CS 513 Final - Q5

Aidan Fischer - 10447681

2023-12-18

Load data

```
rm(list=ls())
library(caTools)
library(class)
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.3.2
```

```
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
## Warning: package 'lattice' was built under R version 4.3.2
```

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.3.2
```

```
## randomForest 4.7-1.1
```

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

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
data = read.csv("Hear_attack.csv")
data = data[complete.cases(data), ]
set.seed(255)
data$Heart_attack = as.factor(data$Heart_attack)
split = sample.split(data$Heart_attack, SplitRatio=0.7)
train = subset(data, split == TRUE)
test = subset(data, split == FALSE)
```

Create the random forest model

```
classifier = randomForest(Heart_attack~.,data=train,method="")
```

Evaluation

```
train_pred <- predict(classifier, newdata=train, type="class")
test_pred  <- predict(classifier, newdata=test, type="class")
```

```r
cm_train <- table(train$Heart_attack, train_pred)
cm_test <- table(test$Heart_attack, test_pred)
```

```r
confusionMatrix(cm_train)
```

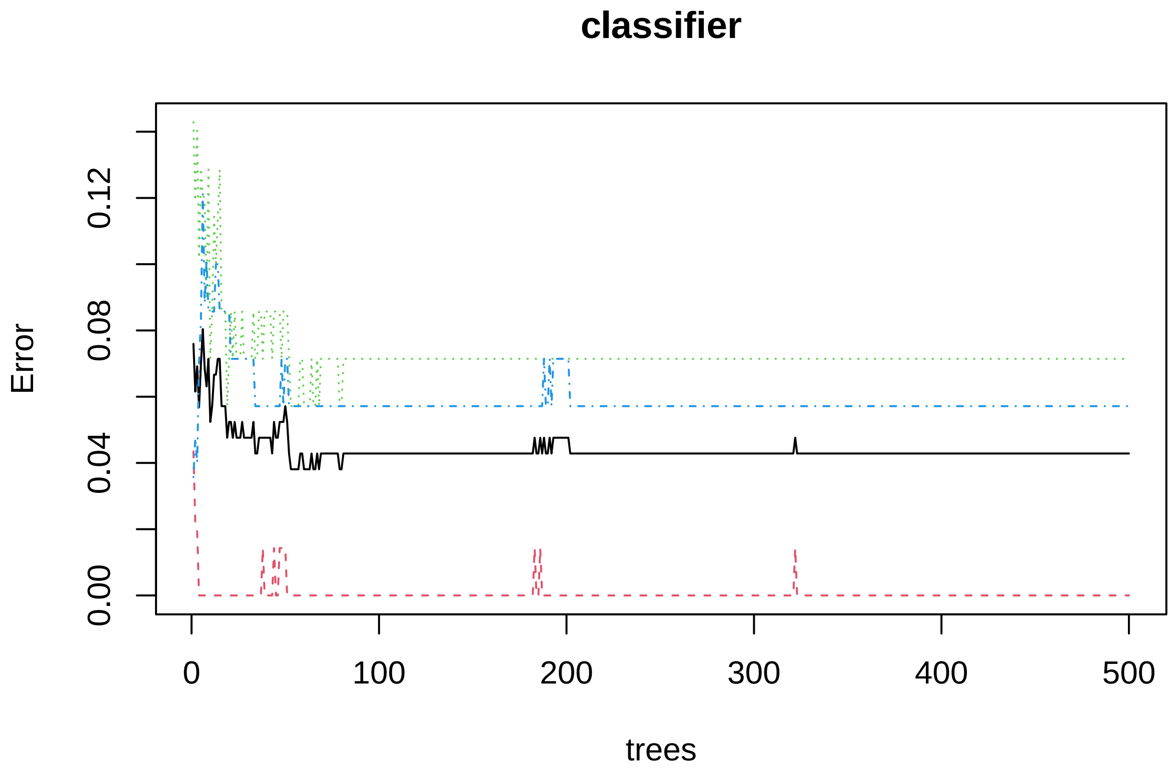
```
## Confusion Matrix and Statistics
##
##          train_pred
##           Light Massive Mild
##   Light      70       0    0
##   Massive     0      70    0
##   Mild        0       0   70
##
## Overall Statistics
##
##                Accuracy : 1
##                  95% CI : (0.9826, 1)
##     No Information Rate : 0.3333
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Light Class: Massive Class: Mild
## Sensitivity                1.0000         1.0000      1.0000
## Specificity                1.0000         1.0000      1.0000
## Pos Pred Value             1.0000         1.0000      1.0000
## Neg Pred Value             1.0000         1.0000      1.0000
## Prevalence                 0.3333         0.3333      0.3333
## Detection Rate             0.3333         0.3333      0.3333
## Detection Prevalence       0.3333         0.3333      0.3333
## Balanced Accuracy          1.0000         1.0000      1.0000
```

```r
confusionMatrix(cm_test)
```

```
## Confusion Matrix and Statistics
##
##          test_pred
##           Light Massive Mild
##   Light      30       0    0
##   Massive     0      29    1
##   Mild        0       3   27
##
## Overall Statistics
##
##                Accuracy : 0.9556
##                  95% CI : (0.8901, 0.9878)
##     No Information Rate : 0.3556
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9333
##
```

```
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Light Class: Massive Class: Mild
## Sensitivity                1.0000         0.9062      0.9643
## Specificity                1.0000         0.9828      0.9516
## Pos Pred Value             1.0000         0.9667      0.9000
## Neg Pred Value             1.0000         0.9500      0.9833
## Prevalence                 0.3333         0.3556      0.3111
## Detection Rate             0.3333         0.3222      0.3000
## Detection Prevalence       0.3333         0.3333      0.3333
## Balanced Accuracy          1.0000         0.9445      0.9579
```

```r
plot(classifier)
```



```r
print(classifier)
```

```
##
## Call:
##  randomForest(formula = Heart_attack ~ ., data = train, method = "")
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 2
##
##         OOB estimate of  error rate: 4.29%
## Confusion matrix:
```

```
##         Light Massive Mild class.error
## Light      70       0    0  0.00000000
## Massive     0      65    5  0.07142857
## Mild        0       4   66  0.05714286
```

Variable Importance

```
classifier["importance"]
```

```
## $importance
##        MeanDecreaseGini
## RestHR        12.380493
## MaxHR          4.999805
## RecHR         68.570332
## BP            53.283323
```

Recovery Heart rate is the most important variable, next to BP. Recovery heart rate seems to increase with the severity of the heart attack, so it seems to be a good indicator of the target variable.