**Practical-Machine-Learning-project.R**

***Mahesha***

***Tue Sep 19 19:39:00 2017***

install.packages("caret")

## Installing package into 'C:/Users/Mahesha/Documents/R/win-library/3.4'

## (as 'lib' is unspecified)

## package 'caret' successfully unpacked and MD5 sums checked

##

## The downloaded binary packages are in

## C:\Users\Mahesha\AppData\Local\Temp\Rtmp8Y1oGj\downloaded\_packages

library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

## Import files

setwd("C:\\Users\\Mahesha\\Desktop\\Desktop\\Data\_Science\\Courseera\\Practical Machine learning\\Project")

trainds <- read.csv("pml-training.csv",na.strings=c("NA","#DIV/0!",""))

testds <- read.csv("pml-testing.csv",na.strings=c("NA","#DIV/0!",""))

dim(trainds)

## [1] 19622 160

dim(testds)

## [1] 20 160

## Data explore

str(trainds)

## '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 1323084232 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\_picth\_belt : num NA NA NA NA NA NA NA NA NA NA ...

## $ kurtosis\_yaw\_belt : logi 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 ...

## $ 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 : 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 0.03 ...

## $ 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 0 ...

## $ 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 -0.03 ...

## $ 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 : num NA NA NA NA NA NA NA NA NA NA ...

## $ kurtosis\_picth\_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 ...

## $ skewness\_pitch\_arm : num NA NA NA NA NA NA NA NA NA NA ...

## $ skewness\_yaw\_arm : num NA NA NA NA NA NA 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 : num NA NA NA NA NA NA NA NA NA NA ...

## $ kurtosis\_picth\_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...

## $ kurtosis\_yaw\_dumbbell : logi NA NA NA NA NA NA ...

## $ skewness\_roll\_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...

## $ skewness\_pitch\_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...

## $ skewness\_yaw\_dumbbell : logi 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 : num NA NA NA 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 : num NA NA NA NA NA NA NA NA NA NA ...

## $ amplitude\_roll\_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...

## [list output truncated]

summary(trainds)

## X user\_name raw\_timestamp\_part\_1 raw\_timestamp\_part\_2

## Min. : 1 adelmo :3892 Min. :1.322e+09 Min. : 294

## 1st Qu.: 4906 carlitos:3112 1st Qu.:1.323e+09 1st Qu.:252912

## Median : 9812 charles :3536 Median :1.323e+09 Median :496380

## Mean : 9812 eurico :3070 Mean :1.323e+09 Mean :500656

## 3rd Qu.:14717 jeremy :3402 3rd Qu.:1.323e+09 3rd Qu.:751891

## Max. :19622 pedro :2610 Max. :1.323e+09 Max. :998801

## cvtd\_timestamp new\_window num\_window roll\_belt

## 28/11/2011 14:14: 1498 no :19216 Min. : 1.0 Min. :-28.90

## 05/12/2011 11:24: 1497 yes: 406 1st Qu.:222.0 1st Qu.: 1.10

## 30/11/2011 17:11: 1440 Median :424.0 Median :113.00

## 05/12/2011 11:25: 1425 Mean :430.6 Mean : 64.41

## 02/12/2011 14:57: 1380 3rd Qu.:644.0 3rd Qu.:123.00

## 02/12/2011 13:34: 1375 Max. :864.0 Max. :162.00

## pitch\_belt yaw\_belt total\_accel\_belt kurtosis\_roll\_belt

## Min. :-55.8000 Min. :-180.00 Min. : 0.00 Min. :-2.121

## 1st Qu.: 1.7600 1st Qu.: -88.30 1st Qu.: 3.00 1st Qu.:-1.329

## Median : 5.2800 Median : -13.00 Median :17.00 Median :-0.899

## Mean : 0.3053 Mean : -11.21 Mean :11.31 Mean :-0.220

## 3rd Qu.: 14.9000 3rd Qu.: 12.90 3rd Qu.:18.00 3rd Qu.:-0.219

## Max. : 60.3000 Max. : 179.00 Max. :29.00 Max. :33.000

## kurtosis\_picth\_belt kurtosis\_yaw\_belt skewness\_roll\_belt

## Min. :-2.190 Mode:logical Min. :-5.745

## 1st Qu.:-1.107 NA's:19622 1st Qu.:-0.444

## Median :-0.151 Median : 0.000

## Mean : 4.334 Mean :-0.026

## 3rd Qu.: 3.178 3rd Qu.: 0.417

## Max. :58.000 Max. : 3.595

## skewness\_roll\_belt.1 skewness\_yaw\_belt max\_roll\_belt max\_picth\_belt

## Min. :-7.616 Mode:logical Min. :-94.300 Min. : 3.00

## 1st Qu.:-1.114 NA's:19622 1st Qu.:-88.000 1st Qu.: 5.00

## Median :-0.068 Median : -5.100 Median :18.00

## Mean :-0.296 Mean : -6.667 Mean :12.92

## 3rd Qu.: 0.661 3rd Qu.: 18.500 3rd Qu.:19.00

## Max. : 7.348 Max. :180.000 Max. :30.00

## max\_yaw\_belt min\_roll\_belt min\_pitch\_belt min\_yaw\_belt

## Min. :-2.10 Min. :-180.00 Min. : 0.00 Min. :-2.10

## 1st Qu.:-1.30 1st Qu.: -88.40 1st Qu.: 3.00 1st Qu.:-1.30

## Median :-0.90 Median : -7.85 Median :16.00 Median :-0.90

## Mean :-0.22 Mean : -10.44 Mean :10.76 Mean :-0.22

## 3rd Qu.:-0.20 3rd Qu.: 9.05 3rd Qu.:17.00 3rd Qu.:-0.20

## Max. :33.00 Max. : 173.00 Max. :23.00 Max. :33.00

## amplitude\_roll\_belt amplitude\_pitch\_belt amplitude\_yaw\_belt

## Min. : 0.000 Min. : 0.000 Min. :0

## 1st Qu.: 0.300 1st Qu.: 1.000 1st Qu.:0

## Median : 1.000 Median : 1.000 Median :0

## Mean : 3.769 Mean : 2.167 Mean :0

## 3rd Qu.: 2.083 3rd Qu.: 2.000 3rd Qu.:0

## Max. :360.000 Max. :12.000 Max. :0

## var\_total\_accel\_belt avg\_roll\_belt stddev\_roll\_belt var\_roll\_belt

## Min. : 0.000 Min. :-27.40 Min. : 0.000 Min. : 0.000

## 1st Qu.: 0.100 1st Qu.: 1.10 1st Qu.: 0.200 1st Qu.: 0.000

## Median : 0.200 Median :116.35 Median : 0.400 Median : 0.100

## Mean : 0.926 Mean : 68.06 Mean : 1.337 Mean : 7.699

## 3rd Qu.: 0.300 3rd Qu.:123.38 3rd Qu.: 0.700 3rd Qu.: 0.500

## Max. :16.500 Max. :157.40 Max. :14.200 Max. :200.700

## avg\_pitch\_belt stddev\_pitch\_belt var\_pitch\_belt avg\_yaw\_belt

## Min. :-51.400 Min. :0.000 Min. : 0.000 Min. :-138.300

## 1st Qu.: 2.025 1st Qu.:0.200 1st Qu.: 0.000 1st Qu.: -88.175

## Median : 5.200 Median :0.400 Median : 0.100 Median : -6.550

## Mean : 0.520 Mean :0.603 Mean : 0.766 Mean : -8.831

## 3rd Qu.: 15.775 3rd Qu.:0.700 3rd Qu.: 0.500 3rd Qu.: 14.125

## Max. : 59.700 Max. :4.000 Max. :16.200 Max. : 173.500

## stddev\_yaw\_belt var\_yaw\_belt gyros\_belt\_x

## Min. : 0.000 Min. : 0.000 Min. :-1.040000

## 1st Qu.: 0.100 1st Qu.: 0.010 1st Qu.:-0.030000

## Median : 0.300 Median : 0.090 Median : 0.030000

## Mean : 1.341 Mean : 107.487 Mean :-0.005592

## 3rd Qu.: 0.700 3rd Qu.: 0.475 3rd Qu.: 0.110000

## Max. :176.600 Max. :31183.240 Max. : 2.220000

## gyros\_belt\_y gyros\_belt\_z accel\_belt\_x accel\_belt\_y

## Min. :-0.64000 Min. :-1.4600 Min. :-120.000 Min. :-69.00

## 1st Qu.: 0.00000 1st Qu.:-0.2000 1st Qu.: -21.000 1st Qu.: 3.00

## Median : 0.02000 Median :-0.1000 Median : -15.000 Median : 35.00

## Mean : 0.03959 Mean :-0.1305 Mean : -5.595 Mean : 30.15

## 3rd Qu.: 0.11000 3rd Qu.:-0.0200 3rd Qu.: -5.000 3rd Qu.: 61.00

## Max. : 0.64000 Max. : 1.6200 Max. : 85.000 Max. :164.00

## accel\_belt\_z magnet\_belt\_x magnet\_belt\_y magnet\_belt\_z

## Min. :-275.00 Min. :-52.0 Min. :354.0 Min. :-623.0

## 1st Qu.:-162.00 1st Qu.: 9.0 1st Qu.:581.0 1st Qu.:-375.0

## Median :-152.00 Median : 35.0 Median :601.0 Median :-320.0

## Mean : -72.59 Mean : 55.6 Mean :593.7 Mean :-345.5

## 3rd Qu.: 27.00 3rd Qu.: 59.0 3rd Qu.:610.0 3rd Qu.:-306.0

## Max. : 105.00 Max. :485.0 Max. :673.0 Max. : 293.0

## roll\_arm pitch\_arm yaw\_arm total\_accel\_arm

## Min. :-180.00 Min. :-88.800 Min. :-180.0000 Min. : 1.00

## 1st Qu.: -31.77 1st Qu.:-25.900 1st Qu.: -43.1000 1st Qu.:17.00

## Median : 0.00 Median : 0.000 Median : 0.0000 Median :27.00

## Mean : 17.83 Mean : -4.612 Mean : -0.6188 Mean :25.51

## 3rd Qu.: 77.30 3rd Qu.: 11.200 3rd Qu.: 45.8750 3rd Qu.:33.00

## Max. : 180.00 Max. : 88.500 Max. : 180.0000 Max. :66.00

## var\_accel\_arm avg\_roll\_arm stddev\_roll\_arm var\_roll\_arm

## Min. : 0.00 Min. :-166.67 Min. : 0.000 Min. : 0.000

## 1st Qu.: 9.03 1st Qu.: -38.37 1st Qu.: 1.376 1st Qu.: 1.898

## Median : 40.61 Median : 0.00 Median : 5.702 Median : 32.517

## Mean : 53.23 Mean : 12.68 Mean : 11.201 Mean : 417.264

## 3rd Qu.: 75.62 3rd Qu.: 76.33 3rd Qu.: 14.921 3rd Qu.: 222.647

## Max. :331.70 Max. : 163.33 Max. :161.964 Max. :26232.208

## avg\_pitch\_arm stddev\_pitch\_arm var\_pitch\_arm avg\_yaw\_arm

## Min. :-81.773 Min. : 0.000 Min. : 0.000 Min. :-173.440

## 1st Qu.:-22.770 1st Qu.: 1.642 1st Qu.: 2.697 1st Qu.: -29.198

## Median : 0.000 Median : 8.133 Median : 66.146 Median : 0.000

## Mean : -4.901 Mean :10.383 Mean : 195.864 Mean : 2.359

## 3rd Qu.: 8.277 3rd Qu.:16.327 3rd Qu.: 266.576 3rd Qu.: 38.185

## Max. : 75.659 Max. :43.412 Max. :1884.565 Max. : 152.000

## stddev\_yaw\_arm var\_yaw\_arm gyros\_arm\_x

## Min. : 0.000 Min. : 0.000 Min. :-6.37000

## 1st Qu.: 2.577 1st Qu.: 6.642 1st Qu.:-1.33000

## Median : 16.682 Median : 278.309 Median : 0.08000

## Mean : 22.270 Mean : 1055.933 Mean : 0.04277

## 3rd Qu.: 35.984 3rd Qu.: 1294.850 3rd Qu.: 1.57000

## Max. :177.044 Max. :31344.568 Max. : 4.87000

## gyros\_arm\_y gyros\_arm\_z accel\_arm\_x accel\_arm\_y

## Min. :-3.4400 Min. :-2.3300 Min. :-404.00 Min. :-318.0

## 1st Qu.:-0.8000 1st Qu.:-0.0700 1st Qu.:-242.00 1st Qu.: -54.0

## Median :-0.2400 Median : 0.2300 Median : -44.00 Median : 14.0

## Mean :-0.2571 Mean : 0.2695 Mean : -60.24 Mean : 32.6

## 3rd Qu.: 0.1400 3rd Qu.: 0.7200 3rd Qu.: 84.00 3rd Qu.: 139.0

## Max. : 2.8400 Max. : 3.0200 Max. : 437.00 Max. : 308.0

## accel\_arm\_z magnet\_arm\_x magnet\_arm\_y magnet\_arm\_z

## Min. :-636.00 Min. :-584.0 Min. :-392.0 Min. :-597.0

## 1st Qu.:-143.00 1st Qu.:-300.0 1st Qu.: -9.0 1st Qu.: 131.2

## Median : -47.00 Median : 289.0 Median : 202.0 Median : 444.0

## Mean : -71.25 Mean : 191.7 Mean : 156.6 Mean : 306.5

## 3rd Qu.: 23.00 3rd Qu.: 637.0 3rd Qu.: 323.0 3rd Qu.: 545.0

## Max. : 292.00 Max. : 782.0 Max. : 583.0 Max. : 694.0

## kurtosis\_roll\_arm kurtosis\_picth\_arm kurtosis\_yaw\_arm skewness\_roll\_arm

## Min. :-1.809 Min. :-2.084 Min. :-2.103 Min. :-2.541

## 1st Qu.:-1.345 1st Qu.:-1.280 1st Qu.:-1.220 1st Qu.:-0.561

## Median :-0.894 Median :-1.010 Median :-0.733 Median : 0.040

## Mean :-0.366 Mean :-0.542 Mean : 0.406 Mean : 0.068

## 3rd Qu.:-0.038 3rd Qu.:-0.379 3rd Qu.: 0.115 3rd Qu.: 0.671

## Max. :21.456 Max. :19.751 Max. :56.000 Max. : 4.394

## skewness\_pitch\_arm skewness\_yaw\_arm max\_roll\_arm max\_picth\_arm

## Min. :-4.565 Min. :-6.708 Min. :-73.100 Min. :-173.000

## 1st Qu.:-0.618 1st Qu.:-0.743 1st Qu.: -0.175 1st Qu.: -1.975

## Median :-0.035 Median :-0.133 Median : 4.950 Median : 23.250

## Mean :-0.065 Mean :-0.229 Mean : 11.236 Mean : 35.751

## 3rd Qu.: 0.454 3rd Qu.: 0.344 3rd Qu.: 26.775 3rd Qu.: 95.975

## Max. : 3.043 Max. : 7.483 Max. : 85.500 Max. : 180.000

## max\_yaw\_arm min\_roll\_arm min\_pitch\_arm min\_yaw\_arm

## Min. : 4.00 Min. :-89.10 Min. :-180.00 Min. : 1.00

## 1st Qu.:29.00 1st Qu.:-41.98 1st Qu.: -72.62 1st Qu.: 8.00

## Median :34.00 Median :-22.45 Median : -33.85 Median :13.00

## Mean :35.46 Mean :-21.22 Mean : -33.92 Mean :14.66

## 3rd Qu.:41.00 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.:19.00

## Max. :65.00 Max. : 66.40 Max. : 152.00 Max. :38.00

## amplitude\_roll\_arm amplitude\_pitch\_arm amplitude\_yaw\_arm

## Min. : 0.000 Min. : 0.000 Min. : 0.00

## 1st Qu.: 5.425 1st Qu.: 9.925 1st Qu.:13.00

## Median : 28.450 Median : 54.900 Median :22.00

## Mean : 32.452 Mean : 69.677 Mean :20.79

## 3rd Qu.: 50.960 3rd Qu.:115.175 3rd Qu.:28.75

## Max. :119.500 Max. :360.000 Max. :52.00

## roll\_dumbbell pitch\_dumbbell yaw\_dumbbell

## Min. :-153.71 Min. :-149.59 Min. :-150.871

## 1st Qu.: -18.49 1st Qu.: -40.89 1st Qu.: -77.644

## Median : 48.17 Median : -20.96 Median : -3.324

## Mean : 23.84 Mean : -10.78 Mean : 1.674

## 3rd Qu.: 67.61 3rd Qu.: 17.50 3rd Qu.: 79.643

## Max. : 153.55 Max. : 149.40 Max. : 154.952

## kurtosis\_roll\_dumbbell kurtosis\_picth\_dumbbell kurtosis\_yaw\_dumbbell

## Min. :-2.174 Min. :-2.200 Mode:logical

## 1st Qu.:-0.682 1st Qu.:-0.721 NA's:19622

## Median :-0.033 Median :-0.133

## Mean : 0.452 Mean : 0.286

## 3rd Qu.: 0.940 3rd Qu.: 0.584

## Max. :54.998 Max. :55.628

## skewness\_roll\_dumbbell skewness\_pitch\_dumbbell skewness\_yaw\_dumbbell

## Min. :-7.384 Min. :-7.447 Mode:logical

## 1st Qu.:-0.581 1st Qu.:-0.526 NA's:19622

## Median :-0.076 Median :-0.091

## Mean :-0.115 Mean :-0.035

## 3rd Qu.: 0.400 3rd Qu.: 0.505

## Max. : 1.958 Max. : 3.769

## max\_roll\_dumbbell max\_picth\_dumbbell max\_yaw\_dumbbell min\_roll\_dumbbell

## Min. :-70.10 Min. :-112.90 Min. :-2.20 Min. :-149.60

## 1st Qu.:-27.15 1st Qu.: -66.70 1st Qu.:-0.70 1st Qu.: -59.67

## Median : 14.85 Median : 40.05 Median : 0.00 Median : -43.55

## Mean : 13.76 Mean : 32.75 Mean : 0.45 Mean : -41.24

## 3rd Qu.: 50.58 3rd Qu.: 133.22 3rd Qu.: 0.90 3rd Qu.: -25.20

## Max. :137.00 Max. : 155.00 Max. :55.00 Max. : 73.20

## min\_pitch\_dumbbell min\_yaw\_dumbbell amplitude\_roll\_dumbbell

## Min. :-147.00 Min. :-2.20 Min. : 0.00

## 1st Qu.: -91.80 1st Qu.:-0.70 1st Qu.: 14.97

## Median : -66.15 Median : 0.00 Median : 35.05

## Mean : -33.18 Mean : 0.45 Mean : 55.00

## 3rd Qu.: 21.20 3rd Qu.: 0.90 3rd Qu.: 81.04

## Max. : 120.90 Max. :55.00 Max. :256.48

## amplitude\_pitch\_dumbbell amplitude\_yaw\_dumbbell total\_accel\_dumbbell

## Min. : 0.00 Min. :0 Min. : 0.00

## 1st Qu.: 17.06 1st Qu.:0 1st Qu.: 4.00

## Median : 41.73 Median :0 Median :10.00

## Mean : 65.93 Mean :0 Mean :13.72

## 3rd Qu.: 99.55 3rd Qu.:0 3rd Qu.:19.00

## Max. :273.59 Max. :0 Max. :58.00

## var\_accel\_dumbbell avg\_roll\_dumbbell stddev\_roll\_dumbbell

## Min. : 0.000 Min. :-128.96 Min. : 0.000

## 1st Qu.: 0.378 1st Qu.: -12.33 1st Qu.: 4.639

## Median : 1.000 Median : 48.23 Median : 12.204

## Mean : 4.388 Mean : 23.86 Mean : 20.761

## 3rd Qu.: 3.434 3rd Qu.: 64.37 3rd Qu.: 26.356

## Max. :230.428 Max. : 125.99 Max. :123.778

## var\_roll\_dumbbell avg\_pitch\_dumbbell stddev\_pitch\_dumbbell

## Min. : 0.00 Min. :-70.73 Min. : 0.000

## 1st Qu.: 21.52 1st Qu.:-42.00 1st Qu.: 3.482

## Median : 148.95 Median :-19.91 Median : 8.089

## Mean : 1020.27 Mean :-12.33 Mean :13.147

## 3rd Qu.: 694.65 3rd Qu.: 13.21 3rd Qu.:19.238

## Max. :15321.01 Max. : 94.28 Max. :82.680

## var\_pitch\_dumbbell avg\_yaw\_dumbbell stddev\_yaw\_dumbbell

## Min. : 0.00 Min. :-117.950 Min. : 0.000

## 1st Qu.: 12.12 1st Qu.: -76.696 1st Qu.: 3.885

## Median : 65.44 Median : -4.505 Median : 10.264

## Mean : 350.31 Mean : 0.202 Mean : 16.647

## 3rd Qu.: 370.11 3rd Qu.: 71.234 3rd Qu.: 24.674

## Max. :6836.02 Max. : 134.905 Max. :107.088

## var\_yaw\_dumbbell gyros\_dumbbell\_x gyros\_dumbbell\_y

## Min. : 0.00 Min. :-204.0000 Min. :-2.10000

## 1st Qu.: 15.09 1st Qu.: -0.0300 1st Qu.:-0.14000

## Median : 105.35 Median : 0.1300 Median : 0.03000

## Mean : 589.84 Mean : 0.1611 Mean : 0.04606

## 3rd Qu.: 608.79 3rd Qu.: 0.3500 3rd Qu.: 0.21000

## Max. :11467.91 Max. : 2.2200 Max. :52.00000

## gyros\_dumbbell\_z accel\_dumbbell\_x accel\_dumbbell\_y accel\_dumbbell\_z

## Min. : -2.380 Min. :-419.00 Min. :-189.00 Min. :-334.00

## 1st Qu.: -0.310 1st Qu.: -50.00 1st Qu.: -8.00 1st Qu.:-142.00

## Median : -0.130 Median : -8.00 Median : 41.50 Median : -1.00

## Mean : -0.129 Mean : -28.62 Mean : 52.63 Mean : -38.32

## 3rd Qu.: 0.030 3rd Qu.: 11.00 3rd Qu.: 111.00 3rd Qu.: 38.00

## Max. :317.000 Max. : 235.00 Max. : 315.00 Max. : 318.00

## magnet\_dumbbell\_x magnet\_dumbbell\_y magnet\_dumbbell\_z roll\_forearm

## Min. :-643.0 Min. :-3600 Min. :-262.00 Min. :-180.0000

## 1st Qu.:-535.0 1st Qu.: 231 1st Qu.: -45.00 1st Qu.: -0.7375

## Median :-479.0 Median : 311 Median : 13.00 Median : 21.7000

## Mean :-328.5 Mean : 221 Mean : 46.05 Mean : 33.8265

## 3rd Qu.:-304.0 3rd Qu.: 390 3rd Qu.: 95.00 3rd Qu.: 140.0000

## Max. : 592.0 Max. : 633 Max. : 452.00 Max. : 180.0000

## pitch\_forearm yaw\_forearm kurtosis\_roll\_forearm

## Min. :-72.50 Min. :-180.00 Min. :-1.879

## 1st Qu.: 0.00 1st Qu.: -68.60 1st Qu.:-1.398

## Median : 9.24 Median : 0.00 Median :-1.119

## Mean : 10.71 Mean : 19.21 Mean :-0.689

## 3rd Qu.: 28.40 3rd Qu.: 110.00 3rd Qu.:-0.618

## Max. : 89.80 Max. : 180.00 Max. :40.060

## kurtosis\_picth\_forearm kurtosis\_yaw\_forearm skewness\_roll\_forearm

## Min. :-2.098 Mode:logical Min. :-2.297

## 1st Qu.:-1.376 NA's:19622 1st Qu.:-0.402

## Median :-0.890 Median : 0.003

## Mean : 0.419 Mean :-0.009

## 3rd Qu.: 0.054 3rd Qu.: 0.370

## Max. :33.626 Max. : 5.856

## skewness\_pitch\_forearm skewness\_yaw\_forearm max\_roll\_forearm

## Min. :-5.241 Mode:logical Min. :-66.60

## 1st Qu.:-0.881 NA's:19622 1st Qu.: 0.00

## Median :-0.156 Median : 26.80

## Mean :-0.223 Mean : 24.49

## 3rd Qu.: 0.514 3rd Qu.: 45.95

## Max. : 4.464 Max. : 89.80

## max\_picth\_forearm max\_yaw\_forearm min\_roll\_forearm min\_pitch\_forearm

## Min. :-151.00 Min. :-1.900 Min. :-72.500 Min. :-180.00

## 1st Qu.: 0.00 1st Qu.:-1.400 1st Qu.: -6.075 1st Qu.:-175.00

## Median : 113.00 Median :-1.100 Median : 0.000 Median : -61.00

## Mean : 81.49 Mean :-0.689 Mean : -0.167 Mean : -57.57

## 3rd Qu.: 174.75 3rd Qu.:-0.600 3rd Qu.: 12.075 3rd Qu.: 0.00

## Max. : 180.00 Max. :40.100 Max. : 62.100 Max. : 167.00

## min\_yaw\_forearm amplitude\_roll\_forearm amplitude\_pitch\_forearm

## Min. :-1.900 Min. : 0.000 Min. : 0.0

## 1st Qu.:-1.400 1st Qu.: 1.125 1st Qu.: 2.0

## Median :-1.100 Median : 17.770 Median : 83.7

## Mean :-0.689 Mean : 24.653 Mean :139.1

## 3rd Qu.:-0.600 3rd Qu.: 39.875 3rd Qu.:350.0

## Max. :40.100 Max. :126.000 Max. :360.0

## amplitude\_yaw\_forearm total\_accel\_forearm var\_accel\_forearm

## Min. :0 Min. : 0.00 Min. : 0.000

## 1st Qu.:0 1st Qu.: 29.00 1st Qu.: 6.759

## Median :0 Median : 36.00 Median : 21.165

## Mean :0 Mean : 34.72 Mean : 33.502

## 3rd Qu.:0 3rd Qu.: 41.00 3rd Qu.: 51.240

## Max. :0 Max. :108.00 Max. :172.606

## avg\_roll\_forearm stddev\_roll\_forearm var\_roll\_forearm

## Min. :-177.234 Min. : 0.000 Min. : 0.00

## 1st Qu.: -0.909 1st Qu.: 0.428 1st Qu.: 0.18

## Median : 11.172 Median : 8.030 Median : 64.48

## Mean : 33.165 Mean : 41.986 Mean : 5274.10

## 3rd Qu.: 107.132 3rd Qu.: 85.373 3rd Qu.: 7289.08

## Max. : 177.256 Max. :179.171 Max. :32102.24

## avg\_pitch\_forearm stddev\_pitch\_forearm var\_pitch\_forearm

## Min. :-68.17 Min. : 0.000 Min. : 0.000

## 1st Qu.: 0.00 1st Qu.: 0.336 1st Qu.: 0.113

## Median : 12.02 Median : 5.516 Median : 30.425

## Mean : 11.79 Mean : 7.977 Mean : 139.593

## 3rd Qu.: 28.48 3rd Qu.:12.866 3rd Qu.: 165.532

## Max. : 72.09 Max. :47.745 Max. :2279.617

## avg\_yaw\_forearm stddev\_yaw\_forearm var\_yaw\_forearm gyros\_forearm\_x

## Min. :-155.06 Min. : 0.000 Min. : 0.00 Min. :-22.000

## 1st Qu.: -26.26 1st Qu.: 0.524 1st Qu.: 0.27 1st Qu.: -0.220

## Median : 0.00 Median : 24.743 Median : 612.21 Median : 0.050

## Mean : 18.00 Mean : 44.854 Mean : 4639.85 Mean : 0.158

## 3rd Qu.: 85.79 3rd Qu.: 85.817 3rd Qu.: 7368.41 3rd Qu.: 0.560

## Max. : 169.24 Max. :197.508 Max. :39009.33 Max. : 3.970

## gyros\_forearm\_y gyros\_forearm\_z accel\_forearm\_x accel\_forearm\_y

## Min. : -7.02000 Min. : -8.0900 Min. :-498.00 Min. :-632.0

## 1st Qu.: -1.46000 1st Qu.: -0.1800 1st Qu.:-178.00 1st Qu.: 57.0

## Median : 0.03000 Median : 0.0800 Median : -57.00 Median : 201.0

## Mean : 0.07517 Mean : 0.1512 Mean : -61.65 Mean : 163.7

## 3rd Qu.: 1.62000 3rd Qu.: 0.4900 3rd Qu.: 76.00 3rd Qu.: 312.0

## Max. :311.00000 Max. :231.0000 Max. : 477.00 Max. : 923.0

## accel\_forearm\_z magnet\_forearm\_x magnet\_forearm\_y magnet\_forearm\_z

## Min. :-446.00 Min. :-1280.0 Min. :-896.0 Min. :-973.0

## 1st Qu.:-182.00 1st Qu.: -616.0 1st Qu.: 2.0 1st Qu.: 191.0

## Median : -39.00 Median : -378.0 Median : 591.0 Median : 511.0

## Mean : -55.29 Mean : -312.6 Mean : 380.1 Mean : 393.6

## 3rd Qu.: 26.00 3rd Qu.: -73.0 3rd Qu.: 737.0 3rd Qu.: 653.0

## Max. : 291.00 Max. : 672.0 Max. :1480.0 Max. :1090.0

## classe

## A:5580

## B:3797

## C:3422

## D:3216

## E:3607

##

## [ reached getOption("max.print") -- omitted 1 row ]

## Data cleaning

# remove variables with nearly zero variance

nzv <- nearZeroVar(trainds)

trainds <- trainds[, -nzv]

# remove variables that are mostly NA

mostlyNA <- sapply(trainds, function(x) mean(is.na(x))) > 0.75

trainds <- trainds[, mostlyNA==F]

# remove variables which won't contribute much for prediction, in this case 1:5 variables are of no use.

trainds <- trainds[, -(1:5)]

dim(trainds)

## [1] 19622 54

# Data Split

set.seed(123)

traindssplit <- createDataPartition(y=trainds$classe, p=0.6, list=F)

trainds1 <- trainds[traindssplit, ]

trainds2 <- trainds[-traindssplit, ]

dim(trainds1)

## [1] 11776 54

dim(trainds2)

## [1] 7846 54

##Build model , i am using Random forest, Decision Trees and Boosting

## Model with Random forest

install.packages("randomForest")

## Installing package into 'C:/Users/Mahesha/Documents/R/win-library/3.4'

## (as 'lib' is unspecified)

## package 'randomForest' successfully unpacked and MD5 sums checked

##

## The downloaded binary packages are in

## C:\Users\Mahesha\AppData\Local\Temp\Rtmp8Y1oGj\downloaded\_packages

library(randomForest)

## 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

# use 3-fold CV to select optimal tuning parameters

fitControl <- trainControl(method="cv", number=3, verboseIter=F)

# model on trainds1

install.packages("e1071")

## Installing package into 'C:/Users/Mahesha/Documents/R/win-library/3.4'

## (as 'lib' is unspecified)

## package 'e1071' successfully unpacked and MD5 sums checked

##

## The downloaded binary packages are in

## C:\Users\Mahesha\AppData\Local\Temp\Rtmp8Y1oGj\downloaded\_packages

library(e1071)

model1 <- train(classe ~ ., data=trainds1, method="rf", trControl=fitControl)

# print final model to see tuning parameters it chose

model1$finalModel

##

## Call:

## randomForest(x = x, y = y, mtry = param$mtry)

## Type of random forest: classification

## Number of trees: 500

## No. of variables tried at each split: 27

##

## OOB estimate of error rate: 0.35%

## Confusion matrix:

## A B C D E class.error

## A 3346 1 0 0 1 0.0005973716

## B 7 2270 1 1 0 0.0039491005

## C 0 7 2045 2 0 0.0043816943

## D 0 0 13 1916 1 0.0072538860

## E 0 1 0 6 2158 0.0032332564

# use model1 to predict classe in validation set (trainds2)

preds1 <- predict(model1, newdata=trainds2)

# show confusion matrix to get estimate of out-of-sample error

confusionMatrix(trainds2$classe, preds1)

## Confusion Matrix and Statistics

##

## Reference

## Prediction A B C D E

## A 2232 0 0 0 0

## B 1 1514 3 0 0

## C 0 7 1358 3 0

## D 0 0 7 1279 0

## E 0 0 0 0 1442

##

## Overall Statistics

##

## Accuracy : 0.9973

## 95% CI : (0.9959, 0.9983)

## No Information Rate : 0.2846

## P-Value [Acc > NIR] : < 2.2e-16

##

## Kappa : 0.9966

## Mcnemar's Test P-Value : NA

##

## Statistics by Class:

##

## Class: A Class: B Class: C Class: D Class: E

## Sensitivity 0.9996 0.9954 0.9927 0.9977 1.0000

## Specificity 1.0000 0.9994 0.9985 0.9989 1.0000

## Pos Pred Value 1.0000 0.9974 0.9927 0.9946 1.0000

## Neg Pred Value 0.9998 0.9989 0.9985 0.9995 1.0000

## Prevalence 0.2846 0.1939 0.1744 0.1634 0.1838

## Detection Rate 0.2845 0.1930 0.1731 0.1630 0.1838

## Detection Prevalence 0.2845 0.1935 0.1744 0.1639 0.1838

## Balanced Accuracy 0.9998 0.9974 0.9956 0.9983 1.0000

##################################

## Model with Decision Trees

library(rpart)

set.seed(123)

model2 <- rpart(classe ~ ., data=trainds1, method="class")

preds2 <- predict(model2, trainds2, type = "class")

confusionMatrix(trainds2$classe, preds2)

## Confusion Matrix and Statistics

##

## Reference

## Prediction A B C D E

## A 2031 90 0 86 25

## B 243 1044 76 60 95

## C 43 135 1043 129 18

## D 62 110 42 993 79

## E 36 62 5 121 1218

##

## Overall Statistics

##

## Accuracy : 0.8067

## 95% CI : (0.7977, 0.8153)

## No Information Rate : 0.3078

## P-Value [Acc > NIR] : < 2.2e-16

##

## Kappa : 0.7547

## Mcnemar's Test P-Value : < 2.2e-16

##

## Statistics by Class:

##

## Class: A Class: B Class: C Class: D Class: E

## Sensitivity 0.8410 0.7245 0.8945 0.7149 0.8488

## Specificity 0.9630 0.9260 0.9513 0.9546 0.9651

## Pos Pred Value 0.9099 0.6877 0.7624 0.7722 0.8447

## Neg Pred Value 0.9316 0.9373 0.9810 0.9396 0.9661

## Prevalence 0.3078 0.1837 0.1486 0.1770 0.1829

## Detection Rate 0.2589 0.1331 0.1329 0.1266 0.1552

## Detection Prevalence 0.2845 0.1935 0.1744 0.1639 0.1838

## Balanced Accuracy 0.9020 0.8252 0.9229 0.8348 0.9069

#######################################

## Model with GBM (Generalised Boosting)

install.packages("gbm")

## Installing package into 'C:/Users/Mahesha/Documents/R/win-library/3.4'

## (as 'lib' is unspecified)

## package 'gbm' successfully unpacked and MD5 sums checked

##

## The downloaded binary packages are in

## C:\Users\Mahesha\AppData\Local\Temp\Rtmp8Y1oGj\downloaded\_packages

library(gbm)

## 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 2.1.3

set.seed(123)

fitControl <- trainControl(method = "repeatedcv",

number = 3,

repeats = 1)

model3 <- train(classe ~ ., data=trainds1, method = "gbm",

trControl = fitControl,

verbose = FALSE)

gbmFinMod3 <- model3$finalModel

preds3 <- predict(model3, newdata=trainds2)

confusionMatrix(trainds2$classe, preds3)

## Confusion Matrix and Statistics

##

## Reference

## Prediction A B C D E

## A 2228 3 0 1 0

## B 10 1494 12 2 0

## C 0 17 1347 3 1

## D 0 7 16 1261 2

## E 1 3 4 9 1425

##

## Overall Statistics

##

## Accuracy : 0.9884

## 95% CI : (0.9858, 0.9907)

## No Information Rate : 0.2854

## P-Value [Acc > NIR] : < 2.2e-16

##

## Kappa : 0.9853

## Mcnemar's Test P-Value : NA

##

## Statistics by Class:

##

## Class: A Class: B Class: C Class: D Class: E

## Sensitivity 0.9951 0.9803 0.9768 0.9882 0.9979

## Specificity 0.9993 0.9962 0.9968 0.9962 0.9974

## Pos Pred Value 0.9982 0.9842 0.9846 0.9806 0.9882

## Neg Pred Value 0.9980 0.9953 0.9951 0.9977 0.9995

## Prevalence 0.2854 0.1942 0.1758 0.1626 0.1820

## Detection Rate 0.2840 0.1904 0.1717 0.1607 0.1816

## Detection Prevalence 0.2845 0.1935 0.1744 0.1639 0.1838

## Balanced Accuracy 0.9972 0.9883 0.9868 0.9922 0.9976

## Models Comparison

## Considering the Accuracy , model1 ( Random Forest) looks better compared to model2 and model3.

## Prediction for test dataset using model1

## Data cleaning for test dataset just like training dataset

# remove variables with nearly zero variance

nzv2 <- nearZeroVar(testds)

testds <- testds[, -nzv2]

# remove variables that are mostly NA

mostlyNA2 <- sapply(testds, function(x) mean(is.na(x))) > 0.75

testds <- testds[, mostlyNA2==F]

# remove variables which won't contribute much for prediction, in this case 1:5 variables are of no use.

testds <- testds[, -(1:5)]

dim(testds)

## [1] 20 54

## Final prediction for test dataset

predsfinal <- predict(model1, newdata=testds)

predsfinal

## [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

## writing to a file

write.table(predsfinal,"Outputfinal1.txt")

write.table(predsfinal,"Outputfinal2.csv")