```
library(tidyverse)
library(caret)
library(ggplot2)
library(utils)
library(skimr)
(scipen=999)

temp = tempfile()
download. file('https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv',temp)
temp1 = tempfile()
download. file('https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv',temp1)

df_fit <- read.csv(temp)
test_fit <- read.csv(temp1)
```

Los datos que vamos a utilizar fueron obtenidos de http://groupware.les.inf.puc-rio.br/har.

Al efectuar un EDA a los datasets, encuentro valores NA en varias de sus columnas, valores n_missing, los cuales no nos sirven para nuestro modelo de machine learning, también observo que son demasiados para imputarlos mediante la técnica de Knn, por lo cual los voy a excluir.

Conjunto de Prueba:

> str(df_fit)

```
> str(df_fit)
'data.frame':  19622 obs. of  160 variables:
 $ X                       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ user_name               : chr  "carlitos" "carlitos" "carlitos" "carlitos" ...
 $ raw_timestamp_part_1    : int  1323084231 1323084231 1323084231 1323084232 13230842
32 1323084232 1323084232 1323084232 1323084232 1323084232 ...
 $ raw_timestamp_part_2    : int  788290 808298 820366 120339 196328 304277 368296 440
390 484323 484434 ...
 $ cvtd_timestamp          : chr  "05/12/2011 11:23" "05/12/2011 11:23" "05/12/2011 11
:23" "05/12/2011 11:23" ...
 $ new_window              : chr  "no" "no" "no" "no" ...
 $ 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      : chr  "" "" "" "" ...
 $ kurtosis_picth_belt     : chr  "" "" "" "" ...
 $ kurtosis_yaw_belt       : chr  "" "" "" "" ...
 $ skewness_roll_belt      : chr  "" "" "" "" ...
 $ skewness_roll_belt.1    : chr  "" "" "" "" ...
 $ skewness_yaw_belt       : chr  "" "" "" "" ...
 $ 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            : chr  "" "" "" "" ...
 $ 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            : chr  "" "" "" "" ...
```

```
 $ 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      : chr  "" "" "" "" ...
 $ 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 -0.0
2 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       : chr  "" "" "" "" ...
 $ kurtosis_picth_arm      : chr  "" "" "" "" ...
 $ kurtosis_yaw_arm        : chr  "" "" "" "" ...
 $ skewness_roll_arm       : chr  "" "" "" "" ...
 $ skewness_pitch_arm      : chr  "" "" "" "" ...
 $ skewness_yaw_arm        : chr  "" "" "" "" ...
 $ 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  : chr  "" "" "" "" ...
 $ kurtosis_picth_dumbbell : chr  "" "" "" "" ...
 $ kurtosis_yaw_dumbbell   : chr  "" "" "" "" ...
 $ skewness_roll_dumbbell  : chr  "" "" "" "" ...
 $ skewness_pitch_dumbbell : chr  "" "" "" "" ...
```

```
 $ skewness_yaw_dumbbell   : chr  "" "" "" "" ...
 $ 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        : chr  "" "" "" "" ...
 $ 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        : chr  "" "" "" "" ...
 $ amplitude_roll_dumbbell : num  NA NA NA NA NA NA NA NA NA NA ...
  [list output truncated]
```

```
> skim(df_fit)
── Data Summary ────────────────────────
                           Values
Name                       df_fit
Number of rows             19622
Number of columns          160
_______________________
Column type frequency:
  character                37
  numeric                  123
_______________________
Group variables            None

── Variable type: character ──────────────────────────────────────────────────────────────
   skim_variable           n_missing complete_rate min max empty n_unique whitespace
 1 user_name                       0             1   5   8     0        6          0
 2 cvtd_timestamp                  0             1  16  16     0       20          0
 3 new_window                      0             1   2   3     0        2          0
 4 kurtosis_roll_belt              0             1   0   9 19216      397          0
 5 kurtosis_picth_belt             0             1   0   9 19216      317          0
 6 kurtosis_yaw_belt               0             1   0   7 19216        2          0
 7 skewness_roll_belt              0             1   0   9 19216      395          0
 8 skewness_roll_belt.1            0             1   0   9 19216      338          0
 9 skewness_yaw_belt               0             1   0   7 19216        2          0
10 max_yaw_belt                    0             1   0   7 19216       68          0
11 min_yaw_belt                    0             1   0   7 19216       68          0
12 amplitude_yaw_belt              0             1   0   7 19216        4          0
13 kurtosis_roll_arm               0             1   0   8 19216      330          0
14 kurtosis_picth_arm              0             1   0   8 19216      328          0
15 kurtosis_yaw_arm                0             1   0   8 19216      395          0
16 skewness_roll_arm               0             1   0   8 19216      331          0
17 skewness_pitch_arm              0             1   0   8 19216      328          0
18 skewness_yaw_arm                0             1   0   8 19216      395          0
19 kurtosis_roll_dumbbell          0             1   0   7 19216      398          0
20 kurtosis_picth_dumbbell         0             1   0   7 19216      401          0
21 kurtosis_yaw_dumbbell           0             1   0   7 19216        2          0
22 skewness_roll_dumbbell          0             1   0   7 19216      401          0
23 skewness_pitch_dumbbell         0             1   0   7 19216      402          0
24 skewness_yaw_dumbbell           0             1   0   7 19216        2          0
25 max_yaw_dumbbell                0             1   0   7 19216       73          0
26 min_yaw_dumbbell                0             1   0   7 19216       73          0
27 amplitude_yaw_dumbbell          0             1   0   7 19216        3          0
28 kurtosis_roll_forearm           0             1   0   7 19216      322          0
29 kurtosis_picth_forearm          0             1   0   7 19216      323          0
30 kurtosis_yaw_forearm            0             1   0   7 19216        2          0
31 skewness_roll_forearm           0             1   0   7 19216      323          0
32 skewness_pitch_forearm          0             1   0   7 19216      319          0
33 skewness_yaw_forearm            0             1   0   7 19216        2          0
34 max_yaw_forearm                 0             1   0   7 19216       45          0
35 min_yaw_forearm                 0             1   0   7 19216       45          0
36 amplitude_yaw_forearm           0             1   0   7 19216        3          0
37 classe                          0             1   1   1     0        5          0

── Variable type: numeric ────────────────────────────────────────────────────────────────
    skim_variable             n_missing complete_rate     mean          sd
  1 X                                 0             1  9.81e+3   5665.
  2 raw_timestamp_part_1              0             1  1.32e+9 204928.
  3 raw_timestamp_part_2              0             1  5.01e+5 288223.
  4 num_window                        0             1  4.31e+2    248.
  5 roll_belt                         0             1  6.44e+1     62.8
  6 pitch_belt                        0             1  3.05e-1     22.4
  7 yaw_belt                          0             1 -1.12e+1     95.2
  8 total_accel_belt                  0             1  1.13e+1      7.74
  9 max_roll_belt                 19216             0.0207 -6.67e+0     94.6
 10 max_picth_belt                19216             0.0207  1.29e+1      8.01
```

```
11 min_roll_belt                19216         0.0207 -1.04e+1     93.6
12 min_pitch_belt               19216         0.0207  1.08e+1      7.47
13 amplitude_roll_belt          19216         0.0207  3.77e+0     25.3
14 amplitude_pitch_belt         19216         0.0207  2.17e+0      2.36
15 var_total_accel_belt         19216         0.0207  9.26e-1      2.22
16 avg_roll_belt                19216         0.0207  6.81e+1     63.1
17 stddev_roll_belt             19216         0.0207  1.34e+0      2.44
18 var_roll_belt                19216         0.0207  7.70e+0     23.2
19 avg_pitch_belt               19216         0.0207  5.20e-1     22.4
20 stddev_pitch_belt            19216         0.0207  6.03e-1      0.639
21 var_pitch_belt               19216         0.0207  7.66e-1      1.76
22 avg_yaw_belt                 19216         0.0207 -8.83e+0     93.5
23 stddev_yaw_belt              19216         0.0207  1.34e+0     10.3
24 var_yaw_belt                 19216         0.0207  1.07e+2   1656.
25 gyros_belt_x                     0              1 -5.59e-3      0.207
26 gyros_belt_y                     0              1  3.96e-2      0.0782
27 gyros_belt_z                     0              1 -1.31e-1      0.241
28 accel_belt_x                     0              1 -5.59e+0     29.6
29 accel_belt_y                     0              1  3.02e+1     28.6
30 accel_belt_z                     0              1 -7.26e+1    100.
31 magnet_belt_x                    0              1  5.56e+1     64.2
32 magnet_belt_y                    0              1  5.94e+2     35.7
33 magnet_belt_z                    0              1 -3.45e+2     65.2
34 roll_arm                         0              1  1.78e+1     72.7
35 pitch_arm                        0              1 -4.61e+0     30.7
36 yaw_arm                          0              1 -6.19e-1     71.4
37 total_accel_arm                  0              1  2.55e+1     10.5
38 var_accel_arm                19216         0.0207  5.32e+1     54.0
39 avg_roll_arm                 19216         0.0207  1.27e+1     68.6
40 stddev_roll_arm              19216         0.0207  1.12e+1     17.1
41 var_roll_arm                 19216         0.0207  4.17e+2   2007.
42 avg_pitch_arm                19216         0.0207 -4.90e+0     26.8
43 stddev_pitch_arm             19216         0.0207  1.04e+1      9.40
44 var_pitch_arm                19216         0.0207  1.96e+2    293.
45 avg_yaw_arm                  19216         0.0207  2.36e+0     61.3
46 stddev_yaw_arm               19216         0.0207  2.23e+1     23.7
47 var_yaw_arm                  19216         0.0207  1.06e+3   2722.
48 gyros_arm_x                      0              1  4.28e-2      1.99
49 gyros_arm_y                      0              1 -2.57e-1      0.851
50 gyros_arm_z                      0              1  2.69e-1      0.553
51 accel_arm_x                      0              1 -6.02e+1    182.
52 accel_arm_y                      0              1  3.26e+1    110.
53 accel_arm_z                      0              1 -7.12e+1    135.
54 magnet_arm_x                     0              1  1.92e+2    444.
55 magnet_arm_y                     0              1  1.57e+2    202.
56 magnet_arm_z                     0              1  3.06e+2    327.
57 max_roll_arm                 19216         0.0207  1.12e+1     26.9
58 max_picth_arm                19216         0.0207  3.58e+1     69.6
59 max_yaw_arm                  19216         0.0207  3.55e+1     10.4
60 min_roll_arm                 19216         0.0207 -2.12e+1     28.7
61 min_pitch_arm                19216         0.0207 -3.39e+1     60.8
62 min_yaw_arm                  19216         0.0207  1.47e+1      9.11
63 amplitude_roll_arm           19216         0.0207  3.25e+1     27.4
64 amplitude_pitch_arm          19216         0.0207  6.97e+1     67.0
65 amplitude_yaw_arm            19216         0.0207  2.08e+1     12.3
66 roll_dumbbell                    0              1  2.38e+1     69.9
67 pitch_dumbbell                   0              1 -1.08e+1     37.0
68 yaw_dumbbell                     0              1  1.67e+0     82.5
69 max_roll_dumbbell            19216         0.0207  1.38e+1     48.3
70 max_picth_dumbbell           19216         0.0207  3.27e+1     93.4
71 min_roll_dumbbell            19216         0.0207 -4.12e+1     34.7
72 min_pitch_dumbbell           19216         0.0207 -3.32e+1     74.3
73 amplitude_roll_dumbbell      19216         0.0207  5.50e+1     54.9
74 amplitude_pitch_dumbbell     19216         0.0207  6.59e+1     65.2
75 total_accel_dumbbell             0              1  1.37e+1     10.2
76 var_accel_dumbbell           19216         0.0207  4.39e+0     13.5
77 avg_roll_dumbbell            19216         0.0207  2.39e+1     62.9
78 stddev_roll_dumbbell         19216         0.0207  2.08e+1     24.3
79 var_roll_dumbbell            19216         0.0207  1.02e+3   2263.
80 avg_pitch_dumbbell           19216         0.0207 -1.23e+1     32.1
81 stddev_pitch_dumbbell        19216         0.0207  1.31e+1     13.3
82 var_pitch_dumbbell           19216         0.0207  3.50e+2    674.
83 avg_yaw_dumbbell             19216         0.0207  2.02e-1     78.2
84 stddev_yaw_dumbbell          19216         0.0207  1.66e+1     17.7
85 var_yaw_dumbbell             19216         0.0207  5.90e+2   1245.
86 gyros_dumbbell_x                 0              1  1.61e-1      1.51
```

```
 87 gyros_dumbbell_y                0        1       4.61e-2      0.610
 88 gyros_dumbbell_z                0        1      -1.29e-1      2.29
 89 accel_dumbbell_x                0        1      -2.86e+1     67.3
 90 accel_dumbbell_y                0        1       5.26e+1     80.8
 91 accel_dumbbell_z                0        1      -3.83e+1    109.
 92 magnet_dumbbell_x               0        1      -3.28e+2    340.
 93 magnet_dumbbell_y               0        1       2.21e+2    327.
 94 magnet_dumbbell_z               0        1       4.61e+1    140.
 95 roll_forearm                    0        1       3.38e+1    108.
 96 pitch_forearm                   0        1       1.07e+1     28.1
 97 yaw_forearm                     0        1       1.92e+1    103.
 98 max_roll_forearm            19216        0.0207  2.45e+1     31.0
 99 max_picth_forearm           19216        0.0207  8.15e+1     95.5
100 min_roll_forearm            19216        0.0207 -1.67e-1     22.6
101 min_pitch_forearm           19216        0.0207 -5.76e+1    111.
102 amplitude_roll_forearm      19216        0.0207  2.47e+1     25.9
103 amplitude_pitch_forearm     19216        0.0207  1.39e+2    148.
104 total_accel_forearm             0        1       3.47e+1     10.1
105 var_accel_forearm           19216        0.0207  3.35e+1     34.0
106 avg_roll_forearm            19216        0.0207  3.32e+1     79.5
107 stddev_roll_forearm         19216        0.0207  4.20e+1     59.3
108 var_roll_forearm            19216        0.0207  5.27e+3   9177.
109 avg_pitch_forearm           19216        0.0207  1.18e+1     24.8
110 stddev_pitch_forearm        19216        0.0207  7.98e+0      8.73
111 var_pitch_forearm           19216        0.0207  1.40e+2    266.
112 avg_yaw_forearm             19216        0.0207  1.80e+1     77.6
113 stddev_yaw_forearm          19216        0.0207  4.49e+1     51.3
114 var_yaw_forearm             19216        0.0207  4.64e+3   7285.
115 gyros_forearm_x                 0        1       1.58e-1      0.649
116 gyros_forearm_y                 0        1       7.52e-2      3.10
117 gyros_forearm_z                 0        1       1.51e-1      1.75
118 accel_forearm_x                 0        1      -6.17e+1    181.
119 accel_forearm_y                 0        1       1.64e+2    200.
120 accel_forearm_z                 0        1      -5.53e+1    138.
121 magnet_forearm_x                0        1      -3.13e+2    347.
122 magnet_forearm_y                0        1       3.80e+2    509.
123 magnet_forearm_z                0        1       3.94e+2    369.
```

Selecciono y excluyo todas aquellas columnas que contengan un porcentaje de NA mayor al 20%.

- df_fit <- df_fit[, -which(colMeans(is.na(df_fit))>=0.2)]

Selecciono y excluyo todas aquellas columnas que empiezen con 'kurt','ske','max','min','ampli', por ser columnas con valores vacíos.

- df_fit <- df_fit %>% select(-starts_with(c('kurt','ske','max','min','ampli')))

Convierto la variable classe a factor, ya que es numérica.

- df_fit <- df_fit %>% mutate (classe = as.factor(classe))

Aplico una conversión de tipo a la variable cvtd_timestamp, ya que es tipo carácter y expresa una fecha.

df_fit <- df_fit %>% mutate (cvtd_timestamp = as.Date(cvtd_timestamp,"%d/%m/%y"))

Conjunto de Testing:

```
> skim(test_fit)
── Data Summary ────────────────────────
                           Values
Name                       test_fit
Number of rows             20
Number of columns          160
_______________________
Column type frequency:
  character                3
  logical                  100
  numeric                  57
________________________
```

```
Group variables            None

── Variable type: character ──────────────────────────────────────────────────
  skim_variable  n_missing complete_rate min max empty n_unique whitespace
1 user_name              0             1   5   8     0        6          0
2 cvtd_timestamp         0             1  16  16     0       11          0
3 new_window             0             1   2   2     0        1          0

── Variable type: logical ────────────────────────────────────────────────────
    skim_variable            n_missing complete_rate mean count
  1 kurtosis_roll_belt              20             0  NaN ": "
  2 kurtosis_picth_belt             20             0  NaN ": "
  3 kurtosis_yaw_belt               20             0  NaN ": "
  4 skewness_roll_belt              20             0  NaN ": "
  5 skewness_roll_belt.1            20             0  NaN ": "
  6 skewness_yaw_belt               20             0  NaN ": "
  7 max_roll_belt                   20             0  NaN ": "
  8 max_picth_belt                  20             0  NaN ": "
  9 max_yaw_belt                    20             0  NaN ": "
 10 min_roll_belt                   20             0  NaN ": "
 11 min_pitch_belt                  20             0  NaN ": "
 12 min_yaw_belt                    20             0  NaN ": "
 13 amplitude_roll_belt             20             0  NaN ": "
 14 amplitude_pitch_belt            20             0  NaN ": "
 15 amplitude_yaw_belt              20             0  NaN ": "
 16 var_total_accel_belt            20             0  NaN ": "
 17 avg_roll_belt                   20             0  NaN ": "
 18 stddev_roll_belt                20             0  NaN ": "
 19 var_roll_belt                   20             0  NaN ": "
 20 avg_pitch_belt                  20             0  NaN ": "
 21 stddev_pitch_belt               20             0  NaN ": "
 22 var_pitch_belt                  20             0  NaN ": "
 23 avg_yaw_belt                    20             0  NaN ": "
 24 stddev_yaw_belt                 20             0  NaN ": "
 25 var_yaw_belt                    20             0  NaN ": "
 26 var_accel_arm                   20             0  NaN ": "
 27 avg_roll_arm                    20             0  NaN ": "
 28 stddev_roll_arm                 20             0  NaN ": "
 29 var_roll_arm                    20             0  NaN ": "
 30 avg_pitch_arm                   20             0  NaN ": "
 31 stddev_pitch_arm                20             0  NaN ": "
 32 var_pitch_arm                   20             0  NaN ": "
 33 avg_yaw_arm                     20             0  NaN ": "
 34 stddev_yaw_arm                  20             0  NaN ": "
 35 var_yaw_arm                     20             0  NaN ": "
 36 kurtosis_roll_arm               20             0  NaN ": "
 37 kurtosis_picth_arm              20             0  NaN ": "
 38 kurtosis_yaw_arm                20             0  NaN ": "
 39 skewness_roll_arm               20             0  NaN ": "
 40 skewness_pitch_arm              20             0  NaN ": "
 41 skewness_yaw_arm                20             0  NaN ": "
 42 max_roll_arm                    20             0  NaN ": "
 43 max_picth_arm                   20             0  NaN ": "
 44 max_yaw_arm                     20             0  NaN ": "
 45 min_roll_arm                    20             0  NaN ": "
 46 min_pitch_arm                   20             0  NaN ": "
 47 min_yaw_arm                     20             0  NaN ": "
 48 amplitude_roll_arm              20             0  NaN ": "
 49 amplitude_pitch_arm             20             0  NaN ": "
 50 amplitude_yaw_arm               20             0  NaN ": "
 51 kurtosis_roll_dumbbell          20             0  NaN ": "
 52 kurtosis_picth_dumbbell         20             0  NaN ": "
 53 kurtosis_yaw_dumbbell           20             0  NaN ": "
 54 skewness_roll_dumbbell          20             0  NaN ": "
 55 skewness_pitch_dumbbell         20             0  NaN ": "
 56 skewness_yaw_dumbbell           20             0  NaN ": "
 57 max_roll_dumbbell               20             0  NaN ": "
 58 max_picth_dumbbell              20             0  NaN ": "
 59 max_yaw_dumbbell                20             0  NaN ": "
 60 min_roll_dumbbell               20             0  NaN ": "
 61 min_pitch_dumbbell              20             0  NaN ": "
 62 min_yaw_dumbbell                20             0  NaN ": "
 63 amplitude_roll_dumbbell         20             0  NaN ": "
 64 amplitude_pitch_dumbbell        20             0  NaN ": "
 65 amplitude_yaw_dumbbell          20             0  NaN ": "
 66 var_accel_dumbbell              20             0  NaN ": "
```

```
 67 avg_roll_dumbbell               20             0  NaN ": "
 68 stddev_roll_dumbbell            20             0  NaN ": "
 69 var_roll_dumbbell               20             0  NaN ": "
 70 avg_pitch_dumbbell              20             0  NaN ": "
 71 stddev_pitch_dumbbell           20             0  NaN ": "
 72 var_pitch_dumbbell              20             0  NaN ": "
 73 avg_yaw_dumbbell                20             0  NaN ": "
 74 stddev_yaw_dumbbell             20             0  NaN ": "
 75 var_yaw_dumbbell                20             0  NaN ": "
 76 kurtosis_roll_forearm           20             0  NaN ": "
 77 kurtosis_picth_forearm          20             0  NaN ": "
 78 kurtosis_yaw_forearm            20             0  NaN ": "
 79 skewness_roll_forearm           20             0  NaN ": "
 80 skewness_pitch_forearm          20             0  NaN ": "
 81 skewness_yaw_forearm            20             0  NaN ": "
 82 max_roll_forearm                20             0  NaN ": "
 83 max_picth_forearm               20             0  NaN ": "
 84 max_yaw_forearm                 20             0  NaN ": "
 85 min_roll_forearm                20             0  NaN ": "
 86 min_pitch_forearm               20             0  NaN ": "
 87 min_yaw_forearm                 20             0  NaN ": "
 88 amplitude_roll_forearm          20             0  NaN ": "
 89 amplitude_pitch_forearm         20             0  NaN ": "
 90 amplitude_yaw_forearm           20             0  NaN ": "
 91 var_accel_forearm               20             0  NaN ": "
 92 avg_roll_forearm                20             0  NaN ": "
 93 stddev_roll_forearm             20             0  NaN ": "
 94 var_roll_forearm                20             0  NaN ": "
 95 avg_pitch_forearm               20             0  NaN ": "
 96 stddev_pitch_forearm            20             0  NaN ": "
 97 var_pitch_forearm               20             0  NaN ": "
 98 avg_yaw_forearm                 20             0  NaN ": "
 99 stddev_yaw_forearm              20             0  NaN ": "
100 var_yaw_forearm                 20             0  NaN ": "

── Variable type: numeric ──────────────────────────────────────────────────────
   skim_variable        n_missing complete_rate     mean          sd          p0
 1 X                            0             1  1.05e+1      5.92           1
 2 raw_timestamp_part_1         0             1  1.32e+9 230560.     1322489635
 3 raw_timestamp_part_2         0             1  5.12e+5 303068.           36553
 4 num_window                   0             1  3.80e+2    219.             48
 5 roll_belt                    0             1  3.13e+1     54.3            -5.92
 6 pitch_belt                   0             1  5.82e+0     14.6           -41.6
 7 yaw_belt                     0             1 -5.93e+1     62.4           -93.7
 8 total_accel_belt             0             1  7.55e+0      6.86           2
 9 gyros_belt_x                 0             1 -4.5 e-2      0.196         -0.5
10 gyros_belt_y                 0             1  1   e-2      0.0397        -0.05
11 gyros_belt_z                 0             1 -1.01e-1      0.167         -0.48
12 accel_belt_x                 0             1 -1.35e+1     19.8           -48
13 accel_belt_y                 0             1  1.84e+1     28.0           -16
14 accel_belt_z                 0             1 -1.76e+1     90.7          -187
15 magnet_belt_x                0             1  3.52e+1     40.7           -13
16 magnet_belt_y                0             1  6.01e+2     27.1           566
17 magnet_belt_z                0             1 -3.47e+2     51.0          -426
18 roll_arm                     0             1  1.64e+1     71.3          -137
19 pitch_arm                    0             1 -3.95e+0     23.5           -63.8
20 yaw_arm                      0             1 -2.8 e+0     94.7          -167
21 total_accel_arm              0             1  2.64e+1     11.2             3
22 gyros_arm_x                  0             1  7.7 e-2      1.90           -3.71
23 gyros_arm_y                  0             1 -1.60e-1      0.923          -2.09
24 gyros_arm_z                  0             1  1.20e-1      0.533          -0.69
25 accel_arm_x                  0             1 -1.35e+2    152.            -341
26 accel_arm_y                  0             1  1.03e+2     92.8            -65
27 accel_arm_z                  0             1 -8.78e+1    110.            -404
28 magnet_arm_x                 0             1 -3.90e+1    430.            -428
29 magnet_arm_y                 0             1  2.39e+2    211.            -307
30 magnet_arm_z                 0             1  3.70e+2    288.            -499
31 roll_dumbbell                0             1  3.38e+1     62.3           -111.
32 pitch_dumbbell               0             1 -1.95e+1     43.4            -55.0
33 yaw_dumbbell                 0             1 -9.38e-1     83.7           -103.
34 total_accel_dumbbell         0             1  1.72e+1     11.7             1
35 gyros_dumbbell_x             0             1  2.69e-1      0.480          -1.03
36 gyros_dumbbell_y             0             1  6.05e-2      0.642          -1.11
37 gyros_dumbbell_z             0             1 -2.66e-1      0.495          -1.18
38 accel_dumbbell_x             0             1 -4.76e+1     93.4           -159
39 accel_dumbbell_y             0             1  7.06e+1     74.1            -30
```

```
40 accel_dumbbell_z            0             1 -6   e+1    130.          -221
41 magnet_dumbbell_x           0             1 -3.04e+2    394.          -576
42 magnet_dumbbell_y           0             1  1.89e+2    318.          -558
43 magnet_dumbbell_z           0             1  7.14e+1    156.          -164
44 roll_forearm                0             1  3.87e+1    134.          -176
45 pitch_forearm               0             1  7.10e+0     32.9          -63.5
46 yaw_forearm                 0             1  2.19e+0    114.          -168
47 total_accel_forearm         0             1  3.20e+1      8.36          21
48 gyros_forearm_x             0             1 -2   e-2      0.687         -1.06
49 gyros_forearm_y             0             1 -4.15e-2      2.87          -5.97
50 gyros_forearm_z             0             1  2.61e-1      0.790         -1.26
51 accel_forearm_x             0             1  3.88e+1    157.          -212
52 accel_forearm_y             0             1  1.25e+2    191.          -331
53 accel_forearm_z             0             1 -9.37e+1    149.          -282
54 magnet_forearm_x            0             1 -1.59e+2    362.          -714
55 magnet_forearm_y            0             1  1.92e+2    619.          -787
56 magnet_forearm_z            0             1  4.60e+2    282.           -32
57 problem_id                  0             1  1.05e+1      5.92           1
```

El dataset correspondiente al conjunto de testing, lo voy a descartar ya que carece de la variable a predecir y no es posible estimarla ya que correría el riesgo de introducir un sesgo indeseado, además el dataset de prueba es bastante extenso y puedo hacer una partición que sirva para esos fines.

Preparar Datos:

Antes de lanzar el modelo de machine learning, debemos realizar varios pasos, que nos permiten crear un modelo de forma óptima.

Feature Selection,encontrar variables con varianza cero:

```
num_cols <- sapply(df_fit, is.numeric)

varianza <- nearZeroVar(df_fit[num_cols],saveMetrics = T)

varianza

table(varianza$nzv)

FALSE
  56
```

Como vemos no tenemos variables con varianza cero que debamos excluir.

Buscar variables correlacionadas:

```
train_fit_cor <- cor(df_fit[num_cols])

eliminate <- findCorrelation(train_fit_cor,verbose = T,names = T)
```

> findCorrelation(train_fit_cor,verbose = T,names = T)

```
Compare row 14  and column  5 with corr  0.992

  Means:  0.261 vs 0.157 so flagging column 14

Compare row 5  and column  13 with corr  0.925

  Means:  0.241 vs 0.154 so flagging column 5

Compare row 13  and column  8 with corr  0.928

  Means:  0.225 vs 0.151 so flagging column 13
```

Compare row 12 and column 6 with corr 0.966

  Means: 0.233 vs 0.147 so flagging column 12

Compare row 23 and column 22 with corr 0.918

  Means: 0.087 vs 0.147 so flagging column 22

Compare row 50 and column 35 with corr 0.914

  Means: 0.094 vs 0.15 so flagging column 35

Compare row 50 and column 37 with corr 0.933

  Means: 0.077 vs 0.153 so flagging column 37

Como Podemos apreciar estas columnas las debemos eliminar por estar altamente correlacionadas.

También buscamos variables que sean combinaciones lineales.

findLinearCombos(train_fit_cor)

> findLinearCombos(train_fit_cor)

$linearCombos

list()

$remove

NULL

En consecuencia, eliminamos las columnas que son correlaciones.

df_fit <- df_fit %>% select (-eliminate [1:7])

Aplicando un análisis de componentes principales PCA con el objetivo de reducir dimensiones y lograr asi un mejor proceso de nuestro algoritmo.

pre_pca <- preProcess(df_fit,method = "pca",thresh = 0.8)

df_preProc <- predict(pre_pca,df_fit)

> pre_pca

Created from 19622 samples and 60 variables

 Pre-processing:

  - centered (56)

  - ignored (4)

  - principal component signal extraction (56)

  - scaled (56)

PCA needed 14 components to capture 80 percent of the variance

> dim(df_preProc)

[1] 19622   19

Nuestro dataframe ahora consta de 19 predictores incluyendo la variable classe que es lo que vamos a predecir, con esto concluimos la transformación de datos.

Crear partición de datos en train y test:

```
intrain <- createDataPartition(y = df_preProc$classe,p = 0.85,list = F)
training <- df_preProc[intrain,]
testing <- df_preProc[-intrain,]
```

#Paralelizacion

```
library(doParallel)
cl=makePSOCKcluster(5)
registerDoParallel(cl)
```

#Modelización

```
set.seed(1235)

cross_valid <- trainControl(method = "repeatedcv",
                            number = 10,
                            repeats = 10)

model_rf <- train (classe~., data = training,method = "rf",trControl = cross_valid)
```

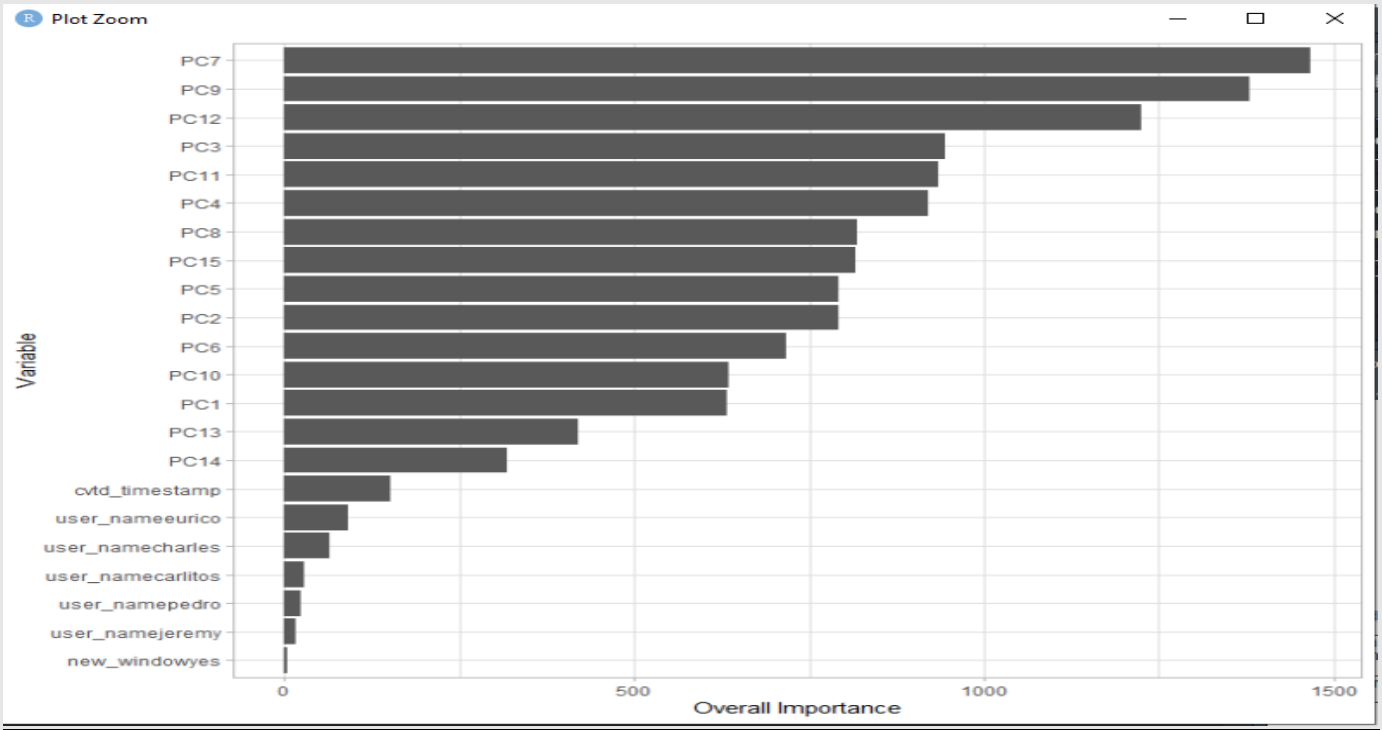
#Importancia de Variables

> var

rf variable importance

only 20 most important variables shown (out of 22)

	Overall
PC7	1466.44
PC9	1379.63
PC12	1224.25
PC3	945.09
PC11	934.86
PC4	921.03
PC8	817.74
PC15	815.65
PC5	791.95
PC2	791.64
PC6	717.25
PC10	635.90
PC1	632.16
PC13	420.17
PC14	317.40
cvtd_timestamp	150.29
user_nameeurico	90.21
user_namecharles	63.48
user_namecarlitos	27.73
user_namepedro	22.12



> model_rf

Random Forest

16680 samples

18 predictor

5 classes: 'A', 'B', 'C', 'D', 'E'

No pre-processing

Resampling: Cross-Validated (10 fold, repeated 10 times)

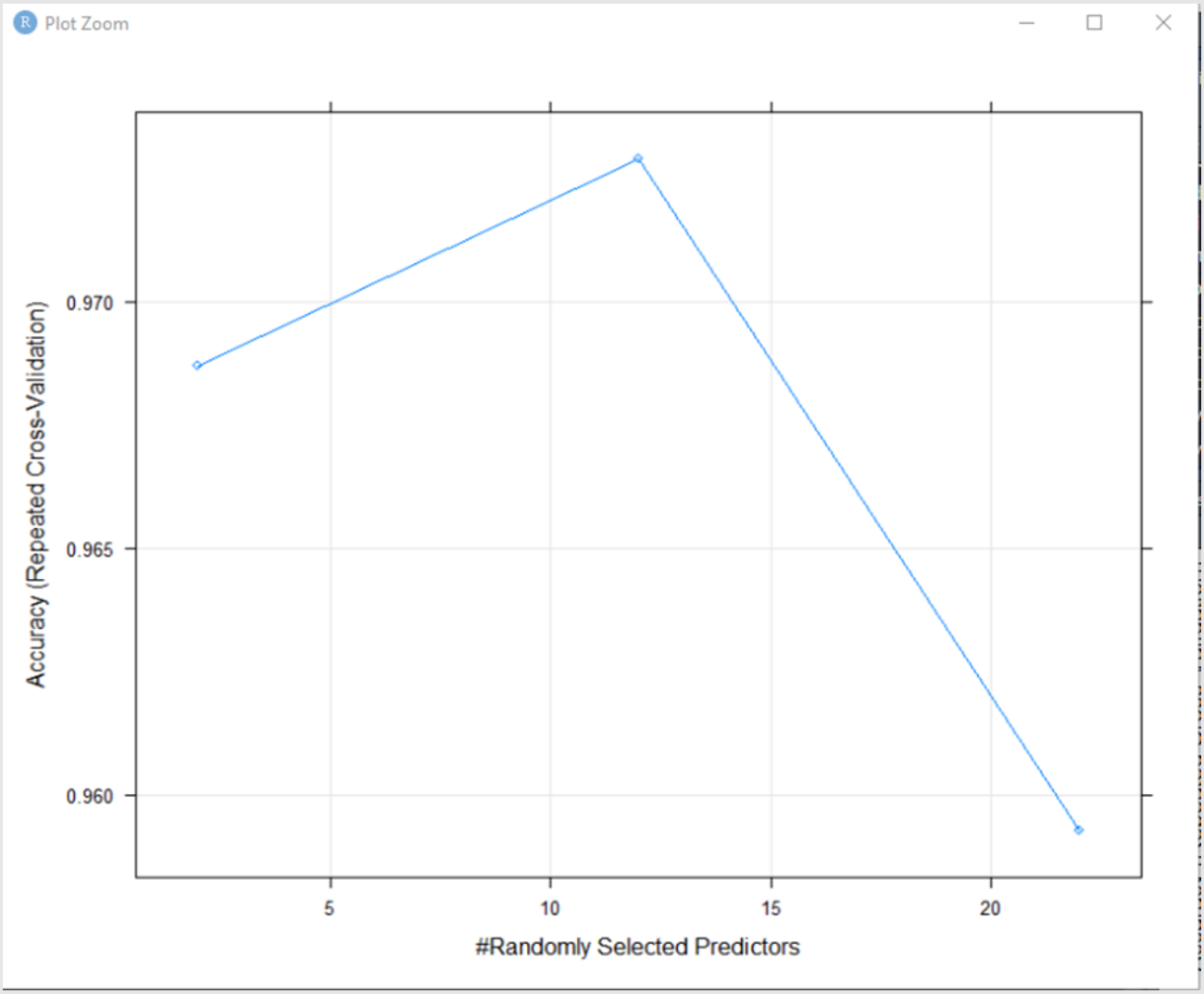
Summary of sample sizes: 15010, 15011, 15011, 15011, 15013, 15012, ...

Resampling results across tuning parameters:

mtry	Accuracy	Kappa
2	0.9672303	0.9585482
12	0.9717568	0.9642798
22	0.9574043	0.9461222

Accuracy was used to select the optimal model using the largest value.

The final value used for the model was mtry = 12.



#Vector de predicciones

```
pred <- predict(model_rf,testing)
conf_matr <- confusionMatrix(pred,testing$classe)
```

> conf_matr

```
Confusion Matrix and Statistics
          Reference
Prediction   A   B   C   D   E
         A 826   9   1   0   1
         B   7 549   3   1   1
         C   2   7 506  23   0
         D   1   3   3 456   3
         E   1   1   0   2 536
```

Overall Statistics

```
               Accuracy : 0.9765
                 95% CI : (0.9704, 0.9817)
    No Information Rate : 0.2845
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9703

 Mcnemar's Test P-Value : NA
```

Statistics by Class:

```
                     Class: A Class: B Class: C Class: D Class: E
Sensitivity            0.9869   0.9649   0.9864   0.9461   0.9908
Specificity            0.9948   0.9949   0.9868   0.9959   0.9983
Pos Pred Value         0.9869   0.9786   0.9405   0.9785   0.9926
Neg Pred Value         0.9948   0.9916   0.9971   0.9895   0.9979
Prevalence             0.2845   0.1934   0.1744   0.1638   0.1839
Detection Rate         0.2808   0.1866   0.1720   0.1550   0.1822
```

Detection Prevalence	0.2845	0.1907	0.1829	0.1584	0.1835
Balanced Accuracy	0.9908	0.9799	0.9866	0.9710	0.9945

Conclusion:

Como resultado de aplicar el algoritmo de Random Forest obtenemos una muy buena performance con un Accuracy : 0.9765 y un valor Kappa : 0.9703 ,los cuales indican que el modelo etiqueta correctamente y clasifica con un alto desempeño, también podemos ver que el parámetro optimizable mtry, indica el número máximo de variables en el modelo creado, alcanza un valor optimo en 12, caret estima este valor automáticamente.

Finalmente tomamos una muestra aleatoria de 20 valores del vector de predicciones:

predicciones <- sample(x=pred,size = 20)