

PML01

GgYy

20160710

1. Install needed packages

```
#install.packages("data.table")
#install.packages("caret")
#install.packages("randomForest")
#install.packages("foreach")
#install.packages("rpart")
#install.packages("rpart.plot")
#install.packages("corrplot")
```

2. Load needed packages

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.2.5
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.2.5
```

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

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

```
## Warning: package 'ggplot2' was built under R version 3.2.5
```

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 3.2.5
```

```
## randomForest 4.6-12
```

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

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

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

```
library(foreach)
```

```
## Warning: package 'foreach' was built under R version 3.2.5
```

```
library(rpart)
```

```
## Warning: package 'rpart' was built under R version 3.2.5
```

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

```
## Warning: package 'rpart.plot' was built under R version 3.2.5
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 3.2.5
```

3.Read datas

```
training_data <- read.csv("pml-training.csv", na.strings=c("#DIV/0!", " ", "",  
"NA", "NAs", "NULL"))  
testing_data <- read.csv("pml-testing.csv", na.strings=c("#DIV/0!", " ", "", "N  
A", "NAs", "NULL"))
```

4.Clean datas Drop NAs, Drop highly corelated variables, drop variables whose contents are the same.

```
#4.1 Drop columns with NAs  
str(training_data)
```

```

## '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 2 2 ...
##  $ raw_timestamp_part_1     : int  1323084231 1323084231 1323084231 132308423
2 1323084232 1323084232 1323084232 1323084232 1323084232 1323084232 ...
##  $ raw_timestamp_part_2     : int  788290 808298 820366 120339 196328 304277
368296 440390 484323 484434 ...
##  $ cvtd_timestamp           : Factor w/ 20 levels "02/12/2011 13:32",...: 9 9
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 -94.4 ...
##  $ total_accel_belt          : int  3 3 3 3 3 3 3 3 3 3 ...
##  $ kurtosis_roll_belt        : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ kurtosis_pitch_belt       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ kurtosis_yaw_belt         : logi  NA NA NA NA NA NA NA ...
##  $ skewness_roll_belt        : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ skewness_roll_belt.1      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ skewness_yaw_belt         : logi  NA NA NA NA NA NA NA ...
##  $ max_roll_belt              : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_pitch_belt            : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_yaw_belt               : num  NA NA NA 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               : num  NA NA NA 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         : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ 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 0.0
3 ...
##  $ 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 0 ...
## $ accel_belt_x      : int   -21 -22 -20 -22 -21 -21 -22 -22 -20 -2
1 ...
## $ 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 60
9 ...
## $ magnet_belt_z     : int  -313 -311 -305 -310 -302 -312 -311 -313 -3
12 -308 ...
## $ roll_arm          : num  -128 -128 -128 -128 -128 -128 -128 -128 -1
28 -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 -1
61 -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 0.0
2 ...
## $ gyros_arm_y       : num    0 -0.02 -0.02 -0.03 -0.03 -0.03 -0.03 -0.0
2 -0.03 -0.03 ...
## $ gyros_arm_z       : num  -0.02 -0.02 -0.02 0.02 0 0 0 0 -0.02 -0.0
2 ...
## $ accel_arm_x       : int  -288 -290 -289 -289 -289 -289 -289 -289 -2
88 -288 ...
## $ accel_arm_y       : int   109 110 110 111 111 111 111 111 109 11
0 ...
## $ accel_arm_z       : int  -123 -125 -126 -123 -123 -122 -125 -124 -1
22 -124 ...
## $ magnet_arm_x      : int  -368 -369 -368 -372 -374 -369 -373 -372 -3
69 -376 ...
## $ magnet_arm_y      : int   337 337 344 344 337 342 336 338 341 33
4 ...
## $ magnet_arm_z      : int   516 513 513 512 506 513 509 510 518 51
6 ...
## $ kurtosis_roll_arm : num   NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_arm : num   NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_arm  : num   NA NA NA NA NA NA NA NA NA NA ...

```

```
## $ skewness_roll_arm      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_pitch_arm    : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_arm      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_roll_arm          : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_arm         : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_arm           : int   NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_arm          : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_arm         : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_arm           : int   NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_arm    : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_arm   : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_arm     : int   NA 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 : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_dumbbell : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_dumbbell  : logi   NA NA NA NA NA NA NA ...
## $ skewness_roll_dumbbell : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_pitch_dumbbell : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_dumbbell  : logi   NA NA NA NA NA NA NA ...
## $ max_roll_dumbbell     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_dumbbell    : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_dumbbell      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_dumbbell     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_dumbbell    : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_dumbbell      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_dumbbell : num  NA NA NA NA NA NA NA NA NA NA NA ...
## [list output truncated]
```

```
cleantraining <- training_data[, -which(names(training_data) %in% c("X", "user_
name", "raw_timestamp_part_1", "raw_timestamp_part_2", "cvtd_timestamp", "new_w
indow", "num_window"))]
cleantraining = cleantraining[, colSums(is.na(cleantraining)) == 0]
#4.2 Drop variables with same content
zerovariance =nearZeroVar(cleantraining[apply(cleantraining, is.numeric)], sav
eMetrics=TRUE)
cleantraining = cleantraining[, zerovariance[, 'nzv'] == 0]
#4.3.1 Return the correlation matrix in matrix format
correlationmatrix <- cor(na.omit(cleantraining[apply(cleantraining, is.numeri
c)]))
dim(correlationmatrix)
```

```
## [1] 52 52
```

```
correlationmatrixdegreesoffreedom <- expand.grid(row = 1:52, col = 1:52)
correlationmatrixdegreesoffreedom$correlation <- as.vector(correlationmatrix)
#4.3.2 Remove highly correlated variables (up to 0.7)
removehighcorrelation <- findCorrelation(correlationmatrix, cutoff = .7, verbose = TRUE)
```

```
## Compare row 10 and column 1 with corr 0.992
## Means: 0.27 vs 0.168 so flagging column 10
## Compare row 1 and column 9 with corr 0.925
## Means: 0.25 vs 0.164 so flagging column 1
## Compare row 9 and column 22 with corr 0.722
## Means: 0.233 vs 0.161 so flagging column 9
## Compare row 22 and column 4 with corr 0.759
## Means: 0.224 vs 0.158 so flagging column 22
## Compare row 4 and column 3 with corr 0.762
## Means: 0.2 vs 0.155 so flagging column 4
## Compare row 3 and column 8 with corr 0.708
## Means: 0.2 vs 0.153 so flagging column 3
## Compare row 36 and column 29 with corr 0.849
## Means: 0.257 vs 0.151 so flagging column 36
## Compare row 8 and column 2 with corr 0.966
## Means: 0.229 vs 0.146 so flagging column 8
## Compare row 2 and column 11 with corr 0.884
## Means: 0.212 vs 0.143 so flagging column 2
## Compare row 37 and column 38 with corr 0.769
## Means: 0.198 vs 0.139 so flagging column 37
## Compare row 35 and column 30 with corr 0.773
## Means: 0.195 vs 0.137 so flagging column 35
## Compare row 38 and column 5 with corr 0.781
## Means: 0.177 vs 0.134 so flagging column 38
## Compare row 21 and column 24 with corr 0.814
## Means: 0.176 vs 0.133 so flagging column 21
## Compare row 34 and column 28 with corr 0.808
## Means: 0.176 vs 0.13 so flagging column 34
## Compare row 23 and column 26 with corr 0.779
## Means: 0.137 vs 0.129 so flagging column 23
## Compare row 25 and column 24 with corr 0.792
## Means: 0.145 vs 0.128 so flagging column 25
## Compare row 12 and column 13 with corr 0.779
## Means: 0.122 vs 0.127 so flagging column 13
## Compare row 48 and column 51 with corr 0.772
## Means: 0.145 vs 0.127 so flagging column 48
## Compare row 19 and column 18 with corr 0.918
## Means: 0.095 vs 0.127 so flagging column 18
## Compare row 46 and column 45 with corr 0.846
## Means: 0.131 vs 0.129 so flagging column 46
## Compare row 45 and column 31 with corr 0.71
## Means: 0.098 vs 0.129 so flagging column 31
## Compare row 45 and column 33 with corr 0.716
## Means: 0.078 vs 0.132 so flagging column 33
## All correlations <= 0.7
```

```

cleantraining <- cleantraining[, -removehighcorrelation]
#4.4 Generally drop blanks
for(i in c(8:ncol(cleantraining)-1)) {cleantraining[,i] = as.numeric(as.character(cleantraining[,i]))}

for(i in c(8:ncol(testing_data)-1)) {testing_data[,i] = as.numeric(as.character(testing_data[,i]))}
#4.5 Redefine to be used data
featureset <- colnames(cleantraining[colSums(is.na(cleantraining)) == 0])[-(1:7)]
modeldata <- cleantraining[featureset]
featureset

```

```

## [1] "yaw_arm"          "total_accel_arm"    "gyros_arm_y"
## [4] "gyros_arm_z"      "magnet_arm_x"       "magnet_arm_z"
## [7] "roll_dumbbell"    "pitch_dumbbell"     "yaw_dumbbell"
## [10] "total_accel_dumbbell" "gyros_dumbbell_y"   "magnet_dumbbell_z"
## [13] "roll_forearm"     "pitch_forearm"      "yaw_forearm"
## [16] "total_accel_forearm" "gyros_forearm_x"    "gyros_forearm_y"
## [19] "accel_forearm_x"   "accel_forearm_z"    "magnet_forearm_x"
## [22] "magnet_forearm_y"  "magnet_forearm_z"   "classe"

```

5.Build model Split 60% for training and 40% for testing.

```

idx <- createDataPartition(modeldata$classe, p=0.6, list=FALSE )
training <- modeldata[idx,]
testing <- modeldata[-idx,]

```

5 fold cross validation is used.

```

control <- trainControl(method="cv", 5)
modelRF<- train(classe ~ ., data=training, method="rf", trControl=control)
modelLDA<-train(classe ~ ., data=training, method="lda", trControl=control)

```

```
## Loading required package: MASS
```

```
modelGBM<-train(classe ~ ., data=training, method="gbm", trControl=control)
```

```
## Loading required package: gbm
```

```
## Warning: package 'gbm' was built under R version 3.2.5
```

```
## Loading required package: survival
```



```
##  
## Attaching package: 'survival'
```

```
## The following object is masked from 'package:caret':  
##  
##      cluster
```

```
## Loading required package: splines
```

```
## Loading required package: parallel
```

```
## Loaded gbm 2.1.1
```

```
## Loading required package: plyr
```

##	Iter	TrainDeviance	ValidDeviance	StepSize	Improve
##	1	1.6094	nan	0.1000	0.0985
##	2	1.5534	nan	0.1000	0.0643
##	3	1.5157	nan	0.1000	0.0480
##	4	1.4875	nan	0.1000	0.0402
##	5	1.4646	nan	0.1000	0.0301
##	6	1.4466	nan	0.1000	0.0350
##	7	1.4266	nan	0.1000	0.0324
##	8	1.4074	nan	0.1000	0.0259
##	9	1.3916	nan	0.1000	0.0239
##	10	1.3775	nan	0.1000	0.0235
##	20	1.2687	nan	0.1000	0.0134
##	40	1.1432	nan	0.1000	0.0084
##	60	1.0623	nan	0.1000	0.0053
##	80	1.0040	nan	0.1000	0.0024
##	100	0.9601	nan	0.1000	0.0026
##	120	0.9230	nan	0.1000	0.0013
##	140	0.8915	nan	0.1000	0.0016
##	150	0.8773	nan	0.1000	0.0009
##					
##	Iter	TrainDeviance	ValidDeviance	StepSize	Improve
##	1	1.6094	nan	0.1000	0.1372
##	2	1.5313	nan	0.1000	0.0963
##	3	1.4755	nan	0.1000	0.0751
##	4	1.4304	nan	0.1000	0.0623
##	5	1.3933	nan	0.1000	0.0586
##	6	1.3592	nan	0.1000	0.0478
##	7	1.3318	nan	0.1000	0.0437
##	8	1.3057	nan	0.1000	0.0333
##	9	1.2852	nan	0.1000	0.0355
##	10	1.2621	nan	0.1000	0.0414
##	20	1.1102	nan	0.1000	0.0132
##	40	0.9375	nan	0.1000	0.0091
##	60	0.8331	nan	0.1000	0.0054
##	80	0.7538	nan	0.1000	0.0035
##	100	0.6932	nan	0.1000	0.0035
##	120	0.6448	nan	0.1000	0.0023
##	140	0.6034	nan	0.1000	0.0024
##	150	0.5844	nan	0.1000	0.0028
##					
##	Iter	TrainDeviance	ValidDeviance	StepSize	Improve
##	1	1.6094	nan	0.1000	0.1660
##	2	1.5151	nan	0.1000	0.1115
##	3	1.4494	nan	0.1000	0.0981
##	4	1.3936	nan	0.1000	0.0920
##	5	1.3404	nan	0.1000	0.0660
##	6	1.2999	nan	0.1000	0.0611
##	7	1.2635	nan	0.1000	0.0484

##	8	1.2335	nan	0.1000	0.0384
##	9	1.2099	nan	0.1000	0.0443
##	10	1.1827	nan	0.1000	0.0477
##	20	0.9972	nan	0.1000	0.0171
##	40	0.8075	nan	0.1000	0.0081
##	60	0.6942	nan	0.1000	0.0078
##	80	0.6128	nan	0.1000	0.0031
##	100	0.5486	nan	0.1000	0.0032
##	120	0.4965	nan	0.1000	0.0018
##	140	0.4540	nan	0.1000	0.0014
##	150	0.4357	nan	0.1000	0.0011

##

##	Iter	TrainDeviance	ValidDeviance	StepSize	Improve
##	1	1.6094	nan	0.1000	0.0999
##	2	1.5544	nan	0.1000	0.0675
##	3	1.5171	nan	0.1000	0.0510
##	4	1.4876	nan	0.1000	0.0404
##	5	1.4643	nan	0.1000	0.0378
##	6	1.4425	nan	0.1000	0.0287
##	7	1.4255	nan	0.1000	0.0275
##	8	1.4088	nan	0.1000	0.0266
##	9	1.3923	nan	0.1000	0.0237
##	10	1.3779	nan	0.1000	0.0220
##	20	1.2677	nan	0.1000	0.0153
##	40	1.1403	nan	0.1000	0.0058
##	60	1.0616	nan	0.1000	0.0039
##	80	1.0023	nan	0.1000	0.0026
##	100	0.9579	nan	0.1000	0.0036
##	120	0.9198	nan	0.1000	0.0012
##	140	0.8871	nan	0.1000	0.0014
##	150	0.8715	nan	0.1000	0.0009

##

##	Iter	TrainDeviance	ValidDeviance	StepSize	Improve
##	1	1.6094	nan	0.1000	0.1374
##	2	1.5317	nan	0.1000	0.0998
##	3	1.4761	nan	0.1000	0.0753
##	4	1.4318	nan	0.1000	0.0616
##	5	1.3960	nan	0.1000	0.0510
##	6	1.3648	nan	0.1000	0.0409
##	7	1.3395	nan	0.1000	0.0410
##	8	1.3148	nan	0.1000	0.0422
##	9	1.2885	nan	0.1000	0.0478
##	10	1.2615	nan	0.1000	0.0317
##	20	1.1065	nan	0.1000	0.0186
##	40	0.9293	nan	0.1000	0.0099
##	60	0.8244	nan	0.1000	0.0052
##	80	0.7486	nan	0.1000	0.0026
##	100	0.6854	nan	0.1000	0.0035
##	120	0.6349	nan	0.1000	0.0022

```

##      140      0.5947      nan      0.1000      0.0013
##      150      0.5782      nan      0.1000      0.0013
##
## Iter    TrainDeviance    ValidDeviance    StepSize    Improve
##      1      1.6094      nan      0.1000      0.1678
##      2      1.5142      nan      0.1000      0.1182
##      3      1.4437      nan      0.1000      0.0890
##      4      1.3919      nan      0.1000      0.0764
##      5      1.3464      nan      0.1000      0.0611
##      6      1.3095      nan      0.1000      0.0667
##      7      1.2690      nan      0.1000      0.0494
##      8      1.2390      nan      0.1000      0.0543
##      9      1.2063      nan      0.1000      0.0556
##     10      1.1722      nan      0.1000      0.0452
##     20      0.9915      nan      0.1000      0.0250
##     40      0.8013      nan      0.1000      0.0102
##     60      0.6879      nan      0.1000      0.0054
##     80      0.6063      nan      0.1000      0.0044
##    100      0.5442      nan      0.1000      0.0029
##    120      0.4944      nan      0.1000      0.0018
##    140      0.4527      nan      0.1000      0.0013
##    150      0.4346      nan      0.1000      0.0019
##
## Iter    TrainDeviance    ValidDeviance    StepSize    Improve
##      1      1.6094      nan      0.1000      0.0961
##      2      1.5532      nan      0.1000      0.0655
##      3      1.5149      nan      0.1000      0.0513
##      4      1.4865      nan      0.1000      0.0391
##      5      1.4645      nan      0.1000      0.0344
##      6      1.4442      nan      0.1000      0.0362
##      7      1.4233      nan      0.1000      0.0289
##      8      1.4062      nan      0.1000      0.0280
##      9      1.3895      nan      0.1000      0.0231
##     10      1.3747      nan      0.1000      0.0239
##     20      1.2623      nan      0.1000      0.0130
##     40      1.1344      nan      0.1000      0.0076
##     60      1.0546      nan      0.1000      0.0045
##     80      0.9995      nan      0.1000      0.0027
##    100      0.9548      nan      0.1000      0.0025
##    120      0.9170      nan      0.1000      0.0018
##    140      0.8860      nan      0.1000      0.0008
##    150      0.8711      nan      0.1000      0.0014
##
## Iter    TrainDeviance    ValidDeviance    StepSize    Improve
##      1      1.6094      nan      0.1000      0.1288
##      2      1.5318      nan      0.1000      0.0995
##      3      1.4758      nan      0.1000      0.0768
##      4      1.4302      nan      0.1000      0.0687
##      5      1.3893      nan      0.1000      0.0574

```

##	6	1.3552	nan	0.1000	0.0424
##	7	1.3294	nan	0.1000	0.0400
##	8	1.3049	nan	0.1000	0.0432
##	9	1.2789	nan	0.1000	0.0357
##	10	1.2564	nan	0.1000	0.0364
##	20	1.0959	nan	0.1000	0.0145
##	40	0.9299	nan	0.1000	0.0078
##	60	0.8208	nan	0.1000	0.0042
##	80	0.7503	nan	0.1000	0.0032
##	100	0.6859	nan	0.1000	0.0028
##	120	0.6346	nan	0.1000	0.0017
##	140	0.5908	nan	0.1000	0.0018
##	150	0.5722	nan	0.1000	0.0011

##

##	Iter	TrainDeviance	ValidDeviance	StepSize	Improve
##	1	1.6094	nan	0.1000	0.1686
##	2	1.5126	nan	0.1000	0.1130
##	3	1.4441	nan	0.1000	0.0947
##	4	1.3882	nan	0.1000	0.0814
##	5	1.3405	nan	0.1000	0.0704
##	6	1.2989	nan	0.1000	0.0584
##	7	1.2628	nan	0.1000	0.0479
##	8	1.2333	nan	0.1000	0.0540
##	9	1.2018	nan	0.1000	0.0410
##	10	1.1756	nan	0.1000	0.0357
##	20	0.9906	nan	0.1000	0.0212
##	40	0.8009	nan	0.1000	0.0099
##	60	0.6827	nan	0.1000	0.0089
##	80	0.5998	nan	0.1000	0.0023
##	100	0.5410	nan	0.1000	0.0033
##	120	0.4935	nan	0.1000	0.0016
##	140	0.4504	nan	0.1000	0.0017
##	150	0.4302	nan	0.1000	0.0021

##

##	Iter	TrainDeviance	ValidDeviance	StepSize	Improve
##	1	1.6094	nan	0.1000	0.1014
##	2	1.5525	nan	0.1000	0.0644
##	3	1.5146	nan	0.1000	0.0513
##	4	1.4854	nan	0.1000	0.0392
##	5	1.4623	nan	0.1000	0.0337
##	6	1.4428	nan	0.1000	0.0351
##	7	1.4219	nan	0.1000	0.0263
##	8	1.4052	nan	0.1000	0.0230
##	9	1.3904	nan	0.1000	0.0263
##	10	1.3733	nan	0.1000	0.0243
##	20	1.2615	nan	0.1000	0.0132
##	40	1.1361	nan	0.1000	0.0063
##	60	1.0568	nan	0.1000	0.0042
##	80	1.0009	nan	0.1000	0.0038

```

##      100      0.9545      nan      0.1000      0.0026
##      120      0.9163      nan      0.1000      0.0016
##      140      0.8823      nan      0.1000      0.0017
##      150      0.8681      nan      0.1000      0.0009
##
## Iter    TrainDeviance    ValidDeviance    StepSize    Improve
##      1      1.6094      nan      0.1000      0.1342
##      2      1.5309      nan      0.1000      0.0959
##      3      1.4753      nan      0.1000      0.0751
##      4      1.4312      nan      0.1000      0.0646
##      5      1.3938      nan      0.1000      0.0568
##      6      1.3610      nan      0.1000      0.0425
##      7      1.3354      nan      0.1000      0.0405
##      8      1.3107      nan      0.1000      0.0397
##      9      1.2875      nan      0.1000      0.0398
##     10      1.2636      nan      0.1000      0.0405
##     20      1.1060      nan      0.1000      0.0214
##     40      0.9377      nan      0.1000      0.0107
##     60      0.8278      nan      0.1000      0.0041
##     80      0.7510      nan      0.1000      0.0034
##    100      0.6938      nan      0.1000      0.0020
##    120      0.6409      nan      0.1000      0.0024
##    140      0.5976      nan      0.1000      0.0018
##    150      0.5801      nan      0.1000      0.0023
##
## Iter    TrainDeviance    ValidDeviance    StepSize    Improve
##      1      1.6094      nan      0.1000      0.1685
##      2      1.5141      nan      0.1000      0.1162
##      3      1.4459      nan      0.1000      0.0985
##      4      1.3875      nan      0.1000      0.0710
##      5      1.3457      nan      0.1000      0.0787
##      6      1.2997      nan      0.1000      0.0676
##      7      1.2610      nan      0.1000      0.0451
##      8      1.2330      nan      0.1000      0.0492
##      9      1.2038      nan      0.1000      0.0486
##     10      1.1752      nan      0.1000      0.0396
##     20      0.9845      nan      0.1000      0.0209
##     40      0.8023      nan      0.1000      0.0104
##     60      0.6882      nan      0.1000      0.0051
##     80      0.6061      nan      0.1000      0.0046
##    100      0.5423      nan      0.1000      0.0047
##    120      0.4875      nan      0.1000      0.0012
##    140      0.4477      nan      0.1000      0.0013
##    150      0.4312      nan      0.1000      0.0025
##
## Iter    TrainDeviance    ValidDeviance    StepSize    Improve
##      1      1.6094      nan      0.1000      0.0957
##      2      1.5537      nan      0.1000      0.0673
##      3      1.5162      nan      0.1000      0.0494

```

```

##      4      1.4874      nan      0.1000      0.0390
##      5      1.4641      nan      0.1000      0.0307
##      6      1.4447      nan      0.1000      0.0333
##      7      1.4257      nan      0.1000      0.0276
##      8      1.4090      nan      0.1000      0.0269
##      9      1.3928      nan      0.1000      0.0265
##     10      1.3771      nan      0.1000      0.0235
##     20      1.2690      nan      0.1000      0.0121
##     40      1.1423      nan      0.1000      0.0081
##     60      1.0627      nan      0.1000      0.0049
##     80      1.0034      nan      0.1000      0.0030
##    100      0.9575      nan      0.1000      0.0019
##    120      0.9209      nan      0.1000      0.0022
##    140      0.8880      nan      0.1000      0.0011
##    150      0.8735      nan      0.1000      0.0004
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1      1.6094      nan      0.1000      0.1351
##      2      1.5335      nan      0.1000      0.0976
##      3      1.4772      nan      0.1000      0.0727
##      4      1.4351      nan      0.1000      0.0590
##      5      1.3998      nan      0.1000      0.0638
##      6      1.3631      nan      0.1000      0.0461
##      7      1.3347      nan      0.1000      0.0409
##      8      1.3092      nan      0.1000      0.0457
##      9      1.2829      nan      0.1000      0.0364
##     10      1.2597      nan      0.1000      0.0357
##     20      1.1027      nan      0.1000      0.0167
##     40      0.9337      nan      0.1000      0.0114
##     60      0.8302      nan      0.1000      0.0072
##     80      0.7512      nan      0.1000      0.0040
##    100      0.6890      nan      0.1000      0.0031
##    120      0.6343      nan      0.1000      0.0034
##    140      0.5913      nan      0.1000      0.0011
##    150      0.5746      nan      0.1000      0.0011
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1      1.6094      nan      0.1000      0.1620
##      2      1.5156      nan      0.1000      0.1191
##      3      1.4454      nan      0.1000      0.0880
##      4      1.3930      nan      0.1000      0.0811
##      5      1.3447      nan      0.1000      0.0688
##      6      1.3032      nan      0.1000      0.0591
##      7      1.2652      nan      0.1000      0.0648
##      8      1.2280      nan      0.1000      0.0444
##      9      1.1996      nan      0.1000      0.0388
##     10      1.1761      nan      0.1000      0.0364
##     20      0.9917      nan      0.1000      0.0143
##     40      0.8013      nan      0.1000      0.0075

```

```
##      60      0.6856      nan      0.1000      0.0051
##      80      0.6053      nan      0.1000      0.0017
##     100      0.5473      nan      0.1000      0.0032
##     120      0.4951      nan      0.1000      0.0020
##     140      0.4532      nan      0.1000      0.0006
##     150      0.4344      nan      0.1000      0.0018
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1      1.6094      nan      0.1000      0.1648
##      2      1.5162      nan      0.1000      0.1156
##      3      1.4490      nan      0.1000      0.0996
##      4      1.3895      nan      0.1000      0.0776
##      5      1.3436      nan      0.1000      0.0654
##      6      1.3033      nan      0.1000      0.0610
##      7      1.2672      nan      0.1000      0.0577
##      8      1.2345      nan      0.1000      0.0465
##      9      1.2065      nan      0.1000      0.0458
##     10      1.1780      nan      0.1000      0.0398
##     20      0.9922      nan      0.1000      0.0171
##     40      0.8048      nan      0.1000      0.0092
##     60      0.6882      nan      0.1000      0.0053
##     80      0.6076      nan      0.1000      0.0041
##    100      0.5466      nan      0.1000      0.0030
##    120      0.4960      nan      0.1000      0.0022
##    140      0.4595      nan      0.1000      0.0019
##    150      0.4405      nan      0.1000      0.0016
```

```
modelRpart<-train(classe ~ ., data=training, method="rpart", trControl=control)
```

Review the performrance of the model

```
predictRF<-predict(modelRF,testing)
predictLDA<-predict(modelLDA,testing)
predictGBM<-predict(modelGBM,testing)
predictRpart<-predict(modelRpart,testing)

confusionMatrix(testing$classe,predictRF)
```



```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A     B     C     D     E
##           A 2224     1     2     4     1
##           B  42 1464     9     1     2
##           C   0   25 1330     8     5
##           D   1    0   55 1216    14
##           E   2   11    7   16 1406
##
## Overall Statistics
##
##           Accuracy : 0.9737
##           95% CI : (0.97, 0.9772)
##           No Information Rate : 0.2892
##           P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.9668
##           McNemar's Test P-Value : 1.072e-15
##
## Statistics by Class:
##
##           Class: A Class: B Class: C Class: D Class: E
## Sensitivity          0.9802   0.9753   0.9480   0.9767   0.9846
## Specificity          0.9986   0.9915   0.9941   0.9894   0.9944
## Pos Pred Value       0.9964   0.9644   0.9722   0.9456   0.9750
## Neg Pred Value       0.9920   0.9942   0.9887   0.9956   0.9966
## Prevalence           0.2892   0.1913   0.1788   0.1587   0.1820
## Detection Rate       0.2835   0.1866   0.1695   0.1550   0.1792
## Detection Prevalence 0.2845   0.1935   0.1744   0.1639   0.1838
## Balanced Accuracy     0.9894   0.9834   0.9710   0.9831   0.9895
```

```
confusionMatrix(testing$classe,predictLDA)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A     B     C     D     E
##           A 1468  273  241  167   83
##           B  377  553  153  159  276
##           C  261  120  765   91  131
##           D   94  147  180  691  174
##           E  149  343  285  246  419
##
## Overall Statistics
##
##           Accuracy : 0.4966
##           95% CI : (0.4854, 0.5077)
##           No Information Rate : 0.2994
##           P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.3625
##           McNemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##           Class: A Class: B Class: C Class: D Class: E
## Sensitivity          0.6249  0.38510  0.4711  0.51034  0.3869
## Specificity          0.8610  0.84945  0.9031  0.90835  0.8487
## Pos Pred Value       0.6577  0.36430  0.5592  0.53733  0.2906
## Neg Pred Value       0.8431  0.86046  0.8674  0.89893  0.8963
## Prevalence           0.2994  0.18302  0.2070  0.17257  0.1380
## Detection Rate       0.1871  0.07048  0.0975  0.08807  0.0534
## Detection Prevalence 0.2845  0.19347  0.1744  0.16391  0.1838
## Balanced Accuracy     0.7430  0.61728  0.6871  0.70934  0.6178
```

```
confusionMatrix(testing$classe,predictGBM)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A     B     C     D     E
##           A 2119    26    30    43    14
##           B  138 1193    84    25    78
##           C   20   74 1191    36    47
##           D   22   30  105 1090    39
##           E   13  102   67  139 1121
##
## Overall Statistics
##
##           Accuracy : 0.8557
##           95% CI : (0.8478, 0.8634)
##           No Information Rate : 0.2947
##           P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.8174
##           McNemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##           Class: A Class: B Class: C Class: D Class: E
## Sensitivity          0.9165    0.8372    0.8064    0.8177    0.8630
## Specificity          0.9796    0.9494    0.9722    0.9699    0.9510
## Pos Pred Value       0.9494    0.7859    0.8706    0.8476    0.7774
## Neg Pred Value       0.9656    0.9633    0.9559    0.9630    0.9722
## Prevalence           0.2947    0.1816    0.1882    0.1699    0.1656
## Detection Rate       0.2701    0.1521    0.1518    0.1389    0.1429
## Detection Prevalence 0.2845    0.1935    0.1744    0.1639    0.1838
## Balanced Accuracy     0.9481    0.8933    0.8893    0.8938    0.9070
```

```
confusionMatrix(testing$classe,predictRpart)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A     B     C     D     E
##           A 1800    20    66   334    12
##           B  689   198    94   300   237
##           C  600    17   354   381    16
##           D  316    68    23   857    22
##           E  447   169    98   406   322
##
## Overall Statistics
##
##           Accuracy : 0.45
##           95% CI : (0.439, 0.4611)
##           No Information Rate : 0.491
##           P-Value [Acc > NIR] : 1
##
##           Kappa : 0.2883
##           McNemar's Test P-Value : <2e-16
##
## Statistics by Class:
##
##           Class: A Class: B Class: C Class: D Class: E
## Sensitivity          0.4673  0.41949  0.55748  0.3762  0.52874
## Specificity          0.8918  0.82099  0.85938  0.9230  0.84524
## Pos Pred Value       0.8065  0.13043  0.25877  0.6664  0.22330
## Neg Pred Value       0.6345  0.95670  0.95662  0.7834  0.95518
## Prevalence           0.4910  0.06016  0.08093  0.2903  0.07762
## Detection Rate       0.2294  0.02524  0.04512  0.1092  0.04104
## Detection Prevalence 0.2845  0.19347  0.17436  0.1639  0.18379
## Balanced Accuracy     0.6796  0.62024  0.70843  0.6496  0.68699
```

```
accuracyRF<-postResample(predictRF,testing$classe)
accuracyLDA<-postResample(predictLDA,testing$classe)
accuracyGBM<-postResample(predictGBM,testing$classe)
accuracyRpart<-postResample(predictRpart,testing$classe)

accuracyRF
```

```
## Accuracy      Kappa
## 0.9737446 0.9667688
```

```
accuracyLDA
```

```
## Accuracy      Kappa
## 0.4965588 0.3624725
```

```
accuracyGBM
```

```
## Accuracy      Kappa
## 0.8557227 0.8173557
```

```
accuracyRpart
```

```
## Accuracy      Kappa
## 0.4500382 0.2882895
```

It shows that the highest accuracy of random forest model is 97.50%.

6. Predict Predict by using testing data

```
model <- train(classe ~ ., data=training, method="rf", trControl=control, ntree
=250)
result <- predict(model, training[, -length(names(training))])
result
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

