Machine learning algorithm to predict quality of physical excercise

Introduction:

The goal of this project is to design an algorithm to accurately predict, using data collected from accelerometers fitted to various devices, the quality of the weight lifting exercise performed by an user wearing the device.[1]

The data from the accelerometers consists of observations recording acceleration and other attributes relating to the movement of dumbbells, belts and other items worn by the user on which the accelerometers were embedded.

Data Collection: The data was downloaded on 15 February 2015 from the following URL: https://class.coursera.org/predmachlearn-011/human_grading/view/courses/973546/assessments/4/submissions The downloaded data has been cached, so that updates do not affect our analysis.

Predictive analysis: Predictive analysis was used to:

- Construct, based on a training set, a model that helps predict the quality of the exercise
- Apply the model on a different data set (the test set) to test the accuracy of the model The R statistical package was used to perform our analysis. The caret and randomForest libraries for R were used.

Modeling: A Random Forest model trained on the training set was used to predict the quality of the exercise.[2]

```
x = read.csv("pml-training.csv", na.strings = c("", "NA"))
# changes all 'NA' and blanks in the dataset to NA

str(x)
```

```
## 'data.frame':    19622 obs. of  160 variables:
##  $ X                       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ user_name               : Factor w/ 6 levels "adelmo","carlitos",..: 2 2 2 2 2 2 2 2
##  $ raw_timestamp_part_1    : int  1323084231 1323084231 1323084231 1323084232 132308423
##  $ raw_timestamp_part_2    : int  788290 808298 820366 120339 196328 304277 368296 44039
##  $ cvtd_timestamp          : Factor w/ 20 levels "02/12/2011 13:32",..: 9 9 9 9 9 9 9 9
##  $ new_window              : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ num_window              : int  11 11 11 12 12 12 12 12 12 12 ...
##  $ roll_belt               : num  1.41 1.41 1.42 1.48 1.48 1.45 1.42 1.42 1.43 1.45 ...
##  $ pitch_belt              : num  8.07 8.07 8.07 8.05 8.07 8.06 8.09 8.13 8.16 8.17 ...
##  $ yaw_belt                : num  -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4
##  $ total_accel_belt        : int  3 3 3 3 3 3 3 3 3 3 ...
##  $ kurtosis_roll_belt      : Factor w/ 396 levels "-0.016850","-0.021024",..: NA NA NA N
##  $ kurtosis_picth_belt     : Factor w/ 316 levels "-0.021887","-0.060755",..: NA NA NA N
##  $ kurtosis_yaw_belt       : Factor w/ 1 level "#DIV/0!": NA NA NA NA NA NA NA NA NA NA
##  $ skewness_roll_belt      : Factor w/ 394 levels "-0.003095","-0.010002",..: NA NA NA N
##  $ skewness_roll_belt.1    : Factor w/ 337 levels "-0.005928","-0.005960",..: NA NA NA N
##  $ skewness_yaw_belt       : Factor w/ 1 level "#DIV/0!": NA NA NA NA NA NA NA NA NA NA
##  $ max_roll_belt           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_picth_belt          : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_yaw_belt            : Factor w/ 67 levels "-0.1","-0.2",..: NA NA NA NA NA NA NA
##  $ min_roll_belt           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_pitch_belt          : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_yaw_belt            : Factor w/ 67 levels "-0.1","-0.2",..: NA NA NA NA NA NA NA
##  $ amplitude_roll_belt     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_pitch_belt    : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_yaw_belt      : Factor w/ 3 levels "#DIV/0!","0.00",..: NA NA NA NA NA NA N
##  $ var_total_accel_belt    : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_roll_belt           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ stddev_roll_belt        : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_roll_belt           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_pitch_belt          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ stddev_pitch_belt       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_pitch_belt          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_yaw_belt            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ stddev_yaw_belt         : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_yaw_belt            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ gyros_belt_x            : num  0 0.02 0 0.02 0.02 0.02 0.02 0.02 0.02 0.03 ...
##  $ gyros_belt_y            : num  0 0 0 0 0.02 0 0 0 0 0 ...
##  $ gyros_belt_z            : num  -0.02 -0.02 -0.02 -0.03 -0.02 -0.02 -0.02 -0.02 -0.02
##  $ accel_belt_x            : int  -21 -22 -20 -22 -21 -21 -22 -22 -20 -21 ...
##  $ accel_belt_y            : int  4 4 5 3 2 4 3 4 2 4 ...
##  $ accel_belt_z            : int  22 22 23 21 24 21 21 21 24 22 ...
##  $ magnet_belt_x           : int  -3 -7 -2 -6 -6 0 -4 -2 1 -3 ...
##  $ magnet_belt_y           : int  599 608 600 604 600 603 599 603 602 609 ...
##  $ magnet_belt_z           : int  -313 -311 -305 -310 -302 -312 -311 -313 -312 -308 ...
##  $ roll_arm                : num  -128 -128 -128 -128 -128 -128 -128 -128 -128 -128 ...
##  $ pitch_arm               : num  22.5 22.5 22.5 22.1 22.1 22 21.9 21.8 21.7 21.6 ...
##  $ yaw_arm                 : num  -161 -161 -161 -161 -161 -161 -161 -161 -161 -161 ...
##  $ total_accel_arm         : int  34 34 34 34 34 34 34 34 34 34 ...
##  $ var_accel_arm           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_roll_arm            : num  NA NA NA NA NA NA NA NA NA NA ...
```

```
##  $ stddev_roll_arm         : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_roll_arm            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_pitch_arm           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ stddev_pitch_arm        : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_pitch_arm           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ avg_yaw_arm             : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ stddev_yaw_arm          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_yaw_arm             : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ gyros_arm_x             : num  0 0.02 0.02 0.02 0 0.02 0 0.02 0.02 0.02 ...
##  $ gyros_arm_y             : num  0 -0.02 -0.02 -0.03 -0.03 -0.03 -0.03 -0.02 -0.03 -0.0
##  $ gyros_arm_z             : num  -0.02 -0.02 -0.02 0.02 0 0 0 0 -0.02 -0.02 ...
##  $ accel_arm_x             : int  -288 -290 -289 -289 -289 -289 -289 -289 -288 -288 ...
##  $ accel_arm_y             : int  109 110 110 111 111 111 111 111 109 110 ...
##  $ accel_arm_z             : int  -123 -125 -126 -123 -123 -122 -125 -124 -122 -124 ...
##  $ magnet_arm_x            : int  -368 -369 -368 -372 -374 -369 -373 -372 -369 -376 ...
##  $ magnet_arm_y            : int  337 337 344 344 337 342 336 338 341 334 ...
##  $ magnet_arm_z            : int  516 513 513 512 506 513 509 510 518 516 ...
##  $ kurtosis_roll_arm       : Factor w/ 329 levels "-0.02438","-0.04190",..: NA NA NA NA
##  $ kurtosis_picth_arm      : Factor w/ 327 levels "-0.00484","-0.01311",..: NA NA NA NA
##  $ kurtosis_yaw_arm        : Factor w/ 394 levels "-0.01548","-0.01749",..: NA NA NA NA
##  $ skewness_roll_arm       : Factor w/ 330 levels "-0.00051","-0.00696",..: NA NA NA NA
##  $ skewness_pitch_arm      : Factor w/ 327 levels "-0.00184","-0.01185",..: NA NA NA NA
##  $ skewness_yaw_arm        : Factor w/ 394 levels "-0.00311","-0.00562",..: NA NA NA NA
##  $ max_roll_arm            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_picth_arm           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_yaw_arm             : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_roll_arm            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_pitch_arm           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_yaw_arm             : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_roll_arm      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_pitch_arm     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_yaw_arm       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ roll_dumbbell           : num  13.1 13.1 12.9 13.4 13.4 ...
##  $ pitch_dumbbell          : num  -70.5 -70.6 -70.3 -70.4 -70.4 ...
##  $ yaw_dumbbell            : num  -84.9 -84.7 -85.1 -84.9 -84.9 ...
##  $ kurtosis_roll_dumbbell  : Factor w/ 397 levels "-0.0035","-0.0073",..: NA NA NA NA NA
##  $ kurtosis_picth_dumbbell : Factor w/ 400 levels "-0.0163","-0.0233",..: NA NA NA NA NA
##  $ kurtosis_yaw_dumbbell   : Factor w/ 1 level "#DIV/0!": NA NA NA NA NA NA NA NA NA NA
##  $ skewness_roll_dumbbell  : Factor w/ 400 levels "-0.0082","-0.0096",..: NA NA NA NA NA
##  $ skewness_pitch_dumbbell : Factor w/ 401 levels "-0.0053","-0.0084",..: NA NA NA NA NA
##  $ skewness_yaw_dumbbell   : Factor w/ 1 level "#DIV/0!": NA NA NA NA NA NA NA NA NA NA
##  $ max_roll_dumbbell       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_picth_dumbbell      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_yaw_dumbbell        : Factor w/ 72 levels "-0.1","-0.2",..: NA NA NA NA NA NA NA
##  $ min_roll_dumbbell       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_pitch_dumbbell      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_yaw_dumbbell        : Factor w/ 72 levels "-0.1","-0.2",..: NA NA NA NA NA NA NA
##  $ amplitude_roll_dumbbell : num  NA NA NA NA NA NA NA NA NA NA ...
##   [list output truncated]
```

```
# we can see that columns are mostly numeric, some integers and some are
# factors, including the all-important 'classe' column

sum(complete.cases(x))
```

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

```
# we can see that very very few rows are complete..most have many 'NA's-
# hence, the data needs to be cleaned of the missing observations

head(x)
```

```
##   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
## 1               <NA>                <NA>              <NA>
## 2               <NA>                <NA>              <NA>
## 3               <NA>                <NA>              <NA>
## 4               <NA>                <NA>              <NA>
## 5               <NA>                <NA>              <NA>
## 6               <NA>                <NA>              <NA>
##   skewness_roll_belt skewness_roll_belt.1 skewness_yaw_belt max_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
## 6               <NA>                <NA>              <NA>            NA
##   max_picth_belt max_yaw_belt min_roll_belt min_pitch_belt min_yaw_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>
## 6             NA         <NA>            NA             NA         <NA>
##   amplitude_roll_belt amplitude_pitch_belt amplitude_yaw_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>
##   var_total_accel_belt avg_roll_belt stddev_roll_belt var_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
## 6                   NA            NA                NA            NA
##   avg_pitch_belt stddev_pitch_belt var_pitch_belt avg_yaw_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
##   stddev_yaw_belt var_yaw_belt gyros_belt_x gyros_belt_y gyros_belt_z
## 1              NA           NA         0.00         0.00        -0.02
## 2              NA           NA         0.02         0.00        -0.02
## 3              NA           NA         0.00         0.00        -0.02
## 4              NA           NA         0.02         0.00        -0.03
## 5              NA           NA         0.02         0.02        -0.02
## 6              NA           NA         0.02         0.00        -0.02
##   accel_belt_x accel_belt_y accel_belt_z magnet_belt_x magnet_belt_y
## 1          -21            4           22             -3           599
## 2          -22            4           22             -7           608
## 3          -20            5           23             -2           600
## 4          -22            3           21             -6           604
## 5          -21            2           24             -6           600
## 6          -21            4           21              0           603
##   magnet_belt_z roll_arm pitch_arm yaw_arm total_accel_arm var_accel_arm
## 1          -313     -128      22.5    -161              34            NA
## 2          -311     -128      22.5    -161              34            NA
## 3          -305     -128      22.5    -161              34            NA
## 4          -310     -128      22.1    -161              34            NA
## 5          -302     -128      22.1    -161              34            NA
## 6          -312     -128      22.0    -161              34            NA
##   avg_roll_arm stddev_roll_arm var_roll_arm avg_pitch_arm stddev_pitch_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_pitch_arm avg_yaw_arm stddev_yaw_arm var_yaw_arm gyros_arm_x
## 1            NA          NA             NA          NA        0.00
## 2            NA          NA             NA          NA        0.02
## 3            NA          NA             NA          NA        0.02
## 4            NA          NA             NA          NA        0.02
## 5            NA          NA             NA          NA        0.00
## 6            NA          NA             NA          NA        0.02
##   gyros_arm_y gyros_arm_z accel_arm_x accel_arm_y accel_arm_z magnet_arm_x
## 1        0.00       -0.02        -288         109        -123         -368
## 2       -0.02       -0.02        -290         110        -125         -369
## 3       -0.02       -0.02        -289         110        -126         -368
## 4       -0.03        0.02        -289         111        -123         -372
## 5       -0.03        0.00        -289         111        -123         -374
## 6       -0.03        0.00        -289         111        -122         -369
##   magnet_arm_y magnet_arm_z kurtosis_roll_arm kurtosis_picth_arm
## 1          337          516              <NA>               <NA>
## 2          337          513              <NA>               <NA>
## 3          344          513              <NA>               <NA>
## 4          344          512              <NA>               <NA>
## 5          337          506              <NA>               <NA>
```

```
## 6          342          513              <NA>               <NA>
##   kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_yaw_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>
##   max_roll_arm max_picth_arm max_yaw_arm min_roll_arm min_pitch_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
##   min_yaw_arm amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_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
##   roll_dumbbell pitch_dumbbell yaw_dumbbell kurtosis_roll_dumbbell
## 1         13.05         -70.49       -84.87                  <NA>
## 2         13.13         -70.64       -84.71                  <NA>
## 3         12.85         -70.28       -85.14                  <NA>
## 4         13.43         -70.39       -84.87                  <NA>
## 5         13.38         -70.43       -84.85                  <NA>
## 6         13.38         -70.82       -84.47                  <NA>
##   kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_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>
##   skewness_pitch_dumbbell skewness_yaw_dumbbell max_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
##   max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell min_pitch_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
##   min_yaw_dumbbell amplitude_roll_dumbbell amplitude_pitch_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
##   amplitude_yaw_dumbbell total_accel_dumbbell var_accel_dumbbell
## 1                   <NA>                    37                 NA
## 2                   <NA>                    37                 NA
## 3                   <NA>                    37                 NA
## 4                   <NA>                    37                 NA
## 5                   <NA>                    37                 NA
## 6                   <NA>                    37                 NA
##   avg_roll_dumbbell stddev_roll_dumbbell var_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
##   avg_pitch_dumbbell stddev_pitch_dumbbell var_pitch_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
##   avg_yaw_dumbbell stddev_yaw_dumbbell var_yaw_dumbbell gyros_dumbbell_x
## 1               NA                  NA               NA                0
## 2               NA                  NA               NA                0
## 3               NA                  NA               NA                0
## 4               NA                  NA               NA                0
## 5               NA                  NA               NA                0
## 6               NA                  NA               NA                0
##   gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y
## 1            -0.02             0.00             -234               47
## 2            -0.02             0.00             -233               47
## 3            -0.02             0.00             -232               46
## 4            -0.02            -0.02             -232               48
## 5            -0.02             0.00             -233               48
## 6            -0.02             0.00             -234               48
##   accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z
## 1             -271              -559               293               -65
## 2             -269              -555               296               -64
## 3             -270              -561               298               -63
## 4             -269              -552               303               -60
## 5             -270              -554               292               -68
## 6             -269              -558               294               -66
##   roll_forearm pitch_forearm yaw_forearm kurtosis_roll_forearm
## 1         28.4         -63.9        -153                  <NA>
## 2         28.3         -63.9        -153                  <NA>
## 3         28.3         -63.9        -152                  <NA>
## 4         28.1         -63.9        -152                  <NA>
```

```
## 5         28.0         -63.9        -152                <NA>
## 6         27.9         -63.9        -152                <NA>
##   kurtosis_picth_forearm kurtosis_yaw_forearm skewness_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>
##   skewness_pitch_forearm skewness_yaw_forearm max_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
##   max_picth_forearm max_yaw_forearm min_roll_forearm min_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
## 6                NA            <NA>               NA                NA
##   min_yaw_forearm amplitude_roll_forearm amplitude_pitch_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_yaw_forearm total_accel_forearm var_accel_forearm
## 1                  <NA>                  36                NA
## 2                  <NA>                  36                NA
## 3                  <NA>                  36                NA
## 4                  <NA>                  36                NA
## 5                  <NA>                  36                NA
## 6                  <NA>                  36                NA
##   avg_roll_forearm stddev_roll_forearm var_roll_forearm avg_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
## 6               NA                  NA                NA                NA
##   stddev_pitch_forearm var_pitch_forearm avg_yaw_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
##   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
```

```
tail(x)
```

```
##           X user_name raw_timestamp_part_1 raw_timestamp_part_2
## 19617 19617    adelmo           1322832937               588324
## 19618 19618    adelmo           1322832937               588376
## 19619 19619    adelmo           1322832937               596287
## 19620 19620    adelmo           1322832937               636283
## 19621 19621    adelmo           1322832937               964299
## 19622 19622    adelmo           1322832937               972293
##         cvtd_timestamp new_window num_window roll_belt pitch_belt yaw_belt
## 19617 02/12/2011 13:35         no        864       148      -34.7      129
## 19618 02/12/2011 13:35         no        864       147      -34.8      129
## 19619 02/12/2011 13:35         no        864       145      -35.3      130
## 19620 02/12/2011 13:35         no        864       145      -35.5      130
## 19621 02/12/2011 13:35         no        864       143      -35.9      131
## 19622 02/12/2011 13:35        yes        864       143      -36.0      132
##       total_accel_belt kurtosis_roll_belt kurtosis_picth_belt
## 19617               21               <NA>                <NA>
## 19618               21               <NA>                <NA>
## 19619               19               <NA>                <NA>
## 19620               19               <NA>                <NA>
## 19621               18               <NA>                <NA>
## 19622               18          -1.175902           -1.063259
##       kurtosis_yaw_belt skewness_roll_belt skewness_roll_belt.1
## 19617              <NA>               <NA>                <NA>
## 19618              <NA>               <NA>                <NA>
## 19619              <NA>               <NA>                <NA>
## 19620              <NA>               <NA>                <NA>
## 19621              <NA>               <NA>                <NA>
## 19622           #DIV/0!           0.196860           -0.572396
##       skewness_yaw_belt max_roll_belt max_picth_belt max_yaw_belt
## 19617              <NA>            NA             NA         <NA>
## 19618              <NA>            NA             NA         <NA>
## 19619              <NA>            NA             NA         <NA>
## 19620              <NA>            NA             NA         <NA>
## 19621              <NA>            NA             NA         <NA>
## 19622           #DIV/0!           132             25         -1.2
##       min_roll_belt min_pitch_belt min_yaw_belt amplitude_roll_belt
## 19617            NA             NA         <NA>                  NA
## 19618            NA             NA         <NA>                  NA
## 19619            NA             NA         <NA>                  NA
## 19620            NA             NA         <NA>                  NA
## 19621            NA             NA         <NA>                  NA
## 19622           123             18         -1.2                   9
##       amplitude_pitch_belt amplitude_yaw_belt var_total_accel_belt
## 19617                   NA               <NA>                   NA
## 19618                   NA               <NA>                   NA
## 19619                   NA               <NA>                   NA
## 19620                   NA               <NA>                   NA
## 19621                   NA               <NA>                   NA
## 19622                    7               0.00                5.627
##       avg_roll_belt stddev_roll_belt var_roll_belt avg_pitch_belt
## 19617            NA               NA            NA             NA
## 19618            NA               NA            NA             NA
```

```
## 19619            NA               NA            NA             NA
## 19620            NA               NA            NA             NA
## 19621            NA               NA            NA             NA
## 19622         151.1            4.753         22.59          -33.63
##       stddev_pitch_belt var_pitch_belt avg_yaw_belt stddev_yaw_belt
## 19617                NA             NA           NA              NA
## 19618                NA             NA           NA              NA
## 19619                NA             NA           NA              NA
## 19620                NA             NA           NA              NA
## 19621                NA             NA           NA              NA
## 19622             1.395          1.947        126.9            2.75
##       var_yaw_belt gyros_belt_x gyros_belt_y gyros_belt_z accel_belt_x
## 19617           NA         0.37         0.00        -0.62           49
## 19618           NA         0.37        -0.02        -0.67           50
## 19619           NA         0.39        -0.02        -0.67           47
## 19620           NA         0.37         0.00        -0.64           47
## 19621           NA         0.37        -0.02        -0.59           46
## 19622        7.564         0.35        -0.02        -0.57           42
##       accel_belt_y accel_belt_z magnet_belt_x magnet_belt_y magnet_belt_z
## 19617           25         -195           191           540          -415
## 19618           26         -193           190           552          -412
## 19619           15         -179           192           558          -389
## 19620           13         -177           191           560          -386
## 19621           18         -172           190           565          -370
## 19622           25         -171           194           566          -349
##       roll_arm pitch_arm yaw_arm total_accel_arm var_accel_arm
## 19617    -99.1     -33.7    79.4              48            NA
## 19618    -99.4     -33.8    79.0              47            NA
## 19619    -99.6     -34.5    77.3              45            NA
## 19620    -99.6     -35.1    76.3              44            NA
## 19621    -98.6     -36.7    73.5              41            NA
## 19622    -97.6     -37.7    71.5              41         54.26
##       avg_roll_arm stddev_roll_arm var_roll_arm avg_pitch_arm
## 19617           NA              NA           NA            NA
## 19618           NA              NA           NA            NA
## 19619           NA              NA           NA            NA
## 19620           NA              NA           NA            NA
## 19621           NA              NA           NA            NA
## 19622       -91.65           9.169        84.06        -37.65
##       stddev_pitch_arm var_pitch_arm avg_yaw_arm stddev_yaw_arm
## 19617               NA            NA          NA             NA
## 19618               NA            NA          NA             NA
## 19619               NA            NA          NA             NA
## 19620               NA            NA          NA             NA
## 19621               NA            NA          NA             NA
## 19622            3.616         13.08       66.31          15.48
##       var_yaw_arm gyros_arm_x gyros_arm_y gyros_arm_z accel_arm_x
## 19617          NA        0.31       -0.45        0.28          67
## 19618          NA        0.55       -0.51        0.25          75
## 19619          NA        0.88       -0.71        0.21          52
## 19620          NA        0.98       -0.82        0.23          62
## 19621          NA        1.35       -1.00        0.49          70
```

```
## 19622       239.6        1.51       -1.06        0.59          58
##       accel_arm_y accel_arm_z magnet_arm_x magnet_arm_y magnet_arm_z
## 19617        -181        -432          268        -138        -566
## 19618        -184        -415          272        -134        -562
## 19619        -163        -406          288        -112        -559
## 19620        -167        -391          309        -103        -541
## 19621        -164        -359          339         -91        -543
## 19622        -152        -365          362         -84        -539
##       kurtosis_roll_arm kurtosis_picth_arm kurtosis_yaw_arm
## 19617              <NA>               <NA>             <NA>
## 19618              <NA>               <NA>             <NA>
## 19619              <NA>               <NA>             <NA>
## 19620              <NA>               <NA>             <NA>
## 19621              <NA>               <NA>             <NA>
## 19622          -1.32631            0.50959         -0.62736
##       skewness_roll_arm skewness_pitch_arm skewness_yaw_arm max_roll_arm
## 19617              <NA>               <NA>             <NA>           NA
## 19618              <NA>               <NA>             <NA>           NA
## 19619              <NA>               <NA>             <NA>           NA
## 19620              <NA>               <NA>             <NA>           NA
## 19621              <NA>               <NA>             <NA>           NA
## 19622          -0.51721           -1.26872         -0.77150        -33.7
##       max_picth_arm max_yaw_arm min_roll_arm min_pitch_arm min_yaw_arm
## 19617            NA          NA           NA            NA          NA
## 19618            NA          NA           NA            NA          NA
## 19619            NA          NA           NA            NA          NA
## 19620            NA          NA           NA            NA          NA
## 19621            NA          NA           NA            NA          NA
## 19622          79.5          49        -43.5          27.5          25
##       amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_arm
## 19617                 NA                  NA                NA
## 19618                 NA                  NA                NA
## 19619                 NA                  NA                NA
## 19620                 NA                  NA                NA
## 19621                 NA                  NA                NA
## 19622                9.8                  52                24
##       roll_dumbbell pitch_dumbbell yaw_dumbbell kurtosis_roll_dumbbell
## 19617         38.61         -22.79       -111.6                  <NA>
## 19618         36.41         -22.86       -113.5                  <NA>
## 19619         35.15         -22.97       -114.5                  <NA>
## 19620         30.06         -20.99       -120.0                  <NA>
## 19621         22.86         -21.76       -125.2                  <NA>
## 19622         20.80         -19.70       -128.2               -1.1322
##       kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell
## 19617                    <NA>                  <NA>                   <NA>
## 19618                    <NA>                  <NA>                   <NA>
## 19619                    <NA>                  <NA>                   <NA>
## 19620                    <NA>                  <NA>                   <NA>
## 19621                    <NA>                  <NA>                   <NA>
## 19622                 -0.7225               #DIV/0!                 0.0955
##       skewness_pitch_dumbbell skewness_yaw_dumbbell max_roll_dumbbell
## 19617                    <NA>                  <NA>                NA
```

```
## 19618                    <NA>                  <NA>                  NA
## 19619                    <NA>                  <NA>                  NA
## 19620                    <NA>                  <NA>                  NA
## 19621                    <NA>                  <NA>                  NA
## 19622                  0.1057                #DIV/0!                -19.7
##       max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell
## 19617                 NA             <NA>                 NA
## 19618                 NA             <NA>                 NA
## 19619                 NA             <NA>                 NA
## 19620                 NA             <NA>                 NA
## 19621                 NA             <NA>                 NA
## 19622                -92             -1.1              -33.1
##       min_pitch_dumbbell min_yaw_dumbbell amplitude_roll_dumbbell
## 19617                 NA             <NA>                      NA
## 19618                 NA             <NA>                      NA
## 19619                 NA             <NA>                      NA
## 19620                 NA             <NA>                      NA
## 19621                 NA             <NA>                      NA
## 19622              -128.2             -1.1                   13.41
##       amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
## 19617                       NA                  <NA>                   19
## 19618                       NA                  <NA>                   19
## 19619                       NA                  <NA>                   18
## 19620                       NA                  <NA>                   19
## 19621                       NA                  <NA>                   19
## 19622                     36.2                  0.00                   19
##       var_accel_dumbbell avg_roll_dumbbell stddev_roll_dumbbell
## 19617                 NA                NA                   NA
## 19618                 NA                NA                   NA
## 19619                 NA                NA                   NA
## 19620                 NA                NA                   NA
## 19621                 NA                NA                   NA
## 19622             0.4217             37.34                9.783
##       var_roll_dumbbell avg_pitch_dumbbell stddev_pitch_dumbbell
## 19617                NA                NA                   NA
## 19618                NA                NA                   NA
## 19619                NA                NA                   NA
## 19620                NA                NA                   NA
## 19621                NA                NA                   NA
## 19622              95.7             -26.82                 4.01
##       var_pitch_dumbbell avg_yaw_dumbbell stddev_yaw_dumbbell
## 19617                 NA                NA                  NA
## 19618                 NA                NA                  NA
## 19619                 NA                NA                  NA
## 19620                 NA                NA                  NA
## 19621                 NA                NA                  NA
## 19622              16.08              -110               9.748
##       var_yaw_dumbbell gyros_dumbbell_x gyros_dumbbell_y gyros_dumbbell_z
## 19617               NA             0.34            -0.31            -0.51
## 19618               NA             0.32            -0.26            -0.36
## 19619               NA             0.24            -0.24             0.05
## 19620               NA             0.22            -0.27             0.21
```

```
## 19621              NA             0.13            -0.14             0.34
## 19622           95.01             0.02             0.02             0.36
##       accel_dumbbell_x accel_dumbbell_y accel_dumbbell_z magnet_dumbbell_x
## 19617              -42               70             -167              -624
## 19618              -42               66             -168              -618
## 19619              -41               62             -164              -618
## 19620              -38               54             -170              -621
## 19621              -40               42             -176              -628
## 19622              -36               38             -176              -627
##       magnet_dumbbell_y magnet_dumbbell_z roll_forearm pitch_forearm
## 19617               127                 8            0             0
## 19618               134                 0            0             0
## 19619               116                 7            0             0
## 19620               113                -9            0             0
## 19621               116                 0            0             0
## 19622               119                 2            0             0
##       yaw_forearm kurtosis_roll_forearm kurtosis_picth_forearm
## 19617           0                  <NA>                  <NA>
## 19618           0                  <NA>                  <NA>
## 19619           0                  <NA>                  <NA>
## 19620           0                  <NA>                  <NA>
## 19621           0                  <NA>                  <NA>
## 19622           0               #DIV/0!               #DIV/0!
##       kurtosis_yaw_forearm skewness_roll_forearm skewness_pitch_forearm
## 19617                 <NA>                  <NA>                  <NA>
## 19618                 <NA>                  <NA>                  <NA>
## 19619                 <NA>                  <NA>                  <NA>
## 19620                 <NA>                  <NA>                  <NA>
## 19621                 <NA>                  <NA>                  <NA>
## 19622              #DIV/0!               #DIV/0!               #DIV/0!
##       skewness_yaw_forearm max_roll_forearm max_picth_forearm
## 19617                 <NA>               NA                NA
## 19618                 <NA>               NA                NA
## 19619                 <NA>               NA                NA
## 19620                 <NA>               NA                NA
## 19621                 <NA>               NA                NA
## 19622              #DIV/0!                0                 0
##       max_yaw_forearm min_roll_forearm min_pitch_forearm min_yaw_forearm
## 19617            <NA>               NA                NA            <NA>
## 19618            <NA>               NA                NA            <NA>
## 19619            <NA>               NA                NA            <NA>
## 19620            <NA>               NA                NA            <NA>
## 19621            <NA>               NA                NA            <NA>
## 19622         #DIV/0!                0                 0         #DIV/0!
##       amplitude_roll_forearm amplitude_pitch_forearm amplitude_yaw_forearm
## 19617                     NA                      NA                  <NA>
## 19618                     NA                      NA                  <NA>
## 19619                     NA                      NA                  <NA>
## 19620                     NA                      NA                  <NA>
## 19621                     NA                      NA                  <NA>
## 19622                      0                       0               #DIV/0!
##       total_accel_forearm var_accel_forearm avg_roll_forearm
```

```
## 19617                   27                 NA                 NA
## 19618                   29                 NA                 NA
## 19619                   29                 NA                 NA
## 19620                   29                 NA                 NA
## 19621                   32                 NA                 NA
## 19622                   33              30.11                  0
##       stddev_roll_forearm var_roll_forearm avg_pitch_forearm
## 19617                  NA                NA                NA
## 19618                  NA                NA                NA
## 19619                  NA                NA                NA
## 19620                  NA                NA                NA
## 19621                  NA                NA                NA
## 19622                   0                 0                 0
##       stddev_pitch_forearm var_pitch_forearm avg_yaw_forearm
## 19617                   NA                NA               NA
## 19618                   NA                NA               NA
## 19619                   NA                NA               NA
## 19620                   NA                NA               NA
## 19621                   NA                NA               NA
## 19622                    0                 0                0
##       stddev_yaw_forearm var_yaw_forearm gyros_forearm_x gyros_forearm_y
## 19617                 NA               NA            1.75           -1.91
## 19618                 NA               NA            1.73           -1.75
## 19619                 NA               NA            1.59           -1.36
## 19620                 NA               NA            1.54           -1.20
## 19621                 NA               NA            1.48           -0.90
## 19622                  0                0            1.38           -0.64
##       gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z
## 19617           -0.38            -255             -50             -30
## 19618           -0.25            -271             -68             -37
## 19619            0.00            -271             -91             -43
## 19620            0.05            -263             -99             -45
## 19621            0.05            -270            -141             -51
## 19622            0.08            -278            -159             -52
##       magnet_forearm_x magnet_forearm_y magnet_forearm_z classe
## 19617             -226             -570               27      E
## 19618             -205             -587                6      E
## 19619             -151             -635              -36      E
## 19620             -116             -654              -70      E
## 19621              -68             -678              -98      E
## 19622              -60             -686             -110      E
```

```
# we can observe that most of the missing data is in particular columns
# and that there are quite a few columns which do not seem to have any
# missing data

y = t(x)
sum(complete.cases(y))
```

```
## [1] 60
```

```
# we see that as maany as around 60 columns of the original input data
# have complete data. That's enough predictors to build a decent model!

# Now, we remove those columns of the input data that do not have a lot of
# observations
cc2 = complete.cases(y)
y = x[, cc2]
```

```
# Further, from reading the head and tail, we could identify some columns
# as being metadata (such as timestamps). Given the nature of data
# (weightlifting excercises, we rule out the possibility of the data being
# a time series.

# There remained a question on whether the 'user_name' column could have
# any value as a predcitor. After all,the quality of excercise could very
# well depend on the person doing it. The following table was run to do a
# bit of fishing..
round(prop.table(table(y$user_name, y$classe), 1), 2)
```

```
##
##               A    B    C    D    E
##   adelmo   0.30 0.20 0.19 0.13 0.18
##   carlitos 0.27 0.22 0.16 0.16 0.20
##   charles  0.25 0.21 0.15 0.18 0.20
##   eurico   0.28 0.19 0.16 0.19 0.18
##   jeremy   0.35 0.14 0.19 0.15 0.17
##   pedro    0.25 0.19 0.19 0.18 0.19
```

```
# It was decided to keep the username column, as we are using a tree-based
# model anyway and could rely on the model to not consider a predictor
# with not much predictive value
```

```
y = y[, -c(1, 3:7)]
# the metadata such as timestamps mentioned above arer removed.
# 'User_name' is kept for now. The column numbers were easy enough to spot
# from runnig head(x) and tail(x).
dim(y)
```

```
## [1] 19622    54
```

```
# At this point, it may be useful to save the processed data set (this is
# just a record keeping step!)
write.csv(y, file = "y.csv", quote = FALSE)

# store the number of predictors in a separate variable for ease of use
# (as the last column is being predicted and the rest used to predict it,
# we know that the number of predictors is no of cols -1)
z = dim(y)[2] - 1

# Forcing all the integer fields to numeric. Leaving factors as factors
# (the first & the last columns).
for (i in 2:z) {
    y[, i] = as.numeric(y[, i])
}
# we could also use apply functions here instead of a for-loop..

y[, z + 1] = as.factor(y[, z + 1])
# using coercion to make sure that the 'classe' is a factor. This is not
# really a required step.
```

Cross validation:

The model was cross-validated on a test set spliced off from the training set provided. 95% of the training set was used to train the model while 5% was used as the test set.

Further, the model built was also cross validated by testing it on randomly generated chunks of 20 and 100 observations from the training set.

```
# Now that the procesing and exploration is over, we go to the model
# building part..

# Importing libraries
library(caret)
library(randomForest)

# Slice the training set data into two so that we can cross-validate..
set.seed(9994)
trainIndex = createDataPartition(y$classe, p = 0.95, list = FALSE)  #create a 95-5 split..
training = y[trainIndex, ]
testing = y[-trainIndex, ]
dim(training)
```

```
## [1] 18643    54
```

```
dim(testing)
```

```
## [1] 979  54
```

Choice of model: Since the data being predicted is not continuous, a tree-based classification model, rather

than a regression based model was considered more appropriate.

The RandomForest method was chosen as the method generates many trees and averages its predictions across the trees, helping smooth out the randomness that a single iteration of the tree model can throw up. The RandomForest method is known to provide good accuracy in classfication problems.

```
set.seed(9)
model <- randomForest(classe ~ ., data = training, mtry = 3)
# Build a predictor model using randomForest.Having a greater 'mtry'' puts
# stress on the memory needed to run R. See ?randomForest for more
# details.
```

Out of sample error: Since the value being predicted is a set of 5 classes (i.e. not continuous), Accuracy was considered the appopriate measure of the effectiveness of the model.

The model built was tested on the test set and the results were satisfactory (Accuracy > 0.99). Based on the testing on the spliced-off test set, we expect an out of sample error rate of < 0.01 for predictions generated by the model. The in-sample accuracy of the model (i.e. the accuracy on the data on which it was trained) was 1.0. However, the low out of sample error rate allays fears of overfitting.

```
pred = predict(model, testing)
confusionMatrix(testing$classe, pred)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A   B   C   D   E
##          A 279   0   0   0   0
##          B   2 187   0   0   0
##          C   0   0 171   0   0
##          D   0   0   2 158   0
##          E   0   0   0   0 180
##
## Overall Statistics
##
##                Accuracy : 0.996
##                  95% CI : (0.99, 0.999)
##     No Information Rate : 0.287
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.995
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity             0.993    1.000    0.988    1.000    1.000
## Specificity             1.000    0.997    1.000    0.998    1.000
## Pos Pred Value          1.000    0.989    1.000    0.988    1.000
## Neg Pred Value          0.997    1.000    0.998    1.000    1.000
## Prevalence              0.287    0.191    0.177    0.161    0.184
## Detection Rate          0.285    0.191    0.175    0.161    0.184
## Detection Prevalence    0.285    0.193    0.175    0.163    0.184
## Balanced Accuracy       0.996    0.999    0.994    0.999    1.000
```

Further testing of the model: The model was also tested on randomly pulled smaller sets of observations from the input data. The results were satisfactory (Accuracy being 1.0 on most occasions, and around 0.99 on a few).

```
# Pick a sample of 20 from the input data and predict..20 was chosen as
# the ultimate test set that had to be predicted was of size 20!
M = y[sample(dim(y)[1], 20, replace = FALSE), ]
predM = predict(model, M)
confusionMatrix(M$classe, predM)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction A B C D E
##          A 6 0 0 0 0
##          B 0 4 0 0 0
##          C 0 0 5 0 0
##          D 0 0 0 3 0
##          E 0 0 0 0 2
##
## Overall Statistics
##
##                Accuracy : 1
##                  95% CI : (0.832, 1)
##     No Information Rate : 0.3
##     P-Value [Acc > NIR] : 3.49e-11
##
##                   Kappa : 1
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity               1.0      1.0     1.00     1.00      1.0
## Specificity               1.0      1.0     1.00     1.00      1.0
## Pos Pred Value            1.0      1.0     1.00     1.00      1.0
## Neg Pred Value            1.0      1.0     1.00     1.00      1.0
## Prevalence                0.3      0.2     0.25     0.15      0.1
## Detection Rate            0.3      0.2     0.25     0.15      0.1
## Detection Prevalence      0.3      0.2     0.25     0.15      0.1
## Balanced Accuracy         1.0      1.0     1.00     1.00      1.0
```

```
# Do it a few more times..
M = y[sample(dim(y)[1], 20, replace = FALSE), ]
predM = predict(model, M)
confusionMatrix(M$classe, predM)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction A B C D E
##          A 5 0 0 0 0
##          B 0 4 0 0 0
##          C 0 0 4 0 0
##          D 0 0 0 5 0
##          E 0 0 0 0 2
##
## Overall Statistics
##
##                Accuracy : 1
##                  95% CI : (0.832, 1)
##     No Information Rate : 0.25
##     P-Value [Acc > NIR] : 9.09e-13
##
##                   Kappa : 1
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity              1.00      1.0      1.0     1.00      1.0
## Specificity              1.00      1.0      1.0     1.00      1.0
## Pos Pred Value           1.00      1.0      1.0     1.00      1.0
## Neg Pred Value           1.00      1.0      1.0     1.00      1.0
## Prevalence               0.25      0.2      0.2     0.25      0.1
## Detection Rate           0.25      0.2      0.2     0.25      0.1
## Detection Prevalence     0.25      0.2      0.2     0.25      0.1
## Balanced Accuracy        1.00      1.0      1.0     1.00      1.0
```

```
M = y[sample(dim(y)[1], 20, replace = FALSE), ]
predM = predict(model, M)
confusionMatrix(M$classe, predM)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction A B C D E
##          A 6 0 0 0 0
##          B 0 6 0 0 0
##          C 0 0 4 0 0
##          D 0 0 0 1 0
##          E 0 0 0 0 3
##
## Overall Statistics
##
##                Accuracy : 1
##                  95% CI : (0.832, 1)
##     No Information Rate : 0.3
##     P-Value [Acc > NIR] : 3.49e-11
##
##                   Kappa : 1
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity               1.0      1.0      1.0     1.00     1.00
## Specificity               1.0      1.0      1.0     1.00     1.00
## Pos Pred Value            1.0      1.0      1.0     1.00     1.00
## Neg Pred Value            1.0      1.0      1.0     1.00     1.00
## Prevalence                0.3      0.3      0.2     0.05     0.15
## Detection Rate            0.3      0.3      0.2     0.05     0.15
## Detection Prevalence      0.3      0.3      0.2     0.05     0.15
## Balanced Accuracy         1.0      1.0      1.0     1.00     1.00
```

```
# Similarly, try a few runs with random sample of 100.. test it for 'M' of
# size 100:
M = y[sample(dim(y)[1], 100, replace = FALSE), ]
predM = predict(model, M)
confusionMatrix(M$classe, predM)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  A  B  C  D  E
##          A 27  0  0  0  0
##          B  1 21  0  0  0
##          C  0  0 20  0  0
##          D  0  0  0 14  0
##          E  0  0  0  0 17
##
## Overall Statistics
##
##                Accuracy : 0.99
##                  95% CI : (0.946, 1)
##     No Information Rate : 0.28
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.987
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity             0.964    1.000      1.0     1.00     1.00
## Specificity             1.000    0.987      1.0     1.00     1.00
## Pos Pred Value          1.000    0.955      1.0     1.00     1.00
## Neg Pred Value          0.986    1.000      1.0     1.00     1.00
## Prevalence              0.280    0.210      0.2     0.14     0.17
## Detection Rate          0.270    0.210      0.2     0.14     0.17
## Detection Prevalence    0.270    0.220      0.2     0.14     0.17
## Balanced Accuracy       0.982    0.994      1.0     1.00     1.00
```

```
# On the last few observations of the set..(just because the tail is
# easier to observe!)
predM = predict(model, y[19601:19622, ])
confusionMatrix(y$classe[19601:19622], predM)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  A  B  C  D  E
##          A  0  0  0  0  0
##          B  0  0  0  0  0
##          C  0  0  0  0  0
##          D  0  0  0  0  0
##          E  0  0  0  0 22
##
## Overall Statistics
##
##                Accuracy : 1
##                  95% CI : (0.846, 1)
##     No Information Rate : 1
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : NaN
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity                NA       NA       NA       NA        1
## Specificity                 1        1        1        1       NA
## Pos Pred Value             NA       NA       NA       NA       NA
## Neg Pred Value             NA       NA       NA       NA       NA
## Prevalence                  0        0        0        0        1
## Detection Rate              0        0        0        0        1
## Detection Prevalence        0        0        0        0        1
## Balanced Accuracy          NA       NA       NA       NA       NA
```

```
# On some observations from the middle of the data..
predM = predict(model, y[2000:2869, ])
confusionMatrix(y$classe[2000:2869], predM)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A   B   C   D   E
##          A 870   0   0   0   0
##          B   0   0   0   0   0
##          C   0   0   0   0   0
##          D   0   0   0   0   0
##          E   0   0   0   0   0
##
## Overall Statistics
##
##                Accuracy : 1
##                  95% CI : (0.996, 1)
##     No Information Rate : 1
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : NaN
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity                 1       NA       NA       NA       NA
## Specificity                NA        1        1        1        1
## Pos Pred Value             NA       NA       NA       NA       NA
## Neg Pred Value             NA       NA       NA       NA       NA
## Prevalence                  1        0        0        0        0
## Detection Rate              1        0        0        0        0
## Detection Prevalence        1        0        0        0        0
## Balanced Accuracy          NA       NA       NA       NA       NA
```

```
# THESE TESTS ABOVE ARE NOT REALLY NECESSARY BUT WERE DOEN JUST TO SEE THE
# MODEL RUN!
```

Result: Finally, the model was run on the final test set of 20 observations for which the class of exercise were not known and had to be predicted by the model. The model predicted all the 20 observations accurately (verified by the grader scripts at the URL: https://class.coursera.org/predmachlearn-011/assignment), in line with the expectation of an accuracy of 0.99 (formed on the basis of the prediction on the internal test set of 5%). The error rate on the 20 observations was 0.0. However, given a larger set, error rates of ~0.01 are reasonable to expect.

```
t = read.csv("pml-testing.csv", na.strings = c("", "NA"))
str(t)
head(t)
# Processing of this set was not considered necessary as the model used
# predictors that were already complete (without missing data) and were in
# numeric/integer and factor format. No errors were expected if the model
# tried to coerce integers to numeric and none were got!

predANS = predict(model, t)  #Finally, the predictions

# Since the project required generation of different files for each
# observation, a few files were written to the working directory..
for (i in 1:20) {
    write.table(as.character(predANS[i]), file = paste(i, ".txt"), quote = FALSE,
        row.names = FALSE, col.names = FALSE)
}
```

References: 1. Description of the weight lifting exercise. http://groupware.les.inf.puc-rio.br/har#weight_lifting_exercises 2. Random Forest: http://en.wikipedia.org/wiki/Random_forest