

# Rui\_Peng\_Week\_6.R

raypeng

2025-04-19

*#Step 1: Use the Decision Tree / Random Forest / Decision Tree code from Week 5 as a Starting Point*

```
library( rpart )
library( rpart.plot )
library( ROCR )
library( MASS )
library( randomForest )
```

```
## randomForest 4.7-1.2
```

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

```
library( gbm )
```

```
## Loaded gbm 2.2.2
```

```
## This version of gbm is no longer under development. Consider transitioning to gbm3, https://github.com
```

```
SEED = 1
set.seed( SEED )

TARGET = "TARGET_BAD_FLAG"

PATH = "/Users/raypeng/Documents/IS 5213 Data science and big data/HMEQ_Scrubbed"
FILE_NAME = "HMEQ_Scrubbed.csv"

INFILE = paste(PATH, FILE_NAME, sep = "/")

setwd(PATH)
df = read.csv(FILE_NAME)
str(df)
```

```
## 'data.frame': 5960 obs. of 29 variables:
## $ TARGET_BAD_FLAG : int 1 1 1 1 0 1 1 1 1 1 ...
## $ TARGET_LOSS_AMT : int 641 1109 767 1425 0 335 1841 373 1217 1523 ...
## $ LOAN : int 1100 1300 1500 1500 1700 1700 1800 1800 2000 2000 ...
## $ IMP_MORTDUE : num 25860 70053 13500 65000 97800 ...
## $ M_MORTDUE : int 0 0 0 1 0 0 0 0 0 1 ...
## $ IMP_VALUE : num 39025 68400 16700 89000 112000 ...
```

```

## $ M_VALUE      : int  0 0 0 1 0 0 0 0 0 0 ...
## $ IMP_YOJ      : num  10.5 7 4 7 3 9 5 11 3 16 ...
## $ M_YOJ       : int  0 0 0 1 0 0 0 0 0 0 ...
## $ IMP_DEROG    : int  0 0 0 1 0 0 3 0 0 0 ...
## $ M_DEROG      : int  0 0 0 1 0 0 0 0 0 0 ...
## $ IMP_DELINQ   : int  0 2 0 1 0 0 2 0 2 0 ...
## $ M_DELINQ     : int  0 0 0 1 0 0 0 0 0 0 ...
## $ IMP_CLAGE    : num  94.4 121.8 149.5 174 93.3 ...
## $ M_CLAGE      : int  0 0 0 1 0 0 0 0 0 0 ...
## $ IMP_NINQ     : int  1 0 1 1 0 1 1 0 1 0 ...
## $ M_NINQ       : int  0 0 0 1 0 0 0 0 0 0 ...
## $ IMP_CLNO     : int  9 14 10 20 14 8 17 8 12 13 ...
## $ M_CLNO       : int  0 0 0 1 0 0 0 0 0 0 ...
## $ IMP_DEBTINC  : num  35 35 35 35 35 ...
## $ M_DEBTINC    : int  1 1 1 1 1 0 1 0 1 1 ...
## $ FLAG.Job.Mgr : int  0 0 0 0 0 0 0 0 0 0 ...
## $ FLAG.Job.Office : int  0 0 0 0 1 0 0 0 0 0 ...
## $ FLAG.Job.Other : int  1 1 1 0 0 1 1 1 1 0 ...
## $ FLAG.Job.ProfExe : int  0 0 0 0 0 0 0 0 0 0 ...
## $ FLAG.Job.Sales : int  0 0 0 0 0 0 0 0 0 1 ...
## $ FLAG.Job.Self  : int  0 0 0 0 0 0 0 0 0 0 ...
## $ FLAG.Reason.DebtCon: int  0 0 0 0 0 0 0 0 0 0 ...
## $ FLAG.Reason.HomeImp: int  1 1 1 0 1 1 1 1 1 1 ...

```

```
summary(df)
```

```

## TARGET_BAD_FLAG TARGET_LOSS_AMT      LOAN      IMP_MORTDUE
## Min.   :0.0000   Min.   :    0   Min.   : 1100   Min.   : 2063
## 1st Qu.:0.0000   1st Qu.:    0   1st Qu.:11100   1st Qu.: 48139
## Median :0.0000   Median :    0   Median :16300   Median : 65000
## Mean   :0.1995   Mean   : 2676   Mean   :18608   Mean   : 72999
## 3rd Qu.:0.0000   3rd Qu.:    0   3rd Qu.:23300   3rd Qu.: 88200
## Max.   :1.0000   Max.   :78987   Max.   :89900   Max.   :399550
##      M_MORTDUE      IMP_VALUE      M_VALUE      IMP_YOJ
## Min.   :0.000000   Min.   : 8000   Min.   :0.000000   Min.   : 0.000
## 1st Qu.:0.000000   1st Qu.: 66490   1st Qu.:0.000000   1st Qu.: 3.000
## Median :0.000000   Median : 89000   Median :0.000000   Median : 7.000
## Mean   :0.08691   Mean   :101536   Mean   :0.01879   Mean   : 8.756
## 3rd Qu.:0.000000   3rd Qu.:119005   3rd Qu.:0.000000   3rd Qu.:12.000
## Max.   :1.000000   Max.   :855909   Max.   :1.000000   Max.   :41.000
##      M_YOJ      IMP_DEROG      M_DEROG      IMP_DELINQ
## Min.   :0.000000   Min.   : 0.0000   Min.   :0.0000   Min.   : 0.000
## 1st Qu.:0.000000   1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.: 0.000
## Median :0.000000   Median : 0.0000   Median :0.0000   Median : 0.000
## Mean   :0.08641   Mean   : 0.3431   Mean   :0.1188   Mean   : 0.503
## 3rd Qu.:0.000000   3rd Qu.: 0.0000   3rd Qu.:0.0000   3rd Qu.: 1.000
## Max.   :1.000000   Max.   :10.0000   Max.   :1.0000   Max.   :15.000
##      M_DELINQ      IMP_CLAGE      M_CLAGE      IMP_NINQ
## Min.   :0.000000   Min.   :    0.0   Min.   :0.000000   Min.   : 0.00
## 1st Qu.:0.000000   1st Qu.: 117.4   1st Qu.:0.000000   1st Qu.: 0.00
## Median :0.000000   Median : 174.0   Median :0.000000   Median : 1.00
## Mean   :0.09732   Mean   : 179.5   Mean   :0.05168   Mean   : 1.17
## 3rd Qu.:0.000000   3rd Qu.: 227.1   3rd Qu.:0.000000   3rd Qu.: 2.00
## Max.   :1.000000   Max.   :1168.2   Max.   :1.000000   Max.   :17.00

```

```
##      M_NINQ      IMP_CLNO      M_CLNO      IMP_DEBTINC
## Min.   :0.00000  Min.   : 0.00  Min.   :0.00000  Min.   : 0.5245
## 1st Qu.:0.00000  1st Qu.:15.00  1st Qu.:0.00000  1st Qu.: 30.7632
## Median :0.00000  Median :20.00  Median :0.00000  Median : 35.0000
## Mean   :0.08557  Mean   :21.25  Mean   :0.03725  Mean   : 34.0393
## 3rd Qu.:0.00000  3rd Qu.:26.00  3rd Qu.:0.00000  3rd Qu.: 37.9499
## Max.   :1.00000  Max.   :71.00  Max.   :1.00000  Max.   :203.3122
##      M_DEBTINC      FLAG.Job.Mgr      FLAG.Job.Office      FLAG.Job.Other
## Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000
## 1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000
## Median :0.0000  Median :0.0000  Median :0.0000  Median :0.0000
## Mean   :0.2126  Mean   :0.1287  Mean   :0.1591  Mean   :0.4007
## 3rd Qu.:0.0000  3rd Qu.:0.0000  3rd Qu.:0.0000  3rd Qu.:1.0000
## Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000
## FLAG.Job.ProfExe FLAG.Job.Sales      FLAG.Job.Self      FLAG.Reason.DebtCon
## Min.   :0.0000  Min.   :0.00000  Min.   :0.00000  Min.   :0.0000
## 1st Qu.:0.0000  1st Qu.:0.00000  1st Qu.:0.00000  1st Qu.:0.0000
## Median :0.0000  Median :0.00000  Median :0.00000  Median :1.0000
## Mean   :0.2141  Mean   :0.01829  Mean   :0.03238  Mean   :0.6591
## 3rd Qu.:0.0000  3rd Qu.:0.00000  3rd Qu.:0.00000  3rd Qu.:1.0000
## Max.   :1.0000  Max.   :1.00000  Max.   :1.00000  Max.   :1.0000
## FLAG.Reason.HomeImp
## Min.   :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean   :0.2987
## 3rd Qu.:1.0000
## Max.   :1.0000
```

```
head(df)
```

```
##      TARGET_BAD_FLAG TARGET_LOSS_AMT LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE
## 1          1          641 1100      25860      0      39025      0
## 2          1         1109 1300      70053      0      68400      0
## 3          1          767 1500      13500      0      16700      0
## 4          1         1425 1500      65000      1      89000      1
## 5          0           0 1700      97800      0     112000      0
## 6          1          335 1700      30548      0      40320      0
##      IMP_YOJ M_YOJ IMP_DEROG M_DEROG IMP_DELINQ M_DELINQ IMP_CLAGE M_CLAGE
## 1     10.5    0          0          0          0          0 94.36667      0
## 2      7.0    0          0          0          2          0 121.83333      0
## 3      4.0    0          0          0          0          0 149.46667      0
## 4      7.0    1          1          1          1          1 174.00000      1
## 5      3.0    0          0          0          0          0 93.33333      0
## 6      9.0    0          0          0          0          0 101.46600      0
##      IMP_NINQ M_NINQ IMP_CLNO M_CLNO IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr
## 1          1      0          9          0      35.00000      1          0
## 2          0      0         14          0      35.00000      1          0
## 3          1      0         10          0      35.00000      1          0
## 4          1      1         20          1      35.00000      1          0
## 5          0      0         14          0      35.00000      1          0
## 6          1      0          8          0      37.11361      0          0
##      FLAG.Job.Office FLAG.Job.Other FLAG.Job.ProfExe FLAG.Job.Sales FLAG.Job.Self
## 1          0          1          0          0          0
```

```
## 2      0      1      0      0      0
## 3      0      1      0      0      0
## 4      0      0      0      0      0
## 5      1      0      0      0      0
## 6      0      1      0      0      0
##  FLAG.Reason.DebtCon FLAG.Reason.HomeImp
## 1      0      1
## 2      0      1
## 3      0      1
## 4      0      0
## 5      0      1
## 6      0      1
```

```
df_flag = df
df_flag$TARGET_LOSS_AMT = NULL

FLAG = sample( c(TRUE, FALSE), nrow(df_flag), replace = TRUE,
               prob = c(0.7,0.3) )
df_train = df_flag[FLAG, ]
df_test = df_flag[!FLAG, ]

dim(df_flag)
```

```
## [1] 5960 28
```

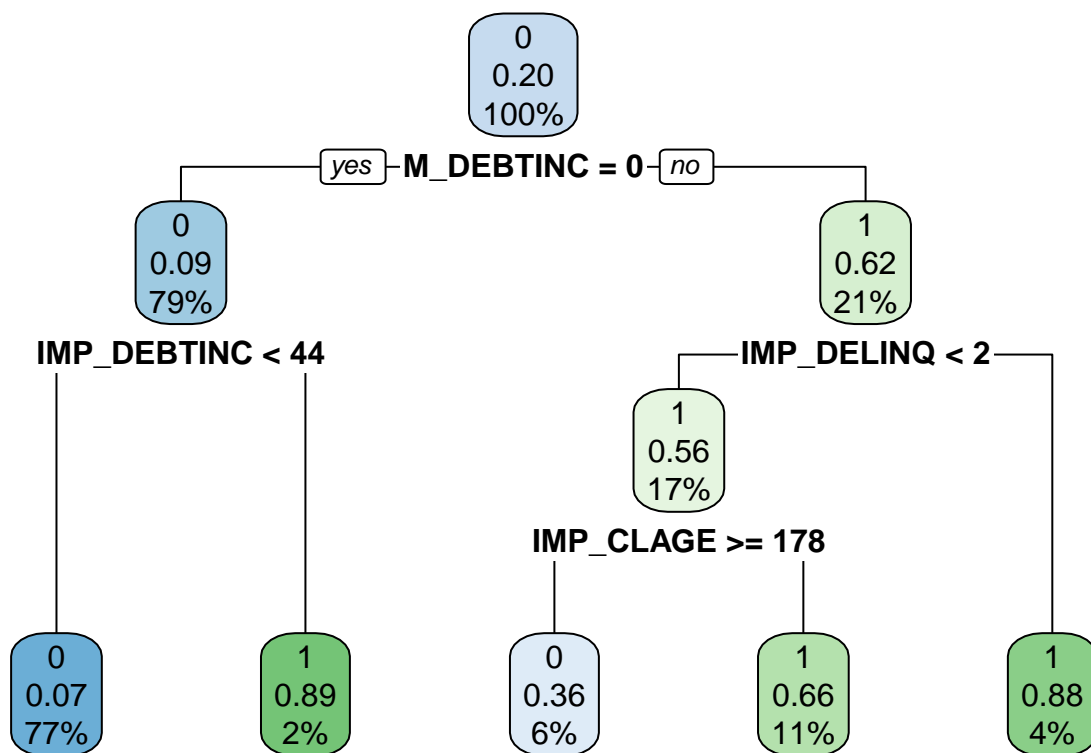
```
dim(df_train)
```

```
## [1] 4142 28
```

```
dim(df_test)
```

```
## [1] 1818 28
```

```
#Decision Tree Model
tr_set = rpart.control( maxdepth = 10 )
tr_model = rpart( data = df_train, TARGET_BAD_FLAG ~ .,
                  control = tr_set, method = "class", parms = list(split = 'information'))
rpart.plot( tr_model )
```



```
tr_model$variable.importance
```

```
##      M_DEBTINC IMP_DEBTINC  IMP_DELINQ  IMP_CLAGE      LOAN      M_VALUE
## 533.397481  134.588883  46.494397  30.749923  24.521888  22.199895
##      IMP_VALUE IMP_MORTDUE   IMP_CLNO   IMP_YOJ
##    7.967967    5.783975    2.459994    2.090995
```

```
pt = predict( tr_model, df_test, type = "prob" )
head( pt )
```

```
##           0           1
## 4  0.3354839 0.66451613
## 6  0.9315112 0.06848885
## 7  0.1206897 0.87931034
## 15 0.3354839 0.66451613
## 17 0.1206897 0.87931034
## 18 0.9315112 0.06848885
```

```
pt2 = prediction( pt[,2], df_test$TARGET_BAD_FLAG )
pt3 = performance( pt2, "tpr", "fpr" )
```

```
#Random Forest Model
```

```
rf_model = randomForest( data = df_train, TARGET_BAD_FLAG ~ .,
                          ntree = 100, importance = TRUE )
```

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

```
pr = predict( rf_model, df_test )
head( pr )
```

```
##          4          6          7          15          17          18
## 0.7853333 0.8671667 0.9460000 0.8203333 0.9213333 0.3466667
```

```
pr2 = prediction( pr, df_test$TARGET_BAD_FLAG )
pr3 = performance( pr2, "tpr", "fpr" )
```

```
#Gradient Boosting Model
```

```
gb_model = gbm( data = df_train, TARGET_BAD_FLAG ~ ., n.trees = 100,
               distribution = "bernoulli" )
```

```
pg = predict( gb_model, df_test, type = "response" )
```

```
## Using 100 trees...
```

```
head(pg)
```

```
## [1] 0.8970211 0.1571409 0.9489942 0.6561628 0.9719267 0.5539409
```

```
pg2 = prediction( pg, df_test$TARGET_BAD_FLAG )
pg3 = performance( pg2, "tpr", "fpr" )
```

```
#Step 2: Classification Models
```

```
#Logistic Reg All and Backward
```

```
theUpper_LR = glm( TARGET_BAD_FLAG ~ ., family = "binomial", data = df_train )
```

```
lr_model = stepAIC( theUpper_LR, direction = "backward" )
```

```
## Start: AIC=2304.21
```

```
## TARGET_BAD_FLAG ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
## M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
## M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO +
## M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
## FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
## FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
```

```
##
##           Df Deviance    AIC
## - FLAG.Reason.HomeImp  1    2248.2 2302.2
## - FLAG.Reason.DebtCon  1    2248.3 2302.3
## - M_NINQ                1    2248.5 2302.5
## <none>                  2248.2 2304.2
## - M_MORTDUE             1    2250.3 2304.3
## - LOAN                  1    2250.8 2304.8
## - IMP_MORTDUE           1    2251.2 2305.2
## - M_CLAGE               1    2253.0 2307.0
## - M_DELINQ              1    2253.7 2307.7
## - IMP_YOJ                1    2254.2 2308.2
```

```

## - IMP_CLNO          1    2254.7 2308.7
## - FLAG.Job.Office   1    2257.4 2311.4
## - IMP_VALUE         1    2258.2 2312.2
## - M_YOJ             1    2259.3 2313.3
## - M_CLNO           1    2261.9 2315.9
## - FLAG.Job.ProfExe  1    2267.8 2321.8
## - FLAG.Job.Mgr      1    2270.2 2324.2
## - FLAG.Job.Self     1    2271.0 2325.0
## - IMP_NINQ          1    2271.1 2325.1
## - FLAG.Job.Other    1    2272.4 2326.4
## - FLAG.Job.Sales    1    2284.4 2338.4
## - IMP_CLAGE         1    2298.9 2352.9
## - IMP_DEROG         1    2317.7 2371.7
## - M_DEROG           1    2320.9 2374.9
## - IMP_DEBTINC       1    2356.6 2410.6
## - M_VALUE           1    2359.2 2413.2
## - IMP_DELINQ        1    2483.2 2537.2
## - M_DEBTINC         1    2866.2 2920.2
##
## Step:  AIC=2302.24
## TARGET_BAD_FLAG ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO +
##     M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##     FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.DebtCon
##
##              Df Deviance    AIC
## - M_NINQ          1    2248.5 2300.5
## - FLAG.Reason.DebtCon 1    2248.6 2300.6
## <none>              1    2248.2 2302.2
## - M_MORTDUE       1    2250.3 2302.3
## - LOAN            1    2250.8 2302.8
## - IMP_MORTDUE     1    2251.2 2303.2
## - M_CLAGE         1    2253.0 2305.0
## - M_DELINQ        1    2253.7 2305.7
## - IMP_YOJ         1    2254.3 2306.3
## - IMP_CLNO        1    2254.7 2306.7
## - FLAG.Job.Office  1    2257.8 2309.8
## - IMP_VALUE       1    2258.2 2310.2
## - M_YOJ           1    2259.3 2311.3
## - M_CLNO          1    2261.9 2313.9
## - FLAG.Job.ProfExe 1    2268.8 2320.8
## - FLAG.Job.Mgr     1    2271.1 2323.1
## - IMP_NINQ        1    2271.2 2323.2
## - FLAG.Job.Self    1    2271.8 2323.8
## - FLAG.Job.Other   1    2273.9 2325.9
## - FLAG.Job.Sales   1    2285.5 2337.5
## - IMP_CLAGE       1    2299.0 2351.0
## - IMP_DEROG       1    2317.7 2369.7
## - M_DEROG         1    2320.9 2372.9
## - IMP_DEBTINC     1    2356.6 2408.6
## - M_VALUE         1    2360.5 2412.5
## - IMP_DELINQ      1    2483.8 2535.8

```

```

## - M_DEBTINC          1    2866.3 2918.3
##
## Step:   AIC=2300.54
## TARGET_BAD_FLAG ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##      M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##      M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO +
##      IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##      FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##      FLAG.Reason.DebtCon
##
##              Df Deviance    AIC
## - FLAG.Reason.DebtCon  1    2248.9 2298.9
## <none>                  2248.5 2300.5
## - M_MORTDUE            1    2250.6 2300.6
## - LOAN                 1    2251.0 2301.0
## - IMP_MORTDUE          1    2251.5 2301.5
## - M_CLAGE              1    2253.4 2303.4
## - IMP_YOJ              1    2254.6 2304.6
## - IMP_CLNO             1    2255.0 2305.0
## - FLAG.Job.Office      1    2258.1 2308.1
## - M_DELINQ             1    2258.1 2308.1
## - IMP_VALUE            1    2258.4 2308.4
## - M_YOJ                1    2260.5 2310.5
## - M_CLNO               1    2262.3 2312.3
## - FLAG.Job.ProfExe     1    2269.1 2319.1
## - FLAG.Job.Mgr         1    2271.4 2321.4
## - IMP_NINQ             1    2271.8 2321.8
## - FLAG.Job.Self        1    2271.8 2321.8
## - FLAG.Job.Other       1    2274.2 2324.2
## - FLAG.Job.Sales       1    2286.2 2336.2
## - IMP_CLAGE            1    2299.2 2349.2
## - IMP_DEROG            1    2317.7 2367.7
## - M_DEROG              1    2321.2 2371.2
## - IMP_DEBTINC          1    2359.1 2409.1
## - M_VALUE              1    2362.1 2412.1
## - IMP_DELINQ           1    2483.8 2533.8
## - M_DEBTINC            1    2868.3 2918.3
##
## Step:   AIC=2298.92
## TARGET_BAD_FLAG ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##      M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##      M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO +
##      IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##      FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self
##
##              Df Deviance    AIC
## <none>                  2248.9 2298.9
## - M_MORTDUE            1    2251.4 2299.4
## - LOAN                 1    2252.0 2300.0
## - IMP_MORTDUE          1    2252.0 2300.0
## - M_CLAGE              1    2253.6 2301.6
## - IMP_YOJ              1    2254.9 2302.9
## - IMP_CLNO             1    2255.9 2303.9
## - FLAG.Job.Office      1    2258.4 2306.4

```



```
## - M_DELINQ      1  2258.4 2306.4
## - IMP_VALUE     1  2259.2 2307.2
## - M_YOJ         1  2261.2 2309.2
## - M_CLNO        1  2263.1 2311.1
## - FLAG.Job.ProfExe 1  2269.5 2317.5
## - FLAG.Job.Mgr   1  2271.6 2319.6
## - IMP_NINQ      1  2271.8 2319.8
## - FLAG.Job.Self  1  2272.6 2320.6
## - FLAG.Job.Other 1  2274.5 2322.5
## - FLAG.Job.Sales 1  2286.2 2334.2
## - IMP_CLAGE     1  2299.3 2347.3
## - IMP_DEROG     1  2318.3 2366.3
## - M_DEROG       1  2322.1 2370.1
## - IMP_DEBTINC   1  2360.8 2408.8
## - M_VALUE       1  2363.3 2411.3
## - IMP_DELINQ    1  2484.9 2532.9
## - M_DEBTINC     1  2869.4 2917.4
```

```
pla = predict( theUpper_LR, df_test, type = "response" )
pla2 = prediction( pla, df_test$TARGET_BAD_FLAG )
pla3 = performance( pla2, "tpr", "fpr" )

plr = predict( lr_model, df_test, type = "response" )
plr2 = prediction( plr, df_test$TARGET_BAD_FLAG )
plr3 = performance( plr2, "tpr", "fpr" )

#LR forward step tree
treeVars = tr_model$variable.importance
treeVars = names(treeVars)
treeVarsPlus = paste( treeVars, collapse = "+" )
F = as.formula( paste( "TARGET_BAD_FLAG ~", treeVarsPlus ) )

tree_LR = glm( F, family = "binomial", data = df_train )
theLower_LR = glm( TARGET_BAD_FLAG ~ 1, family = "binomial", data = df_train )
#summary( tree_LR )
#summary( theLower_LR )

lrt_model = stepAIC( theLower_LR, direction = "forward",
                     scope = list(lower = theLower_LR, upper = tree_LR ) )
```

```
## Start: AIC=4157.3
## TARGET_BAD_FLAG ~ 1
##
##           Df Deviance   AIC
## + M_DEBTINC  1  3088.5 3092.5
## + IMP_DELINQ  1  3776.4 3780.4
## + M_VALUE    1  3942.8 3946.8
## + IMP_CLAGE  1  4020.9 4024.9
## + IMP_DEBTINC 1  4051.1 4055.1
## + LOAN       1  4133.3 4137.3
## + IMP_YOJ     1  4144.0 4148.0
## + IMP_MORTDUE 1  4150.4 4154.4
## + IMP_VALUE   1  4153.1 4157.1
```

```

## <none>          4155.3 4157.3
## + IMP_CLNO      1    4155.3 4159.3
##
## Step:  AIC=3092.5
## TARGET_BAD_FLAG ~ M_DEBTINC
##
##           Df Deviance    AIC
## + IMP_DELINQ  1    2861.2 2867.2
## + M_VALUE     1    2972.0 2978.0
## + IMP_DEBTINC 1    2977.4 2983.4
## + IMP_CLAGE   1    3006.3 3012.3
## + IMP_YOJ     1    3075.2 3081.2
## <none>        3088.5 3092.5
## + LOAN        1    3086.7 3092.7
## + IMP_CLNO    1    3088.2 3094.2
## + IMP_VALUE   1    3088.3 3094.3
## + IMP_MORTDUE 1    3088.5 3094.5
##
## Step:  AIC=2867.17
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ
##
##           Df Deviance    AIC
## + IMP_CLAGE   1    2751.2 2759.2
## + IMP_DEBTINC 1    2753.8 2761.8
## + M_VALUE     1    2778.9 2786.9
## + IMP_YOJ     1    2836.1 2844.1
## + IMP_CLNO    1    2850.8 2858.8
## <none>        2861.2 2867.2
## + LOAN        1    2859.8 2867.8
## + IMP_VALUE   1    2860.8 2868.8
## + IMP_MORTDUE 1    2861.1 2869.1
##
## Step:  AIC=2759.17
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE
##
##           Df Deviance    AIC
## + IMP_DEBTINC 1    2650.4 2660.4
## + M_VALUE     1    2672.4 2682.4
## + IMP_YOJ     1    2741.4 2751.4
## + IMP_VALUE   1    2743.9 2753.9
## <none>        2751.2 2759.2
## + IMP_MORTDUE 1    2749.2 2759.2
## + IMP_CLNO    1    2751.2 2761.2
## + LOAN        1    2751.2 2761.2
##
## Step:  AIC=2660.38
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC
##
##           Df Deviance    AIC
## + M_VALUE     1    2554.2 2566.2
## + IMP_YOJ     1    2644.2 2656.2
## + IMP_VALUE   1    2646.5 2658.5
## + IMP_CLNO    1    2648.2 2660.2
## <none>        2650.4 2660.4

```

```

## + LOAN          1    2650.0 2662.0
## + IMP_MORTDUE   1    2650.1 2662.1
##
## Step:  AIC=2566.19
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE
##
##           Df Deviance    AIC
## + IMP_YOJ      1    2547.7 2561.7
## + IMP_VALUE    1    2549.6 2563.6
## <none>          2554.2 2566.2
## + IMP_CLNO     1    2552.5 2566.5
## + LOAN         1    2553.3 2567.3
## + IMP_MORTDUE  1    2554.2 2568.2
##
## Step:  AIC=2561.67
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + IMP_YOJ
##
##           Df Deviance    AIC
## + IMP_VALUE    1    2542.8 2558.8
## <none>          2547.7 2561.7
## + IMP_CLNO     1    2545.9 2561.9
## + LOAN         1    2547.2 2563.2
## + IMP_MORTDUE  1    2547.7 2563.7
##
## Step:  AIC=2558.79
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + IMP_YOJ + IMP_VALUE
##
##           Df Deviance    AIC
## + IMP_MORTDUE  1    2538.0 2556.0
## + IMP_CLNO     1    2539.4 2557.4
## + LOAN         1    2540.5 2558.5
## <none>          2542.8 2558.8
##
## Step:  AIC=2556.04
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + IMP_YOJ + IMP_VALUE + IMP_MORTDUE
##
##           Df Deviance    AIC
## + LOAN         1    2535.7 2555.7
## + IMP_CLNO     1    2536.0 2556.0
## <none>          2538.0 2556.0
##
## Step:  AIC=2555.73
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + IMP_YOJ + IMP_VALUE + IMP_MORTDUE + LOAN
##
##           Df Deviance    AIC
## <none>          2535.7 2555.7
## + IMP_CLNO     1    2533.8 2555.8

```

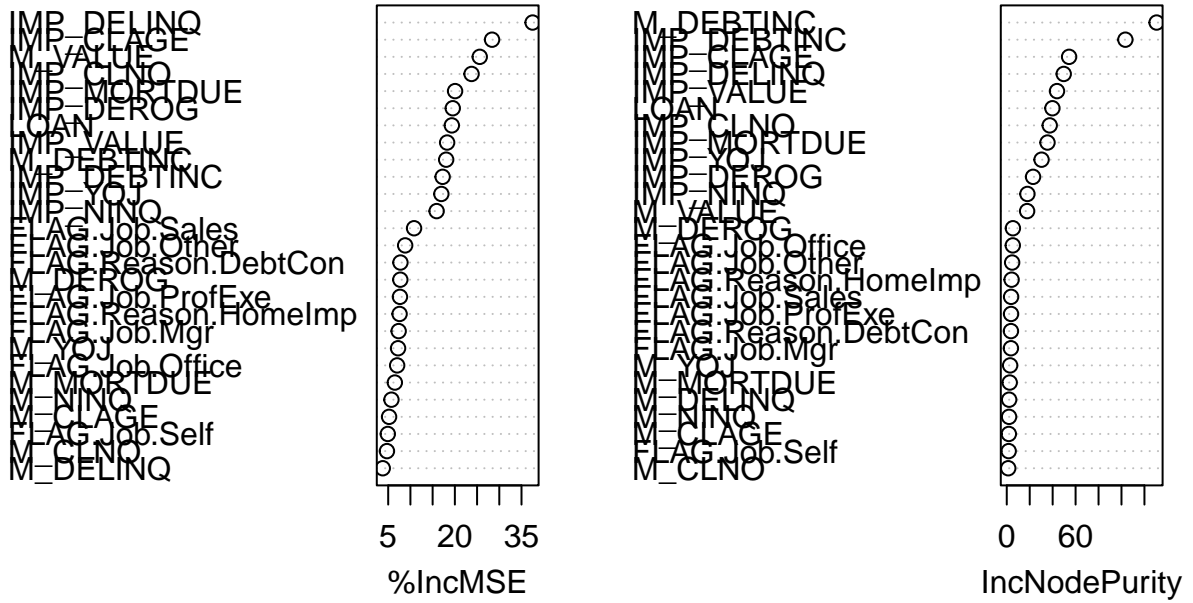
```
plrt = predict( lrt_model, df_test, type = "response" )
plrt2 = prediction( plrt, df_test$TARGET_BAD_FLAG )
plrt3 = performance( plrt2, "tpr", "fpr" )
```

```
#Compare and list the important variables of RF/GB/LOGIT
importance( rf_model )
```

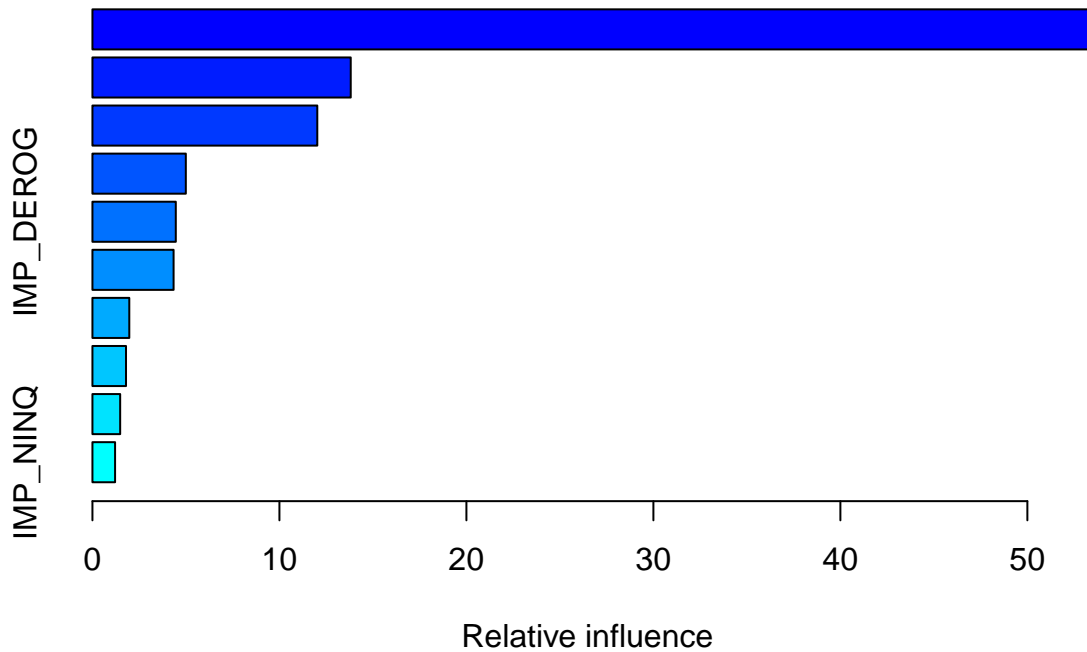
##	%IncMSE	IncNodePurity
## LOAN	19.316810	39.777176
## IMP_MORTDUE	20.052341	35.518005
## M_MORTDUE	6.499260	2.582903
## IMP_VALUE	18.264039	43.787225
## M_VALUE	25.602248	17.743863
## IMP_YOJ	16.963341	30.322241
## M_YOJ	7.220923	2.943678
## IMP_DEROG	19.530010	22.914985
## M_DEROG	7.729186	5.338867
## IMP_DELINQ	37.494775	49.645744
## M_DELINQ	3.780398	2.083958
## IMP_CLAGE	28.371343	54.329981
## M_CLAGE	5.201136	1.794735
## IMP_NINQ	15.901124	17.995586
## M_NINQ	5.727939	2.007916
## IMP_CLNO	23.760935	37.449920
## M_CLNO	4.719829	1.292987
## IMP_DEBTINC	17.246641	103.381548
## M_DEBTINC	18.045350	130.614728
## FLAG.Job.Mgr	7.353261	3.612471
## FLAG.Job.Office	7.020243	5.254487
## FLAG.Job.Other	8.815549	4.689699
## FLAG.Job.ProfExe	7.649276	3.742588
## FLAG.Job.Sales	10.823283	3.779676
## FLAG.Job.Self	4.910753	1.675588
## FLAG.Reason.DebtCon	7.756487	3.643055
## FLAG.Reason.HomeImp	7.565180	4.219131

```
varImpPlot( rf_model )
```

## rf\_model



```
summary.gbm( gb_model, cBars = 10 )
```



```
##          var    rel.inf
## M_DEBTINC      M_DEBTINC 53.4761627
## IMP_DEBTINC    IMP_DEBTINC 13.8130176
## IMP_DELINQ     IMP_DELINQ 12.0257247
## M_VALUE        M_VALUE   4.9940416
## IMP_DEROG      IMP_DEROG  4.4580798
## IMP_CLAGE      IMP_CLAGE  4.3403733
## IMP_VALUE      IMP_VALUE  1.9703911
## LOAN           LOAN      1.7950994
## IMP_CLNO       IMP_CLNO   1.4838466
## IMP_NINQ       IMP_NINQ   1.2106400
## FLAG.Job.Sales FLAG.Job.Sales 0.2176216
## M_DEROG        M_DEROG    0.2150017
## IMP_MORTDUE    IMP_MORTDUE 0.0000000
## M_MORTDUE      M_MORTDUE  0.0000000
## IMP_YOJ        IMP_YOJ    0.0000000
## M_YOJ          M_YOJ      0.0000000
## M_DELINQ       M_DELINQ   0.0000000
## M_CLAGE        M_CLAGE    0.0000000
## M_NINQ         M_NINQ     0.0000000
## M_CLNO         M_CLNO     0.0000000
## FLAG.Job.Mgr   FLAG.Job.Mgr 0.0000000
## FLAG.Job.Office FLAG.Job.Office 0.0000000
## FLAG.Job.Other  FLAG.Job.Other 0.0000000
## FLAG.Job.ProfExe FLAG.Job.ProfExe 0.0000000
## FLAG.Job.Self   FLAG.Job.Self 0.0000000
```

```
## FLAG.Reason.DebtCon FLAG.Reason.DebtCon 0.0000000
## FLAG.Reason.HomeImp FLAG.Reason.HomeImp 0.0000000
```

```
summary( theUpper_LR )
```

```
##
## Call:
## glm(formula = TARGET_BAD_FLAG ~ ., family = "binomial", data = df_train)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.817e+00  6.558e-01 -10.395 < 2e-16 ***
## LOAN        -8.996e-06  5.694e-06  -1.580 0.114122
## IMP_MORTDUE  -3.482e-06  2.092e-06  -1.664 0.096053 .
## M_MORTDUE     3.647e-01  2.491e-01   1.464 0.143231
## IMP_VALUE     4.356e-06  1.495e-06   2.914 0.003572 **
## M_VALUE       4.945e+00  6.167e-01   8.017 1.08e-15 ***
## IMP_YOJ      -2.010e-02  8.271e-03  -2.430 0.015103 *
## M_YOJ        -7.720e-01  2.427e-01  -3.181 0.001466 **
## IMP_DEROG     5.869e-01  7.617e-02   7.706 1.30e-14 ***
## M_DEROG      -2.655e+00  3.603e-01  -7.367 1.74e-13 ***
## IMP_DELINQ     8.182e-01  6.264e-02  13.062 < 2e-16 ***
## M_DELINQ     -1.101e+00  4.859e-01  -2.266 0.023465 *
## IMP_CLAGE     -5.466e-03  8.103e-04  -6.746 1.52e-11 ***
## M_CLAGE       9.834e-01  4.403e-01   2.233 0.025529 *
## IMP_NINQ       1.528e-01  3.175e-02   4.813 1.48e-06 ***
## M_NINQ        -2.494e-01  4.538e-01  -0.550 0.582542
## IMP_CLNO      -1.600e-02  6.331e-03  -2.527 0.011513 *
## M_CLNO        2.693e+00  7.552e-01   3.565 0.000363 ***
## IMP_DEBTINC    9.231e-02  1.017e-02   9.075 < 2e-16 ***
## M_DEBTINC     2.679e+00  1.144e-01  23.425 < 2e-16 ***
## FLAG.Job.Mgr   2.091e+00  4.989e-01   4.191 2.77e-05 ***
## FLAG.Job.Office 1.414e+00  5.005e-01   2.824 0.004736 **
## FLAG.Job.Other  2.081e+00  4.825e-01   4.314 1.60e-05 ***
## FLAG.Job.ProfExe 1.972e+00  4.965e-01   3.971 7.16e-05 ***
## FLAG.Job.Sales  3.240e+00  5.794e-01   5.592 2.24e-08 ***
## FLAG.Job.Self   2.518e+00  5.650e-01   4.456 8.34e-06 ***
## FLAG.Reason.DebtCon -1.318e-01  3.661e-01  -0.360 0.718960
## FLAG.Reason.HomeImp -5.659e-02  3.728e-01  -0.152 0.879329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 4155.3  on 4141  degrees of freedom
## Residual deviance: 2248.2  on 4114  degrees of freedom
## AIC: 2304.2
##
## Number of Fisher Scoring iterations: 6
```

```
summary( lr_model )
```

```
##
```

```
## Call:
## glm(formula = TARGET_BAD_FLAG ~ LOAN + IMP_MORTDUE + M_MORTDUE +
##     IMP_VALUE + M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG +
##     IMP_DELINQ + M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ +
##     IMP_CLNO + M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr +
##     FLAG.Job.Office + FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales +
##     FLAG.Job.Self, family = "binomial", data = df_train)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -6.908e+00  6.124e-01 -11.280  < 2e-16 ***
## LOAN          -9.617e-06  5.566e-06  -1.728  0.084025 .
## IMP_MORTDUE    -3.545e-06  2.090e-06  -1.697  0.089768 .
## M_MORTDUE      3.857e-01  2.451e-01   1.574  0.115585
## IMP_VALUE      4.405e-06  1.486e-06   2.964  0.003041 **
## M_VALUE        4.999e+00  6.115e-01   8.176  2.93e-16 ***
## IMP_YOJ        -2.005e-02  8.265e-03  -2.426  0.015280 *
## M_YOJ          -7.974e-01  2.385e-01  -3.343  0.000829 ***
## IMP_DEROG      5.815e-01  7.534e-02   7.718  1.19e-14 ***
## M_DEROG       -2.654e+00  3.589e-01  -7.397  1.40e-13 ***
## IMP_DELINQ     8.185e-01  6.249e-02  13.098  < 2e-16 ***
## M_DELINQ      -1.227e+00  4.247e-01  -2.889  0.003870 **
## IMP_CLAGE      -5.421e-03  8.065e-04  -6.722  1.79e-11 ***
## M_CLAGE        9.770e-01  4.397e-01   2.222  0.026279 *
## IMP_NINQ       1.520e-01  3.160e-02   4.811  1.50e-06 ***
## IMP_CLNO      -1.646e-02  6.273e-03  -2.624  0.008697 **
## M_CLNO         2.599e+00  7.180e-01   3.620  0.000295 ***
## IMP_DEBTINC     9.295e-02  1.013e-02   9.179  < 2e-16 ***
## M_DEBTINC      2.682e+00  1.143e-01  23.456  < 2e-16 ***
## FLAG.Job.Mgr    2.061e+00  4.828e-01   4.269  1.97e-05 ***
## FLAG.Job.Office 1.387e+00  4.841e-01   2.864  0.004178 **
## FLAG.Job.Other  2.053e+00  4.642e-01   4.424  9.71e-06 ***
## FLAG.Job.ProfExe 1.947e+00  4.793e-01   4.061  4.89e-05 ***
## FLAG.Job.Sales  3.206e+00  5.627e-01   5.698  1.21e-08 ***
## FLAG.Job.Self   2.495e+00  5.469e-01   4.562  5.06e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 4155.3  on 4141  degrees of freedom
## Residual deviance: 2248.9  on 4117  degrees of freedom
## AIC: 2298.9
##
## Number of Fisher Scoring iterations: 6
```

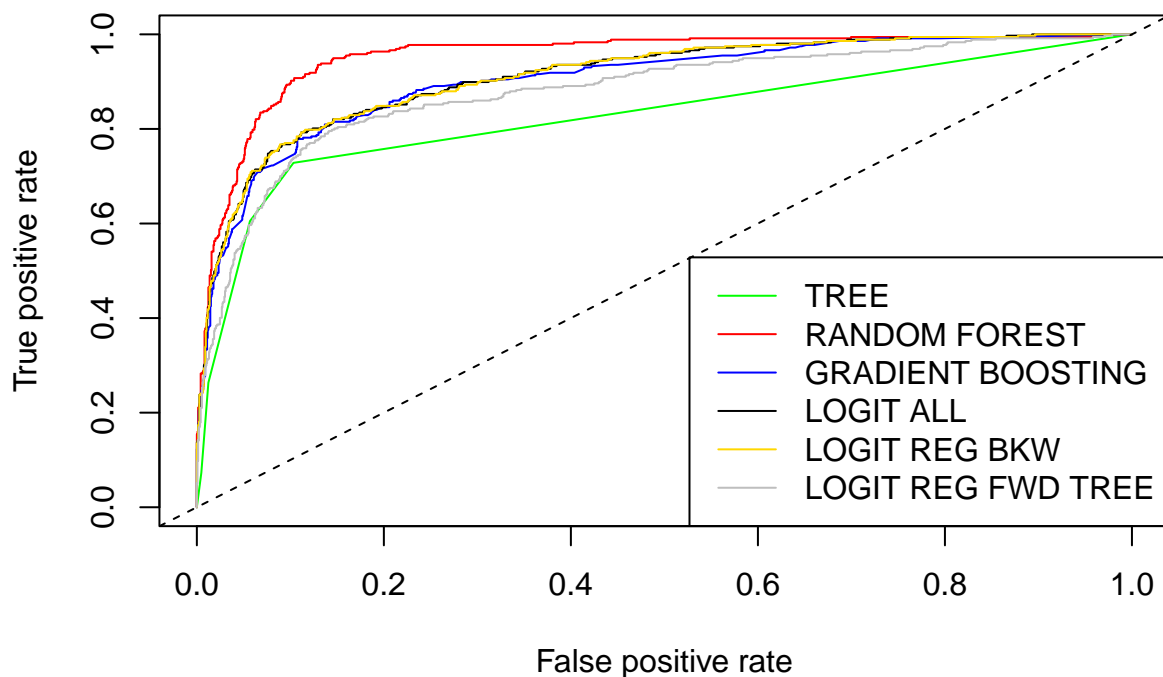
```
summary( lrt_model )
```

```
##
## Call:
## glm(formula = TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE +
##     IMP_DEBTINC + M_VALUE + IMP_YOJ + IMP_VALUE + IMP_MORTDUE +
##     LOAN, family = "binomial", data = df_train)
##
```



```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.723e+00  3.904e-01 -12.097 < 2e-16 ***
## M_DEBTINC    2.777e+00  1.064e-01  26.096 < 2e-16 ***
## IMP_DELINQ    6.697e-01  5.095e-02  13.145 < 2e-16 ***
## IMP_CLAGE     -6.506e-03  7.373e-04  -8.823 < 2e-16 ***
## IMP_DEBTINC    8.820e-02  9.577e-03   9.210 < 2e-16 ***
## M_VALUE       3.897e+00  4.959e-01   7.858 3.91e-15 ***
## IMP_YOJ       -2.093e-02  7.882e-03  -2.655 0.00793 **
## IMP_VALUE      4.432e-06  1.370e-06   3.236 0.00121 **
## IMP_MORTDUE   -4.005e-06  1.890e-06  -2.119 0.03408 *
## LOAN          -7.589e-06  5.054e-06  -1.502 0.13323
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 4155.3  on 4141  degrees of freedom
## Residual deviance: 2535.7  on 4132  degrees of freedom
## AIC: 2555.7
##
## Number of Fisher Scoring iterations: 6
```

```
#Plot the ROC curve
plot(pt3, col = "green")
plot( pr3, col = "red", add = TRUE )
plot( pg3, col = "blue", add = TRUE )
plot( pla3, col = "black", add = TRUE )
plot( plr3, col = "gold", add = TRUE )
plot( plrt3, col = "gray", add = TRUE )
abline(0,1,lty = 2)
legend( "bottomright", c("TREE", "RANDOM FOREST", "GRADIENT BOOSTING", "LOGIT ALL",
                        "LOGIT REG BKW", "LOGIT REG FWD TREE"),
      col = c("green", "red", "blue","black", "gold", "gray"),
      bty = "y", lty = 1)
```



```
aucT = performance( pt2, "auc" )@y.values
aucR = performance( pr2, "auc" )@y.values
aucG = performance( pg2, "auc" )@y.values
aucLRA = performance( pla2, "auc" )@y.values
aucLRB = performance( plr2, "auc" )@y.values
aucLRT = performance( plrt2, "auc" )@y.values

print( paste("TREE AUC = ", aucT ))
```

```
## [1] "TREE AUC = 0.826618121581281"
```

```
print( paste("RF AUC = ", aucR ))
```

```
## [1] "RF AUC = 0.95370673936926"
```

```
print( paste("GB AUC = ", aucG ))
```

```
## [1] "GB AUC = 0.903390103474655"
```

```
print( paste("LR ALL AUC = ", aucLRA ))
```

```
## [1] "LR ALL AUC = 0.910306627784587"
```

```
print( paste("LR BKW AUC = ", aucLRB ))
```

```
## [1] "LR BKW AUC = 0.910356476608441"
```

```
print( paste("LRT FWD AUC = ", aucLRT ))
```

```
## [1] "LRT FWD AUC = 0.879705201724766"
```

```
#Random Forest model performs best with AUC > 0.95.
```

```
#Step 3: Linear Regression
```

```
#Linear Regression Decision Tree of last week
```

```
df_amt = df
```

```
df_amt$TARGET_BAD_FLAG = NULL
```

```
FLAG = sample( c( TRUE, FALSE ), nrow(df_amt),  
               replace = TRUE, prob = c(0.7,0.3) )
```

```
df_train = df_amt[FLAG, ]
```

```
df_test = df_amt[!FLAG, ]
```

```
mean( df_amt$TARGET_LOSS_AMT )
```

```
## [1] 2676.163
```

```
mean( df_train$TARGET_LOSS_AMT )
```

```
## [1] 2699.644
```

```
mean( df_test$TARGET_LOSS_AMT )
```

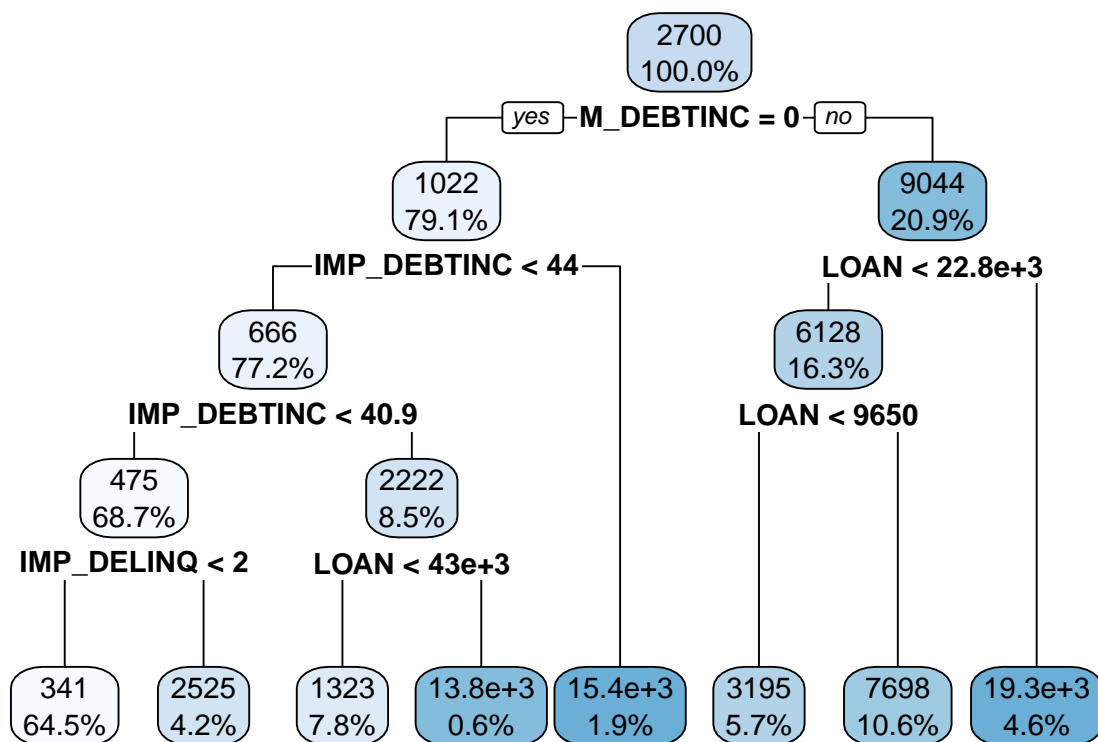
```
## [1] 2623.669
```

```
#Decision Tree Model
```

```
tr_set = rpart.control( maxdepth = 10 )
```

```
tr_model = rpart( data = df_train, TARGET_LOSS_AMT ~ .,  
                  control = tr_set, method = "poisson" )
```

```
rpart.plot( tr_model, digits = 3, extra = 100 )
```



```
tr_model$variable.importance
```

```
##          M_DEBTINC          IMP_DEBTINC          LOAN          IMP_DELTINC
##      12363304.653      5716473.850      4157682.841      1119800.042
##          IMP_VALUE          M_VALUE          IMP_MORTDUE          IMP_DEROG
##      512294.607      358980.971      312273.555      215388.583
##          IMP_YOJ FLAG.Reason.HomeImp          IMP_NINQ          IMP_CLAGE
##      103404.145      47981.060      32855.648      12564.283
## FLAG.Reason.DebtCon          IMP_CLNO
##      11995.265      7197.159
```

```
pt = predict( tr_model, df_test )
head(pt)
```

```
##          1          2          6          8          9         10
## 3194.7642 3194.7642 341.1016 341.1016 3194.7642 3194.7642
```

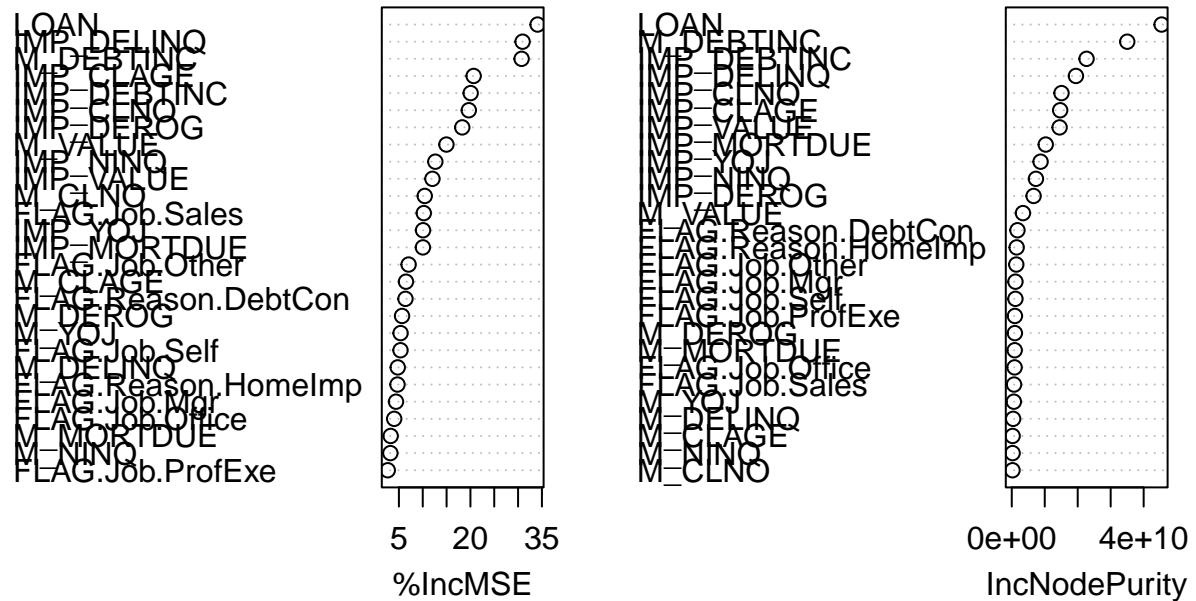
```
RMSEt = sqrt( mean( (df_test$TARGET_LOSS_AMT - pt )^2 ) )
```

```
#Random Forest Model
rf_model = randomForest( data = df_train, TARGET_LOSS_AMT ~ .,
                          ntree = 200, importance = TRUE )
importance( rf_model )
```

##	%IncMSE	IncNodePurity
## LOAN	34.095576	45449205487
## IMP_MORTDUE	10.031460	10299000629
## M_MORTDUE	3.272255	889087966
## IMP_VALUE	12.031425	14509861931
## M_VALUE	14.966826	3397182762
## IMP_YOJ	10.062998	8741387415
## M_YOJ	5.350942	631963812
## IMP_DEROG	18.283119	6623629143
## M_DEROG	5.657055	934822320
## IMP_DELINQ	30.946549	19455210412
## M_DELINQ	4.752220	417930848
## IMP_CLAGE	20.679289	14679814958
## M_CLAGE	6.480203	252365511
## IMP_NINQ	12.639399	7331302802
## M_NINQ	3.211669	252070959
## IMP_CLNO	19.663777	15079995783
## M_CLNO	10.417420	223326328
## IMP_DEBTINC	20.055383	22682382302
## M_DEBTINC	30.740130	35032540126
## FLAG.Job.Mgr	4.388293	1114118538
## FLAG.Job.Office	4.010110	856388951
## FLAG.Job.Other	7.036022	1375449784
## FLAG.Job.ProfExe	2.662832	948947638
## FLAG.Job.Sales	10.216173	692099911
## FLAG.Job.Self	5.334734	1101319723
## FLAG.Reason.DebtCon	6.373739	1729522425
## FLAG.Reason.HomeImp	4.702067	1487337399

```
varImpPlot( rf_model )
```

rf\_model



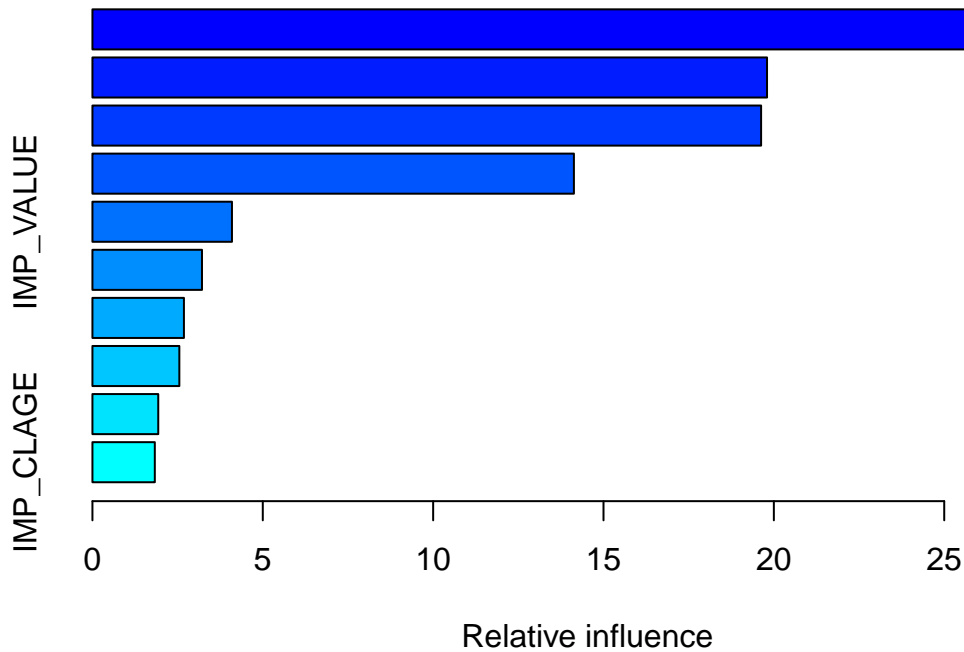
```
pr = predict( rf_model, df_test )
head(pr)
```

##	1	2	6	8	9	10
##	2177.7672	2577.7647	876.1110	896.7708	2886.8135	2295.3772

```
RMSEr = sqrt( mean( (df_test$TARGET_LOSS_AMT - pr ) ^2 ) )
```

### #Gradient Boosting Model

```
gb_model = gbm( data = df_train, TARGET_LOSS_AMT ~ .,
                 n.trees = 200, distribution = "poisson" )
summary.gbm( gb_model, cBars = 10 )
```



##	var	rel.inf
## M_DEBTINC	M_DEBTINC	29.34815372
## IMP_DEBTINC	IMP_DEBTINC	19.79951192
## LOAN	LOAN	19.62604762
## IMP_DELTINQ	IMP_DELTINQ	14.13141517
## IMP_VALUE	IMP_VALUE	4.09495584
## IMP_DEROG	IMP_DEROG	3.21537820
## IMP_CLNO	IMP_CLNO	2.68520122
## M_VALUE	M_VALUE	2.55105222
## IMP_NINQ	IMP_NINQ	1.93134692
## IMP_CLAGE	IMP_CLAGE	1.83083110
## FLAG.Job.Self	FLAG.Job.Self	0.20529953
## FLAG.Reason.HomeImp	FLAG.Reason.HomeImp	0.19693045
## FLAG.Job.Sales	FLAG.Job.Sales	0.19309198
## IMP_MORTDUE	IMP_MORTDUE	0.11048387
## FLAG.Reason.DebtCon	FLAG.Reason.DebtCon	0.08030023
## M_MORTDUE	M_MORTDUE	0.00000000
## IMP_YOJ	IMP_YOJ	0.00000000
## M_YOJ	M_YOJ	0.00000000
## M_DEROG	M_DEROG	0.00000000
## M_DELTINQ	M_DELTINQ	0.00000000
## M_CLAGE	M_CLAGE	0.00000000
## M_NINQ	M_NINQ	0.00000000
## M_CLNO	M_CLNO	0.00000000
## FLAG.Job.Mgr	FLAG.Job.Mgr	0.00000000
## FLAG.Job.Office	FLAG.Job.Office	0.00000000

```
## FLAG.Job.Other          FLAG.Job.Other  0.00000000
## FLAG.Job.ProfExe        FLAG.Job.ProfExe 0.00000000
```

```
pg = predict( gb_model, df_test, type = "response" )
```

```
## Using 200 trees...
```

```
head(pg)
```

```
## [1] 3057.7739 4763.5344 370.8974 387.9944 3201.9446 3624.5679
```

```
RMSEg = sqrt( mean( (df_test$TARGET_LOSS_AMT - pg )^2 ) )
```

```
#LINEAR ALL & BACKWARD TREE
```

```
theUpper_LR = lm( TARGET_LOSS_AMT ~ ., data = df_train )
```

```
theLower_LR = lm( TARGET_LOSS_AMT ~ 1, data = df_train )
```

```
summary( theUpper_LR )
```

```
##
```

```
## Call:
```

```
## lm(formula = TARGET_LOSS_AMT ~ ., data = df_train)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -30121  -2557   -371    1629   57953
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -7.316e+03  7.711e+02  -9.488  < 2e-16 ***
## LOAN           1.452e-01  8.723e-03  16.641  < 2e-16 ***
## IMP_MORTDUE    -8.730e-03  3.633e-03  -2.403  0.016311 *
## M_MORTDUE      1.304e+03  3.622e+02   3.600  0.000322 ***
## IMP_VALUE      1.218e-02  2.756e-03   4.422  1.00e-05 ***
## M_VALUE        6.245e+03  6.888e+02   9.067  < 2e-16 ***
## IMP_YOJ        -3.288e+01  1.267e+01  -2.595  0.009505 **
## M_YOJ          -9.320e+02  3.501e+02  -2.662  0.007788 **
## IMP_DEROG      9.398e+02  1.291e+02   7.280  3.98e-13 ***
## M_DEROG        -2.337e+03  4.443e+02  -5.261  1.51e-07 ***
## IMP_DELINQ     1.634e+03  8.837e+01  18.487  < 2e-16 ***
## M_DELINQ       -1.634e+03  5.788e+02  -2.824  0.004772 **
## IMP_CLAGE      -9.813e+00  1.122e+00  -8.743  < 2e-16 ***
## M_CLAGE        8.580e+02  7.962e+02   1.078  0.281266
## IMP_NINQ       2.792e+02  5.560e+01   5.022  5.34e-07 ***
## M_NINQ         8.525e+01  5.557e+02   0.153  0.878085
## IMP_CLNO       6.327e+01  1.027e+01   6.163  7.85e-10 ***
## M_CLNO        2.439e+03  1.029e+03   2.370  0.017827 *
## IMP_DEBTINC    1.175e+02  1.267e+01   9.275  < 2e-16 ***
## M_DEBTINC      6.248e+03  2.305e+02  27.102  < 2e-16 ***
## FLAG.Job.Mgr    1.092e+03  5.672e+02   1.924  0.054381 .
## FLAG.Job.Office 7.475e+02  5.578e+02   1.340  0.180319
## FLAG.Job.Other  1.360e+03  5.276e+02   2.578  0.009979 **
```



```
## FLAG.Job.ProfExe      1.145e+03  5.525e+02   2.072 0.038324 *
## FLAG.Job.Sales       3.851e+03  8.569e+02   4.494 7.20e-06 ***
## FLAG.Job.Self        2.497e+03  7.251e+02   3.444 0.000579 ***
## FLAG.Reason.DebtCon -2.055e+02  5.024e+02  -0.409 0.682467
## FLAG.Reason.HomeImp -8.583e+02  5.106e+02  -1.681 0.092814 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5637 on 4090 degrees of freedom
## Multiple R-squared:  0.419, Adjusted R-squared:  0.4151
## F-statistic: 109.2 on 27 and 4090 DF, p-value: < 2.2e-16
```

```
summary( theLower_LR )
```

```
##
## Call:
## lm(formula = TARGET_LOSS_AMT ~ 1, data = df_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2700  -2700  -2700  -2700   71246
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2699.6      114.9    23.51  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7370 on 4117 degrees of freedom
```

```
lr_model = stepAIC( theUpper_LR, direction = "backward",
                    scope = list(lower = theLower_LR, upper = theUpper_LR))
```

```
## Start:  AIC=71162.42
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##      M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##      M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO +
##      M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##      FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##      FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
##              Df Sum of Sq      RSS   AIC
## - M_NINQ      1 7.4768e+05 1.2994e+11 71160
## - FLAG.Reason.DebtCon 1 5.3179e+06 1.2995e+11 71161
## - M_CLAGE      1 3.6894e+07 1.2998e+11 71162
## - FLAG.Job.Office 1 5.7047e+07 1.3000e+11 71162
## <none>                  1.2994e+11 71162
## - FLAG.Reason.HomeImp 1 8.9789e+07 1.3003e+11 71163
## - FLAG.Job.Mgr      1 1.1765e+08 1.3006e+11 71164
## - FLAG.Job.ProfExe  1 1.3640e+08 1.3008e+11 71165
## - M_CLNO          1 1.7847e+08 1.3012e+11 71166
## - IMP_MORTDUE      1 1.8344e+08 1.3012e+11 71166
## - FLAG.Job.Other   1 2.1111e+08 1.3015e+11 71167
```

```

## - IMP_YOJ          1 2.1387e+08 1.3015e+11 71167
## - M_YOJ            1 2.2520e+08 1.3017e+11 71168
## - M_DELINQ         1 2.5329e+08 1.3019e+11 71168
## - FLAG.Job.Self    1 3.7684e+08 1.3032e+11 71172
## - M_MORTDUE        1 4.1178e+08 1.3035e+11 71173
## - IMP_VALUE        1 6.2122e+08 1.3056e+11 71180
## - FLAG.Job.Sales   1 6.4152e+08 1.3058e+11 71181
## - IMP_NINQ         1 8.0116e+08 1.3074e+11 71186
## - M_DEROG          1 8.7929e+08 1.3082e+11 71188
## - IMP_CLNO         1 1.2065e+09 1.3115e+11 71198
## - IMP_DEROG        1 1.6838e+09 1.3162e+11 71213
## - IMP_CLAGE        1 2.4288e+09 1.3237e+11 71237
## - M_VALUE          1 2.6120e+09 1.3255e+11 71242
## - IMP_DEBTINC      1 2.7331e+09 1.3267e+11 71246
## - LOAN             1 8.7976e+09 1.3874e+11 71430
## - IMP_DELINQ       1 1.0859e+10 1.4080e+11 71491
## - M_DEBTINC        1 2.3335e+10 1.5328e+11 71841
##
## Step:  AIC=71160.45
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO +
##     IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##     FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
##              Df  Sum of Sq      RSS   AIC
## - FLAG.Reason.DebtCon  1 5.3545e+06 1.2995e+11 71159
## - M_CLAGE              1 3.6694e+07 1.2998e+11 71160
## - FLAG.Job.Office      1 5.6718e+07 1.3000e+11 71160
## <none>                  1.2994e+11 71160
## - FLAG.Reason.HomeImp  1 8.9811e+07 1.3003e+11 71161
## - FLAG.Job.Mgr         1 1.1733e+08 1.3006e+11 71162
## - FLAG.Job.ProfExe     1 1.3608e+08 1.3008e+11 71163
## - IMP_MORTDUE          1 1.8357e+08 1.3013e+11 71164
## - M_CLNO              1 1.9192e+08 1.3013e+11 71165
## - FLAG.Job.Other       1 2.1068e+08 1.3015e+11 71165
## - IMP_YOJ              1 2.1419e+08 1.3016e+11 71165
## - M_YOJ                1 2.2448e+08 1.3017e+11 71166
## - M_DELINQ             1 3.1692e+08 1.3026e+11 71168
## - FLAG.Job.Self        1 3.7903e+08 1.3032e+11 71170
## - M_MORTDUE            1 4.1398e+08 1.3036e+11 71172
## - IMP_VALUE            1 6.2225e+08 1.3056e+11 71178
## - FLAG.Job.Sales       1 6.4078e+08 1.3058e+11 71179
## - IMP_NINQ             1 8.0117e+08 1.3074e+11 71184
## - M_DEROG              1 8.8010e+08 1.3082e+11 71186
## - IMP_CLNO             1 1.2083e+09 1.3115e+11 71197
## - IMP_DEROG            1 1.6917e+09 1.3163e+11 71212
## - IMP_CLAGE            1 2.4289e+09 1.3237e+11 71235
## - M_VALUE              1 2.6136e+09 1.3256e+11 71240
## - IMP_DEBTINC          1 2.7439e+09 1.3269e+11 71245
## - LOAN                 1 8.8031e+09 1.3874e+11 71428
## - IMP_DELINQ           1 1.0893e+10 1.4083e+11 71490
## - M_DEBTINC            1 2.3340e+10 1.5328e+11 71839

```

```

##
## Step: AIC=71158.62
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO +
##     IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##     FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.HomeImp
##
##           Df Sum of Sq      RSS   AIC
## - M_CLAGE      1 3.9387e+07 1.2999e+11 71158
## - FLAG.Job.Office      1 5.2120e+07 1.3000e+11 71158
## <none>                                1.2995e+11 71159
## - FLAG.Job.Mgr      1 1.1201e+08 1.3006e+11 71160
## - FLAG.Job.ProfExe      1 1.3074e+08 1.3008e+11 71161
## - IMP_MORTDUE      1 1.8739e+08 1.3013e+11 71163
## - M_CLNO      1 1.9370e+08 1.3014e+11 71163
## - FLAG.Job.Other      1 2.0611e+08 1.3015e+11 71163
## - IMP_YOJ      1 2.1222e+08 1.3016e+11 71163
## - M_YOJ      1 2.1914e+08 1.3017e+11 71164
## - M_DELINQ      1 3.1527e+08 1.3026e+11 71167
## - FLAG.Reason.HomeImp      1 3.4001e+08 1.3029e+11 71167
## - FLAG.Job.Self      1 3.7388e+08 1.3032e+11 71168
## - M_MORTDUE      1 4.2639e+08 1.3037e+11 71170
## - IMP_VALUE      1 6.3444e+08 1.3058e+11 71177
## - FLAG.Job.Sales      1 6.3825e+08 1.3059e+11 71177
## - IMP_NINQ      1 7.9879e+08 1.3075e+11 71182
## - M_DEROG      1 8.9068e+08 1.3084e+11 71185
## - IMP_CLNO      1 1.2113e+09 1.3116e+11 71195
## - IMP_DEROG      1 1.6951e+09 1.3164e+11 71210
## - IMP_CLAGE      1 2.4236e+09 1.3237e+11 71233
## - M_VALUE      1 2.6307e+09 1.3258e+11 71239
## - IMP_DEBTINC      1 2.7403e+09 1.3269e+11 71243
## - LOAN      1 8.8285e+09 1.3878e+11 71427
## - IMP_DELINQ      1 1.0952e+10 1.4090e+11 71490
## - M_DEBTINC      1 2.3335e+10 1.5328e+11 71837
##
## Step: AIC=71157.87
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
##     M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office + FLAG.Job.Other +
##     FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.HomeImp
##
##           Df Sum of Sq      RSS   AIC
## - FLAG.Job.Office      1 4.1848e+07 1.3003e+11 71157
## <none>                                1.2999e+11 71158
## - FLAG.Job.Mgr      1 9.8120e+07 1.3008e+11 71159
## - FLAG.Job.ProfExe      1 1.1536e+08 1.3010e+11 71160
## - IMP_MORTDUE      1 1.7803e+08 1.3016e+11 71162
## - FLAG.Job.Other      1 1.8810e+08 1.3017e+11 71162
## - M_YOJ      1 2.1795e+08 1.3020e+11 71163
## - IMP_YOJ      1 2.2327e+08 1.3021e+11 71163
## - M_DELINQ      1 3.2444e+08 1.3031e+11 71166

```

```

## - FLAG.Reason.HomeImp 1 3.5024e+08 1.3034e+11 71167
## - FLAG.Job.Self 1 3.5493e+08 1.3034e+11 71167
## - M_MORTDUE 1 4.2940e+08 1.3042e+11 71169
## - FLAG.Job.Sales 1 6.1610e+08 1.3060e+11 71175
## - IMP_VALUE 1 6.2806e+08 1.3061e+11 71176
## - M_CLNO 1 7.9248e+08 1.3078e+11 71181
## - IMP_NINQ 1 8.1131e+08 1.3080e+11 71181
## - M_DEROG 1 8.8809e+08 1.3087e+11 71184
## - IMP_CLNO 1 1.1720e+09 1.3116e+11 71193
## - IMP_DEROG 1 1.7033e+09 1.3169e+11 71209
## - IMP_CLAGE 1 2.3991e+09 1.3239e+11 71231
## - M_VALUE 1 2.6264e+09 1.3261e+11 71238
## - IMP_DEBTINC 1 2.7047e+09 1.3269e+11 71241
## - LOAN 1 8.8378e+09 1.3882e+11 71427
## - IMP_DELINQ 1 1.0936e+10 1.4092e+11 71489
## - M_DEBTINC 1 2.3442e+10 1.5343e+11 71839
##
## Step: AIC=71157.19
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
## M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
## M_DELINQ + IMP_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
## M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Other + FLAG.Job.ProfExe +
## FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.HomeImp
##
##           Df Sum of Sq      RSS   AIC
## <none>                 1.3003e+11 71157
## - FLAG.Job.Mgr 1 6.4173e+07 1.3009e+11 71157
## - FLAG.Job.ProfExe 1 9.7102e+07 1.3013e+11 71158
## - IMP_MORTDUE 1 1.8286e+08 1.3021e+11 71161
## - IMP_YOJ 1 2.2837e+08 1.3026e+11 71162
## - M_YOJ 1 2.3157e+08 1.3026e+11 71163
## - FLAG.Job.Other 1 2.7922e+08 1.3031e+11 71164
## - M_DELINQ 1 3.1192e+08 1.3034e+11 71165
## - FLAG.Reason.HomeImp 1 3.3502e+08 1.3036e+11 71166
## - FLAG.Job.Self 1 3.6532e+08 1.3039e+11 71167
## - M_MORTDUE 1 4.0900e+08 1.3044e+11 71168
## - IMP_VALUE 1 6.3198e+08 1.3066e+11 71175
## - FLAG.Job.Sales 1 6.4261e+08 1.3067e+11 71175
## - M_CLNO 1 7.7267e+08 1.3080e+11 71180
## - IMP_NINQ 1 8.0146e+08 1.3083e+11 71180
## - M_DEROG 1 8.9321e+08 1.3092e+11 71183
## - IMP_CLNO 1 1.1951e+09 1.3122e+11 71193
## - IMP_DEROG 1 1.6944e+09 1.3172e+11 71209
## - IMP_CLAGE 1 2.3737e+09 1.3240e+11 71230
## - M_VALUE 1 2.6236e+09 1.3265e+11 71237
## - IMP_DEBTINC 1 2.8245e+09 1.3285e+11 71244
## - LOAN 1 8.8314e+09 1.3886e+11 71426
## - IMP_DELINQ 1 1.0963e+10 1.4099e+11 71489
## - M_DEBTINC 1 2.3545e+10 1.5357e+11 71841

```

```
summary( lr_model )
```

```

##
## Call:

```

```
## lm(formula = TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE +
##     IMP_VALUE + M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG +
##     IMP_DELINQ + M_DELINQ + IMP_CLAGE + IMP_NINQ + IMP_CLNO +
##     M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Other +
##     FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.HomeImp,
##     data = df_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -30145  -2533   -365    1632   57948
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.864e+03  5.298e+02 -12.956 < 2e-16 ***
## LOAN           1.448e-01  8.685e-03  16.675 < 2e-16 ***
## IMP_MORTDUE    -8.686e-03  3.620e-03  -2.399 0.016464 *
## M_MORTDUE      1.288e+03  3.588e+02   3.589 0.000336 ***
## IMP_VALUE      1.225e-02  2.746e-03   4.461 8.39e-06 ***
## M_VALUE        6.245e+03  6.871e+02   9.089 < 2e-16 ***
## IMP_YOJ       -3.389e+01  1.264e+01  -2.681 0.007359 **
## M_YOJ         -9.311e+02  3.448e+02  -2.700 0.006958 **
## IMP_DEROG      9.412e+02  1.289e+02   7.304 3.34e-13 ***
## M_DEROG      -2.346e+03  4.424e+02  -5.303 1.20e-07 ***
## IMP_DELINQ     1.637e+03  8.810e+01  18.579 < 2e-16 ***
## M_DELINQ     -1.573e+03  5.021e+02  -3.134 0.001738 **
## IMP_CLAGE     -9.668e+00  1.118e+00  -8.645 < 2e-16 ***
## IMP_NINQ      2.784e+02  5.543e+01   5.023 5.29e-07 ***
## IMP_CLNO      6.178e+01  1.007e+01   6.134 9.36e-10 ***
## M_CLNO       3.029e+03  6.142e+02   4.932 8.45e-07 ***
## IMP_DEBTINC    1.178e+02  1.249e+01   9.430 < 2e-16 ***
## M_DEBTINC     6.265e+03  2.301e+02  27.227 < 2e-16 ***
## FLAG.Job.Mgr   4.527e+02  3.185e+02   1.421 0.155264
## FLAG.Job.Other 7.296e+02  2.461e+02   2.965 0.003044 **
## FLAG.Job.ProfExe 5.018e+02  2.870e+02   1.749 0.080450 .
## FLAG.Job.Sales 3.189e+03  7.089e+02   4.498 7.05e-06 ***
## FLAG.Job.Self  1.864e+03  5.495e+02   3.392 0.000702 ***
## FLAG.Reason.HomeImp -6.591e+02  2.029e+02  -3.248 0.001172 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5636 on 4094 degrees of freedom
## Multiple R-squared:  0.4186, Adjusted R-squared:  0.4153
## F-statistic: 128.1 on 23 and 4094 DF, p-value: < 2.2e-16
```

```
plr = predict( lr_model, df_test )
head(plr)
```

```
##           1           2           6           8           9          10
## 3573.531 6735.413 -2460.353 -2645.221 6266.903 6970.125
```

```
RMSElr = sqrt( mean( ( df_test$TARGET_LOSS_AMT - plr )^2 ) )
```

```
#LINEAR REG STEPWISE FORWARD TREE
treeVars = tr_model$variable.importance
treeVars = names(treeVars)
treeVarsPlus = paste( treeVars, collapse = "+")
F = as.formula( paste( "TARGET_LOSS_AMT ~", treeVarsPlus ))

tree_LR = lm( F, data = df_train )
theLower_LR = lm( TARGET_LOSS_AMT ~ 1, data = df_train )
summary( tree_LR )
```

```
##
## Call:
## lm(formula = F, data = df_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26099  -2623   -443    1654   57737
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.539e+03  6.346e+02 -10.303 < 2e-16 ***
## M_DEBTINC      6.648e+03  2.314e+02  28.732 < 2e-16 ***
## IMP_DEBTINC     1.133e+02  1.245e+01   9.100 < 2e-16 ***
## LOAN           1.489e-01  8.602e-03  17.315 < 2e-16 ***
## IMP_DELTINC     1.480e+03  8.772e+01  16.871 < 2e-16 ***
## IMP_VALUE       1.255e-02  2.696e-03   4.657 3.32e-06 ***
## M_VALUE         6.576e+03  6.963e+02   9.445 < 2e-16 ***
## IMP_MORTDUE     -8.359e-03  3.567e-03  -2.343  0.0192 *
## IMP_DEROG       6.056e+02  1.246e+02   4.862 1.21e-06 ***
## IMP_YOJ        -2.884e+01  1.268e+01  -2.275  0.0230 *
## FLAG.Reason.HomeImp -3.457e+02  4.704e+02  -0.735  0.4624
## IMP_NINQ        2.798e+02  5.588e+01   5.008 5.72e-07 ***
## IMP_CLAGE       -9.984e+00  1.120e+00  -8.912 < 2e-16 ***
## FLAG.Reason.DebtCon  1.083e+02  4.577e+02   0.237  0.8130
## IMP_CLNO        5.816e+01  9.918e+00   5.864 4.88e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5728 on 4103 degrees of freedom
## Multiple R-squared:  0.398, Adjusted R-squared:  0.3959
## F-statistic: 193.7 on 14 and 4103 DF, p-value: < 2.2e-16
```

```
summary( theLower_LR )
```

```
##
## Call:
## lm(formula = TARGET_LOSS_AMT ~ 1, data = df_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##  -2700  -2700  -2700  -2700   71246
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2699.6      114.9   23.51  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7370 on 4117 degrees of freedom
```

```
lrt_model = stepAIC( theLower_LR, direction = "forward",
                     scope = list(lower = theLower_LR, upper = tree_LR ))
```

```
## Start:  AIC=73344.23
```

```
## TARGET_LOSS_AMT ~ 1
```

```
##
##           Df Sum of Sq      RSS   AIC
## + M_DEBTINC      1 4.3823e+10 1.7981e+11 72448
## + IMP_DELTINC      1 2.6109e+10 1.9753e+11 72835
## + LOAN            1 1.0504e+10 2.1313e+11 73148
## + M_VALUE         1 1.0478e+10 2.1316e+11 73149
## + IMP_DEROG       1 1.0060e+10 2.1358e+11 73157
## + IMP_DEBTINC     1 9.1093e+09 2.1453e+11 73175
## + IMP_NINQ        1 7.9667e+09 2.1567e+11 73197
## + IMP_CLNO        1 4.7198e+09 2.1892e+11 73258
## + IMP_CLAGE       1 3.6286e+09 2.2001e+11 73279
## + IMP_VALUE       1 3.2056e+09 2.2043e+11 73287
## + IMP_MORTDUE     1 1.7214e+09 2.2191e+11 73314
## + FLAG.Reason.HomeImp 1 1.2742e+09 2.2236e+11 73323
## + FLAG.Reason.DebtCon 1 1.2507e+09 2.2238e+11 73323
## + IMP_YOJ         1 2.3882e+08 2.2340e+11 73342
## <none>                2.2364e+11 73344
##
```

```
## Step:  AIC=72448.09
```

```
## TARGET_LOSS_AMT ~ M_DEBTINC
```

```
##
##           Df Sum of Sq      RSS   AIC
## + LOAN            1 1.5131e+10 1.6468e+11 72088
## + IMP_DELTINC     1 1.4412e+10 1.6540e+11 72106
## + IMP_DEBTINC     1 6.9254e+09 1.7289e+11 72288
## + M_VALUE         1 5.0749e+09 1.7474e+11 72332
## + IMP_DEROG       1 4.8147e+09 1.7500e+11 72338
## + IMP_CLNO        1 4.7851e+09 1.7503e+11 72339
## + IMP_VALUE       1 4.4977e+09 1.7531e+11 72346
## + IMP_NINQ        1 3.6814e+09 1.7613e+11 72365
## + IMP_MORTDUE     1 2.7334e+09 1.7708e+11 72387
## + FLAG.Reason.HomeImp 1 1.8065e+09 1.7801e+11 72409
## + FLAG.Reason.DebtCon 1 1.7322e+09 1.7808e+11 72410
## + IMP_CLAGE       1 1.6394e+09 1.7817e+11 72412
## + IMP_YOJ         1 1.3375e+08 1.7968e+11 72447
## <none>                1.7981e+11 72448
##
```

```
## Step:  AIC=72088.11
```

```
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN
```

```
##
##           Df Sum of Sq      RSS   AIC
## + IMP_DELTINC     1 1.4954e+10 1.4973e+11 71698
```

```

## + IMP_DEBTINC          1 5.2054e+09 1.5948e+11 71958
## + IMP_DEROG            1 5.0416e+09 1.5964e+11 71962
## + M_VALUE              1 4.6970e+09 1.5998e+11 71971
## + IMP_CLNO             1 3.7752e+09 1.6091e+11 71995
## + IMP_NINQ             1 2.8043e+09 1.6188e+11 72019
## + IMP_CLAGE            1 2.4940e+09 1.6219e+11 72027
## + IMP_VALUE            1 8.2119e+08 1.6386e+11 72070
## + IMP_MORTDUE          1 7.6348e+08 1.6392e+11 72071
## + FLAG.Reason.HomeImp  1 6.4309e+08 1.6404e+11 72074
## + IMP_YOJ              1 6.1221e+08 1.6407e+11 72075
## + FLAG.Reason.DebtCon  1 5.0593e+08 1.6418e+11 72077
## <none>                  1.6468e+11 72088
##
## Step:  AIC=71698.1
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ
##
##              Df  Sum of Sq      RSS    AIC
## + IMP_DEBTINC    1 4535862036 1.4519e+11 71573
## + IMP_CLAGE      1 2944175768 1.4678e+11 71618
## + M_VALUE        1 2781723807 1.4695e+11 71623
## + IMP_NINQ       1 2412561513 1.4732e+11 71633
## + IMP_DEROG      1 1915299259 1.4781e+11 71647
## + IMP_CLNO       1 1662525224 1.4807e+11 71654
## + IMP_VALUE      1 875074480 1.4885e+11 71676
## + IMP_YOJ        1 800605127 1.4893e+11 71678
## + FLAG.Reason.DebtCon 1 667577479 1.4906e+11 71682
## + FLAG.Reason.HomeImp 1 636528953 1.4909e+11 71683
## + IMP_MORTDUE    1 631702620 1.4910e+11 71683
## <none>            1.4973e+11 71698
##
## Step:  AIC=71573.42
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ + IMP_DEBTINC
##
##              Df  Sum of Sq      RSS    AIC
## + M_VALUE        1 3076182532 1.4212e+11 71487
## + IMP_CLAGE      1 2567917197 1.4262e+11 71502
## + IMP_DEROG      1 1821188197 1.4337e+11 71523
## + IMP_NINQ       1 1753706518 1.4344e+11 71525
## + IMP_CLNO       1 909590426 1.4428e+11 71550
## + IMP_YOJ        1 618423128 1.4457e+11 71558
## + FLAG.Reason.HomeImp 1 549913533 1.4464e+11 71560
## + IMP_VALUE      1 499507101 1.4469e+11 71561
## + FLAG.Reason.DebtCon 1 456095096 1.4474e+11 71562
## + IMP_MORTDUE    1 269099314 1.4492e+11 71568
## <none>            1.4519e+11 71573
##
## Step:  AIC=71487.24
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ + IMP_DEBTINC +
##      M_VALUE
##
##              Df  Sum of Sq      RSS    AIC
## + IMP_CLAGE      1 2486603627 1.3963e+11 71417
## + IMP_NINQ       1 1783308127 1.4033e+11 71437
## + IMP_DEROG      1 1455722617 1.4066e+11 71447

```



```

## + IMP_CLNO          1 1016968275 1.4110e+11 71460
## + IMP_YOJ           1  595414896 1.4152e+11 71472
## + IMP_VALUE         1  565249253 1.4155e+11 71473
## + FLAG.Reason.HomeImp 1  536609294 1.4158e+11 71474
## + FLAG.Reason.DebtCon 1  521026447 1.4159e+11 71474
## + IMP_MORTDUE       1  279772996 1.4184e+11 71481
## <none>              1.4212e+11 71487
##
## Step:  AIC=71416.55
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ + IMP_DEBTINC +
##      M_VALUE + IMP_CLAGE
##
##              Df  Sum of Sq      RSS   AIC
## + IMP_CLNO          1 1967247956 1.3766e+11 71360
## + IMP_NINQ          1 1390889449 1.3824e+11 71377
## + IMP_DEROG         1 1238631444 1.3839e+11 71382
## + IMP_VALUE         1 1011406646 1.3862e+11 71389
## + IMP_MORTDUE       1  516812268 1.3911e+11 71403
## + FLAG.Reason.HomeImp 1  389075651 1.3924e+11 71407
## + FLAG.Reason.DebtCon 1  349064803 1.3928e+11 71408
## + IMP_YOJ           1  250246504 1.3938e+11 71411
## <none>              1.3963e+11 71417
##
## Step:  AIC=71360.11
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ + IMP_DEBTINC +
##      M_VALUE + IMP_CLAGE + IMP_CLNO
##
##              Df  Sum of Sq      RSS   AIC
## + IMP_NINQ          1 1136177630 1.3653e+11 71328
## + IMP_DEROG         1 1060659382 1.3660e+11 71330
## + IMP_VALUE         1  517517321 1.3714e+11 71347
## + IMP_YOJ           1  262234393 1.3740e+11 71354
## + FLAG.Reason.HomeImp 1  213969494 1.3745e+11 71356
## + FLAG.Reason.DebtCon 1  169132753 1.3749e+11 71357
## + IMP_MORTDUE       1  121966700 1.3754e+11 71358
## <none>              1.3766e+11 71360
##
## Step:  AIC=71327.99
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ + IMP_DEBTINC +
##      M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_NINQ
##
##              Df  Sum of Sq      RSS   AIC
## + IMP_DEROG         1 783559891 1.3574e+11 71306
## + IMP_VALUE         1 568936248 1.3596e+11 71313
## + IMP_YOJ           1 209787943 1.3632e+11 71324
## + FLAG.Reason.HomeImp 1 128045674 1.3640e+11 71326
## + IMP_MORTDUE       1 113815270 1.3641e+11 71327
## + FLAG.Reason.DebtCon 1  92156702 1.3643e+11 71327
## <none>              1.3653e+11 71328
##
## Step:  AIC=71306.28
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ + IMP_DEBTINC +
##      M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_NINQ + IMP_DEROG
##

```

```

##              Df Sum of Sq      RSS   AIC
## + IMP_VALUE      1 625228188 1.3512e+11 71289
## + IMP_YOJ        1 165089312 1.3558e+11 71303
## + IMP_MORTDUE     1 136909486 1.3561e+11 71304
## + FLAG.Reason.HomeImp 1 117148409 1.3562e+11 71305
## + FLAG.Reason.DebtCon 1  99647938 1.3564e+11 71305
## <none>              1.3574e+11 71306
##
## Step:  AIC=71289.27
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ + IMP_DEBTINC +
##      M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_NINQ + IMP_DEROG + IMP_VALUE
##
##              Df Sum of Sq      RSS   AIC
## + FLAG.Reason.HomeImp 1 166811780 1.3495e+11 71286
## + FLAG.Reason.DebtCon 1 147139021 1.3497e+11 71287
## + IMP_YOJ            1 142771612 1.3497e+11 71287
## + IMP_MORTDUE        1 135064472 1.3498e+11 71287
## <none>              1.3512e+11 71289
##
## Step:  AIC=71286.18
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ + IMP_DEBTINC +
##      M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_NINQ + IMP_DEROG + IMP_VALUE +
##      FLAG.Reason.HomeImp
##
##              Df Sum of Sq      RSS   AIC
## + IMP_MORTDUE      1 141188318 1.3481e+11 71284
## + IMP_YOJ          1 131616641 1.3482e+11 71284
## <none>              1.3495e+11 71286
## + FLAG.Reason.DebtCon 1  1147632 1.3495e+11 71288
##
## Step:  AIC=71283.87
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ + IMP_DEBTINC +
##      M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_NINQ + IMP_DEROG + IMP_VALUE +
##      FLAG.Reason.HomeImp + IMP_MORTDUE
##
##              Df Sum of Sq      RSS   AIC
## + IMP_YOJ          1 169812561 1.3464e+11 71281
## <none>              1.3481e+11 71284
## + FLAG.Reason.DebtCon 1  1849685 1.3481e+11 71286
##
## Step:  AIC=71280.68
## TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ + IMP_DEBTINC +
##      M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_NINQ + IMP_DEROG + IMP_VALUE +
##      FLAG.Reason.HomeImp + IMP_MORTDUE + IMP_YOJ
##
##              Df Sum of Sq      RSS   AIC
## <none>              1.3464e+11 71281
## + FLAG.Reason.DebtCon 1  1837235 1.3464e+11 71283

```

```
summary( lrt_model )
```

```

##
## Call:
## lm(formula = TARGET_LOSS_AMT ~ M_DEBTINC + LOAN + IMP_DELINQ +

```

```
##      IMP_DEBTINC + M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_NINQ +
##      IMP_DEROG + IMP_VALUE + FLAG.Reason.HomeImp + IMP_MORTDUE +
##      IMP_YOJ, data = df_train)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -26084   -2627    -438     1654    57733
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.447e+03  5.023e+02 -12.834 < 2e-16 ***
## M_DEBTINC      6.649e+03  2.313e+02  28.747 < 2e-16 ***
## LOAN           1.491e-01  8.581e-03  17.375 < 2e-16 ***
## IMP_DELTINC    1.479e+03  8.758e+01  16.885 < 2e-16 ***
## IMP_DEBTINC    1.136e+02  1.241e+01   9.152 < 2e-16 ***
## M_VALUE        6.565e+03  6.946e+02   9.452 < 2e-16 ***
## IMP_CLAGE     -9.999e+00  1.118e+00  -8.943 < 2e-16 ***
## IMP_CLNO       5.829e+01  9.900e+00   5.888 4.22e-09 ***
## IMP_NINQ       2.803e+02  5.583e+01   5.021 5.36e-07 ***
## IMP_DEROG      6.036e+02  1.242e+02   4.858 1.23e-06 ***
## IMP_VALUE      1.253e-02  2.694e-03   4.652 3.40e-06 ***
## FLAG.Reason.HomeImp -4.463e+02  2.016e+02  -2.214  0.0269 *
## IMP_MORTDUE    -8.338e-03  3.566e-03  -2.338  0.0194 *
## IMP_YOJ        -2.884e+01  1.267e+01  -2.275  0.0230 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5728 on 4104 degrees of freedom
## Multiple R-squared:  0.398, Adjusted R-squared:  0.396
## F-statistic: 208.7 on 13 and 4104 DF, p-value: < 2.2e-16
```

```
plr_tree = predict( tree_LR, df_test )
head( plr_tree )
```

```
##           1           2           6           8           9          10
## 3726.604  6553.318 -2701.449 -2872.021  6027.274  3405.274
```

```
RMSElr_tree = sqrt( mean( ( df_test$TARGET_LOSS_AMT - plr_tree )^2 ) )
```

```
plr_tree_step = predict( lrt_model, df_test )
head( plr_tree_step )
```

```
##           1           2           6           8           9          10
## 3727.527  6552.095 -2701.271 -2872.270  6024.594  3406.346
```

```
RMSElr_tree_step = sqrt( mean( ( df_test$TARGET_LOSS_AMT - plr_tree_step )^2 ) )
```

```
print( paste( "Decision Tree RMSE =", RMSEt ) )
```

```
## [1] "Decision Tree RMSE = 5288.36862187847"
```

```

print( paste( "Random Forest RMSE =", RMSEr ))

## [1] "Random Forest RMSE = 4201.17775586498"

print( paste( "Gradient Boosting RMSE =", RMSEg ))

## [1] "Gradient Boosting RMSE = 5344.21076557694"

print( paste("LR BACK RMSE = ", RMSElr ))

## [1] "LR BACK RMSE = 5091.68878446595"

print( paste("LR TREE RMSE = ", RMSElr_tree ))

## [1] "LR TREE RMSE = 5257.37010242306"

print( paste("LR TREE STEP RMSE = ", RMSElr_tree_step ))

## [1] "LR TREE STEP RMSE = 5257.36931414712"

#Random Forest performs best with lowest RMSE around 4201.

#Step 4: Probability / Severity Model Model
#I choose LOGIT Backward to predict TARGET_BAD_FLAG

df_flag = df
df_flag$TARGET_LOSS_AMT = NULL

FLAG = sample( c(TRUE, FALSE), nrow(df_flag), replace = TRUE, prob = c(0.7, 0.3))
df_flag_train = df_flag[FLAG, ]
df_flag_test = df_flag[!FLAG, ]

tr_set = rpart.control( maxdepth = 10 )

#LOGIT Backward Model 2
theUpper_LR2 = glm( TARGET_BAD_FLAG ~ ., family = "binomial", data = df_flag_train )
lr_model2 = stepAIC( theUpper_LR2, direction = "backward")

## Start: AIC=2351.93
## TARGET_BAD_FLAG ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
## M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
## M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO +
## M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
## FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
## FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
## Df Deviance AIC
## - FLAG.Reason.DebtCon 1 2296.0 2350.0

```

```

## - M_NINQ          1    2296.0 2350.0
## - LOAN            1    2296.1 2350.1
## - FLAG.Reason.HomeImp 1    2296.4 2350.4
## <none>            1    2295.9 2351.9
## - IMP_CLNO        1    2299.0 2353.0
## - M_MORTDUE        1    2299.9 2353.9
## - IMP_YOJ          1    2300.4 2354.4
## - FLAG.Job.Office  1    2300.9 2354.9
## - IMP_MORTDUE      1    2302.3 2356.3
## - IMP_VALUE        1    2303.0 2357.0
## - M_CLAGE          1    2305.0 2359.0
## - M_DELINQ         1    2306.2 2360.2
## - M_YOJ            1    2307.3 2361.3
## - FLAG.Job.ProfExe 1    2310.7 2364.7
## - FLAG.Job.Self    1    2311.4 2365.4
## - FLAG.Job.Other   1    2314.3 2368.3
## - M_CLNO           1    2314.6 2368.6
## - FLAG.Job.Mgr     1    2315.6 2369.6
## - IMP_NINQ         1    2320.9 2374.9
## - FLAG.Job.Sales   1    2328.8 2382.8
## - IMP_CLAGE        1    2349.1 2403.1
## - IMP_DEROG        1    2361.6 2415.6
## - M_DEROG          1    2377.8 2431.8
## - M_VALUE          1    2410.2 2464.2
## - IMP_DEBTINC      1    2413.0 2467.0
## - IMP_DELINQ       1    2540.6 2594.6
## - M_DEBTINC        1    2859.3 2913.3
##
## Step:  AIC=2350.03
## TARGET_BAD_FLAG ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO +
##     M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##     FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.HomeImp
##
##           Df Deviance    AIC
## - M_NINQ          1    2296.2 2348.2
## - LOAN            1    2296.2 2348.2
## - FLAG.Reason.HomeImp 1    2297.6 2349.6
## <none>            1    2296.0 2350.0
## - IMP_CLNO        1    2299.1 2351.1
## - M_MORTDUE        1    2300.0 2352.0
## - IMP_YOJ          1    2300.5 2352.5
## - FLAG.Job.Office  1    2301.9 2353.9
## - IMP_MORTDUE      1    2302.4 2354.4
## - IMP_VALUE        1    2303.1 2355.1
## - M_CLAGE          1    2305.0 2357.0
## - M_DELINQ         1    2306.3 2358.3
## - M_YOJ            1    2307.7 2359.7
## - FLAG.Job.ProfExe 1    2312.7 2364.7
## - FLAG.Job.Self    1    2312.9 2364.9
## - M_CLNO           1    2314.7 2366.7
## - FLAG.Job.Other   1    2317.3 2369.3

```

```

## - FLAG.Job.Mgr          1    2318.3 2370.3
## - IMP_NINQ              1    2321.1 2373.1
## - FLAG.Job.Sales        1    2331.4 2383.4
## - IMP_CLAGE             1    2349.8 2401.8
## - IMP_DEROG            1    2361.7 2413.7
## - M_DEROG              1    2377.8 2429.8
## - M_VALUE              1    2411.0 2463.0
## - IMP_DEBTINC          1    2413.0 2465.0
## - IMP_DELINQ           1    2540.6 2592.6
## - M_DEBTINC            1    2859.3 2911.3
##
## Step:  AIC=2348.15
## TARGET_BAD_FLAG ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO +
##     IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##     FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.HomeImp
##
##              Df Deviance    AIC
## - LOAN                1    2296.3 2346.3
## - FLAG.Reason.HomeImp  1    2297.7 2347.7
## <none>                 1    2296.2 2348.2
## - IMP_CLNO            1    2299.2 2349.2
## - M_MORTDUE           1    2300.1 2350.1
## - IMP_YOJ             1    2300.7 2350.7
## - FLAG.Job.Office     1    2302.0 2352.0
## - IMP_MORTDUE         1    2302.5 2352.5
## - IMP_VALUE           1    2303.1 2353.1
## - M_CLAGE             1    2305.2 2355.2
## - M_YOJ               1    2308.5 2358.5
## - M_DELINQ            1    2311.4 2361.4
## - FLAG.Job.ProfExe    1    2312.8 2362.8
## - FLAG.Job.Self       1    2312.9 2362.9
## - M_CLNO              1    2315.8 2365.8
## - FLAG.Job.Other      1    2317.4 2367.4
## - FLAG.Job.Mgr        1    2318.3 2368.3
## - IMP_NINQ            1    2321.4 2371.4
## - FLAG.Job.Sales      1    2331.8 2381.8
## - IMP_CLAGE           1    2349.8 2399.8
## - IMP_DEROG           1    2362.2 2412.2
## - M_DEROG             1    2378.2 2428.2
## - M_VALUE             1    2412.2 2462.2
## - IMP_DEBTINC         1    2415.7 2465.7
## - IMP_DELINQ          1    2540.8 2590.8
## - M_DEBTINC           1    2862.0 2912.0
##
## Step:  AIC=2346.31
## TARGET_BAD_FLAG ~ IMP_MORTDUE + M_MORTDUE + IMP_VALUE + M_VALUE +
##     IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ + M_DELINQ +
##     IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
##     M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office + FLAG.Job.Other +
##     FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.HomeImp
##

```

```

##                               Df Deviance    AIC
## - FLAG.Reason.HomeImp      1    2298.1  2346.1
## <none>                      2296.3  2346.3
## - IMP_CLNO                  1    2299.4  2347.4
## - M_MORTDUE                 1    2300.1  2348.1
## - IMP_YOJ                   1    2301.1  2349.1
## - FLAG.Job.Office           1    2302.1  2350.1
## - IMP_MORTDUE               1    2302.7  2350.7
## - IMP_VALUE                 1    2303.2  2351.2
## - M_CLAGE                   1    2305.3  2353.3
## - M_YOJ                     1    2309.1  2357.1
## - M_DELINQ                  1    2312.1  2360.1
## - FLAG.Job.Self             1    2312.9  2360.9
## - FLAG.Job.ProfExe          1    2313.0  2361.0
## - M_CLNO                    1    2316.6  2364.6
## - FLAG.Job.Other            1    2317.6  2365.6
## - FLAG.Job.Mgr              1    2318.5  2366.5
## - IMP_NINQ                  1    2321.4  2369.4
## - FLAG.Job.Sales            1    2332.3  2380.3
## - IMP_CLAGE                 1    2350.6  2398.6
## - IMP_DEROG                 1    2362.4  2410.4
## - M_DEROG                   1    2378.3  2426.3
## - M_VALUE                   1    2412.4  2460.4
## - IMP_DEBTINC               1    2415.7  2463.7
## - IMP_DELINQ                1    2541.4  2589.4
## - M_DEBTINC                 1    2871.9  2919.9
##
## Step:  AIC=2346.13
## TARGET_BAD_FLAG ~ IMP_MORTDUE + M_MORTDUE + IMP_VALUE + M_VALUE +
##     IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ + M_DELINQ +
##     IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
##     M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office + FLAG.Job.Other +
##     FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self
##
##                               Df Deviance    AIC
## <none>                      2298.1  2346.1
## - IMP_CLNO                  1    2302.1  2348.1
## - M_MORTDUE                 1    2302.7  2348.7
## - IMP_YOJ                   1    2302.8  2348.8
## - FLAG.Job.Office           1    2304.2  2350.2
## - IMP_MORTDUE               1    2304.7  2350.7
## - IMP_VALUE                 1    2305.4  2351.4
## - M_CLAGE                   1    2306.6  2352.6
## - M_YOJ                     1    2312.1  2358.1
## - M_DELINQ                  1    2314.3  2360.3
## - FLAG.Job.ProfExe          1    2315.7  2361.7
## - FLAG.Job.Self             1    2316.0  2362.0
## - M_CLNO                    1    2319.9  2365.9
## - FLAG.Job.Other            1    2320.2  2366.2
## - FLAG.Job.Mgr              1    2320.9  2366.9
## - IMP_NINQ                  1    2322.0  2368.0
## - FLAG.Job.Sales            1    2334.3  2380.3
## - IMP_CLAGE                 1    2352.2  2398.2
## - IMP_DEROG                 1    2364.9  2410.9

```

```
## - M_DEROG      1    2380.8 2426.8
## - M_VALUE      1    2413.7 2459.7
## - IMP_DEBTINC  1    2418.1 2464.1
## - IMP_DELINQ   1    2545.3 2591.3
## - M_DEBTINC    1    2878.7 2924.7
```

```
summary( theUpper_LR2 )
```

```
##
## Call:
## glm(formula = TARGET_BAD_FLAG ~ ., family = "binomial", data = df_flag_train)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.906e+00  6.372e-01 -10.838  < 2e-16 ***
## LOAN          -2.388e-06  5.764e-06  -0.414  0.678730
## IMP_MORTDUE   -5.030e-06  2.068e-06  -2.432  0.014999 *
## M_MORTDUE      4.777e-01  2.365e-01   2.020  0.043426 *
## IMP_VALUE      3.652e-06  1.452e-06   2.515  0.011917 *
## M_VALUE        5.194e+00  6.345e-01   8.185  2.72e-16 ***
## IMP_YOJ       -1.699e-02  8.124e-03  -2.092  0.036471 *
## M_YOJ         -7.661e-01  2.366e-01  -3.238  0.001204 **
## IMP_DEROG      5.919e-01  7.915e-02   7.478  7.56e-14 ***
## M_DEROG       -2.795e+00  3.616e-01  -7.728  1.09e-14 ***
## IMP_DELINQ     8.624e-01  6.641e-02  12.985  < 2e-16 ***
## M_DELINQ      -1.499e+00  4.866e-01  -3.081  0.002066 **
## IMP_CLAGE     -5.624e-03  8.122e-04  -6.925  4.37e-12 ***
## M_CLAGE        1.215e+00  3.973e-01   3.059  0.002219 **
## IMP_NINQ       1.549e-01  3.056e-02   5.069  3.99e-07 ***
## M_NINQ        -1.552e-01  4.633e-01  -0.335  0.737651
## IMP_CLNO      -1.128e-02  6.427e-03  -1.755  0.079255 .
## M_CLNO         3.025e+00  7.364e-01   4.108  3.99e-05 ***
## IMP_DEBTINC    9.616e-02  1.020e-02   9.430  < 2e-16 ***
## M_DEBTINC     2.579e+00  1.141e-01  22.595  < 2e-16 ***
## FLAG.Job.Mgr   1.884e+00  4.676e-01   4.028  5.62e-05 ***
## FLAG.Job.Office 1.005e+00  4.716e-01   2.132  0.033036 *
## FLAG.Job.Other  1.745e+00  4.514e-01   3.865  0.000111 ***
## FLAG.Job.ProfExe 1.636e+00  4.647e-01   3.521  0.000429 ***
## FLAG.Job.Sales  3.094e+00  5.635e-01   5.490  4.02e-08 ***
## FLAG.Job.Self   1.997e+00  5.328e-01   3.748  0.000178 ***
## FLAG.Reason.DebtCon 1.253e-01  3.998e-01   0.313  0.754027
## FLAG.Reason.HomeImp 2.771e-01  4.028e-01   0.688  0.491574
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 4168.9  on 4216  degrees of freedom
## Residual deviance: 2295.9  on 4189  degrees of freedom
## AIC: 2351.9
##
## Number of Fisher Scoring iterations: 6
```



```
summary( lr_model2 )
```

```
##
## Call:
## glm(formula = TARGET_BAD_FLAG ~ IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO +
##     IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##     FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self,
##     family = "binomial", data = df_flag_train)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -6.810e+00  5.813e-01 -11.714  < 2e-16 ***
## IMP_MORTDUE    -5.091e-06  2.054e-06  -2.479  0.013182 *
## M_MORTDUE       5.012e-01  2.324e-01   2.157  0.031036 *
## IMP_VALUE       3.543e-06  1.374e-06   2.578  0.009939 **
## M_VALUE        5.206e+00  6.277e-01   8.294  < 2e-16 ***
## IMP_YOJ        -1.735e-02  8.069e-03  -2.150  0.031523 *
## M_YOJ          -8.276e-01  2.314e-01  -3.576  0.000349 ***
## IMP_DEROG       5.882e-01  7.786e-02   7.555  4.19e-14 ***
## M_DEROG        -2.798e+00  3.607e-01  -7.757  8.66e-15 ***
## IMP_DELINQ      8.685e-01  6.627e-02  13.105  < 2e-16 ***
## M_DELINQ       -1.612e+00  4.335e-01  -3.719  0.000200 ***
## IMP_CLAGE      -5.637e-03  8.075e-04  -6.980  2.95e-12 ***
## M_CLAGE        1.166e+00  3.950e-01   2.953  0.003148 **
## IMP_NINQ        1.495e-01  3.017e-02   4.955  7.23e-07 ***
## IMP_CLNO       -1.256e-02  6.373e-03  -1.972  0.048661 *
## M_CLNO         3.063e+00  6.895e-01   4.442  8.89e-06 ***
## IMP_DEBTINC     9.658e-02  1.018e-02   9.488  < 2e-16 ***
## M_DEBTINC      2.591e+00  1.131e-01  22.914  < 2e-16 ***
## FLAG.Job.Mgr    1.943e+00  4.487e-01   4.330  1.49e-05 ***
## FLAG.Job.Office 1.060e+00  4.534e-01   2.338  0.019392 *
## FLAG.Job.Other  1.815e+00  4.305e-01   4.215  2.50e-05 ***
## FLAG.Job.ProfExe 1.710e+00  4.463e-01   3.832  0.000127 ***
## FLAG.Job.Sales  3.165e+00  5.451e-01   5.806  6.38e-09 ***
## FLAG.Job.Self   2.075e+00  5.135e-01   4.040  5.35e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 4168.9  on 4216  degrees of freedom
## Residual deviance: 2298.1  on 4193  degrees of freedom
## AIC: 2346.1
##
## Number of Fisher Scoring iterations: 6

pla2 = predict( theUpper_LR2, df_flag_test, type = "response" )
pla22 = prediction( pla2, df_flag_test$TARGET_BAD_FLAG )
pla23 = performance( pla22, "tpr", "fpr" )

plr2 = predict( lr_model2, df_flag_test, type = "response" )
```

```

plr22 = prediction( plr2, df_flag_test$TARGET_BAD_FLAG )
plr23 = performance( plr22, "tpr", "fpr" )

#I choose Linear Reg backward model to predict TARGET_LOSS_AMT

df_amt_2 = subset( df, TARGET_BAD_FLAG == 1)
df_amt_2$TARGET_BAD_FLAG = NULL
head(df_amt_2)

```

```

##      TARGET_LOSS_AMT LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE IMP_YOJ M_YOJ
## 1           641 1100      25860      0      39025      0      10.5      0
## 2           1109 1300      70053      0      68400      0       7.0      0
## 3           767 1500      13500      0      16700      0       4.0      0
## 4          1425 1500      65000      1      89000      1       7.0      1
## 6           335 1700      30548      0      40320      0       9.0      0
## 7          1841 1800      48649      0      57037      0       5.0      0
##      IMP_DEROG M_DEROG IMP_DELIQ M_DELIQ IMP_CLAGE M_CLAGE IMP_NINQ M_NINQ
## 1           0      0      0      0  94.36667      0      1      0
## 2           0      0      2      0 121.83333      0      0      0
## 3           0      0      0      0 149.46667      0      1      0
## 4           1      1      1      1 174.00000      1      1      1
## 6           0      0      0      0 101.46600      0      1      0
## 7           3      0      2      0  77.10000      0      1      0
##      IMP_CLNO M_CLNO IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr FLAG.Job.Office
## 1           9      0    35.00000      1      0      0
## 2          14      0    35.00000      1      0      0
## 3          10      0    35.00000      1      0      0
## 4          20      1    35.00000      1      0      0
## 6           8      0    37.11361      0      0      0
## 7          17      0    35.00000      1      0      0
##      FLAG.Job.Other FLAG.Job.ProfExe FLAG.Job.Sales FLAG.Job.Self
## 1           1      0      0      0
## 2           1      0      0      0
## 3           1      0      0      0
## 4           0      0      0      0
## 6           1      0      0      0
## 7           1      0      0      0
##      FLAG.Reason.DebtCon FLAG.Reason.HomeImp
## 1           0      1
## 2           0      1
## 3           0      1
## 4           0      0
## 6           0      1
## 7           0      1

```

```

FLAG = sample( c( TRUE, FALSE ), nrow(df_amt_2),
               replace = TRUE, prob = c(0.7,0.3) )
df_amt_train = df_amt_2[FLAG, ]
df_amt_test = df_amt_2[!FLAG, ]

mean( df_amt_2$TARGET_LOSS_AMT )

```

```
## [1] 13414.58
```

```
mean( df_amt_train$TARGET_LOSS_AMT )
```

```
## [1] 13387.46
```

```
mean( df_amt_test$TARGET_LOSS_AMT )
```

```
## [1] 13472.3
```

```
#Linear Reg Backward Model
```

```
theUpper_LR3 = lm( TARGET_LOSS_AMT ~ ., data = df_amt_train )
```

```
theLower_LR3 = lm( TARGET_LOSS_AMT ~ 1, data = df_amt_train )
```

```
summary( theUpper_LR3 )
```

```
##
```

```
## Call:
```

```
## lm(formula = TARGET_LOSS_AMT ~ ., data = df_amt_train)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -19462.6   -998.1     93.0   1582.0  13428.2
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)   -9.605e+03  1.320e+03  -7.278 8.24e-13 ***  
## LOAN           8.192e-01  1.393e-02  58.812 < 2e-16 ***  
## IMP_MORTDUE    1.129e-02  3.947e-03   2.860 0.004352 **  
## M_MORTDUE     -8.853e+02  4.906e+02  -1.805 0.071519 .  
## IMP_VALUE     -1.519e-02  2.780e-03  -5.464 6.26e-08 ***  
## M_VALUE      -1.981e+02  4.783e+02  -0.414 0.678824  
## IMP_YOJ       -9.106e+01  1.891e+01  -4.815 1.77e-06 ***  
## M_YOJ         5.097e+02  6.124e+02   0.832 0.405484  
## IMP_DEROG     3.112e+02  9.286e+01   3.351 0.000844 ***  
## M_DEROG       1.272e+03  7.620e+02   1.670 0.095414 .  
## IMP_DELINQ    7.746e+02  7.739e+01  10.009 < 2e-16 ***  
## M_DELINQ     -4.642e+02  9.935e+02  -0.467 0.640455  
## IMP_CLAGE     -1.861e+01  1.495e+00 -12.446 < 2e-16 ***  
## M_CLAGE      -4.802e+03  9.325e+02  -5.150 3.30e-07 ***  
## IMP_NINQ     -3.484e+01  5.823e+01  -0.598 0.549810  
## M_NINQ       -1.769e+03  9.937e+02  -1.780 0.075479 .  
## IMP_CLNO      1.962e+02  1.302e+01  15.065 < 2e-16 ***  
## M_CLNO       6.659e+03  1.821e+03   3.657 0.000272 ***  
## IMP_DEBTINC   1.060e+02  1.228e+01   8.631 < 2e-16 ***  
## M_DEBTINC     5.391e+03  2.743e+02  19.659 < 2e-16 ***  
## FLAG.Job.Mgr  -5.813e+02  1.120e+03  -0.519 0.604004  
## FLAG.Job.Office -3.497e+02  1.142e+03  -0.306 0.759510  
## FLAG.Job.Other -2.137e+02  1.085e+03  -0.197 0.843874  
## FLAG.Job.ProfExe -9.517e+02  1.118e+03  -0.852 0.394717  
## FLAG.Job.Sales  1.098e+03  1.271e+03   0.864 0.388044  
## FLAG.Job.Self  2.221e+03  1.216e+03   1.826 0.068192 .  
## FLAG.Reason.DebtCon 1.718e+03  6.626e+02   2.593 0.009691 **  
## FLAG.Reason.HomeImp 4.705e+02  6.848e+02   0.687 0.492256
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3505 on 781 degrees of freedom
## Multiple R-squared:  0.8993, Adjusted R-squared:  0.8958
## F-statistic: 258.3 on 27 and 781 DF,  p-value: < 2.2e-16

summary( theLower_LR3 )

##
## Call:
## lm(formula = TARGET_LOSS_AMT ~ 1, data = df_amt_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13067  -7687  -2401   4175   65600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13387.5      381.7    35.07  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10860 on 808 degrees of freedom

lr_model3 = stepAIC( theUpper_LR3, direction = "backward",
                    scope = list(lower = theLower_LR3, upper = theUpper_LR3))

## Start:  AIC=13233.42
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##      M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##      M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO +
##      M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##      FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##      FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
##              Df Sum of Sq      RSS   AIC
## - FLAG.Job.Other    1 4.7673e+05 9.5937e+09 13232
## - FLAG.Job.Office    1 1.1519e+06 9.5944e+09 13232
## - M_VALUE            1 2.1076e+06 9.5953e+09 13232
## - M_DELINQ           1 2.6817e+06 9.5959e+09 13232
## - FLAG.Job.Mgr       1 3.3069e+06 9.5965e+09 13232
## - IMP_NINQ           1 4.3970e+06 9.5976e+09 13232
## - FLAG.Reason.HomeImp 1 5.7981e+06 9.5990e+09 13232
## - M_YOJ              1 8.5093e+06 9.6017e+09 13232
## - FLAG.Job.ProfExe   1 8.9073e+06 9.6021e+09 13232
## - FLAG.Job.Sales     1 9.1620e+06 9.6024e+09 13232
## <none>                                9.5932e+09 13233
## - M_DEROG           1 3.4237e+07 9.6275e+09 13234
## - M_NINQ             1 3.8914e+07 9.6321e+09 13235
## - M_MORTDUE          1 4.0002e+07 9.6332e+09 13235
## - FLAG.Job.Self      1 4.0968e+07 9.6342e+09 13235
## - FLAG.Reason.DebtCon 1 8.2591e+07 9.6758e+09 13238
```

```

## - IMP_MORTDUE      1 1.0046e+08 9.6937e+09 13240
## - IMP_DEROG        1 1.3794e+08 9.7312e+09 13243
## - M_CLNO           1 1.6431e+08 9.7575e+09 13245
## - IMP_YOJ          1 2.8476e+08 9.8780e+09 13255
## - M_CLAGE          1 3.2575e+08 9.9190e+09 13258
## - IMP_VALUE        1 3.6675e+08 9.9600e+09 13262
## - IMP_DEBTINC      1 9.1506e+08 1.0508e+10 13305
## - IMP_DELINQ       1 1.2304e+09 1.0824e+10 13329
## - IMP_CLAGE        1 1.9027e+09 1.1496e+10 13378
## - IMP_CLNO         1 2.7878e+09 1.2381e+10 13438
## - M_DEBTINC        1 4.7471e+09 1.4340e+10 13557
## - LOAN             1 4.2486e+10 5.2079e+10 14600
##
## Step: AIC=13231.46
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO +
##     M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##     FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.DebtCon +
##     FLAG.Reason.HomeImp
##
##              Df Sum of Sq      RSS   AIC
## - FLAG.Job.Office      1 1.3522e+06 9.5951e+09 13230
## - M_VALUE              1 1.9036e+06 9.5956e+09 13230
## - M_DELINQ             1 2.7492e+06 9.5965e+09 13230
## - IMP_NINQ             1 4.3475e+06 9.5981e+09 13230
## - FLAG.Reason.HomeImp  1 5.3732e+06 9.5991e+09 13230
## - M_YOJ               1 9.3374e+06 9.6030e+09 13230
## - FLAG.Job.Mgr         1 1.2286e+07 9.6060e+09 13230
## <none>                  9.5937e+09 13232
## - M_DEROG             1 3.4167e+07 9.6279e+09 13232
## - M_NINQ              1 3.9517e+07 9.6332e+09 13233
## - M_MORTDUE           1 4.0075e+07 9.6338e+09 13233
## - FLAG.Job.Sales      1 4.4993e+07 9.6387e+09 13233
## - FLAG.Job.ProfExe    1 4.9474e+07 9.6432e+09 13234
## - FLAG.Reason.DebtCon 1 8.3066e+07 9.6768e+09 13236
## - IMP_MORTDUE         1 1.0046e+08 9.6942e+09 13238
## - IMP_DEROG           1 1.4179e+08 9.7355e+09 13241
## - M_CLNO              1 1.6995e+08 9.7636e+09 13244
## - FLAG.Job.Self       1 1.8119e+08 9.7749e+09 13245
## - IMP_YOJ             1 2.8715e+08 9.8809e+09 13253
## - M_CLAGE             1 3.2574e+08 9.9194e+09 13256
## - IMP_VALUE           1 3.6636e+08 9.9601e+09 13260
## - IMP_DEBTINC         1 9.1459e+08 1.0508e+10 13303
## - IMP_DELINQ          1 1.2302e+09 1.0824e+10 13327
## - IMP_CLAGE           1 1.9046e+09 1.1498e+10 13376
## - IMP_CLNO            1 2.7880e+09 1.2382e+10 13436
## - M_DEBTINC           1 4.7497e+09 1.4343e+10 13555
## - LOAN                1 4.2487e+10 5.2080e+10 14598
##
## Step: AIC=13229.57
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO +

```

```

##      M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.ProfExe +
##      FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
##              Df  Sum of Sq      RSS   AIC
## - M_VALUE      1 1.9829e+06 9.5970e+09 13228
## - M_DELINQ      1 2.7378e+06 9.5978e+09 13228
## - IMP_NINQ      1 4.2839e+06 9.5993e+09 13228
## - FLAG.Reason.HomeImp 1 5.0352e+06 9.6001e+09 13228
## - M_YOJ        1 1.0266e+07 9.6053e+09 13228
## - FLAG.Job.Mgr  1 1.1022e+07 9.6061e+09 13228
## <none>                                9.5951e+09 13230
## - M_DEROG      1 3.4487e+07 9.6295e+09 13230
## - M_MORTDUE     1 3.9103e+07 9.6342e+09 13231
## - M_NINQ        1 3.9770e+07 9.6348e+09 13231
## - FLAG.Job.Sales 1 4.8038e+07 9.6431e+09 13232
## - FLAG.Job.ProfExe 1 4.8583e+07 9.6436e+09 13232
## - FLAG.Reason.DebtCon 1 8.1896e+07 9.6770e+09 13234
## - IMP_MORTDUE   1 9.9742e+07 9.6948e+09 13236
## - IMP_DEROG     1 1.4208e+08 9.7371e+09 13240
## - M_CLNO        1 1.6947e+08 9.7645e+09 13242
## - FLAG.Job.Self  1 1.8944e+08 9.7845e+09 13243
## - IMP_YOJ        1 2.8714e+08 9.8822e+09 13251
## - M_CLAGE       1 3.2597e+08 9.9210e+09 13255
## - IMP_VALUE     1 3.6598e+08 9.9610e+09 13258
## - IMP_DEBTINC   1 9.1471e+08 1.0510e+10 13301
## - IMP_DELINQ    1 1.2303e+09 1.0825e+10 13325
## - IMP_CLAGE     1 1.9051e+09 1.1500e+10 13374
## - IMP_CLNO      1 2.7867e+09 1.2382e+10 13434
## - M_DEBTINC     1 4.7536e+09 1.4349e+10 13553
## - LOAN          1 4.2547e+10 5.2142e+10 14597
##
## Step:  AIC=13227.74
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##      IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ + M_DELINQ +
##      IMP_CLAGE + M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO + M_CLNO +
##      IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.ProfExe +
##      FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
##              Df  Sum of Sq      RSS   AIC
## - M_DELINQ      1 3.3535e+06 9.6004e+09 13226
## - IMP_NINQ      1 3.8455e+06 9.6009e+09 13226
## - FLAG.Reason.HomeImp 1 6.0617e+06 9.6031e+09 13226
## - M_YOJ        1 1.0576e+07 9.6076e+09 13227
## - FLAG.Job.Mgr  1 1.1336e+07 9.6084e+09 13227
## <none>                                9.5970e+09 13228
## - M_DEROG      1 3.3412e+07 9.6305e+09 13228
## - M_NINQ        1 3.8991e+07 9.6360e+09 13229
## - M_MORTDUE     1 4.0535e+07 9.6376e+09 13229
## - FLAG.Job.ProfExe 1 4.8169e+07 9.6452e+09 13230
## - FLAG.Job.Sales 1 4.8973e+07 9.6460e+09 13230
## - FLAG.Reason.DebtCon 1 8.7495e+07 9.6845e+09 13233
## - IMP_MORTDUE   1 9.8128e+07 9.6952e+09 13234
## - IMP_DEROG     1 1.4046e+08 9.7375e+09 13238
## - M_CLNO        1 1.7122e+08 9.7683e+09 13240

```

```

## - FLAG.Job.Self          1 1.8894e+08 9.7860e+09 13242
## - IMP_YOJ                1 2.9238e+08 9.8894e+09 13250
## - M_CLAGE                1 3.2413e+08 9.9212e+09 13253
## - IMP_VALUE              1 3.6473e+08 9.9618e+09 13256
## - IMP_DEBTINC            1 9.2498e+08 1.0522e+10 13300
## - IMP_DELINQ             1 1.2306e+09 1.0828e+10 13323
## - IMP_CLAGE              1 1.9070e+09 1.1504e+10 13372
## - IMP_CLNO               1 2.7918e+09 1.2389e+10 13432
## - M_DEBTINC              1 4.7524e+09 1.4349e+10 13551
## - LOAN                   1 4.3032e+10 5.2629e+10 14602
##
## Step: AIC=13226.02
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ + IMP_CLAGE +
##     M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
##     M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.ProfExe + FLAG.Job.Sales +
##     FLAG.Job.Self + FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
##              Df Sum of Sq      RSS   AIC
## - IMP_NINQ          1 3.7015e+06 9.6041e+09 13224
## - FLAG.Reason.HomeImp 1 6.0991e+06 9.6065e+09 13224
## - M_YOJ              1 1.0302e+07 9.6107e+09 13225
## - FLAG.Job.Mgr       1 1.1251e+07 9.6116e+09 13225
## <none>                9.6004e+09 13226
## - M_DEROG           1 3.0083e+07 9.6305e+09 13227
## - M_NINQ             1 3.8027e+07 9.6384e+09 13227
## - M_MORTDUE          1 4.2159e+07 9.6426e+09 13228
## - FLAG.Job.Sales     1 4.6273e+07 9.6467e+09 13228
## - FLAG.Job.ProfExe   1 4.8081e+07 9.6485e+09 13228
## - FLAG.Reason.DebtCon 1 8.7280e+07 9.6877e+09 13231
## - IMP_MORTDUE        1 9.8621e+07 9.6990e+09 13232
## - IMP_DEROG          1 1.3933e+08 9.7397e+09 13236
## - M_CLNO             1 1.8816e+08 9.7886e+09 13240
## - FLAG.Job.Self      1 1.8832e+08 9.7887e+09 13240
## - IMP_YOJ            1 2.9098e+08 9.8914e+09 13248
## - M_CLAGE            1 3.2280e+08 9.9232e+09 13251
## - IMP_VALUE          1 3.6971e+08 9.9701e+09 13255
## - IMP_DEBTINC        1 9.2869e+08 1.0529e+10 13299
## - IMP_DELINQ         1 1.2383e+09 1.0839e+10 13322
## - IMP_CLAGE          1 1.9071e+09 1.1507e+10 13371
## - IMP_CLNO           1 2.8039e+09 1.2404e+10 13431
## - M_DEBTINC          1 4.7559e+09 1.4356e+10 13550
## - LOAN               1 4.3217e+10 5.2817e+10 14603
##
## Step: AIC=13224.33
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ + IMP_CLAGE +
##     M_CLAGE + M_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC + M_DEBTINC +
##     FLAG.Job.Mgr + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
##              Df Sum of Sq      RSS   AIC
## - FLAG.Reason.HomeImp 1 6.4099e+06 9.6105e+09 13223
## - M_YOJ              1 1.1003e+07 9.6151e+09 13223

```

```

## - FLAG.Job.Mgr          1 1.1328e+07 9.6154e+09 13223
## <none>                  9.6041e+09 13224
## - M_DEROG              1 3.0406e+07 9.6345e+09 13225
## - M_NINQ               1 3.7077e+07 9.6412e+09 13225
## - M_MORTDUE            1 4.2848e+07 9.6469e+09 13226
## - FLAG.Job.Sales       1 4.6649e+07 9.6507e+09 13226
## - FLAG.Job.ProfExe     1 4.7716e+07 9.6518e+09 13226
## - FLAG.Reason.DebtCon  1 8.6965e+07 9.6911e+09 13230
## - IMP_MORTDUE          1 9.6944e+07 9.7010e+09 13230
## - IMP_DEROG            1 1.3638e+08 9.7405e+09 13234
## - M_CLNO               1 1.8602e+08 9.7901e+09 13238
## - FLAG.Job.Self        1 1.8783e+08 9.7919e+09 13238
## - IMP_YOJ              1 2.8919e+08 9.8933e+09 13246
## - M_CLAGE              1 3.1960e+08 9.9237e+09 13249
## - IMP_VALUE            1 3.6750e+08 9.9716e+09 13253
## - IMP_DEBTINC          1 9.3696e+08 1.0541e+10 13298
## - IMP_DELINQ           1 1.2442e+09 1.0848e+10 13321
## - IMP_CLAGE            1 1.9175e+09 1.1522e+10 13370
## - IMP_CLNO             1 2.8037e+09 1.2408e+10 13430
## - M_DEBTINC            1 4.7584e+09 1.4362e+10 13548
## - LOAN                 1 4.3435e+10 5.3039e+10 14605
##
## Step:  AIC=13222.87
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ + IMP_CLAGE +
##     M_CLAGE + M_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC + M_DEBTINC +
##     FLAG.Job.Mgr + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.DebtCon
##
##              Df  Sum of Sq      RSS   AIC
## - M_YOJ        1 1.0691e+07 9.6212e+09 13222
## - FLAG.Job.Mgr  1 1.1177e+07 9.6217e+09 13222
## <none>          9.6105e+09 13223
## - M_DEROG      1 3.1600e+07 9.6421e+09 13224
## - M_NINQ       1 3.8771e+07 9.6493e+09 13224
## - M_MORTDUE    1 4.1200e+07 9.6517e+09 13224
## - FLAG.Job.Sales 1 4.7489e+07 9.6580e+09 13225
## - FLAG.Job.ProfExe 1 4.7556e+07 9.6581e+09 13225
## - IMP_MORTDUE   1 9.5986e+07 9.7065e+09 13229
## - IMP_DEROG     1 1.3483e+08 9.7453e+09 13232
## - M_CLNO        1 1.8517e+08 9.7957e+09 13236
## - FLAG.Job.Self  1 1.8679e+08 9.7973e+09 13236
## - FLAG.Reason.DebtCon 1 2.6153e+08 9.8720e+09 13243
## - IMP_YOJ       1 2.9087e+08 9.9014e+09 13245
## - M_CLAGE       1 3.2048e+08 9.9310e+09 13247
## - IMP_VALUE     1 3.6329e+08 9.9738e+09 13251
## - IMP_DEBTINC   1 9.3520e+08 1.0546e+10 13296
## - IMP_DELINQ    1 1.2389e+09 1.0849e+10 13319
## - IMP_CLAGE     1 1.9126e+09 1.1523e+10 13368
## - IMP_CLNO      1 2.8097e+09 1.2420e+10 13428
## - M_DEBTINC     1 4.7521e+09 1.4363e+10 13546
## - LOAN          1 4.3712e+10 5.3322e+10 14607
##
## Step:  AIC=13221.77

```



```

## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     IMP_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ + IMP_CLAGE +
##     M_CLAGE + M_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC + M_DEBTINC +
##     FLAG.Job.Mgr + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.DebtCon
##
##           Df  Sum of Sq      RSS   AIC
## - FLAG.Job.Mgr      1 1.3027e+07 9.6342e+09 13221
## <none>                                9.6212e+09 13222
## - M_DEROG           1 3.0704e+07 9.6519e+09 13222
## - M_NINQ            1 3.1943e+07 9.6531e+09 13222
## - M_MORTDUE         1 3.5895e+07 9.6571e+09 13223
## - FLAG.Job.Sales    1 4.5171e+07 9.6664e+09 13224
## - FLAG.Job.ProfExe  1 5.1604e+07 9.6728e+09 13224
## - IMP_MORTDUE       1 9.6476e+07 9.7177e+09 13228
## - IMP_DEROG         1 1.3361e+08 9.7548e+09 13231
## - M_CLNO            1 1.8230e+08 9.8035e+09 13235
## - FLAG.Job.Self     1 1.9256e+08 9.8138e+09 13236
## - FLAG.Reason.DebtCon 1 2.6453e+08 9.8857e+09 13242
## - IMP_YOJ           1 2.9276e+08 9.9140e+09 13244
## - M_CLAGE           1 3.2660e+08 9.9478e+09 13247
## - IMP_VALUE         1 3.7093e+08 9.9921e+09 13250
## - IMP_DEBTINC       1 9.3356e+08 1.0555e+10 13295
## - IMP_DELINQ        1 1.2368e+09 1.0858e+10 13318
## - IMP_CLAGE         1 1.9030e+09 1.1524e+10 13366
## - IMP_CLNO          1 2.8032e+09 1.2424e+10 13427
## - M_DEBTINC         1 4.7461e+09 1.4367e+10 13544
## - LOAN              1 4.3788e+10 5.3410e+10 14606
##
## Step:  AIC=13220.87
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     IMP_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ + IMP_CLAGE +
##     M_CLAGE + M_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC + M_DEBTINC +
##     FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.DebtCon
##
##           Df  Sum of Sq      RSS   AIC
## <none>                                9.6342e+09 13221
## - M_DEROG           1 3.2307e+07 9.6665e+09 13222
## - M_MORTDUE         1 3.2980e+07 9.6672e+09 13222
## - M_NINQ            1 3.4339e+07 9.6686e+09 13222
## - FLAG.Job.ProfExe  1 4.1451e+07 9.6757e+09 13222
## - FLAG.Job.Sales    1 5.2870e+07 9.6871e+09 13223
## - IMP_MORTDUE       1 8.9604e+07 9.7238e+09 13226
## - IMP_DEROG         1 1.3168e+08 9.7659e+09 13230
## - M_CLNO            1 1.8678e+08 9.8210e+09 13234
## - FLAG.Job.Self     1 2.1093e+08 9.8451e+09 13236
## - FLAG.Reason.DebtCon 1 2.6439e+08 9.8986e+09 13241
## - IMP_YOJ           1 2.9959e+08 9.9338e+09 13244
## - M_CLAGE           1 3.3063e+08 9.9649e+09 13246
## - IMP_VALUE         1 3.7121e+08 1.0005e+10 13250
## - IMP_DEBTINC       1 9.3758e+08 1.0572e+10 13294
## - IMP_DELINQ        1 1.2282e+09 1.0862e+10 13316
## - IMP_CLAGE         1 1.8940e+09 1.1528e+10 13364
## - IMP_CLNO          1 2.7909e+09 1.2425e+10 13425

```

```
## - M_DEBTINC          1 4.7434e+09 1.4378e+10 13543
## - LOAN                1 4.3923e+10 5.3558e+10 14607
```

```
summary( lr_model3 )
```

```
##
## Call:
## lm(formula = TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE +
##     IMP_VALUE + IMP_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     IMP_CLAGE + M_CLAGE + M_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
##     M_DEBTINC + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.DebtCon, data = df_amt_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19948  -1036    120    1590   13496
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -9.429e+03  6.374e+02 -14.793 < 2e-16 ***
## LOAN           8.182e-01  1.364e-02  59.976 < 2e-16 ***
## IMP_MORTDUE     1.053e-02  3.887e-03   2.709  0.00690 **
## M_MORTDUE      -7.845e+02  4.774e+02  -1.643  0.10069
## IMP_VALUE      -1.508e-02  2.735e-03  -5.514  4.76e-08 ***
## IMP_YOJ        -9.275e+01  1.873e+01  -4.953  8.93e-07 ***
## IMP_DEROG       2.996e+02  9.125e+01   3.284  0.00107 **
## M_DEROG        1.176e+03  7.229e+02   1.627  0.10422
## IMP_DELINQ      7.696e+02  7.673e+01  10.029 < 2e-16 ***
## IMP_CLAGE      -1.836e+01  1.474e+00 -12.454 < 2e-16 ***
## M_CLAGE        -4.810e+03  9.244e+02  -5.204  2.49e-07 ***
## M_NINQ         -1.609e+03  9.597e+02  -1.677  0.09395 .
## IMP_CLNO        1.935e+02  1.280e+01  15.118 < 2e-16 ***
## M_CLNO         6.264e+03  1.602e+03   3.911  9.98e-05 ***
## IMP_DEBTINC     1.057e+02  1.206e+01   8.763 < 2e-16 ***
## M_DEBTINC       5.357e+03  2.718e+02  19.709 < 2e-16 ***
## FLAG.Job.ProfExe -6.301e+02  3.420e+02  -1.842  0.06578 .
## FLAG.Job.Sales   1.373e+03  6.600e+02   2.081  0.03777 *
## FLAG.Job.Self     2.560e+03  6.160e+02   4.156  3.59e-05 ***
## FLAG.Reason.DebtCon 1.298e+03  2.790e+02   4.653  3.83e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3494 on 789 degrees of freedom
## Multiple R-squared:  0.8989, Adjusted R-squared:  0.8964
## F-statistic: 369 on 19 and 789 DF, p-value: < 2.2e-16
```

```
plr3 = predict( lr_model3, df_amt_test )
head(plr3)
```

```
##           9           13           15           18           19           21
## 504.2771 3887.5131 1694.7098 -7333.8501 2877.3824 2165.0179
```

```
RMSElr3 = sqrt( mean( ( df_amt_test$TARGET_LOSS_AMT - plr3 )^2 ) )

summary( lr_model2 )
```

```
##
## Call:
## glm(formula = TARGET_BAD_FLAG ~ IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO +
##     IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##     FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self,
##     family = "binomial", data = df_flag_train)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -6.810e+00  5.813e-01 -11.714 < 2e-16 ***
## IMP_MORTDUE    -5.091e-06  2.054e-06  -2.479 0.013182 *
## M_MORTDUE       5.012e-01  2.324e-01   2.157 0.031036 *
## IMP_VALUE       3.543e-06  1.374e-06   2.578 0.009939 **
## M_VALUE        5.206e+00  6.277e-01   8.294 < 2e-16 ***
## IMP_YOJ        -1.735e-02  8.069e-03  -2.150 0.031523 *
## M_YOJ          -8.276e-01  2.314e-01  -3.576 0.000349 ***
## IMP_DEROG       5.882e-01  7.786e-02   7.555 4.19e-14 ***
## M_DEROG        -2.798e+00  3.607e-01  -7.757 8.66e-15 ***
## IMP_DELINQ      8.685e-01  6.627e-02  13.105 < 2e-16 ***
## M_DELINQ       -1.612e+00  4.335e-01  -3.719 0.000200 ***
## IMP_CLAGE      -5.637e-03  8.075e-04  -6.980 2.95e-12 ***
## M_CLAGE        1.166e+00  3.950e-01   2.953 0.003148 **
## IMP_NINQ        1.495e-01  3.017e-02   4.955 7.23e-07 ***
## IMP_CLNO       -1.256e-02  6.373e-03  -1.972 0.048661 *
## M_CLNO         3.063e+00  6.895e-01   4.442 8.89e-06 ***
## IMP_DEBTINC     9.658e-02  1.018e-02   9.488 < 2e-16 ***
## M_DEBTINC      2.591e+00  1.131e-01  22.914 < 2e-16 ***
## FLAG.Job.Mgr    1.943e+00  4.487e-01   4.330 1.49e-05 ***
## FLAG.Job.Office 1.060e+00  4.534e-01   2.338 0.019392 *
## FLAG.Job.Other  1.815e+00  4.305e-01   4.215 2.50e-05 ***
## FLAG.Job.ProfExe 1.710e+00  4.463e-01   3.832 0.000127 ***
## FLAG.Job.Sales  3.165e+00  5.451e-01   5.806 6.38e-09 ***
## FLAG.Job.Self   2.075e+00  5.135e-01   4.040 5.35e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 4168.9  on 4216  degrees of freedom
## Residual deviance: 2298.1  on 4193  degrees of freedom
## AIC: 2346.1
##
## Number of Fisher Scoring iterations: 6
```

```
summary( lr_model3 )
```

```
##
```

```
## Call:
## lm(formula = TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE +
##     IMP_VALUE + IMP_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     IMP_CLAGE + M_CLAGE + M_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
##     M_DEBTINC + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.DebtCon, data = df_amt_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19948  -1036    120    1590   13496
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -9.429e+03  6.374e+02 -14.793 < 2e-16 ***
## LOAN           8.182e-01  1.364e-02  59.976 < 2e-16 ***
## IMP_MORTDUE     1.053e-02  3.887e-03   2.709  0.00690 **
## M_MORTDUE      -7.845e+02  4.774e+02  -1.643  0.10069
## IMP_VALUE      -1.508e-02  2.735e-03  -5.514  4.76e-08 ***
## IMP_YOJ        -9.275e+01  1.873e+01  -4.953  8.93e-07 ***
## IMP_DEROG       2.996e+02  9.125e+01   3.284  0.00107 **
## M_DEROG        1.176e+03  7.229e+02   1.627  0.10422
## IMP_DELINQ      7.696e+02  7.673e+01  10.029 < 2e-16 ***
## IMP_CLAGE      -1.836e+01  1.474e+00 -12.454 < 2e-16 ***
## M_CLAGE        -4.810e+03  9.244e+02  -5.204  2.49e-07 ***
## M_NINQ         -1.609e+03  9.597e+02  -1.677  0.09395 .
## IMP_CLNO        1.935e+02  1.280e+01  15.118 < 2e-16 ***
## M_CLNO         6.264e+03  1.602e+03   3.911  9.98e-05 ***
## IMP_DEBTINC     1.057e+02  1.206e+01   8.763 < 2e-16 ***
## M_DEBTINC       5.357e+03  2.718e+02  19.709 < 2e-16 ***
## FLAG.Job.ProfExe -6.301e+02  3.420e+02  -1.842  0.06578 .
## FLAG.Job.Sales   1.373e+03  6.600e+02   2.081  0.03777 *
## FLAG.Job.Self    2.560e+03  6.160e+02   4.156  3.59e-05 ***
## FLAG.Reason.DebtCon 1.298e+03  2.790e+02   4.653  3.83e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3494 on 789 degrees of freedom
## Multiple R-squared:  0.8989, Adjusted R-squared:  0.8964
## F-statistic: 369 on 19 and 789 DF, p-value: < 2.2e-16
```

```
p_loss_amt = plr2 * plr3
```

```
## Warning in plr2 * plr3: longer object length is not a multiple of shorter
## object length
```

```
head(p_loss_amt)
```

```
##           1           5           10           11           14           15
## 290.3101 1347.9192 1415.6322 -7249.1379 1375.0521 1469.0758
```

```
RMSE2 = sqrt( mean( (df$TARGET_LOSS_AMT - p_loss_amt )^2 ))
```

```
## Warning in df$TARGET_LOSS_AMT ~ p_loss_amt: longer object length is not a  
## multiple of shorter object length
```

```
print(RMSE2)
```

```
## [1] 8998.083
```

```
#I suggest the model of step 3. Step 4 has a very large RMSE over 8000.
```