Practical Machine Learning Project

Shreyash Mishra

August 6, 2020

Loading required libraries

```
library(caret)
```

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

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

```
library(rpart)
library(rpart.plot)
library(randomForest)
```

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

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

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

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

```
library(rattle)
```

```
## 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.
```

```
##
## Attaching package: 'rattle'
```

```
## The following object is masked from 'package:randomForest':
##
##     importance
```

Taking in the data

```
set.seed(1)
training <- read.csv("pml-training.csv")
testing <- read.csv("pml-testing.csv")
```

Exploring the data

```
dim(training)
```

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

```
dim(testing)
```

```
## [1]  20 160
```

```
head(training)
```

```
##   X user_name raw_timestamp_part_1 raw_timestamp_part_2   cvtd_timestamp
## 1 1  carlitos           1323084231               788290 05/12/2011 11:23
## 2 2  carlitos           1323084231               808298 05/12/2011 11:23
## 3 3  carlitos           1323084231               820366 05/12/2011 11:23
## 4 4  carlitos           1323084232               120339 05/12/2011 11:23
## 5 5  carlitos           1323084232               196328 05/12/2011 11:23
## 6 6  carlitos           1323084232               304277 05/12/2011 11:23
##   new_window num_window roll_belt pitch_belt yaw_belt total_accel_belt
## 1         no         11      1.41       8.07    -94.4                3
## 2         no         11      1.41       8.07    -94.4                3
## 3         no         11      1.42       8.07    -94.4                3
## 4         no         12      1.48       8.05    -94.4                3
## 5         no         12      1.48       8.07    -94.4                3
## 6         no         12      1.45       8.06    -94.4                3
##   kurtosis_roll_belt kurtosis_picth_belt kurtosis_yaw_belt
skewness_roll_belt
## 1
## 2
## 3
## 4
## 5
## 6
##   skewness_roll_belt.1 skewness_yaw_belt max_roll_belt max_picth_belt
## 1                                                   NA             NA
## 2                                                   NA             NA
## 3                                                   NA             NA
## 4                                                   NA             NA
## 5                                                   NA             NA
## 6                                                   NA             NA
##   max_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt
amplitude_roll_belt
## 1                           NA             NA
NA
## 2                           NA             NA
NA
## 3                           NA             NA
NA
## 4                           NA             NA
NA
```

```
## 5                         NA             NA
NA
## 6                         NA             NA
NA
##   amplitude_pitch_belt amplitude_yaw_belt var_total_accel_belt
avg_roll_belt
## 1                   NA                                     NA
NA
## 2                   NA                                     NA
NA
## 3                   NA                                     NA
NA
## 4                   NA                                     NA
NA
## 5                   NA                                     NA
NA
## 6                   NA                                     NA
NA
##   stddev_roll_belt var_roll_belt avg_pitch_belt stddev_pitch_belt
## 1               NA            NA             NA                NA
## 2               NA            NA             NA                NA
## 3               NA            NA             NA                NA
## 4               NA            NA             NA                NA
## 5               NA            NA             NA                NA
## 6               NA            NA             NA                NA
##   var_pitch_belt avg_yaw_belt stddev_yaw_belt var_yaw_belt gyros_belt_x
## 1             NA           NA               NA           NA         0.00
## 2             NA           NA               NA           NA         0.02
## 3             NA           NA               NA           NA         0.00
## 4             NA           NA               NA           NA         0.02
## 5             NA           NA               NA           NA         0.02
## 6             NA           NA               NA           NA         0.02
##   gyros_belt_y gyros_belt_z accel_belt_x accel_belt_y accel_belt_z
## 1         0.00        -0.02          -21            4           22
## 2         0.00        -0.02          -22            4           22
## 3         0.00        -0.02          -20            5           23
## 4         0.00        -0.03          -22            3           21
## 5         0.02        -0.02          -21            2           24
## 6         0.00        -0.02          -21            4           21
##   magnet_belt_x magnet_belt_y magnet_belt_z roll_arm pitch_arm yaw_arm
## 1            -3           599          -313     -128      22.5    -161
## 2            -7           608          -311     -128      22.5    -161
## 3            -2           600          -305     -128      22.5    -161
## 4            -6           604          -310     -128      22.1    -161
## 5            -6           600          -302     -128      22.1    -161
## 6             0           603          -312     -128      22.0    -161
##   total_accel_arm var_accel_arm avg_roll_arm stddev_roll_arm var_roll_arm
## 1              34            NA           NA              NA           NA
## 2              34            NA           NA              NA           NA
## 3              34            NA           NA              NA           NA
```

```
## 4              34            NA           NA               NA           NA
## 5              34            NA           NA               NA           NA
## 6              34            NA           NA               NA           NA
##   avg_pitch_arm stddev_pitch_arm var_pitch_arm avg_yaw_arm stddev_yaw_arm
## 1            NA               NA            NA          NA              NA
## 2            NA               NA            NA          NA              NA
## 3            NA               NA            NA          NA              NA
## 4            NA               NA            NA          NA              NA
## 5            NA               NA            NA          NA              NA
## 6            NA               NA            NA          NA              NA
##   var_yaw_arm gyros_arm_x gyros_arm_y gyros_arm_z accel_arm_x accel_arm_y
## 1          NA        0.00        0.00       -0.02        -288         109
## 2          NA        0.02       -0.02       -0.02        -290         110
## 3          NA        0.02       -0.02       -0.02        -289         110
## 4          NA        0.02       -0.03        0.02        -289         111
## 5          NA        0.00       -0.03        0.00        -289         111
## 6          NA        0.02       -0.03        0.00        -289         111
##   accel_arm_z magnet_arm_x magnet_arm_y magnet_arm_z kurtosis_roll_arm
## 1        -123         -368          337          516
## 2        -125         -369          337          513
## 3        -126         -368          344          513
## 4        -123         -372          344          512
## 5        -123         -374          337          506
## 6        -122         -369          342          513
##   kurtosis_picth_arm kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm
## 1
## 2
## 3
## 4
## 5
## 6
##   skewness_yaw_arm max_roll_arm max_picth_arm max_yaw_arm min_roll_arm
## 1                           NA            NA          NA           NA
## 2                           NA            NA          NA           NA
## 3                           NA            NA          NA           NA
## 4                           NA            NA          NA           NA
## 5                           NA            NA          NA           NA
## 6                           NA            NA          NA           NA
##   min_pitch_arm min_yaw_arm amplitude_roll_arm amplitude_pitch_arm
## 1            NA          NA                NA                 NA
## 2            NA          NA                NA                 NA
## 3            NA          NA                NA                 NA
## 4            NA          NA                NA                 NA
## 5            NA          NA                NA                 NA
## 6            NA          NA                NA                 NA
##   amplitude_yaw_arm roll_dumbbell pitch_dumbbell yaw_dumbbell
## 1                NA      13.05217      -70.49400    -84.87394
## 2                NA      13.13074      -70.63751    -84.71065
## 3                NA      12.85075      -70.27812    -85.14078
## 4                NA      13.43120      -70.39379    -84.87363
```

```
## 5                NA      13.37872      -70.42856    -84.85306
## 6                NA      13.38246      -70.81759    -84.46500
##   kurtosis_roll_dumbbell kurtosis_picth_dumbbell kurtosis_yaw_dumbbell
## 1
## 2
## 3
## 4
## 5
## 6
##   skewness_roll_dumbbell skewness_pitch_dumbbell skewness_yaw_dumbbell
## 1
## 2
## 3
## 4
## 5
## 6
##   max_roll_dumbbell max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell
## 1                NA                 NA                                NA
## 2                NA                 NA                                NA
## 3                NA                 NA                                NA
## 4                NA                 NA                                NA
## 5                NA                 NA                                NA
## 6                NA                 NA                                NA
##   min_pitch_dumbbell min_yaw_dumbbell amplitude_roll_dumbbell
## 1                 NA                                      NA
## 2                 NA                                      NA
## 3                 NA                                      NA
## 4                 NA                                      NA
## 5                 NA                                      NA
## 6                 NA                                      NA
##   amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
## 1                       NA                                         37
## 2                       NA                                         37
## 3                       NA                                         37
## 4                       NA                                         37
## 5                       NA                                         37
## 6                       NA                                         37
##   var_accel_dumbbell avg_roll_dumbbell stddev_roll_dumbbell
var_roll_dumbbell
## 1                 NA                NA                    NA
NA
## 2                 NA                NA                    NA
NA
## 3                 NA                NA                    NA
NA
## 4                 NA                NA                    NA
NA
## 5                 NA                NA                    NA
NA
## 6                 NA                NA                    NA
```

```
NA
##   avg_pitch_dumbbell stddev_pitch_dumbbell var_pitch_dumbbell
avg_yaw_dumbbell
## 1                 NA                    NA                 NA
NA
## 2                 NA                    NA                 NA
NA
## 3                 NA                    NA                 NA
NA
## 4                 NA                    NA                 NA
NA
## 5                 NA                    NA                 NA
NA
## 6                 NA                    NA                 NA
NA
##   stddev_yaw_dumbbell var_yaw_dumbbell gyros_dumbbell_x gyros_dumbbell_y
## 1                  NA                NA                 0            -0.02
## 2                  NA                NA                 0            -0.02
## 3                  NA                NA                 0            -0.02
## 4                  NA                NA                 0            -0.02
## 5                  NA                NA                 0            -0.02
## 6                  NA                NA                 0            -0.02
##   gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y accel_dumbbell_z
## 1             0.00             -234               47             -271
## 2             0.00             -233               47             -269
## 3             0.00             -232               46             -270
## 4            -0.02             -232               48             -269
## 5             0.00             -233               48             -270
## 6             0.00             -234               48             -269
##   magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z roll_forearm
## 1              -559               293               -65         28.4
## 2              -555               296               -64         28.3
## 3              -561               298               -63         28.3
## 4              -552               303               -60         28.1
## 5              -554               292               -68         28.0
## 6              -558               294               -66         27.9
##   pitch_forearm yaw_forearm kurtosis_roll_forearm kurtosis_picth_forearm
## 1         -63.9        -153
## 2         -63.9        -153
## 3         -63.9        -152
## 4         -63.9        -152
## 5         -63.9        -152
## 6         -63.9        -152
##   kurtosis_yaw_forearm skewness_roll_forearm skewness_pitch_forearm
## 1
## 2
## 3
## 4
## 5
## 6
```

```
##   skewness_yaw_forearm max_roll_forearm max_picth_forearm max_yaw_forearm
## 1                                   NA                 NA
## 2                                   NA                 NA
## 3                                   NA                 NA
## 4                                   NA                 NA
## 5                                   NA                 NA
## 6                                   NA                 NA
##   min_roll_forearm min_pitch_forearm min_yaw_forearm
amplitude_roll_forearm
## 1               NA                NA
NA
## 2               NA                NA
NA
## 3               NA                NA
NA
## 4               NA                NA
NA
## 5               NA                NA
NA
## 6               NA                NA
NA
##   amplitude_pitch_forearm amplitude_yaw_forearm total_accel_forearm
## 1                      NA                                       36
## 2                      NA                                       36
## 3                      NA                                       36
## 4                      NA                                       36
## 5                      NA                                       36
## 6                      NA                                       36
##   var_accel_forearm avg_roll_forearm stddev_roll_forearm var_roll_forearm
## 1                NA               NA                  NA               NA
## 2                NA               NA                  NA               NA
## 3                NA               NA                  NA               NA
## 4                NA               NA                  NA               NA
## 5                NA               NA                  NA               NA
## 6                NA               NA                  NA               NA
##   avg_pitch_forearm stddev_pitch_forearm var_pitch_forearm avg_yaw_forearm
## 1                NA                   NA                NA              NA
## 2                NA                   NA                NA              NA
## 3                NA                   NA                NA              NA
## 4                NA                   NA                NA              NA
## 5                NA                   NA                NA              NA
## 6                NA                   NA                NA              NA
##   stddev_yaw_forearm var_yaw_forearm gyros_forearm_x gyros_forearm_y
## 1                 NA              NA            0.03            0.00
## 2                 NA              NA            0.02            0.00
## 3                 NA              NA            0.03           -0.02
## 4                 NA              NA            0.02           -0.02
## 5                 NA              NA            0.02            0.00
## 6                 NA              NA            0.02           -0.02
##   gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z
```

```
## 1           -0.02             192             203            -215
## 2           -0.02             192             203            -216
## 3            0.00             196             204            -213
## 4            0.00             189             206            -214
## 5           -0.02             189             206            -214
## 6           -0.03             193             203            -215
##   magnet_forearm_x magnet_forearm_y magnet_forearm_z classe
## 1              -17              654              476      A
## 2              -18              661              473      A
## 3              -18              658              469      A
## 4              -16              658              469      A
## 5              -17              655              473      A
## 6               -9              660              478      A
```

Checking if the training dataset has any null values,

```
sum(complete.cases(training))
```

```
## [1] 406
```

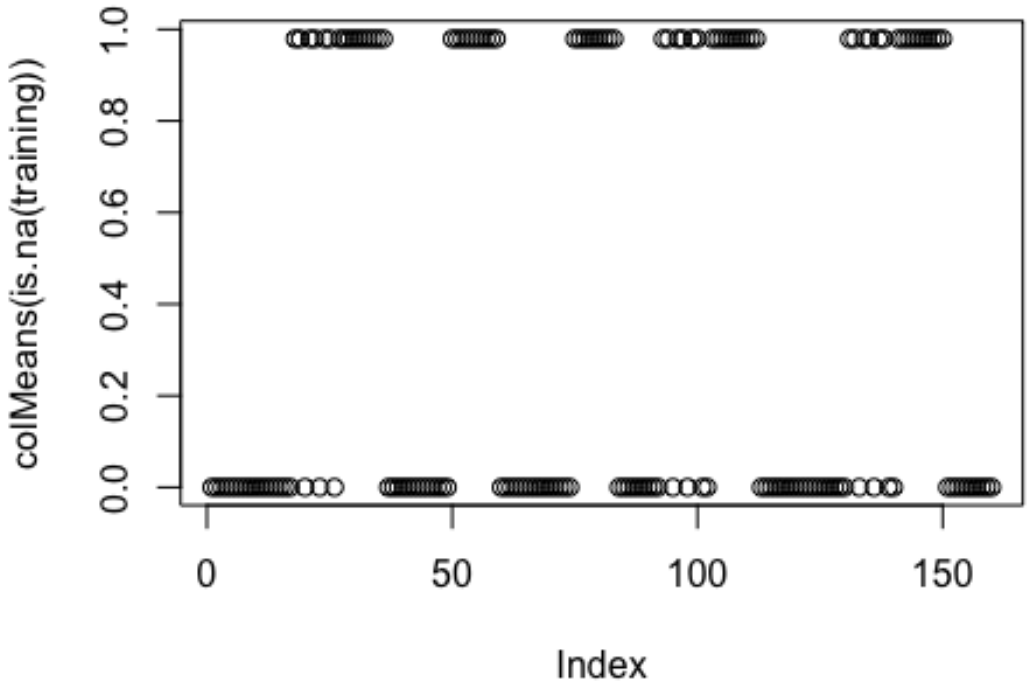
```
colnames(training)
```

```
##   [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"

plot(colMeans(is.na(training)))
```



There are columns with a lot of missing values.
We will reatain only the columns without NA values

Cleaning the data

```
features <- names(testing[,colSums(is.na(testing)) == 0])[8:59]
trainclasse <- training[,c(features,"classe")]
testproblem <- testing[,c(features,"problem_id")]
```

Partitioning the data

In order to evaluate our model before submitting it for grading, we'll designate a partition of it for validation.

```
inTrain <- createDataPartition(trainclasse$classe, p=0.7, list = FALSE)
myTraining <- trainclasse[inTrain,]
myTesting <- trainclasse[-inTrain,]
```

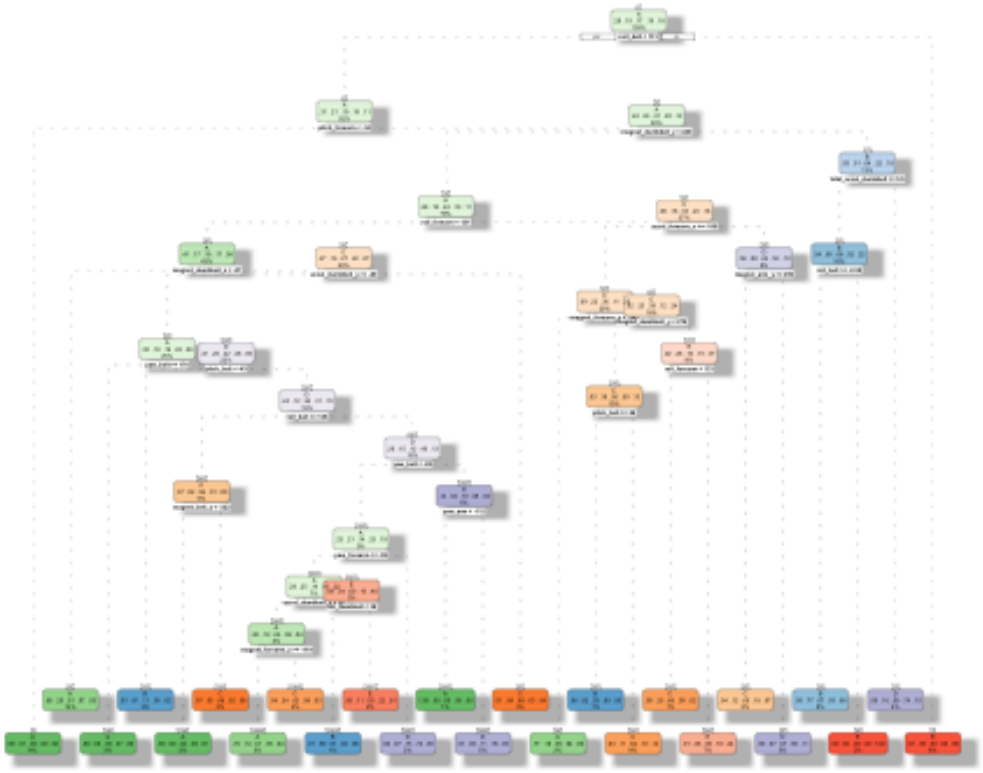
Modeling the data

We will fit a model using **Decision Tree** and **Random Forest**

Decision Tree Prediction

```
set.seed(12345)
DTmodel <- rpart(classe ~ ., data = myTraining, method = "class")
fancyRpartPlot(DTmodel)

## Warning: labs do not fit even at cex 0.15, there may be some overplotting
```




```
DTpredict <- predict(DTmodel, myTesting, type = "class")
confusionMatrix(DTpredict, myTesting$classe)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1554  230   16   73   36
##          B   30  564   47   14   60
##          C   53  150  844  109  109
##          D   20   81   81  669   92
##          E   17  114   38   99  785
```

```
##
## Overall Statistics
##
##                Accuracy : 0.7504
##                  95% CI : (0.7391, 0.7614)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6831
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9283  0.49517   0.8226   0.6940   0.7255
## Specificity            0.9157  0.96818   0.9134   0.9443   0.9442
## Pos Pred Value         0.8140  0.78881   0.6672   0.7094   0.7455
## Neg Pred Value         0.9698  0.88878   0.9606   0.9403   0.9385
## Prevalence             0.2845  0.19354   0.1743   0.1638   0.1839
## Detection Rate         0.2641  0.09584   0.1434   0.1137   0.1334
## Detection Prevalence   0.3244  0.12150   0.2150   0.1602   0.1789
## Balanced Accuracy      0.9220  0.73168   0.8680   0.8192   0.8349
```

Random Forest Prediction

```
RFmodel <- randomForest(classe ~ ., data = myTraining)
RFpredict <- predict(RFmodel, myTesting, type = "class")
confusionMatrix(RFpredict, myTesting$classe)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1672    1    0    0    0
##          B    2 1138    9    0    0
##          C    0    0 1012   15    0
##          D    0    0    5  948    2
##          E    0    0    0    1 1080
##
## Overall Statistics
##
##                Accuracy : 0.9941
##                  95% CI : (0.9917, 0.9959)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9925
##
##  Mcnemar's Test P-Value : NA
##
```

```
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9988   0.9991   0.9864   0.9834   0.9982
## Specificity            0.9998   0.9977   0.9969   0.9986   0.9998
## Pos Pred Value         0.9994   0.9904   0.9854   0.9927   0.9991
## Neg Pred Value         0.9995   0.9998   0.9971   0.9968   0.9996
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2841   0.1934   0.1720   0.1611   0.1835
## Detection Prevalence   0.2843   0.1952   0.1745   0.1623   0.1837
## Balanced Accuracy      0.9993   0.9984   0.9916   0.9910   0.9990
```

Since the random forest model's accuracy was 99.3%, the out of sample error is 0.007. We will use the random forest model to submit our predictions.

```
FinalPredict <- predict(RFmodel, testing, type = "class")
FinalPredict
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  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
```

```
pml_write_files = function(x){
    n = length(x)
    for(i in 1:n){
        filename = paste0("problem_id_",i,".txt")

write.table(x[i],file=filename,quote=FALSE,row.names=FALSE,col.names=FALSE)
    }
}

pml_write_files(FinalPredict)
```