Logistic Regression

Davit Gadyan

August 21, 2018

Logistic Regression

```
credit<-read.csv('C:\\Users\\Gaya\\Desktop\\R\\EDX_Scripts\\Principles-of-Machine-Learning-R-master\\Mod

str(credit)
```

```
## 'data.frame':    1000 obs. of  22 variables:
##  $ Customer_ID             : int  1122334 6156361 2051359 8740590 3924540 3115687 8251714 2272783 18
##  $ checking_account_status : chr  "< 0 DM" "0 - 200 DM" "none" "< 0 DM" ...
##  $ loan_duration_mo        : int  6 48 12 42 24 36 24 36 12 30 ...
##  $ credit_history          : chr  "critical account - other non-bank loans" "current loans paid" "cr
##  $ purpose                 : chr  "radio/television" "radio/television" "education" "furniture/equip
##  $ loan_amount             : int  1169 5951 2096 7882 4870 9055 2835 6948 3059 5234 ...
##  $ savings_account_balance : chr  "unknown/none" "< 100 DM" "< 100 DM" "< 100 DM" ...
##  $ time_employed_yrs       : chr  ">= 7 years" "1 - 4 years" "4 - 7 years" "4 - 7 years" ...
##  $ payment_pcnt_income     : int  4 2 2 2 3 2 3 2 2 4 ...
##  $ gender_status           : chr  "male-single" "female-divorced/separated/married" "male-single" "ma
##  $ other_signators         : chr  "none" "none" "none" "guarantor" ...
##  $ time_in_residence       : int  4 2 3 4 4 4 4 2 4 2 ...
##  $ property                : chr  "real estate" "real estate" "real estate" "building society saving
##  $ age_yrs                 : int  67 22 49 45 53 35 53 35 61 28 ...
##  $ other_credit_outstanding: chr  "none" "none" "none" "none" ...
##  $ home_ownership          : chr  "own" "own" "own" "for free" ...
##  $ number_loans            : int  2 1 1 1 2 1 1 1 1 2 ...
##  $ job_category            : chr  "skilled" "skilled" "unskilled-resident" "skilled" ...
##  $ dependents              : int  1 1 2 2 2 2 1 1 1 1 ...
##  $ telephone               : chr  "yes" "none" "none" "none" ...
##  $ foreign_worker          : chr  "yes" "yes" "yes" "yes" ...
##  $ bad_credit              : int  0 1 0 0 1 0 0 0 0 1 ...
```

```
library(caret)
```

```
## Loading required package: lattice

## Loading required package: ggplot2
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

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

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

```
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```r
library(Metrics)
```

```
##
## Attaching package: 'Metrics'

## The following objects are masked from 'package:caret':
##
##     precision, recall
```

IMPORTANT before modeling anything in classification we need to look at class distribution in our target variable

Let's make factorize bad credits feature for avoiding future problems

```r
credit$bad_credit<-factor(credit$bad_credit, levels = c(0, 1), labels = c('No','Yes'))

levels(credit$bad_credit)
```

```
## [1] "No"  "Yes"
```

```r
as.numeric(credit$bad_credit)
```

```
##    [1] 1 2 1 1 2 1 1 1 1 2 2 2 1 2 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
##   [35] 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 1 1 2 2 1 1 1 1
##   [69] 2 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 2 1 1 2 1 1 1 1 1 1
##  [103] 1 1 1 2 2 1 1 1 1 2 1 1 1 2 1 1 2 1 2 1 2 1 1 1 2 1 1 2 1 2 1 2 1 1
##  [137] 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1
##  [171] 1 2 2 1 2 1 2 2 1 1 1 1 2 2 2 1 2 1 2 1 2 1 2 2 2 1 2 2 1 2 1 2 1 1
##  [205] 1 2 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 2 1 1 1 1 2
##  [239] 2 2 1 1 2 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2
##  [273] 1 1 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 2 2
##  [307] 1 2 1 1 2 2 1 1 1 1 2 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 2 2 2 2 2 1
##  [341] 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 1 2 1 2 1 2 1 1 1 1 2 1 1 1 2 1 1
##  [375] 1 1 1 2 2 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1
##  [409] 2 1 1 1 2 1 1 2 1 2 1 2 1 1 2 1 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 1 1 1
##  [443] 2 1 1 1 2 2 1 2 1 1 2 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 2 2
##  [477] 1 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 2 1 1 1 2 1 1 2 2 2
##  [511] 1 2 1 1 2 1 1 1 1 1 1 2 1 1 1 2 2 1 1 1 1 2 1 1 2 1 1 1 2 1 1 2 1 2
##  [545] 1 2 2 1 2 1 1 2 1 1 1 2 1 1 2 2 2 2 2 1 2 1 2 1 1 2 1 1 2 2 1 1 1 1
##  [579] 1 1 1 2 1 2 1 1 2 1 2 1 1 2 2 1 1 1 2 2 2 2 2 2 1 1 2 2 1 1 1 2 1 1
##  [613] 2 2 1 1 2 1 1 1 2 1 1 2 2 1 2 1 1 2 1 1 1 2 1 2 2 1 1 1 1 2 2 1 2 1
##  [647] 1 2 1 2 2 2 1 2 2 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1
##  [681] 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 2 1 1
##  [715] 1 2 1 1 2 1 1 1 1 1 2 2 2 1 2 1 1 2 2 1 1 2 1 1 1 1 2 1 1 2 1 1 1 1
##  [749] 1 1 1 2 1 1 2 1 1 2 2 2 2 1 2 1 2 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1
##  [783] 1 2 1 1 1 1 2 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 2 2 1
##  [817] 1 1 1 2 1 1 2 1 1 1 2 2 2 1 1 2 2 1 2 2 1 1 1 1 2 1 2 1 1 1 2 1 1 2
##  [851] 2 1 1 2 1 1 1 1 2 1 1 2 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1
##  [885] 2 2 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1 2
##  [919] 2 2 1 1 2 1 2 2 1 2 1 1 1 2 1 1 1 2 2 1 2 1 1 1 1 1 1 1 2 1 2 2 1 2
##  [953] 2 2 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 1 1 1 2 2 1 1 1 1 2 2 2 2 1 2 1 1
```

```
##  [987] 1 1 1 1 1 1 1 1 1 1 1 1 2 1
```

```r
str(credit$bad_credit)
```

```
##  Factor w/ 2 levels "No","Yes": 1 2 1 1 2 1 1 1 1 2 ...
```

```r
table(credit$bad_credit)
```

```
##
##  No Yes
## 700 300
```

Not good there are more than twice 'good' bank customers as we have 'bad' customer (i.e. bad means defaulted and from bank perspective they are 'bad').. This is really nasty situation as we shoud predict more precisely bad customers as risks and costs asssociated with a bad customers are higher compared to leaving oner good customer.In practice we will use imputation and other balancing techniques. . . . or another approach may be adjusting model at the ned towards precisely predicting exactly this class!!!! We will use second approach in this case!!

```r
ind<-createDataPartition(credit$bad_credit,p=0.70,list = F)

credit1<-credit%>%
  dplyr::select(-Customer_ID)

dim(credit1)
```

```
## [1] 1000   21
```

As we see no big resultfrom using this nearZeroVar funtion from caret but anyway useful feature!!!!

```r
credit_train<-credit1[ind,]
credit_test<-credit1[-ind,]

colnames(credit_train)
```

```
##  [1] "checking_account_status"  "loan_duration_mo"
##  [3] "credit_history"           "purpose"
##  [5] "loan_amount"              "savings_account_balance"
##  [7] "time_employed_yrs"        "payment_pcnt_income"
##  [9] "gender_status"            "other_signators"
## [11] "time_in_residence"        "property"
## [13] "age_yrs"                  "other_credit_outstanding"
## [15] "home_ownership"           "number_loans"
## [17] "job_category"             "dependents"
## [19] "telephone"                "foreign_worker"
## [21] "bad_credit"
```

```r
mod1<-glm(bad_credit~.,data=credit_train,family = 'binomial')

mod2<-stepAIC(mod1,method='both')
```

```
## Start:  AIC=715.95
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + time_employed_yrs +
##     payment_pcnt_income + gender_status + other_signators + time_in_residence +
##     property + age_yrs + other_credit_outstanding + home_ownership +
##     number_loans + job_category + dependents + telephone + foreign_worker
##
##                            Df Deviance    AIC
```

```
## - job_category             3   618.51 710.51
## - time_employed_yrs        4   623.31 713.31
## - time_in_residence        1   618.16 714.16
## - telephone                1   618.35 714.35
## - property                 3   622.38 714.38
## - dependents               1   618.78 714.78
## - other_signators          2   621.59 715.59
## - home_ownership           2   621.94 715.94
## <none>                         617.95 715.95
## - age_yrs                  1   621.10 717.10
## - other_credit_outstanding  2   623.17 717.17
## - gender_status            3   625.42 717.42
## - payment_pcnt_income      1   621.74 717.74
## - foreign_worker           1   621.86 717.86
## - number_loans             1   622.53 718.53
## - loan_amount              1   625.44 721.44
## - loan_duration_mo         1   625.84 721.84
## - savings_account_balance   4   633.29 723.29
## - purpose                  9   651.08 731.08
## - credit_history           4   641.17 731.17
## - checking_account_status   3   655.62 747.62
##
## Step:  AIC=710.51
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + time_employed_yrs +
##     payment_pcnt_income + gender_status + other_signators + time_in_residence +
##     property + age_yrs + other_credit_outstanding + home_ownership +
##     number_loans + dependents + telephone + foreign_worker
##
##                            Df Deviance    AIC
## - time_employed_yrs         4   623.83 707.83
## - time_in_residence         1   618.64 708.64
## - property                  3   622.78 708.78
## - telephone                 1   619.21 709.21
## - dependents                1   619.30 709.30
## - other_signators           2   622.06 710.06
## - home_ownership            2   622.34 710.34
## <none>                          618.51 710.51
## - age_yrs                   1   621.37 711.37
## - gender_status             3   625.73 711.73
## - other_credit_outstanding  2   624.05 712.05
## - foreign_worker            1   622.27 712.27
## - payment_pcnt_income       1   622.42 712.42
## - number_loans              1   623.00 713.00
## - loan_duration_mo          1   626.32 716.32
## - loan_amount               1   626.35 716.35
## - savings_account_balance   4   634.34 718.34
## - credit_history            4   641.45 725.45
## - purpose                   9   652.30 726.30
## - checking_account_status   3   656.75 742.75
##
## Step:  AIC=707.83
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
```

```
##     gender_status + other_signators + time_in_residence + property +
##     age_yrs + other_credit_outstanding + home_ownership + number_loans +
##     dependents + telephone + foreign_worker
##
##                            Df Deviance    AIC
## - time_in_residence         1   624.00 706.00
## - property                  3   628.41 706.41
## - dependents                1   624.50 706.50
## - telephone                 1   624.88 706.88
## - home_ownership            2   627.42 707.42
## - other_signators           2   627.72 707.72
## <none>                          623.83 707.83
## - age_yrs                   1   626.49 708.49
## - foreign_worker            1   627.40 709.40
## - other_credit_outstanding  2   629.79 709.79
## - payment_pcnt_income       1   628.01 710.01
## - number_loans              1   628.07 710.07
## - gender_status             3   632.23 710.23
## - loan_duration_mo          1   630.64 712.64
## - loan_amount               1   631.75 713.75
## - savings_account_balance   4   640.28 716.28
## - purpose                   9   656.17 722.17
## - credit_history            4   647.06 723.06
## - checking_account_status   3   661.97 739.97
##
## Step:  AIC=706
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + property + age_yrs + other_credit_outstanding +
##     home_ownership + number_loans + dependents + telephone +
##     foreign_worker
##
##                            Df Deviance    AIC
## - property                  3   628.55 704.55
## - dependents                1   624.69 704.69
## - telephone                 1   625.14 705.14
## - home_ownership            2   627.44 705.44
## - other_signators           2   627.86 705.86
## <none>                          624.00 706.00
## - age_yrs                   1   627.17 707.17
## - foreign_worker            1   627.54 707.54
## - other_credit_outstanding  2   630.10 708.10
## - number_loans              1   628.17 708.17
## - payment_pcnt_income       1   628.21 708.21
## - gender_status             3   632.40 708.40
## - loan_duration_mo          1   630.76 710.76
## - loan_amount               1   632.09 712.09
## - savings_account_balance   4   640.77 714.77
## - purpose                   9   656.63 720.63
## - credit_history            4   647.29 721.29
## - checking_account_status   3   661.98 737.98
##
## Step:  AIC=704.55
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
```

```
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + age_yrs + other_credit_outstanding +
##     home_ownership + number_loans + dependents + telephone +
##     foreign_worker
##
##                            Df Deviance    AIC
## - telephone                 1   629.20 703.20
## - home_ownership            2   631.25 703.25
## - dependents                1   629.42 703.42
## <none>                          628.55 704.55
## - other_signators           2   633.23 705.23
## - age_yrs                   1   631.90 705.90
## - foreign_worker            1   631.98 705.98
## - number_loans              1   632.19 706.19
## - gender_status             3   636.83 706.83
## - other_credit_outstanding  2   635.12 707.12
## - payment_pcnt_income       1   633.55 707.55
## - loan_duration_mo          1   636.56 710.56
## - loan_amount               1   638.46 712.46
## - savings_account_balance   4   645.20 713.20
## - credit_history            4   652.34 720.34
## - purpose                   9   663.11 721.11
## - checking_account_status   3   668.16 738.16
##
## Step:  AIC=703.2
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + age_yrs + other_credit_outstanding +
##     home_ownership + number_loans + dependents + foreign_worker
##
##                            Df Deviance    AIC
## - home_ownership            2   631.92 701.92
## - dependents                1   630.08 702.08
## <none>                          629.20 703.20
## - other_signators           2   633.72 703.72
## - foreign_worker            1   632.39 704.39
## - number_loans              1   632.64 704.64
## - age_yrs                   1   633.00 705.00
## - gender_status             3   637.53 705.53
## - other_credit_outstanding  2   635.67 705.67
## - payment_pcnt_income       1   633.83 705.83
## - loan_duration_mo          1   637.59 709.59
## - loan_amount               1   638.47 710.47
## - savings_account_balance   4   646.06 712.06
## - credit_history            4   652.81 718.81
## - purpose                   9   664.29 720.29
## - checking_account_status   3   669.46 737.46
##
## Step:  AIC=701.92
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + age_yrs + other_credit_outstanding +
##     number_loans + dependents + foreign_worker
##
```

```
##                           Df Deviance    AIC
## - dependents               1   632.79 700.79
## <none>                         631.92 701.92
## - other_signators          2   636.68 702.68
## - foreign_worker           1   635.03 703.03
## - number_loans             1   635.50 703.50
## - other_credit_outstanding  2   637.65 703.65
## - age_yrs                  1   636.20 704.20
## - payment_pcnt_income      1   636.34 704.34
## - gender_status            3   641.31 705.31
## - loan_duration_mo         1   640.36 708.36
## - loan_amount              1   641.26 709.26
## - savings_account_balance   4   648.47 710.47
## - credit_history           4   656.69 718.69
## - purpose                  9   666.90 718.90
## - checking_account_status   3   674.96 738.96
##
## Step:  AIC=700.79
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + age_yrs + other_credit_outstanding +
##     number_loans + foreign_worker
##
##                           Df Deviance    AIC
## <none>                         632.79 700.79
## - foreign_worker           1   635.85 701.85
## - other_signators          2   638.10 702.10
## - number_loans             1   636.10 702.10
## - other_credit_outstanding  2   638.36 702.36
## - age_yrs                  1   637.36 703.36
## - payment_pcnt_income      1   638.01 704.01
## - gender_status            3   644.57 706.57
## - loan_duration_mo         1   641.08 707.08
## - loan_amount              1   642.46 708.46
## - savings_account_balance   4   649.96 709.96
## - purpose                  9   667.28 717.28
## - credit_history           4   657.82 717.82
## - checking_account_status   3   675.72 737.72
```

```
mod2$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + time_employed_yrs +
##     payment_pcnt_income + gender_status + other_signators + time_in_residence +
##     property + age_yrs + other_credit_outstanding + home_ownership +
##     number_loans + job_category + dependents + telephone + foreign_worker
##
## Final Model:
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + age_yrs + other_credit_outstanding +
```

```
##     number_loans + foreign_worker
##
##
##                 Step Df  Deviance Resid. Df Resid. Dev      AIC
## 1                                       651   617.9466 715.9466
## 2      - job_category  3 0.5670417       654   618.5136 710.5136
## 3 - time_employed_yrs  4 5.3126512       658   623.8262 707.8262
## 4 - time_in_residence  1 0.1773350       659   624.0036 706.0036
## 5          - property  3 4.5481008       662   628.5517 704.5517
## 6         - telephone  1 0.6452700       663   629.1970 703.1970
## 7    - home_ownership  2 2.7182470       665   631.9152 701.9152
## 8        - dependents  1 0.8712115       666   632.7864 700.7864
```

```r
summary(mod2)
```

```
##
## Call:
## glm(formula = bad_credit ~ checking_account_status + loan_duration_mo +
##     credit_history + purpose + loan_amount + savings_account_balance +
##     payment_pcnt_income + gender_status + other_signators + age_yrs +
##     other_credit_outstanding + number_loans + foreign_worker,
##     family = "binomial", data = credit_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0234  -0.7093  -0.3930   0.7061   2.6261
##
## Coefficients:
##                                                        Estimate
## (Intercept)                                          -2.026e-01
## checking_account_status> 200 DM or salary assignment  -6.276e-01
## checking_account_status0 - 200 DM                    -3.814e-01
## checking_account_statusnone                          -1.681e+00
## loan_duration_mo                                      3.070e-02
## credit_historycritical account - other non-bank loans -2.058e+00
## credit_historycurrent loans paid                     -1.210e+00
## credit_historyno credit - paid                       -3.798e-01
## credit_historypast payment delays                    -2.009e+00
## purposecar (new)                                      8.086e-01
## purposecar (used)                                    -1.205e+00
## purposedomestic appliances                           -1.106e-01
## purposeeducation                                      1.019e+00
## purposefurniture/equipment                            1.433e-01
## purposeother                                         -1.307e+00
## purposeradio/television                              -1.480e-01
## purposerepairs                                        4.380e-01
## purposeretraining                                     7.217e-02
## loan_amount                                           1.457e-04
## savings_account_balance>= 1000 DM                    -1.577e+00
## savings_account_balance100 - 500 DM                  -2.200e-01
## savings_account_balance500 - 1000 DM                 -4.978e-01
## savings_account_balanceunknown/none                  -9.620e-01
## payment_pcnt_income                                   2.289e-01
## gender_statusmale-divorced/separated                  3.841e-01
## gender_statusmale-married/widowed                    -2.101e-01
```

```
## gender_statusmale-single                              -6.575e-01
## other_signatorsguarantor                              -1.333e+00
## other_signatorsnone                                   -4.622e-01
## age_yrs                                               -2.020e-02
## other_credit_outstandingnone                          -5.125e-01
## other_credit_outstandingstores                         2.682e-01
## number_loans                                           3.939e-01
## foreign_workeryes                                      1.072e+00
##                                                       Std. Error z value
## (Intercept)                                            1.052e+00  -0.193
## checking_account_status> 200 DM or salary assignment   4.293e-01  -1.462
## checking_account_status0 - 200 DM                      2.537e-01  -1.503
## checking_account_statusnone                            2.795e-01  -6.014
## loan_duration_mo                                       1.073e-02   2.861
## credit_historycritical account - other non-bank loans  5.318e-01  -3.869
## credit_historycurrent loans paid                       4.704e-01  -2.572
## credit_historyno credit - paid                         6.700e-01  -0.567
## credit_historypast payment delays                      5.744e-01  -3.497
## purposecar (new)                                       3.887e-01   2.081
## purposecar (used)                                      5.107e-01  -2.359
## purposedomestic appliances                             1.086e+00  -0.102
## purposeeducation                                       5.298e-01   1.924
## purposefurniture/equipment                             4.079e-01   0.351
## purposeother                                           9.805e-01  -1.333
## purposeradio/television                                3.951e-01  -0.375
## purposerepairs                                         7.142e-01   0.613
## purposeretraining                                      1.289e+00   0.056
## loan_amount                                            4.757e-05   3.062
## savings_account_balance>= 1000 DM                      6.133e-01  -2.571
## savings_account_balance100 - 500 DM                    3.406e-01  -0.646
## savings_account_balance500 - 1000 DM                   5.124e-01  -0.972
## savings_account_balanceunknown/none                    3.107e-01  -3.096
## payment_pcnt_income                                    1.011e-01   2.264
## gender_statusmale-divorced/separated                   4.460e-01   0.861
## gender_statusmale-married/widowed                      3.693e-01  -0.569
## gender_statusmale-single                               2.334e-01  -2.817
## other_signatorsguarantor                               6.292e-01  -2.118
## other_signatorsnone                                    4.738e-01  -0.976
## age_yrs                                                9.588e-03  -2.106
## other_credit_outstandingnone                           2.865e-01  -1.789
## other_credit_outstandingstores                         4.672e-01   0.574
## number_loans                                           2.165e-01   1.819
## foreign_workeryes                                      6.590e-01   1.627
##                                                       Pr(>|z|)
## (Intercept)                                           0.847284
## checking_account_status> 200 DM or salary assignment  0.143723
## checking_account_status0 - 200 DM                     0.132819
## checking_account_statusnone                           1.81e-09 ***
## loan_duration_mo                                      0.004221 **
## credit_historycritical account - other non-bank loans 0.000109 ***
## credit_historycurrent loans paid                      0.010111 *
## credit_historyno credit - paid                        0.570789
## credit_historypast payment delays                     0.000470 ***
## purposecar (new)                                      0.037467 *
```

```
## purposecar (used)                                  0.018314 *
## purposedomestic appliances                         0.918928
## purposeeducation                                   0.054407 .
## purposefurniture/equipment                         0.725435
## purposeother                                       0.182636
## purposeradio/television                            0.708005
## purposerepairs                                     0.539672
## purposeretraining                                  0.955370
## loan_amount                                        0.002198 **
## savings_account_balance>= 1000 DM                  0.010154 *
## savings_account_balance100 - 500 DM                0.518333
## savings_account_balance500 - 1000 DM               0.331237
## savings_account_balanceunknown/none                0.001962 **
## payment_pcnt_income                                0.023573 *
## gender_statusmale-divorced/separated               0.389208
## gender_statusmale-married/widowed                  0.569469
## gender_statusmale-single                           0.004843 **
## other_signatorsguarantor                           0.034149 *
## other_signatorsnone                                0.329268
## age_yrs                                            0.035176 *
## other_credit_outstandingnone                       0.073648 .
## other_credit_outstandingstores                     0.566003
## number_loans                                       0.068857 .
## foreign_workeryes                                  0.103712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 855.21  on 699  degrees of freedom
## Residual deviance: 632.79  on 666  degrees of freedom
## AIC: 700.79
##
## Number of Fisher Scoring iterations: 5
```

Let's analyse the model that we got

```r
unique(credit_train$payment_pcnt_income)
```

```
## [1] 4 2 3 1
```

```r
colnames(credit_train)
```

```
##  [1] "checking_account_status"  "loan_duration_mo"
##  [3] "credit_history"           "purpose"
##  [5] "loan_amount"              "savings_account_balance"
##  [7] "time_employed_yrs"        "payment_pcnt_income"
##  [9] "gender_status"            "other_signators"
## [11] "time_in_residence"        "property"
## [13] "age_yrs"                  "other_credit_outstanding"
## [15] "home_ownership"           "number_loans"
## [17] "job_category"             "dependents"
## [19] "telephone"                "foreign_worker"
## [21] "bad_credit"
```

```r
mod3<-glm(bad_credit~checking_account_status+loan_duration_mo+credit_history+purpose+loan_amount+
            other_signators+number_loans+gender_status,data=credit_train,family='binomial')

summary(mod3)
```

```
##
## Call:
## glm(formula = bad_credit ~ checking_account_status + loan_duration_mo +
##     credit_history + purpose + loan_amount + other_signators +
##     number_loans + gender_status, family = "binomial", data = credit_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0914  -0.7471  -0.4375   0.7530   2.5430
##
## Coefficients:
##                                                       Estimate
## (Intercept)                                          5.223e-01
## checking_account_status> 200 DM or salary assignment  -7.822e-01
## checking_account_status0 - 200 DM                    -5.395e-01
## checking_account_statusnone                          -1.843e+00
## loan_duration_mo                                     3.914e-02
## credit_historycritical account - other non-bank loans -2.149e+00
## credit_historycurrent loans paid                     -1.370e+00
## credit_historyno credit - paid                       -3.721e-01
## credit_historypast payment delays                    -2.019e+00
## purposecar (new)                                     5.818e-01
## purposecar (used)                                    -1.354e+00
## purposedomestic appliances                           -4.556e-01
## purposeeducation                                     9.598e-01
## purposefurniture/equipment                           8.094e-02
## purposeother                                         -1.199e+00
## purposeradio/television                              -1.683e-01
## purposerepairs                                       2.445e-01
## purposeretraining                                    -9.501e-02
## loan_amount                                          9.368e-05
## other_signatorsguarantor                             -1.258e+00
## other_signatorsnone                                  -4.559e-01
## number_loans                                         3.564e-01
## gender_statusmale-divorced/separated                 1.666e-01
## gender_statusmale-married/widowed                    -1.459e-01
## gender_statusmale-single                             -6.257e-01
##                                                      Std. Error z value
## (Intercept)                                           7.776e-01   0.672
## checking_account_status> 200 DM or salary assignment   4.082e-01  -1.916
## checking_account_status0 - 200 DM                     2.397e-01  -2.251
## checking_account_statusnone                           2.630e-01  -7.007
## loan_duration_mo                                      1.007e-02   3.886
## credit_historycritical account - other non-bank loans  4.845e-01  -4.435
## credit_historycurrent loans paid                      4.271e-01  -3.208
## credit_historyno credit - paid                        6.404e-01  -0.581
## credit_historypast payment delays                     5.432e-01  -3.716
## purposecar (new)                                      3.735e-01   1.557
## purposecar (used)                                     4.886e-01  -2.772
```

```
## purposedomestic appliances                           1.091e+00  -0.418
## purposeeducation                                     5.080e-01   1.889
## purposefurniture/equipment                           3.952e-01   0.205
## purposeother                                         9.580e-01  -1.251
## purposeradio/television                              3.811e-01  -0.442
## purposerepairs                                       6.888e-01   0.355
## purposeretraining                                    1.196e+00  -0.079
## loan_amount                                          4.329e-05   2.164
## other_signatorsguarantor                             6.202e-01  -2.028
## other_signatorsnone                                  4.651e-01  -0.980
## number_loans                                         2.061e-01   1.729
## gender_statusmale-divorced/separated                 4.236e-01   0.393
## gender_statusmale-married/widowed                    3.559e-01  -0.410
## gender_statusmale-single                             2.210e-01  -2.831
##                                                     Pr(>|z|)
## (Intercept)                                         0.501768
## checking_account_status> 200 DM or salary assignment 0.055364 .
## checking_account_status0 - 200 DM                   0.024415 *
## checking_account_statusnone                         2.44e-12 ***
## loan_duration_mo                                    0.000102 ***
## credit_historycritical account - other non-bank loans 9.19e-06 ***
## credit_historycurrent loans paid                    0.001334 **
## credit_historyno credit - paid                      0.561171
## credit_historypast payment delays                   0.000202 ***
## purposecar (new)                                    0.119355
## purposecar (used)                                   0.005575 **
## purposedomestic appliances                          0.676178
## purposeeducation                                    0.058868 .
## purposefurniture/equipment                          0.837733
## purposeother                                        0.210916
## purposeradio/television                             0.658832
## purposerepairs                                      0.722643
## purposeretraining                                   0.936675
## loan_amount                                         0.030460 *
## other_signatorsguarantor                            0.042549 *
## other_signatorsnone                                 0.327004
## number_loans                                        0.083721 .
## gender_statusmale-divorced/separated                0.694113
## gender_statusmale-married/widowed                   0.681870
## gender_statusmale-single                            0.004645 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 855.21  on 699  degrees of freedom
## Residual deviance: 668.79  on 675  degrees of freedom
## AIC: 718.79
##
## Number of Fisher Scoring iterations: 5
```

```r
train_pred<-predict(mod3,newdata = credit_train,type='response')

train_pred<-ifelse(train_pred>0.5,1,0)
```

```r
table(train_pred,credit_train$bad_credit)
```

```
##
## train_pred  No Yes
##          0 445 111
##          1  45  99
```

```r
table(train_pred)
```

```
## train_pred
##   0   1
## 556 144
```

```r
train_pred<-factor(train_pred, levels = c(0, 1), labels = c('No','Yes'))

levels(train_pred)
```

```
## [1] "No"  "Yes"
```

```r
as.numeric(train_pred)
```

```
##   [1] 1 2 1 2 1 1 2 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1
##  [36] 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1
##  [71] 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 2 2 1 1 1 2 1 1 1
## [106] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1
## [141] 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 1
## [176] 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
## [211] 1 1 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1
## [246] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 2 1 1 2 1 1 1 1 1 2 1 1 2 2 2 1 1 1 1 1
## [281] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1 2 2 1 1 1 1
## [316] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1
## [351] 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2
## [386] 1 2 2 2 1 1 1 1 2 1 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2
## [421] 2 1 1 1 2 1 2 1 1 1 2 1 1 2 1 2 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 2 2 1 1
## [456] 1 1 1 2 1 2 2 1 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1
## [491] 1 1 1 1 1 1 1 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 2 1 2 2 2
## [526] 1 2 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1
## [561] 1 1 1 1 2 2 1 1 2 2 1 1 1 2 2 2 1 1 2 1 1 2 1 2 2 1 1 1 2 1 1 1 1 1 1
## [596] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 1 2 1 1 1 1 2 1 1 1
## [631] 1 1 2 2 1 1 1 1 1 1 2 2 1 2 1 1 1 1 1 2 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1
## [666] 2 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1
```

```r
str(train_pred)
```

```
##  Factor w/ 2 levels "No","Yes": 1 2 1 2 1 1 2 1 2 1 ...
##  - attr(*, "names")= chr [1:700] "1" "2" "3" "4" ...
```

```r
accuracy(actual = credit_train$bad_credit,predicted = train_pred)
```

```
## [1] 0.7771429
```

```r
confusionMatrix(train_pred,credit_train$bad_credit,positive = 'Yes')
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No Yes
##        No  445 111
```

```
##        Yes  45  99
##
##                Accuracy : 0.7771
##                  95% CI : (0.7445, 0.8075)
##     No Information Rate : 0.7
##     P-Value [Acc > NIR] : 2.912e-06
##
##                   Kappa : 0.417
##  Mcnemar's Test P-Value : 1.949e-07
##
##             Sensitivity : 0.4714
##             Specificity : 0.9082
##          Pos Pred Value : 0.6875
##          Neg Pred Value : 0.8004
##              Prevalence : 0.3000
##          Detection Rate : 0.1414
##    Detection Prevalence : 0.2057
##       Balanced Accuracy : 0.6898
##
##        'Positive' Class : Yes
##
```

As we know our priority should be the correct classification of bad credits, which have high risks of defaulting. The measure of bad credits with Yes status is Sensitivity and we should improve it from the current result of 52%

```r
credit_train<-credit1[ind,]
credit_test<-credit1[-ind,]

colnames(credit_train)
```

```
##  [1] "checking_account_status"  "loan_duration_mo"
##  [3] "credit_history"           "purpose"
##  [5] "loan_amount"              "savings_account_balance"
##  [7] "time_employed_yrs"        "payment_pcnt_income"
##  [9] "gender_status"            "other_signators"
## [11] "time_in_residence"        "property"
## [13] "age_yrs"                  "other_credit_outstanding"
## [15] "home_ownership"           "number_loans"
## [17] "job_category"             "dependents"
## [19] "telephone"                "foreign_worker"
## [21] "bad_credit"
```

```r
mod1<-glm(bad_credit~.,data=credit_train,family = 'binomial')

mod2<-stepAIC(mod1,method='both')
```

```
## Start:  AIC=715.95
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + time_employed_yrs +
##     payment_pcnt_income + gender_status + other_signators + time_in_residence +
##     property + age_yrs + other_credit_outstanding + home_ownership +
##     number_loans + job_category + dependents + telephone + foreign_worker
##
##                            Df Deviance    AIC
## - job_category              3   618.51 710.51
```

```
## - time_employed_yrs         4   623.31 713.31
## - time_in_residence         1   618.16 714.16
## - telephone                 1   618.35 714.35
## - property                  3   622.38 714.38
## - dependents                1   618.78 714.78
## - other_signators           2   621.59 715.59
## - home_ownership            2   621.94 715.94
## <none>                          617.95 715.95
## - age_yrs                   1   621.10 717.10
## - other_credit_outstanding  2   623.17 717.17
## - gender_status             3   625.42 717.42
## - payment_pcnt_income       1   621.74 717.74
## - foreign_worker            1   621.86 717.86
## - number_loans              1   622.53 718.53
## - loan_amount               1   625.44 721.44
## - loan_duration_mo          1   625.84 721.84
## - savings_account_balance   4   633.29 723.29
## - purpose                   9   651.08 731.08
## - credit_history            4   641.17 731.17
## - checking_account_status   3   655.62 747.62
##
## Step:  AIC=710.51
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + time_employed_yrs +
##     payment_pcnt_income + gender_status + other_signators + time_in_residence +
##     property + age_yrs + other_credit_outstanding + home_ownership +
##     number_loans + dependents + telephone + foreign_worker
##
##                            Df Deviance    AIC
## - time_employed_yrs         4   623.83 707.83
## - time_in_residence         1   618.64 708.64
## - property                  3   622.78 708.78
## - telephone                 1   619.21 709.21
## - dependents                1   619.30 709.30
## - other_signators           2   622.06 710.06
## - home_ownership            2   622.34 710.34
## <none>                          618.51 710.51
## - age_yrs                   1   621.37 711.37
## - gender_status             3   625.73 711.73
## - other_credit_outstanding  2   624.05 712.05
## - foreign_worker            1   622.27 712.27
## - payment_pcnt_income       1   622.42 712.42
## - number_loans              1   623.00 713.00
## - loan_duration_mo          1   626.32 716.32
## - loan_amount               1   626.35 716.35
## - savings_account_balance   4   634.34 718.34
## - credit_history            4   641.45 725.45
## - purpose                   9   652.30 726.30
## - checking_account_status   3   656.75 742.75
##
## Step:  AIC=707.83
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + time_in_residence + property +
```

```
##     age_yrs + other_credit_outstanding + home_ownership + number_loans +
##     dependents + telephone + foreign_worker
##
##                            Df Deviance    AIC
## - time_in_residence         1   624.00 706.00
## - property                  3   628.41 706.41
## - dependents                1   624.50 706.50
## - telephone                 1   624.88 706.88
## - home_ownership            2   627.42 707.42
## - other_signators           2   627.72 707.72
## <none>                          623.83 707.83
## - age_yrs                   1   626.49 708.49
## - foreign_worker            1   627.40 709.40
## - other_credit_outstanding  2   629.79 709.79
## - payment_pcnt_income       1   628.01 710.01
## - number_loans              1   628.07 710.07
## - gender_status             3   632.23 710.23
## - loan_duration_mo          1   630.64 712.64
## - loan_amount               1   631.75 713.75
## - savings_account_balance   4   640.28 716.28
## - purpose                   9   656.17 722.17
## - credit_history            4   647.06 723.06
## - checking_account_status   3   661.97 739.97
##
## Step:  AIC=706
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + property + age_yrs + other_credit_outstanding +
##     home_ownership + number_loans + dependents + telephone +
##     foreign_worker
##
##                            Df Deviance    AIC
## - property                  3   628.55 704.55
## - dependents                1   624.69 704.69
## - telephone                 1   625.14 705.14
## - home_ownership            2   627.44 705.44
## - other_signators           2   627.86 705.86
## <none>                          624.00 706.00
## - age_yrs                   1   627.17 707.17
## - foreign_worker            1   627.54 707.54
## - other_credit_outstanding  2   630.10 708.10
## - number_loans              1   628.17 708.17
## - payment_pcnt_income       1   628.21 708.21
## - gender_status             3   632.40 708.40
## - loan_duration_mo          1   630.76 710.76
## - loan_amount               1   632.09 712.09
## - savings_account_balance   4   640.77 714.77
## - purpose                   9   656.63 720.63
## - credit_history            4   647.29 721.29
## - checking_account_status   3   661.98 737.98
##
## Step:  AIC=704.55
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
```

```
##     gender_status + other_signators + age_yrs + other_credit_outstanding +
##     home_ownership + number_loans + dependents + telephone +
##     foreign_worker
##
##                            Df Deviance    AIC
## - telephone                 1   629.20 703.20
## - home_ownership            2   631.25 703.25
## - dependents                1   629.42 703.42
## <none>                          628.55 704.55
## - other_signators           2   633.23 705.23
## - age_yrs                   1   631.90 705.90
## - foreign_worker            1   631.98 705.98
## - number_loans              1   632.19 706.19
## - gender_status             3   636.83 706.83
## - other_credit_outstanding  2   635.12 707.12
## - payment_pcnt_income       1   633.55 707.55
## - loan_duration_mo          1   636.56 710.56
## - loan_amount               1   638.46 712.46
## - savings_account_balance   4   645.20 713.20
## - credit_history            4   652.34 720.34
## - purpose                   9   663.11 721.11
## - checking_account_status   3   668.16 738.16
##
## Step:  AIC=703.2
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + age_yrs + other_credit_outstanding +
##     home_ownership + number_loans + dependents + foreign_worker
##
##                            Df Deviance    AIC
## - home_ownership            2   631.92 701.92
## - dependents                1   630.08 702.08
## <none>                          629.20 703.20
## - other_signators           2   633.72 703.72
## - foreign_worker            1   632.39 704.39
## - number_loans              1   632.64 704.64
## - age_yrs                   1   633.00 705.00
## - gender_status             3   637.53 705.53
## - other_credit_outstanding  2   635.67 705.67
## - payment_pcnt_income       1   633.83 705.83
## - loan_duration_mo          1   637.59 709.59
## - loan_amount               1   638.47 710.47
## - savings_account_balance   4   646.06 712.06
## - credit_history            4   652.81 718.81
## - purpose                   9   664.29 720.29
## - checking_account_status   3   669.46 737.46
##
## Step:  AIC=701.92
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + age_yrs + other_credit_outstanding +
##     number_loans + dependents + foreign_worker
##
##                            Df Deviance    AIC
```

```
## - dependents               1   632.79 700.79
## <none>                         631.92 701.92
## - other_signators          2   636.68 702.68
## - foreign_worker           1   635.03 703.03
## - number_loans             1   635.50 703.50
## - other_credit_outstanding  2   637.65 703.65
## - age_yrs                  1   636.20 704.20
## - payment_pcnt_income      1   636.34 704.34
## - gender_status            3   641.31 705.31
## - loan_duration_mo         1   640.36 708.36
## - loan_amount              1   641.26 709.26
## - savings_account_balance  4   648.47 710.47
## - credit_history           4   656.69 718.69
## - purpose                  9   666.90 718.90
## - checking_account_status  3   674.96 738.96
##
## Step:  AIC=700.79
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + age_yrs + other_credit_outstanding +
##     number_loans + foreign_worker
##
##                           Df Deviance    AIC
## <none>                         632.79 700.79
## - foreign_worker           1   635.85 701.85
## - other_signators          2   638.10 702.10
## - number_loans             1   636.10 702.10
## - other_credit_outstanding  2   638.36 702.36
## - age_yrs                  1   637.36 703.36
## - payment_pcnt_income      1   638.01 704.01
## - gender_status            3   644.57 706.57
## - loan_duration_mo         1   641.08 707.08
## - loan_amount              1   642.46 708.46
## - savings_account_balance  4   649.96 709.96
## - purpose                  9   667.28 717.28
## - credit_history           4   657.82 717.82
## - checking_account_status  3   675.72 737.72
```

```r
mod2$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + time_employed_yrs +
##     payment_pcnt_income + gender_status + other_signators + time_in_residence +
##     property + age_yrs + other_credit_outstanding + home_ownership +
##     number_loans + job_category + dependents + telephone + foreign_worker
##
## Final Model:
## bad_credit ~ checking_account_status + loan_duration_mo + credit_history +
##     purpose + loan_amount + savings_account_balance + payment_pcnt_income +
##     gender_status + other_signators + age_yrs + other_credit_outstanding +
##     number_loans + foreign_worker
```

```
##
##
##                  Step Df   Deviance Resid. Df Resid. Dev      AIC
## 1                                        651   617.9466 715.9466
## 2      - job_category  3 0.5670417       654   618.5136 710.5136
## 3 - time_employed_yrs  4 5.3126512       658   623.8262 707.8262
## 4 - time_in_residence  1 0.1773350       659   624.0036 706.0036
## 5          - property  3 4.5481008       662   628.5517 704.5517
## 6         - telephone  1 0.6452700       663   629.1970 703.1970
## 7    - home_ownership  2 2.7182470       665   631.9152 701.9152
## 8        - dependents  1 0.8712115       666   632.7864 700.7864
```

```r
summary(mod2)
```

```
##
## Call:
## glm(formula = bad_credit ~ checking_account_status + loan_duration_mo +
##     credit_history + purpose + loan_amount + savings_account_balance +
##     payment_pcnt_income + gender_status + other_signators + age_yrs +
##     other_credit_outstanding + number_loans + foreign_worker,
##     family = "binomial", data = credit_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0234  -0.7093  -0.3930   0.7061   2.6261
##
## Coefficients:
##                                                        Estimate
## (Intercept)                                          -2.026e-01
## checking_account_status> 200 DM or salary assignment  -6.276e-01
## checking_account_status0 - 200 DM                    -3.814e-01
## checking_account_statusnone                          -1.681e+00
## loan_duration_mo                                      3.070e-02
## credit_historycritical account - other non-bank loans -2.058e+00
## credit_historycurrent loans paid                     -1.210e+00
## credit_historyno credit - paid                       -3.798e-01
## credit_historypast payment delays                    -2.009e+00
## purposecar (new)                                      8.086e-01
## purposecar (used)                                    -1.205e+00
## purposedomestic appliances                           -1.106e-01
## purposeeducation                                      1.019e+00
## purposefurniture/equipment                            1.433e-01
## purposeother                                         -1.307e+00
## purposeradio/television                              -1.480e-01
## purposerepairs                                        4.380e-01
## purposeretraining                                     7.217e-02
## loan_amount                                           1.457e-04
## savings_account_balance>= 1000 DM                    -1.577e+00
## savings_account_balance100 - 500 DM                  -2.200e-01
## savings_account_balance500 - 1000 DM                 -4.978e-01
## savings_account_balanceunknown/none                  -9.620e-01
## payment_pcnt_income                                   2.289e-01
## gender_statusmale-divorced/separated                  3.841e-01
## gender_statusmale-married/widowed                    -2.101e-01
## gender_statusmale-single                             -6.575e-01
```

```
## other_signatorsguarantor                             -1.333e+00
## other_signatorsnone                                  -4.622e-01
## age_yrs                                              -2.020e-02
## other_credit_outstandingnone                         -5.125e-01
## other_credit_outstandingstores                        2.682e-01
## number_loans                                          3.939e-01
## foreign_workeryes                                     1.072e+00
##                                                      Std. Error z value
## (Intercept)                                           1.052e+00  -0.193
## checking_account_status> 200 DM or salary assignment  4.293e-01  -1.462
## checking_account_status0 - 200 DM                     2.537e-01  -1.503
## checking_account_statusnone                           2.795e-01  -6.014
## loan_duration_mo                                      1.073e-02   2.861
## credit_historycritical account - other non-bank loans  5.318e-01  -3.869
## credit_historycurrent loans paid                      4.704e-01  -2.572
## credit_historyno credit - paid                        6.700e-01  -0.567
## credit_historypast payment delays                     5.744e-01  -3.497
## purposecar (new)                                      3.887e-01   2.081
## purposecar (used)                                     5.107e-01  -2.359
## purposedomestic appliances                            1.086e+00  -0.102
## purposeeducation                                      5.298e-01   1.924
## purposefurniture/equipment                            4.079e-01   0.351
## purposeother                                          9.805e-01  -1.333
## purposeradio/television                               3.951e-01  -0.375
## purposerepairs                                        7.142e-01   0.613
## purposeretraining                                     1.289e+00   0.056
## loan_amount                                           4.757e-05   3.062
## savings_account_balance>= 1000 DM                     6.133e-01  -2.571
## savings_account_balance100 - 500 DM                   3.406e-01  -0.646
## savings_account_balance500 - 1000 DM                  5.124e-01  -0.972
## savings_account_balanceunknown/none                   3.107e-01  -3.096
## payment_pcnt_income                                   1.011e-01   2.264
## gender_statusmale-divorced/separated                  4.460e-01   0.861
## gender_statusmale-married/widowed                     3.693e-01  -0.569
## gender_statusmale-single                              2.334e-01  -2.817
## other_signatorsguarantor                              6.292e-01  -2.118
## other_signatorsnone                                   4.738e-01  -0.976
## age_yrs                                               9.588e-03  -2.106
## other_credit_outstandingnone                          2.865e-01  -1.789
## other_credit_outstandingstores                        4.672e-01   0.574
## number_loans                                          2.165e-01   1.819
## foreign_workeryes                                     6.590e-01   1.627
##                                                      Pr(>|z|)
## (Intercept)                                          0.847284
## checking_account_status> 200 DM or salary assignment 0.143723
## checking_account_status0 - 200 DM                    0.132819
## checking_account_statusnone                          1.81e-09 ***
## loan_duration_mo                                     0.004221 **
## credit_historycritical account - other non-bank loans 0.000109 ***
## credit_historycurrent loans paid                     0.010111 *
## credit_historyno credit - paid                       0.570789
## credit_historypast payment delays                    0.000470 ***
## purposecar (new)                                     0.037467 *
## purposecar (used)                                    0.018314 *
```

```
## purposedomestic appliances                              0.918928
## purposeeducation                                        0.054407 .
## purposefurniture/equipment                              0.725435
## purposeother                                            0.182636
## purposeradio/television                                 0.708005
## purposerepairs                                          0.539672
## purposeretraining                                       0.955370
## loan_amount                                             0.002198 **
## savings_account_balance>= 1000 DM                       0.010154 *
## savings_account_balance100 - 500 DM                     0.518333
## savings_account_balance500 - 1000 DM                    0.331237
## savings_account_balanceunknown/none                     0.001962 **
## payment_pcnt_income                                     0.023573 *
## gender_statusmale-divorced/separated                    0.389208
## gender_statusmale-married/widowed                       0.569469
## gender_statusmale-single                                0.004843 **
## other_signatorsguarantor                                0.034149 *
## other_signatorsnone                                     0.329268
## age_yrs                                                 0.035176 *
## other_credit_outstandingnone                            0.073648 .
## other_credit_outstandingstores                          0.566003
## number_loans                                            0.068857 .
## foreign_workeryes                                       0.103712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 855.21  on 699  degrees of freedom
## Residual deviance: 632.79  on 666  degrees of freedom
## AIC: 700.79
##
## Number of Fisher Scoring iterations: 5
```

Let's analyse the model that we got

```
unique(credit_train$payment_pcnt_income)
```

```
## [1] 4 2 3 1
```

```
colnames(credit_train)
```

```
##  [1] "checking_account_status"  "loan_duration_mo"
##  [3] "credit_history"           "purpose"
##  [5] "loan_amount"              "savings_account_balance"
##  [7] "time_employed_yrs"        "payment_pcnt_income"
##  [9] "gender_status"            "other_signators"
## [11] "time_in_residence"        "property"
## [13] "age_yrs"                  "other_credit_outstanding"
## [15] "home_ownership"           "number_loans"
## [17] "job_category"             "dependents"
## [19] "telephone"                "foreign_worker"
## [21] "bad_credit"
```

```
mod3<-glm(bad_credit~checking_account_status+loan_duration_mo+credit_history+purpose+loan_amount+
            other_signators+number_loans+gender_status,data=credit_train,family='binomial')
```

```r
summary(mod3)
```

```
##
## Call:
## glm(formula = bad_credit ~ checking_account_status + loan_duration_mo +
##     credit_history + purpose + loan_amount + other_signators +
##     number_loans + gender_status, family = "binomial", data = credit_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0914  -0.7471  -0.4375   0.7530   2.5430
##
## Coefficients:
##                                                       Estimate
## (Intercept)                                          5.223e-01
## checking_account_status> 200 DM or salary assignment  -7.822e-01
## checking_account_status0 - 200 DM                     -5.395e-01
## checking_account_statusnone                           -1.843e+00
## loan_duration_mo                                       3.914e-02
## credit_historycritical account - other non-bank loans -2.149e+00
## credit_historycurrent loans paid                      -1.370e+00
## credit_historyno credit - paid                        -3.721e-01
## credit_historypast payment delays                     -2.019e+00
## purposecar (new)                                       5.818e-01
## purposecar (used)                                     -1.354e+00
## purposedomestic appliances                            -4.556e-01
## purposeeducation                                       9.598e-01
## purposefurniture/equipment                             8.094e-02
## purposeother                                          -1.199e+00
## purposeradio/television                               -1.683e-01
## purposerepairs                                         2.445e-01
## purposeretraining                                     -9.501e-02
## loan_amount                                            9.368e-05
## other_signatorsguarantor                              -1.258e+00
## other_signatorsnone                                   -4.559e-01
## number_loans                                           3.564e-01
## gender_statusmale-divorced/separated                   1.666e-01
## gender_statusmale-married/widowed                     -1.459e-01
## gender_statusmale-single                              -6.257e-01
##                                                       Std. Error z value
## (Intercept)                                            7.776e-01   0.672
## checking_account_status> 200 DM or salary assignment   4.082e-01  -1.916
## checking_account_status0 - 200 DM                      2.397e-01  -2.251
## checking_account_statusnone                            2.630e-01  -7.007
## loan_duration_mo                                       1.007e-02   3.886
## credit_historycritical account - other non-bank loans  4.845e-01  -4.435
## credit_historycurrent loans paid                       4.271e-01  -3.208
## credit_historyno credit - paid                         6.404e-01  -0.581
## credit_historypast payment delays                      5.432e-01  -3.716
## purposecar (new)                                       3.735e-01   1.557
## purposecar (used)                                      4.886e-01  -2.772
## purposedomestic appliances                             1.091e+00  -0.418
## purposeeducation                                       5.080e-01   1.889
```

```
## purposefurniture/equipment                              3.952e-01   0.205
## purposeother                                            9.580e-01  -1.251
## purposeradio/television                                 3.811e-01  -0.442
## purposerepairs                                          6.888e-01   0.355
## purposeretraining                                       1.196e+00  -0.079
## loan_amount                                             4.329e-05   2.164
## other_signatorsguarantor                                6.202e-01  -2.028
## other_signatorsnone                                     4.651e-01  -0.980
## number_loans                                            2.061e-01   1.729
## gender_statusmale-divorced/separated                    4.236e-01   0.393
## gender_statusmale-married/widowed                       3.559e-01  -0.410
## gender_statusmale-single                                2.210e-01  -2.831
##                                                        Pr(>|z|)
## (Intercept)                                            0.501768
## checking_account_status> 200 DM or salary assignment   0.055364 .
## checking_account_status0 - 200 DM                      0.024415 *
## checking_account_statusnone                            2.44e-12 ***
## loan_duration_mo                                       0.000102 ***
## credit_historycritical account - other non-bank loans 9.19e-06 ***
## credit_historycurrent loans paid                       0.001334 **
## credit_historyno credit - paid                         0.561171
## credit_historypast payment delays                      0.000202 ***
## purposecar (new)                                       0.119355
## purposecar (used)                                      0.005575 **
## purposedomestic appliances                             0.676178
## purposeeducation                                       0.058868 .
## purposefurniture/equipment                             0.837733
## purposeother                                           0.210916
## purposeradio/television                                0.658832
## purposerepairs                                         0.722643
## purposeretraining                                      0.936675
## loan_amount                                            0.030460 *
## other_signatorsguarantor                               0.042549 *
## other_signatorsnone                                    0.327004
## number_loans                                           0.083721 .
## gender_statusmale-divorced/separated                   0.694113
## gender_statusmale-married/widowed                      0.681870
## gender_statusmale-single                               0.004645 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 855.21  on 699  degrees of freedom
## Residual deviance: 668.79  on 675  degrees of freedom
## AIC: 718.79
##
## Number of Fisher Scoring iterations: 5
```

```r
train_pred<-predict(mod3,newdata = credit_train,type='response')

library(ROCR)
```

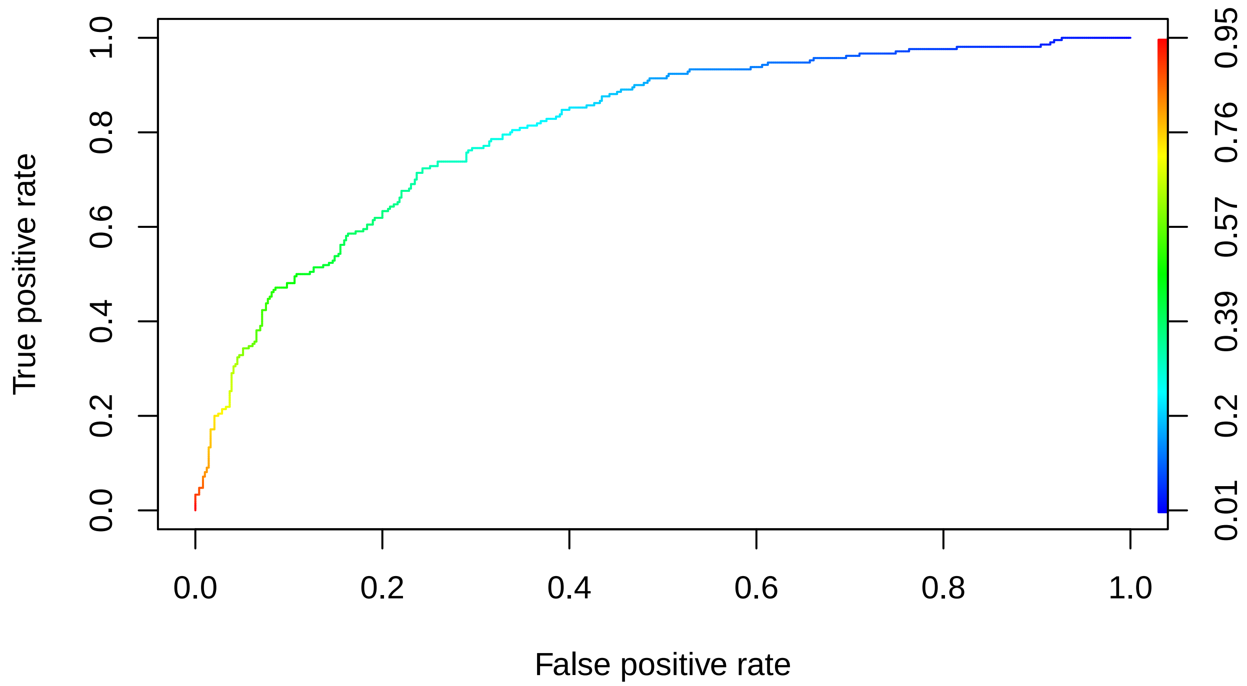
```
## Loading required package: gplots
```

```
##
## Attaching package: 'gplots'

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

```r
P_train<-prediction(train_pred,credit_train$bad_credit)
perf<-performance(P_train,'tpr','fpr')

plot(perf,colorize=T)
```




```r
performance(P_train,'auc')@y.values
```

```
## [[1]]
## [1] 0.8074538
```

```r
train_pred<-ifelse(train_pred>0.25,1,0)

table(train_pred,credit_train$bad_credit)
```

```
##
## train_pred  No Yes
##          0 325  42
##          1 165 168
```

```r
table(train_pred)
```

```
## train_pred
```

```
##   0   1
## 367 333
```

```r
train_pred<-factor(train_pred, levels = c(0, 1), labels = c('No','Yes'))

levels(train_pred)
```

```
## [1] "No"  "Yes"
```

```r
as.numeric(train_pred)
```

```
##   [1] 1 2 1 2 2 1 2 2 2 2 2 2 2 1 2 2 1 1 2 1 1 1 1 2 1 2 2 2 1 1 1 1 1 2 1
##  [36] 1 1 1 1 1 2 1 1 2 2 2 1 2 2 1 2 1 2 2 1 2 1 1 2 2 1 1 1 1 2 2 2 2 1 1
##  [71] 2 2 1 1 1 1 2 1 1 2 2 2 1 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 1
## [106] 2 1 1 2 1 1 2 1 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 2 1 1 1 2 2 2 1 2 1 2 1
## [141] 1 2 2 1 2 1 2 1 1 2 1 1 2 2 1 1 2 1 1 1 2 2 1 2 2 1 1 2 2 2 1 1 1 1 2
## [176] 2 1 1 1 2 1 1 2 2 1 1 1 1 2 1 1 1 2 2 1 1 2 1 2 1 2 2 1 1 2 2 1 1 1 2
## [211] 1 1 2 2 1 2 1 2 1 2 1 2 2 2 1 2 1 2 1 1 2 1 1 1 2 2 2 2 2 1 1 2 1 1 1
## [246] 2 1 1 1 2 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 2 1 1 2 2 2 1 1 1 1 1
## [281] 2 2 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1 2 1 1 2 1 1 1 2 1 2 1 1 2 2 1 2 1 1
## [316] 1 2 1 1 1 2 2 2 1 1 1 1 1 1 2 2 2 1 2 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 1
## [351] 2 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 2 2 1 1 2 1 1 1 1 1 1 1 1 2 1 2
## [386] 1 2 2 2 1 1 2 1 2 2 2 2 1 2 2 1 2 1 1 2 1 1 1 2 1 2 2 1 1 2 2 1 2 2 2
## [421] 2 1 1 1 2 2 2 1 1 2 2 2 1 2 2 2 1 1 1 2 2 1 2 1 1 2 1 1 2 1 1 2 2 2 1
## [456] 1 1 2 2 1 2 2 2 2 2 1 1 1 1 1 2 2 2 1 1 1 1 1 1 2 2 1 1 2 2 2 2 2 1 1
## [491] 1 1 2 1 2 2 2 1 2 2 1 1 2 1 1 1 1 1 1 2 2 1 1 2 2 1 1 1 2 2 2 1 2 2 2
## [526] 1 2 2 1 2 1 2 1 2 1 2 1 1 2 1 1 1 1 1 2 2 1 2 1 2 1 1 1 1 2 2 2 1 1 2
## [561] 1 1 1 1 2 2 1 1 2 2 1 1 2 2 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 1 1 1 2 2 1
## [596] 1 1 1 2 1 1 2 1 2 1 1 1 1 2 2 1 2 2 1 2 1 1 1 2 2 1 2 2 1 1 1 2 1 1 1
## [631] 1 1 2 2 1 2 1 2 2 1 2 2 1 2 2 1 1 2 2 2 2 2 1 2 2 1 2 1 1 1 1 2 1 2 2
## [666] 2 2 2 1 1 2 2 2 1 1 2 1 1 2 2 1 2 2 1 2 1 2 2 2 1 2 1 1 1 2 2 1 2 1 1
```

```r
str(train_pred)
```

```
##  Factor w/ 2 levels "No","Yes": 1 2 1 2 2 1 2 2 2 2 ...
##  - attr(*, "names")= chr [1:700] "1" "2" "3" "4" ...
```

```r
accuracy(actual = credit_train$bad_credit,predicted = train_pred)
```

```
## [1] 0.7042857
```

```r
confusionMatrix(train_pred,credit_train$bad_credit,positive = 'Yes')
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No Yes
##        No  325  42
##        Yes 165 168
##
##                Accuracy : 0.7043
##                  95% CI : (0.669, 0.7379)
##     No Information Rate : 0.7
##     P-Value [Acc > NIR] : 0.4204
##
##                   Kappa : 0.3969
##  Mcnemar's Test P-Value : <2e-16
##
```

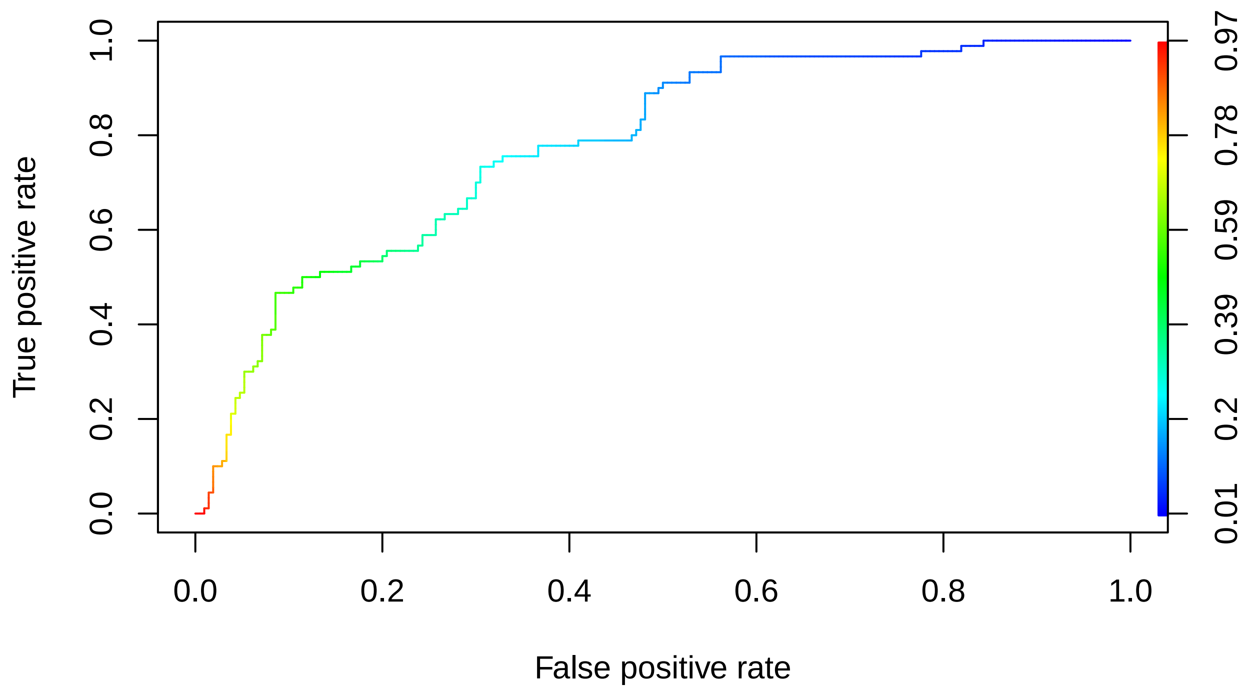
```
##              Sensitivity : 0.8000
##              Specificity : 0.6633
##           Pos Pred Value : 0.5045
##           Neg Pred Value : 0.8856
##               Prevalence : 0.3000
##           Detection Rate : 0.2400
##     Detection Prevalence : 0.4757
##        Balanced Accuracy : 0.7316
##
##         'Positive' Class : Yes
##
```

Let's eventually test our model

```r
pred_test<-predict(mod3,newdata = credit_test,type='response')

library(ROCR)
P_test<-prediction(pred_test,credit_test$bad_credit)
perf<-performance(P_test,'tpr','fpr')

plot(perf,colorize=T)
```




```r
performance(P_test,'auc')@y.values
```

```
## [[1]]
## [1] 0.7782011
```

```r
pred_test<-ifelse(pred_test>0.25,1,0)

table(pred_test,credit_test$bad_credit)
```

```
##
## pred_test  No Yes
##         0 141  22
##         1  69  68
```

```r
table(pred_test)
```

```
## pred_test
##   0   1
## 163 137
```

```r
pred_test<-factor(pred_test, levels = c(0, 1), labels = c('No','Yes'))

levels(pred_test)
```

```
## [1] "No"  "Yes"
```

```r
as.numeric(pred_test)
```

```
##   [1] 2 1 1 1 2 2 2 1 2 1 2 1 1 1 1 1 1 1 2 2 1 1 1 2 1 1 1 2 1 2 1 1 2 1 1
##  [36] 2 1 1 2 1 1 1 1 2 2 1 2 2 1 1 1 2 2 2 2 2 2 2 2 1 1 2 1 1 1 2 2 1 1
##  [71] 1 2 1 1 2 1 2 1 1 2 2 2 1 2 1 1 1 2 1 2 2 1 1 1 1 1 2 2 2 2 2 1 2 2 1
## [106] 2 2 2 1 1 2 2 1 2 1 2 1 2 1 2 1 1 1 1 1 1 2 1 2 1 1 2 1 2 2 1 1 2 1 2
## [141] 1 1 2 2 1 1 1 1 1 1 1 2 2 2 2 1 1 2 1 2 2 1 1 2 1 2 2 2 2 1 2 1 2 2 1
## [176] 2 1 1 2 2 1 2 1 2 2 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 1 1 1 1 1 2 1 1 2 2
## [211] 1 2 1 2 2 2 2 1 2 2 1 2 1 2 2 2 1 2 2 1 1 1 1 2 1 1 2 1 2 1 1 1 2 1 2
## [246] 2 2 1 1 2 1 1 2 1 1 1 2 1 1 2 2 1 2 2 2 1 2 1 1 1 1 2 1 1 2 2 1 1 1 1
## [281] 2 1 1 2 2 1 1 1 1 1 1 1 1 2 2 2 1 1 1 2
```

```r
str(pred_test)
```

```
##  Factor w/ 2 levels "No","Yes": 2 1 1 1 2 2 2 1 2 1 ...
##  - attr(*, "names")= chr [1:300] "5" "7" "9" "21" ...
```

```r
accuracy(actual = credit_test$bad_credit,predicted = pred_test)
```

```
## [1] 0.6966667
```

```r
confusionMatrix(pred_test,credit_test$bad_credit,positive = 'Yes')
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No Yes
##        No  141  22
##        Yes  69  68
##
##                Accuracy : 0.6967
##                  95% CI : (0.6412, 0.7482)
##     No Information Rate : 0.7
##     P-Value [Acc > NIR] : 0.5781
##
##                   Kappa : 0.3715
##  Mcnemar's Test P-Value : 1.42e-06
```

```
##
##             Sensitivity : 0.7556
##             Specificity : 0.6714
##          Pos Pred Value : 0.4964
##          Neg Pred Value : 0.8650
##              Prevalence : 0.3000
##          Detection Rate : 0.2267
##    Detection Prevalence : 0.4567
##       Balanced Accuracy : 0.7135
##
##        'Positive' Class : Yes
##
```

As a matter of fact confusion matrix that we get in Python is different in positions so not be fooled!!!!