Analyst Price Target Analysis using Random Forest model

```r
#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)]
```

```r
finaldf[names(finaldf)] <- lapply(finaldf[names(finaldf)],factor)
```

```r
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(417)
#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,]
```

```
h2o.init(
    nthreads=-1,            ## -1: use all available threads
    max_mem_size = "2G")    ## specify the memory size for the H2O cloud
```

```
##
## H2O is not running yet, starting it now...
##
## Note:  In case of errors look at the following log files:
##     /var/folders/0c/5jjqfkcj5qz84mm0xwqqh_zr0000gn/T//Rtmp4JPVad/h2o_virajbhalala_sta
rted_from_r.out
##     /var/folders/0c/5jjqfkcj5qz84mm0xwqqh_zr0000gn/T//Rtmp4JPVad/h2o_virajbhalala_sta
rted_from_r.err
##
##
## Starting H2O JVM and connecting: ... Connection successful!
##
## R is connected to the H2O cluster:
##     H2O cluster uptime:         4 seconds 199 milliseconds
##     H2O cluster version:        3.10.4.6
##     H2O cluster version age:    1 month and 16 days
##     H2O cluster name:           H2O_started_from_R_virajbhalala_ybo673
##     H2O cluster total nodes:    1
##     H2O cluster total memory:   1.78 GB
##     H2O cluster total cores:    4
##     H2O cluster allowed cores:  4
##     H2O cluster healthy:        TRUE
##     H2O Connection ip:          localhost
##     H2O Connection port:        54321
##     H2O Connection proxy:       NA
##     H2O Internal Security:      FALSE
##     R Version:                  R version 3.3.2 (2016-10-31)
```

```r
h2o.removeAll()            # Clean slate - just in case the cluster was already running
```

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

```r
system.time({
    training <- as.h2o(training, destination_frame="training")
    testing <- as.h2o(testing, destination_frame="testing")

    ## assign the first result the R variable train and the H2O name train.hex
    train <- h2o.assign(training, "train.hex")
    test <- h2o.assign(testing, "test.hex")     ## R test, H2O test.hex
})
```

```
##
  |
  |                                                                 |   0%
  |
  |=================================================================| 100%
##
  |
  |                                                                 |   0%
  |
  |=================================================================| 100%
```

```
##    user  system elapsed
##   1.181   0.054   3.503
```

```r
#delete headers
train <- train[-1, ]
test <- test[-1, ]
```

```
#system.time({
## run our first predictive model
## FOR EACH MODEL, NEED TO CHANGE PREDICTOR AND TARGET COLUMN!
rf1 <- h2o.randomForest(         ## h2o.randomForest function
  training_frame = train,        ## the H2O frame for training
  #validation_frame = valid,     ## the H2O frame for validation (not required)
  # x=-3,                        ## the predictor columns, by column index - if x is mis
sing, then all columns except y are used
  y=1,                           ## the target index (what we are predicting)
  model_id = "rf_covType_v1",    ## name the model in H2O
                                 ##   not required, but helps use Flow
  ntrees = 50,                  ## use a maximum of 200 trees to create the
                                 ##  random forest model. The default is 50.
                                 ##  I have increased it because I will let
                                 ##  the early stopping criteria decide when
                                 ##  the random forest is sufficiently accurate
  stopping_rounds = 2,           ## Stop fitting new trees when the 2-tree
                                 ##  average is within 0.001 (default) of
                                 ##  the prior two 2-tree averages.
                                 ##  Can be thought of as a convergence setting
  score_each_iteration = T,      ## Predict against training and validation for
                                 ##  each tree. Default will skip several.
  seed = 1000000)                ## Set the random seed so that this can be
```

```
##
  |
  |                                                                      |   0%
  |
  |======                                                                |  10%
  |
  |=====================                                                 |  30%
  |
  |================================                                      |  46%
  |
  |======================================================================| 100%
```

```
                                 ##  reproduced.
#})
```

```
system.time(rf.probs <- as.data.frame(h2o.predict(rf1, test)$p1))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
##    user  system elapsed
##   0.144   0.011   2.158
```

```
nrow(rf.probs) == nrow(test) # verifiy that this is equivalent to nrow(testing)
```

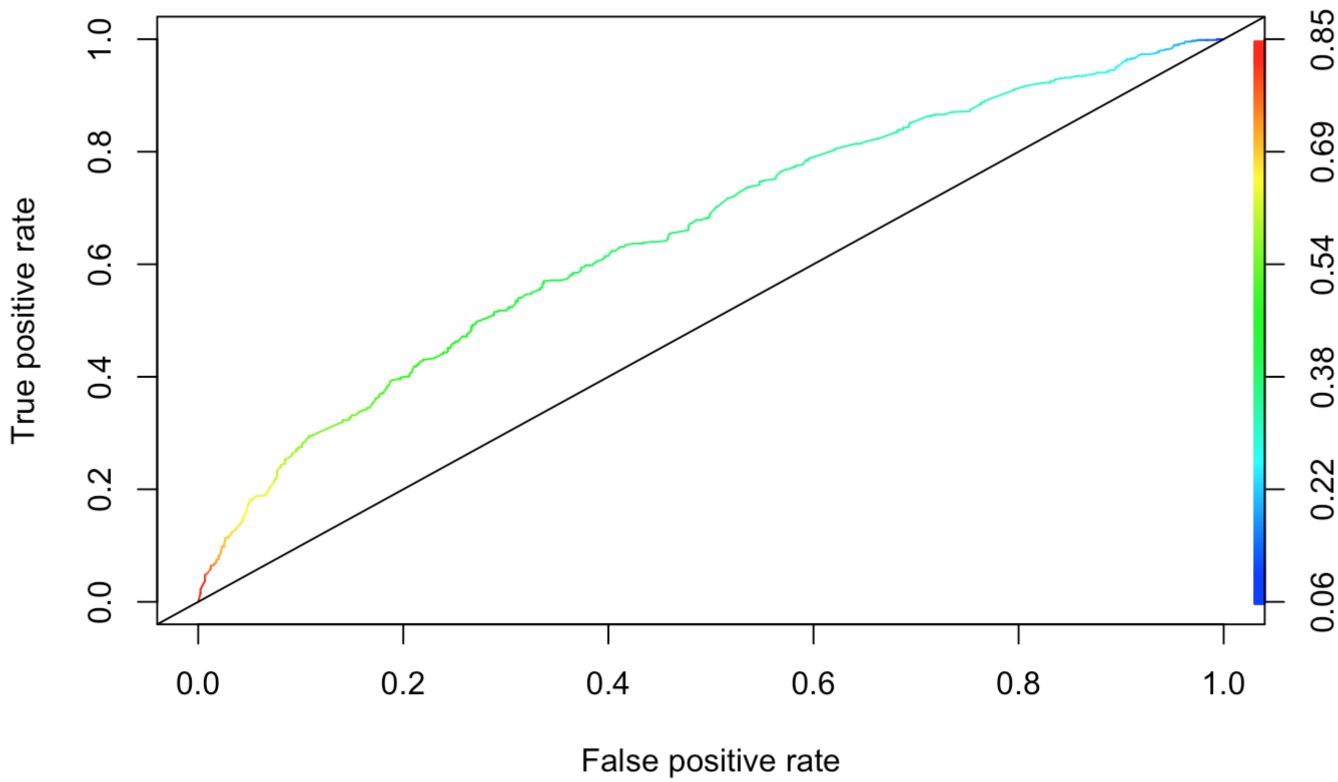
```
## [1] TRUE
```

```
# set cutoff value (selection is detailed in the Analysis section)
```

```
rf.pred <- as.numeric(rf.probs > 0.65)
# set positive class as "1"
# use precision/recall mode



#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


output_vector_testing <- as.data.frame(test)$TargetAchieved
confusionMatrix(rf.pred, output_vector_testing, positive="1", mode="prec_recall")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 865 563
##          1  43 109
##
##                Accuracy : 0.6165
##                  95% CI : (0.592, 0.6405)
##     No Information Rate : 0.5747
##     P-Value [Acc > NIR] : 0.0004072
##
##                   Kappa : 0.1277
##  Mcnemar's Test P-Value : < 2.2e-16
##
##               Precision : 0.71711
##                  Recall : 0.16220
##                      F1 : 0.26456
##              Prevalence : 0.42532
##          Detection Rate : 0.06899
##    Detection Prevalence : 0.09620
##       Balanced Accuracy : 0.55742
##
##        'Positive' Class : 1
##
```

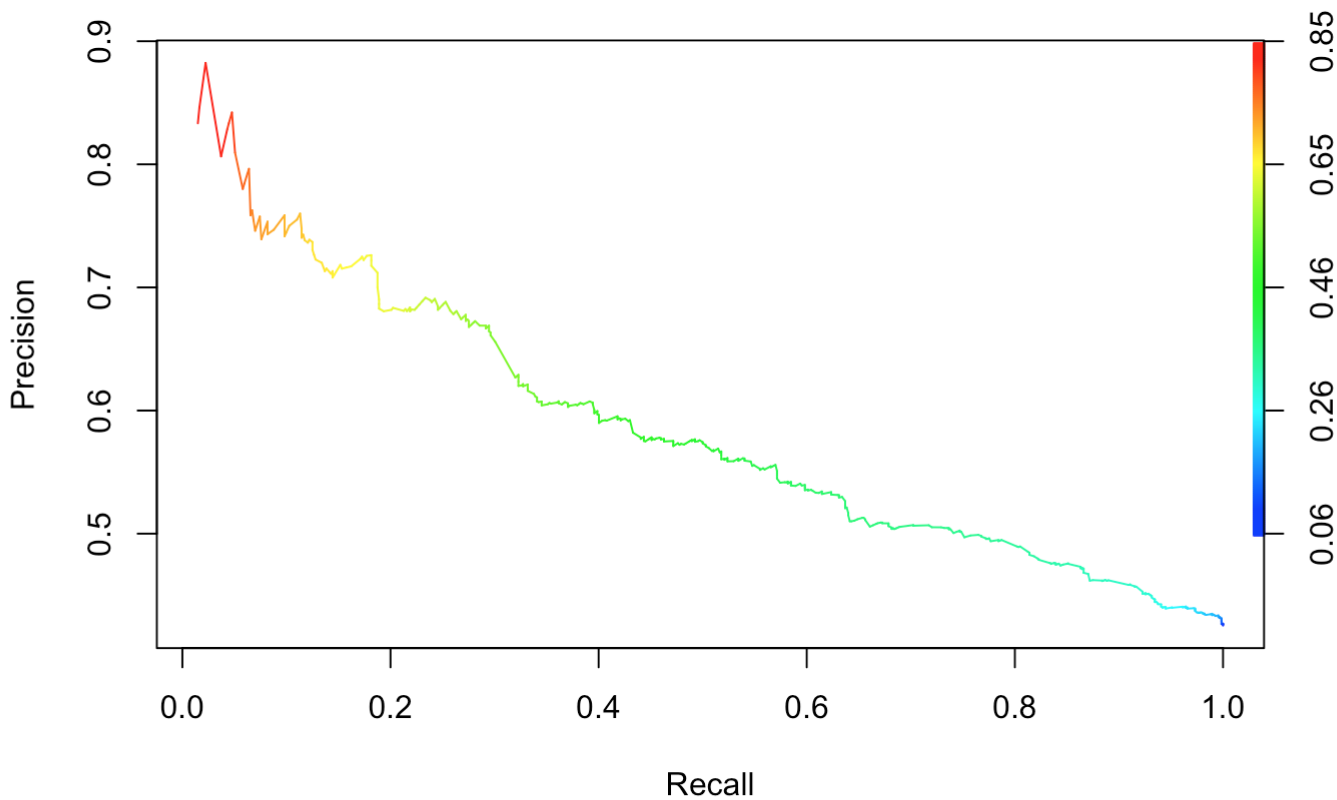
```
pred<- prediction(rf.probs, output_vector_testing)
perf <- performance(pred,"tpr","fpr")
plot(perf, colorize=TRUE)
abline(a=0, b=1)
```




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

```
## [[1]]
## [1] 0.6536917
```

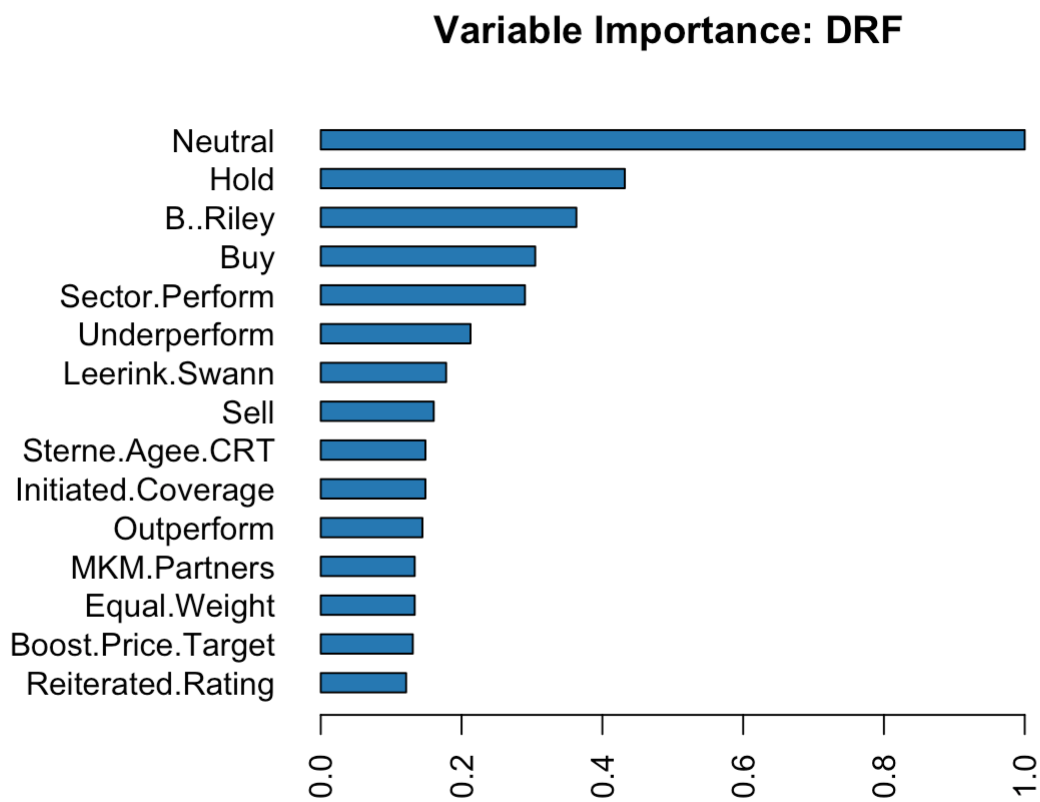
```
perf <- performance(pred,"prec","rec")
plot(perf, colorize=TRUE)
```




```
importance <- h2o.varimp(rf1)
head(importance, 25)
```

```
## Variable Importances:
##         variable relative_importance scaled_importance percentage
## 1        Neutral          340.788849          1.000000   0.149966
## 2           Hold          147.175522          0.431867   0.064765
## 3       B..Riley          123.742058          0.363105   0.054453
## 4            Buy          103.824883          0.304660   0.045689
## 5 Sector.Perform           98.848679          0.290058   0.043499
##
## ---
##                variable relative_importance scaled_importance percentage
## 20     Deutsche.Bank.AG           27.070045          0.079433   0.011912
## 21              Upgrade           25.804388          0.075720   0.011355
## 22 Royal.Bank.Of.Canada           23.935081          0.070234   0.010533
## 23            Downgrade           23.807281          0.069859   0.010477
## 24      Robert.W..Baird           23.783762          0.069790   0.010466
## 25              FBR...Co          23.212437          0.068114   0.010215
```

```
h2o.varimp_plot(rf1, num_of_features =  15)
```




```
# All done, shutdown H2O
h2o.shutdown(prompt=FALSE)
```

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