Churn_Analysis_best_case

2024-02-03

```
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
```

```
library('caret')
```

```
## Loading required package: ggplot2

## Loading required package: lattice
```

```
churn <- read.csv('Telco_customer_churn_cleaned.csv')
#head(churn)

#Eliminated Churn Label, Churn Score, and CLTV as we are not using it
churn <- churn[ -c(30, 32, 33) ]
#head(churn)
```

```
# Create the best case (not churned) and worst case (churned group)

unknown_churn <- filter(churn, churn$Tenure < 12 & churn$Churn_val== 0) #1070
unknown_churn_best <- filter(churn, churn$Tenure < 12 & churn$Churn_val== 0) #1070
unknown_churn_worst <- filter(churn, churn$Tenure < 12 & churn$Churn_val== 0) #1070
```

```
known_churn <- churn %>%
      filter(! CustomerID %in% unknown_churn$CustomerID) #5973

unknown_churn_best[ , 'churn_12month'] = 0
unknown_churn_worst[ , 'churn_12month'] = 1
known_churn[ , 'churn_12month'] = known_churn$Churn_val

best_case <- rbind(known_churn, unknown_churn_best)
worst_case <- rbind(known_churn, unknown_churn_worst)

# eliminate churn_val since we substitute them with 12 month churn_val using two cases
best_case <- best_case[-best_case$Churn_val]
worst_case <- worst_case[-worst_case$Churn_val]

dim(best_case)
```

```
## [1] 7043   31
```

```
dim(worst_case)
```

```
## [1] 7043   31
```

```
#write.csv(best_case, "C:\\Users\\user\\Desktop\\24 WI\\DATA #557\\Course_Proj\\best_case.csv", row.nam
#write.csv(worst_case, "C:\\Users\\user\\Desktop\\24 WI\\DATA #557\\Course_Proj\\worst_case.csv", row.n
```

```
# divide customer into churned and not churned group
churned <- best_case[best_case$churn_12month == 1,]
not_churned <- best_case[best_case$churn_12month == 0,]
n_churned = 1869
n_not_churned = 5174
```

```
# Code that manually checks the confidence interval
mean_churned = mean(churned$Tenure)
mean_not_churned = mean(not_churned$Tenure)

se_churned = sd(churned$Tenure)/sqrt(n_churned)
se_not_churned = sd(not_churned$Tenure)/sqrt(n_not_churned)

churned_CI = c(mean_churned - (se_churned * 1.96), mean_churned + (se_churned * 1.96))
not_churned_CI = c(mean_not_churned - (se_not_churned * 1.96), mean_not_churned + (se_not_churned * 1.9

churned_CI
not_churned_CI

s = sqrt((var(churned$Tenure)+var(not_churned$Tenure))/(n_churned + n_not_churned -2))
two_mean_se = s * (sqrt(1/n_churned+1/n_not_churned))
mean_diff = mean_churned - mean_not_churned

diff_CI = c(mean_diff- (1.96*two_mean_se), mean_diff+(1.96*two_mean_se))
diff_CI
```

```r
sum(is.na(best_case)) # sanity check for na values
```

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

```r
# Confidence interval for tenure months based on churned_12 month or not
t.test(churned$Tenure)$conf
```

```
## [1] 17.09310 18.86517
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(not_churned$Tenure)$conf
```

```
## [1] 36.91276 38.22717
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(churned$Tenure-not_churned$Tenure)$conf
```

```
## Warning in churned$Tenure - not_churned$Tenure: longer object length is not a
## multiple of shorter object length
```

```
## [1] -20.49165 -18.80908
## attr(,"conf.level")
## [1] 0.95
```

```r
# Confidence interval for total payment based on churned_12 month or not
# Think about way to standardize this -- total.chargs/tenure.month creats 11 null values
t.test(churned$Total.Charges)$conf
```

```
## [1] 1446.018 1617.574
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(not_churned$Total.Charges)$conf
```

```
## [1] 2486.410 2613.413
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(churned$Total.Charges-not_churned$Total.Charges)$conf
```

```
## Warning in churned$Total.Charges - not_churned$Total.Charges: longer object
## length is not a multiple of shorter object length
```

```
## [1] -1106.240  -942.207
## attr(,"conf.level")
## [1] 0.95
```

```r
# Contingency table for churn_12month and non-demographic qualitative variables
phone_service <- table(best_case$churn_12month, best_case$Phone.Service)
phone_service
```

```
##
##       No  Yes
##   0  512 4662
##   1  170 1699
```

```r
multi_lines <- table(best_case$churn_12month, best_case$Multiple.Lines)
multi_lines
```

```
##
##       No No phone service  Yes
##   0 2541              512 2121
##   1  849              170  850
```

```r
internet_service <- table(best_case$churn_12month, best_case$Internet.Service)
internet_service
```

```
##
##      DSL Fiber optic   No
##   0 1962        1799 1413
##   1  459        1297  113
```

```r
online_security <- table(best_case$churn_12month, best_case$Online.Security)
online_security
```

```
##
##       No No internet service  Yes
##   0 2037                1413 1724
##   1 1461                 113  295
```

```r
online_backup <- table(best_case$churn_12month, best_case$Online.Backup)
online_backup
```

```
##
##       No No internet service  Yes
##   0 1855                1413 1906
##   1 1233                 113  523
```

```r
device_protect <- table(best_case$churn_12month, best_case$Device.Protection)
device_protect
```

```
##
##       No No internet service  Yes
##   0 1884                1413 1877
##   1 1211                 113  545
```

```r
tech_support <- table(best_case$churn_12month, best_case$Tech.Support)
stream_tv <- table(best_case$churn_12month, best_case$Streaming.TV)
stream_tv
```

```
##
##       No No internet service  Yes
##   0 1868                1413 1893
##   1  942                 113  814
```

```r
stream_movies <- table(best_case$churn_12month, best_case$Streaming.Movies)
stream_movies
```

```
##
##       No No internet service  Yes
##   0 1847                1413 1914
##   1  938                 113  818
```

```r
payment <- table(best_case$churn_12month, best_case$Payment.Method)
payment
```

```
##
##     Bank transfer (automatic) Credit card (automatic) Electronic check
##   0                      1286                    1290             1294
##   1                       258                     232             1071
##
##     Mailed check
##   0         1304
##   1          308
```

```r
paperless <- table(best_case$churn_12month, best_case$Paperless.Billing)
paperless
```

```
##
##       No  Yes
##   0 2403 2771
##   1  469 1400
```

```r
contract <- table(best_case$churn_12month, best_case$Contract)
contract
```

```
##
##     Month-to-month One year Two year
##   0           2220     1307     1647
##   1           1655      166       48
```

```r
# just interested
age_paperless <- table(best_case$Senior.Citizen, best_case$Paperless.Billing)
age_paperless
```

```
##
##         No  Yes
##   No  2606 3295
##   Yes  266  876
```

```
# Separation of data set into training and test data set with stratification using churn_12 month (Resp
train.index <- createDataPartition(best_case$churn_12month, p = .7, list = FALSE)
train_best <- best_case[ train.index,]
#churned_train <- train_best[train_best$churn_12month == 1,]
#print(length(churned_train$churn_12month)/length(train_best$churn_12month))

test_best  <- best_case[-train.index,]
#churned_test <- test_best[test_best$churn_12month == 1,]
#print(length(churned_test$churn_12month)/length(test_best$churn_12month))

#churned <- best_case[best_case$churn_12month == 1,]
#print(length(churned$churn_12month)/length(best_case$churn_12month))
```

```
# Manually created model by me - includes relevent look-like factors
model1 <- glm(churn_12month ~ I(Senior.Citizen) + Tenure+I(Internet.Service)+I(Contract) + Total.Charges
summary(model1)
```

```
##
## Call:
## glm(formula = churn_12month ~ I(Senior.Citizen) + Tenure + I(Internet.Service) +
##     I(Contract) + Total.Charges + I(Payment.Method), family = "binomial",
##     data = train_best)
##
## Coefficients:
##                                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)                         -3.260e-01  1.277e-01  -2.552 0.010698
## I(Senior.Citizen)Yes                 3.821e-01  9.658e-02   3.957 7.59e-05
## Tenure                              -5.493e-02  6.928e-03  -7.929 2.22e-15
## I(Internet.Service)Fiber optic       8.761e-01  1.046e-01   8.379  < 2e-16
## I(Internet.Service)No               -7.955e-01  1.458e-01  -5.456 4.87e-08
## I(Contract)One year                 -7.605e-01  1.250e-01  -6.083 1.18e-09
## I(Contract)Two year                 -1.688e+00  2.124e-01  -7.949 1.88e-15
## Total.Charges                        2.673e-04  7.332e-05   3.645 0.000267
## I(Payment.Method)Credit card (automatic) -1.455e-01  1.345e-01  -1.082 0.279060
## I(Payment.Method)Electronic check    3.655e-01  1.100e-01   3.323 0.000890
## I(Payment.Method)Mailed check       -1.342e-01  1.338e-01  -1.003 0.315959
##
## (Intercept)                         *
## I(Senior.Citizen)Yes                ***
## Tenure                              ***
## I(Internet.Service)Fiber optic      ***
## I(Internet.Service)No               ***
## I(Contract)One year                 ***
## I(Contract)Two year                 ***
## Total.Charges                       ***
## I(Payment.Method)Credit card (automatic)
## I(Payment.Method)Electronic check   ***
## I(Payment.Method)Mailed check
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5705.0  on 4930  degrees of freedom
## Residual deviance: 4151.2  on 4920  degrees of freedom
## AIC: 4173.2
##
## Number of Fisher Scoring iterations: 6
```

```
# Did AIC for testing, it seems like AIC did not recommend to exclude any of factors
model1_step <- stepAIC(model1, trace = TRUE, direction = 'both')
```

```
## Start:  AIC=4173.15
## churn_12month ~ I(Senior.Citizen) + Tenure + I(Internet.Service) +
##     I(Contract) + Total.Charges + I(Payment.Method)
##
##                        Df Deviance    AIC
## <none>                      4151.2 4173.2
## - Total.Charges         1   4165.2 4185.2
## - I(Senior.Citizen)     1   4166.8 4186.8
## - I(Payment.Method)     3   4184.0 4200.0
## - Tenure                1   4226.0 4246.0
## - I(Contract)           2   4243.4 4261.4
## - I(Internet.Service)   2   4282.2 4300.2
```

```
model1_step$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## churn_12month ~ I(Senior.Citizen) + Tenure + I(Internet.Service) +
##     I(Contract) + Total.Charges + I(Payment.Method)
##
## Final Model:
## churn_12month ~ I(Senior.Citizen) + Tenure + I(Internet.Service) +
##     I(Contract) + Total.Charges + I(Payment.Method)
##
##
##   Step Df Deviance Resid. Df Resid. Dev      AIC
## 1                       4920   4151.154 4173.154
```

Including Interaction might depend on the collinearity

```
# model with every possible variable
model_2 <- glm(churn_12month ~ I(Senior.Citizen) + Tenure+I(Internet.Service)+I(Contract) + Total.Charg
summary(model_2)
```

```
##
## Call:
## glm(formula = churn_12month ~ I(Senior.Citizen) + Tenure + I(Internet.Service) +
##     I(Contract) + Total.Charges + I(Payment.Method) + Gender +
##     Partner + Dependents + Phone.Service + Multiple.Lines + Internet.Service +
##     Online.Security + Online.Backup + Contract + Paperless.Billing,
##     family = "binomial", data = train_best)
##
## Coefficients: (7 not defined because of singularities)
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             0.3189442  0.1983379   1.608 0.107817
## I(Senior.Citizen)Yes                    0.1721905  0.0994820   1.731 0.083475
## Tenure                                 -0.0622585  0.0072502  -8.587  < 2e-16
## I(Internet.Service)Fiber optic          0.7674294  0.1194613   6.424 1.33e-10
## I(Internet.Service)No                  -0.6559147  0.1628933  -4.027 5.66e-05
## I(Contract)One year                    -0.7004403  0.1287725  -5.439 5.35e-08
## I(Contract)Two year                    -1.5022323  0.2177329  -6.899 5.22e-12
## Total.Charges                           0.0003452  0.0000784   4.403 1.07e-05
## I(Payment.Method)Credit card (automatic) -0.1545693  0.1382910  -1.118 0.263691
## I(Payment.Method)Electronic check       0.2720337  0.1139167   2.388 0.016940
## I(Payment.Method)Mailed check          -0.1255264  0.1392040  -0.902 0.367193
## GenderMale                              0.0788522  0.0789550   0.999 0.317941
## PartnerYes                              0.2887544  0.0904309   3.193 0.001408
## DependentsYes                          -1.5667278  0.1449949 -10.805  < 2e-16
## Phone.ServiceYes                       -0.6896279  0.1624686  -4.245 2.19e-05
## Multiple.LinesNo phone service                 NA         NA      NA       NA
## Multiple.LinesYes                       0.2575738  0.0969418   2.657 0.007884
## Internet.ServiceFiber optic                    NA         NA      NA       NA
## Internet.ServiceNo                             NA         NA      NA       NA
## Online.SecurityNo internet service             NA         NA      NA       NA
## Online.SecurityYes                     -0.4526867  0.1035087  -4.373 1.22e-05
## Online.BackupNo internet service               NA         NA      NA       NA
## Online.BackupYes                       -0.1479743  0.0949943  -1.558 0.119300
## ContractOne year                               NA         NA      NA       NA
## ContractTwo year                               NA         NA      NA       NA
## Paperless.BillingYes                    0.3067274  0.0899632   3.409 0.000651
##
## (Intercept)
## I(Senior.Citizen)Yes                   .
## Tenure                                 ***
## I(Internet.Service)Fiber optic         ***
## I(Internet.Service)No                  ***
## I(Contract)One year                    ***
## I(Contract)Two year                    ***
## Total.Charges                          ***
## I(Payment.Method)Credit card (automatic)
## I(Payment.Method)Electronic check      *
## I(Payment.Method)Mailed check
## GenderMale
## PartnerYes                             **
## DependentsYes                          ***
## Phone.ServiceYes                       ***
## Multiple.LinesNo phone service
## Multiple.LinesYes                      **
```

```
## Internet.ServiceFiber optic
## Internet.ServiceNo
## Online.SecurityNo internet service
## Online.SecurityYes                        ***
## Online.BackupNo internet service
## Online.BackupYes
## ContractOne year
## ContractTwo year
## Paperless.BillingYes                      ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5705.0  on 4930  degrees of freedom
## Residual deviance: 3938.2  on 4912  degrees of freedom
## AIC: 3976.2
##
## Number of Fisher Scoring iterations: 6
```

```
model2_step <- stepAIC(model_2, trace = FALSE,direction = 'both')
model2_step$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## churn_12month ~ I(Senior.Citizen) + Tenure + I(Internet.Service) +
##     I(Contract) + Total.Charges + I(Payment.Method) + Gender +
##     Partner + Dependents + Phone.Service + Multiple.Lines + Internet.Service +
##     Online.Security + Online.Backup + Contract + Paperless.Billing
##
## Final Model:
## churn_12month ~ I(Senior.Citizen) + Tenure + I(Internet.Service) +
##     I(Contract) + Total.Charges + I(Payment.Method) + Partner +
##     Dependents + Multiple.Lines + Online.Security + Online.Backup +
##     Paperless.Billing
##
##
##                 Step Df  Deviance Resid. Df Resid. Dev      AIC
## 1                                      4912   3938.234 3976.234
## 2         - Contract  0 0.0000000      4912   3938.234 3976.234
## 3 - Internet.Service  0 0.0000000      4912   3938.234 3976.234
## 4    - Phone.Service  0 0.0000000      4912   3938.234 3976.234
## 5           - Gender  1 0.9977828      4913   3939.231 3975.231
```

```
null <- glm(churn_12month ~ 1,  data = train_best, family = "binomial")
step(null, scope = list(lower=null,upper=model_2),
     direction="both", criterion = "AIC", trace = FALSE)
```

```
##
## Call:  glm(formula = churn_12month ~ I(Contract) + I(Internet.Service) +
##     Tenure + Dependents + I(Payment.Method) + Multiple.Lines +
```

```
##     Online.Security + Total.Charges + Paperless.Billing + Partner +
##     I(Senior.Citizen) + Online.Backup, family = "binomial", data = train_best)
##
## Coefficients:
##                                (Intercept)
##                                  -0.330995
##                       I(Contract)One year
##                                  -0.702242
##                       I(Contract)Two year
##                                  -1.505769
##           I(Internet.Service)Fiber optic
##                                   0.765591
##                     I(Internet.Service)No
##                                  -0.658372
##                                     Tenure
##                                  -0.062146
##                              DependentsYes
##                                  -1.569440
## I(Payment.Method)Credit card (automatic)
##                                  -0.152205
##        I(Payment.Method)Electronic check
##                                   0.272995
##            I(Payment.Method)Mailed check
##                                  -0.119672
##           Multiple.LinesNo phone service
##                                   0.686948
##                           Multiple.LinesYes
##                                   0.256884
##       Online.SecurityNo internet service
##                                         NA
##                         Online.SecurityYes
##                                  -0.456454
##                              Total.Charges
##                                   0.000345
##                       Paperless.BillingYes
##                                   0.306784
##                                 PartnerYes
##                                   0.288897
##                       I(Senior.Citizen)Yes
##                                   0.171157
##         Online.BackupNo internet service
##                                         NA
##                           Online.BackupYes
##                                  -0.148385
##
## Degrees of Freedom: 4930 Total (i.e. Null);  4913 Residual
## Null Deviance:       5705
## Residual Deviance: 3939  AIC: 3975
```

```r
#named as AIC but k = log(n) makes it calculate BIC
null <- glm(churn_12month ~ 1,  data = train_best, family = "binomial")
step(null, scope = list(lower=null,upper=model_2),
     direction="both", criterion = "BIC", k = log(4931),trace = FALSE)
```

```
##
## Call:  glm(formula = churn_12month ~ I(Contract) + I(Internet.Service) +
##     Tenure + Dependents + Online.Security + Paperless.Billing +
##     Total.Charges + Partner + Multiple.Lines, family = "binomial",
##     data = train_best)
##
## Coefficients:
##                          (Intercept)                 I(Contract)One year
##                            -0.311502                          -0.770436
##                I(Contract)Two year      I(Internet.Service)Fiber optic
##                            -1.617275                           0.880678
##               I(Internet.Service)No                              Tenure
##                            -0.727095                          -0.061779
##                        DependentsYes  Online.SecurityNo internet service
##                            -1.620881                                 NA
##                   Online.SecurityYes               Paperless.BillingYes
##                            -0.493775                           0.341377
##                        Total.Charges                          PartnerYes
##                             0.000325                           0.306415
##     Multiple.LinesNo phone service                   Multiple.LinesYes
##                             0.721530                           0.280686
##
## Degrees of Freedom: 4930 Total (i.e. Null);  4918 Residual
## Null Deviance:       5705
## Residual Deviance: 3966  AIC: 3992
```

```
model_AIC <- glm(churn_12month ~ I(Contract) + I(Internet.Service) +
    Tenure + Dependents + Online.Security + Multiple.Lines +
    I(Payment.Method) + Paperless.Billing + Total.Charges + Partner,
    data = train_best, family = "binomial")

model_BIC <- glm(churn_12month ~ I(Contract) + I(Internet.Service) + Tenure +
    Dependents + Multiple.Lines + Paperless.Billing + Total.Charges +
    Online.Security + I(Payment.Method) + Partner, data = train_best, family = "binomial")
```

```
# what data set... this? need to check if this is on training or test data set
cutoff= 0.25
DF <- model.frame(model_AIC)
DF$prob <- predict(model_AIC, type = "response")
DF$flag <- ifelse(DF$prob > cutoff, 1, 0)
actual_va <- train_best$churn_12month
ta <- table(DF$flag, actual_va)
print(ta)
```

```
##    actual_va
##        0    1
##   0 2625  234
##   1  998 1074
```

```
sensitivity(ta)
```

```
## [1] 0.7245377
```

```r
specificity(ta)
```

```
## [1] 0.8211009
```

```r
# Creation of Confusion Matrix
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```r
err_metric=function(CM)
{
  TN =CM[1,1]
  TP =CM[2,2]
  FP =CM[1,2]
  FN =CM[2,1]
  precision =(TP)/(TP+FP)
  recall_score =(FP)/(FP+TN)
  f1_score=2*((precision*recall_score)/(precision+recall_score))
  accuracy_model  =(TP+TN)/(TP+TN+FP+FN)
  False_positive_rate =(FP)/(FP+TN)
  False_negative_rate =(FN)/(FN+TP)
  print(paste("Precision value of the model: ",round(precision,2)))
  print(paste("Accuracy of the model: ",round(accuracy_model,2)))
  print(paste("Recall value of the model: ",round(recall_score,2)))
  print(paste("False Positive rate of the model: ",round(False_positive_rate,2)))
  print(paste("False Negative rate of the model: ",round(False_negative_rate,2)))
  print(paste("f1 score of the model: ",round(f1_score,2)))
}
```

```r
# precision higher than 50
probability = c(0.3, 0.4, 0.5)
# ROC curve without probability cutoff
length(test_best$CustomerID)
```

```
## [1] 2112
```

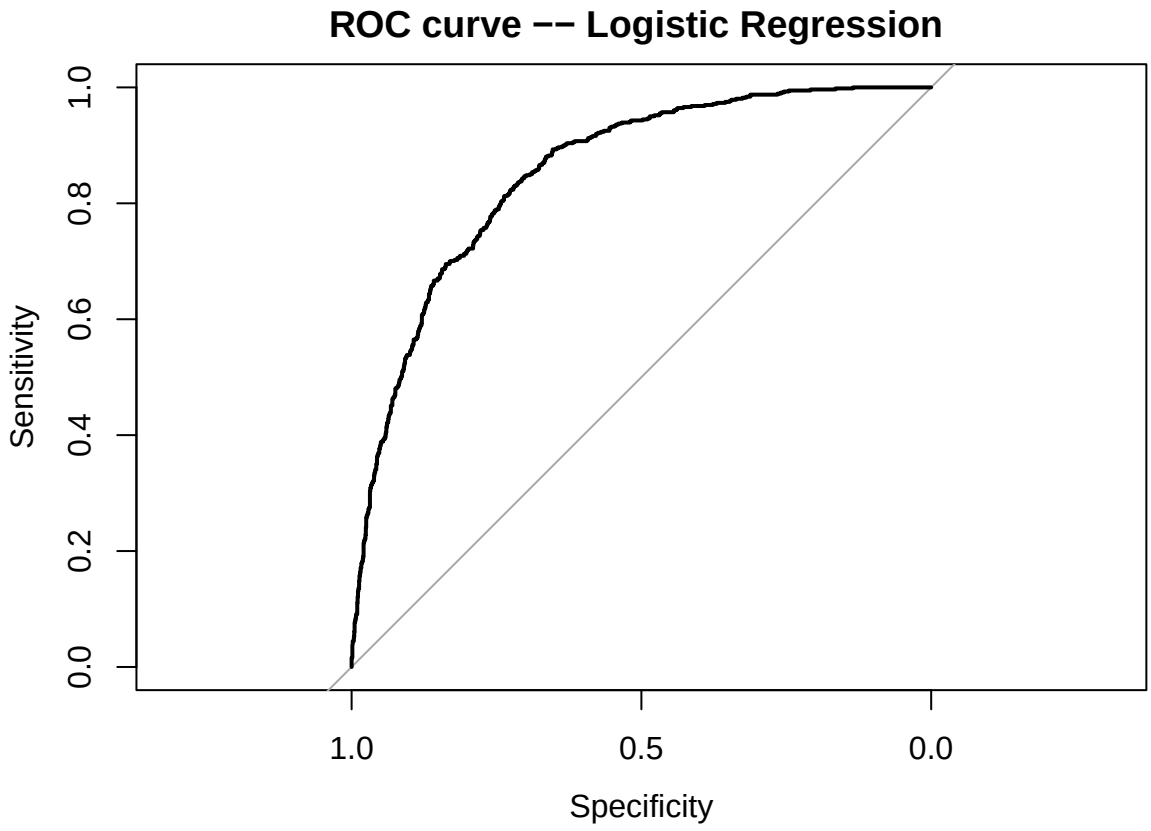
```r
pred_set <- test_best[-test_best$churn_12month]

logit_P = predict(model_AIC , newdata = test_best[-test_best$churn_12month] ,type = 'response' )
roc_score=roc(test_best$churn_12month, logit_P) #AUC score
```

```
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
```

```r
plot(roc_score ,main ="ROC curve -- Logistic Regression ")
```




```r
print(roc_score)
```

```
##
## Call:
## roc.default(response = test_best$churn_12month, predictor = logit_P)
##
## Data: logit_P in 1551 controls (test_best$churn_12month 0) < 561 cases (test_best$churn_12month 1).
## Area under the curve: 0.8525
```

```r
# ROC curve with cut off
for (i in 1:length(probability)){
   cut_off = probability[i]
   logit_P_cutoff <- ifelse(logit_P > cut_off, 1,0) # Probability check
   roc_curve <- roc(test_best$churn_12month, logit_P_cutoff)
   if(i==1) {

   plot(roc_curve,col=i)

  }else {
  lines(roc_curve,col=i) # Need to add legend here
  }
  cut_off = probability[i]
```

```
  logit_P_cutoff <- ifelse(logit_P > cut_off, 1,0) # Probability check
  CM= table(test_best$churn_12month, logit_P_cutoff)
  print(cut_off)
  roc_score_cutoff=roc(test_best$churn_12month, logit_P_cutoff) #AUC score
  print(roc_score_cutoff)
  err_metric(CM)
}
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
## [1] 0.3
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
##
## Call:
## roc.default(response = test_best$churn_12month, predictor = logit_P_cutoff)
##
## Data: logit_P_cutoff in 1551 controls (test_best$churn_12month 0) < 561 cases (test_best$churn_12mon
## Area under the curve: 0.7627
## [1] "Precision value of the model:  0.54"
## [1] "Accuracy of the model:  0.77"
## [1] "Recall value of the model:  0.23"
## [1] "False Positive rate of the model:  0.23"
## [1] "False Negative rate of the model:  0.24"
## [1] "f1 score of the model:  0.32"
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## [1] 0.4
```

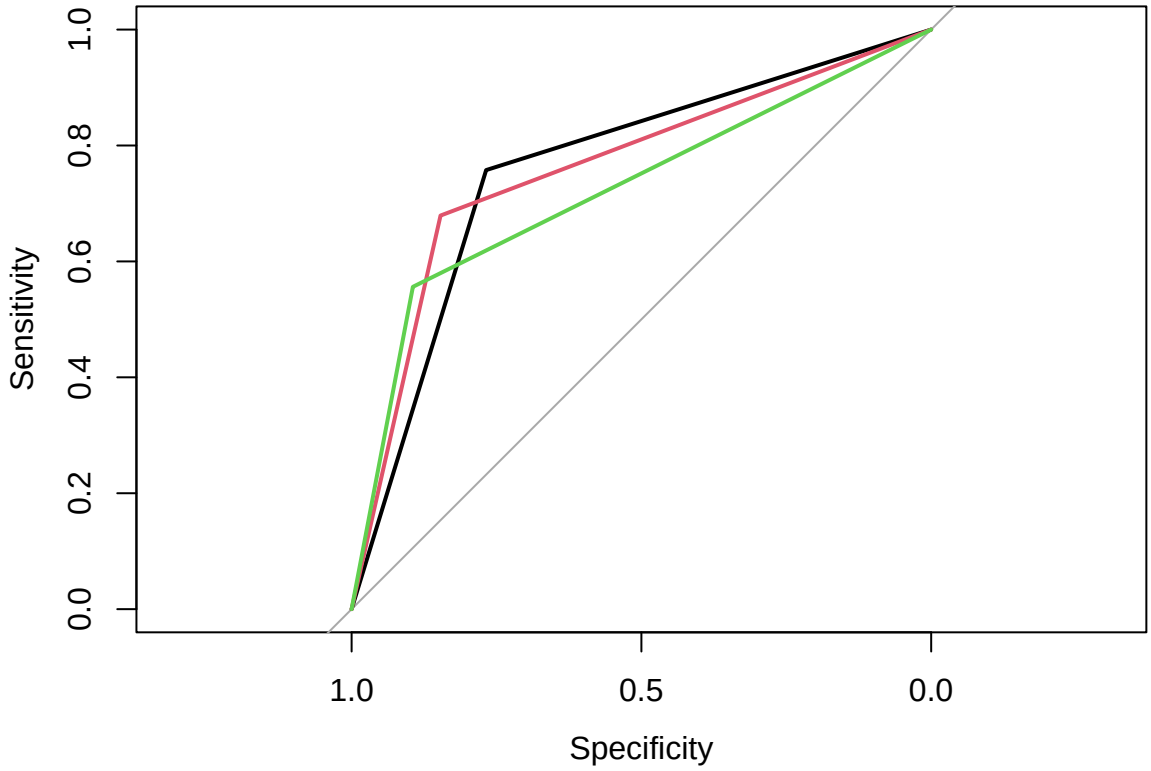
```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
##
## Call:
## roc.default(response = test_best$churn_12month, predictor = logit_P_cutoff)
##
## Data: logit_P_cutoff in 1551 controls (test_best$churn_12month 0) < 561 cases (test_best$churn_12mon
## Area under the curve: 0.7628
## [1] "Precision value of the model:  0.62"
## [1] "Accuracy of the model:  0.8"
## [1] "Recall value of the model:  0.15"
## [1] "False Positive rate of the model:  0.15"
## [1] "False Negative rate of the model:  0.32"
## [1] "f1 score of the model:  0.25"
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases

## [1] 0.5

## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```




```
##
## Call:
## roc.default(response = test_best$churn_12month, predictor = logit_P_cutoff)
##
## Data: logit_P_cutoff in 1551 controls (test_best$churn_12month 0) < 561 cases (test_best$churn_12mon
## Area under the curve: 0.7252
## [1] "Precision value of the model:  0.66"
## [1] "Accuracy of the model:  0.8"
## [1] "Recall value of the model:  0.11"
## [1] "False Positive rate of the model:  0.11"
## [1] "False Negative rate of the model:  0.44"
## [1] "f1 score of the model:  0.18"
```

```
# Bootstrap and CI for AUC
# Get test sample with replacement, calculate AUC, and do bootstrap?

# Need to ask why and how about this part
```