hw1

ParkJiHeon

2021 3 22

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## √ ggplot2 3.3.3 √ purrr 0.3.4  
## √ tibble 3.1.0 √ dplyr 1.0.5  
## √ tidyr 1.1.3 √ stringr 1.4.0  
## √ readr 1.4.0 √ forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(VIM)

## Loading required package: colorspace

## Loading required package: grid

## VIM is ready to use.

## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues

##   
## Attaching package: 'VIM'

## The following object is masked from 'package:datasets':  
##   
## sleep

library(cowplot)  
library(ggcorrplot)  
library(GGally)

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(ggthemes)

##   
## Attaching package: 'ggthemes'

## The following object is masked from 'package:cowplot':  
##   
## theme\_map

library(MASS)

##   
## Attaching package: 'MASS'

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

library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

library(caret)

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library(ROCR)  
library(e1071)  
library(gmodels)  
library(splitstackshape)  
load("telData20210317.Rda")  
#summary(finalData)  
data <- read.csv("ChurnData.csv",stringsAsFactors = T)  
data <- data[1:1000,]  
summary(data)

## 고객ID 성별 연령 서비스기간 단선횟수   
## C101180: 1 : 0 Min. :13.00 Min. : 9.667 Min. : 0.000   
## C101300: 1 남:524 1st Qu.:30.00 1st Qu.:22.867 1st Qu.: 1.000   
## C101770: 1 여:476 Median :35.00 Median :34.633 Median : 1.000   
## C101780: 1 Mean :37.54 Mean :34.304 Mean : 2.588   
## C101800: 1 3rd Qu.:44.00 3rd Qu.:46.200 3rd Qu.: 2.000   
## C101810: 1 Max. :71.00 Max. :58.200 Max. :15.000   
## (Other):994   
## 주간통화횟수 주간통화시간\_분 야간통화횟수 야간통화시간\_분  
## Min. : 0.0 Min. : 0.0 Min. : 0.0 Min. : 0.0   
## 1st Qu.: 177.8 1st Qu.: 606.0 1st Qu.: 27.0 1st Qu.:134.7   
## Median : 372.5 Median : 848.1 Median : 68.0 Median :255.9   
## Mean : 416.4 Mean : 924.0 Mean : 96.5 Mean :272.8   
## 3rd Qu.: 581.0 3rd Qu.:1150.3 3rd Qu.:145.0 3rd Qu.:383.2   
## Max. :1581.0 Max. :2846.4 Max. :500.0 Max. :883.8   
##   
## 주말통화횟수 주말통화시간\_분 국제통화시간\_분 국내통화횟수   
## Min. : 0.00 Min. : 0.00 Min. : 0.1943 Min. : 16.0   
## 1st Qu.: 4.00 1st Qu.: 21.90 1st Qu.:212.3943 1st Qu.: 277.0   
## Median :10.00 Median : 43.00 Median :262.4638 Median : 487.0   
## Mean :15.92 Mean : 49.73 Mean :290.5778 Mean : 528.8   
## 3rd Qu.:23.00 3rd Qu.: 72.00 3rd Qu.:345.2667 3rd Qu.: 732.2   
## Max. :92.00 Max. :195.50 Max. :892.5571 Max. :1802.0   
##   
## 국내통화시간\_분 총통화시간\_분 통화량구분 총통화요금   
## Min. : 67.69 Min. : 69.1 : 0 Min. : 0.0907   
## 1st Qu.: 887.48 1st Qu.:1120.8 고 : 90 1st Qu.: 78.6070   
## Median :1159.40 Median :1442.9 저 : 7 Median :100.2785   
## Mean :1246.57 Mean :1537.1 중 :575 Mean :114.1718   
## 3rd Qu.:1560.05 3rd Qu.:1892.3 중고:294 3rd Qu.:129.6040   
## Max. :3192.90 Max. :3899.8 중저: 34 Max. :410.3089   
##   
## 부과요금 통화품질불만 이탈여부   
## Min. : 60.03 Mode :logical : 0   
## 1st Qu.:206.91 FALSE:976 유지:700   
## Median :235.47 TRUE :24 이탈:300   
## Mean :247.41   
## 3rd Qu.:278.29   
## Max. :560.31   
##

pSeed<-12345  
set.seed(pSeed)  
training800 <-stratified(finalData,"이탈여부",0.8)  
testing200 <-finalData[!finalData$고객ID%in%training800$고객ID,]  
str(finalData)

## 'data.frame': 1000 obs. of 24 variables:  
## $ 고객ID : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ 성별 : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 1 1 2 ...  
## $ 통화품질불만 : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ 이탈여부 : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 1 2 1 ...  
## $ d.dummy.고 : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ d.dummy.중고 : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ d.dummy.중 : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 2 2 2 ...  
## $ d.dummy.중저 : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ d.dummy.저 : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 1 1 1 ...  
## $ 연령 : int 37 38 30 33 40 44 26 45 34 27 ...  
## $ 서비스기간 : num 20.3 35 17.5 26 51.5 ...  
## $ 단선횟수 : int 1 1 0 0 2 1 0 2 2 5 ...  
## $ 주간통화횟수 : int 29 25 2 10 7 4 7 234 337 345 ...  
## $ 주간통화시간\_분: num 46.8 33 67.8 52.2 25 ...  
## $ 야간통화횟수 : int 17 20 6 48 24 17 1 185 11 75 ...  
## $ 야간통화시간\_분: num 47.4 19.8 30.6 90.3 39.7 ...  
## $ 주말통화횟수 : int 3 2 8 12 2 24 27 46 29 6 ...  
## $ 주말통화시간\_분: num 30 37.22 75 19.5 2.95 ...  
## $ 국제통화시간\_분: num 14.56 12.09 42.72 19.3 1.41 ...  
## $ 국내통화횟수 : int 49 47 16 70 33 45 35 465 377 426 ...  
## $ 국내통화시간\_분: num 124.2 90.1 173.4 162 67.7 ...  
## $ 총통화시간\_분 : num 138.8 102.2 216.1 181.3 69.1 ...  
## $ 총통화요금 : num 5.822 4.837 17.086 7.72 0.565 ...  
## $ 부과요금 : num 65.8 64.8 77 67.7 60.5 ...

table(training800$이탈여부)

##   
## 0 1   
## 560 240

table(testing200$이탈여부)

##   
## 0 1   
## 140 60

training800<-training800[,-1]  
model1\_1 <- glm(이탈여부~.,data = training800,family = binomial(link = "logit"))

## Warning: glm.fit: algorithm did not converge

summary(model1\_1)

##   
## Call:  
## glm(formula = 이탈여부 ~ ., family = binomial(link = "logit"),   
## data = training800)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.1865 -0.7519 -0.6585 1.1403 2.0592   
##   
## Coefficients: (3 not defined because of singularities)  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.807e+14 1.291e+14 -2.174 0.02969 \*   
## 성별1 3.574e-01 1.676e-01 2.133 0.03294 \*   
## 통화품질불만1 3.413e+00 1.063e+00 3.212 0.00132 \*\*  
## d.dummy.고1 2.807e+14 1.291e+14 2.174 0.02969 \*   
## d.dummy.중고1 2.807e+14 1.291e+14 2.174 0.02969 \*   
## d.dummy.중1 2.807e+14 1.291e+14 2.174 0.02969 \*   
## d.dummy.중저1 2.807e+14 1.291e+14 2.174 0.02969 \*   
## d.dummy.저1 2.807e+14 1.291e+14 2.174 0.02969 \*   
## 연령 -6.902e-03 8.722e-03 -0.791 0.42879   
## 서비스기간 6.622e-04 6.031e-03 0.110 0.91257   
## 단선횟수 1.773e-02 2.767e-02 0.641 0.52160   
## 주간통화횟수 -1.069e-04 6.757e-04 -0.158 0.87434   
## 주간통화시간\_분 -3.131e-03 1.889e-03 -1.657 0.09751 .   
## 야간통화횟수 2.503e-05 1.322e-03 0.019 0.98490   
## 야간통화시간\_분 -2.637e-03 1.770e-03 -1.490 0.13628   
## 주말통화횟수 -2.466e-03 7.370e-03 -0.335 0.73794   
## 주말통화시간\_분 -3.294e-03 3.626e-03 -0.908 0.36363   
## 국제통화시간\_분 -7.528e-03 6.107e-03 -1.233 0.21773   
## 국내통화횟수 NA NA NA NA   
## 국내통화시간\_분 NA NA NA NA   
## 총통화시간\_분 NA NA NA NA   
## 총통화요금 -1.171e-02 6.984e-03 -1.677 0.09357 .   
## 부과요금 3.547e-02 2.427e-02 1.461 0.14398   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 977.38 on 799 degrees of freedom  
## Residual deviance: 903.07 on 780 degrees of freedom  
## AIC: 943.07  
##   
## Number of Fisher Scoring iterations: 25

model1\_2 <- stepAIC(model1\_1,direction = "both")

## Start: AIC=943.07  
## 이탈여부 ~ 성별 + 통화품질불만 + d.dummy.고 + d.dummy.중고 +   
## d.dummy.중 + d.dummy.중저 + d.dummy.저 + 연령 + 서비스기간 +   
## 단선횟수 + 주간통화횟수 + 주간통화시간\_분 + 야간통화횟수 +   
## 야간통화시간\_분 + 주말통화횟수 + 주말통화시간\_분 + 국제통화시간\_분 +   
## 국내통화횟수 + 국내통화시간\_분 + 총통화시간\_분 + 총통화요금 +   
## 부과요금

## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge

## Warning in stepAIC(model1\_1, direction = "both"): 0 df terms are changing AIC

## Warning: glm.fit: algorithm did not converge

##   
## Step: AIC=939.92  
## 이탈여부 ~ 성별 + 통화품질불만 + d.dummy.고 + d.dummy.중고 +   
## d.dummy.중 + d.dummy.중저 + d.dummy.저 + 연령 + 서비스기간 +   
## 단선횟수 + 주간통화횟수 + 주간통화시간\_분 + 야간통화횟수 +   
## 야간통화시간\_분 + 주말통화횟수 + 주말통화시간\_분 + 국제통화시간\_분 +   
## 국내통화횟수 + 국내통화시간\_분 + 총통화시간\_분 + 총통화요금

## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge

## Warning in stepAIC(model1\_1, direction = "both"): 0 df terms are changing AIC

##   
## Step: AIC=935.7  
## 이탈여부 ~ 성별 + 통화품질불만 + d.dummy.고 + d.dummy.중고 +   
## d.dummy.중 + d.dummy.중저 + d.dummy.저 + 연령 + 서비스기간 +   
## 단선횟수 + 주간통화횟수 + 주간통화시간\_분 + 야간통화횟수 +   
## 주말통화횟수 + 주말통화시간\_분 + 국제통화시간\_분 + 국내통화횟수 +   
## 국내통화시간\_분 + 총통화시간\_분 + 총통화요금

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge

## Warning in stepAIC(model1\_1, direction = "both"): 0 df terms are changing AIC

##   
## Step: AIC=935.7  
## 이탈여부 ~ 성별 + 통화품질불만 + d.dummy.고 + d.dummy.중고 +   
## d.dummy.중 + d.dummy.중저 + d.dummy.저 + 연령 + 서비스기간 +   
## 단선횟수 + 주간통화횟수 + 주간통화시간\_분 + 야간통화횟수 +   
## 주말통화횟수 + 주말통화시간\_분 + 국제통화시간\_분 + 국내통화시간\_분 +   
## 총통화시간\_분 + 총통화요금

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge  
  
## Warning: glm.fit: algorithm did not converge

## Warning in stepAIC(model1\_1, direction = "both"): 0 df terms are changing AIC

##   
## Step: AIC=934.33  
## 이탈여부 ~ 성별 + 통화품질불만 + d.dummy.고 + d.dummy.중고 +   
## d.dummy.중 + d.dummy.중저 + d.dummy.저 + 연령 + 서비스기간 +   
## 단선횟수 + 주간통화횟수 + 주간통화시간\_분 + 야간통화횟수 +   
## 주말통화횟수 + 주말통화시간\_분 + 국제통화시간\_분 + 국내통화시간\_분 +   
## 총통화시간\_분

## Warning: glm.fit: algorithm did not converge  
  
## Warning: 0 df terms are changing AIC

##   
## Step: AIC=934.33  
## 이탈여부 ~ 성별 + 통화품질불만 + d.dummy.고 + d.dummy.중고 +   
## d.dummy.중 + d.dummy.중저 + d.dummy.저 + 연령 + 서비스기간 +   
## 단선횟수 + 주간통화횟수 + 주간통화시간\_분 + 야간통화횟수 +   
## 주말통화횟수 + 주말통화시간\_분 + 국내통화시간\_분 + 총통화시간\_분

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge

## Warning in stepAIC(model1\_1, direction = "both"): 0 df terms are changing AIC

## Warning: glm.fit: algorithm did not converge

##   
## Step: AIC=935.08  
## 이탈여부 ~ 성별 + 통화품질불만 + d.dummy.고 + d.dummy.중고 +   
## d.dummy.중 + d.dummy.중저 + d.dummy.저 + 연령 + 서비스기간 +   
## 단선횟수 + 주간통화횟수 + 주간통화시간\_분 + 야간통화횟수 +   
## 주말통화횟수 + 주말통화시간\_분 + 국내통화시간\_분

summary(model1\_2)

##   
## Call:  
## glm(formula = 이탈여부 ~ 성별 + 통화품질불만 + d.dummy.고 + d.dummy.중고 +   
## d.dummy.중 + d.dummy.중저 + d.dummy.저 + 연령 + 서비스기간 +   
## 단선횟수 + 주간통화횟수 + 주간통화시간\_분 + 야간통화횟수 +   
## 주말통화횟수 + 주말통화시간\_분 + 국내통화시간\_분, family = binomial(link = "logit"),   
## data = training800)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2668 -0.7953 -0.7083 1.1518 1.9294   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.602e+13 7.078e+13 -0.791 0.42870   
## 성별1 3.583e-01 1.667e-01 2.150 0.03156 \*   
## 통화품질불만1 3.369e+00 1.042e+00 3.233 0.00123 \*\*  
## d.dummy.고1 5.602e+13 7.078e+13 0.791 0.42870   
## d.dummy.중고1 5.602e+13 7.078e+13 0.791 0.42870   
## d.dummy.중1 5.602e+13 7.078e+13 0.791 0.42870   
## d.dummy.중저1 5.602e+13 7.078e+13 0.791 0.42870   
## d.dummy.저1 5.602e+13 7.078e+13 0.791 0.42870   
## 연령 -6.793e-03 8.676e-03 -0.783 0.43362   
## 서비스기간 9.164e-04 5.998e-03 0.153 0.87857   
## 단선횟수 1.813e-02 2.748e-02 0.660 0.50950   
## 주간통화횟수 4.585e-05 6.663e-04 0.069 0.94514   
## 주간통화시간\_분 -3.256e-04 7.873e-04 -0.414 0.67924   
## 야간통화횟수 3.482e-05 1.318e-03 0.026 0.97892   
## 주말통화횟수 -1.575e-03 7.337e-03 -0.215 0.83002   
## 주말통화시간\_분 -8.712e-04 3.288e-03 -0.265 0.79105   
## 국내통화시간\_분 -3.749e-04 7.922e-04 -0.473 0.63600   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 977.38 on 799 degrees of freedom  
## Residual deviance: 901.08 on 783 degrees of freedom  
## AIC: 935.08  
##   
## Number of Fisher Scoring iterations: 25

formula(model1\_2)

## 이탈여부 ~ 성별 + 통화품질불만 + d.dummy.고 + d.dummy.중고 +   
## d.dummy.중 + d.dummy.중저 + d.dummy.저 + 연령 + 서비스기간 +   
## 단선횟수 + 주간통화횟수 + 주간통화시간\_분 + 야간통화횟수 +   
## 주말통화횟수 + 주말통화시간\_분 + 국내통화시간\_분

(vif\_vars <- as.data.frame(vif(model1\_2)))

## vif(model1\_2)  
## 성별 1.112531e+00  
## 통화품질불만 5.834168e-01  
## d.dummy.고 1.422189e+15  
## d.dummy.중고 9.481262e+14  
## d.dummy.중 6.162821e+15  
## d.dummy.중저 4.740631e+14  
## d.dummy.저 9.481262e+14  
## 연령 1.009470e+00  
## 서비스기간 1.015450e+00  
## 단선횟수 1.093504e+00  
## 주간통화횟수 1.864372e+00  
## 주간통화시간\_분 4.729658e+00  
## 야간통화횟수 1.840731e+00  
## 주말통화횟수 2.051327e+00  
## 주말통화시간\_분 2.084938e+00  
## 국내통화시간\_분 3.795819e+00

model1\_3 <- glm(이탈여부~성별+d.dummy.고+d.dummy.중저+연령+서비스기간+단선횟수+주간통화횟수+주간통화시간\_분+야간통화횟수+주말통화횟수+주말통화시간\_분+국내통화시간\_분,data = training800,family = binomial)  
summary(model1\_3)

##   
## Call:  
## glm(formula = 이탈여부 ~ 성별 + d.dummy.고 + d.dummy.중저 + 연령 +   
## 서비스기간 + 단선횟수 + 주간통화횟수 + 주간통화시간\_분 +   
## 야간통화횟수 + 주말통화횟수 + 주말통화시간\_분 + 국내통화시간\_분,   
## family = binomial, data = training800)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.6108 -0.8420 -0.7247 1.2528 2.0014   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.1453107 0.4831869 -0.301 0.7636   
## 성별1 0.3860431 0.1606879 2.402 0.0163 \*   
## d.dummy.고1 0.9550334 0.3823342 2.498 0.0125 \*   
## d.dummy.중저1 1.0106174 0.4581849 2.206 0.0274 \*   
## 연령 -0.0112470 0.0083263 -1.351 0.1768   
## 서비스기간 -0.0004309 0.0057466 -0.075 0.9402   
## 단선횟수 0.0648136 0.0228142 2.841 0.0045 \*\*  
## 주간통화횟수 0.0008591 0.0006280 1.368 0.1713   
## 주간통화시간\_분 -0.0007172 0.0007712 -0.930 0.3524   
## 야간통화횟수 0.0002044 0.0013006 0.157 0.8751   
## 주말통화횟수 -0.0008491 0.0071070 -0.119 0.9049   
## 주말통화시간\_분 -0.0010397 0.0031846 -0.326 0.7441   
## 국내통화시간\_분 -0.0003652 0.0006951 -0.525 0.5993   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 977.38 on 799 degrees of freedom  
## Residual deviance: 933.54 on 787 degrees of freedom  
## AIC: 959.54  
##   
## Number of Fisher Scoring iterations: 4

#model1\_1 AIC 936.85  
#model1\_2 AIC 923.18 lowest 채택  
#model1\_3 AIC 933.06  
  
  
pred <- predict(model1\_2,newdata = testing200[,which(colnames(testing200)!= "이탈여부")],type = "response")  
pred

## 2 11 35 36 37 47 56 57   
## 1.0000000 0.3443393 0.2151747 0.2290902 0.2161243 0.2427841 0.2169052 0.3103905   
## 59 75 83 95 105 119 122 133   
## 0.1982491 0.2101780 0.3154100 0.2597690 0.2193870 0.2637716 0.2701142 0.2735292   
## 136 137 140 141 142 146 148 151   
## 0.2925393 0.2212410 0.2721762 0.2534923 0.2406032 0.2595409 0.2790059 0.2379014   
## 160 166 175 177 182 183 187 191   
## 0.2757462 0.3241738 0.2196149 0.2839871 0.2188791 0.2822584 0.2773229 0.2136556   
## 195 199 200 205 210 215 217 221   
## 0.2975555 0.3119272 0.3329426 0.2713892 0.2696465 0.2797452 0.2171362 0.2409358   
## 224 234 247 249 250 254 257 263   
## 0.1894704 0.2338071 0.2918038 0.2963729 0.1955830 0.2246624 0.2929136 0.2619469   
## 266 286 291 298 302 305 309 311   
## 0.2565378 0.3170087 0.3970932 0.2604326 0.2694929 0.2488741 0.2184067 0.2262109   
## 313 317 326 327 329 330 333 338   
## 0.3015496 0.2230211 0.4024670 0.2283291 0.1883645 0.1856746 0.4523195 0.2319617   
## 348 360 363 370 378 387 388 390   
## 0.2045058 0.2428121 0.2552512 0.2397655 0.3720691 0.2851896 0.2378800 0.2606646   
## 394 398 403 415 418 421 422 428   
## 0.3310678 0.2430762 0.3080738 0.2658250 0.2582119 0.1935970 0.2496778 0.2047884   
## 431 434 436 437 442 448 449 454   
## 0.2918169 0.3145894 0.3095305 0.2991642 0.2295065 0.2347197 0.4515701 0.2463538   
## 455 460 461 462 463 466 487 488   
## 0.2566090 0.2881909 0.2164925 0.2905032 0.2318734 0.2773211 0.2227622 0.2211584   
## 491 492 493 500 504 514 515 520   
## 0.2548352 0.2064111 0.3737310 0.2153343 0.2234739 0.3062947 0.4402306 0.3312028   
## 524 528 533 534 536 540 547 550   
## 0.2473576 0.2495907 0.2003692 0.3268521 0.3069773 0.2860441 0.2884109 0.1925836   
## 555 557 564 596 598 605 606 608   
## 0.3030448 0.2271560 0.1858934 0.2493943 0.3024440 0.2540880 0.3168467 0.2369925   
## 610 611 625 627 630 632 639 643   
## 0.2870896 0.2283922 0.2803028 0.3000195 0.1867506 0.1689534 0.2440003 0.1837311   
## 644 646 651 657 658 660 663 665   
## 0.1980372 0.4304186 0.1976992 0.2480633 0.1925333 0.2357945 0.4658907 0.2633666   
## 671 675 677 679 680 683 685 693   
## 0.2299692 0.3079768 0.2013618 0.2313485 0.2768374 0.2742676 0.2948758 0.2219381   
## 696 697 702 710 714 717 718 719   
## 0.2544833 0.2653346 0.2365539 0.2688990 0.2820346 0.3071793 0.2046290 0.2704777   
## 722 730 738 742 746 748 753 755   
## 0.4515669 0.2621822 0.2720141 0.2519910 0.2050055 0.3128859 0.2063233 0.2464564   
## 763 772 795 796 803 805 812 818   
## 0.3242986 0.1877584 0.3111681 0.2116081 0.2293506 0.2349478 0.1947681 0.1827454   
## 821 827 837 862 863 869 872 879   
## 0.1983089 0.2706561 0.2401131 0.2738439 0.2781530 0.1838791 0.1721913 0.2642789   
## 880 882 884 885 888 895 911 913   
## 0.1829639 0.1737778 0.2674742 0.2665937 0.2485623 0.2719467 0.1963972 0.9813319   
## 928 929 931 934 938 946 949 953   
## 0.2696740 0.1726724 0.1808128 0.2771819 0.1853585 0.1782897 0.1660459 0.2594173   
## 954 960 977 980 982 984 989 997   
## 0.9460790 0.5445031 0.5213212 0.5344488 0.9181357 0.6086160 0.9201546 0.9233628

summary(pred)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1660 0.2212 0.2574 0.2851 0.2926 1.0000

# check model test  
pred\_churn <- factor(ifelse(pred >= 0.50, "Yes", "No"))  
actual\_churn <- factor(ifelse(testing200$이탈여부==1,"Yes","No"))  
pred\_churn

## 2 11 35 36 37 47 56 57 59 75 83 95 105 119 122 133 136 137 140 141   
## Yes No No No No No No No No No No No No No No No No No No No   
## 142 146 148 151 160 166 175 177 182 183 187 191 195 199 200 205 210 215 217 221   
## No No No No No No No No No No No No No No No No No No No No   
## 224 234 247 249 250 254 257 263 266 286 291 298 302 305 309 311 313 317 326 327   
## No No No No No No No No No No No No No No No No No No No No   
## 329 330 333 338 348 360 363 370 378 387 388 390 394 398 403 415 418 421 422 428   
## No No No No No No No No No No No No No No No No No No No No   
## 431 434 436 437 442 448 449 454 455 460 461 462 463 466 487 488 491 492 493 500   
## No No No No No No No No No No No No No No No No No No No No   
## 504 514 515 520 524 528 533 534 536 540 547 550 555 557 564 596 598 605 606 608   
## No No No No No No No No No No No No No No No No No No No No   
## 610 611 625 627 630 632 639 643 644 646 651 657 658 660 663 665 671 675 677 679   
## No No No No No No No No No No No No No No No No No No No No   
## 680 683 685 693 696 697 702 710 714 717 718 719 722 730 738 742 746 748 753 755   
## No No No No No No No No No No No No No No No No No No No No   
## 763 772 795 796 803 805 812 818 821 827 837 862 863 869 872 879 880 882 884 885   
## No No No No No No No No No No No No No No No No No No No No   
## 888 895 911 913 928 929 931 934 938 946 949 953 954 960 977 980 982 984 989 997   
## No No No Yes No No No No No No No No Yes Yes Yes Yes Yes Yes Yes Yes   
## Levels: No Yes

actual\_churn

## [1] Yes No No No No No No No Yes Yes No Yes No Yes No No Yes No   
## [19] No No No Yes No Yes Yes Yes No No Yes Yes No No No No No No   
## [37] No No Yes Yes No No Yes No No No No No No No No No No No   
## [55] Yes No Yes No No No No No No No No Yes Yes No Yes No Yes No   
## [73] No No No Yes No No No No No No No Yes No No No Yes Yes No   
## [91] No No No No No No No No No No Yes Yes No No Yes No No Yes  
## [109] Yes No No No Yes Yes Yes No No No Yes No Yes No No No Yes No   
## [127] No No No No No No Yes No No No Yes No Yes No Yes No Yes No   
## [145] No Yes No No Yes No No No No No Yes No Yes No Yes No No No   
## [163] Yes Yes No No Yes No No No Yes No Yes No No No No No No Yes  
## [181] No Yes No Yes No No No No No No No Yes Yes Yes No No Yes No   
## [199] Yes Yes  
## Levels: No Yes

table(actual\_churn,pred\_churn)

## pred\_churn  
## actual\_churn No Yes  
## No 137 3  
## Yes 53 7

caret::confusionMatrix(pred\_churn,actual\_churn,positive = "Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 137 53  
## Yes 3 7  
##   
## Accuracy : 0.72   
## 95% CI : (0.6523, 0.781)  
## No Information Rate : 0.7   
## P-Value [Acc > NIR] : 0.2972   
##   
## Kappa : 0.125   
##   
## Mcnemar's Test P-Value : 5.835e-11   
##   
## Sensitivity : 0.1167   
## Specificity : 0.9786   
## Pos Pred Value : 0.7000   
## Neg Pred Value : 0.7211   
## Prevalence : 0.3000   
## Detection Rate : 0.0350   
## Detection Prevalence : 0.0500   
## Balanced Accuracy : 0.5476   
##   
## 'Positive' Class : Yes   
##

###로지스틱회귀분석 적중률 72