

MLearnProj

Haarkin898

1/18/2021

Project Introduction

This project looks at how fitness data about personal activity collected from electronic devices can be analyzed using Machine learning techniques. Datasets were obtained from a group of enthusiasts who take regular records of their activities for varying reasons. While most people regularly measure how much of a specific activity they do, how well these activities are carried are barely measured. The aim of this project is to distinguish between correctly done exercises and their incorrect versions using data from <http://groupware.les.inf.puc-rio.br/har> (see the section on the Weight Lifting Exercise Dataset). There are six participants and five activities to be tested. Original dataset is divided into train and test data. The aim is to create a model that predicts the manner in which the exercise or activities are carried out using the “classe” variable in the train set.

Download data

the code below downloads the data and assigns them into R objects which are then explored. Both data have 160 variables however the train data has 19622 observations while the test data is limited to 20 observations.

```
trainUrl <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"
testUrl <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"

trainData <- "./MLProj/trainData.csv"
testData <- "./MLProj/testData.csv"

if (!file.exists("./MLProj")) {
  dir.create("./MLProj")
}
if (!file.exists(trainData)) {
  download.file(trainUrl, destfile=trainData, method="curl")
}
if (!file.exists(testData)) {
  download.file(testUrl, destfile=testData, method="curl")
}
##download.file(trainUrl, destfile = "./trainData.csv", method = "curl")
##download.file(testUrl, destfile = "./testData.csv", method = "curl")

## load raw data

trainData <- read.csv("./MLProj/trainData.csv", na.strings = c("NA", "#DIV/0!", "")) ## identify NAs with
testData <- read.csv("./MLProj/testData.csv", na.strings = c("NA", "#DIV/0!", "")) ## identify NAs with
```

```
## explore loaded data
head(trainData,3)
```

```
## 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
## 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
## kurtosis_roll_belt kurtosis_pitch_belt kurtosis_yaw_belt skewness_roll_belt
## 1 NA NA NA NA
## 2 NA NA NA NA
## 3 NA NA NA NA
## skewness_roll_belt.1 skewness_yaw_belt max_roll_belt max_pitch_belt
## 1 NA NA NA NA
## 2 NA NA NA NA
## 3 NA NA NA NA
## max_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt amplitude_roll_belt
## 1 NA NA NA NA NA
## 2 NA NA NA NA NA
## 3 NA NA NA NA NA
## amplitude_pitch_belt amplitude_yaw_belt var_total_accel_belt avg_roll_belt
## 1 NA NA NA NA
## 2 NA NA NA NA
## 3 NA NA NA NA
## stddev_roll_belt var_roll_belt avg_pitch_belt stddev_pitch_belt
## 1 NA NA NA NA
## 2 NA NA NA NA
## 3 NA NA NA NA
## var_pitch_belt avg_yaw_belt stddev_yaw_belt var_yaw_belt gyros_belt_x
## 1 NA NA NA NA 0.00
## 2 NA NA NA NA 0.02
## 3 NA NA NA NA 0.00
## gyros_belt_y gyros_belt_z accel_belt_x accel_belt_y accel_belt_z
## 1 0 -0.02 -21 4 22
## 2 0 -0.02 -22 4 22
## 3 0 -0.02 -20 5 23
## magnet_belt_x magnet_belt_y magnet_belt_z roll_arm pitch_arm yaw_arm
## 1 -3 599 -313 -128 22.5 -161
## 2 -7 608 -311 -128 22.5 -161
## 3 -2 600 -305 -128 22.5 -161
## total_accel_arm var_accel_arm avg_roll_arm stddev_roll_arm var_roll_arm
## 1 34 NA NA NA NA
## 2 34 NA NA NA NA
## 3 34 NA NA NA NA
## avg_pitch_arm stddev_pitch_arm var_pitch_arm avg_yaw_arm stddev_yaw_arm
## 1 NA NA NA NA NA
## 2 NA NA NA NA NA
## 3 NA NA NA NA NA
## var_yaw_arm gyros_arm_x gyros_arm_y gyros_arm_z accel_arm_x accel_arm_y
## 1 NA 0.00 0.00 -0.02 -288 109
```

## 2	NA	0.02	-0.02	-0.02	-290	110
## 3	NA	0.02	-0.02	-0.02	-289	110
##	accel_arm_z	magnet_arm_x	magnet_arm_y	magnet_arm_z	kurtosis_roll_arm	
## 1	-123	-368	337	516	NA	
## 2	-125	-369	337	513	NA	
## 3	-126	-368	344	513	NA	
##	kurtosis_picth_arm	kurtosis_yaw_arm	skewness_roll_arm	skewness_pitch_arm		
## 1	NA	NA	NA	NA	NA	
## 2	NA	NA	NA	NA	NA	
## 3	NA	NA	NA	NA	NA	
##	skewness_yaw_arm	max_roll_arm	max_picth_arm	max_yaw_arm	min_roll_arm	
## 1	NA	NA	NA	NA	NA	
## 2	NA	NA	NA	NA	NA	
## 3	NA	NA	NA	NA	NA	
##	min_pitch_arm	min_yaw_arm	amplitude_roll_arm	amplitude_pitch_arm		
## 1	NA	NA	NA	NA	NA	
## 2	NA	NA	NA	NA	NA	
## 3	NA	NA	NA	NA	NA	
##	amplitude_yaw_arm	roll_dumbbell	pitch_dumbbell	yaw_dumbbell		
## 1	NA	13.05217	-70.49400	-84.87394		
## 2	NA	13.13074	-70.63751	-84.71065		
## 3	NA	12.85075	-70.27812	-85.14078		
##	kurtosis_roll_dumbbell	kurtosis_picth_dumbbell	kurtosis_yaw_dumbbell			
## 1	NA	NA	NA	NA		
## 2	NA	NA	NA	NA		
## 3	NA	NA	NA	NA		
##	skewness_roll_dumbbell	skewness_pitch_dumbbell	skewness_yaw_dumbbell			
## 1	NA	NA	NA	NA		
## 2	NA	NA	NA	NA		
## 3	NA	NA	NA	NA		
##	max_roll_dumbbell	max_picth_dumbbell	max_yaw_dumbbell	min_roll_dumbbell		
## 1	NA	NA	NA	NA	NA	
## 2	NA	NA	NA	NA	NA	
## 3	NA	NA	NA	NA	NA	
##	min_pitch_dumbbell	min_yaw_dumbbell	amplitude_roll_dumbbell			
## 1	NA	NA	NA	NA		
## 2	NA	NA	NA	NA		
## 3	NA	NA	NA	NA		
##	amplitude_pitch_dumbbell	amplitude_yaw_dumbbell	total_accel_dumbbell			
## 1	NA	NA	37			
## 2	NA	NA	37			
## 3	NA	NA	37			
##	var_accel_dumbbell	avg_roll_dumbbell	stddev_roll_dumbbell	var_roll_dumbbell		
## 1	NA	NA	NA	NA	NA	
## 2	NA	NA	NA	NA	NA	
## 3	NA	NA	NA	NA	NA	
##	avg_pitch_dumbbell	stddev_pitch_dumbbell	var_pitch_dumbbell	avg_yaw_dumbbell		
## 1	NA	NA	NA	NA	NA	
## 2	NA	NA	NA	NA	NA	
## 3	NA	NA	NA	NA	NA	
##	stddev_yaw_dumbbell	var_yaw_dumbbell	gyros_dumbbell_x	gyros_dumbbell_y		
## 1	NA	NA	0	-0.02		
## 2	NA	NA	0	-0.02		
## 3	NA	NA	0	-0.02		

```

## gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y accel_dumbbell_z
## 1 0 -234 47 -271
## 2 0 -233 47 -269
## 3 0 -232 46 -270
## magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z roll_forearm
## 1 -559 293 -65 28.4
## 2 -555 296 -64 28.3
## 3 -561 298 -63 28.3
## pitch_forearm yaw_forearm kurtosis_roll_forearm kurtosis_pitch_forearm
## 1 -63.9 -153 NA NA
## 2 -63.9 -153 NA NA
## 3 -63.9 -152 NA NA
## kurtosis_yaw_forearm skewness_roll_forearm skewness_pitch_forearm
## 1 NA NA NA
## 2 NA NA NA
## 3 NA NA NA
## skewness_yaw_forearm max_roll_forearm max_pitch_forearm max_yaw_forearm
## 1 NA NA NA NA
## 2 NA NA NA NA
## 3 NA NA NA NA
## min_roll_forearm min_pitch_forearm min_yaw_forearm amplitude_roll_forearm
## 1 NA NA NA NA
## 2 NA NA NA NA
## 3 NA NA NA NA
## amplitude_pitch_forearm amplitude_yaw_forearm total_accel_forearm
## 1 NA NA 36
## 2 NA NA 36
## 3 NA NA 36
## var_accel_forearm avg_roll_forearm stddev_roll_forearm var_roll_forearm
## 1 NA NA NA NA
## 2 NA NA NA NA
## 3 NA NA NA NA
## avg_pitch_forearm stddev_pitch_forearm var_pitch_forearm avg_yaw_forearm
## 1 NA NA NA NA
## 2 NA NA NA NA
## 3 NA NA NA NA
## 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
## 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
## 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

```

```
head(testData,3)
```

```

## X user_name raw_timestamp_part_1 raw_timestamp_part_2 cvtd_timestamp
## 1 1 pedro 1323095002 868349 05/12/2011 14:23
## 2 2 jeremy 1322673067 778725 30/11/2011 17:11

```

```

## 3 3      jeremy          1322673075          342967 30/11/2011 17:11
##      new_window num_window roll_belt pitch_belt yaw_belt total_accel_belt
## 1          no          74      123.00      27.00      -4.75          20
## 2          no         431        1.02        4.87     -88.90          4
## 3          no         439        0.87        1.82     -88.50          5
##      kurtosis_roll_belt kurtosis_pitch_belt kurtosis_yaw_belt skewness_roll_belt
## 1              NA              NA              NA              NA
## 2              NA              NA              NA              NA
## 3              NA              NA              NA              NA
##      skewness_roll_belt.1 skewness_yaw_belt max_roll_belt max_pitch_belt
## 1              NA              NA              NA              NA
## 2              NA              NA              NA              NA
## 3              NA              NA              NA              NA
##      max_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt amplitude_roll_belt
## 1              NA              NA              NA              NA
## 2              NA              NA              NA              NA
## 3              NA              NA              NA              NA
##      amplitude_pitch_belt amplitude_yaw_belt var_total_accel_belt avg_roll_belt
## 1              NA              NA              NA              NA
## 2              NA              NA              NA              NA
## 3              NA              NA              NA              NA
##      stddev_roll_belt var_roll_belt avg_pitch_belt stddev_pitch_belt
## 1              NA              NA              NA              NA
## 2              NA              NA              NA              NA
## 3              NA              NA              NA              NA
##      var_pitch_belt avg_yaw_belt stddev_yaw_belt var_yaw_belt gyros_belt_x
## 1              NA              NA              NA              NA      -0.50
## 2              NA              NA              NA              NA      -0.06
## 3              NA              NA              NA              NA       0.05
##      gyros_belt_y gyros_belt_z accel_belt_x accel_belt_y accel_belt_z
## 1      -0.02      -0.46        -38          69        -179
## 2      -0.02      -0.07        -13          11          39
## 3       0.02       0.03          1          -1          49
##      magnet_belt_x magnet_belt_y magnet_belt_z roll_arm pitch_arm yaw_arm
## 1        -13         581        -382      40.7      -27.8      178
## 2         43         636        -309       0.0       0.0       0
## 3         29         631        -312       0.0       0.0       0
##      total_accel_arm var_accel_arm avg_roll_arm stddev_roll_arm var_roll_arm
## 1          10          NA          NA          NA          NA
## 2          38          NA          NA          NA          NA
## 3          44          NA          NA          NA          NA
##      avg_pitch_arm stddev_pitch_arm var_pitch_arm avg_yaw_arm stddev_yaw_arm
## 1          NA          NA          NA          NA          NA
## 2          NA          NA          NA          NA          NA
## 3          NA          NA          NA          NA          NA
##      var_yaw_arm gyros_arm_x gyros_arm_y gyros_arm_z accel_arm_x accel_arm_y
## 1          NA      -1.65        0.48      -0.18         16         38
## 2          NA      -1.17        0.85      -0.43      -290        215
## 3          NA       2.10       -1.36       1.13     -341        245
##      accel_arm_z magnet_arm_x magnet_arm_y magnet_arm_z kurtosis_roll_arm
## 1          93        -326        385         481          NA
## 2         -90        -325        447         434          NA
## 3         -87        -264        474         413          NA
##      kurtosis_pitch_arm kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm

```

##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##	3	NA	NA	NA	NA
##		skewness_yaw_arm	max_roll_arm	max_pitch_arm	max_yaw_arm
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##	3	NA	NA	NA	NA
##		min_pitch_arm	min_yaw_arm	amplitude_roll_arm	amplitude_pitch_arm
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##	3	NA	NA	NA	NA
##		amplitude_yaw_arm	roll_dumbbell	pitch_dumbbell	yaw_dumbbell
##	1	NA	-17.73748	24.96085	126.23596
##	2	NA	54.47761	-53.69758	-75.51480
##	3	NA	57.07031	-51.37303	-75.20287
##		kurtosis_roll_dumbbell	kurtosis_pitch_dumbbell	kurtosis_yaw_dumbbell	
##	1	NA	NA	NA	
##	2	NA	NA	NA	
##	3	NA	NA	NA	
##		skewness_roll_dumbbell	skewness_pitch_dumbbell	skewness_yaw_dumbbell	
##	1	NA	NA	NA	
##	2	NA	NA	NA	
##	3	NA	NA	NA	
##		max_roll_dumbbell	max_pitch_dumbbell	max_yaw_dumbbell	min_roll_dumbbell
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##	3	NA	NA	NA	NA
##		min_pitch_dumbbell	min_yaw_dumbbell	amplitude_roll_dumbbell	
##	1	NA	NA	NA	
##	2	NA	NA	NA	
##	3	NA	NA	NA	
##		amplitude_pitch_dumbbell	amplitude_yaw_dumbbell	total_accel_dumbbell	
##	1	NA	NA	9	
##	2	NA	NA	31	
##	3	NA	NA	29	
##		var_accel_dumbbell	avg_roll_dumbbell	stddev_roll_dumbbell	var_roll_dumbbell
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##	3	NA	NA	NA	NA
##		avg_pitch_dumbbell	stddev_pitch_dumbbell	var_pitch_dumbbell	avg_yaw_dumbbell
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##	3	NA	NA	NA	NA
##		stddev_yaw_dumbbell	var_yaw_dumbbell	gyros_dumbbell_x	gyros_dumbbell_y
##	1	NA	NA	0.64	0.06
##	2	NA	NA	0.34	0.05
##	3	NA	NA	0.39	0.14
##		gyros_dumbbell_z	accel_dumbbell_x	accel_dumbbell_y	accel_dumbbell_z
##	1	-0.61	21	-15	81
##	2	-0.71	-153	155	-205
##	3	-0.34	-141	155	-196
##		magnet_dumbbell_x	magnet_dumbbell_y	magnet_dumbbell_z	roll_forearm
##	1	523	-528	-56	141
##	2	-502	388	-36	109

```

## 3          -506          349          41          131
## pitch_forearm yaw_forearm kurtosis_roll_forearm kurtosis_pitch_forearm
## 1          49.3          156          NA          NA
## 2          -17.6          106          NA          NA
## 3          -32.6          93          NA          NA
## kurtosis_yaw_forearm skewness_roll_forearm skewness_pitch_forearm
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## skewness_yaw_forearm max_roll_forearm max_pitch_forearm max_yaw_forearm
## 1          NA          NA          NA          NA
## 2          NA          NA          NA          NA
## 3          NA          NA          NA          NA
## min_roll_forearm min_pitch_forearm min_yaw_forearm amplitude_roll_forearm
## 1          NA          NA          NA          NA
## 2          NA          NA          NA          NA
## 3          NA          NA          NA          NA
## amplitude_pitch_forearm amplitude_yaw_forearm total_accel_forearm
## 1          NA          NA          33
## 2          NA          NA          39
## 3          NA          NA          34
## var_accel_forearm avg_roll_forearm stddev_roll_forearm var_roll_forearm
## 1          NA          NA          NA          NA
## 2          NA          NA          NA          NA
## 3          NA          NA          NA          NA
## avg_pitch_forearm stddev_pitch_forearm var_pitch_forearm avg_yaw_forearm
## 1          NA          NA          NA          NA
## 2          NA          NA          NA          NA
## 3          NA          NA          NA          NA
## stddev_yaw_forearm var_yaw_forearm gyros_forearm_x gyros_forearm_y
## 1          NA          NA          0.74          -3.34
## 2          NA          NA          1.12          -2.78
## 3          NA          NA          0.18          -0.79
## gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z
## 1          -0.59          -110          267          -149
## 2          -0.18          212          297          -118
## 3          0.28          154          271          -129
## magnet_forearm_x magnet_forearm_y magnet_forearm_z problem_id
## 1          -714          419          617          1
## 2          -237          791          873          2
## 3          -51          698          783          3

```

Data Cleaning

Irrelevant variables are removed and columns with NAs as well. Also column with near zero variations were checked for however there are none in the dataset. The refined datasets have 53 variables.

```

##remove irrelevant columns
trainDataRef <- trainData[, 8:length(trainData)]
testdataRef <- testData[, 8:length(testData)]

##remove columns with NAs
trainDataRef <- trainDataRef[colSums(is.na(trainDataRef)) == 0]

```

```
testdataRef <- testdataRef[colSums(is.na(testdataRef)) == 0]

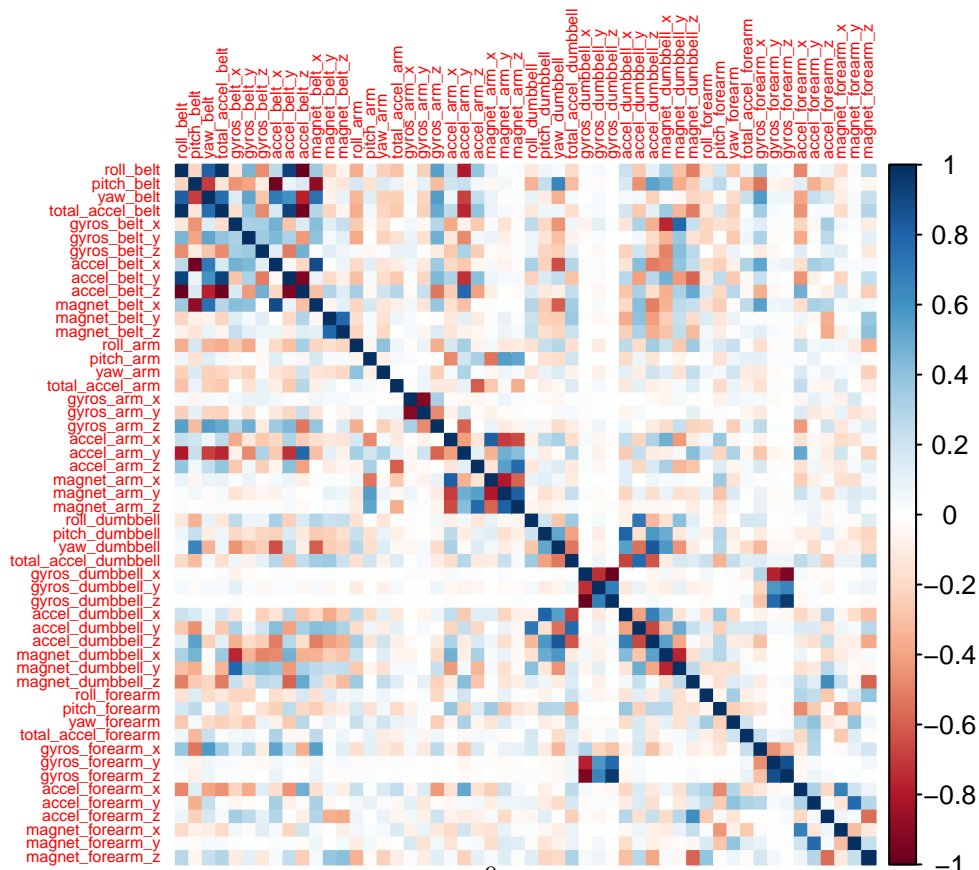
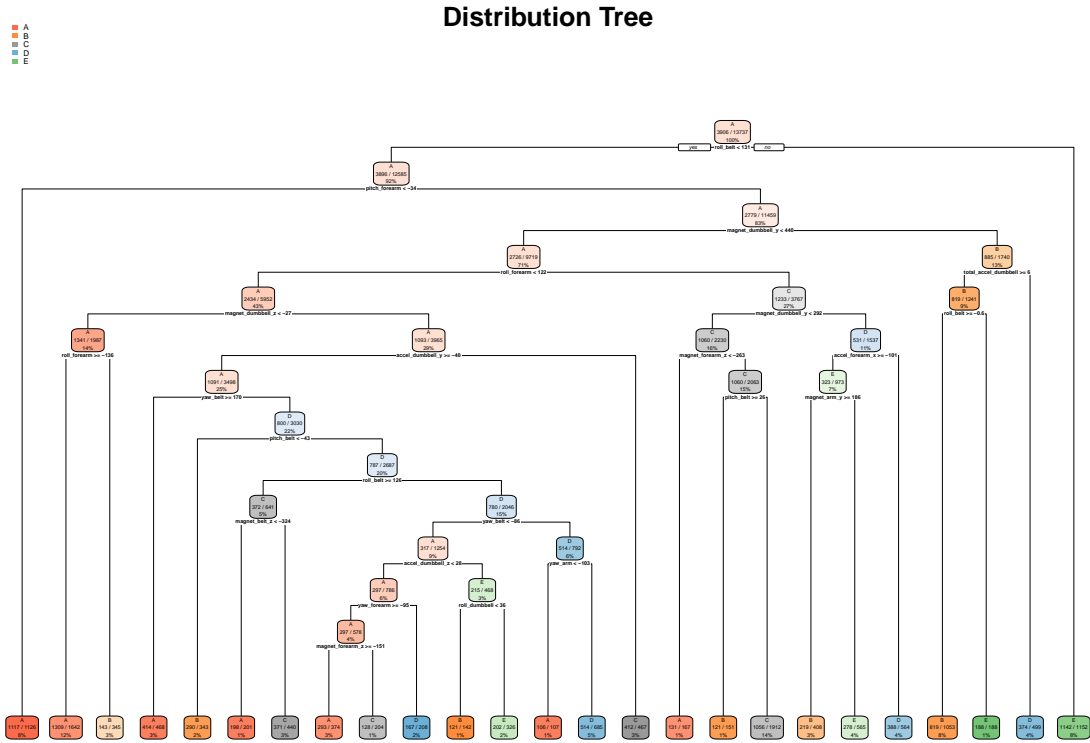
## identify records with near zero variance and remove them
NZV <- nearZeroVar(trainDataRef, saveMetrics = TRUE)
trainDataRef <- trainDataRef[,NZV$nzv==FALSE]
```

(Train)Data Partition

the train data set is partitioned into a train set and a validation set

```
## Train data partition necessary for cross validation
set.seed(1234)
in.training <- createDataPartition(trainDataRef$classe, p = 0.70, list = F)
trainData1 <- trainDataRef[in.training,]
validationData <- trainDataRef[-in.training,]
```


Initialized Exploration Plots



Modeling of the Train Data

A predictive model using the random forest algorithm for activity recognition is used to model the data. A 5 fold cross validation scheme is utilized in the process.

```
##Model generation
rfModel <- train(classe ~ ., data = trainData1, method = "rf", trControl = trainControl(method = "cv",
rfModel
```

```
## Random Forest
##
## 13737 samples
##    52 predictor
##    5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 10989, 10990, 10989, 10990, 10990
## Resampling results across tuning parameters:
##
##  mtry  Accuracy  Kappa
##    2    0.9914827 0.9892247
##   27    0.9908275 0.9883956
##   52    0.9843484 0.9801989
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 2.
```

```
##test model with validation dataset
rfPredict <- predict(rfModel, validationData)
confusionMatrix(table(validationData$classe, rfPredict))
```

```
## Confusion Matrix and Statistics
##
##      rfPredict
##      A      B      C      D      E
## A 1674      0      0      0      0
## B      2 1134      3      0      0
## C      0     12 1014      0      0
## D      0      0     15  948      1
## E      0      0      0      0 1082
##
## Overall Statistics
##
##              Accuracy : 0.9944
##              95% CI : (0.9921, 0.9961)
##      No Information Rate : 0.2848
##      P-Value [Acc > NIR] : < 2.2e-16
##
##              Kappa : 0.9929
##
## Mcnemar's Test P-Value : NA
##
```

```
## Statistics by Class:
##
##           Class: A Class: B Class: C Class: D Class: E
## Sensitivity      0.9988  0.9895  0.9826  1.0000  0.9991
## Specificity      1.0000  0.9989  0.9975  0.9968  1.0000
## Pos Pred Value   1.0000  0.9956  0.9883  0.9834  1.0000
## Neg Pred Value   0.9995  0.9975  0.9963  1.0000  0.9998
## Prevalence       0.2848  0.1947  0.1754  0.1611  0.1840
## Detection Rate   0.2845  0.1927  0.1723  0.1611  0.1839
## Detection Prevalence 0.2845  0.1935  0.1743  0.1638  0.1839
## Balanced Accuracy 0.9994  0.9942  0.9900  0.9984  0.9995

modelAccuracy <- as.numeric(confusionMatrix(table(validationData$classe, rfPredict))$overall[1])
modelAccuracy

## [1] 0.9943925

OSE <- 1 - modelAccuracy
OSE

## [1] 0.005607477

##OSE1 <- 1 - as.numeric(confusionMatrix(table(validationData$classe, rfPredict))$overall[1])
```

The estimated accuracy of the model is 0.9943925 and the estimated out of sample error is 0.0056075.

Test Data Prediction

The model is now applied to the test data however the last variable in the dataset is excluded from the analysis.

```
##Results
testResults <- predict(rfModel, testdataRef[, -length(testdataRef)])
testResults

## [1] B A B A A E D B A A B C B A E E A B B B
## Levels: A B C D E
```