PREDICTING QUALITY OF DUMBBELL BICEPS CURL

by Lorelie Calima-Prall

March 4, 2016

Executive Summary

Based on weight lifting exercises data from http://groupware.les.inf.puc-rio.br/har (http://groupware.les.inf.puc-rio.br/har), the goal of this project is to predict the manner in which the dumbbell bicep curl was performed. This is the "classe" variable in the training set. This report describes how the model was built, how cross validation was implemented, what the expected out of sample error is and why certain choices were made. The resulting model is also used to predict 20 different test cases.

The training data for this project is downloaded from:
https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv
(https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv)

The test data is available from:
https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv
(https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv)

Import and Explore Data

The training data consists of 19622 observations on 127 variables, where classe is the dependent variable and the rest are explanatory variables:

```
trainurl<-"https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"
traindest<-"C:/Users/Lorelie/Documents/Classes/Coursera/Data Science/08_Practical_Machine
_Learning/04_Week/Project/traindata.csv"
download.file(trainurl,traindest)

testurl<-"https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"
testdest<-"C:/Users/Lorelie/Documents/Classes/Coursera/Data Science/08_Practical_Machine_
Learning/04_Week/Project/testdata.csv"
download.file(testurl,testdest)

train <- read.csv(traindest,na.strings=c("#DIV/0!"))
test <- read.csv(testdest,na.strings=c("#DIV/0!"))

require(dplyr)
glimpse(train)
```

```
Observations: 19,622
Variables: 160
$ X                         (int) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12...
$ user_name                 (fctr) carlitos, carlitos, carlitos, carlit...
$ raw_timestamp_part_1      (int) 1323084231, 1323084231, 1323084231, 1...
$ raw_timestamp_part_2      (int) 788290, 808298, 820366, 120339, 19632...
$ cvtd_timestamp            (fctr) 05/12/2011 11:23, 05/12/2011 11:23, ...
$ new_window                (fctr) no, no, no, no, no, no, no, no, no, ...
$ num_window                (int) 11, 11, 11, 12, 12, 12, 12, 12, 12, 1...
$ roll_belt                 (dbl) 1.41, 1.41, 1.42, 1.48, 1.48, 1.45, 1...
$ pitch_belt                (dbl) 8.07, 8.07, 8.07, 8.05, 8.07, 8.06, 8...
$ yaw_belt                  (dbl) -94.4, -94.4, -94.4, -94.4, -94.4, -9...
$ total_accel_belt          (int) 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3...
$ kurtosis_roll_belt        (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ kurtosis_picth_belt       (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ kurtosis_yaw_belt         (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_roll_belt        (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_roll_belt.1      (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_yaw_belt         (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ max_roll_belt             (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ max_picth_belt            (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ max_yaw_belt              (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ min_roll_belt             (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ min_pitch_belt            (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ min_yaw_belt              (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ amplitude_roll_belt       (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ amplitude_pitch_belt      (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ amplitude_yaw_belt        (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ var_total_accel_belt      (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_roll_belt             (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_roll_belt          (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_roll_belt             (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_pitch_belt            (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_pitch_belt         (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_pitch_belt            (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_yaw_belt              (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_yaw_belt           (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_yaw_belt              (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ gyros_belt_x              (dbl) 0.00, 0.02, 0.00, 0.02, 0.02, 0.02, 0...
$ gyros_belt_y              (dbl) 0.00, 0.00, 0.00, 0.00, 0.02, 0.00, 0...
$ gyros_belt_z              (dbl) -0.02, -0.02, -0.02, -0.03, -0.02, -0...
$ accel_belt_x              (int) -21, -22, -20, -22, -21, -21, -22, -2...
$ accel_belt_y              (int) 4, 4, 5, 3, 2, 4, 3, 4, 2, 4, 2, 2, 4...
$ accel_belt_z              (int) 22, 22, 23, 21, 24, 21, 21, 21, 24, 2...
$ 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, 60...
$ magnet_belt_z             (int) -313, -311, -305, -310, -302, -312, -...
$ roll_arm                  (dbl) -128, -128, -128, -128, -128, -128, -...
$ pitch_arm                 (dbl) 22.5, 22.5, 22.5, 22.1, 22.1, 22.0, 2...
$ yaw_arm                   (dbl) -161, -161, -161, -161, -161, -161, -...
```

```
$ total_accel_arm         (int) 34, 34, 34, 34, 34, 34, 34, 34, 34, 3...
$ var_accel_arm           (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_roll_arm            (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_roll_arm         (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_roll_arm            (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_pitch_arm           (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_pitch_arm        (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_pitch_arm           (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_yaw_arm             (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_yaw_arm          (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_yaw_arm             (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ gyros_arm_x             (dbl) 0.00, 0.02, 0.02, 0.02, 0.00, 0.02, 0...
$ gyros_arm_y             (dbl) 0.00, -0.02, -0.02, -0.03, -0.03, -0....
$ gyros_arm_z             (dbl) -0.02, -0.02, -0.02, 0.02, 0.00, 0.00...
$ accel_arm_x             (int) -288, -290, -289, -289, -289, -289, -...
$ accel_arm_y             (int) 109, 110, 110, 111, 111, 111, 111, 11...
$ accel_arm_z             (int) -123, -125, -126, -123, -123, -122, -...
$ magnet_arm_x            (int) -368, -369, -368, -372, -374, -369, -...
$ magnet_arm_y            (int) 337, 337, 344, 344, 337, 342, 336, 33...
$ magnet_arm_z            (int) 516, 513, 513, 512, 506, 513, 509, 51...
$ kurtosis_roll_arm       (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ kurtosis_picth_arm      (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ kurtosis_yaw_arm        (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_roll_arm       (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_pitch_arm      (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_yaw_arm        (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ max_roll_arm            (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ max_picth_arm           (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ max_yaw_arm             (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ min_roll_arm            (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ min_pitch_arm           (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ min_yaw_arm             (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ amplitude_roll_arm      (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ amplitude_pitch_arm     (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ amplitude_yaw_arm       (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ roll_dumbbell           (dbl) 13.05217, 13.13074, 12.85075, 13.4312...
$ pitch_dumbbell          (dbl) -70.49400, -70.63751, -70.27812, -70....
$ yaw_dumbbell            (dbl) -84.87394, -84.71065, -85.14078, -84....
$ kurtosis_roll_dumbbell  (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ kurtosis_picth_dumbbell (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ kurtosis_yaw_dumbbell   (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_roll_dumbbell  (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_pitch_dumbbell (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_yaw_dumbbell   (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ max_roll_dumbbell       (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ max_picth_dumbbell      (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ max_yaw_dumbbell        (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ min_roll_dumbbell       (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ min_pitch_dumbbell      (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ min_yaw_dumbbell        (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
```

```
$ amplitude_roll_dumbbell   (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ amplitude_pitch_dumbbell  (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ amplitude_yaw_dumbbell    (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ total_accel_dumbbell      (int) 37, 37, 37, 37, 37, 37, 37, 37, 37, 3...
$ var_accel_dumbbell        (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_roll_dumbbell         (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_roll_dumbbell      (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_roll_dumbbell         (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_pitch_dumbbell        (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_pitch_dumbbell     (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_pitch_dumbbell        (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_yaw_dumbbell          (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_yaw_dumbbell       (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_yaw_dumbbell          (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ gyros_dumbbell_x          (dbl) 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0...
$ gyros_dumbbell_y          (dbl) -0.02, -0.02, -0.02, -0.02, -0.02, -0...
$ gyros_dumbbell_z          (dbl) 0.00, 0.00, 0.00, -0.02, 0.00, 0.00, ...
$ accel_dumbbell_x          (int) -234, -233, -232, -232, -233, -234, -...
$ accel_dumbbell_y          (int) 47, 47, 46, 48, 48, 48, 47, 46, 47, 4...
$ accel_dumbbell_z          (int) -271, -269, -270, -269, -270, -269, -...
$ magnet_dumbbell_x         (int) -559, -555, -561, -552, -554, -558, -...
$ magnet_dumbbell_y         (int) 293, 296, 298, 303, 292, 294, 295, 30...
$ magnet_dumbbell_z         (dbl) -65, -64, -63, -60, -68, -66, -70, -7...
$ roll_forearm             (dbl) 28.4, 28.3, 28.3, 28.1, 28.0, 27.9, 2...
$ pitch_forearm            (dbl) -63.9, -63.9, -63.9, -63.9, -63.9, -6...
$ yaw_forearm              (dbl) -153, -153, -152, -152, -152, -152, -...
$ kurtosis_roll_forearm     (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ kurtosis_picth_forearm    (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ kurtosis_yaw_forearm      (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_roll_forearm     (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_pitch_forearm    (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ skewness_yaw_forearm      (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ max_roll_forearm         (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ max_picth_forearm        (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ max_yaw_forearm          (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ min_roll_forearm         (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ min_pitch_forearm        (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ min_yaw_forearm          (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ amplitude_roll_forearm   (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ amplitude_pitch_forearm  (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ amplitude_yaw_forearm    (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
$ total_accel_forearm      (int) 36, 36, 36, 36, 36, 36, 36, 36, 36, 3...
$ var_accel_forearm        (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_roll_forearm         (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_roll_forearm      (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_roll_forearm         (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_pitch_forearm        (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ stddev_pitch_forearm     (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_pitch_forearm        (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ avg_yaw_forearm          (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
```

```
$ stddev_yaw_forearm        (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ var_yaw_forearm           (fctr) NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
$ gyros_forearm_x           (dbl) 0.03, 0.02, 0.03, 0.02, 0.02, 0.02, 0...
$ gyros_forearm_y           (dbl) 0.00, 0.00, -0.02, -0.02, 0.00, -0.02...
$ gyros_forearm_z           (dbl) -0.02, -0.02, 0.00, 0.00, -0.02, -0.0...
$ accel_forearm_x           (int) 192, 192, 196, 189, 189, 193, 195, 19...
$ accel_forearm_y           (int) 203, 203, 204, 206, 206, 203, 205, 20...
$ accel_forearm_z           (int) -215, -216, -213, -214, -214, -215, -...
$ magnet_forearm_x          (int) -17, -18, -18, -16, -17, -9, -18, -9,...
$ magnet_forearm_y          (dbl) 654, 661, 658, 658, 655, 660, 659, 66...
$ magnet_forearm_z          (dbl) 476, 473, 469, 469, 473, 478, 470, 47...
$ classe                    (fctr) A, A, A, A, A, A, A, A, A, A, A, A, ...
```

Clean Data

Remove columns with NA values, those that are insignificant and have zero variance.

```
library(caret)
train<-train[,colSums(is.na(train)) ==0]
colnames(train)[colSums(is.na(train)) > 0]
```

```
character(0)
```

```
dropcols<-names(train) %in% c('X', 'user_name', 'raw_timestamp_part_1', 'raw_timestamp_pa
rt_2', 'cvtd_timestamp', 'new_window', 'num_window')
train <- train[!dropcols]

nzv <- nearZeroVar(train)
train <- train[, -nzv]
```

Split Data for Cross-Validation

```
library(caret)
trainIndex <- createDataPartition(train$classe, p =0.8,list = FALSE,times=1)
train <-train[trainIndex,]
valid <-train[-trainIndex,]
```

Random Forests Model

```
library(randomForest)
trainrf<-randomForest(classe~.,data=train,importance=TRUE)
trainrf
```

```
Call:
 randomForest(formula = classe ~ ., data = train, importance = TRUE)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 7

        OOB estimate of  error rate: 0.34%
Confusion matrix:
     A    B    C    D    E  class.error
A 4461    2    0    0    1 0.0006720430
B   10 3024    4    0    0 0.0046082949
C    0   16 2719    3    0 0.0069393718
D    0    0   16 2557    0 0.0062184221
E    0    0    0    2 2884 0.0006930007
```
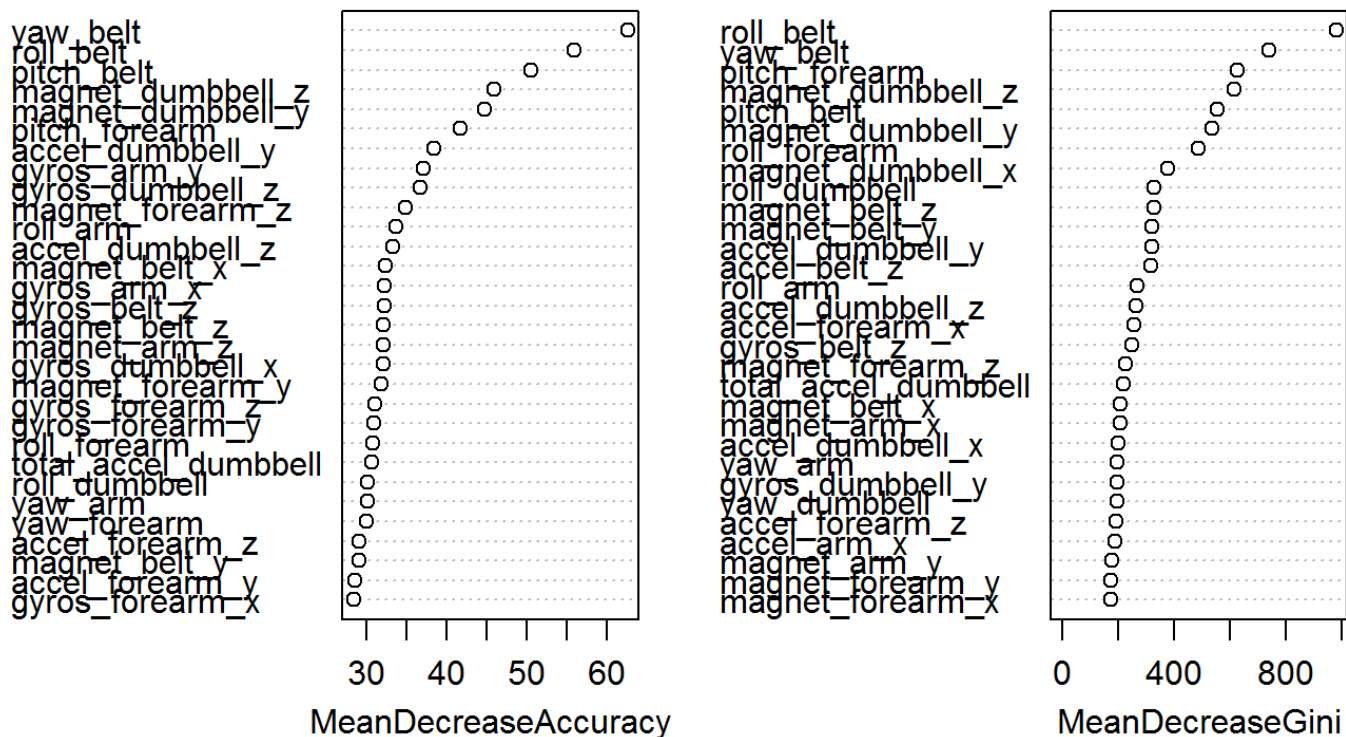
```
varImpPlot(trainrf)
```


trainrf

Cross Validation Testing and Out-of-Sample Error Estimate

```
predvalid<-predict(trainrf,valid)
confusionMatrix(valid$classe, predvalid)
```

```
Confusion Matrix and Statistics

          Reference
Prediction   A   B   C   D   E
         A 874   0   0   0   0
         B   0 608   0   0   0
         C   0   0 578   0   0
         D   0   0   0 501   0
         E   0   0   0   0 574

Overall Statistics

               Accuracy : 1
                 95% CI : (0.9988, 1)
    No Information Rate : 0.2788
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 1
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: A Class: B Class: C Class: D Class: E
Sensitivity            1.0000   1.0000   1.0000   1.0000   1.0000
Specificity            1.0000   1.0000   1.0000   1.0000   1.0000
Pos Pred Value         1.0000   1.0000   1.0000   1.0000   1.0000
Neg Pred Value         1.0000   1.0000   1.0000   1.0000   1.0000
Prevalence             0.2788   0.1939   0.1844   0.1598   0.1831
Detection Rate         0.2788   0.1939   0.1844   0.1598   0.1831
Detection Prevalence   0.2788   0.1939   0.1844   0.1598   0.1831
Balanced Accuracy      1.0000   1.0000   1.0000   1.0000   1.0000
```

```
accuracy <- postResample(valid$classe, predvalid)
accuracy <-accuracy[[1]]
accuracy
```

```
[1] 1
```

```
oose<-1-accuracy
oose
```

```
[1] 0
```

The estimated accuracy of the model is 100% and the estimated out-of-sample error based on our fitted model applied to the cross validation dataset is 0%.

Predicting with Random Forests

```
pred<-predict(trainrf,test)
pred
```

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
 B  A  B  A  A  E  D  B  A  A  B  C  B  A  E  E  A  B  B  B
Levels: A B C D E
```