Loan Analysis for Lending Club Platform

Olya Fomicheva

3/20/2019

For my project I selected the data set that I found on Lending Club's website (https://www.lendingclub.com). The data is provided for potential investors. The data set contains information about loans that were issued from 2007 to the third quarter of 2017.

Lending Club is the world's largest peer-to-peer lending platform that enables borrowers to obtain a loan, and investors to purchase notes backed by payments made on loans.

The goals of the project are

1. To find the equation that best predicts the probability of weather the load will be paid off or not.
2. To understand what might might cause the probability to change.
3. Find the classifier that can predict whether the loan will be paid off or not with higher accuracy

An investor earns money when loan is fully paid of and loses money when loan is charged off. If an investor obtains the results generated by the model that classify loans he would be able to make better investment decisions.

While I was reviewing Landing Club's website I found out that investors can see the information such as loan rate, loan term, interest rate, borrower's FICO score, loan amount and loan purpose. Moreover, they have an ability to filter by borrower's employment length and monthly income.

In order to collect the data I downloaded (data source: https://www.lendingclub.com/info/download-data.action) and merged 11 files that contain data from 2007 to the third quarter of 2017. To reduce the loading time I implemented the following steps.

```
#1. read in a few records of the input file to identify the classes of the input file and assign that c
data_2007_2011 <- read.csv(file="https://cdn-stage.fedweb.org/fed-2/13/LoanStats3a.csv",
                           stringsAsFactors=T, header=T, nrows=5)

data_2012_2013 <- read.csv(file="https://cdn-stage.fedweb.org/fed-2/13/LoanStats3b.csv",
                           stringsAsFactors=T, header=T, nrows=5)

data_2014 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats3c.csv",
                           stringsAsFactors=T, header=T, nrows=5)

data_2015 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats3d.csv",
                           stringsAsFactors=T, header=T, nrows=5)

data_2016_q1 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2016Q1.csv",
                          stringsAsFactors=T, header=T, nrows=5)

data_2016_q2 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2016Q2.csv",
                          stringsAsFactors=T, header=T, nrows=5)

data_2016_q3 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2016Q3.csv",
                          stringsAsFactors=T, header=T, nrows=5)

data_2016_q4 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2016Q4.csv",
                          stringsAsFactors=T, header=T, nrows=5)
```

```
data_2017_q1 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2017Q1.csv",
                          stringsAsFactors=T, header=T, nrows=5)

data_2017_q2 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2017Q2.csv",
                          stringsAsFactors=T, header=T, nrows=5)

data_2017_q3 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2017Q3.csv",
                          stringsAsFactors=T, header=T, nrows=5)


#2. replace all missing values with NAs
data_2007_2011 <- data_2007_2011[is.na(data_2007_2011)]
data_2012_2013 <- data_2012_2013[is.na(data_2012_2013)]
data_2014 <- data_2014[is.na(data_2014)]
data_2015 <- data_2015[is.na(data_2015)]
data_2016_q1 <- data_2016_q1[is.na(data_2016_q1)]
data_2016_q2 <- data_2016_q1[is.na(data_2016_q2)]
data_2016_q3 <- data_2016_q1[is.na(data_2016_q3)]
data_2016_q4 <- data_2016_q1[is.na(data_2016_q4)]
data_2017_q1 <- data_2017_q1[is.na(data_2017_q1)]
data_2017_q2 <- data_2017_q2[is.na(data_2017_q2)]
data_2017_q3 <- data_2017_q3[is.na(data_2017_q3)]


#3. determine classes
data_2007_2011.colclass <- sapply(data_2007_2011,class)
data_2012_2013.colclass <- sapply(data_2012_2013,class)
data_2014.colclass <- sapply(data_2014,class)
data_2015.colclass <- sapply(data_2015,class)
data_2016_q1.colclass <- sapply(data_2016_q1,class)
data_2016_q2.colclass <- sapply(data_2016_q2,class)
data_2016_q3.colclass <- sapply(data_2016_q3,class)
data_2016_q4.colclass <- sapply(data_2016_q4,class)
data_2017_q1.colclass <- sapply(data_2017_q1,class)
data_2017_q2.colclass <- sapply(data_2017_q2,class)
data_2017_q3.colclass <- sapply(data_2017_q3,class)


#4. assign that column class to the input file while reading the entire data set and define comment.cha
data_2007_2011 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats3a.csv",
                           stringsAsFactors=T,
                           header=T,colClasses=data_2007_2011.colclass, comment.char="")

data_2012_2013 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats3b.csv",
                           stringsAsFactors=T,
                           header=T,colClasses=data_2007_2011.colclass, comment.char="")

data_2014 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats3c.csv",
                       stringsAsFactors=T, colClasses=data_2014.colclass, comment.char="")

data_2015 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats3d.csv",
                      stringsAsFactors=T, header=T, colClasses=data_2015.colclass, comment.char="")
```

```
data_2016_q1 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2016Q1.csv",
                         stringsAsFactors=T, header=T,colClasses=data_2016_q1.colclass, comment.char=""

data_2016_q2 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2016Q2.csv",
                          stringsAsFactors=T, header=T,colClasses=data_2016_q2.colclass, comment.char="

data_2016_q3 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2016Q3.csv",
                          stringsAsFactors=T, header=T,colClasses=data_2016_q3.colclass, comment.char="

data_2016_q4 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2016Q4.csv",
                          stringsAsFactors=T, header=T,colClasses=data_2016_q4.colclass, comment.char="

data_2017_q1 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2017Q1.csv",
                          stringsAsFactors=T, header=T,colClasses=data_2017_q1.colclass, comment.char="

data_2017_q2 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2017Q2.csv",
                          stringsAsFactors=T, header=T,colClasses=data_2017_q2.colclass, comment.char="

data_2017_q3 <- read.csv("https://cdn-stage.fedweb.org/fed-2/13/LoanStats_2017Q3.csv",
                          stringsAsFactors=T, header=T,colClasses=data_2017_q3.colclass, comment.char="
```

```
#5. merge csv files
full_data <- rbind(data_2007_2011[1:51],data_2012_2013[1:51],data_2014[1:51],data_2015[1:51],data_2016_
head(full_data)
```

```
##   id member_id loan_amnt funded_amnt funded_amnt_inv       term int_rate
## 1           NA      5000        5000            4975  36 months   10.65%
## 2           NA      2500        2500            2500  60 months   15.27%
## 3           NA      2400        2400            2400  36 months   15.96%
## 4           NA     10000       10000           10000  36 months   13.49%
## 5           NA      3000        3000            3000  60 months   12.69%
## 6           NA      5000        5000            5000  36 months    7.90%
##   installment grade sub_grade                emp_title emp_length
## 1      162.87     B        B2                           10+ years
## 2       59.83     C        C4                    Ryder   < 1 year
## 3       84.33     C        C5                           10+ years
## 4      339.31     C        C1      AIR RESOURCES BOARD  10+ years
## 5       67.79     B        B5 University Medical Group     1 year
## 6      156.46     A        A4     Veolia Transportaton    3 years
##   home_ownership annual_inc verification_status issue_d loan_status
## 1           RENT      24000            Verified  11-Dec  Fully Paid
## 2           RENT      30000     Source Verified  11-Dec Charged Off
## 3           RENT      12252        Not Verified  11-Dec  Fully Paid
## 4           RENT      49200     Source Verified  11-Dec  Fully Paid
## 5           RENT      80000     Source Verified  11-Dec  Fully Paid
## 6           RENT      36000     Source Verified  11-Dec  Fully Paid
##   pymnt_plan url
## 1          n  NA
## 2          n  NA
## 3          n  NA
## 4          n  NA
## 5          n  NA
## 6          n  NA
##
```

```
## 1
## 2   Borrower added on 12/22/11 > I plan to use this money to finance the motorcycle i am looking at.
## 3
## 4
## 5
## 6
##          purpose                                title zip_code addr_state
## 1    credit_card                             Computer    860xx         AZ
## 2            car                                 bike    309xx         GA
## 3 small_business                 real estate business    606xx         IL
## 4          other                             personel    917xx         CA
## 5          other                             Personal    972xx         OR
## 6        wedding My wedding loan I promise to pay back    852xx         AZ
##     dti delinq_2yrs earliest_cr_line inq_last_6mths mths_since_last_delinq
## 1 27.65           0           Jan-85              1                     NA
## 2  1.00           0           Apr-99              5                     NA
## 3  8.72           0            1-Nov              2                     NA
## 4 20.00           0           Feb-96              1                     35
## 5 17.94           0           Jan-96              0                     38
## 6 11.20           0            4-Nov              3                     NA
##   mths_since_last_record open_acc pub_rec revol_bal revol_util total_acc
## 1                     NA        3       0     13648     83.70%         9
## 2                     NA        3       0      1687      9.40%         4
## 3                     NA        2       0      2956     98.50%        10
## 4                     NA       10       0      5598        21%        37
## 5                     NA       15       0     27783     53.90%        38
## 6                     NA        9       0      7963     28.30%        12
##   initial_list_status out_prncp out_prncp_inv total_pymnt total_pymnt_inv
## 1                   f         0             0    5863.155         5833.84
## 2                   f         0             0    1014.530         1014.53
## 3                   f         0             0    3005.667         3005.67
## 4                   f         0             0   12231.890        12231.89
## 5                   f         0             0    4066.908         4066.91
## 6                   f         0             0    5632.210         5632.21
##   total_rec_prncp total_rec_int total_rec_late_fee recoveries
## 1         5000.00        863.16               0.00        0.0
## 2          456.46        435.17               0.00      122.9
## 3         2400.00        605.67               0.00        0.0
## 4        10000.00       2214.92              16.97        0.0
## 5         3000.00       1066.91               0.00        0.0
## 6         5000.00        632.21               0.00        0.0
##   collection_recovery_fee last_pymnt_d last_pymnt_amnt next_pymnt_d
## 1                    0.00       15-Jan          171.62
## 2                    1.11       13-Apr          119.66
## 3                    0.00       14-Jun          649.91
## 4                    0.00       15-Jan          357.48
## 5                    0.00       17-Jan           67.30
## 6                    0.00       15-Jan          161.03
##   last_credit_pull_d collections_12_mths_ex_med
## 1             17-Sep                          0
## 2             16-Oct                          0
## 3             17-Jun                          0
## 4             16-Apr                          0
## 5             17-Jan                          0
## 6            17-Feb                          0
##   mths_since_last_major_derog
## 1                          NA
## 2                          NA
## 3                          NA
## 4                          NA
## 5                          NA
## 6                          NA
```

After that I determined all types of loan statuses.

```
levels(factor(full_data$loan_status))
```

```
##  [1] ""
##  [2] "Charged Off"
##  [3] "Does not meet the credit policy. Status:Charged Off"
##  [4] "Does not meet the credit policy. Status:Fully Paid"
##  [5] "Fully Paid"
##  [6] "Current"
##  [7] "Default"
##  [8] "In Grace Period"
##  [9] "Late (16-30 days)"
## [10] "Late (31-120 days)"
```

I filtered the data so that the data set contain loans with “Fully Paid” or “Charged Off” statuses. I ignored loans with statuses “Current”, “Late (31-120 days)”, “Late (16-30 days)” and “Default” since theoretically borrowers still can pay them off.

```
full_data <- full_data %>% mutate(loan_status=str_replace(loan_status, "Does not meet the credit policy
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

Also, I removed all attributes that investors can’t see on the website and kept only the ones that they can see. Moreover, I converted term and interest rate attribute to numerical format.

```
##   loan_status loan_amnt term int_rate installment grade emp_length
## 1  Fully Paid      5000   36    10.65      162.87     B  10+ years
## 2 Charged Off      2500   60    15.27       59.83     C   < 1 year
## 3  Fully Paid      2400   36    15.96       84.33     C  10+ years
## 4  Fully Paid     10000   36    13.49      339.31     C  10+ years
## 5  Fully Paid      3000   60    12.69       67.79     B     1 year
## 6  Fully Paid      5000   36     7.90      156.46     A    3 years
##   home_ownership annual_inc verification_status   dti delinq_2yrs
## 1           RENT      24000            Verified 27.65           0
## 2           RENT      30000     Source Verified  1.00           0
## 3           RENT      12252        Not Verified  8.72           0
## 4           RENT      49200     Source Verified 20.00           0
## 5           RENT      80000     Source Verified 17.94           0
## 6           RENT      36000     Source Verified 11.20           0
##   inq_last_6mths mths_since_last_delinq mths_since_last_record open_acc
## 1              1                     NA                     NA        3
## 2              5                     NA                     NA        3
## 3              2                     NA                     NA        2
## 4              1                     35                     NA       10
## 5              0                     38                     NA       15
## 6              3                     NA                     NA        9
##   pub_rec revol_bal revol_util total_acc collections_12_mths_ex_med
## 1       0     13648         83         9                          0
## 2       0      1687          9         4                          0
## 3       0      2956         98        10                          0
## 4       0      5598         21        37                          0
## 5       0     27783         53        38                          0
## 6       0      7963         28        12                          0
##   mths_since_last_major_derog
## 1                          NA
## 2                          NA
## 3                          NA
## 4                          NA
## 5                          NA
## 6                          NA
```

Since the response variable "loan status" is a binary categorical variable (that has two possible outcomes - "Paid Off" or "Charged Off") and explanatory variables are numerical and categorical variables I used logistic regression for data set analysis.

In order to find the best fitting model that will provide the best prediction about the response variable one should use non-redundant explanatory variables. In order to decide which explanatory variables to include in multiple logistic regression I checked whether dependent variables are correlated between each other or not.

```
#removing categorical variables
full_data_no_categorical<- full_data %>% select(-loan_status,-grade,-emp_length,-home_ownership,-verifi
cor(full_data_no_categorical,use="complete.obs")
```

```
##                                loan_amnt         term     int_rate
## loan_amnt                    1.000000000  0.422905078  0.252943051
## term                         0.422905078  1.000000000  0.449439920
## int_rate                     0.252943051  0.449439920  1.000000000
## installment                  0.960984857  0.201161598  0.251209631
## annual_inc                   0.396604274  0.077325539 -0.024229579
## dti                         -0.004296079  0.055376313  0.154213650
## delinq_2yrs                  0.004823642 -0.013761784  0.044475316
## inq_last_6mths              -0.019746327 -0.002598566  0.242287085
## mths_since_last_delinq      -0.030847229  0.006928472 -0.050570450
## mths_since_last_record      -0.020720434  0.025732174  0.022453452
## open_acc                     0.145728246  0.086797964  0.068295286
## pub_rec                      0.031823800 -0.024404755  0.006143159
## revol_bal                    0.218459342  0.051399233 -0.006254746
## revol_util                   0.109403712  0.034683183  0.095508757
## total_acc                    0.106496607  0.069087858  0.001666116
## collections_12_mths_ex_med  -0.003320038 -0.009964323  0.003916385
## mths_since_last_major_derog -0.019175444  0.004729816 -0.029554970
##                              installment   annual_inc           dti
## loan_amnt                    0.960984857  0.396604274 -0.0042960793
## term                         0.201161598  0.077325539  0.0553763132
## int_rate                     0.251209631 -0.024229579  0.1542136500
## installment                  1.000000000  0.392374700 -0.0018279538
## annual_inc                   0.392374700  1.000000000 -0.2858904922
## dti                         -0.001827954 -0.285890492  1.0000000000
## delinq_2yrs                  0.011625624  0.042238702 -0.0020443121
## inq_last_6mths               0.009682516  0.040434813  0.0004220809
## mths_since_last_delinq      -0.038752539 -0.076417767  0.0204770180
## mths_since_last_record      -0.026630255 -0.086111918  0.0615709422
## open_acc                     0.139589986  0.092781884  0.2540949157
## pub_rec                     0.042574297  0.082048619 -0.0488774759
## revol_bal                   0.214007210  0.255702205  0.0733672979
## revol_util                  0.117137922  0.050101528  0.1431764418
## total_acc                   0.094812658  0.105119526  0.1645578233
## collections_12_mths_ex_med  0.000183462  0.005928169  0.0064742469
## mths_since_last_major_derog -0.023985298 -0.042258911  0.0328661843
##                              delinq_2yrs inq_last_6mths
## loan_amnt                    0.0048236422  -0.0197463266
## term                        -0.0137617839  -0.0025985660
## int_rate                     0.0444753161   0.2422870854
## installment                  0.0116256241   0.0096825159
## annual_inc                   0.0422387017   0.0404348126
## dti                         -0.0020443121   0.0004220809
## delinq_2yrs                  1.0000000000   0.0282431493
## inq_last_6mths               0.0282431493   1.0000000000
## mths_since_last_delinq      -0.4979427822   0.0032003344
## mths_since_last_record      -0.0124547371  -0.0347099768
## open_acc                     0.0550553163   0.1503937659
## pub_rec                      0.0006722893   0.0040616085
## revol_bal                    0.0064078761  -0.0122525865
## revol_util                   0.0001444110  -0.0794601846
## total_acc                    0.0423192281   0.1703407140
## collections_12_mths_ex_med   0.0847557528   0.0004772454
## mths_since_last_major_derog -0.3953539971   0.0124708097
##                             mths_since_last_delinq mths_since_last_record
## loan_amnt                             -0.030847229           -0.020720434
## term                                   0.006928472            0.025732174
## int_rate                              -0.050570450            0.022453452
## installment                           -0.038752539           -0.026630255
## annual_inc                            -0.076417767           -0.086111918
## dti                                    0.020477018            0.061570942
## delinq_2yrs                           -0.497942782           -0.012454737
## inq_last_6mths                         0.003200334           -0.034709977
## mths_since_last_delinq                 1.000000000           -0.006842418
## mths_since_last_record                -0.006842418            1.000000000
## open_acc                              -0.040857028            0.031370633
## pub_rec                               -0.004742984           -0.269376093
## revol_bal                             -0.019583893           -0.024295713
## revol_util                            -0.015932256            0.041013824
## total_acc                             -0.003720052           -0.144489072
## collections_12_mths_ex_med            -0.097995842           -0.005553399
## mths_since_last_major_derog            0.689480972           -0.007275000
##                                open_acc       pub_rec    revol_bal
## loan_amnt                    0.14572825  0.0318237998  0.218459342
## term                         0.08679796 -0.0244047554  0.051399233
## int_rate                     0.06829529  0.0061431595 -0.006254746
## installment                  0.13958999  0.0425742965  0.214007210
## annual_inc                   0.09278188  0.0820486188  0.255702205
## dti                          0.25409492 -0.0488774759  0.073367298
## delinq_2yrs                  0.05505532  0.0006722893  0.006407876
## inq_last_6mths               0.15039377  0.0040616085 -0.012252587
## mths_since_last_delinq      -0.04085703 -0.0047429835 -0.019583893
## mths_since_last_record       0.03137063 -0.2693760931 -0.024295713
## open_acc                     1.00000000 -0.0184278479  0.162964627
## pub_rec                     -0.01842785  1.0000000000  0.011407285
## revol_bal                    0.16296463  0.0114072853  1.000000000
## revol_util                  -0.07903746 -0.0002151927  0.230578235
## total_acc                    0.59745422 -0.0586662051  0.079748021
## collections_12_mths_ex_med   0.01444761  0.0169394696 -0.007964900
## mths_since_last_major_derog -0.01320742  0.0099003795  0.002334282
##                                 revol_util    total_acc
## loan_amnt                    0.1094037119  0.106496607
## term                         0.0346831827  0.069087858
## int_rate                     0.0955087571  0.001666116
## installment                  0.1171379217  0.094812658
## annual_inc                   0.0501015282  0.105119526
## dti                          0.1431764418  0.164557823
## delinq_2yrs                  0.0001444110  0.042319228
## inq_last_6mths              -0.0794601846  0.170340714
## mths_since_last_delinq      -0.0159322556 -0.003720052
## mths_since_last_record       0.0410138239 -0.144489072
## open_acc                    -0.0790374605  0.597454218
## pub_rec                     -0.0002151927 -0.058666205
## revol_bal                    0.2305782349  0.079748021
## revol_util                   1.0000000000 -0.086334311
## total_acc                   -0.0863343109  1.000000000
## collections_12_mths_ex_med  -0.0258469002 -0.023336431
## mths_since_last_major_derog  0.0221548906 -0.006425679
##                             collections_12_mths_ex_med
## loan_amnt                                -0.0033200383
## term                                     -0.0099643227
## int_rate                                  0.0039163850
## installment                               0.0001834620
## annual_inc                                0.0059281689
## dti                                       0.0064742469
## delinq_2yrs                               0.0847557528
## inq_last_6mths                            0.0004772454
## mths_since_last_delinq                   -0.0979958421
## mths_since_last_record                   -0.0055533989
## open_acc                                  0.0144476057
## pub_rec                                   0.0169394696
## revol_bal                                -0.0079648997
## revol_util                               -0.0258469002
## total_acc                                -0.0233364306
## collections_12_mths_ex_med                1.0000000000
## mths_since_last_major_derog              -0.1270235332
##                             mths_since_last_major_derog
## loan_amnt                                  -0.019175444
## term                                        0.004729816
## int_rate                                   -0.029554970
## installment                                -0.023985298
## annual_inc                                 -0.042258911
## dti                                         0.032866184
## delinq_2yrs                                -0.395353997
## inq_last_6mths                              0.012470810
## mths_since_last_delinq                      0.689480972
## mths_since_last_record                     -0.007275000
## open_acc                                   -0.013207424
## pub_rec                                          0.009900380
## revol_bal                                        0.002334282
## revol_util                                       0.022154891
## total_acc                                       -0.006425679
## collections_12_mths_ex_med                      -0.127023533
## mths_since_last_major_derog                      1.000000000
```

```
#chart.Correlation(full_data_no_categorical, method="spearman",histogram=TRUE)
```

According to the correlation matrix loan amount and installment there are four pairs of variables that are highly correlated. Those pairs are:

1. loan amount and installment
2. the number of open accounts and the number of total accounts
3. the number of months since the borrower's last delinquency and the number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
4. the number of months since the borrower's last delinquency and the number of since most recent 90-day or worse rating

Both highly correlated variables should not be in a final regression model.

In order to find the best regression model I ran the step function that analyses all combination of variables and selects the best regression model based on lowest AIC (Akaike's criterion) value. Lower values of AIC indicate the preferred model, that is, the one with the fewest parameters that still provides an adequate fit to the data.

```
full_data.omit = na.omit(full_data)

model.null = glm(loan_status ~ 1,
                 data = full_data.omit,
                 family = binomial(link="logit")
                 )

model.full = glm(loan_status ~ .,
                 data = full_data.omit,
                 family = binomial(link="logit")
                 )

step(model.null,
     scope = list(upper=model.full),
             direction = "both",
             test = "Chisq",
             data = full_data)
```

```
## Start:  AIC=38611.8
## loan_status ~ 1
##
##                                Df Deviance   AIC     LRT  Pr(>Chi)
## + grade                         6    36736 36750 1873.51 < 2.2e-16 ***
## + int_rate                      1    37092 37096 1517.43 < 2.2e-16 ***
## + term                          1    37555 37559 1054.57 < 2.2e-16 ***
## + dti                           1    37891 37895  719.12 < 2.2e-16 ***
## + loan_amnt                     1    37991 37995  618.82 < 2.2e-16 ***
## + installment                   1    38155 38159  455.20 < 2.2e-16 ***
## + open_acc                      1    38411 38415  199.08 < 2.2e-16 ***
## + verification_status           2    38411 38417  199.28 < 2.2e-16 ***
## + home_ownership                3    38411 38419  199.21 < 2.2e-16 ***
## + revol_util                  1    38528 38532   81.78 < 2.2e-16 ***
## + inq_last_6mths              1    38535 38539   74.47 < 2.2e-16 ***
## + emp_length                 11    38518 38542   92.03 6.667e-15 ***
## + delinq_2yrs                 1    38569 38573   41.13 1.423e-10 ***
## + mths_since_last_delinq      1    38576 38580   33.71 6.407e-09 ***
## + annual_inc                  1    38579 38583   30.62 3.134e-08 ***
## + collections_12_mths_ex_med  1    38587 38591   22.85 1.748e-06 ***
## + revol_bal                   1    38594 38598   15.47 8.404e-05 ***
## + mths_since_last_major_derog  1    38604 38608    5.47   0.01938 *
## + pub_rec                     1    38605 38609    5.12   0.02370 *
## + mths_since_last_record      1    38605 38609    4.54   0.03307 *
## <none>                             38610 38612
## + total_acc                   1    38609 38613    0.77   0.37955
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=36750.29
## loan_status ~ grade
##
##                             Df Deviance   AIC     LRT  Pr(>Chi)
## + dti                         1    36323 36339  413.41 < 2.2e-16 ***
## + term                        1    36533 36549  203.05 < 2.2e-16 ***
## + loan_amnt                   1    36535 36551  201.53 < 2.2e-16 ***
## + home_ownership              3    36576 36596  160.23 < 2.2e-16 ***
## + installment                 1    36612 36628  124.14 < 2.2e-16 ***
## + open_acc                    1    36617 36633  119.23 < 2.2e-16 ***
## + emp_length                 11    36632 36668  104.18 < 2.2e-16 ***
## + verification_status         2    36691 36709   45.74 1.168e-10 ***
## + revol_util                  1    36704 36720   32.30 1.319e-08 ***
## + int_rate                    1    36706 36722   30.59 3.193e-08 ***
## + annual_inc                  1    36710 36726   26.55 2.563e-07 ***
## + collections_12_mths_ex_med  1    36716 36732   20.57 5.752e-06 ***
## + delinq_2yrs                 1    36716 36732   20.31 6.598e-06 ***
## + revol_bal                   1    36719 36735   17.63 2.679e-05 ***
## + mths_since_last_delinq      1    36720 36736   15.84 6.887e-05 ***
## + mths_since_last_record      1    36732 36748    4.31   0.03782 *
## + pub_rec                     1    36733 36749    3.24   0.07175 .
## + inq_last_6mths              1    36733 36749    3.22   0.07266 .
## + mths_since_last_major_derog  1    36734 36750    2.25   0.13362
## <none>                             36736 36750
## + total_acc                   1    36736 36752    0.06   0.80158
## - grade                       6    38610 38612 1873.51 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=36338.87
## loan_status ~ grade + dti
##
##                             Df Deviance   AIC     LRT  Pr(>Chi)
## + loan_amnt                   1    36085 36103  238.24 < 2.2e-16 ***
## + term                        1    36103 36121  219.45 < 2.2e-16 ***
## + installment                 1    36173 36191  149.44 < 2.2e-16 ***
## + home_ownership              3    36178 36200  144.96 < 2.2e-16 ***
## + emp_length                 11    36237 36275   85.53 1.244e-13 ***
## + verification_status          2    36277 36297   45.86 1.102e-10 ***
## + open_acc                     1    36284 36302   39.28 3.673e-10 ***
## + int_rate                     1    36293 36311   29.83 4.707e-08 ***
## + delinq_2yrs                  1    36300 36318   23.33 1.366e-06 ***
## + mths_since_last_delinq       1    36301 36319   21.66 3.262e-06 ***
## + collections_12_mths_ex_med   1    36303 36321   19.93 8.022e-06 ***
## + revol_util                   1    36313 36331   10.04  0.001533 **
## + total_acc                    1    36313 36331    9.71  0.001834 **
## + pub_rec                      1    36314 36332    8.41  0.003732 **
## + revol_bal                    1    36315 36333    7.52  0.006102 **
## + mths_since_last_major_derog  1    36318 36336    5.31  0.021258 *
## <none>                              36323 36339
## + annual_inc                   1    36322 36340    1.02  0.312494
## + mths_since_last_record       1    36322 36340    0.62  0.429703
## + inq_last_6mths               1    36322 36340    0.55  0.459223
## - dti                          1    36736 36750  413.41 < 2.2e-16 ***
## - grade                        6    37891 37895 1567.81 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=36102.63
## loan_status ~ grade + dti + loan_amnt
##
##                               Df Deviance   AIC     LRT  Pr(>Chi)
## + home_ownership               3    35876 35900  209.07 < 2.2e-16 ***
## + term                         1    35985 36005   99.58 < 2.2e-16 ***
## + emp_length                  11    35970 36010  114.82 < 2.2e-16 ***
## + installment                  1    36017 36037   67.85 < 2.2e-16 ***
## + annual_inc                   1    36042 36062   42.43 7.319e-11 ***
## + int_rate                     1    36050 36070   34.53 4.204e-09 ***
## + verification_status          2    36056 36078   28.25 7.347e-07 ***
## + total_acc                    1    36059 36079   25.61 4.178e-07 ***
## + delinq_2yrs                  1    36060 36080   25.06 5.571e-07 ***
## + collections_12_mths_ex_med   1    36063 36083   21.19 4.150e-06 ***
## + mths_since_last_delinq       1    36065 36085   19.73 8.898e-06 ***
## + open_acc                     1    36068 36088   16.60 4.619e-05 ***
## + pub_rec                      1    36078 36098    6.43   0.01119 *
## + mths_since_last_major_derog  1    36080 36100    4.76   0.02920 *
## + revol_util                   1    36082 36102    2.58   0.10856
## <none>                              36085 36103
## + mths_since_last_record       1    36083 36103    1.26   0.26086
## + revol_bal                    1    36084 36104    0.56   0.45461
## + inq_last_6mths               1    36084 36104    0.50   0.47748
## - loan_amnt                    1    36323 36339  238.24 < 2.2e-16 ***
## - dti                          1    36535 36551  450.12 < 2.2e-16 ***
## - grade                        6    37243 37249 1158.75 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35899.56
## loan_status ~ grade + dti + loan_amnt + home_ownership
##
##                               Df Deviance   AIC     LRT  Pr(>Chi)
## + term                         1    35760 35786  115.67 < 2.2e-16 ***
## + installment                  1    35795 35821   80.31 < 2.2e-16 ***
## + emp_length                  11    35778 35824   97.28 6.162e-16 ***
## + int_rate                     1    35839 35865   36.45 1.564e-09 ***
## + delinq_2yrs                  1    35846 35872   29.99 4.342e-08 ***
## + annual_inc                   1    35846 35872   29.75 4.916e-08 ***
## + mths_since_last_delinq       1    35850 35876   25.42 4.605e-07 ***
## + open_acc                     1    35853 35879   22.86 1.741e-06 ***
## + collections_12_mths_ex_med   1    35856 35882   19.21 1.172e-05 ***
## + total_acc                    1    35859 35885   16.60 4.623e-05 ***
## + verification_status          2    35858 35886   17.20 0.0001842 ***
## + revol_util                   1    35868 35894    7.09 0.0077667 **
## + pub_rec                      1    35870 35896    5.75 0.0164890 *
## + mths_since_last_major_derog  1    35870 35896    5.13 0.0235422 *
## <none>                              35876 35900
## + mths_since_last_record       1    35874 35900    1.53 0.2156855
## + inq_last_6mths               1    35874 35900    1.51 0.2189319
## + revol_bal                    1    35876 35902    0.02 0.8815250
## - home_ownership               3    36085 36103  209.07 < 2.2e-16 ***
## - loan_amnt                    1    36178 36200  302.34 < 2.2e-16 ***
## - dti                          1    36312 36334  435.95 < 2.2e-16 ***
## - grade                        6    36965 36977 1089.41 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35785.89
## loan_status ~ grade + dti + loan_amnt + home_ownership + term
##
##                               Df Deviance   AIC    LRT  Pr(>Chi)
## + emp_length                  11    35655 35703 104.45 < 2.2e-16 ***
## + int_rate                     1    35720 35748  40.23 2.260e-10 ***
## + delinq_2yrs                  1    35725 35753  35.21 2.953e-09 ***
## + mths_since_last_delinq       1    35729 35757  30.54 3.277e-08 ***
## + open_acc                     1    35738 35766  21.81 3.004e-06 ***
## + annual_inc                   1    35738 35766  21.62 3.323e-06 ***
## + collections_12_mths_ex_med   1    35739 35767  20.61 5.622e-06 ***
## + total_acc                    1    35740 35768  19.95 7.967e-06 ***
## + verification_status          2    35741 35771  18.75 8.494e-05 ***
## + revol_util                   1    35751 35779   9.11  0.002537 **
## + pub_rec                      1    35751 35779   8.66  0.003257 **
## + mths_since_last_major_derog  1    35754 35782   6.37  0.011587 *
## + inq_last_6mths               1    35754 35782   6.16  0.013033 *
## <none>                              35760 35786
## + installment                  1    35758 35786   1.74  0.187287
## + mths_since_last_record       1    35759 35787   0.64  0.425373
## + revol_bal                    1    35760 35788   0.19  0.663781
## - term                         1    35876 35900 115.67 < 2.2e-16 ***
## - loan_amnt                    1    35918 35942 158.49 < 2.2e-16 ***
## - home_ownership               3    35985 36005 225.17 < 2.2e-16 ***
## - dti                          1    36198 36222 438.08 < 2.2e-16 ***
## - grade                        6    36437 36451 677.60 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35703.44
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length
##
##                               Df Deviance   AIC    LRT  Pr(>Chi)
## + int_rate                     1    35615 35665  40.55 1.914e-10 ***
## + delinq_2yrs                  1    35619 35669  36.30 1.693e-09 ***
## + mths_since_last_delinq       1    35624 35674  31.77 1.738e-08 ***
## + open_acc                     1    35629 35679  26.21 3.062e-07 ***
## + collections_12_mths_ex_med   1    35635 35685  20.21 6.941e-06 ***
## + total_acc                    1    35637 35687  17.99 2.218e-05 ***
## + annual_inc                   1    35642 35692  13.51 0.0002377 ***
## + verification_status          2    35641 35693  14.75 0.0006265 ***
## + revol_util                   1    35644 35694  11.93 0.0005535 ***
## + pub_rec                      1    35647 35697   8.68 0.0032252 **
## + inq_last_6mths               1    35648 35698   7.24 0.0071142 **
## + mths_since_last_major_derog  1    35649 35699   6.13 0.0132866 *
## + installment                  1    35653 35703   2.17 0.1408510
## <none>                              35655 35703
## + mths_since_last_record       1    35655 35705   0.74 0.3892936
## + revol_bal                    1    35655 35705   0.40 0.5268580
## - emp_length                  11    35760 35786 104.45 < 2.2e-16 ***
## - term                         1    35778 35824 122.84 < 2.2e-16 ***
## - loan_amnt                    1    35832 35878 177.01 < 2.2e-16 ***
## - home_ownership               3    35863 35905 207.19 < 2.2e-16 ***
## - dti                          1    36072 36118 416.37 < 2.2e-16 ***
## - grade                        6    36330 36366 674.89 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35664.89
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate
##
##                               Df Deviance   AIC    LRT  Pr(>Chi)
## + delinq_2yrs                  1    35579 35631  36.24 1.747e-09 ***
## + mths_since_last_delinq       1    35581 35633  33.40 7.486e-09 ***
## + open_acc                     1    35590 35642  25.27 4.973e-07 ***
## + total_acc                    1    35595 35647  20.04 7.586e-06 ***
## + collections_12_mths_ex_med   1    35596 35648  19.22 1.164e-05 ***
## + annual_inc                   1    35600 35652  15.12 0.0001008 ***
## + verification_status          2    35598 35652  16.59 0.0002491 ***
## + revol_util                   1    35601 35653  13.66 0.0002190 ***
## + installment                  1    35602 35654  12.82 0.0003428 ***
## + inq_last_6mths               1    35606 35658   8.50 0.0035595 **
## + pub_rec                      1    35607 35659   8.10 0.0044290 **
## + mths_since_last_major_derog  1    35608 35660   7.19 0.0073372 **
## <none>                              35615 35665
## + mths_since_last_record       1    35613 35665   1.51 0.2198582
## + revol_bal                    1    35615 35667   0.36 0.5494714
## - int_rate                     1    35655 35703  40.55 1.914e-10 ***
## - emp_length                  11    35720 35748 104.77 < 2.2e-16 ***
## - term                         1    35742 35790 126.80 < 2.2e-16 ***
## - loan_amnt                    1    35795 35843 179.99 < 2.2e-16 ***
## - home_ownership               3    35825 35869 209.68 < 2.2e-16 ***
## - grade                        6    35882 35920 267.29 < 2.2e-16 ***
## - dti                          1    36031 36079 415.74 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35630.65
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs
##
##                               Df Deviance   AIC    LRT  Pr(>Chi)
## + total_acc                    1    35556 35610  23.00 1.619e-06 ***
## + open_acc                     1    35556 35610  22.16 2.508e-06 ***
## + annual_inc                   1    35561 35615  17.57 2.764e-05 ***
## + verification_status          2    35561 35617  17.66 0.0001463 ***
## + collections_12_mths_ex_med   1    35563 35617  15.21 9.634e-05 ***
## + revol_util                   1    35564 35618  14.24 0.0001608 ***
## + installment                  1    35565 35619  13.87 0.0001956 ***
## + mths_since_last_delinq       1    35568 35622  10.68 0.0010834 **
## + pub_rec                      1    35571 35625   8.13 0.0043559 **
## + inq_last_6mths               1    35571 35625   8.03 0.0046057 **
## <none>                              35579 35631
## + mths_since_last_record       1    35577 35631   1.70 0.1923600
## + revol_bal                    1    35578 35632   0.34 0.5615886
## + mths_since_last_major_derog  1    35579 35633   0.13 0.7210945
## - delinq_2yrs                  1    35615 35665  36.24 1.747e-09 ***
## - int_rate                     1    35619 35669  40.49 1.975e-10 ***
## - emp_length                  11    35685 35715 105.92 < 2.2e-16 ***
## - term                         1    35711 35761 132.52 < 2.2e-16 ***
## - loan_amnt                    1    35759 35809 180.22 < 2.2e-16 ***
## - home_ownership               3    35794 35840 215.72 < 2.2e-16 ***
## - grade                        6    35841 35881 261.98 < 2.2e-16 ***
## - dti                          1    35998 36048 419.09 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35609.65
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc
##
##                               Df Deviance   AIC    LRT  Pr(>Chi)
## + open_acc                     1    35470 35526  86.03 < 2.2e-16 ***
## + verification_status          2    35538 35596  17.97 0.0001251 ***
## + inq_last_6mths               1    35540 35596  15.27 9.337e-05 ***
## + installment                  1    35541 35597  14.71 0.0001253 ***
## + collections_12_mths_ex_med   1    35541 35597  14.25 0.0001604 ***
## + annual_inc                   1    35543 35599  13.12 0.0002928 ***
## + revol_util                   1    35545 35601  10.24 0.0013740 **
## + mths_since_last_delinq       1    35546 35602   9.90 0.0016514 **
## + pub_rec                      1    35549 35605   6.76 0.0093373 **
## <none>                              35556 35610
## + revol_bal                    1    35555 35611   0.57 0.4511160
## + mths_since_last_record       1    35555 35611   0.38 0.5398402
## + mths_since_last_major_derog  1    35556 35612   0.09 0.7668467
## - total_acc                    1    35579 35631  23.00 1.619e-06 ***
## - delinq_2yrs                  1    35595 35647  39.20 3.827e-10 ***
## - int_rate                     1    35598 35650  42.71 6.343e-11 ***
## - emp_length                  11    35659 35691 103.68 < 2.2e-16 ***
## - term                         1    35692 35744 136.72 < 2.2e-16 ***
## - loan_amnt                    1    35746 35798 190.65 < 2.2e-16 ***
## - home_ownership               3    35762 35810 206.09 < 2.2e-16 ***
## - grade                        6    35819 35861 263.20 < 2.2e-16 ***
## - dti                          1    35997 36049 440.89 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35525.62
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc
##
##                               Df Deviance   AIC    LRT  Pr(>Chi)
## + annual_inc                   1    35452 35510  17.68 2.612e-05 ***
## + verification_status          2    35452 35512  17.84 0.0001337 ***
## + revol_util                   1    35454 35512  15.59 7.885e-05 ***
## + installment                  1    35456 35514  13.17 0.0002840 ***
## + collections_12_mths_ex_med   1    35457 35515  12.23 0.0004697 ***
## + inq_last_6mths               1    35459 35517  10.92 0.0009531 ***
## + mths_since_last_delinq       1    35461 35519   8.23 0.0041228 **
## + pub_rec                      1    35464 35522   5.74 0.0166118 *
## <none>                              35470 35526
## + mths_since_last_record       1    35469 35527   0.40 0.5273230
## + revol_bal                    1    35470 35528   0.11 0.7401254
## + mths_since_last_major_derog  1    35470 35528   0.09 0.7693954
## - delinq_2yrs                  1    35505 35559  35.77 2.217e-09 ***
## - int_rate                     1    35514 35568  43.90 3.460e-11 ***
## - open_acc                     1    35556 35610  86.03 < 2.2e-16 ***
## - total_acc                    1    35556 35610  86.87 < 2.2e-16 ***
## - emp_length                  11    35581 35615 111.11 < 2.2e-16 ***
## - term                         1    35609 35663 139.87 < 2.2e-16 ***
## - loan_amnt                    1    35640 35694 170.04 < 2.2e-16 ***
## - home_ownership               3    35679 35729 209.56 < 2.2e-16 ***
## - grade                        6    35732 35776 262.04 < 2.2e-16 ***
## - dti                          1    35827 35881 357.52 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35509.94
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc +
##     annual_inc
##
##                               Df Deviance   AIC     LRT  Pr(>Chi)
## + revol_util                   1    35433 35493  18.588 1.623e-05 ***
## + verification_status          2    35433 35495  18.553 9.359e-05 ***
## + installment                  1    35439 35499  12.701 0.0003654 ***
## + inq_last_6mths               1    35439 35499  12.552 0.0003958 ***
## + collections_12_mths_ex_med   1    35440 35500  12.436 0.0004212 ***
## + mths_since_last_delinq       1    35442 35502   9.649 0.0018949 **
## + pub_rec                      1    35445 35505   7.311 0.0068524 **
## <none>                              35452 35510
## + mths_since_last_record       1    35451 35511   0.771 0.3798338
## + revol_bal                    1    35452 35512   0.368 0.5439891
## + mths_since_last_major_derog  1    35452 35512   0.131 0.7169657
## - annual_inc                   1    35470 35526  17.682 2.612e-05 ***
## - delinq_2yrs                  1    35490 35546  37.789 7.883e-10 ***
## - int_rate                     1    35497 35553  45.528 1.505e-11 ***
## - total_acc                    1    35533 35589  81.387 < 2.2e-16 ***
## - emp_length                  11    35553 35589 101.556 < 2.2e-16 ***
## - open_acc                     1    35543 35599  90.594 < 2.2e-16 ***
## - term                         1    35583 35639 130.879 < 2.2e-16 ***
## - loan_amnt                    1    35637 35693 185.415 < 2.2e-16 ***
## - home_ownership               3    35652 35704 200.255 < 2.2e-16 ***
## - grade                        6    35715 35761 263.176 < 2.2e-16 ***
## - dti                          1    35722 35778 269.725 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35493.35
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc +
##     annual_inc + revol_util
##
##                               Df Deviance   AIC     LRT  Pr(>Chi)
## + verification_status          2    35415 35479  18.678 8.794e-05 ***
## + inq_last_6mths               1    35418 35480  15.288 9.229e-05 ***
## + collections_12_mths_ex_med   1    35420 35482  13.495 0.0002392 ***
## + installment                  1    35421 35483  12.085 0.0005083 ***
## + mths_since_last_delinq       1    35424 35486   9.578 0.0019694 **
## + pub_rec                      1    35426 35488   7.547 0.0060107 **
## <none>                              35433 35493
## + mths_since_last_record       1    35432 35494   1.047 0.3062975
## + mths_since_last_major_derog  1    35433 35495   0.242 0.6224561
## + revol_bal                    1    35433 35495   0.133 0.7149611
## - revol_util                   1    35452 35510  18.588 1.623e-05 ***
## - annual_inc                   1    35454 35512  20.684 5.418e-06 ***
## - delinq_2yrs                  1    35472 35530  38.213 6.343e-10 ***
## - int_rate                     1    35481 35539  47.636 5.131e-12 ***
## - total_acc                    1    35510 35568  76.320 < 2.2e-16 ***
## - emp_length                  11    35538 35576 104.886 < 2.2e-16 ***
## - open_acc                     1    35530 35588  97.048 < 2.2e-16 ***
## - term                         1    35567 35625 133.351 < 2.2e-16 ***
## - loan_amnt                    1    35610 35668 176.639 < 2.2e-16 ***
## - home_ownership               3    35642 35696 208.305 < 2.2e-16 ***
## - dti                          1    35668 35726 234.738 < 2.2e-16 ***
## - grade                        6    35696 35744 262.880 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35478.68
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc +
##     annual_inc + revol_util + verification_status
##
```

```
##                              Df Deviance   AIC     LRT  Pr(>Chi)    
## + inq_last_6mths              1    35400 35466  14.945 0.0001107 ***
## + collections_12_mths_ex_med  1    35402 35468  13.146 0.0002881 ***
## + installment                 1    35403 35469  11.899 0.0005618 ***
## + mths_since_last_delinq      1    35405 35471   9.597 0.0019490 ** 
## + pub_rec                     1    35408 35474   6.977 0.0082583 ** 
## <none>                             35415 35479                      
## + mths_since_last_record      1    35414 35480   0.585 0.4443052    
## + mths_since_last_major_derog  1    35414 35480   0.181 0.6707073    
## + revol_bal                   1    35415 35481   0.115 0.7347541    
## - verification_status         2    35433 35493  18.678 8.794e-05 ***
## - revol_util                  1    35433 35495  18.712 1.520e-05 ***
## - annual_inc                  1    35436 35498  21.459 3.614e-06 ***
## - delinq_2yrs                 1    35454 35516  39.381 3.487e-10 ***
## - int_rate                    1    35465 35527  50.077 1.478e-12 ***
## - total_acc                   1    35491 35553  76.520 < 2.2e-16 ***
## - emp_length                 11    35513 35555  97.873 4.706e-16 ***
## - open_acc                    1    35512 35574  97.035 < 2.2e-16 ***
## - term                        1    35549 35611 134.484 < 2.2e-16 ***
## - loan_amnt                   1    35579 35641 163.909 < 2.2e-16 ***
## - home_ownership              3    35612 35670 196.882 < 2.2e-16 ***
## - dti                         1    35649 35711 234.259 < 2.2e-16 ***
## - grade                       6    35678 35730 263.677 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Step:  AIC=35465.73
## loan_status ~ grade + dti + loan_amnt + home_ownership + term + 
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc + 
##     annual_inc + revol_util + verification_status + inq_last_6mths
## 
##                              Df Deviance   AIC     LRT  Pr(>Chi)    
## + collections_12_mths_ex_med  1    35387 35455  13.199 0.0002800 ***
## + installment                 1    35388 35456  11.435 0.0007208 ***
## + mths_since_last_delinq      1    35389 35457  10.371 0.0012802 ** 
## + pub_rec                     1    35393 35461   6.977 0.0082545 ** 
## <none>                             35400 35466                      
## + mths_since_last_record      1    35399 35467   0.560 0.4543541    
## + mths_since_last_major_derog  1    35399 35467   0.301 0.5832474    
## + revol_bal                   1    35400 35468   0.113 0.7362599    
## - inq_last_6mths              1    35415 35479  14.945 0.0001107 ***
## - verification_status         2    35418 35480  18.334 0.0001044 ***
## - revol_util                  1    35421 35485  21.399 3.729e-06 ***
## - annual_inc                  1    35423 35487  23.687 1.133e-06 ***
## - delinq_2yrs                 1    35439 35503  39.512 3.260e-10 ***
## - int_rate                    1    35452 35516  52.315 4.727e-13 ***
## - emp_length                 11    35499 35543  99.289 2.470e-16 ***
## - total_acc                   1    35483 35547  83.662 < 2.2e-16 ***
## - open_acc                    1    35492 35556  92.606 < 2.2e-16 ***
## - term                        1    35543 35607 143.380 < 2.2e-16 ***
## - loan_amnt                   1    35572 35636 172.021 < 2.2e-16 ***
## - home_ownership              3    35600 35660 200.002 < 2.2e-16 ***
## - dti                         1    35640 35704 240.650 < 2.2e-16 ***
## - grade                       6    35652 35706 252.455 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35454.53
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc +
##     annual_inc + revol_util + verification_status + inq_last_6mths +
##     collections_12_mths_ex_med
##
##                              Df Deviance   AIC     LRT  Pr(>Chi)
## + installment                 1    35375 35445  11.310 0.0007711 ***
## + mths_since_last_delinq      1    35378 35448   8.885 0.0028751 **
## + pub_rec                     1    35380 35450   6.783 0.0092052 **
## <none>                             35387 35455
## + mths_since_last_record      1    35386 35456   0.511 0.4745557
## + revol_bal                   1    35386 35456   0.103 0.7479974
## + mths_since_last_major_derog  1    35387 35457   0.023 0.8803784
## - collections_12_mths_ex_med  1    35400 35466  13.199 0.0002800 ***
## - inq_last_6mths              1    35402 35468  14.998 0.0001076 ***
## - verification_status         2    35405 35469  17.996 0.0001236 ***
## - revol_util                  1    35409 35475  22.509 2.091e-06 ***
## - annual_inc                  1    35411 35477  24.014 9.566e-07 ***
## - delinq_2yrs                 1    35422 35488  35.564 2.468e-09 ***
## - int_rate                    1    35438 35504  51.314 7.870e-13 ***
## - emp_length                 11    35486 35532  99.013 2.800e-16 ***
## - total_acc                   1    35467 35533  80.828 < 2.2e-16 ***
## - open_acc                    1    35477 35543  90.715 < 2.2e-16 ***
## - term                        1    35531 35597 144.192 < 2.2e-16 ***
## - loan_amnt                   1    35559 35625 172.286 < 2.2e-16 ***
## - home_ownership              3    35585 35647 198.856 < 2.2e-16 ***
## - dti                         1    35626 35692 239.435 < 2.2e-16 ***
## - grade                       6    35637 35693 250.729 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35445.22
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc +
##     annual_inc + revol_util + verification_status + inq_last_6mths +
##     collections_12_mths_ex_med + installment
##
##                              Df Deviance   AIC     LRT  Pr(>Chi)
## + mths_since_last_delinq      1    35366 35438   9.000 0.0026992 **
## + pub_rec                     1    35369 35441   6.604 0.0101768 *
## - loan_amnt                   1    35377 35445   1.738 0.1873772
## <none>                             35375 35445
## + mths_since_last_record      1    35375 35447   0.509 0.4757351
## + revol_bal                   1    35375 35447   0.070 0.7907009
## + mths_since_last_major_derog  1    35375 35447   0.026 0.8715877
## - installment                 1    35387 35455  11.310 0.0007711 ***
## - collections_12_mths_ex_med  1    35388 35456  13.074 0.0002994 ***
## - inq_last_6mths              1    35390 35458  14.533 0.0001377 ***
## - verification_status         2    35393 35459  17.811 0.0001356 ***
## - revol_util                  1    35397 35465  21.800 3.026e-06 ***
## - annual_inc                   1    35399 35467  23.379 1.330e-06 ***
## - delinq_2yrs                  1    35412 35480  36.540 1.495e-09 ***
## - int_rate                     1    35437 35505  61.562 4.291e-15 ***
## - term                         1    35445 35513  69.348 < 2.2e-16 ***
## - emp_length                  11    35475 35523 100.066 < 2.2e-16 ***
## - total_acc                    1    35456 35524  81.199 < 2.2e-16 ***
## - open_acc                     1    35464 35532  89.124 < 2.2e-16 ***
## - home_ownership               3    35574 35638 198.608 < 2.2e-16 ***
## - dti                          1    35615 35683 239.476 < 2.2e-16 ***
## - grade                        6    35632 35690 257.064 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35438.22
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc +
##     annual_inc + revol_util + verification_status + inq_last_6mths +
##     collections_12_mths_ex_med + installment + mths_since_last_delinq
##
##                               Df Deviance   AIC     LRT  Pr(>Chi)
## + pub_rec                      1    35360 35434   6.709 0.0095944 **
## + mths_since_last_major_derog  1    35362 35436   4.371 0.0365566 *
## - loan_amnt                    1    35368 35438   1.781 0.1820678
## <none>                              35366 35438
## + mths_since_last_record       1    35366 35440   0.601 0.4383232
## + revol_bal                    1    35366 35440   0.059 0.8086500
## - mths_since_last_delinq       1    35375 35445   9.000 0.0026992 **
## - installment                  1    35378 35448  11.425 0.0007246 ***
## - collections_12_mths_ex_med   1    35378 35448  11.585 0.0006650 ***
## - delinq_2yrs                  1    35381 35451  14.849 0.0001165 ***
## - inq_last_6mths               1    35381 35451  15.247 9.433e-05 ***
## - verification_status          2    35384 35452  17.811 0.0001357 ***
## - revol_util                   1    35388 35458  21.732 3.135e-06 ***
## - annual_inc                   1    35391 35461  24.967 5.831e-07 ***
## - int_rate                     1    35429 35499  62.921 2.151e-15 ***
## - term                         1    35436 35506  70.156 < 2.2e-16 ***
## - total_acc                    1    35445 35515  79.250 < 2.2e-16 ***
## - emp_length                  11    35466 35516 100.169 < 2.2e-16 ***
## - open_acc                     1    35454 35524  87.568 < 2.2e-16 ***
## - home_ownership               3    35567 35633 200.971 < 2.2e-16 ***
## - dti                          1    35607 35677 240.428 < 2.2e-16 ***
## - grade                        6    35624 35684 258.143 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35433.51
## loan_status ~ grade + dti + loan_amnt + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc +
##     annual_inc + revol_util + verification_status + inq_last_6mths +
##     collections_12_mths_ex_med + installment + mths_since_last_delinq +
##     pub_rec
##
##                               Df Deviance   AIC     LRT  Pr(>Chi)
## + mths_since_last_major_derog  1    35355 35431   4.163 0.0413185 *
## - loan_amnt                    1    35361 35433   1.715 0.1903017    
## <none>                              35360 35434                      
## + revol_bal                    1    35359 35435   0.046 0.8301274    
## + mths_since_last_record       1    35360 35436   0.003 0.9563428    
## - pub_rec                      1    35366 35438   6.709 0.0095944 ** 
## - mths_since_last_delinq       1    35369 35441   9.105 0.0025486 ** 
## - installment                  1    35371 35443  11.245 0.0007985 ***
## - collections_12_mths_ex_med   1    35371 35443  11.396 0.0007362 ***
## - delinq_2yrs                  1    35374 35446  14.815 0.0001186 ***
## - inq_last_6mths               1    35375 35447  15.256 9.387e-05 ***
## - verification_status          2    35377 35447  17.271 0.0001777 ***
## - revol_util                   1    35381 35453  21.968 2.773e-06 ***
## - annual_inc                   1    35386 35458  26.767 2.296e-07 ***
## - int_rate                     1    35422 35494  62.159 3.168e-15 ***
## - term                         1    35430 35502  70.525 < 2.2e-16 ***
## - total_acc                    1    35436 35508  76.195 < 2.2e-16 ***
## - emp_length                  11    35459 35511  99.882 < 2.2e-16 ***
## - open_acc                     1    35446 35518  86.769 < 2.2e-16 ***
## - home_ownership               3    35560 35628 200.483 < 2.2e-16 ***
## - dti                          1    35601 35673 241.783 < 2.2e-16 ***
## - grade                        6    35616 35678 256.169 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Step:  AIC=35431.35
## loan_status ~ grade + dti + loan_amnt + home_ownership + term + 
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc + 
##     annual_inc + revol_util + verification_status + inq_last_6mths + 
##     collections_12_mths_ex_med + installment + mths_since_last_delinq + 
##     pub_rec + mths_since_last_major_derog
## 
##                               Df Deviance   AIC     LRT  Pr(>Chi)    
## - loan_amnt                    1    35357 35431   1.707 0.1913507    
## <none>                              35355 35431                      
## + revol_bal                    1    35355 35433   0.048 0.8262588    
## + mths_since_last_record       1    35355 35433   0.001 0.9703426    
## - mths_since_last_major_derog  1    35360 35434   4.163 0.0413185 *  
## - pub_rec                      1    35362 35436   6.501 0.0107836 *  
## - installment                  1    35367 35441  11.253 0.0007950 ***
## - collections_12_mths_ex_med   1    35368 35442  12.529 0.0004006 ***
## - mths_since_last_delinq       1    35369 35443  13.223 0.0002765 ***
## - inq_last_6mths               1    35370 35444  14.980 0.0001087 ***
## - verification_status          2    35373 35445  17.628 0.0001487 ***
## - delinq_2yrs                  1    35371 35445  16.117 5.955e-05 ***
## - revol_util                   1    35377 35451  21.194 4.151e-06 ***
## - annual_inc                   1    35382 35456  27.139 1.894e-07 ***
## - int_rate                     1    35417 35491  61.293 4.919e-15 ***
## - term                         1    35426 35500  70.618 < 2.2e-16 ***
## - total_acc                    1    35431 35505  75.528 < 2.2e-16 ***
## - emp_length                  11    35456 35510 100.299 < 2.2e-16 ***
## - open_acc                     1    35441 35515  85.768 < 2.2e-16 ***
## - home_ownership               3    35557 35627 201.702 < 2.2e-16 ***
## - dti                          1    35596 35670 240.364 < 2.2e-16 ***
## - grade                        6    35609 35673 253.873 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=35431.06
## loan_status ~ grade + dti + home_ownership + term + emp_length +
##     int_rate + delinq_2yrs + total_acc + open_acc + annual_inc +
##     revol_util + verification_status + inq_last_6mths + collections_12_mths_ex_med +
##     installment + mths_since_last_delinq + pub_rec + mths_since_last_major_derog
##
##                               Df Deviance   AIC     LRT  Pr(>Chi)
## <none>                              35357 35431
## + loan_amnt                    1    35355 35431   1.707 0.1913507
## + revol_bal                    1    35357 35433   0.067 0.7962285
## + mths_since_last_record       1    35357 35433   0.001 0.9733566
## - mths_since_last_major_derog  1    35361 35433   4.171 0.0411216 *
## - pub_rec                      1    35364 35436   6.564 0.0104046 *
## - collections_12_mths_ex_med   1    35370 35442  12.585 0.0003888 ***
## - mths_since_last_delinq       1    35370 35442  13.192 0.0002811 ***
## - inq_last_6mths               1    35372 35444  15.327 9.044e-05 ***
## - verification_status          2    35375 35445  17.542 0.0001552 ***
## - delinq_2yrs                  1    35373 35445  15.956 6.482e-05 ***
## - revol_util                   1    35378 35450  21.352 3.821e-06 ***
## - annual_inc                   1    35386 35458  28.465 9.540e-08 ***
## - int_rate                     1    35418 35490  60.728 6.554e-15 ***
## - total_acc                    1    35432 35504  75.417 < 2.2e-16 ***
## - emp_length                  11    35457 35509 100.030 < 2.2e-16 ***
## - open_acc                     1    35443 35515  86.168 < 2.2e-16 ***
## - installment                  1    35539 35611 182.291 < 2.2e-16 ***
## - home_ownership               3    35559 35627 202.234 < 2.2e-16 ***
## - dti                          1    35597 35669 240.002 < 2.2e-16 ***
## - grade                        6    35609 35671 252.185 < 2.2e-16 ***
## - term                         1    35650 35722 293.043 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:  glm(formula = loan_status ~ grade + dti + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc +
##     annual_inc + revol_util + verification_status + inq_last_6mths +
##     collections_12_mths_ex_med + installment + mths_since_last_delinq +
##     pub_rec + mths_since_last_major_derog, family = binomial(link = "logit"),
##     data = full_data.omit)
##
## Coefficients:
##                        (Intercept)                              gradeB
##                          3.434e+00                          -5.621e-01
##                             gradeC                              gradeD
##                         -1.138e+00                          -1.644e+00
##                             gradeE                              gradeF
##                         -2.144e+00                          -2.411e+00
##                             gradeG                                 dti
##                         -2.689e+00                          -2.710e-02
##                  home_ownershipOWN                  home_ownershipRENT
##                         -2.842e-01                          -4.066e-01
##                  home_ownershipANY                                term
##                         -1.058e+00                          -2.481e-02
##                   emp_length1 year                 emp_length10+ years
##                          6.867e-02                           3.027e-01
##                  emp_length2 years                   emp_length3 years
##                          1.534e-01                           1.718e-01
##                  emp_length4 years                   emp_length5 years
##                          1.559e-01                           1.778e-01
##                  emp_length6 years                   emp_length7 years
##                          2.102e-01                           2.829e-01
##                  emp_length8 years                   emp_length9 years
##                          1.571e-01                           1.667e-01
##                      emp_lengthn/a                            int_rate
##                         -1.651e-01                           7.601e-02
##                        delinq_2yrs                           total_acc
##                         -4.956e-02                           1.239e-02
##                           open_acc                          annual_inc
##                         -3.128e-02                           1.922e-06
##                         revol_util  verification_statusSource Verified
##                         -2.941e-03                          -1.512e-01
##        verification_statusVerified                      inq_last_6mths
##                         -1.317e-01                          -4.745e-02
##         collections_12_mths_ex_med                         installment
##                         -2.149e-01                          -8.799e-04
##             mths_since_last_delinq                             pub_rec
##                          3.070e-03                          -3.824e-02
##        mths_since_last_major_derog
##                         -1.607e-03
##
## Degrees of Freedom: 36221 Total (i.e. Null);  36185 Residual
## Null Deviance:       38610
## Residual Deviance: 35360     AIC: 35430
```

The step function ignored redundant variables (the variables that are highly correlated). The best models shown below:

```
final.model <- glm(formula = loan_status ~ grade + dti + home_ownership + term +
    emp_length + int_rate + delinq_2yrs + total_acc + open_acc +
    annual_inc + revol_util + verification_status + inq_last_6mths +
    collections_12_mths_ex_med + installment + mths_since_last_delinq +
    pub_rec + mths_since_last_major_derog, family = binomial(link = "logit"),
    data = full_data.omit)

summary(final.model)
```

```
##
## Call:
## glm(formula = loan_status ~ grade + dti + home_ownership + term +
##     emp_length + int_rate + delinq_2yrs + total_acc + open_acc +
##     annual_inc + revol_util + verification_status + inq_last_6mths +
##     collections_12_mths_ex_med + installment + mths_since_last_delinq +
##     pub_rec + mths_since_last_major_derog, family = binomial(link = "logit"),
##     data = full_data.omit)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -2.5369   0.3621   0.5534   0.7258   2.4048
##
## Coefficients:
##                                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)                         3.434e+00  1.442e-01  23.817  < 2e-16
## gradeB                             -5.621e-01  8.512e-02  -6.603 4.02e-11
## gradeC                             -1.138e+00  9.916e-02 -11.481  < 2e-16
## gradeD                             -1.644e+00  1.240e-01 -13.260  < 2e-16
## gradeE                             -2.144e+00  1.515e-01 -14.150  < 2e-16
## gradeF                             -2.411e+00  1.911e-01 -12.618  < 2e-16
## gradeG                             -2.689e+00  2.348e-01 -11.452  < 2e-16
## dti                                -2.710e-02  1.755e-03 -15.446  < 2e-16
## home_ownershipOWN                  -2.842e-01  4.494e-02  -6.324 2.55e-10
## home_ownershipRENT                 -4.066e-01  2.899e-02 -14.029  < 2e-16
## home_ownershipANY                  -1.058e+00  9.513e-01  -1.113 0.265872
## term                               -2.481e-02  1.436e-03 -17.272  < 2e-16
## emp_length1 year                    6.867e-02  7.433e-02   0.924 0.355566
## emp_length10+ years                 3.027e-01  5.543e-02   5.461 4.72e-08
## emp_length2 years                   1.534e-01  6.833e-02   2.245 0.024757
## emp_length3 years                   1.718e-01  6.924e-02   2.481 0.013086
## emp_length4 years                   1.559e-01  7.310e-02   2.132 0.032993
## emp_length5 years                   1.778e-01  7.349e-02   2.420 0.015540
## emp_length6 years                   2.102e-01  7.920e-02   2.654 0.007956
## emp_length7 years                   2.829e-01  8.161e-02   3.467 0.000527
## emp_length8 years                   1.571e-01  7.973e-02   1.971 0.048715
## emp_length9 years                   1.667e-01  8.529e-02   1.955 0.050639
## emp_lengthn/a                      -1.651e-01  6.911e-02  -2.389 0.016912
## int_rate                            7.601e-02  9.796e-03   7.760 8.51e-15
## delinq_2yrs                        -4.956e-02  1.219e-02  -4.067 4.77e-05
## total_acc                           1.239e-02  1.444e-03   8.586  < 2e-16
## open_acc                           -3.128e-02  3.359e-03  -9.312  < 2e-16
## annual_inc                          1.922e-06  3.750e-07   5.125 2.98e-07
## revol_util                         -2.941e-03  6.367e-04  -4.619 3.85e-06
## verification_statusSource Verified -1.512e-01  3.702e-02  -4.085 4.41e-05
## verification_statusVerified        -1.317e-01  3.934e-02  -3.348 0.000814
## inq_last_6mths                     -4.745e-02  1.208e-02  -3.929 8.51e-05
## collections_12_mths_ex_med         -2.149e-01  5.958e-02  -3.607 0.000310
## installment                        -8.799e-04  6.495e-05 -13.547  < 2e-16
## mths_since_last_delinq              3.070e-03  8.439e-04   3.638 0.000275
## pub_rec                            -3.824e-02  1.475e-02  -2.592 0.009546
## mths_since_last_major_derog        -1.607e-03  7.846e-04  -2.048 0.040558
##
## (Intercept)                        ***
## gradeB                             ***
## gradeC                             ***
## gradeD                             ***
## gradeE                             ***
## gradeF                             ***
## gradeG                             ***
## dti                                ***
## home_ownershipOWN                  ***
## home_ownershipRENT                 ***
## home_ownershipANY
## term                               ***
## emp_length1 year
## emp_length10+ years                ***
## emp_length2 years                  *
## emp_length3 years                  *
## emp_length4 years                  *
## emp_length5 years                  *
## emp_length6 years                  **
## emp_length7 years                  ***
## emp_length8 years                  *
## emp_length9 years                  .
## emp_lengthn/a                      *
## int_rate                           ***
## delinq_2yrs                        ***
## total_acc                          ***
## open_acc                           ***
## annual_inc                         ***
## revol_util                         ***
## verification_statusSource Verified ***
## verification_statusVerified        ***
## inq_last_6mths                     ***
## collections_12_mths_ex_med         ***
## installment                        ***
## mths_since_last_delinq             ***
## pub_rec                            **
## mths_since_last_major_derog        *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 38610  on 36221  degrees of freedom
## Residual deviance: 35357  on 36185  degrees of freedom
## AIC: 35431
##
## Number of Fisher Scoring iterations: 4
```

By looking at the summary statistics, I concluded that the variables "home_ownershipANY", "emp_length1 year" and "emp_length9 years" are not statistically significant (they don't affect the outcome) because theirs p-value is greater that 0.05 level of significance.

Based on the results of the glm function I built the following logit function:

```
#logit(p) = ln(p/(p-1)) =
#logit_p=
#(3.434e+00)-(5.621e-01)*gradeB-(1.138e+00)*gradeC-(1.644e+00)*gradeD-(2.144e+00)*gradeE -(2.144e+00)*g
```

The natural logarithm of the odds ratio is equivalent to a linear function of the independent variables. The antilog of the logit function allows us to find the estimated regression equation.

The estimated regression equation is shown below:

```
#p_hat = e^logit_p/(1+e^logit_p)
```

In order to illustrate how the equation works I will show how grade can affect the probability of whether the loan will be paid off or not.

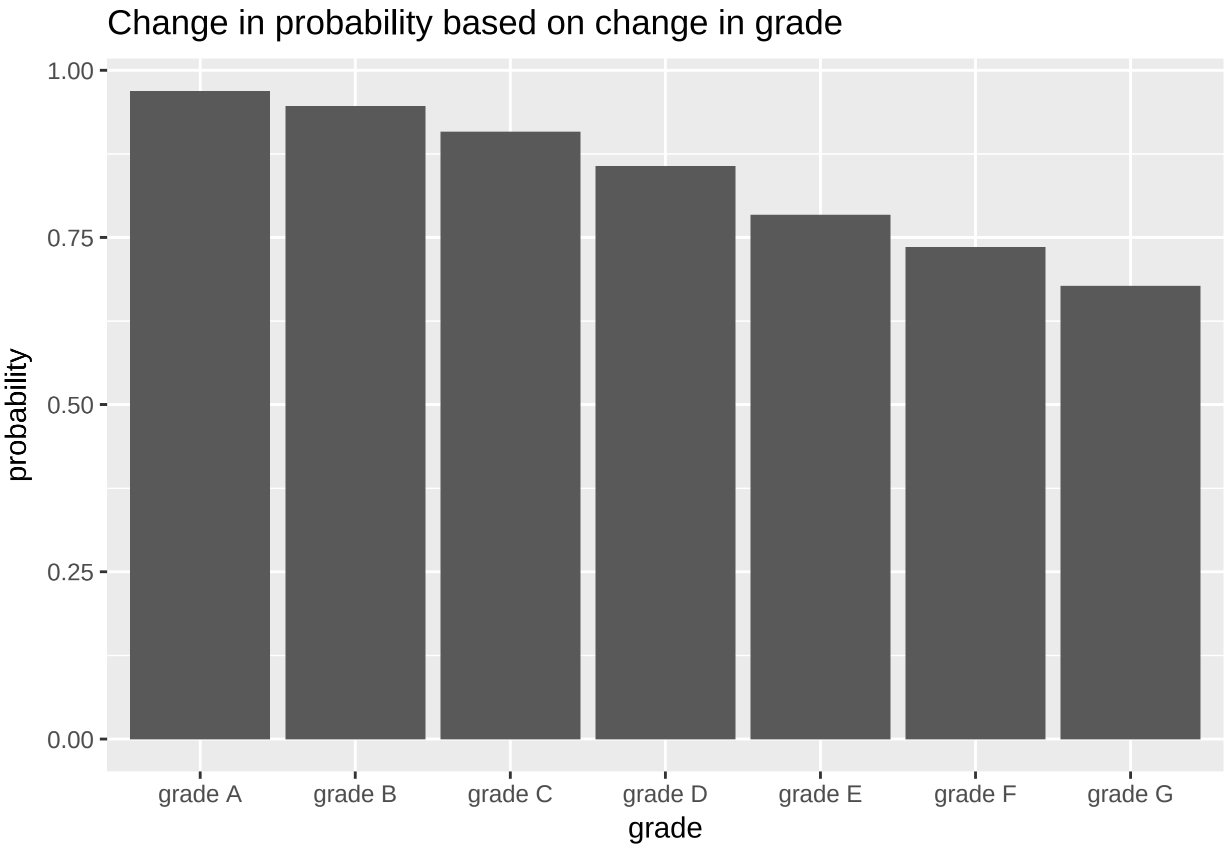
Let's change the loan grade and hold all remaining variables constant.

```r
#create vectors to store grades and probabilities
grade <- c()
probability <- c()

#grade A
logit_p_A = (3.434e+00)-(5.621e-01)*0-(1.138e+00)*0-(1.644e+00)*0-(2.144e+00)*0 -(2.411e+00)*0-(2.689e+
probability[1] = exp(1)^logit_p_A/(1+exp(1)^logit_p_A)
grade[1] <- "grade A"

#grade B
logit_p_B = (3.434e+00)-(5.621e-01)*1-(1.138e+00)*0-(1.644e+00)*0-(2.144e+00)*0 -(2.411e+00)*0-(2.689e+
probability[2] = exp(1)^logit_p_B/(1+exp(1)^logit_p_B)
grade[2] <- "grade B"

#grade C
logit_p_C = (3.434e+00)-(5.621e-01)*0-(1.138e+00)*1-(1.644e+00)*0-(2.144e+00)*0 -(2.411e+00)*0-(2.689e+
probability[3] = exp(1)^logit_p_C/(1+exp(1)^logit_p_C)
grade[3] <- "grade C"

#grade D
logit_p_D = (3.434e+00)-(5.621e-01)*0-(1.138e+00)*0-(1.644e+00)*1-(2.144e+00)*0 -(2.411e+00)*0-(2.689e+
probability[4] = exp(1)^logit_p_D/(1+exp(1)^logit_p_D)
grade[4] <- "grade D"

#grade E
logit_p_E = (3.434e+00)-(5.621e-01)*0-(1.138e+00)*0-(1.644e+00)*0-(2.144e+00)*1 -(2.411e+00)*0-(2.689e+
probability[5] = exp(1)^logit_p_E/(1+exp(1)^logit_p_E)
grade[5] <- "grade E"

#grade F
logit_p_F = (3.434e+00)-(5.621e-01)*0-(1.138e+00)*0-(1.644e+00)*0-(2.144e+00)*0 -(2.411e+00)*1-(2.689e+
probability[6] = exp(1)^logit_p_F/(1+exp(1)^logit_p_F)
grade[6] <- "grade F"

#grade G
logit_p_G = (3.434e+00)-(5.621e-01)*0-(1.138e+00)*0-(1.644e+00)*0-(2.144e+00)*0 -(2.411e+00)*0-(2.689e+
probability[7] = exp(1)^logit_p_G/(1+exp(1)^logit_p_G)
grade[7] <- "grade G"

#table with results
prob_table <- data.frame(grade,probability)
head(prob_table)
```

```
##     grade probability
## 1 grade A   0.9687504
## 2 grade B   0.9464397
## 3 grade C   0.9085452
## 4 grade D   0.8569273
## 5 grade E   0.7841472
## 6 grade F   0.7355566
```

```r
#graph
g <- ggplot(prob_table, aes(grade,probability))
```

```
g + ggtitle("Change in probability based on change in grade") + geom_bar(stat="identity")
```

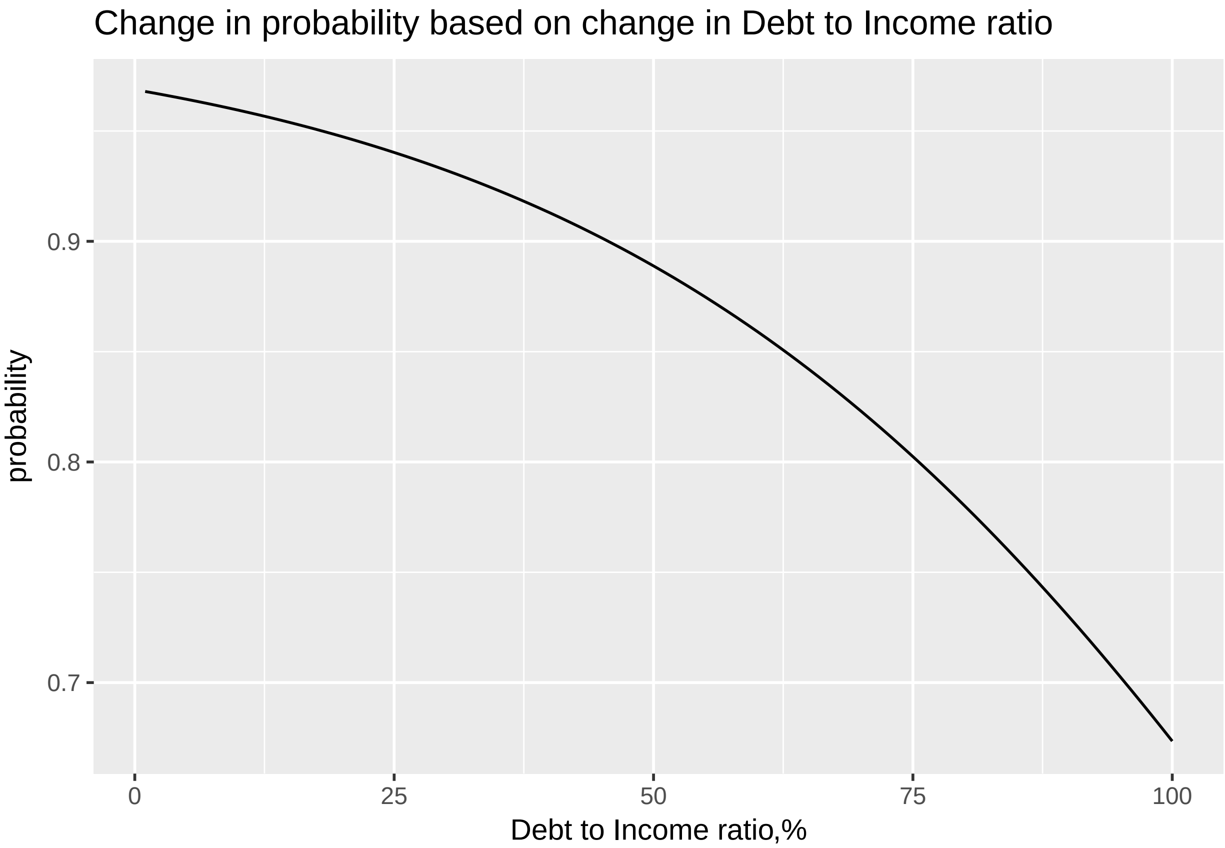


The bar chart above shows how change in grades reflects probability while holding all other variables constant.

```
#create vectors to store change in dti and probabilitiies
dti_value <- c()
probability <- c()

for (i in 1:100){

dti_value[i] <- i
logit_p_dti = (3.434e+00)-(2.710e-02)*i

probability[i] = exp(1)^logit_p_dti/(1+exp(1)^logit_p_dti)

}

dti_table <- data.frame(dti_value,probability)
head(dti_table)
```

```
##   dti_value probability
## 1         1   0.9679195
## 2         2   0.9670672
## 3         3   0.9661931
## 4         4   0.9652967
## 5         5   0.9643773
## 6         6   0.9634345
```

```
#graph
g <- ggplot(data=dti_table, aes(x=dti_value,y=probability))
g + geom_line()+ ggtitle("Change in probability based on change in Debt to Income ratio")+labs(x="Debt
```



The bar chart above shows how change in debt to income ratio reflects probability while holding all other variables constant.

By using logit function investor can verify how change in variable or combinations of variables can affect probability while holding all other variables constant.

The second part of my research focuses on finding a classifier that can predict whether the loan belongs to paid off or charged off class with higher accuracy. I considered J48, Naive Bayes, Ibk classifier and SMO classifier.

Before running classifiers I adjusted the data set, so that, it doesn't contain redundant or statistically insignificant variables.

```
#full_data <- full_data %>% select(-loan_amnt,-revol_bal,-mths_since_last_record)
```

J48 classifier. It uses ID3 algorithm that constructs the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node.

```
#create vectors to store accuracies values and errors values

classifier_name <-c()
Accuracy_mean <- c()
Accuracy_upper <- c()
Accuracy_lower <- c()
Kappa <- c()
```

```
MAE <- c()
RMSE <- c()
RAE <- c()
RRSE <- c()
```

I used 66% of data for training and 34% for testing. In order to reduce the error I generated training set 50 times and ran each classifier 50 times.

```
classifier_name[1] <-c("J48")

#create empty vectors to store accuracies values and errors values for a certain classifier
pctCorrect_vector <- c()
Kappa_vector <- c()
MAE_vector <- c()
RMSE_vector <- c()
RAE_vector <- c()
RRSE_vector <- c()

for (i in 1:50){

    #set the seed of R‘s random number generator
    set.seed(i)
    #create 66% training data set
    training <- full_data[sample(nrow(full_data)),][1:round(0.66*nrow(full_data)),]
    resultJ48 <- J48(loan_status ~., training)
    evaluation <- evaluate_Weka_classifier(resultJ48)$details
    pctCorrect_value <- evaluation["pctCorrect"]
    Kappa_value <- evaluation["kappa"]
    MAE_value <- evaluation["meanAbsoluteError"]
    RMSE_value <- evaluation["rootMeanSquaredError"]
    RAE_value <- evaluation["rootMeanSquaredError"]
    RRSE_value <- evaluation["rootRelativeSquaredError"]

    #add the value generated after each iteration to the vector
    pctCorrect_vector <- c(pctCorrect_vector,pctCorrect_value)
    Kappa_vector <- c(Kappa_vector, Kappa_value)
    MAE_vector <- c(MAE_vector, MAE_value)
    RMSE_vector <- c(RMSE_vector, RMSE_value)
    RAE_vector <- c(RAE_vector, RAE_value)
    RRSE_vector <- c(RRSE_vector, RRSE_value)

}

#calculate mean and standard deviation of accuracies and all errors
Accuracy_mean[1] <- mean(pctCorrect_vector)
pctCorrect_sd <- sd(pctCorrect_vector)
Accuracy_upper[1] <-Accuracy_mean[1] + pctCorrect_sd
Accuracy_lower <-Accuracy_mean[1] - pctCorrect_sd
Kappa[1] <- mean(Kappa_vector)
MAE[1] <- mean(MAE_vector)
RMSE[1] <- mean(RMSE_vector)
RAE[1] <- mean(RAE_vector)
RRSE[1] <- mean(RRSE_vector)
```

Naive Bayes. This classifier is a simple probabilistic classifier that works based on applying Bayes' theorem with strong independence assumptions. Naive Bayes classifiers can handle an arbitrary number of independent variables whether continuous or categorical

```r
classifier_name[2] <-c("Naive Bayes")

#create empty vectors to store accuracies values and errors values
pctCorrect_vector <- c()
Kappa_vector <- c()
MAE_vector <- c()
RMSE_vector <- c()
RAE_vector <- c()
RRSE_vector <- c()
for (i in 1:50){
    set.seed(i)
    #create 66% training data set
    training <- full_data[sample(nrow(full_data)),][1:round(0.66*nrow(full_data)),]
    NB <- make_Weka_classifier("weka/classifiers/bayes/NaiveBayes")
    result_NaiveBayes <- NB(loan_status ~., training)
    evaluation <- evaluate_Weka_classifier(result_NaiveBayes)$details
    pctCorrect_value <- evaluation["pctCorrect"]
    Kappa_value <- evaluation["kappa"]
    MAE_value <- evaluation["meanAbsoluteError"]
    RMSE_value <- evaluation["rootMeanSquaredError"]
    RAE_value <- evaluation["rootMeanSquaredError"]
    RRSE_value <- evaluation["rootRelativeSquaredError"]

    #add the value generated after each iteration to the vector
    pctCorrect_vector <- c(pctCorrect_vector,pctCorrect_value)
    Kappa_vector <- c(Kappa_vector, Kappa_value)
    MAE_vector <- c(MAE_vector, MAE_value)
    RMSE_vector <- c(RMSE_vector, RMSE_value)
    RAE_vector <- c(RAE_vector, RAE_value)
    RRSE_vector <- c(RRSE_vector, RRSE_value)
}

#calculate mean and standard deviation of accuracies and all errors
Accuracy_mean[2] <- mean(pctCorrect_vector)
pctCorrect_sd[2] <- sd(pctCorrect_vector)
Accuracy_upper[2] <-Accuracy_mean[2] + pctCorrect_sd[2]
Accuracy_lower[2] <-Accuracy_mean[2] - pctCorrect_sd[2]
Kappa[2] <- mean(Kappa_vector)
MAE[2] <- mean(MAE_vector)
RMSE[2] <- mean(RMSE_vector)
RAE[2] <- mean(RAE_vector)
RRSE[2] <- mean(RRSE_vector)
```

IBk classifier. It implements the k-nearest neighbor algorithm that stores all available cases and classifies new cases based on a similarity measure such as Euclidean distance, Manhattan Distance or Makowski distance.

```r
classifier_name[3] <-c("Knn")

#ignore the loop for this classifier since it loads very slow
#for (i in 1:50){
    #set.seed(i)
    #create 66% training data set
    training <- full_data[sample(nrow(full_data)),][1:round(0.66*nrow(full_data)),]
    knn <- IBk(loan_status ~., training)
    evaluation <- evaluate_Weka_classifier(knn)$details
    pctCorrect_value <- evaluation["pctCorrect"]
    Kappa_value <- evaluation["Kappa"]
    MAE_value <- evaluation["meanAbsoluteError"]
    RMSE_value <- evaluation["rootMeanSquaredError"]
    RAE_value <- evaluation["rootMeanSquaredError"]
    RRSE_value <- evaluation["rootRelativeSquaredError"]

#}

#calculate mean and standard deviation of accuracies and all errors
Accuracy_mean[3] <- mean(pctCorrect_value)
#pctCorrect_sd[3] <- sd(pctCorrect_value)
Accuracy_upper[3] <- ""
Accuracy_lower[3] <- ""
Kappa[3] <- mean(Kappa_value)
MAE[3] <- mean(MAE_value)
RMSE[3] <- mean(RMSE_value)
RAE[3] <- mean(RAE_value)
RRSE[3] <- mean(RRSE_value)
```

Support Vector Machine (SMO classifier) performs classification by constructing an N-dimensional hyper plane that optimally separates the data into two categories. SVM models are closely related to neural networks. SVM model that is uses a sigmoid kernel function s equivalent to a two-layer perceptron neural network.

```
classifier_name[4] <-c("SVM")

#ignore the loop for this classifier since it loads very slow
#for (i in 1:50){
    #set.seed(i)
    #create 66% training data set
    training <- full_data[sample(nrow(full_data)),][1:round(0.66*nrow(full_data)),]
    SVM <- SMO(loan_status ~., training)
    evaluation <- evaluate_Weka_classifier(SVM)$details
    pctCorrect_value <- evaluation["pctCorrect"]
    Kappa_value <- evaluation["kappa"]
    MAE_value <- evaluation["meanAbsoluteError"]
    RMSE_value <- evaluation["rootMeanSquaredError"]
    RAE_value <- evaluation["rootMeanSquaredError"]
    RRSE_value <- evaluation["rootRelativeSquaredError"]
#}

    #calculate mean and standard deviation of accuracies and all errors
Accuracy_mean[4] <- mean(pctCorrect_value)
#pctCorrect_sd[4] <- sd(pctCorrect_value)
Accuracy_upper[4] <- ""
Accuracy_lower[4] <- ""
Kappa[4] <- mean(Kappa_value)
MAE[4] <- mean(MAE_value)
RMSE[4] <- mean(RMSE_value)
RAE[4] <- mean(RAE_value)
```

```
RRSE[4] <- mean(RRSE_value)
```

```
data_table <- data.frame(classifier_name, Accuracy_mean, Accuracy_upper, Accuracy_lower, Kappa, MAE, RMS
data_table
```

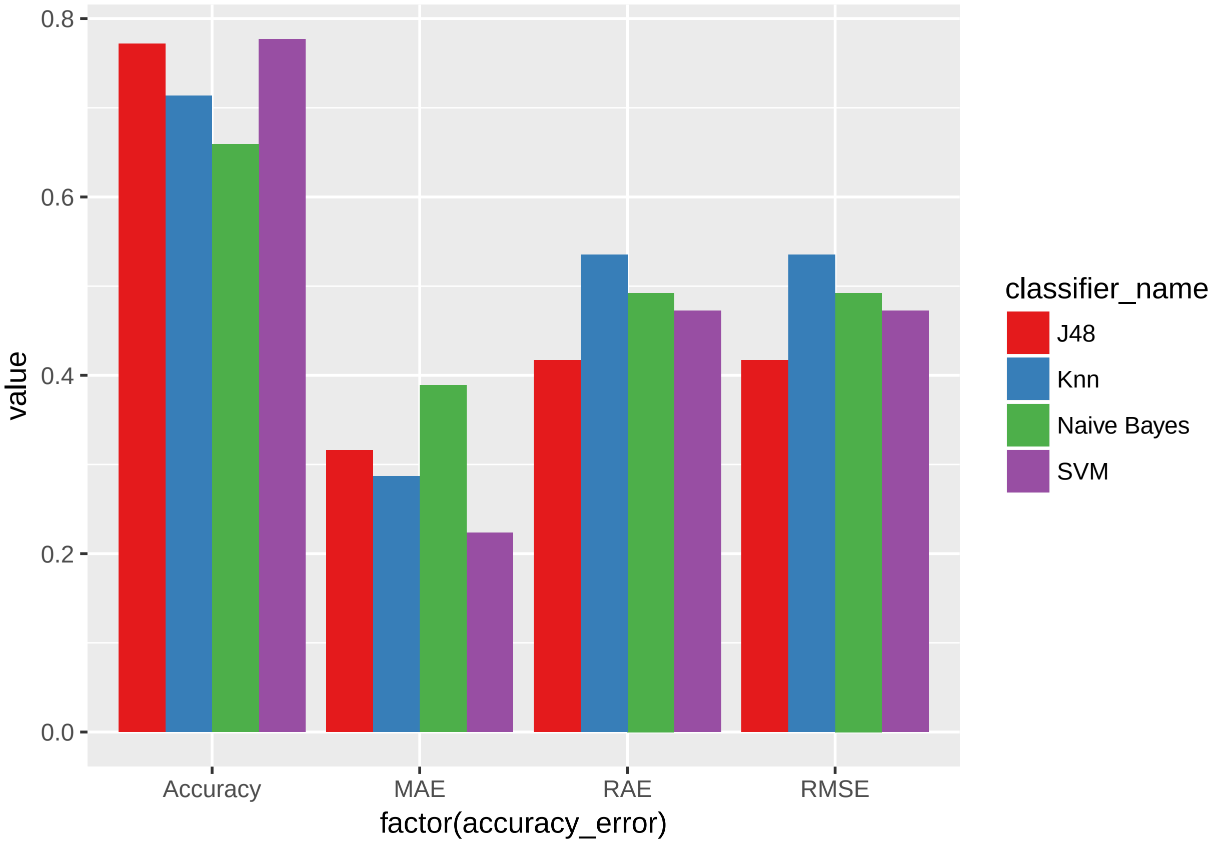
```
##   classifier_name Accuracy_mean   Accuracy_upper   Accuracy_lower
## 1             J48      77.20244  77.474168268397 76.9307023318712
## 2     Naive Bayes      65.88770 67.1079614504185 64.6674348453141
## 3             Knn      71.33875
## 4             SVM      77.66718
##        Kappa       MAE      RMSE       RAE      RRSE
## 1 0.04889376 0.3159595 0.4167403 0.4167403  99.74316
## 2 0.14666984 0.3886717 0.4922871 0.4922871 117.82535
## 3         NA 0.2866303 0.5353396 0.5353396 128.48147
## 4 0.00000000 0.2233282 0.4725761 0.4725761 113.47004
```

I considered the following errors in my analysis:

1. Mean Absolute Error (MAE) shows the deviation between predicted and actual outcome.
2. Root Mean Squared Error (RMSD) measures the error between predicted and actual results.
3. Relative Absolute Error (RAE) measures by normalizing with respect to the performance obtained by predicting the classes' prior probabilities as estimated from the training data with a simple Laplace estimator.

```
#rename accuracy_mean column and divide by 100
#converting wide data format to long data
data <- data_table %>% rename(Accuracy = Accuracy_mean) %>% mutate(Accuracy = Accuracy/100) %>% select(
data <- data %>% gather(accuracy_error,value,colnames(data)[2]:colnames(data)[5])
```

```
ggplot(data, aes(factor(accuracy_error), value, fill = classifier_name)) + geom_bar(stat="identity", po
```



According to the bar chart above, Support Vector Mashine classifier and J48 classifier have higher accuracies while J48 classifier has pretty low MAE, RMSD and RAE. I can conclude that J48 classifier is the optimal classifier because it has the heist accuracy and pretty low error values.

Let's check.

```
#34% data for testing
testing <- full_data[sample(nrow(full_data)),][(round(0.66*nrow(full_data))+1):nrow(full_data),]
J48_pred <- predict(resultJ48, newdata = testing)

table(testing$loan_status, J48_pred)
```

```
##              J48_pred
##               Charged Off Fully Paid
##   Charged Off        9826      43511
##   Fully Paid        15715     194053
```

The confusion matrix shows that the accuracy of prediction is pretty high.

The goals of the project have been achieved. I figured out the equation that estimates the probability of weather the load will be paid off or not, demonstrated how different variables can affect the probability and found the optimal classifier that can predict whether the loan will be paid off or not with higher accuracy.

Each part of the project was a challenge for me since I'm new to statistics and R. The biggest problem was that some of the function were running very slow because the initial merged data set contains 773838 observation.