Analyst Price Target Analysis using XGBoost model

```
#Drop Cols that will not be used in model
finaldf <- read.csv("AnalystRatingData/ratings.csv", check.names = TRUE)
drops <- c("Date","Firm","Action","Rating","Ticker","", "Price.After.One.Year", "Price.T
arget")
finaldf <- finaldf[ , !(names(finaldf) %in% drops)]
```

```
#convert everything to numeric as XGBoost only recognize numeric data
finaldf[names(finaldf)] <- lapply(finaldf[names(finaldf)],as.numeric)
```

```
nzv <- nearZeroVar(finaldf,saveMetrics = TRUE)
sorted <-nzv[order(nzv$freqRatio),]
print(sorted)
```

```
##                                      freqRatio percentUnique zeroVar
## Reiterated.Rating                     1.081721    0.03792907   FALSE
## TargetAchieved                        1.348775    0.03792907   FALSE
## Buy                                   1.465171    0.03792907   FALSE
## Lower.Price.Target                    4.813671    0.03792907   FALSE
## Boost.Price.Target                    4.845898    0.03792907   FALSE
## Outperform                            5.383777    0.03792907   FALSE
## Neutral                               8.466786    0.03792907   FALSE
## Hold                                 10.929864    0.03792907   FALSE
## Overweight                           11.584726    0.03792907   FALSE
## Initiated.Coverage                   12.731771    0.03792907   FALSE
## Royal.Bank.Of.Canada                 17.566901    0.03792907   FALSE
## Deutsche.Bank.AG                     18.675373    0.03792907   FALSE
## Upgrade                              18.749064    0.03792907   FALSE
## Jefferies.Group.LLC                  20.610656    0.03792907   FALSE
## Piper.Jaffray.Companies              21.438298    0.03792907   FALSE
## Downgrade                            22.026201    0.03792907   FALSE
## Barclays.PLC                         24.597087    0.03792907   FALSE
## Credit.Suisse.Group                  26.041026    0.03792907   FALSE
## B..Riley                             27.502703    0.03792907   FALSE
## Cowen.and.Company                    27.657609    0.03792907   FALSE
## Morgan.Stanley                       28.294444    0.03792907   FALSE
## X                                    28.960227    0.03792907   FALSE
## Goldman.Sachs.Group..Inc...The.      31.549383    0.03792907   FALSE
## Market.Perform                       31.956250    0.03792907   FALSE
## J.P.Morgan.Chase...Co                32.163522    0.03792907   FALSE
## Citigroup.Inc.                       37.772059    0.03792907   FALSE
## Equal.Weight                         38.059259    0.03792907   FALSE
## Robert.W..Baird                      39.561538    0.03792907   FALSE
## Canaccord.Genuity                    42.221311    0.03792907   FALSE
## Nomura                               45.663717    0.03792907   FALSE
## Stifel.Nicolaus                      46.080357    0.03792907   FALSE
## Oppenheimer.Holdings..Inc.           46.504505    0.03792907   FALSE
## Bank.of.America.Corporation          46.936364    0.03792907   FALSE
## BMO.Capital.Markets                  48.745283    0.03792907   FALSE
## Susquehanna.Bancshares.Inc           51.730000    0.03792907   FALSE
## Wedbush                              52.806122    0.03792907   FALSE
## Mizuho                               53.360825    0.03792907   FALSE
## Sell                                 55.095745    0.03792907   FALSE
## Leerink.Swann                        56.945055    0.03792907   FALSE
## MKM.Partners                         57.588889    0.03792907   FALSE
## Pacific.Crest                        58.247191    0.03792907   FALSE
## Sector.Perform                       58.247191    0.03792907   FALSE
## FBR...Co                             58.920455    0.03792907   FALSE
## Macquarie                            68.381579    0.03792907   FALSE
## Cantor.Fitzgerald                    70.256757    0.03792907   FALSE
## Brean.Capital                        72.236111    0.03792907   FALSE
## Sanford.C..Bernstein                 73.267606    0.03792907   FALSE
## Evercore.ISI                         75.420290    0.03792907   FALSE
## SunTrust.Banks..Inc.                 75.420290    0.03792907   FALSE
## Sterne.Agee.CRT                      82.698413    0.03792907   FALSE
## Topeka.Capital.Markets               85.442623    0.03792907   FALSE
## Positive                             85.442623    0.03792907   FALSE
```

```
## Argus                              91.508772   0.03792907   FALSE
## Needham...Company.LLC              91.508772   0.03792907   FALSE
## Set.Price.Target                  102.392157   0.03792907   FALSE
## Raymond.James.Financial..Inc.     111.191489   0.03792907   FALSE
## Underperform                      113.630435   0.03792907   FALSE
## BTIG.Research                     116.177778   0.03792907   FALSE
## Pivotal.Research                  124.547619   0.03792907   FALSE
## Underweight                       130.825000   0.03792907   FALSE
## Strong.Buy                        145.472222   0.03792907   FALSE
## Axiom.Securities                  174.766667   0.03792907   FALSE
## JMP.Securities                    174.766667   0.03792907   FALSE
## Telsey.Advisory.Group             174.766667   0.03792907   FALSE
## Summit.Research                   201.807692   0.03792907   FALSE
## Maxim.Group                       218.708333   0.03792907   FALSE
## Wunderlich                        228.260870   0.03792907   FALSE
## Benchmark.Co.                     238.681818   0.03792907   FALSE
## Drexel.Hamilton                   238.681818   0.03792907   FALSE
## Rosenblatt.Securities             262.650000   0.03792907   FALSE
## Buckingham.Research               276.526316   0.03792907   FALSE
## Craig.Hallum                      276.526316   0.03792907   FALSE
## S.P.Equity.Research               291.944444   0.03792907   FALSE
## Global.Equities.Research          328.562500   0.03792907   FALSE
## KeyCorp                           350.533333   0.03792907   FALSE
## Roth.Capital                      375.642857   0.03792907   FALSE
## Moffett.Nathanson                 404.615385   0.03792907   FALSE
## Standpoint.Research               404.615385   0.03792907   FALSE
## Wells.Fargo...Co                  404.615385   0.03792907   FALSE
## Market.Outperform                 404.615385   0.03792907   FALSE
## Monness.Crespi...Hardt            438.416667   0.03792907   FALSE
## Top.Pick                          438.416667   0.03792907   FALSE
## Chardan.Capital                   478.363636   0.03792907   FALSE
## FBN.Securities                    478.363636   0.03792907   FALSE
## HC.Wainwright                     478.363636   0.03792907   FALSE
## Conviction.Buy                    526.300000   0.03792907   FALSE
## Atlantic.Securities               584.888889   0.03792907   FALSE
## BNP.Paribas                       584.888889   0.03792907   FALSE
## Stephens                          584.888889   0.03792907   FALSE
## Avondale.Partners                 658.125000   0.03792907   FALSE
## Gabelli                           658.125000   0.03792907   FALSE
## William.Blair                     658.125000   0.03792907   FALSE
## Compass.Point                     752.285714   0.03792907   FALSE
## Dougherty...Co                    752.285714   0.03792907   FALSE
## HSBC.Holdings.plc                 752.285714   0.03792907   FALSE
## Reduce                            752.285714   0.03792907   FALSE
## Barrington.Research               877.833333   0.03792907   FALSE
## Wolfe.Research                    877.833333   0.03792907   FALSE
## Berenberg.Bank                   1053.600000   0.03792907   FALSE
## Janney.Montgomery.Scott          1053.600000   0.03792907   FALSE
## Northland.Securities             1053.600000   0.03792907   FALSE
## Societe.Generale                 1053.600000   0.03792907   FALSE
## Ascendiant.Capital.Markets       1317.250000   0.03792907   FALSE
## BB.T.Corporation                 1317.250000   0.03792907   FALSE
## DA.Davidson                      1317.250000   0.03792907   FALSE
## Guggenheim                       1317.250000   0.03792907   FALSE
```

```
## Jyske.Bank                             1317.250000    0.03792907   FALSE
## Keefe..Bruyette...Woods                1317.250000    0.03792907   FALSE
## Longbow.Research                       1317.250000    0.03792907   FALSE
## Scotiabank                             1317.250000    0.03792907   FALSE
## Independent.Research.GmbH              1756.666667    0.03792907   FALSE
## UBS.Group.AG                           1756.666667    0.03792907   FALSE
## In.Line                                1756.666667    0.03792907   FALSE
## Mkt.Perform                            1756.666667    0.03792907   FALSE
## Boenning.Scattergood                   2635.500000    0.03792907   FALSE
## Evercore.Partners.Inc                  2635.500000    0.03792907   FALSE
## Hilliard.Lyons                         2635.500000    0.03792907   FALSE
## Imperial.Capital                       2635.500000    0.03792907   FALSE
## T.H..Capital                           2635.500000    0.03792907   FALSE
## BGC.Financial                          5272.000000    0.03792907   FALSE
## CRT.Capital                            5272.000000    0.03792907   FALSE
## Feltl...Co.                            5272.000000    0.03792907   FALSE
## First.Analysis                         5272.000000    0.03792907   FALSE
## Hovde.Group                            5272.000000    0.03792907   FALSE
## Mitsubishi.UFJ.Financial.Group         5272.000000    0.03792907   FALSE
## MLV...Co.                              5272.000000    0.03792907   FALSE
## National.Bank.Financial                5272.000000    0.03792907   FALSE
## Sidoti                                 5272.000000    0.03792907   FALSE
## Taglich.Brothers                       5272.000000    0.03792907   FALSE
## Williams.Capital                       5272.000000    0.03792907   FALSE
## But.Estimates.Decline.As.Orders.Slow   5272.000000    0.03792907   FALSE
## Focus.List                             5272.000000    0.03792907   FALSE
## Overweight.Rating.                     5272.000000    0.03792907   FALSE
## Sector.Outperform                      5272.000000    0.03792907   FALSE
## Weight                                 5272.000000    0.03792907   FALSE
##                                          nzv
## Reiterated.Rating                      FALSE
## TargetAchieved                         FALSE
## Buy                                    FALSE
## Lower.Price.Target                     FALSE
## Boost.Price.Target                     FALSE
## Outperform                             FALSE
## Neutral                                FALSE
## Hold                                   FALSE
## Overweight                             FALSE
## Initiated.Coverage                     FALSE
## Royal.Bank.Of.Canada                   FALSE
## Deutsche.Bank.AG                       FALSE
## Upgrade                                FALSE
## Jefferies.Group.LLC                     TRUE
## Piper.Jaffray.Companies                 TRUE
## Downgrade                               TRUE
## Barclays.PLC                            TRUE
## Credit.Suisse.Group                     TRUE
## B..Riley                                TRUE
## Cowen.and.Company                       TRUE
## Morgan.Stanley                          TRUE
## X                                       TRUE
## Goldman.Sachs.Group..Inc...The.         TRUE
## Market.Perform                          TRUE
```

```
## J.P.Morgan.Chase...Co                   TRUE
## Citigroup.Inc.                          TRUE
## Equal.Weight                            TRUE
## Robert.W..Baird                         TRUE
## Canaccord.Genuity                       TRUE
## Nomura                                  TRUE
## Stifel.Nicolaus                         TRUE
## Oppenheimer.Holdings..Inc.              TRUE
## Bank.of.America.Corporation             TRUE
## BMO.Capital.Markets                     TRUE
## Susquehanna.Bancshares.Inc              TRUE
## Wedbush                                 TRUE
## Mizuho                                  TRUE
## Sell                                    TRUE
## Leerink.Swann                           TRUE
## MKM.Partners                            TRUE
## Pacific.Crest                           TRUE
## Sector.Perform                          TRUE
## FBR...Co                                TRUE
## Macquarie                               TRUE
## Cantor.Fitzgerald                       TRUE
## Brean.Capital                           TRUE
## Sanford.C..Bernstein                    TRUE
## Evercore.ISI                            TRUE
## SunTrust.Banks..Inc.                    TRUE
## Sterne.Agee.CRT                         TRUE
## Topeka.Capital.Markets                  TRUE
## Positive                                TRUE
## Argus                                   TRUE
## Needham...Company.LLC                   TRUE
## Set.Price.Target                        TRUE
## Raymond.James.Financial..Inc.           TRUE
## Underperform                            TRUE
## BTIG.Research                           TRUE
## Pivotal.Research                        TRUE
## Underweight                             TRUE
## Strong.Buy                              TRUE
## Axiom.Securities                        TRUE
## JMP.Securities                          TRUE
## Telsey.Advisory.Group                   TRUE
## Summit.Research                         TRUE
## Maxim.Group                             TRUE
## Wunderlich                              TRUE
## Benchmark.Co.                           TRUE
## Drexel.Hamilton                         TRUE
## Rosenblatt.Securities                   TRUE
## Buckingham.Research                     TRUE
## Craig.Hallum                            TRUE
## S.P.Equity.Research                     TRUE
## Global.Equities.Research                TRUE
## KeyCorp                                 TRUE
## Roth.Capital                            TRUE
## Moffett.Nathanson                       TRUE
## Standpoint.Research                     TRUE
```

```
## Wells.Fargo...Co                       TRUE
## Market.Outperform                      TRUE
## Monness.Crespi...Hardt                 TRUE
## Top.Pick                               TRUE
## Chardan.Capital                        TRUE
## FBN.Securities                         TRUE
## HC.Wainwright                          TRUE
## Conviction.Buy                         TRUE
## Atlantic.Securities                    TRUE
## BNP.Paribas                            TRUE
## Stephens                               TRUE
## Avondale.Partners                      TRUE
## Gabelli                                TRUE
## William.Blair                          TRUE
## Compass.Point                          TRUE
## Dougherty...Co                         TRUE
## HSBC.Holdings.plc                      TRUE
## Reduce                                 TRUE
## Barrington.Research                    TRUE
## Wolfe.Research                         TRUE
## Berenberg.Bank                         TRUE
## Janney.Montgomery.Scott                TRUE
## Northland.Securities                   TRUE
## Societe.Generale                       TRUE
## Ascendiant.Capital.Markets             TRUE
## BB.T.Corporation                       TRUE
## DA.Davidson                            TRUE
## Guggenheim                             TRUE
## Jyske.Bank                             TRUE
## Keefe..Bruyette...Woods                TRUE
## Longbow.Research                       TRUE
## Scotiabank                             TRUE
## Independent.Research.GmbH              TRUE
## UBS.Group.AG                           TRUE
## In.Line                                TRUE
## Mkt.Perform                            TRUE
## Boenning.Scattergood                   TRUE
## Evercore.Partners.Inc                  TRUE
## Hilliard.Lyons                         TRUE
## Imperial.Capital                       TRUE
## T.H..Capital                           TRUE
## BGC.Financial                          TRUE
## CRT.Capital                            TRUE
## Feltl...Co.                            TRUE
## First.Analysis                         TRUE
## Hovde.Group                            TRUE
## Mitsubishi.UFJ.Financial.Group         TRUE
## MLV...Co.                              TRUE
## National.Bank.Financial                TRUE
## Sidoti                                 TRUE
## Taglich.Brothers                       TRUE
## Williams.Capital                       TRUE
## But.Estimates.Decline.As.Orders.Slow   TRUE
## Focus.List                             TRUE
```

```
## Overweight.Rating.                    TRUE
## Sector.Outperform                     TRUE
## Weight                                TRUE
```

```
#remove features that with high freq ratio (which appears less frequently.). freqRatio o
f 700 means it appears once in 700 times. This
#helped us to remove some of analyst/banks that are less known and
remove_features <- rownames(nzv[nzv$freqRatio > 700, ])
#this reduced number of columns from 137 to 94
finaldf <-finaldf[,!names(finaldf)%in% remove_features]
```

```
set.seed(410)
#make.name will replace space with dot in column names
names(finaldf)<-make.names(names(finaldf),unique=TRUE)
finaldf<-x<-na.omit(finaldf)
inTrain <- createDataPartition(y=finaldf$"TargetAchieved",p = 0.7, list=FALSE)
training <- finaldf[inTrain,]
testing <- finaldf[-inTrain,]
output_vector_training <- training$TargetAchieved
output_vector_testing <- testing$TargetAchieved

#drop independent variable form training and testing set as xgboost will understand it a
s a feature
training <- training[ , !names(training) %in% c("TargetAchieved")]
testing <- testing[ , !names(testing) %in% c("TargetAchieved")]
```

```
dtrain <-xgb.DMatrix(as.matrix(training), label = output_vector_training)
dtest <- xgb.DMatrix(as.matrix(testing), label = output_vector_testing)
#just to check if we have either 0 or 1
sumwpos <- sum(output_vector_training == 1)
sumwneg <- sum(output_vector_training == 0)
nrow(training) == sumwpos + sumwneg
```

```
## [1] TRUE
```

```r
xgb_params_1 = list(
  objective = "binary:logistic", # binary classification
  eta = 0.1, # learning rate
  max.depth = 4, # maxtree depth
  eval_metric = "auc", # evaluation metric
  scale_pos_weight = sumwneg / sumwpos
)


system.time(
bst <- xgboost(data = dtrain,
params = xgb_params_1,
nrounds = 250,
verbose = 1,
print_every_n = 10,
early_stopping_rounds = 50) # stop if no improvement within n trees)
)
```

```
## [1]  train-auc:0.614044
## Will train until train_auc hasn't improved in 50 rounds.
##
## [11] train-auc:0.646063
## [21] train-auc:0.678651
## [31] train-auc:0.687306
## [41] train-auc:0.697653
## [51] train-auc:0.701955
## [61] train-auc:0.706757
## [71] train-auc:0.712065
## [81] train-auc:0.716154
## [91] train-auc:0.721212
## [101]    train-auc:0.724315
## [111]    train-auc:0.726147
## [121]    train-auc:0.728632
## [131]    train-auc:0.730410
## [141]    train-auc:0.731798
## [151]    train-auc:0.733245
## [161]    train-auc:0.734244
## [171]    train-auc:0.734893
## [181]    train-auc:0.735968
## [191]    train-auc:0.736963
## [201]    train-auc:0.737952
## [211]    train-auc:0.738404
## [221]    train-auc:0.738973
## [231]    train-auc:0.739307
## [241]    train-auc:0.739862
## [250]    train-auc:0.740617
```

```
##    user  system elapsed
##   2.550   0.067   2.764
```

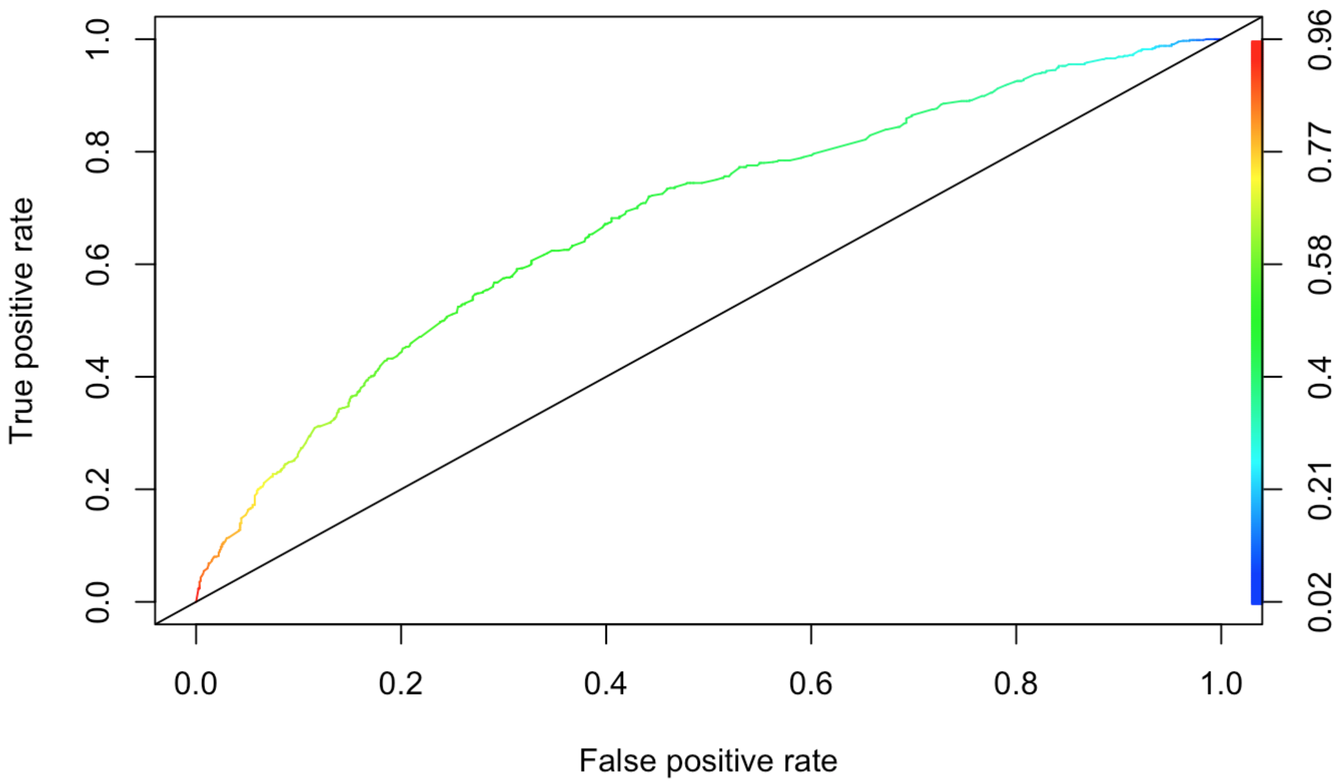
```
xgb.probs <- predict(bst, as.matrix(testing))
length(xgb.probs) == nrow(testing) # verifiy that this is equivalent to nrow(testing)
```

```
## [1] TRUE
```

```
#this creates a list with all 0 with same size as testing
xgb.predict <- rep( 0 ,nrow(testing))


#In our case having higher precission is more important as getting False Positive rate i
s costly
#(When you predict True but its actually false thus you loose money! or cant get your gu
ranteed
#return by following analyst) and False Negative is not important because what you predi
cted false
#and it is actually true. (ignored that rating and that rating was correct)

#Check the threshold graph to see relation between precission and recall in our case

xgb.pred <- as.numeric(xgb.probs > 0.65)
confusionMatrix(xgb.pred, output_vector_testing, positive="1", mode="prec_recall")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 811 483
##          1  97 190
##
##                Accuracy : 0.6331
##                  95% CI : (0.6088, 0.657)
##     No Information Rate : 0.5743
##     P-Value [Acc > NIR] : 1.082e-06
##
##                   Kappa : 0.1896
##  Mcnemar's Test P-Value : < 2.2e-16
##
##               Precision : 0.6620
##                  Recall : 0.2823
##                      F1 : 0.3958
##              Prevalence : 0.4257
##          Detection Rate : 0.1202
##    Detection Prevalence : 0.1815
##       Balanced Accuracy : 0.5877
##
##        'Positive' Class : 1
##
```

```
pred<- prediction(xgb.probs, output_vector_testing)
perf <- performance(pred,"tpr","fpr")
plot(perf, colorize=TRUE)
abline(a=0, b=1)
```



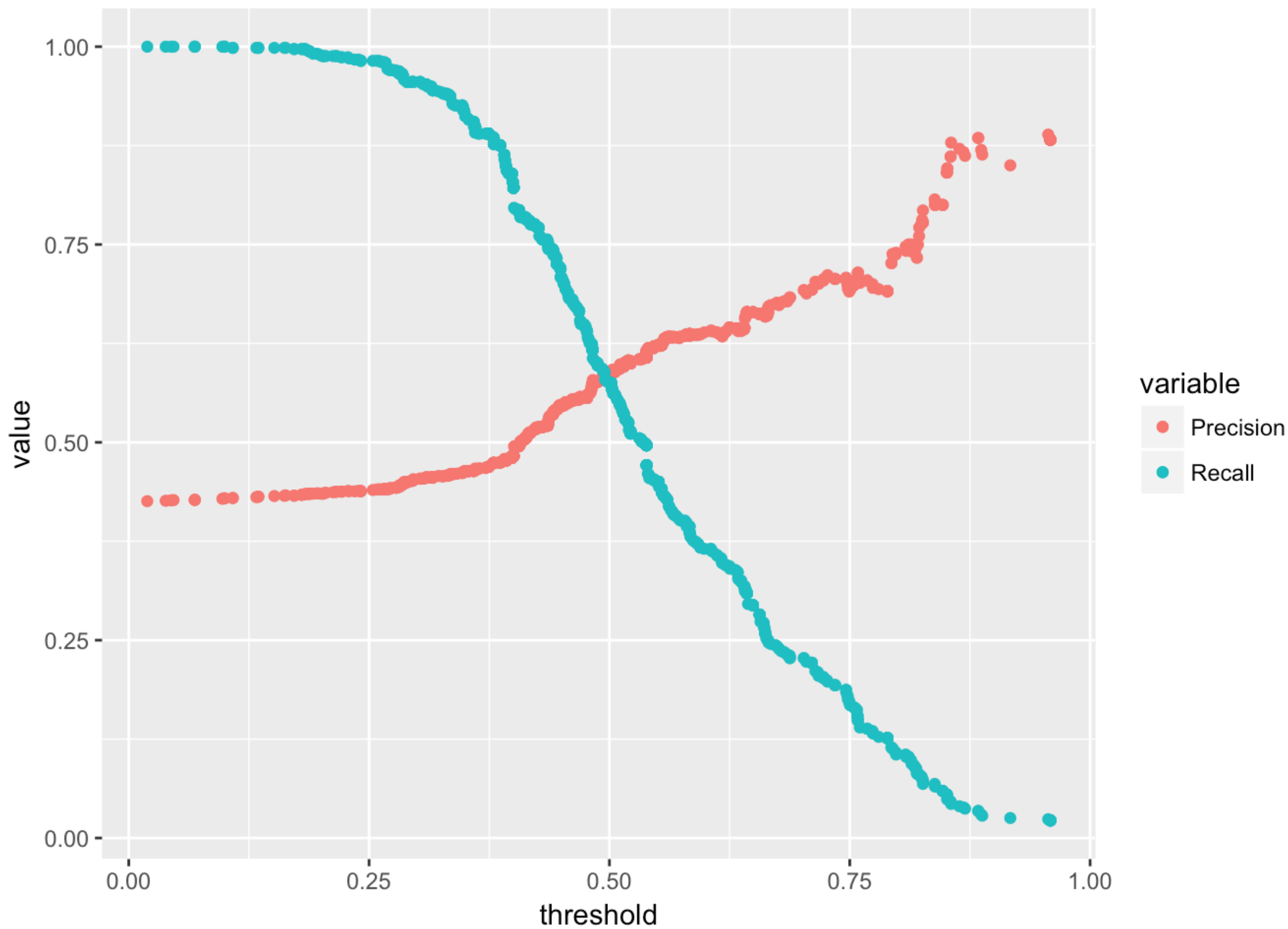

```
#AUC or C-Stats
performance(pred, measure = "auc")@y.values
```

```
## [[1]]
## [1] 0.679864
```

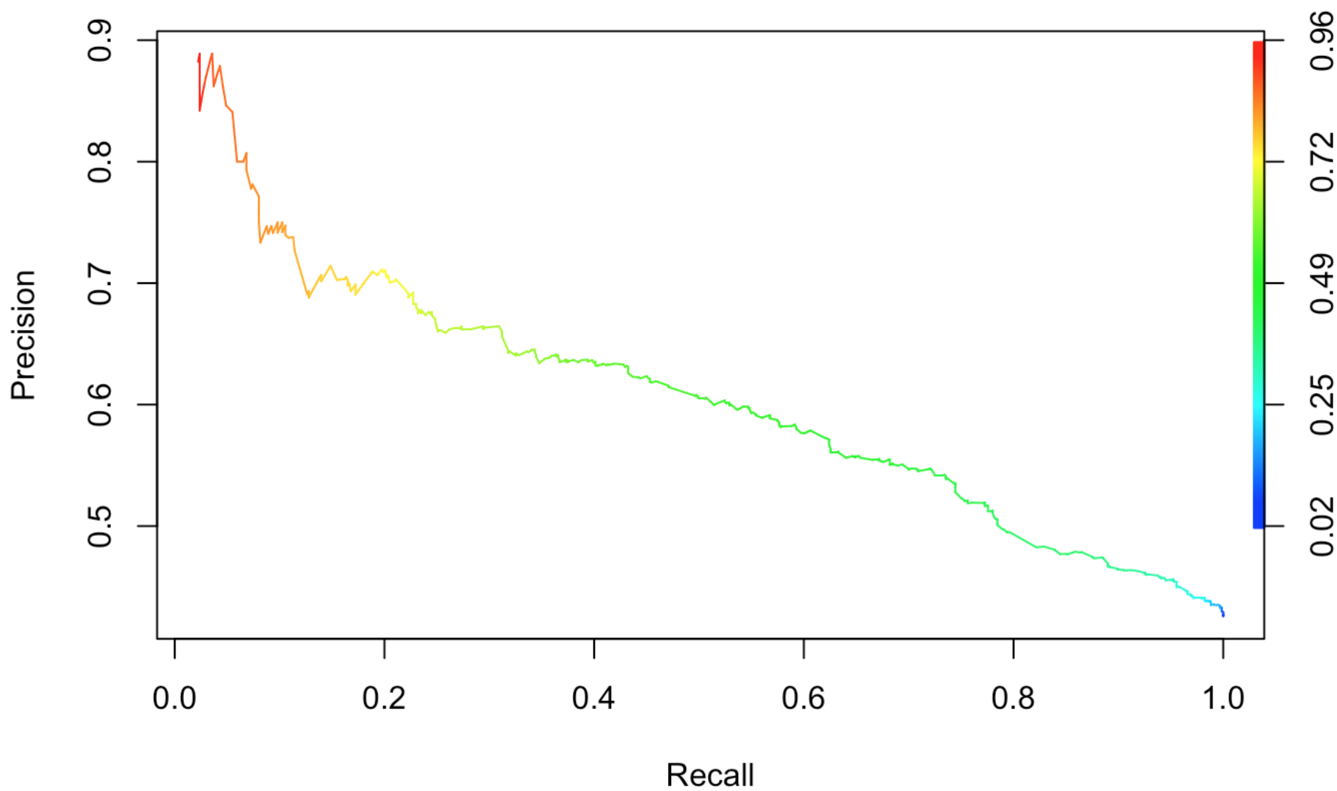
```r
dfLength <- length(seq(0.00,0.999,0.001))
thresholdDF <- data.frame(percent = numeric(dfLength),threshold = numeric(dfLength),
precision = numeric(dfLength), recall = numeric(dfLength), f1 = numeric(dfLength))
dfListIndex <-0
for (i in seq(0.00, 0.999, 0.001)){
  dfListIndex = dfListIndex +1
  #get minimum value from the sorted list which will be our cutoff value
  threshold = min(head(sort(xgb.probs,decreasing=TRUE), n = (length(xgb.probs)*(1-i))
  ))
  thresholdDF$percent[dfListIndex] = (1-i)*100
  thresholdDF$threshold[dfListIndex] = threshold
  # set to 1 if probability is above threshold otehr wise 0
  xgb.pred = rep(0, nrow(testing))
  xgb.pred[xgb.probs >= threshold] = 1
  #create confustion matrix and get pos pred value from the confusion matrix
  cm = confusionMatrix(xgb.pred, output_vector_testing, positive="1",
mode="prec_recall")
  byClass = cm$byClass
  precision = byClass[5]
  recall = byClass[6]
  f1 = byClass[7]
  thresholdDF$precision[dfListIndex] = precision
  thresholdDF$recall[dfListIndex] = recall
  thresholdDF$f1[dfListIndex] = f1
}
```

```
## Warning in confusionMatrix.default(xgb.pred, output_vector_testing,
## positive = "1", : Levels are not in the same order for reference and data.
## Refactoring data to match.
```

```r
# Plot
ggplot(data=thresholdDF,aes(threshold, y=value,color =variable))+ geom_point(aes(y=preci
sion , col = "Precision"))+ geom_point(aes(y=recall , col = "Recall"))
```



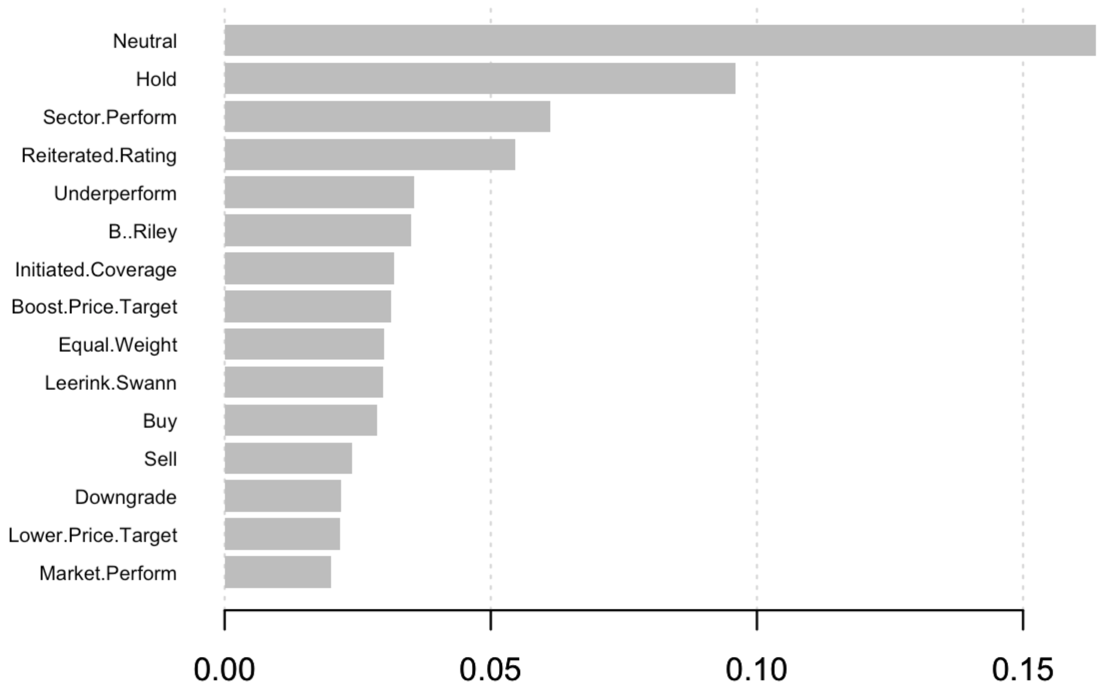

```
perf <- performance(pred,"prec","rec")
plot(perf, colorize=TRUE)
```




```
importance <- xgb.importance(feature_names = colnames(dtrain), model = bst)
head(importance, 25)
```

```
##                         Feature        Gain       Cover   Frequency
##  1:                     Neutral 0.163744211 0.026622026 0.034059946
##  2:                        Hold 0.096002070 0.036820422 0.029064487
##  3:              Sector.Perform 0.061171462 0.029537314 0.014078111
##  4:           Reiterated.Rating 0.054640240 0.016993371 0.108537693
##  5:                Underperform 0.035581780 0.035729617 0.016802906
##  6:                    B..Riley 0.035015884 0.034335775 0.020890100
##  7:          Initiated.Coverage 0.031881785 0.020718009 0.042234332
##  8:          Boost.Price.Target 0.031208389 0.009183818 0.060399637
##  9:                Equal.Weight 0.030039546 0.022822314 0.012261580
## 10:               Leerink.Swann 0.029723777 0.027152911 0.012715713
## 11:                         Buy 0.028614989 0.017330933 0.071752952
## 12:                        Sell 0.024005013 0.029979371 0.014532243
## 13:                   Downgrade 0.021791320 0.023348469 0.039509537
## 14:          Lower.Price.Target 0.021686147 0.002262820 0.033151680
## 15:              Market.Perform 0.019955851 0.028596782 0.015894641
## 16:                     Upgrade 0.017998923 0.004226613 0.030881017
## 17: Bank.of.America.Corporation 0.016158312 0.011965734 0.021344233
## 18:                  Outperform 0.015315279 0.007284579 0.024069028
## 19:              Citigroup.Inc. 0.015147563 0.012615295 0.015894641
## 20:         SunTrust.Banks..Inc. 0.014014662 0.023551335 0.011353315
## 21:         Sanford.C..Bernstein 0.012654713 0.012989764 0.014986376
## 22:                MKM.Partners 0.012178455 0.023849788 0.010899183
## 23:                   Macquarie 0.010213151 0.019716325 0.009082652
## 24:                 Underweight 0.010030488 0.020872581 0.009536785
## 25:           Cantor.Fitzgerald 0.009972372 0.006453986 0.008628520
##                         Feature        Gain       Cover   Frequency
```

```
xgb.plot.importance(importance_matrix = importance, top_n = 15)
```



Neutral
Hold
Sector.Perform
Reiterated.Rating
Underperform
B..Riley
Initiated.Coverage
Boost.Price.Target
Equal.Weight
Leerink.Swann
Buy
Sell
Downgrade
Lower.Price.Target
Market.Perform
0.00
0.05
0.10
0.15