# Practical ML - Prediction Assignment

Sath

12/07/2020

## Predict if a physical activity is performed correctly with self-reported data from individuals.

With devices such as Jawbone Up, Nike FuelBand, and Fitbit it is possible to collect a large amount of data about personal activity. People regularly quantify how much of a particular activity they do, but they rarely quantify how well they do it.

In this exercise, 6 participants were asked to perform barbell lifts correctly and incorrectly in 5 different ways. The goal of this project, is to use data from accelerometers on the belt, forearm, arm, and dumbell and predict the manner in which they did the exercise.

Data is from http://web.archive.org/web/20161224072740/htp://http:/groupware.les.inf.puc-rio.br/har (http://web.archive.org/web/20161224072740/htp://http:/groupware.les.inf.puc-rio.br/har). See citation [1]

Following are steps involved in this exercise:

1. Process the data, for use of this project
2. Explore the data, and remove variables with no predicting power
3. Model selection - try different models
4. Model examination, to see if we get a good accuracy
5. Conclusion on the better ML model that can be used for classification
6. Predicting the classification (‘classe’) on test set using the ML Model.

```
train_data <- read.csv("pml-training.csv", na.strings = c("NA",""))
test_data <- read.csv("pml-testing.csv", na.strings = c("NA", ""))
```

```
dim(train_data)
```

```
## [1] 19622   160
```

```
names(train_data)
```

```
##   [1] "X"                         "user_name"
##   [3] "raw_timestamp_part_1"      "raw_timestamp_part_2"
##   [5] "cvtd_timestamp"            "new_window"
##   [7] "num_window"                "roll_belt"
##   [9] "pitch_belt"                "yaw_belt"
##  [11] "total_accel_belt"          "kurtosis_roll_belt"
##  [13] "kurtosis_picth_belt"       "kurtosis_yaw_belt"
##  [15] "skewness_roll_belt"        "skewness_roll_belt.1"
##  [17] "skewness_yaw_belt"         "max_roll_belt"
##  [19] "max_picth_belt"            "max_yaw_belt"
##  [21] "min_roll_belt"             "min_pitch_belt"
##  [23] "min_yaw_belt"              "amplitude_roll_belt"
##  [25] "amplitude_pitch_belt"      "amplitude_yaw_belt"
##  [27] "var_total_accel_belt"      "avg_roll_belt"
##  [29] "stddev_roll_belt"          "var_roll_belt"
##  [31] "avg_pitch_belt"            "stddev_pitch_belt"
##  [33] "var_pitch_belt"            "avg_yaw_belt"
##  [35] "stddev_yaw_belt"           "var_yaw_belt"
##  [37] "gyros_belt_x"              "gyros_belt_y"
##  [39] "gyros_belt_z"              "accel_belt_x"
##  [41] "accel_belt_y"              "accel_belt_z"
##  [43] "magnet_belt_x"             "magnet_belt_y"
##  [45] "magnet_belt_z"             "roll_arm"
##  [47] "pitch_arm"                 "yaw_arm"
##  [49] "total_accel_arm"           "var_accel_arm"
##  [51] "avg_roll_arm"              "stddev_roll_arm"
##  [53] "var_roll_arm"              "avg_pitch_arm"
##  [55] "stddev_pitch_arm"          "var_pitch_arm"
##  [57] "avg_yaw_arm"               "stddev_yaw_arm"
##  [59] "var_yaw_arm"               "gyros_arm_x"
##  [61] "gyros_arm_y"               "gyros_arm_z"
##  [63] "accel_arm_x"               "accel_arm_y"
##  [65] "accel_arm_z"               "magnet_arm_x"
##  [67] "magnet_arm_y"              "magnet_arm_z"
##  [69] "kurtosis_roll_arm"         "kurtosis_picth_arm"
##  [71] "kurtosis_yaw_arm"          "skewness_roll_arm"
##  [73] "skewness_pitch_arm"        "skewness_yaw_arm"
##  [75] "max_roll_arm"              "max_picth_arm"
##  [77] "max_yaw_arm"               "min_roll_arm"
##  [79] "min_pitch_arm"             "min_yaw_arm"
##  [81] "amplitude_roll_arm"        "amplitude_pitch_arm"
##  [83] "amplitude_yaw_arm"         "roll_dumbbell"
##  [85] "pitch_dumbbell"            "yaw_dumbbell"
##  [87] "kurtosis_roll_dumbbell"    "kurtosis_picth_dumbbell"
##  [89] "kurtosis_yaw_dumbbell"     "skewness_roll_dumbbell"
##  [91] "skewness_pitch_dumbbell"   "skewness_yaw_dumbbell"
##  [93] "max_roll_dumbbell"         "max_picth_dumbbell"
##  [95] "max_yaw_dumbbell"          "min_roll_dumbbell"
##  [97] "min_pitch_dumbbell"        "min_yaw_dumbbell"
##  [99] "amplitude_roll_dumbbell"   "amplitude_pitch_dumbbell"
## [101] "amplitude_yaw_dumbbell"    "total_accel_dumbbell"
## [103] "var_accel_dumbbell"        "avg_roll_dumbbell"
## [105] "stddev_roll_dumbbell"      "var_roll_dumbbell"
## [107] "avg_pitch_dumbbell"        "stddev_pitch_dumbbell"
## [109] "var_pitch_dumbbell"        "avg_yaw_dumbbell"
## [111] "stddev_yaw_dumbbell"       "var_yaw_dumbbell"
## [113] "gyros_dumbbell_x"          "gyros_dumbbell_y"
```

```
## [115] "gyros_dumbbell_z"         "accel_dumbbell_x"
## [117] "accel_dumbbell_y"         "accel_dumbbell_z"
## [119] "magnet_dumbbell_x"        "magnet_dumbbell_y"
## [121] "magnet_dumbbell_z"        "roll_forearm"
## [123] "pitch_forearm"            "yaw_forearm"
## [125] "kurtosis_roll_forearm"    "kurtosis_picth_forearm"
## [127] "kurtosis_yaw_forearm"     "skewness_roll_forearm"
## [129] "skewness_pitch_forearm"   "skewness_yaw_forearm"
## [131] "max_roll_forearm"         "max_picth_forearm"
## [133] "max_yaw_forearm"          "min_roll_forearm"
## [135] "min_pitch_forearm"        "min_yaw_forearm"
## [137] "amplitude_roll_forearm"   "amplitude_pitch_forearm"
## [139] "amplitude_yaw_forearm"    "total_accel_forearm"
## [141] "var_accel_forearm"        "avg_roll_forearm"
## [143] "stddev_roll_forearm"      "var_roll_forearm"
## [145] "avg_pitch_forearm"        "stddev_pitch_forearm"
## [147] "var_pitch_forearm"        "avg_yaw_forearm"
## [149] "stddev_yaw_forearm"       "var_yaw_forearm"
## [151] "gyros_forearm_x"          "gyros_forearm_y"
## [153] "gyros_forearm_z"          "accel_forearm_x"
## [155] "accel_forearm_y"          "accel_forearm_z"
## [157] "magnet_forearm_x"         "magnet_forearm_y"
## [159] "magnet_forearm_z"         "classe"
```

```
# Remove variables like timestamp that offer little or no predicting power
train_data <- train_data[,colSums(is.na(train_data)) == 0]
trainData <- train_data[, -c(1:7)]

test_data <- test_data[,colSums(is.na(test_data)) == 0]
testData  <- test_data[, -c(1:7)]
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.0.2
```

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

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

```
set.seed(7826)
inTrain <- createDataPartition(trainData$classe, p = 0.7, list = FALSE)
train <- trainData[inTrain, ]
valid <- trainData[-inTrain, ]
```

# Classification Tree with k-fold cross validation:

We build a Classification Tree model with a 5 fold cross validation and fit the training data on to the model. We then print the results for review.

```
control   <- trainControl(method = "cv", number = 5)
fit_rpart <- train(classe ~ ., data = train, method = "rpart",
                   trControl = control)
print(fit_rpart, digits = 4)
```

```
## CART
##
## 13737 samples
##    52 predictor
##     5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 10990, 10990, 10990, 10989, 10989
## Resampling results across tuning parameters:
##
##   cp       Accuracy  Kappa
##   0.03102  0.5260    0.38038
##   0.05954  0.3935    0.16982
##   0.11586  0.3168    0.04946
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was cp = 0.03102.
```

Plot the tree for visual inspection.

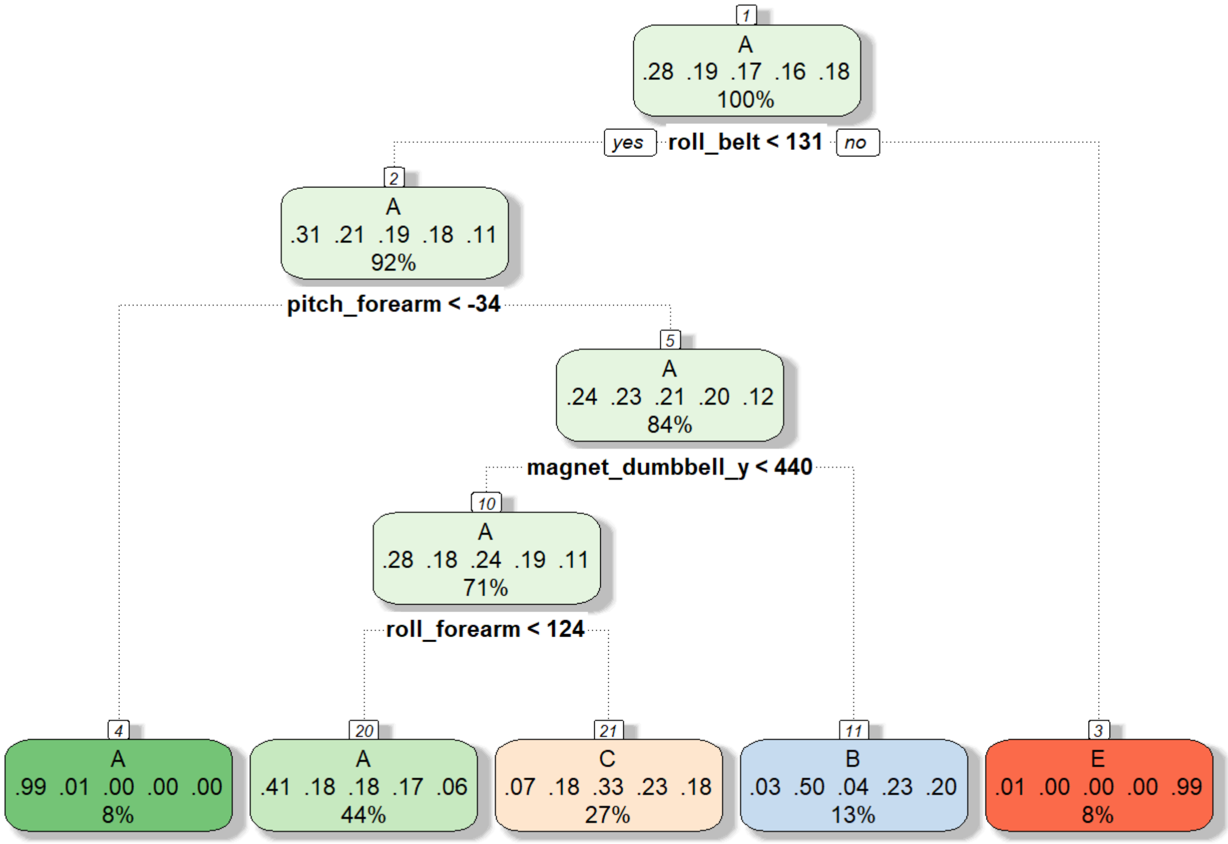
```
## Warning: package 'rpart.plot' was built under R version 4.0.2
```

```
## Warning: package 'rattle' was built under R version 4.0.2
```

```
## Loading required package: tibble
```

```
## Loading required package: bitops
```

```
## Rattle: A free graphical interface for data science with R.
## Version 5.4.0 Copyright (c) 2006-2020 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```




Rattle 2020-Jul-12 15:32:17 Dhuruv

Now we run the validation dataset through the model to see what the prediction looks like using the Classification tree.

```
# predict outcomes using validation set
predict_rpart <- predict(fit_rpart, valid)
predict_rpart
```

```
##    [1] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##   [38] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##   [75] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [112] A A A C A A A A A A A A A A A A A A A A A A A A A A A C A A A A A C A A A
##  [149] A A A A C A A A A A A A A A A A A C A C A A A A A C A A A A A A C C A A A
##  [186] C A C A A A C C C A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [223] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [260] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [297] A A A A A A A A A A A A E E E A A A A A A A A A A A A A A A A A A A A A A
##  [334] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [371] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [408] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [445] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [482] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [519] A A A A A A A C C A A A A A A A A A A A A A C C C C C C C C C A A A A A A
##  [556] A A A A C C C C C C C A A A A A A C C C A C C C C C C C C C C C A A A A A
##  [593] A C C C C C C C C C A A A A A C C C C C C A A A C C C C C C C C C C C A A
##  [630] A A A C C C C C A A A A A A A A A A A C C C C C C C C A A A A A A A A A A
##  [667] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [704] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [741] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [778] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [815] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [852] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [889] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [926] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
##  [963] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1000] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1037] A A A A A A A A A A A A A A A A A A A A A A A A A A A A B A A A A A A A B
## [1074] B B B B A A A A A B B B B B B B B B B A A A A A A A A A A A B B B B B B B
## [1111] B B A A A C A C C A B A A A A A A A A A C C C C A A A A A A A A A A A A A
## [1148] A A A A A C A A A A A A A A A C C C A A A A A C A A A A A A C C C C C A A
## [1185] A A A B A A A A C C C A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1222] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1259] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1296] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1333] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1370] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1407] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1444] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1481] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1518] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [1555] A A A A A A A A A A A A A A A A A C C C C C C A A A A A A A A C C C C C C
## [1592] C C C C C C C C C A A A A A A A C C C A A A A A A A A A A A A A A A A A A
## [1629] A A A A A A A A A A C C C A C C A A A A A A A A A A A A A A A A A A A A A
## [1666] A A A A A A A A A A A A C C A A A A A A A A A A A A A A C C C A A A A A A
## [1703] A A A A A A A A A A A C C A A A A A A A A A A C C A A A A C A C A C C A A
## [1740] A A A A A A C A A C A A A A A C A A A C C C C C C C C A A A A C A A C C C
## [1777] C C C C C C C A C C C A A C C C C C C C C C C C C A A A A A A C C C C C C
## [1814] C C C A A A A C C C C C C C C C C C C C C C C C C C A A A A C C C C C C C
## [1851] C C A A A A A A A C C C C C C C C C C A A A A C C C C C C C C C A A A A A
## [1888] A A A A A C C C C C C C A A A A A A A A A C C C C C C C C A A A A A A C C
## [1925] C C A A A A A C C C C C C C C A A A A A A B B B B B B B B B A A A A A A A
## [1962] A A A A B B B B B B B B B A A B B B B B B A A A A A B A A A A A A A A A A
## [1999] C C C C B B B B B B B B B B C C C C C C A B B C A A A A A A A A A A A C C
## [2036] C A A A B B B B B B C C C A B B B B B C A A A C C C C C C C C C C C A A A
## [2073] A A A A A A C C C C C C C B B B B B C C C C C C C A C C A A A A A C C C B
```

```
## [2110] B B B B B B B B B B B B B B B B B C C C B B B B B B B B B B C C C B B B B
## [2147] B B B B B B C C C C C C B B B B B B B C C C B B B B B B B B B C B B B B B
## [2184] B B C B B B A A A A B B A A A A A A A B B B B B B B B A A A A A A A A A A
## [2221] B B B A A A A A A A A A A B B B B B A A A A A A A A A A A A A A A B B B B
## [2258] B B A A A A A A B B A A A A A A A A A A A A A A A A A A A A B B B B B B A
## [2295] A A A A A A A A A A A A A A A B B B B B B B B B B A A A A A A A A A A A A
## [2332] A B B B B B B B A A A A B A A A A A A A B B B B B B A B B B B B A A A A A
## [2369] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [2406] B B B B B B B B A A A A A A A A A B B B B B B B B B B B B B B B B B B B B
## [2443] B B B B B A A B B B B B B B B B B B B B B B B B B B B B B B B B B A B B B
## [2480] A B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B
## [2517] B B B B B B B B B B B B B B B B A A A A B B B B B B B B B B B B B B B B B
## [2554] B B B B B B B A A A A A B B B B B B B B B B B B B B B B A A A A A A A A A
## [2591] A A A A A A A A A A A A A A A B B B B B B B B B B B A B B B B B B B B B B
## [2628] B B B B B A A A B B C C C C B B B B B C C C C C C C C C C C B B B B B C C
## [2665] C C C C C C C C C C A A A A A A A A A A C A C C C C C C C C C A A C C C C
## [2702] C C C C C C C A A A A A A C A A A A A C C C C C C A A A C C C C A A A A A
## [2739] A A A A A A A A A A A A A A A C B B B B B B B B B C C A B B B C B B B B B
## [2776] B B B B B B B C C C C A A A A A B B B B B B B B B B B B B A A A A A A B B
## [2813] B C C C C C C C C C C C C A A A A C C C C C C C C C C C C C C C C C C C C
## [2850] C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C
## [2887] C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C
## [2924] C C C C C C C C C C C C C C C C C C C C C C A C A A A A A A C C C C C C C
## [2961] C C C A A A A A A A A A A A A A A A A A A C C C C C A A A A A A A A A A A
## [2998] A A A A A A A A C C C C C A A A A A A A A A C C C C A A A A A A A C C C C
## [3035] C A A A A A A A A C C C C C A A A A C C C C A A A A A A A C C C C C C C C
## [3072] C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C
## [3109] C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C
## [3146] C C C C C C C C C C C C C C A C C C C C C C C C C C C C A C C C C C A A A
## [3183] A A A C C C C C C C A A B B B C C C C C C C C C C C C C C C C C C C C C C
## [3220] C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C
## [3257] C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C
## [3294] C C C C C C C C C C C C C C C C C A A A C C C C C C C C C C C C A C C C C
## [3331] C C C C C C C C C B B B B B C B B B B B B B B C C C C C C C C C C B B B B
## [3368] B B B B B B B B B B B A A A A A A A A A A A A A A A A A A A A A A A A A A
## [3405] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [3442] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [3479] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [3516] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [3553] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A B B B B A
## [3590] A A A B A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [3627] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [3664] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [3701] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [3738] A A A A C C C C C C C C C C B C C C C C A A A A A A C C C C C C C C A C C
## [3775] C C C A A C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C
## [3812] C C C C C C C C C A A A C C C C C C C C C C C C C C C C A A A A A A C C C
## [3849] C A A A A A C C C C C C A A C A A A A A A A A A A C C C C C C C A A A A A
## [3886] A A A A A C C C C C C C C C C C C C C A A A A A C A A C A A A A A A C C C
## [3923] C C C C C A A A A A C C C C C C C C C C C A A A A A A C C C C C C C C C A
## [3960] A C C C C C A A A A A A A A A C C C C C C C C C C C C C C C C A A A C C C
## [3997] C C C C A A A A A A C C C C C C A A A A A B B B B B B B B B B B B B B B B
## [4034] B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B
## [4071] B B B B B B B B B B B B B B B B B B B B B B B B C C C C A A A A A A A A C
## [4108] C A A A A A A A C A A A A A C C C C C C A A A A A A A A A A A A C C C C C
## [4145] C C A A C C C C C A A A A A A A A A A A A A C C C C C A A A A A A A C C C
## [4182] C C C A A A A A A A A A A C C C C C C C C C C C A C A A A A A A A A C C A
## [4219] A A C C C C A A A A A C C C C C C C C C C C C C C C C C C C C A A A A A C
```

```
## [4256] C A A A A A A A C C C C C C C A A A A C C C A A A A A A A A A A A C C C C
## [4293] C A A A A A A C C C A A A C C C C C C C C C C C C C A C C C C C C C C C C
## [4330] C C C C C C C B B C C C C A A A C C A A A A A A A C C C C C C C A A A C C
## [4367] C C C C C A A A C C C B B B C C A A C C C B B B B B B C C A C C C C C C C
## [4404] C C C C C C B C C C C C C C C B C C C C C C C C C C C C C C C C C C C C C
## [4441] C C C C A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [4478] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [4515] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [4552] A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A
## [4589] A A A A A A A A A A A A A A A A A A A A B B B B B B B B B B B B B B B B B
## [4626] B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B B
## [4663] B B B B B B B B B B A B B B B B B B B A A A A A A A C C C C C C A C C C C
## [4700] C C C C C C C C C C C B C C C C C C C C C A C C C A C A A C A A A A A A A
## [4737] C C C C C C C A A A A A A A A A A A A C C C A A A C C C C C A A A C C C C
## [4774] C C C C C C A A A A C C C C A A A A A C C C C A A A C C C C A A A A A A A
## [4811] A A A A A A A C C C E E E E E E E E E E E E E E E E E E E E E E E E E E E
## [4848] E E E E E E E E E E E E E E E E A A A B B B A A A A A C C C C B B B B B B
## [4885] B B B B B C C C C A A A A A A B B B B B B B B B B B B C C C A B B B B B B
## [4922] B B B B B B B B B B C A A A B B B B B B B B B B C C C A A A A A A B A A A
## [4959] A A A A C C C C C B A A A A A C C C C C C C C C B B B B B B B B A C C C C
## [4996] C C C C B B B B B B B B B C C C C C C C C C C C C C C C C C C C C C C C C
## [5033] C E E E E E E C C E E E E E E E E E E A A A A A C E E E E E E E E B B B B
## [5070] E E E E E E E E E B B B B B E E E E E E E E E E E E E E E E E E E E E E E
## [5107] E E E E E E E E E E E E E E E E E E E E E E E E E E E E E B B A A A E E E
## [5144] E E E E E E E E E E E E B B E E E E E E E E E E E B B B E E E E E E E E E
## [5181] E E E E E B B B E E E E E E E E E E E E E E E E E E E E E E E E E E E E E
## [5218] E E E E E E E E E E E E E E B B B B B B E E E E E E E E E E E E A A A A A
## [5255] A A A A C C C C C A A A A A A A A A A A A A A A A A C C A A A A C C C C C
## [5292] C A A A C C C C C C A C C C C A C C C C C C C C A A A C A A C C C C C C C
## [5329] C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C A A A A A A
## [5366] A A A A A A A A A A A A A A A A A A A A C C C C C C C C C C C C C C C C C
## [5403] C C C A A C C C C C C C C C A A A A A C C C C C C C C C C C C B B B B C C
## [5440] C C C C B B B C C C C C C C B B B B C C C C C C B B B B B C C C C C C C C
## [5477] C B B B B B B B C C C C C C C B B B B C B C C B B B C C C C C C B B C A A
## [5514] A E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E
## [5551] E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E
## [5588] E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E
## [5625] E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E C C
## [5662] C C C E E E E E E E E E E E E E E E E E E E E E E E E C C E E E E E E E E
## [5699] E E E E E E E E E E E E E E E E E E E E E E E E E C C C C C C C C C C C C
## [5736] C C C C C C C C C E E E E E E E E E E E E E C C C C B B B B B C C C C C C
## [5773] C C C C C C C C C C C C C A A A A A A A A A C C C C C C C C A A A A A A A
## [5810] A A E E E E E E E E E E E E E E E E E E E E E E E E E A A C E E E E E E E
## [5847] E C E E E E C C C B B B B B A A A B B C C C C C A A A A A A B B B E E E E
## [5884] E E
## Levels: A B C D E
```

```
(conf_rpart <- confusionMatrix(as.factor(valid$classe), predict_rpart))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1510   27  134    0    3
##          B  464  410  265    0    0
##          C  467   37  522    0    0
##          D  432  163  369    0    0
##          E  164  144  293    0  481
##
## Overall Statistics
##
##                Accuracy : 0.4967
##                  95% CI : (0.4838, 0.5095)
##     No Information Rate : 0.5161
##     P-Value [Acc > NIR] : 0.9986
##
##                   Kappa : 0.3425
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.4972  0.52497   0.3298       NA  0.99380
## Specificity            0.9424  0.85717   0.8828   0.8362  0.88872
## Pos Pred Value         0.9020  0.35996   0.5088       NA  0.44455
## Neg Pred Value         0.6374  0.92183   0.7816       NA  0.99938
## Prevalence             0.5161  0.13271   0.2690   0.0000  0.08224
## Detection Rate         0.2566  0.06967   0.0887   0.0000  0.08173
## Detection Prevalence   0.2845  0.19354   0.1743   0.1638  0.18386
## Balanced Accuracy      0.7198  0.69107   0.6063       NA  0.94126
```

```
(accuracy_rpart <- conf_rpart$overall[1])
```

```
##  Accuracy
## 0.4966865
```

The sensitivity is poor (accuracy rate of ~ 0.5) and thus classification tree does not predict very well.

# Random Forest

Now, We build a Random Forest model with a 5 fold cross validation and fit the training data on to the model. We then print the results for review.

```
control <- trainControl(method = "cv", number = 5)
fit_rf <- train(classe ~ ., data = train, method = "rf",
                   trControl = control)
print(fit_rf, digits = 4)
```

```
## Random Forest
##
## 13737 samples
##    52 predictor
##     5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 10991, 10990, 10990, 10988, 10989
## Resampling results across tuning parameters:
##
##   mtry  Accuracy  Kappa
##    2    0.9913    0.9889
##   27    0.9921    0.9900
##   52    0.9840    0.9797
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 27.
```

```
# predict outcomes using validation set
predict_rf <- predict(fit_rf, valid)
# Show prediction result
(conf_rf <- confusionMatrix(as.factor(valid$classe), predict_rf))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1671    2    0    0    1
##          B    4 1134    1    0    0
##          C    0    8 1013    5    0
##          D    0    0   16  947    1
##          E    0    1    1    9 1071
##
## Overall Statistics
##
##                Accuracy : 0.9917
##                  95% CI : (0.989, 0.9938)
##     No Information Rate : 0.2846
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9895
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9976   0.9904   0.9825   0.9854   0.9981
## Specificity            0.9993   0.9989   0.9973   0.9965   0.9977
## Pos Pred Value         0.9982   0.9956   0.9873   0.9824   0.9898
## Neg Pred Value         0.9991   0.9977   0.9963   0.9972   0.9996
## Prevalence             0.2846   0.1946   0.1752   0.1633   0.1823
## Detection Rate         0.2839   0.1927   0.1721   0.1609   0.1820
## Detection Prevalence   0.2845   0.1935   0.1743   0.1638   0.1839
## Balanced Accuracy      0.9984   0.9947   0.9899   0.9910   0.9979
```

```
(accuracy_rf <- conf_rf$overall[1])
```

```
##  Accuracy
## 0.9916737
```

Random forest is a better classifier with better accuracy compared to classification tree.

We will now do the predictions for the testing set using random forest.

```
(predict(fit_rf, testData))
```

```
##  [1] B A B A A E D B A A B C B A E E A B B B
## Levels: A B C D E
```

```
R.version
```

```
##                _
## platform       x86_64-w64-mingw32
## arch           x86_64
## os             mingw32
## system         x86_64, mingw32
## status
## major          4
## minor          0.0
## year           2020
## month          04
## day            24
## svn rev        78286
## language       R
## version.string R version 4.0.0 (2020-04-24)
## nickname       Arbor Day
```

References:

[1] Velloso, E.; Bulling, A.; Gellersen, H.; Ugulino, W.; Fuks, H. Qualitative Activity Recognition of Weight Lifting Exercises. Proceedings of 4th International Conference in Cooperation with SIGCHI (Augmented Human '13) . Stuttgart, Germany: ACM SIGCHI, 2013.