

# 15.- Feature Importance\_04\_19\_vacunacion\_completo\_v\_01

June 8, 2023

#

CU04\_Optimización de vacunas

Citizenlab Data Science Methodology > III - Feature Engineering Domain \*\*\* > # 15.- Feature Importance

Feature Importance is the process that assigns scores to the input characteristics to a model, which indicate the relative importance of each characteristic, in order, for example, to be able to select the most important ones.

## 0.1 Tasks

Perform Feature importance from model coefficients

- Linear Regression Feature importance
- Logistic Regression Feature importance

Perform Feature importance from Decision Tree

- CART Feature Importance
- Random Forest Regression Feature Importance

Perform Feature importance from Permutation testing

Evaluate a Logistic Regression model with feature selection

## 0.2 Consideraciones casos CitizenLab programados en R

- Algunas de las tareas de este proceso se han realizado en los notebooks del proceso 05 Data Collection porque eran necesarias para las tareas ETL. En esos casos, en este notebook se referencia al notebook del proceso 05 correspondiente
- Otras tareas típicas de este proceso se realizan en los notebooks del dominio IV al ser más eficiente realizarlas en el propio pipeline de modelización.
- Por tanto en los notebooks de este proceso de manera general se incluyen las comprobaciones necesarias, y comentarios si procede
- Las tareas del proceso se van a aplicar solo a los archivos que forman parte del despliegue, ya que hay muchos archivos intermedios que no procede pasar por este proceso
- El nombre de archivo del notebook hace referencia al nombre de archivo del proceso 05 al que se aplica este proceso, por eso pueden no ser correlativa la numeración
- Las comprobaciones se van a realizar teniendo en cuenta que el lenguaje utilizado en el despliegue de este caso es R

## 0.3 File

- Input File: CU\_04\_08\_20\_vacunacion\_gripe\_train\_and\_test.csv
- Output File: No aplica

### 0.3.1 Encoding

Con la siguiente expresión se evitan problemas con el encoding al ejecutar el notebook. Es posible que deba ser eliminada o adaptada a la máquina en la que se ejecute el código.

```
[ ]: Sys.setlocale(category = "LC_ALL", locale = "es_ES.UTF-8")
```

## 0.4 Settings

### 0.4.1 Libraries to use

```
[1]: library(readr)
library(dplyr)
library(tidyr)
library(forcats)
library(lubridate)
```

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

filter, lag

The following objects are masked from ‘package:base’:

intersect, setdiff, setequal, union

Attaching package: ‘lubridate’

The following objects are masked from ‘package:base’:

date, intersect, setdiff, union

### 0.4.2 Paths

```
[2]: iPath <- "Data/Input/"  
     oPath <- "Data/Output/"
```

## 0.5 Data Load

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Uncomment the line if using this option

```
[3]: # file_data <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

```
[4]: iFile <- "CU_04_08_20_vacunacion_gripe_train_and_test.csv"  
     file_data <- paste0(iPath, iFile)  
  
     if(file.exists(file_data)){  
       cat("Se leerán datos del archivo: ", file_data)  
     } else{  
       warning("Cuidado: el archivo no existe.")  
     }
```

Se leerán datos del archivo:

Data/Input/CU\_04\_08\_20\_vacunacion\_gripe\_train\_and\_test.csv

**Data file to dataframe** Usar la función adecuada según el formato de entrada (xlsx, csv, json, ...)

```
[5]: data <- read_csv(file_data)
```

Rows: 21736 Columns: 49  
Column specification

Delimiter: ","

```
chr (3): GEOCODIGO, DESBDT, nombre_zona  
dbl (45): ano, semana, n_vacunas, n_citas, tmed, prec, velmedia,  
presMax, be...  
lgl (1): is_train
```

Use `spec()` to retrieve the full column specification for this data.

Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

Estructura de los datos:

```
[6]: data |> glimpse()
```

```
Rows: 21,736
Columns: 49
$ GEOCODIGO      <chr> "259", "260", "041", "025", "046",
"159", "065", "09...
$ DESBDT         <chr> "V Centenario", "Valdeacederas",
"Canillejas", "Bara...
$ ano            <dbl> 2022, 2022, 2022, 2022, 2022, 2022,
2022, 2021, 2023...
$ semana         <dbl> 34, 8, 9, 49, 24, 3, 8, 47, 1, 2,
52, 39, 16, 50, 34...
$ n_vacunas      <dbl> 0, 0, 0, 292, 0, 524, 0, 248, 204,
205, NA, 0, 0, 51...
$ n_citas        <dbl> 0, 0, 0, 280, 0, 498, 0, 228, 198,
187, NA, 0, 0, 51...
$ tmed           <dbl> 27.278748, 9.577289, 8.536554,
9.065363, 29.905728, ...
$ prec           <dbl> 0.169955881, 1.264910043,
3.122881160, 7.313886680, ...
$ velmedia       <dbl> 2.297067, 1.890425, 2.418071,
1.562328, 2.564749, 1...
$ presMax        <dbl> 940.0420, 944.1770, 949.7179,
941.8342, 940.5669, 95...
$ benzene        <dbl> 0.1764413, 0.4591543, 0.4099159,
0.4224172, 0.195865...
$ co             <dbl> 0.4987735, 0.3960647, 0.3951587,
NA, 0.2891224, 0.50...
$ no             <dbl> NA, 6.611337, 9.331224, 14.007722,
4.063517, 24.4756...
$ no2            <dbl> 14.21113, 34.67671, 30.29999,
32.54832, 26.06913, 44...
$ nox            <dbl> 18.00109, 48.94660, 45.22346,
56.75574, 30.35311, 74...
$ o3             <dbl> 80.90659, 42.06663, 48.88088,
26.68276, 64.55205, 31...
$ pm10           <dbl> 20.117087, 15.042152, 14.002432,
18.032354, 55.79346...
$ pm2.5          <dbl> 10.628064, 5.539590, 7.124192,
6.793868, 19.520373, ...
$ so2            <dbl> 2.794934, 3.507164, 2.692125,
2.351139, 3.397640, 2...
$ campana        <dbl> NA, NA, NA, 2022, NA, 2021, NA,
2021, 2022, 2021, 20...
$ scampana       <dbl> NA, NA, NA, 14, NA, 20, NA, 12, 18,
19, 17, 4, NA, 1...
$ capacidad_zona <dbl> 7957, 6537, 7167, 5633, 3864,
```

12583, 8544, 5077, 494...  
 \$ prop\_riesgo <dbl> 0.11393237, 0.15763986, 0.25500690,  
 0.14452370, 0.26...  
 \$ tasa\_riesgo <dbl> 0.013477754, 0.015731142,  
 0.009177382, 0.013099129, ...  
 \$ tasa\_mayores <dbl> 0.023033610, 0.032817374,  
 0.028147027, 0.020829657, ...  
 \$ poblacion\_mayores <dbl> 0.10330662, 0.14362062, 0.23161874,  
 0.13058449, 0.24...  
 \$ nombre\_zona <chr> "V Centenario", "Valdeacederas",  
 "Canillejas", "Bara...  
 \$ nsec <dbl> 17, 18, 22, 13, 14, 42, 32, 13, 17,  
 11, NA, 15, 15, ...  
 \$ t3\_1 <dbl> 36.73039, 41.41412, 45.44882,  
 39.78001, 46.13171, 46...  
 \$ t1\_1 <dbl> 31778, 26202, 28658, 22492, 15450,  
 50478, 34148, 202...  
 \$ t2\_1 <dbl> 0.5084658, 0.5329728, 0.5316594,  
 0.5189021, 0.551191...  
 \$ t2\_2 <dbl> 0.4915342, 0.4670272, 0.4683406,  
 0.4810979, 0.448809...  
 \$ t4\_1 <dbl> 0.22551283, 0.12790298, 0.12603707,  
 0.18104432, 0.11...  
 \$ t4\_2 <dbl> 0.6711962, 0.7284970, 0.6423306,  
 0.6883785, 0.641173...  
 \$ t4\_3 <dbl> 0.10330662, 0.14362062, 0.23161874,  
 0.13058449, 0.24...  
 \$ t5\_1 <dbl> 0.1063332, 0.2295250, 0.1655070,  
 0.1266086, 0.165893...  
 \$ t6\_1 <dbl> 0.1706875, 0.3477631, 0.2511757,  
 0.1998911, 0.261480...  
 \$ t7\_1 <dbl> 0.05131106, 0.04606911, 0.04379644,  
 0.05585777, 0.06...  
 \$ t8\_1 <dbl> 0.03892836, 0.03586418, 0.03207779,  
 0.04434976, 0.05...  
 \$ t9\_1 <dbl> 0.5151383, 0.3863876, 0.3129631,  
 0.4611972, 0.701812...  
 \$ t10\_1 <dbl> 0.09258503, 0.13151901, 0.13926119,  
 0.10460043, 0.06...  
 \$ t11\_1 <dbl> 0.6406787, 0.5451465, 0.4600730,  
 0.5920292, 0.471769...  
 \$ t12\_1 <dbl> 0.7028586, 0.6277335, 0.5346482,  
 0.6590530, 0.502531...  
 \$ area <dbl> 2100118.9, 1164622.0, 1597474.5,  
 3816572.0, 870986.8...  
 \$ densidad\_hab\_km <dbl> 15131.52443, 22498.28643,  
 17939.56640, 5893.24662, 1...  
 \$ tuits\_gripe <dbl> 60, 56, 72, 196, 46, 382, 56, 280,

```

24, 508, NA, 126,...
$ interes_gripe    <dbl> 24, 15, 24, 77, 21, 42, 15, 64, 64,
69, NA, 42, 40, ...
$ Target          <dbl> 24, 15, 24, 77, 21, 42, 15, 64, 64,
69, NA, 42, 40, ...
$ is_train         <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE,
TRUE, TRUE, TRUE...

```

Muestra de los primeros datos:

```
[7]: data |> slice_head(n = 5)
```

	GEOCODIGO	DESBDT	ano	semana	n_vacunas	n_citas	tmed	
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
	259	V Centenario	2022	34	0	0	27.278748	0
A spec_tbl_df: 5 × 9	260	Valdeacederas	2022	8	0	0	9.577289	1
	041	Canillejas	2022	9	0	0	8.536554	3
	025	Barajas	2022	49	292	280	9.065363	7
	046	Castelló	2022	24	0	0	29.905728	0

## 0.6 Feature importance from model coefficients

```
[8]: data <- data |>
      na.omit()
```

### 0.6.1 Linear Regression Feature importance

```

[9]: # install the necessary packages if not already installed
if (!require(caret)) {
  install.packages('caret')
}

# load the necessary packages
library(caret)

# create the model
set.seed(123)
train_control <- trainControl(method = "cv", number = 5)
linear_model <- train(Target ~ ., data = data, method = "lm",
                      trControl = train_control)

# extract the coefficients
coef <- summary(linear_model$finalModel)$coefficients

# make a data frame of feature names and their corresponding coefficients
feature_importance <- data.frame(Feature = rownames(coef), Coefficient = coef[,1],
                                  ↪ "Estimate")

```

```

# compute the absolute value of the coefficients
feature_importance$AbsCoefficient <- abs(feature_importance$Coefficient)

# sort the features by the absolute value of their coefficients in descending
  ↪order
feature_importance <-
  ↪feature_importance[order(-feature_importance$AbsCoefficient), ]

# print the feature importance
print(feature_importance)

```

Loading required package: caret

Loading required package: ggplot2

Loading required package: lattice

Warning message in predict.lm(modelFit, newdata):  
 "prediction from a rank-deficient fit may be misleading"  
 Warning message in predict.lm(modelFit, newdata):  
 "prediction from a rank-deficient fit may be misleading"  
 Warning message in predict.lm(modelFit, newdata):  
 "prediction from a rank-deficient fit may be misleading"  
 Warning message in predict.lm(modelFit, newdata):  
 "prediction from a rank-deficient fit may be misleading"  
 Warning message in predict.lm(modelFit, newdata):  
 "prediction from a rank-deficient fit may be misleading"  
 Warning message in summary.lm(linear\_model\$finalModel):  
 "essentially perfect fit: summary may be unreliable"

	Feature	Coefficient	AbsCoefficient
interes_gripe	interes_gripe	1.000000e+00	1.000000e+00
(Intercept)	(Intercept)	-4.412119e-11	4.412119e-11
campana	campana	2.933503e-13	2.933503e-13
ano	ano	-2.711501e-13	2.711501e-13
GEOCODIGO033	GEOCODIGO033	6.115202e-14	6.115202e-14
GEOCODIGO256	GEOCODIGO256	6.094220e-14	6.094220e-14
GEOCODIGO279	GEOCODIGO279	5.639716e-14	5.639716e-14
GEOCODIGO201	GEOCODIGO201	5.551830e-14	5.551830e-14
GEOCODIGO036	GEOCODIGO036	5.539307e-14	5.539307e-14
GEOCODIGO228	GEOCODIGO228	-5.041923e-14	5.041923e-14
GEOCODIGO085	GEOCODIGO085	-4.936885e-14	4.936885e-14
GEOCODIGO275	GEOCODIGO275	4.741706e-14	4.741706e-14
GEOCODIGO241	GEOCODIGO241	4.732671e-14	4.732671e-14
GEOCODIGO049	GEOCODIGO049	-4.530675e-14	4.530675e-14
GEOCODIGO097	GEOCODIGO097	-4.495278e-14	4.495278e-14
GEOCODIGO184	GEOCODIGO184	4.483504e-14	4.483504e-14
GEOCODIGO118	GEOCODIGO118	4.353739e-14	4.353739e-14

GEOCODIG0121	GEOCODIG0121	4.323652e-14	4.323652e-14
GEOCODIG0056	GEOCODIG0056	-3.994818e-14	3.994818e-14
GEOCODIG0055	GEOCODIG0055	-3.934225e-14	3.934225e-14
GEOCODIG0038	GEOCODIG0038	3.822105e-14	3.822105e-14
GEOCODIG0030	GEOCODIG0030	-3.634783e-14	3.634783e-14
GEOCODIG0108	GEOCODIG0108	-3.589486e-14	3.589486e-14
GEOCODIG0059	GEOCODIG0059	3.552827e-14	3.552827e-14
GEOCODIG0180	GEOCODIG0180	3.552245e-14	3.552245e-14
GEOCODIG0105	GEOCODIG0105	-3.504909e-14	3.504909e-14
GEOCODIG0182	GEOCODIG0182	3.498939e-14	3.498939e-14
GEOCODIG0013	GEOCODIG0013	-3.438680e-14	3.438680e-14
GEOCODIG0138	GEOCODIG0138	3.384073e-14	3.384073e-14
GEOCODIG0054	GEOCODIG0054	-3.218474e-14	3.218474e-14
GEOCODIG0080	GEOCODIG0080	3.209474e-14	3.209474e-14
GEOCODIG0230	GEOCODIG0230	3.081440e-14	3.081440e-14
GEOCODIG0215	GEOCODIG0215	-3.038346e-14	3.038346e-14
GEOCODIG0282	GEOCODIG0282	3.024473e-14	3.024473e-14
GEOCODIG0155	GEOCODIG0155	-2.983943e-14	2.983943e-14
GEOCODIG0250	GEOCODIG0250	2.953767e-14	2.953767e-14
GEOCODIG0176	GEOCODIG0176	2.951762e-14	2.951762e-14
GEOCODIG0257	GEOCODIG0257	-2.906245e-14	2.906245e-14
GEOCODIG0168	GEOCODIG0168	-2.905828e-14	2.905828e-14
GEOCODIG0159	GEOCODIG0159	-2.826553e-14	2.826553e-14
GEOCODIG0156	GEOCODIG0156	2.738223e-14	2.738223e-14
GEOCODIG0095	GEOCODIG0095	2.694826e-14	2.694826e-14
GEOCODIG0177	GEOCODIG0177	-2.656150e-14	2.656150e-14
GEOCODIG0115	GEOCODIG0115	2.647622e-14	2.647622e-14
GEOCODIG0120	GEOCODIG0120	2.637973e-14	2.637973e-14
GEOCODIG0064	GEOCODIG0064	2.626559e-14	2.626559e-14
GEOCODIG0251	GEOCODIG0251	-2.616292e-14	2.616292e-14
GEOCODIG0211	GEOCODIG0211	2.600648e-14	2.600648e-14
GEOCODIG0196	GEOCODIG0196	2.541676e-14	2.541676e-14
GEOCODIG0283	GEOCODIG0283	2.516373e-14	2.516373e-14
GEOCODIG0045	GEOCODIG0045	2.457623e-14	2.457623e-14
GEOCODIG0174	GEOCODIG0174	2.385072e-14	2.385072e-14
GEOCODIG0140	GEOCODIG0140	-2.361619e-14	2.361619e-14
GEOCODIG0190	GEOCODIG0190	2.334264e-14	2.334264e-14
GEOCODIG0217	GEOCODIG0217	-2.317229e-14	2.317229e-14
GEOCODIG0017	GEOCODIG0017	2.292853e-14	2.292853e-14
GEOCODIG0046	GEOCODIG0046	2.284804e-14	2.284804e-14
GEOCODIG0181	GEOCODIG0181	-2.277713e-14	2.277713e-14
GEOCODIG0028	GEOCODIG0028	2.236327e-14	2.236327e-14
GEOCODIG0243	GEOCODIG0243	-2.200368e-14	2.200368e-14
GEOCODIG0179	GEOCODIG0179	-2.200012e-14	2.200012e-14
GEOCODIG0161	GEOCODIG0161	2.195473e-14	2.195473e-14
GEOCODIG0144	GEOCODIG0144	2.180516e-14	2.180516e-14
GEOCODIG0048	GEOCODIG0048	2.125090e-14	2.125090e-14
GEOCODIG0222	GEOCODIG0222	2.094436e-14	2.094436e-14



GEOCODIG0089	GEOCODIG0089	2.089226e-14	2.089226e-14
GEOCODIG0219	GEOCODIG0219	2.068343e-14	2.068343e-14
GEOCODIG0258	GEOCODIG0258	2.066934e-14	2.066934e-14
GEOCODIG0267	GEOCODIG0267	2.040372e-14	2.040372e-14
GEOCODIG0160	GEOCODIG0160	2.025745e-14	2.025745e-14
GEOCODIG0096	GEOCODIG0096	2.018002e-14	2.018002e-14
GEOCODIG0212	GEOCODIG0212	-2.011654e-14	2.011654e-14
GEOCODIG0132	GEOCODIG0132	-1.988772e-14	1.988772e-14
GEOCODIG0101	GEOCODIG0101	1.965666e-14	1.965666e-14
GEOCODIG0066	GEOCODIG0066	1.959239e-14	1.959239e-14
GEOCODIG0007	GEOCODIG0007	1.958985e-14	1.958985e-14
GEOCODIG0139	GEOCODIG0139	-1.944571e-14	1.944571e-14
GEOCODIG0135	GEOCODIG0135	-1.916330e-14	1.916330e-14
GEOCODIG0218	GEOCODIG0218	1.896370e-14	1.896370e-14
GEOCODIG0238	GEOCODIG0238	1.874554e-14	1.874554e-14
GEOCODIG0124	GEOCODIG0124	1.861895e-14	1.861895e-14
GEOCODIG0002	GEOCODIG0002	1.853664e-14	1.853664e-14
GEOCODIG0122	GEOCODIG0122	1.851806e-14	1.851806e-14
GEOCODIG0110	GEOCODIG0110	1.849172e-14	1.849172e-14
GEOCODIG0216	GEOCODIG0216	1.815378e-14	1.815378e-14
GEOCODIG0154	GEOCODIG0154	-1.814361e-14	1.814361e-14
GEOCODIG0166	GEOCODIG0166	1.796963e-14	1.796963e-14
GEOCODIG0051	GEOCODIG0051	1.755906e-14	1.755906e-14
GEOCODIG0070	GEOCODIG0070	-1.738501e-14	1.738501e-14
GEOCODIG0015	GEOCODIG0015	1.724818e-14	1.724818e-14
GEOCODIG0242	GEOCODIG0242	1.722469e-14	1.722469e-14
GEOCODIG0014	GEOCODIG0014	-1.715465e-14	1.715465e-14
GEOCODIG0034	GEOCODIG0034	1.704445e-14	1.704445e-14
GEOCODIG0072	GEOCODIG0072	1.699151e-14	1.699151e-14
GEOCODIG0247	GEOCODIG0247	1.693972e-14	1.693972e-14
GEOCODIG0037	GEOCODIG0037	1.679426e-14	1.679426e-14
GEOCODIG0126	GEOCODIG0126	-1.661739e-14	1.661739e-14
GEOCODIG0094	GEOCODIG0094	1.646378e-14	1.646378e-14
GEOCODIG0157	GEOCODIG0157	-1.638246e-14	1.638246e-14
GEOCODIG0252	GEOCODIG0252	-1.590439e-14	1.590439e-14
GEOCODIG0188	GEOCODIG0188	1.585290e-14	1.585290e-14
GEOCODIG0075	GEOCODIG0075	1.574201e-14	1.574201e-14
GEOCODIG0004	GEOCODIG0004	1.558974e-14	1.558974e-14
GEOCODIG0206	GEOCODIG0206	1.558060e-14	1.558060e-14
GEOCODIG0162	GEOCODIG0162	1.556950e-14	1.556950e-14
GEOCODIG0114	GEOCODIG0114	1.534182e-14	1.534182e-14
GEOCODIG0104	GEOCODIG0104	1.502235e-14	1.502235e-14
GEOCODIG0023	GEOCODIG0023	1.476073e-14	1.476073e-14
GEOCODIG0203	GEOCODIG0203	1.470438e-14	1.470438e-14
GEOCODIG0021	GEOCODIG0021	1.441770e-14	1.441770e-14
GEOCODIG0225	GEOCODIG0225	1.421428e-14	1.421428e-14
GEOCODIG0187	GEOCODIG0187	1.420487e-14	1.420487e-14
GEOCODIG0199	GEOCODIG0199	1.398843e-14	1.398843e-14

GEOCODIG0185	GEOCODIG0185	1.397912e-14	1.397912e-14
GEOCODIG0073	GEOCODIG0073	-1.373033e-14	1.373033e-14
GEOCODIG0009	GEOCODIG0009	1.370698e-14	1.370698e-14
GEOCODIG0248	GEOCODIG0248	1.370470e-14	1.370470e-14
GEOCODIG0192	GEOCODIG0192	1.368791e-14	1.368791e-14
GEOCODIG0129	GEOCODIG0129	1.367978e-14	1.367978e-14
GEOCODIG0194	GEOCODIG0194	-1.348813e-14	1.348813e-14
GEOCODIG0043	GEOCODIG0043	1.338216e-14	1.338216e-14
GEOCODIG0172	GEOCODIG0172	1.322763e-14	1.322763e-14
GEOCODIG0164	GEOCODIG0164	1.322676e-14	1.322676e-14
GEOCODIG0082	GEOCODIG0082	1.321655e-14	1.321655e-14
GEOCODIG0137	GEOCODIG0137	1.293705e-14	1.293705e-14
GEOCODIG0204	GEOCODIG0204	1.284380e-14	1.284380e-14
GEOCODIG0163	GEOCODIG0163	1.280443e-14	1.280443e-14
GEOCODIG0200	GEOCODIG0200	1.270519e-14	1.270519e-14
GEOCODIG0268	GEOCODIG0268	1.267619e-14	1.267619e-14
GEOCODIG0060	GEOCODIG0060	1.259206e-14	1.259206e-14
GEOCODIG0276	GEOCODIG0276	1.231351e-14	1.231351e-14
GEOCODIG0240	GEOCODIG0240	-1.227139e-14	1.227139e-14
GEOCODIG0074	GEOCODIG0074	-1.224353e-14	1.224353e-14
GEOCODIG0175	GEOCODIG0175	1.223487e-14	1.223487e-14
GEOCODIG0029	GEOCODIG0029	1.216554e-14	1.216554e-14
GEOCODIG0109	GEOCODIG0109	1.211238e-14	1.211238e-14
GEOCODIG0209	GEOCODIG0209	1.202868e-14	1.202868e-14
GEOCODIG0141	GEOCODIG0141	1.197890e-14	1.197890e-14
GEOCODIG0271	GEOCODIG0271	1.191575e-14	1.191575e-14
GEOCODIG0285	GEOCODIG0285	-1.182411e-14	1.182411e-14
so2	so2	-1.166616e-14	1.166616e-14
GEOCODIG0128	GEOCODIG0128	-1.165566e-14	1.165566e-14
GEOCODIG0170	GEOCODIG0170	1.152202e-14	1.152202e-14
GEOCODIG0053	GEOCODIG0053	1.131479e-14	1.131479e-14
GEOCODIG0058	GEOCODIG0058	1.114406e-14	1.114406e-14
GEOCODIG0260	GEOCODIG0260	1.096997e-14	1.096997e-14
GEOCODIG0125	GEOCODIG0125	1.090624e-14	1.090624e-14
GEOCODIG0117	GEOCODIG0117	1.064512e-14	1.064512e-14
GEOCODIG0231	GEOCODIG0231	1.049574e-14	1.049574e-14
GEOCODIG0178	GEOCODIG0178	1.044514e-14	1.044514e-14
GEOCODIG0261	GEOCODIG0261	1.033291e-14	1.033291e-14
GEOCODIG0266	GEOCODIG0266	-1.026916e-14	1.026916e-14
GEOCODIG0112	GEOCODIG0112	1.014021e-14	1.014021e-14
GEOCODIG0229	GEOCODIG0229	1.013749e-14	1.013749e-14
GEOCODIG0236	GEOCODIG0236	1.010615e-14	1.010615e-14
GEOCODIG0022	GEOCODIG0022	-1.008301e-14	1.008301e-14
GEOCODIG0153	GEOCODIG0153	-1.008064e-14	1.008064e-14
GEOCODIG0061	GEOCODIG0061	1.004014e-14	1.004014e-14
GEOCODIG0151	GEOCODIG0151	9.974774e-15	9.974774e-15
GEOCODIG0039	GEOCODIG0039	9.524276e-15	9.524276e-15
GEOCODIG0193	GEOCODIG0193	9.272592e-15	9.272592e-15

GEOCODIG0084	GEOCODIG0084	9.267628e-15	9.267628e-15
GEOCODIG0173	GEOCODIG0173	9.242158e-15	9.242158e-15
GEOCODIG0205	GEOCODIG0205	-9.209576e-15	9.209576e-15
GEOCODIG0213	GEOCODIG0213	9.147095e-15	9.147095e-15
GEOCODIG0010	GEOCODIG0010	8.974972e-15	8.974972e-15
GEOCODIG0098	GEOCODIG0098	8.855665e-15	8.855665e-15
GEOCODIG0202	GEOCODIG0202	8.842218e-15	8.842218e-15
GEOCODIG0264	GEOCODIG0264	-8.738713e-15	8.738713e-15
GEOCODIG0148	GEOCODIG0148	8.733031e-15	8.733031e-15
GEOCODIG0068	GEOCODIG0068	-8.709746e-15	8.709746e-15
GEOCODIG0134	GEOCODIG0134	8.703883e-15	8.703883e-15
GEOCODIG0277	GEOCODIG0277	-8.537276e-15	8.537276e-15
GEOCODIG0133	GEOCODIG0133	8.450886e-15	8.450886e-15
GEOCODIG0255	GEOCODIG0255	8.408601e-15	8.408601e-15
GEOCODIG0093	GEOCODIG0093	-8.306698e-15	8.306698e-15
GEOCODIG0102	GEOCODIG0102	8.294903e-15	8.294903e-15
GEOCODIG0197	GEOCODIG0197	-8.162224e-15	8.162224e-15
GEOCODIG0239	GEOCODIG0239	8.132926e-15	8.132926e-15
GEOCODIG0183	GEOCODIG0183	-8.069563e-15	8.069563e-15
GEOCODIG0149	GEOCODIG0149	8.042529e-15	8.042529e-15
GEOCODIG0091	GEOCODIG0091	7.974849e-15	7.974849e-15
GEOCODIG0127	GEOCODIG0127	7.969956e-15	7.969956e-15
GEOCODIG0065	GEOCODIG0065	-7.687629e-15	7.687629e-15
GEOCODIG0263	GEOCODIG0263	7.687085e-15	7.687085e-15
GEOCODIG0143	GEOCODIG0143	-7.679146e-15	7.679146e-15
GEOCODIG0186	GEOCODIG0186	7.529804e-15	7.529804e-15
velmedia	velmedia	7.468246e-15	7.468246e-15
GEOCODIG0052	GEOCODIG0052	7.352651e-15	7.352651e-15
GEOCODIG0142	GEOCODIG0142	7.103378e-15	7.103378e-15
GEOCODIG0111	GEOCODIG0111	7.032409e-15	7.032409e-15
GEOCODIG0130	GEOCODIG0130	7.019133e-15	7.019133e-15
GEOCODIG0107	GEOCODIG0107	-6.863170e-15	6.863170e-15
GEOCODIG0189	GEOCODIG0189	6.854650e-15	6.854650e-15
GEOCODIG0273	GEOCODIG0273	6.722653e-15	6.722653e-15
GEOCODIG0005	GEOCODIG0005	-6.649262e-15	6.649262e-15
GEOCODIG0006	GEOCODIG0006	6.641116e-15	6.641116e-15
GEOCODIG0237	GEOCODIG0237	-6.537978e-15	6.537978e-15
GEOCODIG0092	GEOCODIG0092	6.421496e-15	6.421496e-15
GEOCODIG0265	GEOCODIG0265	-6.361985e-15	6.361985e-15
GEOCODIG0083	GEOCODIG0083	6.357894e-15	6.357894e-15
GEOCODIG0026	GEOCODIG0026	6.325071e-15	6.325071e-15
GEOCODIG0214	GEOCODIG0214	6.277091e-15	6.277091e-15
GEOCODIG0220	GEOCODIG0220	6.097355e-15	6.097355e-15
GEOCODIG0227	GEOCODIG0227	-6.075078e-15	6.075078e-15
GEOCODIG0076	GEOCODIG0076	6.040635e-15	6.040635e-15
GEOCODIG0146	GEOCODIG0146	-6.036466e-15	6.036466e-15
GEOCODIG0041	GEOCODIG0041	-5.967446e-15	5.967446e-15
GEOCODIG0234	GEOCODIG0234	5.957902e-15	5.957902e-15

GEOCODIGO284	GEOCODIGO284	5.835275e-15	5.835275e-15
GEOCODIGO119	GEOCODIGO119	-5.826391e-15	5.826391e-15
GEOCODIGO062	GEOCODIGO062	5.793850e-15	5.793850e-15
GEOCODIGO272	GEOCODIGO272	5.768298e-15	5.768298e-15
GEOCODIGO077	GEOCODIGO077	-5.705706e-15	5.705706e-15
GEOCODIGO208	GEOCODIGO208	5.653040e-15	5.653040e-15
tmed	tmed	-5.645617e-15	5.645617e-15
GEOCODIGO008	GEOCODIGO008	5.512154e-15	5.512154e-15
GEOCODIGO087	GEOCODIGO087	-5.451613e-15	5.451613e-15
GEOCODIGO027	GEOCODIGO027	5.430499e-15	5.430499e-15
GEOCODIGO012	GEOCODIGO012	5.367830e-15	5.367830e-15
GEOCODIGO207	GEOCODIGO207	5.347509e-15	5.347509e-15
GEOCODIGO032	GEOCODIGO032	5.326909e-15	5.326909e-15
GEOCODIGO152	GEOCODIGO152	-5.246170e-15	5.246170e-15
semana	semana	-5.181520e-15	5.181520e-15
GEOCODIGO088	GEOCODIGO088	-5.173313e-15	5.173313e-15
GEOCODIGO278	GEOCODIGO278	5.169027e-15	5.169027e-15
GEOCODIGO167	GEOCODIGO167	5.046559e-15	5.046559e-15
GEOCODIGO270	GEOCODIGO270	5.018251e-15	5.018251e-15
GEOCODIGO221	GEOCODIGO221	4.927506e-15	4.927506e-15
GEOCODIGO281	GEOCODIGO281	-4.825751e-15	4.825751e-15
GEOCODIGO103	GEOCODIGO103	4.822206e-15	4.822206e-15
GEOCODIGO249	GEOCODIGO249	-4.728840e-15	4.728840e-15
GEOCODIGO100	GEOCODIGO100	4.709529e-15	4.709529e-15
GEOCODIGO078	GEOCODIGO078	4.577270e-15	4.577270e-15
GEOCODIGO044	GEOCODIGO044	4.441032e-15	4.441032e-15
GEOCODIGO246	GEOCODIGO246	4.286484e-15	4.286484e-15
GEOCODIGO280	GEOCODIGO280	4.184614e-15	4.184614e-15
prec	prec	3.865762e-15	3.865762e-15
pm10	pm10	3.848193e-15	3.848193e-15
GEOCODIGO198	GEOCODIGO198	-3.678668e-15	3.678668e-15
GEOCODIGO223	GEOCODIGO223	3.609401e-15	3.609401e-15
GEOCODIGO147	GEOCODIGO147	-3.601896e-15	3.601896e-15
benzene	benzene	-3.495366e-15	3.495366e-15
GEOCODIGO020	GEOCODIGO020	3.419734e-15	3.419734e-15
GEOCODIGO210	GEOCODIGO210	3.418016e-15	3.418016e-15
GEOCODIGO090	GEOCODIGO090	3.413952e-15	3.413952e-15
GEOCODIGO063	GEOCODIGO063	3.330066e-15	3.330066e-15
GEOCODIGO099	GEOCODIGO099	3.269780e-15	3.269780e-15
GEOCODIGO269	GEOCODIGO269	-3.027114e-15	3.027114e-15
GEOCODIGO233	GEOCODIGO233	2.980871e-15	2.980871e-15
GEOCODIGO131	GEOCODIGO131	-2.957958e-15	2.957958e-15
GEOCODIGO274	GEOCODIGO274	-2.877303e-15	2.877303e-15
GEOCODIGO081	GEOCODIGO081	2.625062e-15	2.625062e-15
GEOCODIGO019	GEOCODIGO019	-2.600442e-15	2.600442e-15
GEOCODIGO067	GEOCODIGO067	-2.473551e-15	2.473551e-15
GEOCODIGO003	GEOCODIGO003	2.463516e-15	2.463516e-15
GEOCODIGO145	GEOCODIGO145	-2.426997e-15	2.426997e-15

GEOCODIGO235	GEOCODIGO235	2.402835e-15	2.402835e-15
GEOCODIGO158	GEOCODIGO158	2.252205e-15	2.252205e-15
GEOCODIGO086	GEOCODIGO086	2.054451e-15	2.054451e-15
GEOCODIGO050	GEOCODIGO050	2.053475e-15	2.053475e-15
GEOCODIGO286	GEOCODIGO286	-1.908625e-15	1.908625e-15
o3	o3	-1.844939e-15	1.844939e-15
GEOCODIGO245	GEOCODIGO245	1.791983e-15	1.791983e-15
GEOCODIGO106	GEOCODIGO106	-1.535206e-15	1.535206e-15
GEOCODIGO195	GEOCODIGO195	-1.469323e-15	1.469323e-15
GEOCODIGO025	GEOCODIGO025	1.427787e-15	1.427787e-15
GEOCODIGO079	GEOCODIGO079	-1.326788e-15	1.326788e-15
GEOCODIGO262	GEOCODIGO262	1.302302e-15	1.302302e-15
GEOCODIGO224	GEOCODIGO224	-1.265762e-15	1.265762e-15
GEOCODIGO040	GEOCODIGO040	-1.257494e-15	1.257494e-15
GEOCODIGO042	GEOCODIGO042	-1.207480e-15	1.207480e-15
co	co	-1.120073e-15	1.120073e-15
GEOCODIGO011	GEOCODIGO011	-1.118055e-15	1.118055e-15
GEOCODIGO169	GEOCODIGO169	1.026639e-15	1.026639e-15
GEOCODIGO136	GEOCODIGO136	-9.893035e-16	9.893035e-16
GEOCODIGO259	GEOCODIGO259	-8.719533e-16	8.719533e-16
GEOCODIGO018	GEOCODIGO018	-8.710594e-16	8.710594e-16
GEOCODIGO035	GEOCODIGO035	-6.956178e-16	6.956178e-16
GEOCODIGO232	GEOCODIGO232	-5.687241e-16	5.687241e-16
GEOCODIGO191	GEOCODIGO191	-5.608885e-16	5.608885e-16
GEOCODIGO069	GEOCODIGO069	-5.375048e-16	5.375048e-16
GEOCODIGO253	GEOCODIGO253	5.323516e-16	5.323516e-16
no2	no2	-4.960132e-16	4.960132e-16
GEOCODIGO047	GEOCODIGO047	4.912500e-16	4.912500e-16
GEOCODIGO016	GEOCODIGO016	4.907245e-16	4.907245e-16
GEOCODIGO244	GEOCODIGO244	-4.305242e-16	4.305242e-16
GEOCODIGO254	GEOCODIGO254	3.896931e-16	3.896931e-16
presMax	presMax	-3.449334e-16	3.449334e-16
GEOCODIGO113	GEOCODIGO113	-3.331622e-16	3.331622e-16
is_trainTRUE	is_trainTRUE	3.101909e-16	3.101909e-16
GEOCODIGO024	GEOCODIGO024	2.857620e-16	2.857620e-16
GEOCODIGO116	GEOCODIGO116	-2.821347e-16	2.821347e-16
GEOCODIGO150	GEOCODIGO150	2.797148e-16	2.797148e-16
GEOCODIGO031	GEOCODIGO031	-2.019895e-16	2.019895e-16
n_vacunas	n_vacunas	1.761763e-16	1.761763e-16
tuits_gripe	tuits_gripe	1.528383e-16	1.528383e-16
GEOCODIGO057	GEOCODIGO057	-1.477188e-16	1.477188e-16
pm2.5	pm2.5	1.422193e-16	1.422193e-16
nox	nox	-5.214295e-17	5.214295e-17
no	no	-4.413159e-17	4.413159e-17
n_citas	n_citas	-1.964112e-17	1.964112e-17
GEOCODIGO071	GEOCODIGO071	1.214984e-17	1.214984e-17

## 0.6.2 Logistic Regression Feature importance

No aplica

## 0.7 Decision Tree Feature Importance

### 0.7.1 CART Regression Feature Importance

```
[10]: # install the necessary packages if not already installed
if (!require(rpart)) {
  install.packages('rpart')
}

# load the necessary packages
library(rpart)

# fit the model
cart_model <- rpart(Target ~ ., data = data, method = "anova")

# compute feature importances
importance <- as.data.frame(varImp(cart_model))

# sort by importance
importance <- importance[order(-importance$Overall), ]

# print the feature importance
print(importance)
```

Loading required package: rpart

```
[1] 3.1708648 1.4384476 1.2730982 1.1835086 0.5392957 0.5263859 0.4977876
[8] 0.4391151 0.3306283 0.2629820 0.1173294 0.0000000 0.0000000 0.0000000
[15] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
[22] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
[29] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
[36] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
[43] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
```

### 0.7.2 CART Classification Feature Importance

No aplica

### 0.7.3 Random Forest Regression Feature Importance

```
[18]: # install the necessary packages if not already installed
if (!require(randomForest)) {
  install.packages('randomForest')
}
```

```

# load the necessary packages
library(randomForest)

# fit the model
rf_model <- randomForest(Target ~ ., data = data, importance = TRUE)

# get feature importances
importance <- rf_model$importance

# make a data frame of feature names and their corresponding importances
feature_importance <- data.frame(Feature = rownames(importance), Importance =
  importance[,1]) # Adjust the index based on your requirement

# sort the features by importance in descending order
feature_importance <- feature_importance[order(-feature_importance$Importance),
  ]

# print the feature importance
print(feature_importance)

```

	Feature	Importance
interes_gripe	interes_gripe	665.041924825
tuits_gripe	tuits_gripe	148.284259879
scampana	scampana	146.581337214
semana	semana	90.136665620
n_vacunas	n_vacunas	61.363162786
n_citas	n_citas	58.783013919
prec	prec	29.412115470
benzene	benzene	12.936875246
tmed	tmed	11.263259346
so2	so2	8.412109795
pm10	pm10	7.820897289
ano	ano	6.614590753
pm2.5	pm2.5	6.428949390
campana	campana	5.185778031
o3	o3	3.662270853
no	no	3.129776410
velmedia	velmedia	2.484920743
nox	nox	1.642691242
no2	no2	1.194571534
co	co	0.696934189
GEOCODIGO	GEOCODIGO	0.317690358
DESBDT	DESBDT	0.265026645
nombre_zona	nombre_zona	0.223058360
prop_riesgo	prop_riesgo	0.173890498
t4_3	t4_3	0.149066106

densidad_hab_km	densidad_hab_km	0.145619155
poblacion_mayores	poblacion_mayores	0.133757698
t2_1	t2_1	0.080100193
area	area	0.067242660
t12_1	t12_1	0.065372391
presMax	presMax	0.063929068
nsec	nsec	0.063551397
capacidad_zona	capacidad_zona	0.053062868
t6_1	t6_1	0.049762605
t10_1	t10_1	0.044360141
t4_1	t4_1	0.043590444
t1_1	t1_1	0.042297268
t4_2	t4_2	0.037693132
t7_1	t7_1	0.034866311
t9_1	t9_1	0.032525707
t3_1	t3_1	0.022146521
t8_1	t8_1	0.015283011
t5_1	t5_1	0.010346115
t2_2	t2_2	0.003709971
t11_1	t11_1	-0.009720072
is_train	is_train	-0.014410352
tasa_mayores	tasa_mayores	-0.015156100
tasa_riesgo	tasa_riesgo	-0.105808303

```
[17]: rf_model$importance
```



	%IncMSE	IncNodePurity
GEOCODIGO	0.317690358	7.478272e-01
DESBDT	0.265026645	5.051162e-01
ano	6.614590753	1.931740e+04
semana	90.136665620	4.710209e+05
n_vacunas	61.363162786	2.002295e+05
n_citas	58.783013919	1.984159e+05
tmed	11.263259346	4.787210e+04
prec	29.412115470	1.168580e+05
velmedia	2.484920743	3.490770e+04
presMax	0.063929068	1.956558e+02
benzene	12.936875246	1.268288e+05
co	0.696934189	1.760577e+03
no	3.129776410	4.136604e+04
no2	1.194571534	6.065672e+03
nox	1.642691242	1.287343e+04
o3	3.662270853	2.339107e+04
pm10	7.820897289	5.183155e+04
pm2.5	6.428949390	4.858137e+04
so2	8.412109795	4.302407e+04
campana	5.185778031	1.414884e+04
scampana	146.581337214	9.183421e+05
capacidad_zona	0.053062868	4.113672e+00
prop_riesgo	0.173890498	3.210903e-01
tasa_riesgo	-0.105808303	2.329904e-01
tasa_mayores	-0.015156100	6.482611e-02
poblacion_mayores	0.133757698	2.669380e+00
nombre_zona	0.223058360	2.140744e+00
nsec	0.063551397	3.292341e+00
t3_1	0.022146521	4.105629e+00
t1_1	0.042297268	3.368476e+00
t2_1	0.080100193	4.973171e+01
t2_2	0.003709971	1.868699e+01
t4_1	0.043590444	1.547176e+01
t4_2	0.037693132	1.856270e+00
t4_3	0.149066106	2.377697e-01
t5_1	0.010346115	1.143387e+01
t6_1	0.049762605	1.600634e+00
t7_1	0.034866311	7.694122e+00
t8_1	0.015283011	8.255059e+00
t9_1	0.032525707	7.166661e+00
t10_1	0.044360141	3.504122e+00
t11_1	-0.009720072	9.065964e+00
t12_1	0.065372391	6.403215e-01
area	0.067242660	2.439165e+01
densidad_hab_km	0.145619155	6.314780e+01
tuits_gripe	148.284259879	9.254715e+05
interes_gripe	665.041924825	2.795582e+06
is_train	-0.014410352	9.835381e-01

A matrix: 48 × 2 of type dbl

## 0.7.4 Random Forest Classification Feature Importance

No aplica

## 0.8 Permutation Feature Importance

### 0.8.1 Permutation Feature Importance for Regression

```
[12]: # install the necessary packages if not already installed
if (!require(ranger)) {
  install.packages('ranger')
}

# load the necessary packages
library(ranger)

# fit the model
rf_model <- ranger(Target ~ ., data = data, importance = 'permutation')

# get feature importances
importance <- data.frame(Feature = names(rf_model$variable.importance),
  Importance = rf_model$variable.importance)

# sort the features by importance in descending order
importance <- importance[order(-importance$Importance), ]

# print the feature importance
print(importance)
```

Loading required package: ranger

Attaching package: 'ranger'

The following object is masked from 'package:randomForest':

importance

	Feature	Importance
interes_gripe	interes_gripe	3.589550e+02
tuits_gripe	tuits_gripe	1.777444e+02
scampana	scampana	1.741863e+02
semana	semana	1.599750e+02
n_vacunas	n_vacunas	1.032406e+02
n_citas	n_citas	9.647796e+01
tmed	tmed	5.065526e+01

prec		prec	4.564980e+01
benzene		benzene	3.644910e+01
ano		ano	2.478358e+01
so2		so2	2.366207e+01
pm10		pm10	2.044931e+01
o3		o3	1.952931e+01
pm2.5		pm2.5	1.813701e+01
no		no	1.749017e+01
campana		campana	1.630669e+01
velmedia		velmedia	1.574798e+01
no2		no2	1.089645e+01
nox		nox	8.764699e+00
co		co	3.307979e+00
presMax		presMax	1.835227e+00
area		area	4.717773e-01
densidad_hab_km	densidad_hab_km		4.399667e-01
t2_1		t2_1	3.219288e-01
t2_2		t2_2	3.056317e-01
t1_1		t1_1	1.566538e-01
t4_1		t4_1	1.097111e-01
capacidad_zona	capacidad_zona		1.039774e-01
poblacion_mayores	poblacion_mayores		9.624182e-02
t9_1		t9_1	9.154759e-02
t3_1		t3_1	6.410058e-02
t6_1		t6_1	5.434363e-02
t8_1		t8_1	5.348162e-02
t4_2		t4_2	4.152267e-02
t7_1		t7_1	3.342118e-02
t10_1		t10_1	2.916619e-02
nsec		nsec	2.334047e-02
t4_3		t4_3	1.915116e-02
t5_1		t5_1	1.833996e-02
GEOCODIGO	GEOCODIGO		1.495837e-02
prop_riesgo	prop_riesgo		1.435767e-02
t11_1		t11_1	1.432857e-02
t12_1		t12_1	5.761710e-03
tasa_riesgo	tasa_riesgo		4.829077e-03
tasa_mayores	tasa_mayores		2.380711e-03
nombre_zona	nombre_zona		4.976843e-04
is_train	is_train		-8.571006e-03
DESBDT	DESBDT		-1.420096e-02

## 0.8.2 Permutation Feature Importance for Classification

[ ]: No aplica

## 0.9 Evaluating a Logistic Regression model with feature selection.

### 0.9.1 Evaluating with all selected features.

[ ]:

### 0.9.2 Evaluating with feature selection performed using feature importance.

Select type of Feature Importance to use

[ ]:

Operation