Group-Project

Carl Saba

11/11/2021

```
setwd("/Users/carlsaba/Desktop/CSE-160/GroupProject")
shot_logs <- read.csv(file = 'shot_logs.csv')
shot_logs[["SHOT_CLOCK"]][is.na(shot_logs[["SHOT_CLOCK"]])] <- 0
shot_logs$SHOT_RESULT <- as.factor(shot_logs$SHOT_RESULT)
shot_logs$PTS_TYPE <- as.factor(shot_logs$PTS_TYPE)
```

Random Forest with 100 Trees

```
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
dfShuffled<-shot_logs[sample(nrow(shot_logs)),]
folds <- cut(seq(1,nrow(dfShuffled)),breaks=10,labels=FALSE)
accuracy <- integer(10)
precision <- integer(10)
recall <- integer(10)
fmeasure <- integer(10)
for (i in 1:10) {
    testIndexes <- which(folds==i,arr.ind=TRUE)
    testData <- dfShuffled[testIndexes,]
    trainData <- dfShuffled[-testIndexes,]
    forest <- randomForest(SHOT_RESULT ~ SHOT_DIST + CLOSE_DEF_DIST + PTS_TYPE + TOUCH_TIME + DRIBBLES
                            data = trainData, importance = TRUE, ntree = 100, na.action=na.roughfix)
    forest.test <- predict(forest, newdata=testData)
    tab <- table(forest.test, testData$SHOT_RESULT, dnn=c("Prediction","Actual"))
    print(paste("Fold:", i))
    print(tab)
    precision[i] <- tab[1,1]/sum(diag(tab))
    accuracy[i] <- sum(diag(tab))/sum(tab)
    recall[i] <- tab[1,1]/(tab[1,1] + tab[2,1])
    fmeasure[i] <- 2*precision[i]*recall[i]/(precision[i]+recall[i])
}
```

```
## [1] "Fold: 1"
##          Actual
## Prediction made missed
##     made   2295   1566
##     missed 3497   5449
## [1] "Fold: 2"
##          Actual
```

```
## Prediction made missed
##      made   2479   1677
##      missed 3454   5197
## [1] "Fold: 3"
##            Actual
## Prediction made missed
##      made   2317   1553
##      missed 3452   5485
## [1] "Fold: 4"
##            Actual
## Prediction made missed
##      made   2368   1605
##      missed 3366   5468
## [1] "Fold: 5"
##            Actual
## Prediction made missed
##      made   2407   1644
##      missed 3386   5370
## [1] "Fold: 6"
##            Actual
## Prediction made missed
##      made   2424   1704
##      missed 3310   5368
## [1] "Fold: 7"
##            Actual
## Prediction made missed
##      made   2417   1650
##      missed 3396   5344
## [1] "Fold: 8"
##            Actual
## Prediction made missed
##      made   2386   1598
##      missed 3363   5460
## [1] "Fold: 9"
##            Actual
## Prediction made missed
##      made   2393   1651
##      missed 3374   5389
## [1] "Fold: 10"
##            Actual
## Prediction made missed
##      made   2373   1668
##      missed 3448   5318
```

```r
cat("Number of observations:", nrow(shot_logs), "\n")
```

```
## Number of observations: 128069
```

```r
cat("Average Random Forest test Accuracy: ", mean(accuracy), "\n")
```

```
## Average Random Forest test Accuracy:  0.6067589
```

```r
cat("Average Random Forest test Precision: ", mean(precision), "\n")
```

```
## Average Random Forest test Precision:  0.3070654
```

```
cat("Average Random Forest test Recall: ", mean(recall), "\n")
```

Average Random Forest test Recall: 0.4120346

```
cat("Average Random Forest test F-measure: ", mean(recall), "\n")
```

Average Random Forest test F-measure: 0.4120346

Naive Bayes with 10 folds

```
library("e1071")
dfShuffled<-shot_logs[sample(nrow(shot_logs)),]
folds <- cut(seq(1,nrow(dfShuffled)),breaks=10,labels=FALSE)
accuracy <- integer(10)
precision <- integer(10)
recall <- integer(10)
fmeasure <- integer(10)
for(i in 1:10){
    testIndexes <- which(folds==i,arr.ind=TRUE)
    testData <- dfShuffled[testIndexes, ]
    trainData <- dfShuffled[-testIndexes, ]
    classifier <- naiveBayes(SHOT_RESULT~ PTS_TYPE + TOUCH_TIME + CLOSE_DEF_DIST + SHOT_DIST + DRIBBLES
    prediction <- predict(classifier, testData, type="class")
    tab <- table(prediction, testData$SHOT_RESULT,dnn=c("Prediction","Actual"))
    print(paste("Fold", i))
    print(tab)
    precision[i] <- tab[1,1]/sum(diag(tab))
    accuracy[i] <- sum(diag(tab))/sum(tab)
    recall[i] <- tab[1,1]/(tab[1,1] + tab[2,1])
    fmeasure[i] <- 2*precision[i]*recall[i]/(precision[i]+recall[i])
}
```

```
## [1] "Fold 1"
##         Actual
## Prediction made missed
##     made   3193   2711
##     missed 2600   4303
## [1] "Fold 2"
##         Actual
## Prediction made missed
##     made   3248   2610
##     missed 2668   4281
## [1] "Fold 3"
##         Actual
## Prediction made missed
##     made   3188   2710
##     missed 2590   4319
## [1] "Fold 4"
##         Actual
## Prediction made missed
##     made   3275   2657
##     missed 2565   4310
## [1] "Fold 5"
##         Actual
```

```
## Prediction made missed
##     made   3105  2710
##     missed 2553  4439
## [1] "Fold 6"
##        Actual
## Prediction made missed
##     made   3077  2706
##     missed 2705  4318
## [1] "Fold 7"
##        Actual
## Prediction made missed
##     made   3079  2655
##     missed 2617  4456
## [1] "Fold 8"
##        Actual
## Prediction made missed
##     made   3180  2626
##     missed 2670  4331
## [1] "Fold 9"
##        Actual
## Prediction made missed
##     made   3089  2721
##     missed 2630  4367
## [1] "Fold 10"
##        Actual
## Prediction made missed
##     made   3131  2622
##     missed 2742  4312
```

```r
# Print Averages
cat("Number of observations:", nrow(dfShuffled), "\n")
```

```
## Number of observations: 128069
```

```r
cat("Average Naive Bayes Test accuracy: ", mean(accuracy), "\n")
```

```
## Average Naive Bayes Test accuracy:  0.5856296
```

```r
cat("Average Naive Bayes Test precision: ", mean(precision), "\n")
```

```
## Average Naive Bayes Test precision:  0.4208446
```

```r
cat("Average Naive Bayes Test recall: ", mean(recall), "\n")
```

```
## Average Naive Bayes Test recall:  0.5451079
```

```r
cat("Average Naive Bayes Test F-measure: ", mean(fmeasure), "\n")
```

```
## Average Naive Bayes Test F-measure:  0.474956
```

Naive Bayes Lebron James with 10 Fold Cross Validation

```r
library("e1071")
shot_logs.player <- subset(shot_logs, player_id == 2544)
dfShuffled<-shot_logs.player[sample(nrow(shot_logs.player)),]
folds <- cut(seq(1,nrow(dfShuffled)),breaks=10,labels=FALSE)
acc <- integer(10)
```

```
prec <- integer(10)
rec <- integer(10)
fmeasure <- integer(10)
for (i in 1:10) {
    testIndexes <- which(folds==i,arr.ind=TRUE)
    testData <- dfShuffled[testIndexes, ]
    trainData <- dfShuffled[-testIndexes, ]
    classifier <- naiveBayes(SHOT_RESULT~ PTS_TYPE + TOUCH_TIME + CLOSE_DEF_DIST + SHOT_DIST + DRIBBLES
    prediction <- predict(classifier, testData, type="class")
    tab <- table(prediction, testData$SHOT_RESULT,dnn=c("Prediction","Actual"))
    print(paste("Fold", i))
    print(tab)
    accuracy[i] = sum(diag(tab))/sum(tab)
    precision[i] = tab[1]/(tab[1]+tab[3])
    recall[i] = tab[1]/(tab[1]+tab[2])
    fmeasure[i] <- 2*precision[i]*recall[i]/(precision[i]+recall[i])
}
```

```
## [1] "Fold 1"
##           Actual
## Prediction made missed
##     made       21     20
##     missed     24     33
## [1] "Fold 2"
##           Actual
## Prediction made missed
##     made       28     20
##     missed     19     31
## [1] "Fold 3"
##           Actual
## Prediction made missed
##     made       31     18
##     missed     20     29
## [1] "Fold 4"
##           Actual
## Prediction made missed
##     made       27     17
##     missed     24     29
## [1] "Fold 5"
##           Actual
## Prediction made missed
##     made       27     10
##     missed     27     34
## [1] "Fold 6"
##           Actual
## Prediction made missed
##     made       30     10
##     missed     18     40
## [1] "Fold 7"
##           Actual
## Prediction made missed
##     made       20     13
##     missed     22     42
## [1] "Fold 8"
```

```
##         Actual
## Prediction made missed
##     made    31    17
##     missed  23    27
## [1] "Fold 9"
##         Actual
## Prediction made missed
##     made    24    14
##     missed  17    43
## [1] "Fold 10"
##         Actual
## Prediction made missed
##     made    27    20
##     missed  18    33
```

```r
cat("Number of observations", nrow(dfShuffled), "\n")
```

```
## Number of observations 978
```

```r
cat("Average Lebron James Naive Bayes Test accuracy: ", mean(accuracy), "\n")
```

```
## Average Lebron James Naive Bayes Test accuracy:  0.6206291
```

```r
cat("Average Lebron James Naive Bayes Test precision: ", mean(precision), "\n")
```

```
## Average Lebron James Naive Bayes Test precision:  0.6279489
```

```r
cat("Average Lebron James Naive Bayes Test recall: ", mean(recall), "\n")
```

```
## Average Lebron James Naive Bayes Test recall:  0.5560297
```

```r
cat("Average Lebron James Naive Bayes Test F-measure: ", mean(fmeasure), "\n")
```

```
## Average Lebron James Naive Bayes Test F-measure:  0.587722
```

Naive Bayes Kyrie Irving with 10 Fold Cross Validation

```r
library("e1071")
shot_logs.player <- subset(shot_logs, player_id == 202681)
dfShuffled<-shot_logs.player[sample(nrow(shot_logs.player)),]
folds <- cut(seq(1,nrow(dfShuffled)),breaks=10,labels=FALSE)
acc <- integer(10)
prec <- integer(10)
rec <- integer(10)
fmeasure <- integer(10)
for (i in 1:10) {
    testIndexes <- which(folds==i,arr.ind=TRUE)
    testData <- dfShuffled[testIndexes, ]
    trainData <- dfShuffled[-testIndexes, ]
    classifier <- naiveBayes(SHOT_RESULT~ PTS_TYPE + TOUCH_TIME + CLOSE_DEF_DIST + SHOT_DIST, data=trai
    prediction <- predict(classifier, testData, type="class")
    tab <- table(prediction, testData$SHOT_RESULT,dnn=c("Prediction","Actual"))
    print(paste("Fold", i))
    print(tab)
    accuracy[i] = sum(diag(tab))/sum(tab)
    precision[i] = tab[1]/(tab[1]+tab[3])
    recall[i] = tab[1]/(tab[1]+tab[2])
```

```
    fmeasure[i] <- 2*precision[i]*recall[i]/(precision[i]+recall[i])

}
```

```
## [1] "Fold 1"
##         Actual
## Prediction made missed
##     made    22     11
##     missed  32     30
## [1] "Fold 2"
##         Actual
## Prediction made missed
##     made    15     21
##     missed  31     27
## [1] "Fold 3"
##         Actual
## Prediction made missed
##     made    25     18
##     missed  16     35
## [1] "Fold 4"
##         Actual
## Prediction made missed
##     made    18     21
##     missed  26     29
## [1] "Fold 5"
##         Actual
## Prediction made missed
##     made    22     18
##     missed  21     33
## [1] "Fold 6"
##         Actual
## Prediction made missed
##     made    20     23
##     missed  22     29
## [1] "Fold 7"
##         Actual
## Prediction made missed
##     made    25     30
##     missed  19     20
## [1] "Fold 8"
##         Actual
## Prediction made missed
##     made    23     24
##     missed  22     25
## [1] "Fold 9"
##         Actual
## Prediction made missed
##     made    19     22
##     missed  17     36
## [1] "Fold 10"
##         Actual
## Prediction made missed
##     made    20     17
##     missed  24     34
```

```
cat("Number of observations:", nrow(dfShuffled), "\n")
```

Number of observations: 942

```
cat("Average Kyrie Irving Average Naive Bayes Test accuracy: ", mean(accuracy), "\n")
```

Average Kyrie Irving Average Naive Bayes Test accuracy: 0.5381747

```
cat("Average Kyrie Irving Average Naive Bayes Test precision: ", mean(precision), "\n")
```

Average Kyrie Irving Average Naive Bayes Test precision: 0.5089246

```
cat("Average Kyrie Irving Average Naive Bayes Test recall: ", mean(recall), "\n")
```

Average Kyrie Irving Average Naive Bayes Test recall: 0.4801776

```
cat("Average Kyrie Irving Average Naive Bayes Test F-measure: ", mean(fmeasure), "\n")
```

Average Kyrie Irving Average Naive Bayes Test F-measure: 0.4893667

Naive Bayes Steph Curry with 10 Fold Cross Validation

```
library("e1071")
shot_logs.player <- subset(shot_logs, player_id == 201939)
dfShuffled<-shot_logs.player[sample(nrow(shot_logs.player)),]
folds <- cut(seq(1,nrow(dfShuffled)),breaks=10,labels=FALSE)
acc <- integer(10)
prec <- integer(10)
rec <- integer(10)
fmeasure <- integer(10)
for (i in 1:10) {
    testIndexes <- which(folds==i,arr.ind=TRUE)
    testData <- dfShuffled[testIndexes, ]
    trainData <- dfShuffled[-testIndexes, ]
    classifier <- naiveBayes(SHOT_RESULT~ PTS_TYPE + TOUCH_TIME + CLOSE_DEF_DIST + SHOT_DIST, data=trai
    prediction <- predict(classifier, testData, type="class")
    tab <- table(prediction, testData$SHOT_RESULT,dnn=c("Prediction","Actual"))
    print(paste("Fold", i))
    print(tab)
    accuracy[i] = sum(diag(tab))/sum(tab)
    precision[i] = tab[1]/(tab[1]+tab[3])
    recall[i] = tab[1]/(tab[1]+tab[2])
    fmeasure[i] <- 2*precision[i]*recall[i]/(precision[i]+recall[i])

}
```

```
## [1] "Fold 1"
##         Actual
## Prediction made missed
##     made    25     18
##     missed  19     35
## [1] "Fold 2"
##         Actual
## Prediction made missed
##     made    19     17
##     missed  19     42
## [1] "Fold 3"
```

```
##          Actual
## Prediction made missed
##     made    29     20
##     missed  16     32
## [1] "Fold 4"
##          Actual
## Prediction made missed
##     made    27     10
##     missed  27     32
## [1] "Fold 5"
##          Actual
## Prediction made missed
##     made    28     30
##     missed  15     24
## [1] "Fold 6"
##          Actual
## Prediction made missed
##     made    37      9
##     missed  20     31
## [1] "Fold 7"
##          Actual
## Prediction made missed
##     made    22     16
##     missed  22     36
## [1] "Fold 8"
##          Actual
## Prediction made missed
##     made    25     13
##     missed  25     34
## [1] "Fold 9"
##          Actual
## Prediction made missed
##     made    21     16
##     missed  28     32
## [1] "Fold 10"
##          Actual
## Prediction made missed
##     made    24     18
##     missed  22     33
```

```r
cat("Number of observations:", nrow(dfShuffled), "\n")
```

```
## Number of observations: 968
```

```r
cat("Average Steph Curry Naive Bayes Test accuracy: ", mean(accuracy), "\n")
```

```
## Average Steph Curry Naive Bayes Test accuracy:  0.607442
```

```r
cat("Average Steph Curry Naive Bayes Test precision: ", mean(precision), "\n")
```

```
## Average Steph Curry Naive Bayes Test precision:  0.6093684
```

```r
cat("Average Steph Curry Naive Bayes Test recall: ", mean(recall), "\n")
```

```
## Average Steph Curry Naive Bayes Test recall:  0.5463222
```

```
cat("Average Steph Curry Naive Bayes Test F-measure: ", mean(fmeasure), "\n")
```

Average Steph Curry Naive Bayes Test F-measure: 0.571015