

Accelerometer Data Analysis and Prediction

2024-01-17

#Overview In this project, we will use data from accelerometers on the belt, forearm, arm, and dumbbell of 6 participants to predict the manner in which they did the exercise. This is the “classe” variable in the training set. We train 3 models: Decision Tree, Random Forest, Support Vector Machine using k-folds cross validation on the training set. We then predict using a validation set randomly selected from the training csv data to obtain the accuracy and out of sample error rate. Based on those numbers, we decide on the best model, and use it to predict 20 cases using the test csv set.

#Background Using devices such as Jawbone Up, Nike FuelBand, and Fitbit it is now possible to collect a large amount of data about personal activity relatively inexpensively. These type of devices are part of the quantified self movement - a group of enthusiasts who take measurements about themselves regularly to improve their health, to find patterns in their behavior, or because they are tech geeks. One thing that people regularly do is quantify how much of a particular activity they do, but they rarely quantify how well they do it. In this project, the goal will be to use data from accelerometers on the belt, forearm, arm, and dumbbell of 6 participants. They were asked to perform barbell lifts correctly and incorrectly in 5 different ways.

The training data for this project are available here:

<https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv>
(<https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv>)

The test data are available here:

<https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv>
(<https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv>)

#Data Processing ##Loading Necessary Libraries

```
# Load Libraries
library(caret)
```

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

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

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

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

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

```
## Warning: package 'rpart.plot' 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
```

```
library(corrplot)
```

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

```
## corrplot 0.92 loaded
```

```
library(ggplot2)
```

##Loading Data

```
# read data from csv files  
train <- read.csv("pml-training.csv", na.strings=c("NA", "#DIV/0!", ""))  
test <- read.csv("pml-testing.csv", na.strings=c("NA", "#DIV/0!", ""))  
  
head(train, n=5)
```

```

## 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
## 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
## kurtosis_roll_belt kurtosis_pitch_belt kurtosis_yaw_belt skewness_roll_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
## skewness_roll_belt.1 skewness_yaw_belt max_roll_belt max_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
## max_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt amplitude_roll_belt
## 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
## amplitude_pitch_belt amplitude_yaw_belt var_total_accel_belt avg_roll_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
## 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
## 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
## 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
## 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
## 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
## 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
## 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
## accel_arm_z magnet_arm_x magnet_arm_y magnet_arm_z kurtosis_roll_arm
## 1      -123      -368      337      516      NA
## 2      -125      -369      337      513      NA
## 3      -126      -368      344      513      NA
## 4      -123      -372      344      512      NA
## 5      -123      -374      337      506      NA
## kurtosis_pitch_arm kurtosis_yaw_arm skewness_roll_arm skewness_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
## skewness_yaw_arm max_roll_arm max_pitch_arm max_yaw_arm min_roll_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
## 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
## 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
## kurtosis_roll_dumbbell kurtosis_pitch_dumbbell kurtosis_yaw_dumbbell
## 1      NA      NA      NA
## 2      NA      NA      NA

```

```

## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## skewness_roll_dumbbell skewness_pitch_dumbbell skewness_yaw_dumbbell
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## max_roll_dumbbell max_pitch_dumbbell max_yaw_dumbbell min_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
## min_pitch_dumbbell min_yaw_dumbbell amplitude_roll_dumbbell
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
## 1          NA          NA          37
## 2          NA          NA          37
## 3          NA          NA          37
## 4          NA          NA          37
## 5          NA          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
## 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
## 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
## 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
## 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
## pitch_forearm yaw_forearm kurtosis_roll_forearm kurtosis_pitch_forearm
## 1          -63.9          -153          NA          NA
## 2          -63.9          -153          NA          NA
## 3          -63.9          -152          NA          NA
## 4          -63.9          -152          NA          NA
## 5          -63.9          -152          NA          NA
## kurtosis_yaw_forearm skewness_roll_forearm skewness_pitch_forearm
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## skewness_yaw_forearm max_roll_forearm max_pitch_forearm max_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
## min_roll_forearm min_pitch_forearm min_yaw_forearm amplitude_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
## amplitude_pitch_forearm amplitude_yaw_forearm total_accel_forearm
## 1          NA          NA          36
## 2          NA          NA          36
## 3          NA          NA          36
## 4          NA          NA          36
## 5          NA          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
## 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
## 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
## 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
## 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 |

```
head(test,n=5)
```

```

## X user_name raw_timestamp_part_1 raw_timestamp_part_2 cvtd_timestamp
## 1 1 pedro 1323095002 868349 05/12/2011 14:23
## 2 2 jeremy 1322673067 778725 30/11/2011 17:11
## 3 3 jeremy 1322673075 342967 30/11/2011 17:11
## 4 4 adelmo 1322832789 560311 02/12/2011 13:33
## 5 5 eurico 1322489635 814776 28/11/2011 14:13
## new_window num_window roll_belt pitch_belt yaw_belt total_accel_belt
## 1 no 74 123.00 27.00 -4.75 20
## 2 no 431 1.02 4.87 -88.90 4
## 3 no 439 0.87 1.82 -88.50 5
## 4 no 194 125.00 -41.60 162.00 17
## 5 no 235 1.35 3.33 -88.60 3
## kurtosis_roll_belt kurtosis_pitch_belt kurtosis_yaw_belt skewness_roll_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
## skewness_roll_belt.1 skewness_yaw_belt max_roll_belt max_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
## max_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt amplitude_roll_belt
## 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
## amplitude_pitch_belt amplitude_yaw_belt var_total_accel_belt avg_roll_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
## 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
## var_pitch_belt avg_yaw_belt stddev_yaw_belt var_yaw_belt gyros_belt_x
## 1 NA NA NA NA -0.50
## 2 NA NA NA NA -0.06
## 3 NA NA NA NA 0.05
## 4 NA NA NA NA 0.11
## 5 NA NA NA NA 0.03
## gyros_belt_y gyros_belt_z accel_belt_x accel_belt_y accel_belt_z
## 1 -0.02 -0.46 -38 69 -179
## 2 -0.02 -0.07 -13 11 39
## 3 0.02 0.03 1 -1 49
## 4 0.11 -0.16 46 45 -156
## 5 0.02 0.00 -8 4 27
## magnet_belt_x magnet_belt_y magnet_belt_z roll_arm pitch_arm yaw_arm

```



```

## 1      -13      581      -382      40.7      -27.80      178
## 2       43      636      -309       0.0       0.00       0
## 3       29      631      -312       0.0       0.00       0
## 4      169      608      -304     -109.0      55.00     -142
## 5       33      566      -418      76.1       2.76      102
##  total_accel_arm var_accel_arm avg_roll_arm stddev_roll_arm var_roll_arm
## 1          10          NA          NA          NA          NA
## 2          38          NA          NA          NA          NA
## 3          44          NA          NA          NA          NA
## 4          25          NA          NA          NA          NA
## 5          29          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
##  var_yaw_arm gyros_arm_x gyros_arm_y gyros_arm_z accel_arm_x accel_arm_y
## 1          NA      -1.65       0.48      -0.18          16          38
## 2          NA      -1.17       0.85      -0.43         -290         215
## 3          NA       2.10      -1.36       1.13         -341         245
## 4          NA       0.22      -0.51       0.92         -238         -57
## 5          NA      -1.96       0.79      -0.54         -197         200
##  accel_arm_z magnet_arm_x magnet_arm_y magnet_arm_z kurtosis_roll_arm
## 1          93      -326       385       481          NA
## 2         -90      -325       447       434          NA
## 3         -87      -264       474       413          NA
## 4           6      -173       257       633          NA
## 5         -30      -170       275       617          NA
##  kurtosis_pitch_arm kurtosis_yaw_arm skewness_roll_arm skewness_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
##  skewness_yaw_arm max_roll_arm max_pitch_arm max_yaw_arm min_roll_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
##  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
##  amplitude_yaw_arm roll_dumbbell pitch_dumbbell yaw_dumbbell
## 1          NA     -17.73748      24.96085     126.23596
## 2          NA     54.47761     -53.69758     -75.51480
## 3          NA     57.07031     -51.37303     -75.20287
## 4          NA     43.10927     -30.04885     -103.32003
## 5          NA    -101.38396     -53.43952     -14.19542
##  kurtosis_roll_dumbbell kurtosis_pitch_dumbbell kurtosis_yaw_dumbbell
## 1          NA          NA          NA
## 2          NA          NA          NA

```

```

## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## skewness_roll_dumbbell skewness_pitch_dumbbell skewness_yaw_dumbbell
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## max_roll_dumbbell max_pitch_dumbbell max_yaw_dumbbell min_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
## min_pitch_dumbbell min_yaw_dumbbell amplitude_roll_dumbbell
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
## 1          NA          NA          9
## 2          NA          NA          31
## 3          NA          NA          29
## 4          NA          NA          18
## 5          NA          NA          4
## 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
## 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
## stddev_yaw_dumbbell var_yaw_dumbbell gyros_dumbbell_x gyros_dumbbell_y
## 1          NA          NA          0.64          0.06
## 2          NA          NA          0.34          0.05
## 3          NA          NA          0.39          0.14
## 4          NA          NA          0.10         -0.02
## 5          NA          NA          0.29         -0.47
## gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y accel_dumbbell_z
## 1         -0.61          21          -15          81
## 2         -0.71         -153          155         -205
## 3         -0.34         -141          155         -196
## 4          0.05          -51           72         -148
## 5         -0.46          -18          -30           -5
## magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z roll_forearm
## 1          523         -528          -56          141
## 2         -502          388          -36          109
## 3         -506          349           41          131
## 4         -576          238           53           0

```

| | -424 | 252 | 312 | -176 |
|--|--------|-------|-------|-------|
| ## 5 | | | | |
| ## pitch_forearm yaw_forearm kurtosis_roll_forearm kurtosis_pitch_forearm | | | | |
| ## 1 | 49.30 | 156.0 | NA | NA |
| ## 2 | -17.60 | 106.0 | NA | NA |
| ## 3 | -32.60 | 93.0 | NA | NA |
| ## 4 | 0.00 | 0.0 | NA | NA |
| ## 5 | -2.16 | -47.9 | NA | NA |
| ## kurtosis_yaw_forearm skewness_roll_forearm skewness_pitch_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 |
| ## skewness_yaw_forearm max_roll_forearm max_pitch_forearm max_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 |
| ## min_roll_forearm min_pitch_forearm min_yaw_forearm amplitude_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 |
| ## amplitude_pitch_forearm amplitude_yaw_forearm total_accel_forearm | | | | |
| ## 1 | NA | NA | 33 | |
| ## 2 | NA | NA | 39 | |
| ## 3 | NA | NA | 34 | |
| ## 4 | NA | NA | 43 | |
| ## 5 | NA | NA | 24 | |
| ## 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 |
| ## 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 |
| ## stddev_yaw_forearm var_yaw_forearm gyros_forearm_x gyros_forearm_y | | | | |
| ## 1 | NA | NA | 0.74 | -3.34 |
| ## 2 | NA | NA | 1.12 | -2.78 |
| ## 3 | NA | NA | 0.18 | -0.79 |
| ## 4 | NA | NA | 1.38 | 0.69 |
| ## 5 | NA | NA | -0.75 | 3.10 |
| ## gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z | | | | |
| ## 1 | -0.59 | -110 | 267 | -149 |
| ## 2 | -0.18 | 212 | 297 | -118 |
| ## 3 | 0.28 | 154 | 271 | -129 |
| ## 4 | 1.80 | -92 | 406 | -39 |
| ## 5 | 0.80 | 131 | -93 | 172 |
| ## magnet_forearm_x magnet_forearm_y magnet_forearm_z problem_id | | | | |

| | | | | |
|------|------|------|-----|---|
| ## 1 | -714 | 419 | 617 | 1 |
| ## 2 | -237 | 791 | 873 | 2 |
| ## 3 | -51 | 698 | 783 | 3 |
| ## 4 | -233 | 783 | 521 | 4 |
| ## 5 | 375 | -787 | 91 | 5 |

```
names(train)
```

```

## [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_pitch_belt" "kurtosis_yaw_belt"
## [15] "skewness_roll_belt" "skewness_roll_belt.1"
## [17] "skewness_yaw_belt" "max_roll_belt"
## [19] "max_pitch_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_pitch_arm"
## [71] "kurtosis_yaw_arm" "skewness_roll_arm"
## [73] "skewness_pitch_arm" "skewness_yaw_arm"
## [75] "max_roll_arm" "max_pitch_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_pitch_dumbbell"
## [89] "kurtosis_yaw_dumbbell" "skewness_roll_dumbbell"
## [91] "skewness_pitch_dumbbell" "skewness_yaw_dumbbell"
## [93] "max_roll_dumbbell" "max_pitch_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_pitch_forearm"
## [127] "kurtosis_yaw_forearm"    "skewness_roll_forearm"
## [129] "skewness_pitch_forearm"  "skewness_yaw_forearm"
## [131] "max_roll_forearm"        "max_pitch_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"
```

```
dim(train)
```

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

```
dim(test)
```

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

The dataset contains a lot of NA and missing values , therefore it is necessary to perform some data cleaning before proceeding with the training of models.

##Data Cleaning

```
# Clean dataset by removing rows and columns with missing values
trainClean <- train[, colSums(is.na(train)) == 0]
testClean <- test[, colSums(is.na(test)) == 0]

trainClean <- trainClean[, -c(1:7)] #removing metadata which is irrelevant to the outcome
nearZeroVar(trainClean) # no variable with near zero variance
```

```
## integer(0)
```

Now that the data is clean and has all the unwanted variables removed, we can split the data for training the model.

```
# split data into train and test from the training set
set.seed(42)
inTrain <- createDataPartition(y=trainClean$classe, p=0.6, list=FALSE)
myTrain <- trainClean[inTrain, ]
myTest <- trainClean[-inTrain, ]
dim(myTrain)
```

```
## [1] 11776    53
```

```
dim(myTest)
```

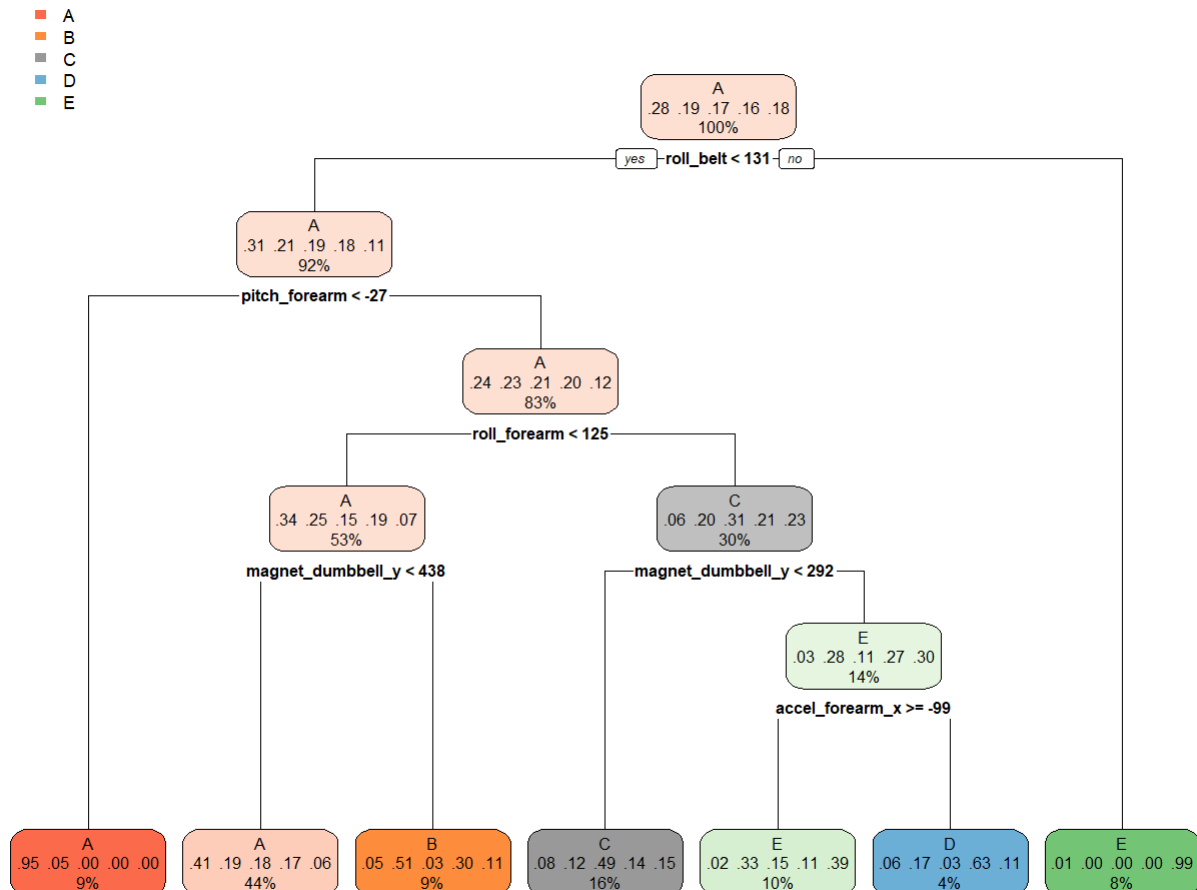
```
## [1] 7846    53
```

##Create Regression Models We will create 3 Regression models and perform a comparative analysis for all three of them:

```
# setting control variable for 4 fold cross validation
control <- trainControl(method="cv", number=4, verboseIter=F)
```

###1.Decision Trees

```
# Fitting Decision Tree Model
mod1 <- train(classe~., data=myTrain, method="rpart", trControl = control, tuneLength = 5)
rpart.plot(mod1$finalModel)
```



```

pred1 <- predict(mod1, myTest)
cm1 <- confusionMatrix(pred1, factor(myTest$classe))
cm1

```

```

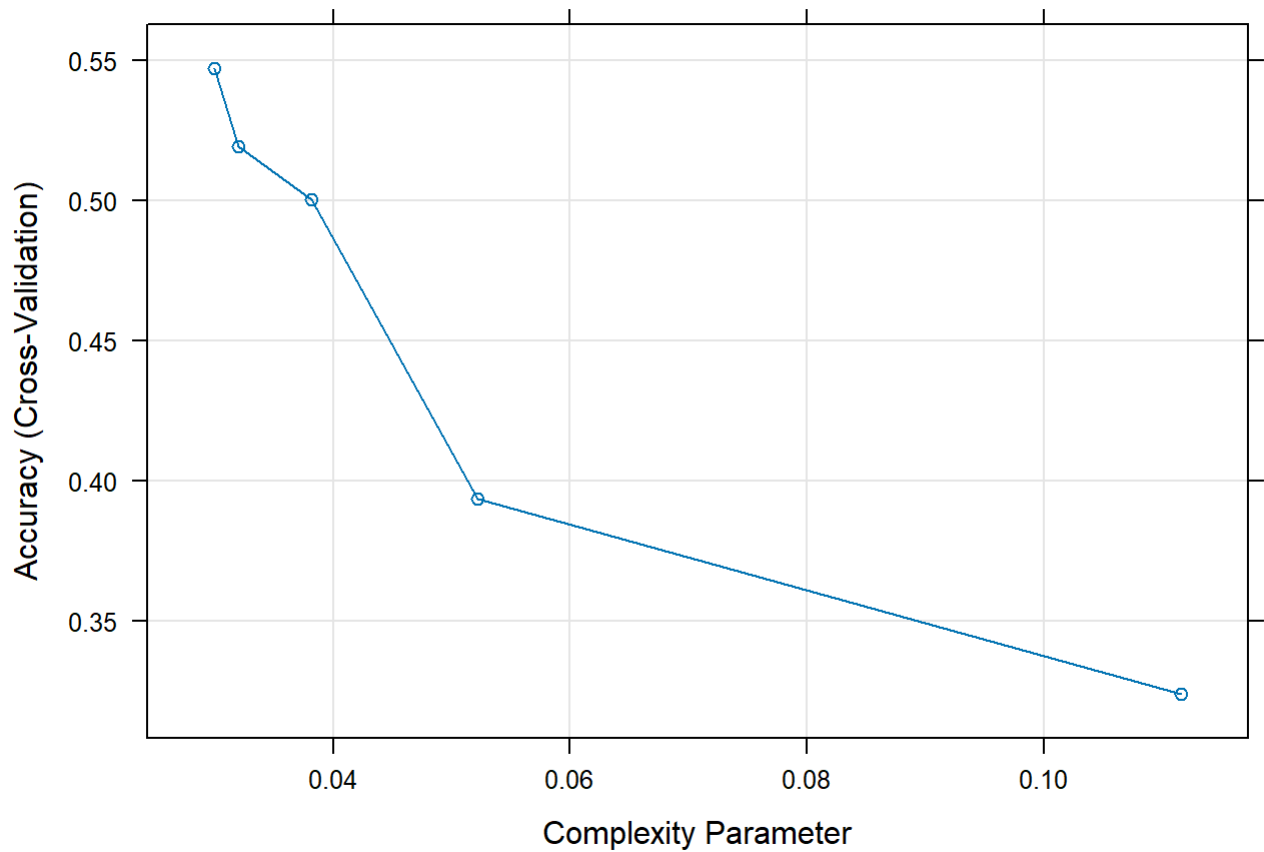
## Confusion Matrix and Statistics
##
##              Reference
## Prediction    A    B    C    D    E
##      A 2027  639  665  576  224
##      B   33  389   20  215   88
##      C  115  165  574  178  153
##      D   21   56   12  241   46
##      E   36  269   97   76  931
##
## Overall Statistics
##
##              Accuracy : 0.5305
##              95% CI : (0.5193, 0.5416)
##      No Information Rate : 0.2845
##      P-Value [Acc > NIR] : < 2.2e-16
##
##              Kappa : 0.3859
##
##  McNemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##              Class: A Class: B Class: C Class: D Class: E
## Sensitivity          0.9082  0.25626  0.41959  0.18740  0.6456
## Specificity          0.6252  0.94374  0.90568  0.97942  0.9254
## Pos Pred Value       0.4907  0.52215  0.48439  0.64096  0.6608
## Neg Pred Value       0.9448  0.84101  0.88080  0.86011  0.9206
## Prevalence          0.2845  0.19347  0.17436  0.16391  0.1838
## Detection Rate       0.2583  0.04958  0.07316  0.03072  0.1187
## Detection Prevalence 0.5265  0.09495  0.15103  0.04792  0.1796
## Balanced Accuracy    0.7667  0.60000  0.66264  0.58341  0.7855

```

```

plot(mod1)

```

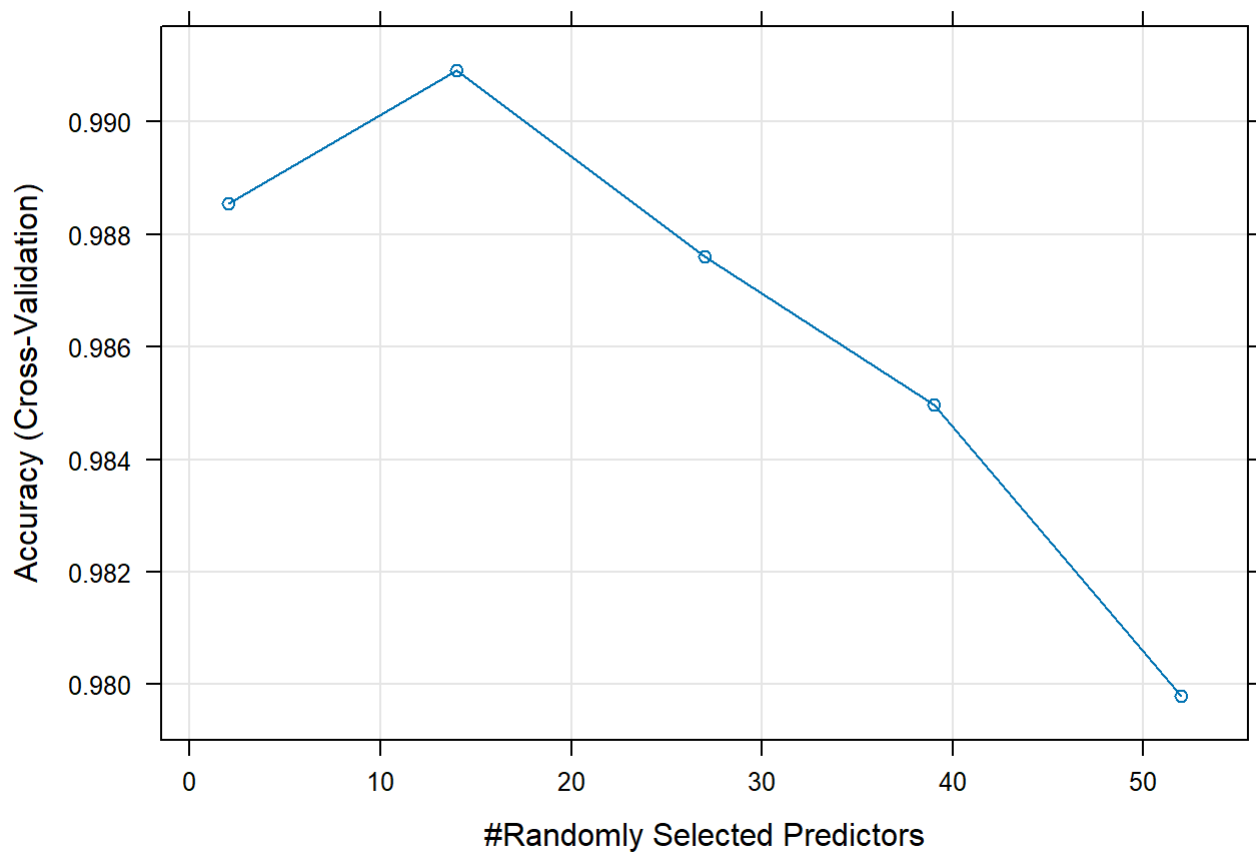



###2.Random Forest

```
# Fitting Random Forest Model
mod2 <- train(classe~., data=myTrain, method="rf", trControl = control, tuneLength = 5)
pred2 <- predict(mod2, myTest)
cm2 <- confusionMatrix(pred2, factor(myTest$classe))
cm2
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A    B    C    D    E
##           A 2226    9    0    0    0
##           B   5 1508    9    0    0
##           C    1    1 1355    8    0
##           D    0    0    4 1273    1
##           E    0    0    0    5 1441
##
## Overall Statistics
##
##           Accuracy : 0.9945
##           95% CI : (0.9926, 0.996)
##           No Information Rate : 0.2845
##           P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.9931
##
##           McNemar's Test P-Value : NA
##
## Statistics by Class:
##
##           Class: A Class: B Class: C Class: D Class: E
## Sensitivity      0.9973  0.9934  0.9905  0.9899  0.9993
## Specificity      0.9984  0.9978  0.9985  0.9992  0.9992
## Pos Pred Value   0.9960  0.9908  0.9927  0.9961  0.9965
## Neg Pred Value   0.9989  0.9984  0.9980  0.9980  0.9998
## Prevalence       0.2845  0.1935  0.1744  0.1639  0.1838
## Detection Rate   0.2837  0.1922  0.1727  0.1622  0.1837
## Detection Prevalence 0.2849  0.1940  0.1740  0.1629  0.1843
## Balanced Accuracy 0.9979  0.9956  0.9945  0.9946  0.9993
```

```
plot(mod2)
```



###2.Support Vector Machine

```
# Fitting SVM Model
mod3 <- train(classe~., data=myTrain, method="svmLinear", trControl = control, tuneLength =
5, verbose = F)
pred3 <- predict(mod3, myTest)
cm3 <- confusionMatrix(pred3, factor(myTest$classe))
cm3
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A    B    C    D    E
##           A 2072  206  116   85   70
##           B   43 1086  129   50  183
##           C   57   77 1064  134  108
##           D   51   39   31  958   68
##           E    9  110   28   59 1013
##
## Overall Statistics
##
##           Accuracy : 0.7893
##           95% CI : (0.7801, 0.7983)
##           No Information Rate : 0.2845
##           P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.7321
##
##           McNemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##           Class: A Class: B Class: C Class: D Class: E
## Sensitivity           0.9283   0.7154   0.7778   0.7449   0.7025
## Specificity           0.9150   0.9360   0.9420   0.9712   0.9678
## Pos Pred Value        0.8129   0.7284   0.7389   0.8352   0.8310
## Neg Pred Value        0.9698   0.9320   0.9525   0.9510   0.9353
## Prevalence            0.2845   0.1935   0.1744   0.1639   0.1838
## Detection Rate        0.2641   0.1384   0.1356   0.1221   0.1291
## Detection Prevalence  0.3249   0.1900   0.1835   0.1462   0.1554
## Balanced Accuracy      0.9217   0.8257   0.8599   0.8581   0.8352
```

##Comparative Analysis Now, we will compare the accuracy and Out of Sample Error for all the three models

```
# Compile Model Results
ModelName <- c("Decision Tree","Random Forest","SVM")
OOSE <- c(1 - as.numeric(cm1$overall[1]),1 - as.numeric(cm2$overall[1]),1 - as.numeric(cm3$overall[1]))
Accuracy <- c(0.5442,0.9944,0.7822)
Model_Summary <- data.frame(ModelName,Accuracy,OOSE)
Model_Summary
```

```
##           ModelName Accuracy      OOSE
## 1 Decision Tree    0.5442 0.4695386
## 2 Random Forest    0.9944 0.0054805
## 3              SVM    0.7822 0.2106806
```

It can be observed that Random Forest model performs the best with the highest accuracy of 99.44%. Therefore, it is chosen as the best model for this dataset and the prediction on testing dataset will be performed using the Random Forest Model

```
# Prediction on Tests set  
pred <- predict(mod2, testClean)  
print(pred)
```

```
## [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
```