```r
library(randomForest)
library(ROCR)

train_tf_idf = read.csv("/Users/ouyamei/Documents/GitHub/kaggle-crisis/data/kaggle_train_tf_idf.csv" )
train_wc = read.csv("/Users/ouyamei/Documents/GitHub/kaggle-crisis/data/kaggle_train_wc.csv" )
test_wc = read.csv("/Users/ouyamei/Documents/GitHub/kaggle-crisis/data/kaggle_test_wc.csv" )

features = train_tf_idf[0:3000,c(-1,-502)]
label = as.factor(train_tf_idf$Predict[0:3000])
features_v = train_tf_idf[3000:4000,c(-1,-502)]
label_v = as.factor(train_tf_idf$Predict[3000:4000])

# find a best mtry parameter
bestmtry <- tuneRF(features,label, ntreeTry=100,
     stepFactor=1.5,improve=0.01, trace=TRUE, plot=TRUE, dobest=FALSE)

# use parameter to train a random forest model
rf <- randomForest(x=features, y=label, mtry=373, ntree=500,
     keep.forest=TRUE, importance=TRUE)

# create more model to compare, actually tried more than this
rf2 <- randomForest(x=features, y=label, mtry=373, ntree=500,
     classwt=c(3293,707), importance=TRUE)


# get the statitic result
rf.pr = predict(rf,newdata=features_v)
error = mean(rf.pr!=label_v)
library(Epi)
ROC(form=label_v~rf.pr, plot="ROC")
important_varibles = importance(rf,type=1)
```