568
## - PAY_0     10    23388 23412
##
## Step:  AIC=21218.4
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5
##
##             Df Deviance   AIC
## + PAY_AMT1   1    21108 21172
## + MARRIAGE   3    21116 21184
## + EDUCATION  6    21112 21186
## + PAY_AMT2   1    21127 21191
## + PAY_2     10    21112 21194
## + PAY_AMT3   1    21134 21198
## + AGE        1    21137 21201
## + PAY_6      9    21122 21202
## + PAY_4     10    21121 21203
## + SEX        1    21143 21207
## + PAY_AMT5   1    21144 21208
## + PAY_AMT6   1    21144 21208
## + BILL_AMT1  1    21146 21210
## + BILL_AMT2  1    21147 21211
## + BILL_AMT3  1    21150 21214
## + PAY_AMT4   1    21150 21214
## + BILL_AMT4  1    21153 21217
## + BILL_AMT5  1    21154 21218
## + BILL_AMT6  1    21154 21218
## <none>            21156 21218
## - PAY_3     10    21281 21323
## - PAY_5      9    21292 21336
## - LIMIT_BAL  1    21307 21367
## - PAY_0     10    23004 23046
##
## Step:  AIC=21171.49
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1
##
##             Df Deviance   AIC
```

```
## + MARRIAGE   3    21068 21138
## + EDUCATION  6    21065 21141
## + BILL_AMT2  1    21083 21149
## + AGE        1    21088 21154
## + BILL_AMT3  1    21090 21156
## + PAY_6      9    21074 21156
## + PAY_4     10    21072 21156
## + PAY_AMT2   1    21091 21157
## + BILL_AMT1  1    21092 21158
## + PAY_2     10    21075 21159
## + SEX        1    21093 21159
## + PAY_AMT3   1    21095 21161
## + BILL_AMT4  1    21097 21163
## + BILL_AMT5  1    21099 21165
## + PAY_AMT5   1    21100 21166
## + BILL_AMT6  1    21100 21166
## + PAY_AMT6   1    21101 21167
## + PAY_AMT4   1    21105 21171
## <none>            21108 21172
## - PAY_AMT1   1    21156 21218
## - LIMIT_BAL  1    21212 21274
## - PAY_3     10    21237 21281
## - PAY_5      9    21239 21285
## - PAY_0     10    22913 22957
##
## Step:  AIC=21137.78
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE
##
##             Df Deviance   AIC
## + EDUCATION  6    21026 21108
## + BILL_AMT2  1    21044 21116
## + PAY_6      9    21034 21122
## + BILL_AMT3  1    21051 21123
## + PAY_4     10    21033 21123
## + PAY_AMT2   1    21052 21124
## + SEX        1    21053 21125
## + PAY_2     10    21035 21125
## + BILL_AMT1  1    21053 21125
## + PAY_AMT3   1    21055 21127
## + BILL_AMT4  1    21057 21129
## + BILL_AMT5  1    21060 21132
## + PAY_AMT5   1    21060 21132
## + BILL_AMT6  1    21061 21133
## + PAY_AMT6   1    21061 21133
## + AGE        1    21063 21135
## + PAY_AMT4   1    21065 21137
## <none>            21068 21138
## - MARRIAGE   3    21108 21172
## - PAY_AMT1   1    21116 21184
```

```
## - PAY_3     10    21197 21247
## - LIMIT_BAL  1    21180 21248
## - PAY_5      9    21199 21251
## - PAY_0     10    22863 22913
##
## Step:  AIC=21107.91
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION
##
##             Df Deviance   AIC
## + BILL_AMT2  1    21001 21085
## + BILL_AMT3  1    21008 21092
## + PAY_6      9    20992 21092
## + PAY_4     10    20991 21093
## + BILL_AMT1  1    21010 21094
## + PAY_AMT2   1    21010 21094
## + SEX        1    21011 21095
## + PAY_2     10    20993 21095
## + PAY_AMT3   1    21014 21098
## + BILL_AMT4  1    21015 21099
## + BILL_AMT5  1    21018 21102
## + PAY_AMT5   1    21018 21102
## + BILL_AMT6  1    21019 21103
## + PAY_AMT6   1    21020 21104
## + AGE        1    21020 21104
## + PAY_AMT4   1    21023 21107
## <none>            21026 21108
## - EDUCATION  6    21068 21138
## - MARRIAGE   3    21065 21141
## - PAY_AMT1   1    21073 21153
## - LIMIT_BAL  1    21131 21211
## - PAY_3     10    21149 21211
## - PAY_5      9    21155 21219
## - PAY_0     10    22823 22885
##
## Step:  AIC=21084.48
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2
##
##             Df Deviance   AIC
## + PAY_6      9    20963 21065
## + PAY_AMT2   1    20982 21068
## + PAY_4     10    20966 21070
## + SEX        1    20986 21072
## + PAY_AMT3   1    20986 21072
## + PAY_AMT5   1    20990 21076
## + PAY_2     10    20972 21076
## + PAY_AMT6   1    20991 21077
## + BILL_AMT1  1    20993 21079
## + AGE        1    20995 21081
```

```
## + PAY_AMT4   1    20996 21082
## + BILL_AMT5  1    20996 21082
## + BILL_AMT4  1    20997 21083
## + BILL_AMT6  1    20997 21083
## <none>            21001 21085
## + BILL_AMT3  1    21000 21086
## - BILL_AMT2  1    21026 21108
## - EDUCATION  6    21044 21116
## - MARRIAGE   3    21039 21117
## - PAY_AMT1   1    21063 21145
## - PAY_3     10    21115 21179
## - PAY_5      9    21129 21195
## - LIMIT_BAL  1    21131 21213
## - PAY_0     10    22762 22826
##
## Step:  AIC=21065.09
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6
##
##             Df Deviance   AIC
## + PAY_4     10    20927 21049
## + PAY_AMT2   1    20945 21049
## + SEX        1    20949 21053
## + PAY_AMT3   1    20950 21054
## + PAY_2     10    20934 21056
## + PAY_AMT5   1    20953 21057
## + PAY_AMT6   1    20954 21058
## + BILL_AMT1  1    20956 21060
## + AGE        1    20958 21062
## + PAY_AMT4   1    20959 21063
## + BILL_AMT4  1    20960 21064
## + BILL_AMT5  1    20960 21064
## + BILL_AMT6  1    20960 21064
## <none>            20963 21065
## + BILL_AMT3  1    20962 21066
## - PAY_6      9    21001 21085
## - BILL_AMT2  1    20992 21092
## - EDUCATION  6    21007 21097
## - MARRIAGE   3    21002 21098
## - PAY_5      9    21018 21102
## - PAY_AMT1   1    21025 21125
## - PAY_3     10    21068 21150
## - LIMIT_BAL  1    21091 21191
## - PAY_0     10    22696 22778
##
## Step:  AIC=21048.91
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4
##
##             Df Deviance   AIC
```

```
## + PAY_AMT2   1    20908 21032
## + SEX        1    20913 21037
## + PAY_AMT3   1    20916 21040
## + PAY_2     10    20899 21041
## + PAY_AMT5   1    20917 21041
## + PAY_AMT6   1    20917 21041
## + BILL_AMT1  1    20920 21044
## + AGE        1    20922 21046
## + PAY_AMT4   1    20923 21047
## + BILL_AMT5  1    20923 21047
## + BILL_AMT4  1    20923 21047
## + BILL_AMT6  1    20924 21048
## <none>            20927 21049
## + BILL_AMT3  1    20926 21050
## - PAY_5      9    20960 21064
## - PAY_4     10    20963 21065
## - PAY_6      9    20966 21070
## - BILL_AMT2  1    20956 21076
## - EDUCATION  6    20971 21081
## - MARRIAGE   3    20965 21081
## - PAY_3     10    20997 21099
## - PAY_AMT1   1    20989 21109
## - LIMIT_BAL  1    21053 21173
## - PAY_0     10    22638 22740
##
## Step:  AIC=21031.61
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4 +
##     PAY_AMT2
##
##             Df Deviance   AIC
## + SEX        1    20893 21019
## + PAY_2     10    20878 21022
## + PAY_AMT5   1    20899 21025
## + PAY_AMT3   1    20900 21026
## + PAY_AMT6   1    20900 21026
## + AGE        1    20902 21028
## + BILL_AMT1  1    20903 21029
## + PAY_AMT4   1    20905 21031
## + BILL_AMT3  1    20905 21031
## <none>            20908 21032
## + BILL_AMT6  1    20906 21032
## + BILL_AMT5  1    20906 21032
## + BILL_AMT4  1    20907 21033
## - PAY_5      9    20941 21047
## - PAY_AMT2   1    20927 21049
## - PAY_4     10    20945 21049
## - PAY_6      9    20946 21052
## - BILL_AMT2  1    20939 21061
## - EDUCATION  6    20951 21063
```

```
## - MARRIAGE   3    20946 21064
## - PAY_3     10    20965 21069
## - PAY_AMT1   1    20955 21077
## - LIMIT_BAL  1    21020 21142
## - PAY_0     10    22614 22718
##
## Step:  AIC=21019.14
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4 +
##     PAY_AMT2 + SEX
##
##             Df Deviance   AIC
## + PAY_2     10    20864 21010
## + PAY_AMT5   1    20885 21013
## + PAY_AMT3   1    20885 21013
## + PAY_AMT6   1    20885 21013
## + BILL_AMT1  1    20888 21016
## + AGE        1    20890 21018
## + PAY_AMT4   1    20891 21019
## + BILL_AMT3  1    20891 21019
## <none>            20893 21019
## + BILL_AMT6  1    20892 21020
## + BILL_AMT5  1    20892 21020
## + BILL_AMT4  1    20893 21021
## - SEX        1    20908 21032
## - PAY_5      9    20926 21034
## - PAY_4     10    20931 21037
## - PAY_AMT2   1    20913 21037
## - PAY_6      9    20931 21039
## - BILL_AMT2  1    20924 21048
## - EDUCATION  6    20936 21050
## - MARRIAGE   3    20932 21052
## - PAY_3     10    20949 21055
## - PAY_AMT1   1    20940 21064
## - LIMIT_BAL  1    21005 21129
## - PAY_0     10    22600 22706
##
## Step:  AIC=21009.8
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4 +
##     PAY_AMT2 + SEX + PAY_2
##
##             Df Deviance   AIC
## + PAY_AMT5   1    20855 21003
## + PAY_AMT3   1    20855 21003
## + PAY_AMT6   1    20856 21004
## + BILL_AMT1  1    20859 21007
## + AGE        1    20860 21008
## + BILL_AMT3  1    20861 21009
## + PAY_AMT4   1    20861 21009
```

```
## <none>              20864 21010
## + BILL_AMT6  1    20863 21011
## + BILL_AMT5  1    20863 21011
## + BILL_AMT4  1    20864 21012
## - PAY_2     10    20893 21019
## - SEX        1    20878 21022
## - PAY_5      9    20894 21022
## - PAY_4     10    20901 21027
## - PAY_AMT2   1    20885 21029
## - PAY_6      9    20903 21031
## - PAY_3     10    20907 21033
## - BILL_AMT2  1    20890 21034
## - EDUCATION  6    20907 21041
## - PAY_AMT1   1    20899 21043
## - MARRIAGE   3    20903 21043
## - LIMIT_BAL  1    20972 21116
## - PAY_0     10    22061 22187
##
## Step:  AIC=21002.62
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4 +
##     PAY_AMT2 + SEX + PAY_2 + PAY_AMT5
##
##             Df Deviance   AIC
## + PAY_AMT3   1    20848 20998
## + PAY_AMT6   1    20848 20998
## + BILL_AMT1  1    20851 21001
## + AGE        1    20851 21001
## + BILL_AMT3  1    20852 21002
## <none>              20855 21003
## + PAY_AMT4   1    20853 21003
## + BILL_AMT5  1    20854 21004
## + BILL_AMT4  1    20854 21004
## + BILL_AMT6  1    20855 21005
## - PAY_AMT5   1    20864 21010
## - PAY_2     10    20885 21013
## - SEX        1    20869 21015
## - PAY_5      9    20886 21016
## - PAY_AMT2   1    20873 21019
## - PAY_4     10    20892 21020
## - PAY_6      9    20894 21024
## - PAY_3     10    20898 21026
## - BILL_AMT2  1    20884 21030
## - PAY_AMT1   1    20887 21033
## - EDUCATION  6    20898 21034
## - MARRIAGE   3    20894 21036
## - LIMIT_BAL  1    20955 21101
## - PAY_0     10    22051 22179
##
## Step:  AIC=20997.66
```

```
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4 +
##     PAY_AMT2 + SEX + PAY_2 + PAY_AMT5 + PAY_AMT3
##
##             Df Deviance   AIC
## + PAY_AMT6   1    20842 20994
## + AGE        1    20844 20996
## + BILL_AMT1  1    20845 20997
## + BILL_AMT3  1    20845 20997
## <none>            20848 20998
## + PAY_AMT4   1    20846 20998
## + BILL_AMT5  1    20848 21000
## + BILL_AMT4  1    20848 21000
## + BILL_AMT6  1    20848 21000
## - PAY_AMT3   1    20855 21003
## - PAY_AMT5   1    20855 21003
## - PAY_2     10    20877 21007
## - SEX        1    20863 21011
## - PAY_AMT2   1    20864 21012
## - PAY_4     10    20883 21013
## - PAY_5      9    20881 21013
## - PAY_6      9    20886 21018
## - PAY_3     10    20891 21021
## - PAY_AMT1   1    20877 21025
## - BILL_AMT2  1    20879 21027
## - EDUCATION  6    20891 21029
## - MARRIAGE   3    20886 21030
## - LIMIT_BAL  1    20942 21090
## - PAY_0     10    22040 22170
##
## Step:  AIC=20993.96
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4 +
##     PAY_AMT2 + SEX + PAY_2 + PAY_AMT5 + PAY_AMT3 + PAY_AMT6
##
##             Df Deviance   AIC
## + AGE        1    20839 20993
## + BILL_AMT1  1    20839 20993
## + BILL_AMT3  1    20840 20994
## <none>            20842 20994
## + PAY_AMT4   1    20841 20995
## + BILL_AMT5  1    20842 20996
## + BILL_AMT4  1    20842 20996
## + BILL_AMT6  1    20842 20996
## - PAY_AMT6   1    20848 20998
## - PAY_AMT3   1    20848 20998
## - PAY_AMT5   1    20849 20999
## - PAY_2     10    20872 21004
## - PAY_AMT2   1    20857 21007
## - SEX        1    20857 21007
```

```
## - PAY_5       9    20875 21009
## - PAY_4      10    20877 21009
## - PAY_6       9    20881 21015
## - PAY_3      10    20885 21017
## - PAY_AMT1    1    20870 21020
## - EDUCATION   6    20885 21025
## - BILL_AMT2   1    20875 21025
## - MARRIAGE    3    20880 21026
## - LIMIT_BAL   1    20932 21082
## - PAY_0      10    22034 22166
##
## Step:  AIC=20992.73
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4 +
##     PAY_AMT2 + SEX + PAY_2 + PAY_AMT5 + PAY_AMT3 + PAY_AMT6 +
##     AGE
##
##             Df Deviance   AIC
## + BILL_AMT1  1    20836 20992
## + BILL_AMT3  1    20836 20992
## <none>            20839 20993
## + PAY_AMT4   1    20838 20994
## - AGE        1    20842 20994
## + BILL_AMT5  1    20839 20995
## + BILL_AMT4  1    20839 20995
## + BILL_AMT6  1    20839 20995
## - PAY_AMT6   1    20844 20996
## - PAY_AMT3   1    20845 20997
## - PAY_AMT5   1    20846 20998
## - PAY_2     10    20868 21002
## - SEX        1    20852 21004
## - PAY_AMT2   1    20854 21006
## - PAY_4     10    20874 21008
## - PAY_5      9    20872 21008
## - PAY_6      9    20878 21014
## - MARRIAGE   3    20866 21014
## - PAY_3     10    20882 21016
## - PAY_AMT1   1    20866 21018
## - BILL_AMT2  1    20871 21023
## - EDUCATION  6    20882 21024
## - LIMIT_BAL  1    20931 21083
## - PAY_0     10    22032 22166
##
## Step:  AIC=20992.18
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4 +
##     PAY_AMT2 + SEX + PAY_2 + PAY_AMT5 + PAY_AMT3 + PAY_AMT6 +
##     AGE + BILL_AMT1
##
##             Df Deviance   AIC
```

```
## + BILL_AMT3  1    20834 20992
## <none>            20836 20992
## - BILL_AMT1  1    20839 20993
## + PAY_AMT4   1    20835 20993
## - AGE        1    20839 20993
## + BILL_AMT5  1    20836 20994
## + BILL_AMT4  1    20836 20994
## + BILL_AMT6  1    20836 20994
## - PAY_AMT6   1    20841 20995
## - PAY_AMT3   1    20842 20996
## - PAY_AMT5   1    20843 20997
## - BILL_AMT2  1    20847 21001
## - PAY_2     10    20865 21001
## - SEX        1    20849 21003
## - PAY_AMT2   1    20850 21004
## - PAY_4     10    20871 21007
## - PAY_5      9    20869 21007
## - PAY_6      9    20875 21013
## - MARRIAGE   3    20863 21013
## - PAY_3     10    20878 21014
## - PAY_AMT1   1    20865 21019
## - EDUCATION  6    20879 21023
## - LIMIT_BAL  1    20924 21078
## - PAY_0     10    22024 22160
##
## Step:  AIC=20991.88
## default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
##     PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4 +
##     PAY_AMT2 + SEX + PAY_2 + PAY_AMT5 + PAY_AMT3 + PAY_AMT6 +
##     AGE + BILL_AMT1 + BILL_AMT3
##
##             Df Deviance   AIC
## <none>            20834 20992
## + BILL_AMT5  1    20832 20992
## - BILL_AMT2  1    20836 20992
## - BILL_AMT3  1    20836 20992
## - BILL_AMT1  1    20836 20992
## + PAY_AMT4   1    20833 20993
## + BILL_AMT6  1    20833 20993
## - AGE        1    20837 20993
## + BILL_AMT4  1    20833 20993
## - PAY_AMT3   1    20839 20995
## - PAY_AMT6   1    20839 20995
## - PAY_AMT5   1    20840 20996
## - PAY_2     10    20864 21002
## - SEX        1    20847 21003
## - PAY_AMT2   1    20850 21006
## - PAY_5      9    20867 21007
## - PAY_4     10    20869 21007
## - PAY_6      9    20873 21013
```

```
## - MARRIAGE   3    20861 21013
## - PAY_3     10    20876 21014
## - PAY_AMT1   1    20860 21016
## - EDUCATION  6    20877 21023
## - LIMIT_BAL  1    20923 21079
## - PAY_0     10    22022 22160

# Lowest AIC = 20991.88
#default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL + PAY_5 +
#  PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 + PAY_4 +
#  PAY_AMT2 + SEX + PAY_2 + PAY_AMT5 + PAY_AMT3 + PAY_AMT6 +
#  AGE + BILL_AMT1 + BILL_AMT3
```

Running model with selected variables

```
model_stepwise_AIC.glm <- glm(default_payment_next_month ~ PAY_0 + PAY_3 + LI
MIT_BAL + PAY_5 +
                                PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 +
PAY_6 + PAY_4 +
                                PAY_AMT2 + SEX + PAY_2 + PAY_AMT5 + PAY_AMT3
+ PAY_AMT6 +
                                AGE + BILL_AMT1 + BILL_AMT3, family=binomial,
data=credit_card_data_train)

model_stepwise_AIC.glm_summary<-summary(model_stepwise_AIC.glm)

model_stepwise_AIC.glm_summary

##
## Call:
## glm(formula = default_payment_next_month ~ PAY_0 + PAY_3 + LIMIT_BAL +
##     PAY_5 + PAY_AMT1 + MARRIAGE + EDUCATION + BILL_AMT2 + PAY_6 +
##     PAY_4 + PAY_AMT2 + SEX + PAY_2 + PAY_AMT5 + PAY_AMT3 + PAY_AMT6 +
##     AGE + BILL_AMT1 + BILL_AMT3, family = binomial, data = credit_card_dat
a_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4765  -0.5989  -0.5099  -0.2951   3.4870
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.607e+01  2.603e+02  -0.062 0.950773
## PAY_0-1      6.151e-01  1.214e-01   5.067 4.04e-07 ***
## PAY_00      -1.845e-01  1.310e-01  -1.408 0.159156
## PAY_01       8.270e-01  9.420e-02   8.779  < 2e-16 ***
## PAY_02       2.109e+00  1.189e-01  17.739  < 2e-16 ***
## PAY_03       1.959e+00  1.860e-01  10.532  < 2e-16 ***
## PAY_04       1.982e+00  3.487e-01   5.684 1.32e-08 ***
## PAY_05       1.446e+00  5.463e-01   2.647 0.008117 **
## PAY_06      -7.208e-01  1.375e+00  -0.524 0.600050
```

```
## PAY_07      -1.262e+01  6.184e+02  -0.020 0.983714
## PAY_08      -1.310e+01  8.827e+02  -0.015 0.988164
## PAY_3-1      7.694e-02  1.223e-01   0.629 0.529391
## PAY_30       1.585e-01  1.411e-01   1.124 0.261223
## PAY_31      -1.264e+01  8.827e+02  -0.014 0.988572
## PAY_32       4.797e-01  1.427e-01   3.361 0.000776 ***
## PAY_33       4.993e-01  2.539e-01   1.966 0.049248 *
## PAY_34      -4.138e-01  4.627e-01  -0.894 0.371074
## PAY_35      -9.316e-01  8.682e-01  -1.073 0.283242
## PAY_36       1.444e+01  4.481e+02   0.032 0.974293
## PAY_37       1.450e-01  1.005e+00   0.144 0.885268
## PAY_38      -2.578e+01  4.975e+02  -0.052 0.958673
## LIMIT_BAL   -1.787e-06  1.938e-07  -9.223  < 2e-16 ***
## PAY_5-1     -7.008e-02  1.193e-01  -0.587 0.556992
## PAY_50       1.475e-01  1.320e-01   1.117 0.264111
## PAY_52       4.564e-01  1.478e-01   3.088 0.002015 **
## PAY_53       1.544e-01  2.728e-01   0.566 0.571342
## PAY_54       1.256e-01  5.287e-01   0.238 0.812149
## PAY_55       9.432e-01  9.618e-01   0.981 0.326746
## PAY_56       3.930e+01  8.146e+02   0.048 0.961520
## PAY_57       1.559e+01  5.717e+02   0.027 0.978248
## PAY_58       4.195e+01  2.098e+03   0.020 0.984044
## PAY_AMT1    -1.157e-05  2.587e-06  -4.474 7.67e-06 ***
## MARRIAGE1    2.051e+00  6.532e-01   3.140 0.001688 **
## MARRIAGE2    1.912e+00  6.533e-01   2.926 0.003431 **
## MARRIAGE3    2.202e+00  6.723e-01   3.276 0.001054 **
## EDUCATION1   1.283e+01  2.603e+02   0.049 0.960710
## EDUCATION2   1.286e+01  2.603e+02   0.049 0.960601
## EDUCATION3   1.279e+01  2.603e+02   0.049 0.960823
## EDUCATION4   1.156e+01  2.603e+02   0.044 0.964578
## EDUCATION5   1.161e+01  2.603e+02   0.045 0.964433
## EDUCATION6   1.259e+01  2.603e+02   0.048 0.961427
## BILL_AMT2    2.430e-06  1.680e-06   1.447 0.148034
## PAY_6-1     -1.102e-01  9.225e-02  -1.195 0.232238
## PAY_60      -3.368e-01  9.757e-02  -3.452 0.000557 ***
## PAY_62      -4.135e-02  1.135e-01  -0.364 0.715491
## PAY_63       5.523e-01  2.615e-01   2.112 0.034706 *
## PAY_64      -3.448e-01  5.233e-01  -0.659 0.509995
## PAY_65       7.694e-01  1.019e+00   0.755 0.450324
## PAY_66      -3.397e-01  1.134e+00  -0.299 0.764584
## PAY_67      -1.275e+01  2.812e+02  -0.045 0.963851
## PAY_68       2.870e+01  1.271e+03   0.023 0.981980
## PAY_4-1     -2.084e-01  1.225e-01  -1.702 0.088815 .
## PAY_40      -2.354e-01  1.368e-01  -1.721 0.085265 .
## PAY_41       2.854e+01  1.248e+03   0.023 0.981761
## PAY_42       4.815e-02  1.461e-01   0.330 0.741707
## PAY_43      -1.454e-01  2.793e-01  -0.520 0.602755
## PAY_44       4.021e-01  5.098e-01   0.789 0.430268
## PAY_45      -1.323e+00  8.438e-01  -1.568 0.116784
## PAY_46      -2.931e+01  7.128e+02  -0.041 0.967200
```

```
## PAY_47      -1.937e+00  6.371e+02  -0.003 0.997574
## PAY_48      -3.131e+01  1.052e+03  -0.030 0.976251
## PAY_AMT2    -8.201e-06  2.261e-06  -3.627 0.000287 ***
## SEX2        -1.323e-01  3.616e-02  -3.659 0.000253 ***
## PAY_2-1     -3.656e-01  1.279e-01  -2.859 0.004248 **
## PAY_20      -1.013e-01  1.550e-01  -0.654 0.513261
## PAY_21      -6.862e-01  5.794e-01  -1.184 0.236333
## PAY_22      -4.595e-02  1.306e-01  -0.352 0.725062
## PAY_23       1.681e-02  2.015e-01   0.083 0.933507
## PAY_24      -5.108e-01  3.632e-01  -1.407 0.159563
## PAY_25       2.329e+00  1.179e+00   1.974 0.048354 *
## PAY_26       1.506e+01  6.184e+02   0.024 0.980575
## PAY_27       9.469e-01  9.899e+02   0.001 0.999237
## PAY_28       1.446e+01  1.135e+03   0.013 0.989830
## PAY_AMT5    -4.157e-06  1.760e-06  -2.362 0.018174 *
## PAY_AMT3    -4.075e-06  1.942e-06  -2.099 0.035838 *
## PAY_AMT6    -3.205e-06  1.492e-06  -2.148 0.031719 *
## AGE          3.911e-03  2.205e-03   1.773 0.076153 .
## BILL_AMT1   -1.914e-06  1.240e-06  -1.543 0.122756
## BILL_AMT3    1.759e-06  1.171e-06   1.501 0.133272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 25495  on 23999  degrees of freedom
## Residual deviance: 20834  on 23921  degrees of freedom
## AIC: 20992
##
## Number of Fisher Scoring iterations: 13

AIC(model_stepwise_AIC.glm) # 20991.88

## [1] 20991.88

BIC(model_stepwise_AIC.glm) # 21630.66

## [1] 21630.66

model_stepwise_AIC.glm$deviance # 20833.88

## [1] 20833.88
```

Calculating Model mean residual deviance (in-sample)

```
model_stepwise_AIC.glm$dev/model_stepwise_AIC.glm$df.residual # 0.8709453

## [1] 0.8709453
```

#———— Stepwise variable selection using BIC ————#

```r
model_stepwise_BIC <- step(null.glm, scope = list(lower= null.glm, upper= ful
l.glm), direction = "both", k = log(nrow(credit_card_data_train)))
```

```
## Start:  AIC=25505.12
## default_payment_next_month ~ 1
##
##              Df Deviance   AIC
## + PAY_0      10    21792 21903
## + PAY_2      10    23061 23172
## + PAY_3      10    23666 23776
## + PAY_4      10    23883 23994
## + PAY_5       9    23991 24092
## + PAY_6       9    24234 24335
## + LIMIT_BAL   1    24889 24909
## + PAY_AMT1    1    25201 25221
## + PAY_AMT2    1    25238 25259
## + PAY_AMT3    1    25308 25328
## + PAY_AMT4    1    25364 25384
## + PAY_AMT5    1    25366 25387
## + PAY_AMT6    1    25384 25405
## + EDUCATION   6    25348 25418
## + SEX         1    25463 25484
## + BILL_AMT1   1    25481 25502
## + MARRIAGE    3    25462 25502
## <none>             25495 25505
## + BILL_AMT2   1    25488 25508
## + BILL_AMT3   1    25489 25509
## + BILL_AMT4   1    25491 25511
## + AGE         1    25491 25511
## + BILL_AMT5   1    25493 25513
## + BILL_AMT6   1    25493 25514
##
## Step:  AIC=21902.5
## default_payment_next_month ~ PAY_0
##
##              Df Deviance   AIC
## + LIMIT_BAL   1    21544 21665
## + PAY_5       9    21466 21668
## + PAY_3      10    21461 21673
## + PAY_4      10    21464 21675
## + PAY_6       9    21530 21732
## + PAY_2      10    21558 21770
## + PAY_AMT2    1    21686 21807
## + PAY_AMT1    1    21689 21810
## + PAY_AMT3    1    21715 21836
## + PAY_AMT5    1    21735 21856
## + PAY_AMT6    1    21741 21862
## + PAY_AMT4    1    21742 21863
## + EDUCATION   6    21707 21879
## + SEX         1    21772 21893
```

```
## <none>             21792 21903
## + BILL_AMT4  1    21784 21905
## + BILL_AMT5  1    21784 21905
## + MARRIAGE   3    21764 21906
## + BILL_AMT3  1    21785 21906
## + BILL_AMT6  1    21785 21906
## + BILL_AMT1  1    21785 21907
## + BILL_AMT2  1    21787 21908
## + AGE        1    21787 21908
## - PAY_0     10    25495 25505
##
## Step:  AIC=21665.43
## default_payment_next_month ~ PAY_0 + LIMIT_BAL
##
##             Df Deviance   AIC
## + PAY_5      9    21281 21493
## + PAY_4     10    21285 21506
## + PAY_3     10    21292 21514
## + PAY_6      9    21339 21551
## + PAY_2     10    21376 21598
## + PAY_AMT2   1    21497 21628
## + PAY_AMT1   1    21498 21630
## + PAY_AMT3   1    21516 21647
## + BILL_AMT2  1    21523 21655
## + MARRIAGE   3    21503 21655
## + BILL_AMT1  1    21524 21655
## + AGE        1    21525 21656
## + SEX        1    21526 21657
## + BILL_AMT3  1    21526 21658
## + BILL_AMT6  1    21528 21659
## + PAY_AMT5   1    21528 21659
## + BILL_AMT4  1    21529 21660
## + BILL_AMT5  1    21529 21660
## + PAY_AMT4   1    21530 21661
## + PAY_AMT6   1    21531 21662
## <none>             21544 21665
## + EDUCATION  6    21487 21668
## - LIMIT_BAL  1    21792 21903
## - PAY_0     10    24889 24909
##
## Step:  AIC=21493.01
## default_payment_next_month ~ PAY_0 + LIMIT_BAL + PAY_5
##
##             Df Deviance   AIC
## + PAY_AMT2   1    21234 21456
## + PAY_AMT1   1    21237 21459
## + PAY_3     10    21156 21469
## + PAY_AMT3   1    21253 21475
## + MARRIAGE   3    21240 21482
## + AGE        1    21262 21484
```

```
## + SEX        1    21266 21488
## + BILL_AMT1  1    21266 21488
## + BILL_AMT2  1    21266 21488
## + PAY_AMT5   1    21268 21490
## + PAY_AMT6   1    21269 21490
## + BILL_AMT3  1    21271 21493
## <none>            21281 21493
## + BILL_AMT4  1    21275 21497
## + BILL_AMT6  1    21276 21498
## + BILL_AMT5  1    21276 21498
## + PAY_AMT4   1    21276 21498
## + EDUCATION  6    21231 21504
## + PAY_2     10    21194 21507
## + PAY_4     10    21203 21515
## + PAY_6      9    21236 21539
## - PAY_5      9    21544 21665
## - LIMIT_BAL  1    21466 21668
## - PAY_0     10    23668 23779
##
## Step:  AIC=21456.19
## default_payment_next_month ~ PAY_0 + LIMIT_BAL + PAY_5 + PAY_AMT2
##
##             Df Deviance   AIC
## + PAY_AMT1   1    21205 21437
## + BILL_AMT3  1    21206 21438
## + BILL_AMT1  1    21214 21446
## + BILL_AMT2  1    21214 21446
## + MARRIAGE   3    21194 21446
## + AGE        1    21215 21447
## + PAY_AMT3   1    21215 21447
## + PAY_3     10    21127 21450
## + SEX        1    21218 21450
## + BILL_AMT4  1    21218 21450
## + BILL_AMT5  1    21222 21454
## + BILL_AMT6  1    21223 21455
## <none>            21234 21456
## + PAY_AMT5   1    21226 21458
## + PAY_AMT6   1    21227 21459
## + PAY_AMT4   1    21233 21465
## + EDUCATION  6    21186 21468
## + PAY_4     10    21155 21478
## + PAY_2     10    21155 21478
## - PAY_AMT2   1    21281 21493
## + PAY_6      9    21191 21504
## - LIMIT_BAL  1    21368 21579
## - PAY_5      9    21497 21628
## - PAY_0     10    23592 23713
##
## Step:  AIC=21437.08
## default_payment_next_month ~ PAY_0 + LIMIT_BAL + PAY_5 + PAY_AMT2 +
```

```
##     PAY_AMT1
##
##             Df Deviance   AIC
## + BILL_AMT3  1    21164 21406
## + BILL_AMT2  1    21168 21410
## + BILL_AMT1  1    21180 21422
## + BILL_AMT4  1    21180 21422
## + PAY_3     10    21091 21424
## + BILL_AMT5  1    21186 21428
## + MARRIAGE   3    21166 21428
## + AGE        1    21186 21428
## + BILL_AMT6  1    21188 21430
## + SEX        1    21189 21431
## + PAY_AMT3   1    21191 21433
## <none>            21205 21437
## + PAY_AMT5   1    21199 21441
## + PAY_AMT6   1    21201 21443
## + PAY_AMT4   1    21204 21447
## + EDUCATION  6    21157 21450
## - PAY_AMT1   1    21234 21456
## + PAY_4     10    21125 21458
## - PAY_AMT2   1    21237 21459
## + PAY_2     10    21137 21469
## + PAY_6      9    21162 21485
## - LIMIT_BAL  1    21315 21537
## - PAY_5      9    21467 21609
## - PAY_0     10    23529 23660
##
## Step:  AIC=21406.02
## default_payment_next_month ~ PAY_0 + LIMIT_BAL + PAY_5 + PAY_AMT2 +
##     PAY_AMT1 + BILL_AMT3
##
##             Df Deviance   AIC
## + MARRIAGE   3    21126 21398
## + AGE        1    21147 21399
## + SEX        1    21148 21400
## + PAY_AMT3   1    21149 21401
## + PAY_3     10    21061 21404
## <none>            21164 21406
## + PAY_AMT5   1    21155 21407
## + PAY_AMT6   1    21156 21408
## + BILL_AMT4  1    21160 21412
## + BILL_AMT5  1    21160 21412
## + BILL_AMT6  1    21161 21413
## + BILL_AMT1  1    21161 21413
## + PAY_AMT4   1    21162 21414
## + BILL_AMT2  1    21164 21416
## + EDUCATION  6    21115 21418
## + PAY_4     10    21089 21432
## - BILL_AMT3  1    21205 21437
```

```
## - PAY_AMT1   1    21206 21438
## - PAY_AMT2   1    21213 21445
## + PAY_6      9    21119 21451
## + PAY_2     10    21109 21452
## - LIMIT_BAL  1    21311 21543
## - PAY_5      9    21406 21558
## - PAY_0     10    23466 23607
##
## Step:  AIC=21397.9
## default_payment_next_month ~ PAY_0 + LIMIT_BAL + PAY_5 + PAY_AMT2 +
##     PAY_AMT1 + BILL_AMT3 + MARRIAGE
##
##             Df Deviance   AIC
## + SEX        1    21109 21391
## + PAY_AMT3   1    21111 21393
## + PAY_3     10    21023 21396
## <none>            21126 21398
## + PAY_AMT5   1    21117 21399
## + PAY_AMT6   1    21118 21400
## + AGE        1    21121 21404
## + BILL_AMT4  1    21122 21404
## + BILL_AMT5  1    21122 21404
## + BILL_AMT1  1    21122 21405
## + BILL_AMT6  1    21122 21405
## - MARRIAGE   3    21164 21406
## + PAY_AMT4   1    21124 21406
## + BILL_AMT2  1    21126 21408
## + EDUCATION  6    21078 21411
## + PAY_4     10    21051 21424
## - BILL_AMT3  1    21166 21428
## - PAY_AMT1   1    21167 21429
## - PAY_AMT2   1    21174 21436
## + PAY_6      9    21080 21443
## + PAY_2     10    21070 21444
## - LIMIT_BAL  1    21281 21543
## - PAY_5      9    21368 21550
## - PAY_0     10    23415 23587
##
## Step:  AIC=21390.89
## default_payment_next_month ~ PAY_0 + LIMIT_BAL + PAY_5 + PAY_AMT2 +
##     PAY_AMT1 + BILL_AMT3 + MARRIAGE + SEX
##
##             Df Deviance   AIC
## + PAY_AMT3   1    21093 21386
## + PAY_3     10    21007 21391
## <none>            21109 21391
## + PAY_AMT5   1    21100 21392
## + PAY_AMT6   1    21101 21393
## + BILL_AMT4  1    21105 21397
## + BILL_AMT1  1    21105 21397
```

```
## + BILL_AMT5  1    21105 21397
## - SEX        1    21126 21398
## + BILL_AMT6  1    21106 21398
## + AGE        1    21106 21398
## + PAY_AMT4   1    21107 21399
## - MARRIAGE   3    21148 21400
## + BILL_AMT2  1    21109 21401
## + EDUCATION  6    21061 21404
## + PAY_4     10    21035 21418
## - BILL_AMT3  1    21148 21420
## - PAY_AMT1   1    21150 21422
## - PAY_AMT2   1    21157 21429
## + PAY_6      9    21063 21436
## + PAY_2     10    21054 21438
## - LIMIT_BAL  1    21263 21535
## - PAY_5      9    21348 21540
## - PAY_0     10    23398 23579
##
## Step:  AIC=21385.83
## default_payment_next_month ~ PAY_0 + LIMIT_BAL + PAY_5 + PAY_AMT2 +
##     PAY_AMT1 + BILL_AMT3 + MARRIAGE + SEX + PAY_AMT3
##
##             Df Deviance   AIC
## <none>            21093 21386
## + PAY_3     10    20996 21389
## + PAY_AMT5   1    21086 21389
## + PAY_AMT6   1    21087 21390
## - PAY_AMT3   1    21109 21391
## - SEX        1    21111 21393
## + AGE        1    21091 21393
## + BILL_AMT1  1    21091 21394
## - MARRIAGE   3    21133 21395
## + PAY_AMT4   1    21092 21395
## + BILL_AMT6  1    21092 21395
## + BILL_AMT5  1    21093 21395
## + BILL_AMT4  1    21093 21396
## + BILL_AMT2  1    21093 21396
## + EDUCATION  6    21047 21400
## - PAY_AMT1   1    21129 21411
## - BILL_AMT3  1    21134 21416
## - PAY_AMT2   1    21137 21419
## + PAY_4     10    21026 21420
## + PAY_6      9    21049 21433
## + PAY_2     10    21041 21434
## - LIMIT_BAL  1    21230 21512
## - PAY_5      9    21332 21534
## - PAY_0     10    23375 23566
```

```
#Lowest AIC = 21385.83
#default_payment_next_month ~ PAY_0 + LIMIT_BAL + PAY_5 + PAY_AMT2 +
#  PAY_AMT1 + BILL_AMT3 + MARRIAGE + SEX + PAY_AMT3

model_stepwise_BIC.glm <- glm(default_payment_next_month ~ PAY_0 + LIMIT_BAL
+ PAY_5 + PAY_AMT2 +
                               PAY_AMT1 + BILL_AMT3 + MARRIAGE + SEX + PAY_AM
T3, family=binomial, data=credit_card_data_train)

model_stepwise_BIC.glm_summary<-summary(model_stepwise_BIC.glm)

model_stepwise_BIC.glm_summary

##
## Call:
## glm(formula = default_payment_next_month ~ PAY_0 + LIMIT_BAL +
##     PAY_5 + PAY_AMT2 + PAY_AMT1 + BILL_AMT3 + MARRIAGE + SEX +
##     PAY_AMT3, family = binomial, data = credit_card_data_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1081  -0.5983  -0.5170  -0.3227   3.4414
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.953e+00  6.464e-01  -4.569 4.90e-06 ***
## PAY_0-1      3.639e-01  8.726e-02   4.170 3.05e-05 ***
## PAY_00      -3.571e-01  8.707e-02  -4.101 4.11e-05 ***
## PAY_01       8.241e-01  8.392e-02   9.820  < 2e-16 ***
## PAY_02       2.093e+00  9.599e-02  21.801  < 2e-16 ***
## PAY_03       2.124e+00  1.697e-01  12.513  < 2e-16 ***
## PAY_04       2.011e+00  3.088e-01   6.511 7.48e-11 ***
## PAY_05       1.182e+00  4.520e-01   2.615 0.008935 **
## PAY_06       6.577e-01  8.255e-01   0.797 0.425608
## PAY_07       1.436e+00  8.676e-01   1.655 0.097945 .
## PAY_08       1.356e+00  6.112e-01   2.219 0.026506 *
## LIMIT_BAL   -2.049e-06  1.807e-07 -11.341  < 2e-16 ***
## PAY_5-1     -3.339e-01  7.025e-02  -4.753 2.00e-06 ***
## PAY_50      -1.321e-01  6.525e-02  -2.025 0.042842 *
## PAY_52       6.507e-01  7.759e-02   8.386  < 2e-16 ***
## PAY_53       5.572e-01  2.135e-01   2.610 0.009059 **
## PAY_54       3.005e-01  3.185e-01   0.944 0.345400
## PAY_55      -3.291e-02  5.667e-01  -0.058 0.953693
## PAY_56       1.069e+01  1.137e+02   0.094 0.925063
## PAY_57       1.238e+00  4.308e-01   2.873 0.004061 **
## PAY_58       1.054e+01  1.970e+02   0.054 0.957332
## PAY_AMT2    -1.183e-05  2.162e-06  -5.470 4.51e-08 ***
## PAY_AMT1    -1.133e-05  2.291e-06  -4.942 7.72e-07 ***
## BILL_AMT3    2.296e-06  3.584e-07   6.407 1.49e-10 ***
## MARRIAGE1    1.945e+00  6.424e-01   3.027 0.002467 **
```

```
## MARRIAGE2    1.768e+00  6.423e-01   2.752 0.005925 **
## MARRIAGE3    2.056e+00  6.617e-01   3.107 0.001892 **
## SEX2        -1.484e-01  3.552e-02  -4.180 2.92e-05 ***
## PAY_AMT3    -7.068e-06  2.015e-06  -3.509 0.000451 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 25495  on 23999  degrees of freedom
## Residual deviance: 21093  on 23971  degrees of freedom
## AIC: 21151
##
## Number of Fisher Scoring iterations: 10

AIC(model_stepwise_BIC.glm) # 21151.34

## [1] 21151.34

BIC(model_stepwise_BIC.glm) # 21385.83

## [1] 21385.83

model_stepwise_BIC.glm$deviance # 21093.34

## [1] 21093.34
```

Calculating Model mean residual deviance (in-sample)

```
model_stepwise_BIC.glm$dev/model_stepwise_BIC.glm$df.residual # 0.8799525

## [1] 0.8799525
```

##———————- Part 3 ———————-##

Performing LASSO variable selection

Creating dummy variable for object datatypes

```
dummy <- model.matrix(~ ., data = credit_card_data)
```

Checking head of dummy matrix

```
head(dummy)

##   (Intercept) LIMIT_BAL SEX2 EDUCATION1 EDUCATION2 EDUCATION3 EDUCATION4
## 1           1     20000    1          0          1          0          0
## 2           1    120000    1          0          1          0          0
## 3           1     90000    1          0          1          0          0
## 4           1     50000    1          0          1          0          0
## 5           1     50000    0          0          1          0          0
## 6           1     50000    0          1          0          0          0
```

```
##   EDUCATION5 EDUCATION6 MARRIAGE1 MARRIAGE2 MARRIAGE3 AGE PAY_0-1 PAY_00 P
AY_01
## 1          0          0         1         0         0  24       0      0
0
## 2          0          0         0         1         0  26       1      0
0
## 3          0          0         0         1         0  34       0      1
0
## 4          0          0         1         0         0  37       0      1
0
## 5          0          0         1         0         0  57       1      0
0
## 6          0          0         0         1         0  37       0      1
0
##   PAY_02 PAY_03 PAY_04 PAY_05 PAY_06 PAY_07 PAY_08 PAY_2-1 PAY_20 PAY_21 P
AY_22
## 1      1      0      0      0      0      0      0       0      0      0
1
## 2      0      0      0      0      0      0      0       0      0      0
1
## 3      0      0      0      0      0      0      0       0      1      0
0
## 4      0      0      0      0      0      0      0       0      1      0
0
## 5      0      0      0      0      0      0      0       0      1      0
0
## 6      0      0      0      0      0      0      0       0      1      0
0
##   PAY_23 PAY_24 PAY_25 PAY_26 PAY_27 PAY_28 PAY_3-1 PAY_30 PAY_31 PAY_32 P
AY_33
## 1      0      0      0      0      0      0       1      0      0      0
0
## 2      0      0      0      0      0      0       0      1      0      0
0
## 3      0      0      0      0      0      0       0      1      0      0
0
## 4      0      0      0      0      0      0       0      1      0      0
0
## 5      0      0      0      0      0      0       1      0      0      0
0
## 6      0      0      0      0      0      0       0      1      0      0
0
##   PAY_34 PAY_35 PAY_36 PAY_37 PAY_38 PAY_4-1 PAY_40 PAY_41 PAY_42 PAY_43 P
AY_44
## 1      0      0      0      0      0       1      0      0      0      0
0
## 2      0      0      0      0      0       0      1      0      0      0
0
## 3      0      0      0      0      0       0      1      0      0      0
0
```

```
## 4      0      0      0      0      0       0      1      0      0      0
0
## 5      0      0      0      0      0       0      1      0      0      0
0
## 6      0      0      0      0      0       0      1      0      0      0
0
##   PAY_45 PAY_46 PAY_47 PAY_48 PAY_5-1 PAY_50 PAY_52 PAY_53 PAY_54 PAY_55 P
AY_56
## 1      0      0      0      0       0      0      0      0      0      0
0
## 2      0      0      0      0       0      1      0      0      0      0
0
## 3      0      0      0      0       0      1      0      0      0      0
0
## 4      0      0      0      0       0      1      0      0      0      0
0
## 5      0      0      0      0       0      1      0      0      0      0
0
## 6      0      0      0      0       0      1      0      0      0      0
0
##   PAY_57 PAY_58 PAY_6-1 PAY_60 PAY_62 PAY_63 PAY_64 PAY_65 PAY_66 PAY_67 P
AY_68
## 1      0      0       0      0      0      0      0      0      0      0
0
## 2      0      0       0      0      1      0      0      0      0      0
0
## 3      0      0       0      1      0      0      0      0      0      0
0
## 4      0      0       0      1      0      0      0      0      0      0
0
## 5      0      0       0      1      0      0      0      0      0      0
0
## 6      0      0       0      1      0      0      0      0      0      0
0
##   BILL_AMT1 BILL_AMT2 BILL_AMT3 BILL_AMT4 BILL_AMT5 BILL_AMT6 PAY_AMT1 PAY
_AMT2
## 1      3913      3102       689         0         0         0        0
689
## 2      2682      1725      2682      3272      3455      3261        0
1000
## 3     29239     14027     13559     14331     14948     15549     1518
1500
## 4     46990     48233     49291     28314     28959     29547     2000
2019
## 5      8617      5670     35835     20940     19146     19131     2000
36681
## 6     64400     57069     57608     19394     19619     20024     2500
1815
##   PAY_AMT3 PAY_AMT4 PAY_AMT5 PAY_AMT6 default_payment_next_month1
## 1        0        0        0        0                           1
```

```
## 2      1000      1000         0      2000                           1
## 3      1000      1000      1000      5000                           0
## 4      1200      1100      1069      1000                           0
## 5     10000      9000       689       679                           0
## 6       657      1000      1000       800                           0
```

```
credit_data_lasso <- data.frame(dummy[,-1])

set.seed(14283873)

index <- sample(nrow(credit_card_data),nrow(credit_card_data)*0.80)

credit_card_data_train_Y = credit_data_lasso[index,"default_payment_next_mont
h1"]

credit_card_data_test_Y = credit_data_lasso[-index,"default_payment_next_mont
h1"]

credit_card_data_train_X = as.matrix(dplyr::select(credit_data_lasso, -defaul
t_payment_next_month1)[index,])

credit_card_data_test_X = as.matrix(dplyr::select(credit_data_lasso, -default
_payment_next_month1)[-index,])
```

Checking dimensions

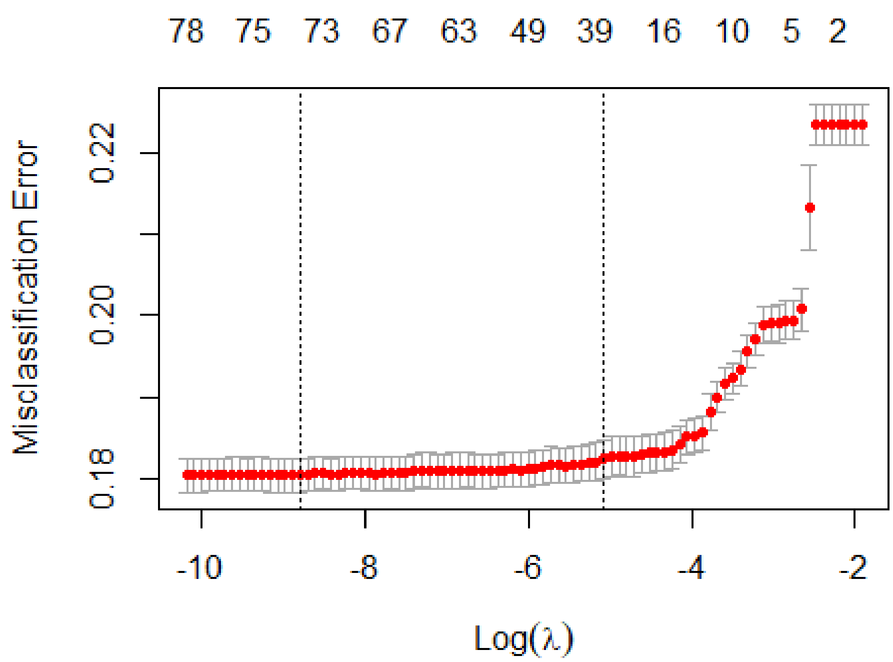
```
str(credit_card_data_train_X)

##  num [1:24000, 1:82] 60000 290000 220000 20000 80000 360000 170000 20000 1
50000 80000 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:24000] "15183" "19813" "10560" "7678" ...
##   ..$ : chr [1:82] "LIMIT_BAL" "SEX2" "EDUCATION1" "EDUCATION2" ...

str(credit_card_data_train_Y)

##  num [1:24000] 0 0 0 0 0 0 0 0 1 1 ...

str(credit_card_data_test_X)

##  num [1:6000, 1:82] 140000 20000 260000 50000 230000 500000 160000 20000 1
80000 500000 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:6000] "9" "10" "12" "29" ...
##   ..$ : chr [1:82] "LIMIT_BAL" "SEX2" "EDUCATION1" "EDUCATION2" ...

str(credit_card_data_test_Y)

##  num [1:6000] 0 0 0 0 0 0 0 1 0 0 ...
```

#LASSO

```
library(glmnet)

credit_default_lasso <- glmnet(x=credit_card_data_train_X, y=credit_card_data
_train_Y, family = "binomial")
```

Performing cross-validation to determine the shrinkage parameter

```
credit_lasso_cv <- cv.glmnet(x=credit_card_data_train_X, y=credit_card_data_t
rain_Y, family = "binomial", type.measure = "class")
plot(credit_lasso_cv)
```



Calculating coefficients using Lambda min

```
coef(credit_default_lasso, s=credit_lasso_cv$lambda.min)

## 83 x 1 sparse Matrix of class "dgCMatrix"
##                         s1
## (Intercept) -1.858093e+00
## LIMIT_BAL   -1.760362e-06
## SEX2        -1.303843e-01
## EDUCATION1   1.078806e-02
## EDUCATION2   4.688392e-02
## EDUCATION3  -2.853131e-02
## EDUCATION4  -1.207079e+00
## EDUCATION5  -1.176967e+00
## EDUCATION6  -1.892908e-01
## MARRIAGE1    6.537543e-01
## MARRIAGE2    5.134560e-01
## MARRIAGE3    7.943517e-01
```

```
## AGE          3.700778e-03
## PAY_0.1      5.705219e-01
## PAY_00      -2.255277e-01
## PAY_01       8.039033e-01
## PAY_02       2.069624e+00
## PAY_03       1.922528e+00
## PAY_04       1.923802e+00
## PAY_05       1.317083e+00
## PAY_06      -4.835447e-01
## PAY_07      -7.312651e-01
## PAY_08      -6.004081e-01
## PAY_2.1     -3.035947e-01
## PAY_20      -2.982134e-02
## PAY_21      -6.526346e-01
## PAY_22       1.187020e-02
## PAY_23       7.991498e-02
## PAY_24      -4.057113e-01
## PAY_25       2.118004e+00
## PAY_26       3.213223e+00
## PAY_27       .
## PAY_28       .
## PAY_3.1      .
## PAY_30       6.073035e-02
## PAY_31      -9.604622e-01
## PAY_32       3.925134e-01
## PAY_33       3.837856e-01
## PAY_34      -5.024126e-01
## PAY_35      -9.625038e-01
## PAY_36       2.726557e+00
## PAY_37       4.772097e-02
## PAY_38      -4.444323e+00
## PAY_4.1     -1.408405e-01
## PAY_40      -1.457141e-01
## PAY_41       6.022485e+00
## PAY_42       1.345145e-01
## PAY_43      -1.399087e-02
## PAY_44       4.913421e-01
## PAY_45      -1.024909e+00
## PAY_46      -6.243302e+00
## PAY_47       .
## PAY_48      -6.605627e+00
## PAY_5.1     -1.149793e-01
## PAY_50       6.831132e-02
## PAY_52       3.764342e-01
## PAY_53       4.964307e-02
## PAY_54       .
## PAY_55       5.924082e-01
## PAY_56       6.656411e+00
## PAY_57       2.548794e+00
## PAY_58       4.954925e+00
```

```
## PAY_6.1     -7.789002e-02
## PAY_60      -2.766505e-01
## PAY_62       1.843543e-02
## PAY_63       5.883395e-01
## PAY_64      -2.132068e-01
## PAY_65       8.530263e-01
## PAY_66       8.747063e-02
## PAY_67      -1.534455e+00
## PAY_68       8.812993e+00
## BILL_AMT1   -1.128568e-06
## BILL_AMT2    1.715863e-06
## BILL_AMT3    2.147493e-06
## BILL_AMT4    .
## BILL_AMT5   -6.728090e-07
## BILL_AMT6    .
## PAY_AMT1    -1.071537e-05
## PAY_AMT2    -8.160608e-06
## PAY_AMT3    -3.492515e-06
## PAY_AMT4    -1.127769e-06
## PAY_AMT5    -4.051189e-06
## PAY_AMT6    -3.115225e-06
```

```
lasso.min.glm <- glm(default_payment_next_month ~ LIMIT_BAL + SEX + EDUCATION
+
                       MARRIAGE +  AGE + PAY_0 + PAY_2 + PAY_3 + PAY_4 + PAY_
5 +PAY_6 +
                       BILL_AMT1 + BILL_AMT2 + BILL_AMT3 + BILL_AMT4 + BILL_A
MT5 + BILL_AMT6 +
                       PAY_AMT1 + PAY_AMT2 + PAY_AMT3 + PAY_AMT4 + PAY_AMT5 +
PAY_AMT6, family = "binomial", data=credit_card_data_train)

lasso.min.glm_summary<-summary(lasso.min.glm)

lasso.min.glm_summary
```

```
##
## Call:
## glm(formula = default_payment_next_month ~ LIMIT_BAL + SEX +
##     EDUCATION + MARRIAGE + AGE + PAY_0 + PAY_2 + PAY_3 + PAY_4 +
##     PAY_5 + PAY_6 + BILL_AMT1 + BILL_AMT2 + BILL_AMT3 + BILL_AMT4 +
##     BILL_AMT5 + BILL_AMT6 + PAY_AMT1 + PAY_AMT2 + PAY_AMT3 +
##     PAY_AMT4 + PAY_AMT5 + PAY_AMT6, family = "binomial", data = credit_car
d_data_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5178  -0.5993  -0.5098  -0.2951   3.5160
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept) -1.608e+01  2.604e+02  -0.062 0.950753
## LIMIT_BAL   -1.757e-06  1.949e-07  -9.017  < 2e-16 ***
## SEX2        -1.314e-01  3.618e-02  -3.631 0.000282 ***
## EDUCATION1   1.283e+01  2.604e+02   0.049 0.960703
## EDUCATION2   1.287e+01  2.604e+02   0.049 0.960595
## EDUCATION3   1.279e+01  2.604e+02   0.049 0.960819
## EDUCATION4   1.156e+01  2.604e+02   0.044 0.964598
## EDUCATION5   1.161e+01  2.604e+02   0.045 0.964429
## EDUCATION6   1.261e+01  2.604e+02   0.048 0.961392
## MARRIAGE1    2.051e+00  6.532e-01   3.141 0.001686 **
## MARRIAGE2    1.912e+00  6.533e-01   2.927 0.003426 **
## MARRIAGE3    2.199e+00  6.723e-01   3.271 0.001071 **
## AGE          3.902e-03  2.205e-03   1.770 0.076796 .
## PAY_0-1      6.146e-01  1.215e-01   5.060 4.19e-07 ***
## PAY_00      -1.853e-01  1.311e-01  -1.414 0.157489
## PAY_01       8.259e-01  9.426e-02   8.762  < 2e-16 ***
## PAY_02       2.109e+00  1.190e-01  17.720  < 2e-16 ***
## PAY_03       1.956e+00  1.860e-01  10.511  < 2e-16 ***
## PAY_04       1.985e+00  3.493e-01   5.682 1.33e-08 ***
## PAY_05       1.438e+00  5.456e-01   2.636 0.008377 **
## PAY_06      -7.054e-01  1.375e+00  -0.513 0.608038
## PAY_07      -1.262e+01  6.187e+02  -0.020 0.983731
## PAY_08      -1.309e+01  8.827e+02  -0.015 0.988169
## PAY_2-1     -3.672e-01  1.279e-01  -2.871 0.004094 **
## PAY_20      -1.044e-01  1.550e-01  -0.673 0.500779
## PAY_21      -6.855e-01  5.795e-01  -1.183 0.236824
## PAY_22      -4.859e-02  1.307e-01  -0.372 0.710042
## PAY_23       1.405e-02  2.016e-01   0.070 0.944440
## PAY_24      -5.142e-01  3.633e-01  -1.416 0.156899
## PAY_25       2.323e+00  1.180e+00   1.969 0.049005 *
## PAY_26       1.507e+01  6.187e+02   0.024 0.980574
## PAY_27       9.470e-01  9.899e+02   0.001 0.999237
## PAY_28       1.447e+01  1.134e+03   0.013 0.989826
## PAY_3-1      7.783e-02  1.224e-01   0.636 0.524836
## PAY_30       1.532e-01  1.412e-01   1.085 0.278116
## PAY_31      -1.265e+01  8.827e+02  -0.014 0.988565
## PAY_32       4.772e-01  1.428e-01   3.341 0.000834 ***
## PAY_33       4.969e-01  2.540e-01   1.956 0.050419 .
## PAY_34      -4.239e-01  4.627e-01  -0.916 0.359540
## PAY_35      -9.385e-01  8.686e-01  -1.080 0.279937
## PAY_36       1.443e+01  4.479e+02   0.032 0.974288
## PAY_37       1.389e-01  1.005e+00   0.138 0.890012
## PAY_38      -2.578e+01  4.974e+02  -0.052 0.958669
## PAY_4-1     -2.209e-01  1.229e-01  -1.796 0.072438 .
## PAY_40      -2.460e-01  1.371e-01  -1.795 0.072700 .
## PAY_41       2.852e+01  1.248e+03   0.023 0.981773
## PAY_42       4.012e-02  1.463e-01   0.274 0.783887
## PAY_43      -1.537e-01  2.794e-01  -0.550 0.582157
## PAY_44       3.930e-01  5.098e-01   0.771 0.440855
## PAY_45      -1.338e+00  8.441e-01  -1.585 0.112860
```

```
## PAY_46      -2.933e+01  7.126e+02  -0.041 0.967170
## PAY_47      -1.951e+00  6.370e+02  -0.003 0.997557
## PAY_48      -3.130e+01  1.052e+03  -0.030 0.976253
## PAY_5-1     -5.877e-02  1.202e-01  -0.489 0.624934
## PAY_50       1.510e-01  1.328e-01   1.136 0.255810
## PAY_52       4.600e-01  1.486e-01   3.095 0.001970 **
## PAY_53       1.542e-01  2.733e-01   0.564 0.572644
## PAY_54       1.306e-01  5.291e-01   0.247 0.804956
## PAY_55       9.381e-01  9.618e-01   0.975 0.329380
## PAY_56       3.930e+01  8.145e+02   0.048 0.961519
## PAY_57       1.556e+01  5.716e+02   0.027 0.978289
## PAY_58       4.186e+01  2.098e+03   0.020 0.984077
## PAY_6-1     -1.054e-01  9.243e-02  -1.141 0.254059
## PAY_60      -3.156e-01  9.907e-02  -3.186 0.001444 **
## PAY_62      -1.543e-02  1.154e-01  -0.134 0.893675
## PAY_63       5.765e-01  2.621e-01   2.199 0.027848 *
## PAY_64      -3.145e-01  5.239e-01  -0.600 0.548240
## PAY_65       8.037e-01  1.019e+00   0.789 0.430157
## PAY_66      -2.789e-01  1.129e+00  -0.247 0.804860
## PAY_67      -1.269e+01  2.813e+02  -0.045 0.964022
## PAY_68       2.879e+01  1.270e+03   0.023 0.981917
## BILL_AMT1   -1.888e-06  1.242e-06  -1.520 0.128527
## BILL_AMT2    2.472e-06  1.676e-06   1.475 0.140315
## BILL_AMT3    2.636e-06  1.458e-06   1.809 0.070482 .
## BILL_AMT4   -3.516e-07  1.467e-06  -0.240 0.810642
## BILL_AMT5   -1.209e-06  1.719e-06  -0.703 0.481785
## BILL_AMT6    4.343e-07  1.368e-06   0.317 0.750872
## PAY_AMT1    -1.159e-05  2.595e-06  -4.467 7.94e-06 ***
## PAY_AMT2    -8.394e-06  2.287e-06  -3.671 0.000242 ***
## PAY_AMT3    -3.247e-06  2.130e-06  -1.524 0.127406
## PAY_AMT4    -9.531e-07  1.912e-06  -0.499 0.618086
## PAY_AMT5    -4.421e-06  2.054e-06  -2.153 0.031328 *
## PAY_AMT6    -3.114e-06  1.512e-06  -2.060 0.039392 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 25495  on 23999  degrees of freedom
## Residual deviance: 20832  on 23917  degrees of freedom
## AIC: 20998
##
## Number of Fisher Scoring iterations: 13
```

```
AIC(lasso.min.glm) # 20997.55
```

```
## [1] 20997.55
```

```
BIC(lasso.min.glm) # 21668.67
```

```
## [1] 21668.67
```

```r
lasso.min.glm$deviance # 20831.55
```

```
## [1] 20831.55
```

Calculating Model mean residual deviance (in-sample)

```r
lasso.min.glm$dev/lasso.min.glm$df.residual # 0.8709933
```

```
## [1] 0.8709933
```

Calculating coefficients using Lambda 1se

```r
coef(credit_default_lasso, s=credit_lasso_cv$lambda.1se)
```

```
## 83 x 1 sparse Matrix of class "dgCMatrix"
##                          s1
## (Intercept) -1.209709e+00
## LIMIT_BAL   -1.420220e-06
## SEX2        -5.861743e-02
## EDUCATION1   .
## EDUCATION2   .
## EDUCATION3   .
## EDUCATION4  -8.413459e-02
## EDUCATION5  -3.889263e-01
## EDUCATION6   .
## MARRIAGE1    4.331142e-02
## MARRIAGE2   -3.001809e-02
## MARRIAGE3    .
## AGE          4.728057e-04
## PAY_0.1      .
## PAY_00      -4.245929e-01
## PAY_01       4.015517e-01
## PAY_02       1.661649e+00
## PAY_03       1.384610e+00
## PAY_04       1.031806e+00
## PAY_05       9.973009e-02
## PAY_06       .
## PAY_07       .
## PAY_08       .
## PAY_2.1     -4.854628e-02
## PAY_20       .
## PAY_21       .
## PAY_22       2.212984e-01
## PAY_23       2.284162e-01
## PAY_24       .
## PAY_25       2.157309e-01
## PAY_26       9.316508e-01
## PAY_27       5.286922e-01
## PAY_28       .
## PAY_3.1     -3.507524e-02
## PAY_30       .
```

```
## PAY_31         .
## PAY_32          3.263531e-01
## PAY_33          1.747145e-01
## PAY_34         .
## PAY_35         .
## PAY_36         .
## PAY_37         .
## PAY_38         .
## PAY_4.1        -9.427373e-03
## PAY_40         .
## PAY_41         .
## PAY_42          2.394056e-01
## PAY_43         .
## PAY_44          6.509339e-02
## PAY_45         .
## PAY_46         .
## PAY_47          3.212301e-03
## PAY_48         .
## PAY_5.1        -6.076462e-02
## PAY_50         .
## PAY_52          3.136286e-01
## PAY_53         .
## PAY_54         .
## PAY_55         .
## PAY_56         .
## PAY_57          5.284801e-01
## PAY_58         .
## PAY_6.1        .
## PAY_60         -3.304292e-02
## PAY_62          1.730696e-01
## PAY_63          4.507172e-01
## PAY_64         .
## PAY_65         .
## PAY_66         .
## PAY_67         .
## PAY_68         .
## BILL_AMT1      .
## BILL_AMT2      .
## BILL_AMT3      .
## BILL_AMT4      .
## BILL_AMT5      .
## BILL_AMT6      .
## PAY_AMT1       -4.030889e-06
## PAY_AMT2       -2.493118e-06
## PAY_AMT3       -1.168346e-06
## PAY_AMT4       .
## PAY_AMT5       -6.271596e-07
## PAY_AMT6       -1.214479e-07
```

```r
lasso.1se.glm <- glm(default_payment_next_month ~ LIMIT_BAL + SEX + EDUCATION
+
                       MARRIAGE +  AGE + PAY_0 + PAY_2 + PAY_3 + PAY_4 + PAY_
5 +PAY_6 +
                       PAY_AMT1 + PAY_AMT2 + PAY_AMT3 + PAY_AMT5 + PAY_AMT6,
family = "binomial", data=credit_card_data_train)

lasso.1se.glm_summary<-summary(lasso.1se.glm)

lasso.1se.glm_summary

##
## Call:
## glm(formula = default_payment_next_month ~ LIMIT_BAL + SEX +
##     EDUCATION + MARRIAGE + AGE + PAY_0 + PAY_2 + PAY_3 + PAY_4 +
##     PAY_5 + PAY_6 + PAY_AMT1 + PAY_AMT2 + PAY_AMT3 + PAY_AMT5 +
##     PAY_AMT6, family = "binomial", data = credit_card_data_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2827  -0.5975  -0.5097  -0.3143   3.3401
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.624e+01  2.611e+02  -0.062 0.950407
## LIMIT_BAL   -1.393e-06  1.756e-07  -7.934 2.12e-15 ***
## SEX2        -1.313e-01  3.612e-02  -3.635 0.000278 ***
## EDUCATION1   1.284e+01  2.611e+02   0.049 0.960788
## EDUCATION2   1.288e+01  2.611e+02   0.049 0.960672
## EDUCATION3   1.280e+01  2.611e+02   0.049 0.960901
## EDUCATION4   1.158e+01  2.611e+02   0.044 0.964622
## EDUCATION5   1.166e+01  2.611e+02   0.045 0.964402
## EDUCATION6   1.265e+01  2.611e+02   0.048 0.961365
## MARRIAGE1    2.071e+00  6.544e-01   3.165 0.001551 **
## MARRIAGE2    1.933e+00  6.545e-01   2.953 0.003148 **
## MARRIAGE3    2.209e+00  6.735e-01   3.280 0.001038 **
## AGE          4.435e-03  2.205e-03   2.012 0.044237 *
## PAY_0-1      6.240e-01  1.209e-01   5.162 2.45e-07 ***
## PAY_00      -1.411e-01  1.301e-01  -1.085 0.277973
## PAY_01       8.579e-01  9.383e-02   9.143  < 2e-16 ***
## PAY_02       2.171e+00  1.181e-01  18.374  < 2e-16 ***
## PAY_03       1.988e+00  1.858e-01  10.696  < 2e-16 ***
## PAY_04       2.047e+00  3.496e-01   5.854 4.79e-09 ***
## PAY_05       1.517e+00  5.488e-01   2.764 0.005713 **
## PAY_06      -6.726e-01  1.370e+00  -0.491 0.623496
## PAY_07      -1.248e+01  6.211e+02  -0.020 0.983969
## PAY_08      -1.289e+01  8.827e+02  -0.015 0.988346
## PAY_2-1     -3.849e-01  1.267e-01  -3.038 0.002380 **
## PAY_20      -9.246e-02  1.526e-01  -0.606 0.544511
## PAY_21      -7.135e-01  5.713e-01  -1.249 0.211655
```

```
## PAY_22      -2.105e-02  1.282e-01  -0.164 0.869548
## PAY_23       4.118e-02  1.995e-01   0.206 0.836449
## PAY_24      -4.687e-01  3.620e-01  -1.295 0.195423
## PAY_25       2.374e+00  1.176e+00   2.019 0.043484 *
## PAY_26       1.524e+01  6.211e+02   0.025 0.980429
## PAY_27       9.890e-01  9.889e+02   0.001 0.999202
## PAY_28       1.463e+01  1.132e+03   0.013 0.989687
## PAY_3-1      7.004e-02  1.213e-01   0.577 0.563631
## PAY_30       2.039e-01  1.391e-01   1.465 0.142888
## PAY_31      -1.275e+01  8.827e+02  -0.014 0.988473
## PAY_32       5.172e-01  1.411e-01   3.666 0.000246 ***
## PAY_33       5.334e-01  2.531e-01   2.108 0.035041 *
## PAY_34      -3.740e-01  4.647e-01  -0.805 0.420886
## PAY_35      -9.184e-01  8.721e-01  -1.053 0.292299
## PAY_36       1.451e+01  4.458e+02   0.033 0.974030
## PAY_37       1.422e-01  1.004e+00   0.142 0.887359
## PAY_38      -2.572e+01  4.973e+02  -0.052 0.958756
## PAY_4-1     -1.967e-01  1.221e-01  -1.612 0.107032
## PAY_40      -1.874e-01  1.355e-01  -1.382 0.166842
## PAY_41       2.911e+01  1.248e+03   0.023 0.981398
## PAY_42       8.627e-02  1.449e-01   0.595 0.551715
## PAY_43      -9.652e-02  2.786e-01  -0.346 0.729023
## PAY_44       4.245e-01  5.103e-01   0.832 0.405487
## PAY_45      -1.345e+00  8.453e-01  -1.591 0.111509
## PAY_46      -2.935e+01  7.088e+02  -0.041 0.966969
## PAY_47      -1.877e+00  6.367e+02  -0.003 0.997647
## PAY_48      -3.119e+01  1.051e+03  -0.030 0.976333
## PAY_5-1     -7.791e-02  1.192e-01  -0.654 0.513206
## PAY_50       1.469e-01  1.316e-01   1.116 0.264330
## PAY_52       4.582e-01  1.473e-01   3.110 0.001872 **
## PAY_53       1.413e-01  2.729e-01   0.518 0.604695
## PAY_54       1.602e-01  5.291e-01   0.303 0.762025
## PAY_55       9.323e-01  9.611e-01   0.970 0.332016
## PAY_56       3.927e+01  8.139e+02   0.048 0.961523
## PAY_57       1.538e+01  5.711e+02   0.027 0.978513
## PAY_58       4.137e+01  2.096e+03   0.020 0.984251
## PAY_6-1     -1.166e-01  9.217e-02  -1.265 0.205919
## PAY_60      -2.984e-01  9.704e-02  -3.075 0.002104 **
## PAY_62      -1.033e-02  1.131e-01  -0.091 0.927245
## PAY_63       5.534e-01  2.612e-01   2.118 0.034142 *
## PAY_64      -3.318e-01  5.243e-01  -0.633 0.526818
## PAY_65       7.696e-01  1.012e+00   0.760 0.447111
## PAY_66      -1.649e-01  1.094e+00  -0.151 0.880186
## PAY_67      -1.259e+01  2.813e+02  -0.045 0.964294
## PAY_68       2.912e+01  1.268e+03   0.023 0.981676
## PAY_AMT1    -7.725e-06  2.174e-06  -3.553 0.000381 ***
## PAY_AMT2    -6.315e-06  1.918e-06  -3.291 0.000997 ***
## PAY_AMT3    -3.692e-06  1.850e-06  -1.996 0.045937 *
## PAY_AMT5    -3.353e-06  1.678e-06  -1.997 0.045781 *
## PAY_AMT6    -2.585e-06  1.452e-06  -1.780 0.075051 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 25495  on 23999  degrees of freedom
## Residual deviance: 20871  on 23924  degrees of freedom
## AIC: 21023
##
## Number of Fisher Scoring iterations: 13
```

```
AIC(lasso.1se.glm) # 21023.12
```

```
## [1] 21023.12
```

```
BIC(lasso.1se.glm) # 21637.65
```

```
## [1] 21637.65
```

```
lasso.1se.glm$deviance # 20871.12
```

```
## [1] 20871.12
```

```
# Calculating Model mean residual deviance (in-sample)

lasso.1se.glm$dev/lasso.1se.glm$df.residual # 0.8723928
```

```
## [1] 0.8723928
```

Plotting graph for LASSO

#install.packages('glmnet')

```
#library(glmnet)

str(credit_card_data_train)
```

```
## 'data.frame':    24000 obs. of  24 variables:
##  $ LIMIT_BAL                    : int  60000 290000 220000 20000 80000 360000
170000 20000 150000 80000 ...
##  $ SEX                          : Factor w/ 2 levels "1","2": 2 2 2 2 2 2 1 1
2 1 ...
##  $ EDUCATION                    : Factor w/ 7 levels "0","1","2","3",..: 3 3
2 4 4 2 4 4 3 3 ...
##  $ MARRIAGE                     : Factor w/ 4 levels "0","1","2","3": 3 3 3 2
2 3 2 3 3 3 ...
##  $ AGE                          : int  25 41 27 23 48 29 31 48 28 32 ...
##  $ PAY_0                        : Factor w/ 11 levels "-2","-1","0",..: 3 2 2
3 3 3 5 2 5 3 ...
##  $ PAY_2                        : Factor w/ 11 levels "-2","-1","0",..: 3 3 2
3 3 3 3 2 3 3 ...
##  $ PAY_3                        : Factor w/ 11 levels "-2","-1","0",..: 3 3 2
```

```
3 3 3 3 1 3 3 ...
##  $ PAY_4                     : Factor w/ 11 levels "-2","-1","0",..: 3 3 3
3 3 3 3 2 3 3 ...
##  $ PAY_5                     : Factor w/ 10 levels "-2","-1","0",..: 3 3 3
4 3 3 3 2 3 3 ...
##  $ PAY_6                     : Factor w/ 10 levels "-2","-1","0",..: 3 3 2
4 3 3 3 2 3 3 ...
##  $ BILL_AMT1                 : int  6234 32192 471 20030 44508 27370 17048
5 170 89336 78239 ...
##  $ BILL_AMT2                 : int  7402 32595 5261 18808 43709 20750 2298
05 -220 90337 80426 ...
##  $ BILL_AMT3                 : int  8270 14688 16436 17080 52660 14898 167
524 -610 84905 81767 ...
##  $ BILL_AMT4                 : int  9287 17105 16071 18293 48532 7524 1854
14 390 86814 78340 ...
##  $ BILL_AMT5                 : int  10076 18875 96 18752 44469 7524 121092
0 73827 36895 ...
##  $ BILL_AMT6                 : int  10612 21304 480 19195 45068 0 117320 1
9993 60284 39079 ...
##  $ PAY_AMT1                  : int  1276 2500 5272 1600 1785 1328 8000 0 4
027 4500 ...
##  $ PAY_AMT2                  : int  1144 13500 17571 1300 10000 2000 8000
0 4004 5009 ...
##  $ PAY_AMT3                  : int  1161 3000 0 1500 1541 150 5000 1390 50
16 5000 ...
##  $ PAY_AMT4                  : int  950 3000 0 900 1610 0 5000 0 5004 2000
...
##  $ PAY_AMT5                  : int  700 3000 549 900 1604 0 4500 20773 300
0 4000 ...
##  $ PAY_AMT6                  : int  1000 2000 0 787 1485 0 5000 400 5005 1
500 ...
##  $ default_payment_next_month: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1
2 2 ...

## Converting to integer for plotting

credit_card_data_train$SEX <- as.integer(credit_card_data_train$SEX)
credit_card_data_train$MARRIAGE <- as.integer(credit_card_data_train$MARRIAGE
)
credit_card_data_train$EDUCATION <- as.integer(credit_card_data_train$EDUCATI
ON)
credit_card_data_train$default_payment_next_month <- as.integer(credit_card_d
ata_train$default_payment_next_month)
credit_card_data_train$PAY_0 <- as.integer(credit_card_data_train$PAY_0)
credit_card_data_train$PAY_2 <- as.integer(credit_card_data_train$PAY_2)
credit_card_data_train$PAY_3 <- as.integer(credit_card_data_train$PAY_3)
credit_card_data_train$PAY_4 <- as.integer(credit_card_data_train$PAY_4)
credit_card_data_train$PAY_5 <- as.integer(credit_card_data_train$PAY_5)
credit_card_data_train$PAY_6 <- as.integer(credit_card_data_train$PAY_6)
```

Checking structure to confirm conversion

```
str(credit_card_data_train)

## 'data.frame':    24000 obs. of  24 variables:
##  $ LIMIT_BAL                 : int  60000 290000 220000 20000 80000 360000
170000 20000 150000 80000 ...
##  $ SEX                       : int  2 2 2 2 2 2 1 1 2 1 ...
##  $ EDUCATION                 : int  3 3 2 4 4 2 4 4 3 3 ...
##  $ MARRIAGE                  : int  3 3 3 2 2 3 2 3 3 3 ...
##  $ AGE                       : int  25 41 27 23 48 29 31 48 28 32 ...
##  $ PAY_0                     : int  3 2 2 3 3 3 5 2 5 3 ...
##  $ PAY_2                     : int  3 3 2 3 3 3 3 2 3 3 ...
##  $ PAY_3                     : int  3 3 2 3 3 3 3 1 3 3 ...
##  $ PAY_4                     : int  3 3 3 3 3 3 3 2 3 3 ...
##  $ PAY_5                     : int  3 3 3 4 3 3 3 2 3 3 ...
##  $ PAY_6                     : int  3 3 2 4 3 3 3 2 3 3 ...
##  $ BILL_AMT1                 : int  6234 32192 471 20030 44508 27370 17048
5 170 89336 78239 ...
##  $ BILL_AMT2                 : int  7402 32595 5261 18808 43709 20750 2298
05 -220 90337 80426 ...
##  $ BILL_AMT3                 : int  8270 14688 16436 17080 52660 14898 167
524 -610 84905 81767 ...
##  $ BILL_AMT4                 : int  9287 17105 16071 18293 48532 7524 1854
14 390 86814 78340 ...
##  $ BILL_AMT5                 : int  10076 18875 96 18752 44469 7524 121092
0 73827 36895 ...
##  $ BILL_AMT6                 : int  10612 21304 480 19195 45068 0 117320 1
9993 60284 39079 ...
##  $ PAY_AMT1                  : int  1276 2500 5272 1600 1785 1328 8000 0 4
027 4500 ...
##  $ PAY_AMT2                  : int  1144 13500 17571 1300 10000 2000 8000
0 4004 5009 ...
##  $ PAY_AMT3                  : int  1161 3000 0 1500 1541 150 5000 1390 50
16 5000 ...
##  $ PAY_AMT4                  : int  950 3000 0 900 1610 0 5000 0 5004 2000
...
##  $ PAY_AMT5                  : int  700 3000 549 900 1604 0 4500 20773 300
0 4000 ...
##  $ PAY_AMT6                  : int  1000 2000 0 787 1485 0 5000 400 5005 1
500 ...
##  $ default_payment_next_month: int  1 1 1 1 1 1 1 1 2 2 ...
```

performing lasso

```
lasso_fit <- glmnet(x = as.matrix(credit_card_data_train[, -c(which(colnames(
credit_card_data_train)=='default_payment_next_month'))]), y = credit_card_da
ta_train$default_payment_next_month, alpha = 1)

#lambda = 0.5
coef(lasso_fit,s=0.5)
```

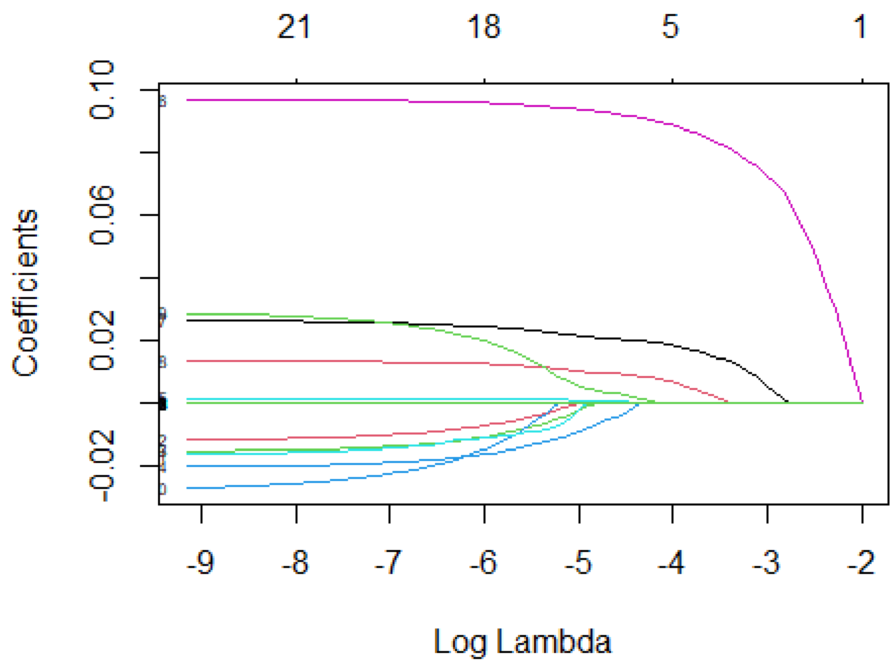
```
## 24 x 1 sparse Matrix of class "dgCMatrix"
##                   s1
## (Intercept) 1.223375
## LIMIT_BAL   .
## SEX         .
## EDUCATION   .
## MARRIAGE    .
## AGE         .
## PAY_0       .
## PAY_2       .
## PAY_3       .
## PAY_4       .
## PAY_5       .
## PAY_6       .
## BILL_AMT1   .
## BILL_AMT2   .
## BILL_AMT3   .
## BILL_AMT4   .
## BILL_AMT5   .
## BILL_AMT6   .
## PAY_AMT1    .
## PAY_AMT2    .
## PAY_AMT3    .
## PAY_AMT4    .
## PAY_AMT5    .
## PAY_AMT6    .
```

```
#lambda = 1
coef(lasso_fit,s=1)
```

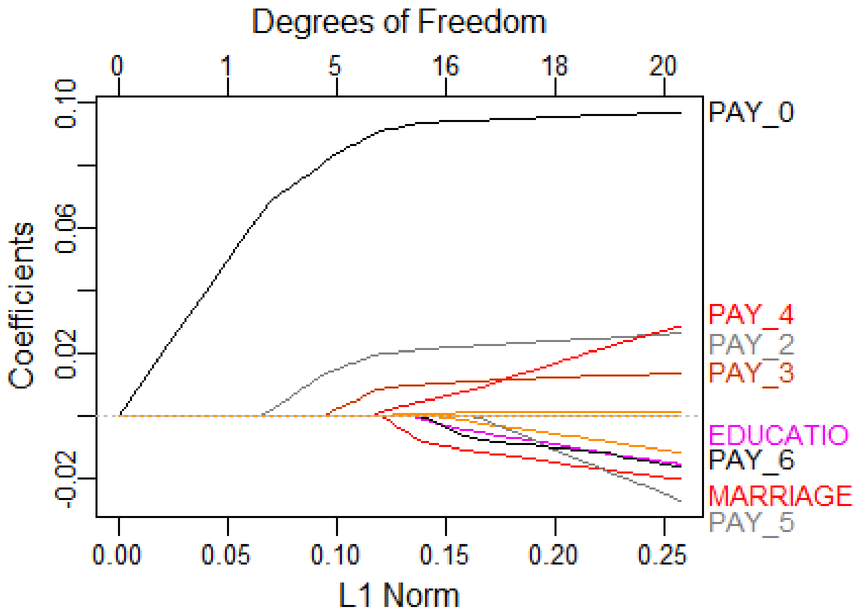
```
## 24 x 1 sparse Matrix of class "dgCMatrix"
##                   s1
## (Intercept) 1.223375
## LIMIT_BAL   .
## SEX         .
## EDUCATION   .
## MARRIAGE    .
## AGE         .
## PAY_0       .
## PAY_2       .
## PAY_3       .
## PAY_4       .
## PAY_5       .
## PAY_6       .
## BILL_AMT1   .
## BILL_AMT2   .
## BILL_AMT3   .
## BILL_AMT4   .
## BILL_AMT5   .
## BILL_AMT6   .
```

```
## PAY_AMT1    .
## PAY_AMT2    .
## PAY_AMT3    .
## PAY_AMT4    .
## PAY_AMT5    .
## PAY_AMT6    .
```

```r
plot(lasso_fit, xvar = "lambda", label = TRUE)
```

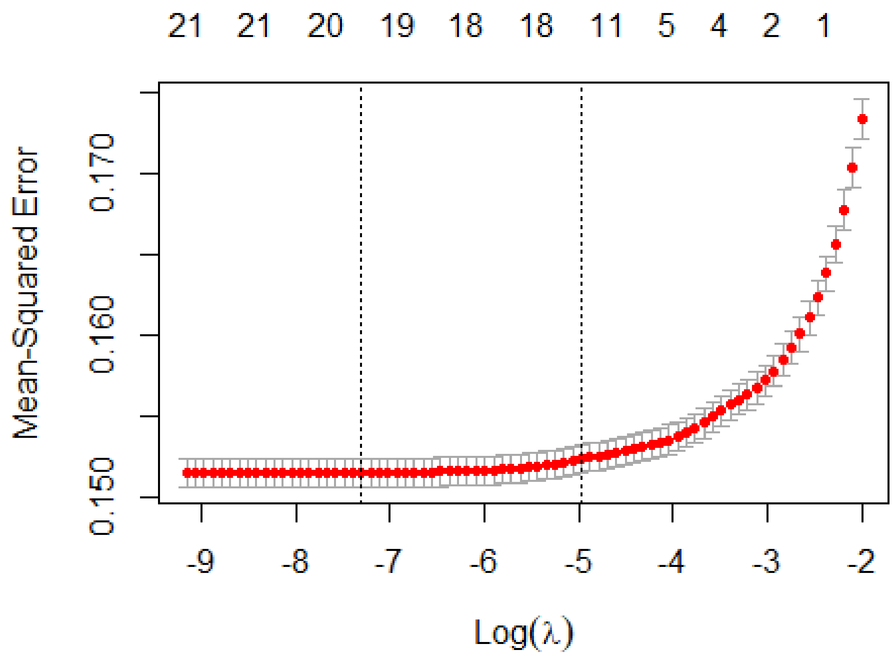



```r
library(plotmo)

plot_glmnet(lasso_fit, label=8, xvar ="norm")  # label the 8 biggest final co
efs
```



#use 5-fold cross validation to pick lambda

```
cv_lasso_fit <- cv.glmnet(x = as.matrix(credit_card_data_train[, -c(which(colnames(credit_card_data_train)=='default_payment_next_month'))]), y = credit_card_data_train$default_payment_next_month, alpha = 1, nfolds = 5)
plot(cv_lasso_fit)
```




```
cv_lasso_fit$lambda.min

## [1] 0.000674677
```

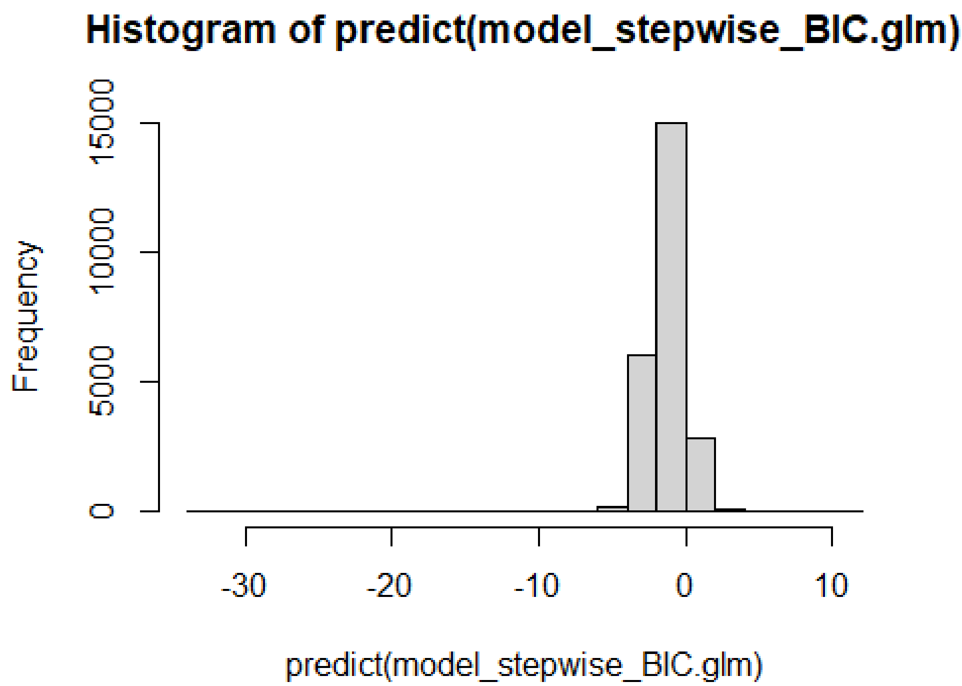
#—————————————-Credit Card Default Case Study Part AB#—————————————-

head(credit_card_data_train) ## Best Logistic Regression model chosen was with Stepwise BIC - below are the variables

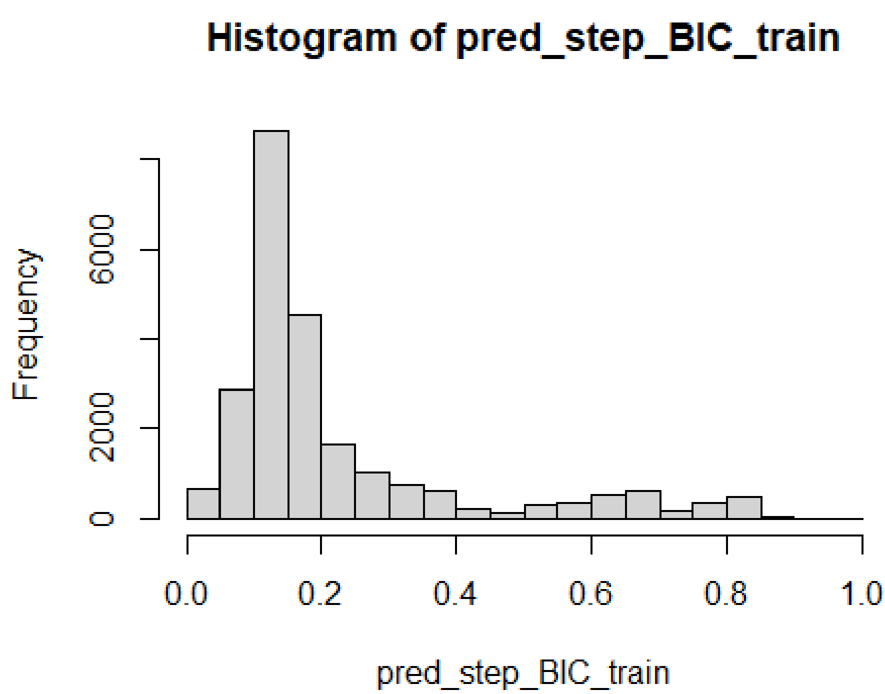
```
#model_stepwise_BIC.glm <- glm(default_payment_next_month ~ PAY_0 + LIMIT_BAL
+ PAY_5 + PAY_AMT2 +
#                                 PAY_AMT1 + BILL_AMT3 + MARRIAGE + SEX + PAY_
AMT3, family=binomial, #data=credit_card_data_train)
```

Drawing histogram of the prediction obtained using best model

```
hist(predict(model_stepwise_BIC.glm))
```



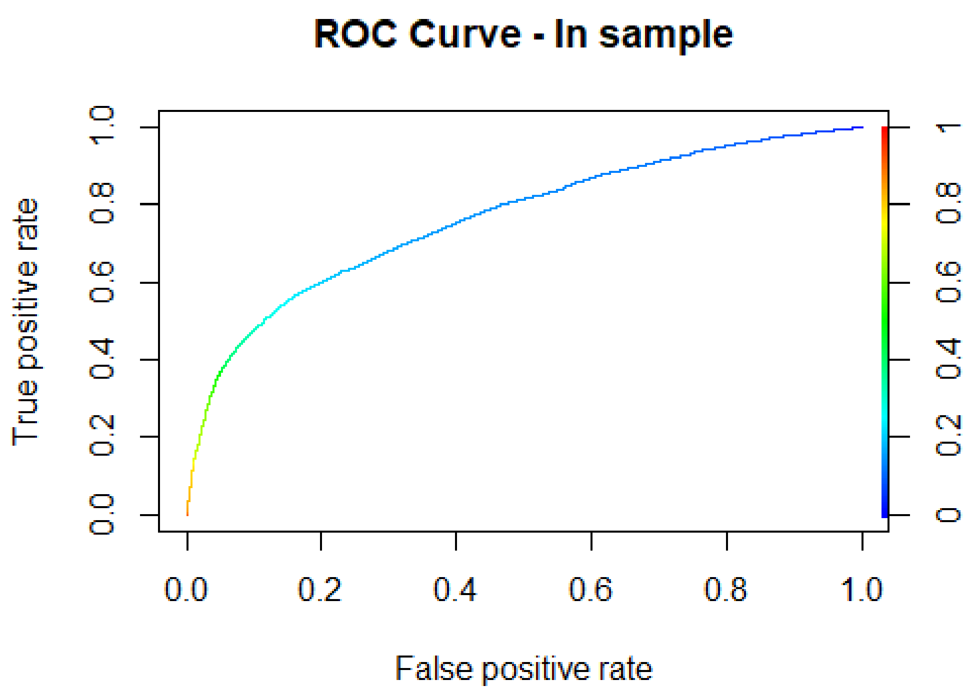

```
pred_step_BIC_train <- predict(model_stepwise_BIC.glm ,type="response")
hist(pred_step_BIC_train)
```



Drawing ROC curve

#install.packages("PRROC")

```
#library(ROCR)
#library(PRROC)

pred <- prediction(pred_step_BIC_train, credit_card_data_train$default_paymen
t_next_month)
perf <- performance(pred, "tpr", "fpr")
plot(perf, colorize=TRUE, main="ROC Curve - In sample")
```



Reporting the AUC

```
unlist(slot(performance(pred, "auc"), "y.values")) # 0.7633874

## [1] 0.7633874
```

Drawing the 2x2 Misclassification Rate Table, Default Cutoff Probability = 1/2 and Symmetric cost

```
table(credit_card_data_train$default_payment_next_month, (pred_step_BIC_train
> 0.5)*1, dnn=c("Actual values","Predicted values"))

##               Predicted values
## Actual values     0     1
##             1 17729   910
##             2  3407  1954
```

#Symmetric cost

```r
pcut <- 1/2
cost1 <- function(r, pi){
  mean(((r==0)&(pi>pcut)) | ((r==1)&(pi<pcut)))
}

cost1(r = credit_card_data_train$default_payment_next_month, pi = pred_step_B
IC_train) # 0.179875

## [1] 0.7387083

table(credit_card_data_train$default_payment_next_month, (pred_step_BIC_train
> 0.5)*1, dnn=c("Actual values","Predicted values"))

##               Predicted values
## Actual values     0     1
##             1 17729   910
##             2  3407  1954
```

Drawing the 2x2 Misclassification Rate Table, Default Cutoff Probability = 1/6 and Symmetric cost

```r
pcut <- 1/(5+1) # cost ratio 5:1
cost2 <- function(r, pi){
  weight1 <- 5
  weight0 <- 1
  c1 <- (r==1)&(pi<pcut) # logical vector - true if actual 1 but predict 0
  c0 <-(r==0)&(pi>pcut) # logical vector - true if actual 0 but predict 1
  return(mean(weight1*c1+weight0*c0))}

cost2(r = credit_card_data_train$default_payment_next_month, pi = pred_step_B
IC_train) # 0.5890417

## [1] 2.6775

table(credit_card_data_train$default_payment_next_month, (pred_step_BIC_train
> 0.167)*1, dnn=c("Actual values","Predicted values"))

##               Predicted values
## Actual values     0     1
##             1 12879  5760
##             2  1678  3683
```

#———— Out of Sample Testing ————#

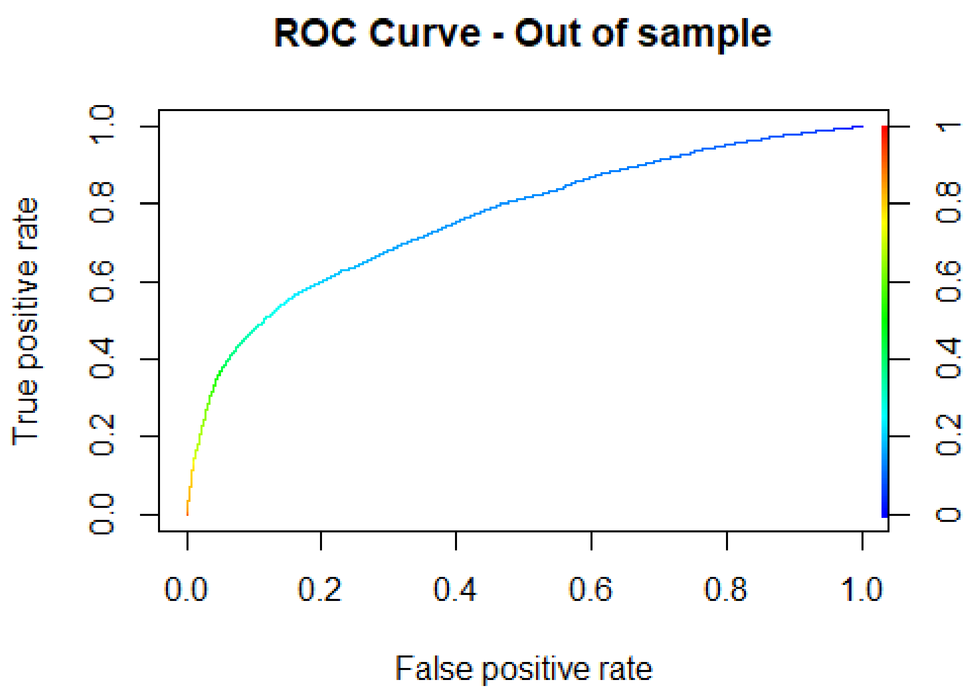
Testing on remaining 20% of the data

```r
pred_step_BIC_test<- predict(model_stepwise_BIC.glm, credit_card_data_test, t
ype="response")
```

Out-of-sample ROC curve

```r
pred_test <- prediction(pred_step_BIC_test, credit_card_data_test$default_pay
ment_next_month)
```

```
perf_test <- performance(pred, "tpr", "fpr")
plot(perf_test, colorize=TRUE, main="ROC Curve - Out of sample")
```



Reporting the AUC

```
unlist(slot(performance(pred_test, "auc"), "y.values")) # 0.7574783

## [1] 0.7574783
```

Drawing the 2x2 Misclassification Rate Table

```
pcut <- 1/2
cost3 <- function(r, pi){
  mean(((r==0)&(pi>pcut)) | ((r==1)&(pi<pcut)))
}

cost3(r = credit_card_data_test$default_payment_next_month, pi = pred_step_BI
C_test) # 0.1751667

## [1] 0.1751667

table(credit_card_data_test$default_payment_next_month, (pred_step_BIC_test >
0.5)*1, dnn=c("Actual values","Predicted values"))

##              Predicted values
## Actual values    0    1
##             0 4476  249
##             1  802  473
```

Drawing the 2x2 Misclassification Rate Table, Default Cutoff Probability = 1/6 and Symmetric cost

```
pcut <- 1/(5+1) # cost ratio 5:1
cost4 <- function(r, pi){
  weight1 <- 5
  weight0 <- 1
  c1 <- (r==1)&(pi<pcut) # logical vector - true if actual 1 but predict 0
  c0 <-(r==0)&(pi>pcut) # logical vector - true if actual 0 but predict 1
  return(mean(weight1*c1+weight0*c0))}

cost4(r = credit_card_data_test$default_payment_next_month, pi = pred_step_BI
C_test) # 0.565

## [1] 0.565

table(credit_card_data_test$default_payment_next_month, (pred_step_BIC_test >
0.167)*1, dnn=c("Actual values","Predicted values"))

##                Predicted values
## Actual values    0    1
##              0 3201 1524
##              1  375  900
```

Performing 5-Fold Cross Validation

```
library(MASS)
library(boot)
```

Importing the dataset with a new name

```
credit_card_data_cv <- read.csv("D:/UC/Classes/Spring/Data Mining 1/Credit Ca
rd Data/default_of_credit_card_clients.csv")

credit_card_data_cv <- rename(credit_card_data_cv, default_payment_next_month
= default.payment.next.month)

credit_card_data_cv = subset(credit_card_data_cv, select = -c(ID))

#install.packages("caret")

#install.packages("ROCR")

#install.packages("pROC")

library(caret)
library(ROCR)
library(pROC)
```

—————— AUC on full model without cross validation

```
model_full_wcv <- glm(default_payment_next_month ~ ., family=binomial, data=c
redit_card_data_cv)
```

print cv scores

```
log_predict_full_wcv <- predict(model_full_wcv, newdata = credit_card_data_cv
, type = "response")
log_predict_full_wcv <- ifelse(log_predict_full_wcv > 0.5,1,0)
auc(credit_card_data_cv$default_payment_next_month,log_predict_full_wcv) # 0.
6067

## Area under the curve: 0.6067

pcut <- 1/(5+1) # cost ratio 5:1
cost_full_wcv <- function(r, pi){
  weight1 <- 5
  weight0 <- 1
  c1 <- (r==1)&(pi<pcut) # logical vector - true if actual 1 but predict 0
  c0 <-(r==0)&(pi>pcut) # logical vector - true if actual 0 but predict 1
  return(mean(weight1*c1+weight0*c0))}

cost_full_wcv(r = credit_card_data_cv$default_payment_next_month, pi = log_pr
edict_full_wcv) # 0.8611667

## [1] 0.8611667
```

Calculating symmetric score - symmetric

```
pcut <- 1/2
cost_full_s <- function(r, pi){
  mean(((r==0)&(pi>pcut)) | ((r==1)&(pi<pcut)))
}

cost_full_s(r = credit_card_data_cv$default_payment_next_month, pi = log_pred
ict_full_wcv) # 0.189

## [1] 0.1890333
```

—————— AUC on full model with cross validation

define training control

```
train_control <- trainControl(method = "cv", number = 5)
```

train the model on training set

```
model_full <- train(default_payment_next_month ~ .,
               data = credit_card_data_cv,
               trControl = train_control,
               method = "glm",
               family=binomial(link="logit"), na.action=na.omit)
```

print cv scores

```
log_predict_full <- predict(model_full, newdata = credit_card_data_cv)
log_predict_full <- ifelse(log_predict_full > 0.5,1,0)
auc(credit_card_data_cv$default_payment_next_month,log_predict_full) # 0.6063

## Area under the curve: 0.6067
```

Calculating symmetric score - asymmetric

```
pcut <- 1/(5+1) # cost ratio 5:1
cost_full <- function(r, pi){
  weight1 <- 5
  weight0 <- 1
  c1 <- (r==1)&(pi<pcut) # logical vector - true if actual 1 but predict 0
  c0 <-(r==0)&(pi>pcut) # logical vector - true if actual 0 but predict 1
  return(mean(weight1*c1+weight0*c0))}

cost_full(r = credit_card_data_cv$default_payment_next_month, pi = log_predic
t_full) # 0.862

## [1] 0.8611667
```

Calculating symmetric score - symmetric

```
pcut <- 1/2
cost_full_s <- function(r, pi){
  mean(((r==0)&(pi>pcut)) | ((r==1)&(pi<pcut)))
}

cost_full_s(r = credit_card_data_cv$default_payment_next_month, pi = log_pred
ict_full) # 0.189

## [1] 0.1890333
```

Training the model using Stepwise BIC

```
#model_stepwise_BIC.glm <- glm(default_payment_next_month ~ PAY_0 + LIMIT_BAL
+ PAY_5 + PAY_AMT2 +
#                                 PAY_AMT1 + BILL_AMT3 + MARRIAGE + SEX + PAY_
AMT3, family=binomial, #data=credit_card_data_train)

model_stepwise_BIC.glm_summary<-summary(model_stepwise_BIC.glm)

model_stepwise_BIC.glm_summary

##
## Call:
## glm(formula = default_payment_next_month ~ PAY_0 + LIMIT_BAL +
##     PAY_5 + PAY_AMT2 + PAY_AMT1 + BILL_AMT3 + MARRIAGE + SEX +
##     PAY_AMT3, family = binomial, data = credit_card_data_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1081  -0.5983  -0.5170  -0.3227   3.4414
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.953e+00  6.464e-01  -4.569 4.90e-06 ***
## PAY_0-1      3.639e-01  8.726e-02   4.170 3.05e-05 ***
## PAY_00      -3.571e-01  8.707e-02  -4.101 4.11e-05 ***
```

```
## PAY_01       8.241e-01  8.392e-02   9.820  < 2e-16 ***
## PAY_02       2.093e+00  9.599e-02  21.801  < 2e-16 ***
## PAY_03       2.124e+00  1.697e-01  12.513  < 2e-16 ***
## PAY_04       2.011e+00  3.088e-01   6.511 7.48e-11 ***
## PAY_05       1.182e+00  4.520e-01   2.615 0.008935 **
## PAY_06       6.577e-01  8.255e-01   0.797 0.425608
## PAY_07       1.436e+00  8.676e-01   1.655 0.097945 .
## PAY_08       1.356e+00  6.112e-01   2.219 0.026506 *
## LIMIT_BAL   -2.049e-06  1.807e-07 -11.341  < 2e-16 ***
## PAY_5-1     -3.339e-01  7.025e-02  -4.753 2.00e-06 ***
## PAY_50      -1.321e-01  6.525e-02  -2.025 0.042842 *
## PAY_52       6.507e-01  7.759e-02   8.386  < 2e-16 ***
## PAY_53       5.572e-01  2.135e-01   2.610 0.009059 **
## PAY_54       3.005e-01  3.185e-01   0.944 0.345400
## PAY_55      -3.291e-02  5.667e-01  -0.058 0.953693
## PAY_56       1.069e+01  1.137e+02   0.094 0.925063
## PAY_57       1.238e+00  4.308e-01   2.873 0.004061 **
## PAY_58       1.054e+01  1.970e+02   0.054 0.957332
## PAY_AMT2    -1.183e-05  2.162e-06  -5.470 4.51e-08 ***
## PAY_AMT1    -1.133e-05  2.291e-06  -4.942 7.72e-07 ***
## BILL_AMT3    2.296e-06  3.584e-07   6.407 1.49e-10 ***
## MARRIAGE1    1.945e+00  6.424e-01   3.027 0.002467 **
## MARRIAGE2    1.768e+00  6.423e-01   2.752 0.005925 **
## MARRIAGE3    2.056e+00  6.617e-01   3.107 0.001892 **
## SEX2        -1.484e-01  3.552e-02  -4.180 2.92e-05 ***
## PAY_AMT3    -7.068e-06  2.015e-06  -3.509 0.000451 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 25495  on 23999  degrees of freedom
## Residual deviance: 21093  on 23971  degrees of freedom
## AIC: 21151
##
## Number of Fisher Scoring iterations: 10
```

#————— 1. Using 80% Training Data

```
#install.packages("rpart")

library(rpart)
```

Fit a classification tree (CART) on your 80% training data using >rpart with default arguments. Please use 5:1

cost ratio throughout. That is, you will specify “loss=matrix(c(0,5,10))” as in lab 6B notes. You may use the default

cp argument without further pruning.

```
credit_rpart <- rpart(formula = default_payment_next_month ~ . , data = credi
t_card_data_train, method = "class",
                      parms = list(loss=matrix(c(0,5,1,0), nrow = 2)))


credit_train_prob = predict(credit_rpart, credit_card_data_train, type="prob"
)
```

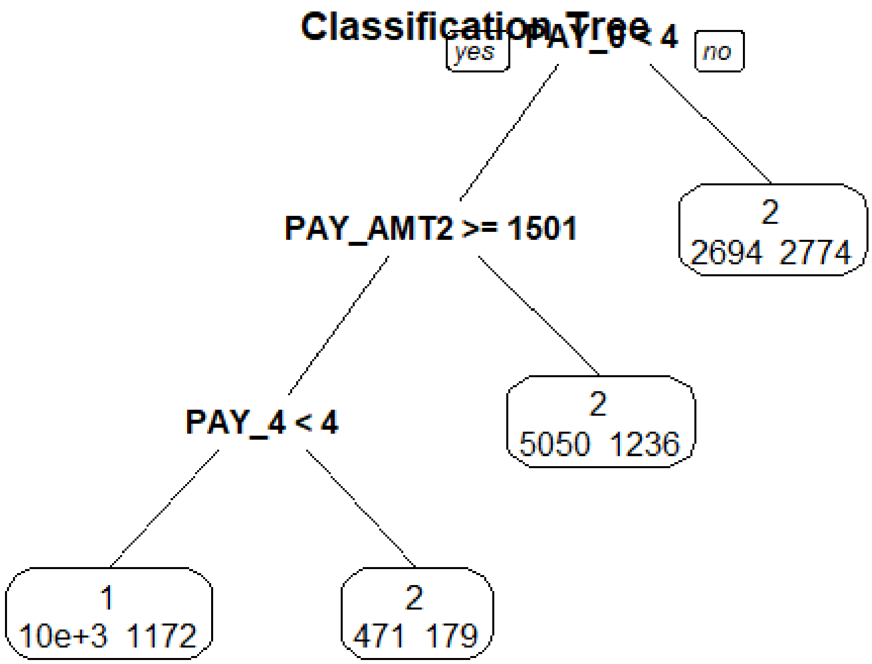
Calculating values for summary table

```
cost <- function(r, phat){
  weight1 <- 5
  weight0 <- 1
  pcut <- weight0/(weight1+weight0)
  c1 <- (r==1)&(phat<pcut) #logical vector - true if actual 1 but predict 0
  c0 <-(r==0)&(phat>pcut) #logical vector - true if actual 0 but predict 1
  return(mean(weight1*c1+weight0*c0))
}

cost(credit_card_data_train$default_payment_next_month, credit_train_prob) #
0.681

## [1] 1.085833
```

#———— 2. Plotting Classification tree

Plot your classification tree. Please give proper title and nice presentation of the tree figure output.

```
prp(credit_rpart, extra = 1) + ggplot((title('Classification Tree')))
```




```
## NULL
```

#———— 3. Interpreting main predictor variables

Please interpret some of the main predictor variables used to split the tree. Pick one terminal node and

interpret briefly of the outputs and number of observations in that particular node.

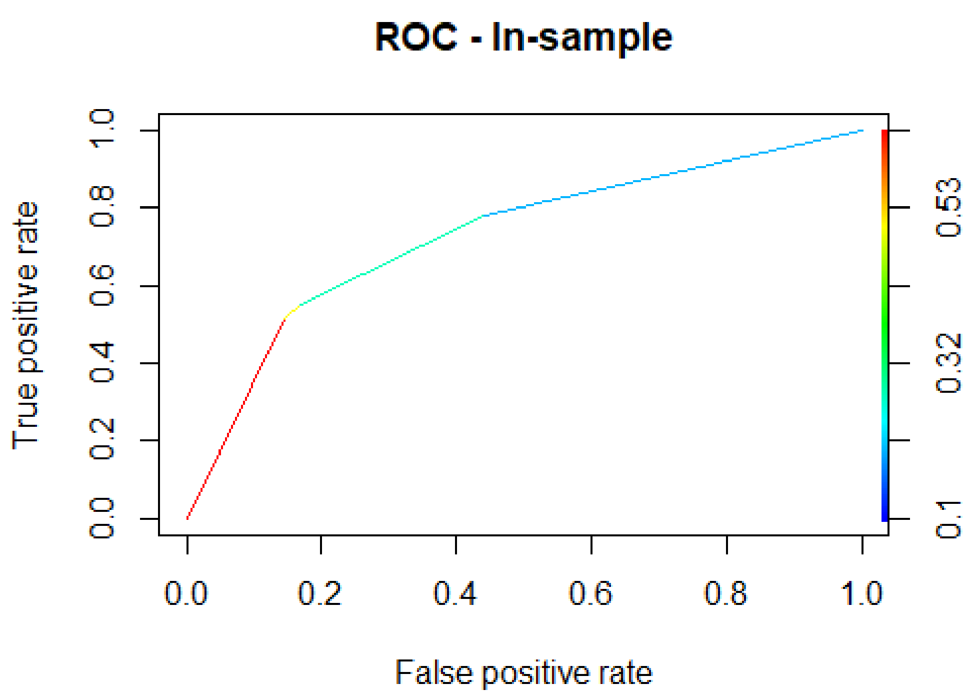
#———— 4. ROC Curve

```
#install.packages("ROCR")

#library(ROCR)
```

Draw ROC curve

```
cost <- function(r, phat){
  weight1 <- 5
  weight0 <- 1
  pcut <- weight0/(weight1+weight0)
  c1 <- (r==1)&(phat<pcut) #logical vector - true if actual 1 but predict 0
  c0 <-(r==0)&(phat>pcut) #logical vector - true if actual 0 but predict 1
  return(mean(weight1*c1+weight0*c0))
}
```

```
cost(credit_card_data_train$default_payment_next_month, predict(credit_rpart,
credit_card_data_train, type="prob")) # 0.6815

## [1] 1.085833

credit_train_pred_tree <- predict(credit_rpart, credit_card_data_train, type=
"prob")

pred = prediction(credit_train_pred_tree[,2], credit_card_data_train$default_
payment_next_month)

perf = performance(pred, "tpr", "fpr")

plot(perf, colorize=TRUE) + ggplot((title('ROC - In-sample')))
```




```
## NULL
```

#———— 5. AUC

Report the AUC. Is it “>0.7” with satisfactory discrimatory power?

```
slot(performance(pred, "auc"), "y.values")[[1]] # 0.7295294

## [1] 0.7294911
```

END