Churn_Analysis_worst_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
worst_case <- worst_case[-worst_case$Churn_val]
worst_case <- worst_case[-worst_case$Churn_val]

dim(best_case)
```

```
## [1] 7043   32
```

```
dim(worst_case)
```

```
## [1] 7043   30
```

```
#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 <- worst_case[worst_case$churn_12month == 1,]
not_churned <- worst_case[worst_case$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(worst_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] 12.60479 13.82869
## attr(,"conf.level")
## [1] 0.95
```

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

```
## [1] 45.49224 46.68417
## 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] -32.37647 -30.79312
## 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] 1002.750 1121.133
## attr(,"conf.level")
## [1] 0.95
```

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

```
## [1] 3082.869 3220.797
## 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] -2046.505 -1874.720
## attr(,"conf.level")
## [1] 0.95
```

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

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

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

```
##
##       No No phone service  Yes
##   0 1736              407 1961
##   1 1654              275 1010
```

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

```
##
##      DSL Fiber optic   No
##   0 1543        1538 1023
##   1  878        1558  503
```

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

```
##
##       No No internet service  Yes
##   0 1523                1023 1558
##   1 1975                 503  461
```

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

```
##
##       No No internet service  Yes
##   0 1337                1023 1744
##   1 1751                 503  685
```

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

```
##
##       No No internet service  Yes
##   0 1337                1023 1744
##   1 1758                 503  678
```

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

```
##
##       No No internet service  Yes
##   0 1348                1023 1733
##   1 1462                 503  974
```

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

```
##
##       No No internet service  Yes
##   0 1335                1023 1746
##   1 1450                 503  986
```

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

```
##
##     Bank transfer (automatic) Credit card (automatic) Electronic check
##   0                      1178                    1160              949
##   1                       366                     362             1416
##
##     Mailed check
##   0          817
##   1          795
```

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

```
##
##       No  Yes
##   0 1838 2266
##   1 1034 1905
```

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

```
##
##     Month-to-month One year Two year
##   0           1303     1213     1588
##   1           2572      260      107
```

```r
# just interested
age_paperless <- table(worst_case$Senior.Citizen, worst_case$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(worst_case$churn_12month, p = .7, list = FALSE)
train_worst <- worst_case[ train.index,]
#churned_train <- train_worst[train_worst$churn_12month == 1,]
#print(length(churned_train$churn_12month)/length(train_worst$churn_12month))

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

#churned <- worst_case[worst_case$churn_12month == 1,]
#print(length(churned$churn_12month)/length(worst_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_worst)
##
## Coefficients:
##                                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)                          2.2754886  0.1504502  15.125  < 2e-16
## I(Senior.Citizen)Yes                 0.2985376  0.1186951   2.515 0.011898
## Tenure                              -0.1979369  0.0093317 -21.211  < 2e-16
## I(Internet.Service)Fiber optic      -0.0653684  0.1297459  -0.504 0.614389
## I(Internet.Service)No                0.3681613  0.1470931   2.503 0.012318
## I(Contract)One year                 -0.8452798  0.1201093  -7.038 1.96e-12
## I(Contract)Two year                 -1.2534351  0.1759794  -7.123 1.06e-12
## Total.Charges                        0.0013202  0.0000925  14.273  < 2e-16
## I(Payment.Method)Credit card (automatic)  0.0200091  0.1442813   0.139 0.889702
## I(Payment.Method)Electronic check    0.4220211  0.1241382   3.400 0.000675
## I(Payment.Method)Mailed check        0.1867978  0.1445087   1.293 0.196135
##
## (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: 6703.1  on 4930  degrees of freedom
## Residual deviance: 3368.8  on 4920  degrees of freedom
## AIC: 3390.8
##
## Number of Fisher Scoring iterations: 7
```

```
# 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=3390.77
## churn_12month ~ I(Senior.Citizen) + Tenure + I(Internet.Service) +
##     I(Contract) + Total.Charges + I(Payment.Method)
##
##                        Df Deviance    AIC
## <none>                      3368.8 3390.8
## - I(Internet.Service)  2   3375.9 3393.9
## - I(Senior.Citizen)    1   3375.1 3395.1
## - I(Payment.Method)    3   3384.8 3400.8
## - I(Contract)          2   3447.2 3465.2
## - Total.Charges        1   3613.6 3633.6
## - Tenure               1   4153.1 4173.1
```

```
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   3368.766 3390.766
```

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_worst)
##
## Coefficients: (7 not defined because of singularities)
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             3.6051243  0.2632022  13.697  < 2e-16
## I(Senior.Citizen)Yes                    0.0809988  0.1209761   0.670 0.503148
## Tenure                                 -0.2175140  0.0100357 -21.674  < 2e-16
## I(Internet.Service)Fiber optic         -0.1260234  0.1444974  -0.872 0.383127
## I(Internet.Service)No                   0.5908078  0.1740565   3.394 0.000688
## I(Contract)One year                    -0.8484285  0.1252875  -6.772 1.27e-11
## I(Contract)Two year                    -1.1017607  0.1818146  -6.060 1.36e-09
## Total.Charges                           0.0015484  0.0001014  15.268  < 2e-16
## I(Payment.Method)Credit card (automatic) -0.0106374  0.1479148  -0.072 0.942669
## I(Payment.Method)Electronic check       0.3454852  0.1274679   2.710 0.006721
## I(Payment.Method)Mailed check           0.1614995  0.1500810   1.076 0.281891
## GenderMale                              0.0455544  0.0893091   0.510 0.609998
## PartnerYes                              0.2778741  0.1002935   2.771 0.005595
## DependentsYes                          -1.1200719  0.1266221  -8.846  < 2e-16
## Phone.ServiceYes                       -1.1316339  0.2065215  -5.479 4.27e-08
## Multiple.LinesNo phone service                 NA         NA      NA       NA
## Multiple.LinesYes                      -0.1444850  0.1097857  -1.316 0.188153
## Internet.ServiceFiber optic                    NA         NA      NA       NA
## Internet.ServiceNo                             NA         NA      NA       NA
## Online.SecurityNo internet service             NA         NA      NA       NA
## Online.SecurityYes                     -0.4816981  0.1133618  -4.249 2.15e-05
## Online.BackupNo internet service               NA         NA      NA       NA
## Online.BackupYes                       -0.1107502  0.1076969  -1.028 0.303785
## ContractOne year                               NA         NA      NA       NA
## ContractTwo year                               NA         NA      NA       NA
## Paperless.BillingYes                    0.1590552  0.1023492   1.554 0.120174
##
## (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: 6703.1  on 4930  degrees of freedom
## Residual deviance: 3222.9  on 4912  degrees of freedom
## AIC: 3260.9
##
## Number of Fisher Scoring iterations: 7
```

```
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 ~ Tenure + I(Contract) + Total.Charges + I(Payment.Method) +
##     Partner + Dependents + Multiple.Lines + Online.Security +
##     Paperless.Billing
##
##
##                     Step Df  Deviance Resid. Df Resid. Dev      AIC
## 1                                         4912   3222.927 3260.927
## 2            - Contract  0 0.0000000      4912   3222.927 3260.927
## 3    - Internet.Service  0 0.0000000      4912   3222.927 3260.927
## 4       - Phone.Service  0 0.0000000      4912   3222.927 3260.927
## 5             - Gender  1 0.2602003      4913   3223.187 3259.187
## 6   - I(Senior.Citizen)  1 0.4655466      4914   3223.653 3257.653
## 7 - I(Internet.Service)  1 0.6699494      4915   3224.323 3256.323
## 8       - Online.Backup  1 0.9752785      4916   3225.298 3255.298
```

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

```
##
```

```
## Call:  glm(formula = churn_12month ~ Tenure + Total.Charges + Dependents +
##     I(Contract) + Online.Security + Phone.Service + Partner +
##     I(Payment.Method) + Paperless.Billing, family = "binomial",
##     data = train_worst)
##
## Coefficients:
##                                 (Intercept)
##                                    3.567334
##                                      Tenure
##                                   -0.213356
##                                Total.Charges
##                                     0.001477
##                                DependentsYes
##                                    -1.126183
##                          I(Contract)One year
##                                    -0.828717
##                          I(Contract)Two year
##                                    -1.082496
##       Online.SecurityNo internet service
##                                     0.661985
##                           Online.SecurityYes
##                                    -0.460441
##                             Phone.ServiceYes
##                                    -1.203281
##                                 PartnerYes
##                                     0.279338
## I(Payment.Method)Credit card (automatic)
##                                    -0.008198
##        I(Payment.Method)Electronic check
##                                     0.342256
##             I(Payment.Method)Mailed check
##                                     0.179462
##                         Paperless.BillingYes
##                                     0.143375
##
## Degrees of Freedom: 4930 Total (i.e. Null);  4917 Residual
## Null Deviance:       6703
## Residual Deviance: 3227  AIC: 3255
```

```r
#named as AIC but k = log(n) makes it calculate BIC
null <- glm(churn_12month ~ 1,  data = train_worst, 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 ~ Tenure + Total.Charges + Dependents +
##     I(Contract) + Phone.Service + Online.Security, family = "binomial",
##     data = train_worst)
##
## Coefficients:
##                          (Intercept)                                Tenure
##                             3.953279                             -0.215365
##                         Total.Charges                         DependentsYes
##                             0.001517                             -1.028242
```

```
##                I(Contract)One year                I(Contract)Two year
##                          -0.879270                          -1.172150
##                   Phone.ServiceYes  Online.SecurityNo internet service
##                          -1.209431                           0.574539
##                 Online.SecurityYes
##                          -0.501265
##
## Degrees of Freedom: 4930 Total (i.e. Null);  4922 Residual
## Null Deviance:       6703
## Residual Deviance: 3249  AIC: 3267
```

```
model_AIC <- glm(churn_12month ~ I(Contract) + Phone.Service +
    Tenure + Dependents + Online.Security + Senior.Citizen +
    I(Payment.Method) + Paperless.Billing + Total.Charges + Partner,
    data = train_worst, family = "binomial")

model_BIC <- glm(churn_12month ~ I(Contract) + Phone.Service + Tenure +
    Dependents + Paperless.Billing + Total.Charges +
    Online.Security, data = train_worst, 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$prob > cutoff, 1, 0)
actual_va <- train_worst$churn_12month
ta <- table(DF$flag, actual_va)
print(ta)
```

```
##    actual_va
##        0    1
##   0 2122  180
##   1  747 1882
```

```
sensitivity(ta)
```

```
## [1] 0.7396305
```

```
specificity(ta)
```

```
## [1] 0.9127061
```

```
# 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
```

```
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)))
}
```

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

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

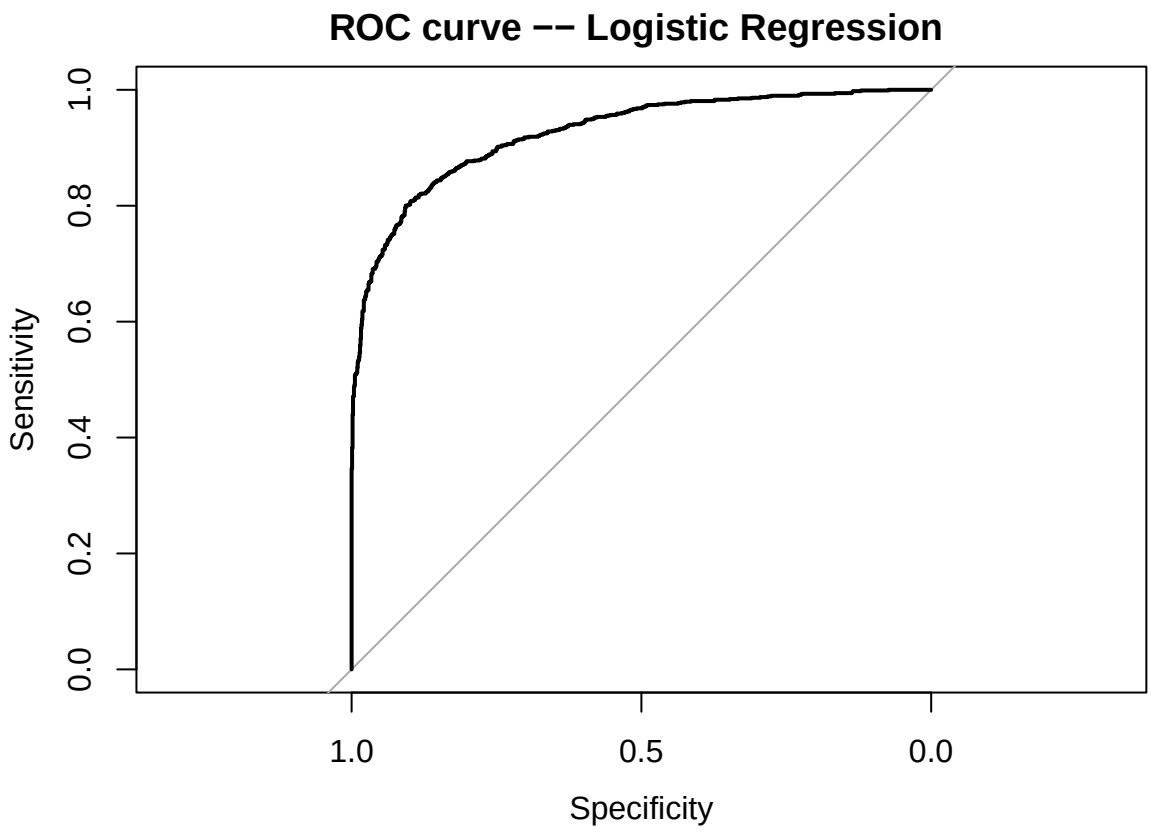
```
pred_set <- test_worst[-test_worst$churn_12month]

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

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

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

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




```
print(roc_score)
```

```
##
## Call:
## roc.default(response = test_worst$churn_12month, predictor = logit_P)
##
## Data: logit_P in 1235 controls (test_worst$churn_12month 0) < 877 cases (test_worst$churn_12month 1)
## Area under the curve: 0.9255
```

```
# 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_worst$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_worst$churn_12month, logit_P_cutoff)
  print(cut_off)
```

```
  roc_score_cutoff=roc(test_worst$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_worst$churn_12month, predictor = logit_P_cutoff)
##
## Data: logit_P_cutoff in 1235 controls (test_worst$churn_12month 0) < 877 cases (test_worst$churn_12m
## Area under the curve: 0.8261
## [1] "Precision value of the model:  0.73"
## [1] "Accuracy of the model:  0.82"
## [1] "Recall value of the model:  0.23"
## [1] "False Positive rate of the model:  0.23"
## [1] "False Negative rate of the model:  0.12"
## [1] "f1 score of the model:  0.35"
```

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

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

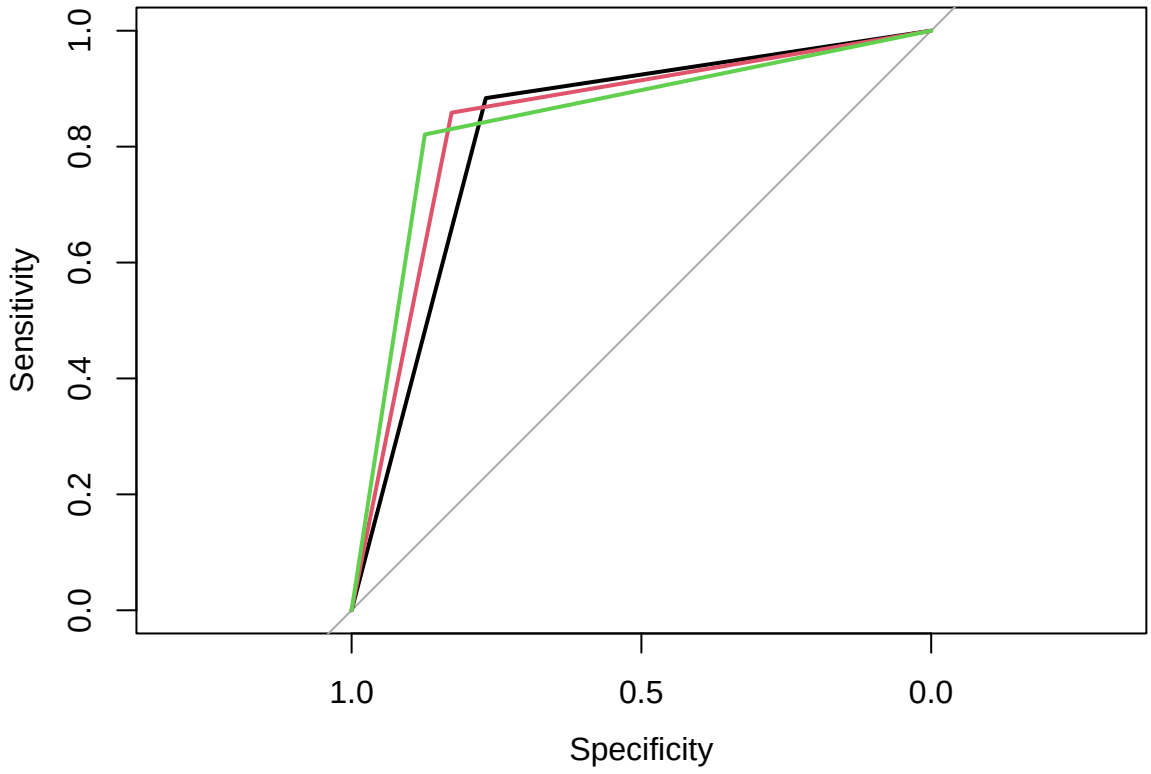
```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
##
## Call:
## roc.default(response = test_worst$churn_12month, predictor = logit_P_cutoff)
##
## Data: logit_P_cutoff in 1235 controls (test_worst$churn_12month 0) < 877 cases (test_worst$churn_12m
## Area under the curve: 0.8431
## [1] "Precision value of the model:  0.78"
## [1] "Accuracy of the model:  0.84"
## [1] "Recall value of the model:  0.17"
## [1] "False Positive rate of the model:  0.17"
## [1] "False Negative rate of the model:  0.14"
## [1] "f1 score of the model:  0.28"
```

```
## 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_worst$churn_12month, predictor = logit_P_cutoff)
##
## Data: logit_P_cutoff in 1235 controls (test_worst$churn_12month 0) < 877 cases (test_worst$churn_12m
## Area under the curve: 0.8473
## [1] "Precision value of the model:  0.82"
## [1] "Accuracy of the model:  0.85"
## [1] "Recall value of the model:  0.13"
## [1] "False Positive rate of the model:  0.13"
## [1] "False Negative rate of the model:  0.18"
## [1] "f1 score of the model:  0.22"
```

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

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