

TP 1

Tecnologia Digital VI

Piñera - Saintnom - Dediego

2023-08-15

Ejercicio 1

/indent Seleccionamos un dataset que incluye información para luego estimar el nivel de obesidad de individuos basado en sus hábitos alimenticios y su condición física. Los países de donde se extrajeron los datos son Mexico, Peru y Colombia. La información contiene 17 atributos y 2111 observaciones. Los registros están etiquetados con la variable de clase NObesity (Obesity Level), lo que permite la clasificación de los datos utilizando los valores de Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II y Obesity Type III. /indent Vamos a aplicar una transformación del dataset para convertirlo en una clasificación binaria. Para eso vamos a cambiar las etiquetas anteriormente mencionadas (Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II) por 0 indicando que la persona no es obesa y (Obesity Type I, Obesity Type II y Obesity Type III) por 1 indicando que el individuo es obeso.

/indent Generamos una tabla de frecuencias que muestra cuántas veces aparece cada categoría única en la columna "NObesidad".

```
knitr::opts_chunk$set(echo = TRUE)
data <- read.csv("./data/ObesityDataSet_raw_and_data_synthetic.csv")

columna_categorias <- "NObesidad"

tabla_frecuencias <- table(data[, columna_categorias])
tabla_frecuencias
```

```
##
## Insufficient_Weight      Normal_Weight      Obesity_Type_I      Obesity_Type_II
##           272           287           351           297
## Obesity_Type_III Overweight_Level_I Overweight_Level_II
##           324           290           290
```

Hacemos la transformación

```
data <- read.csv("./data/ObesityDataSet_raw_and_data_synthetic.csv")

reemplazar_categorias <- function(categoria) {
  if (categoria %in% c("Insufficient_Weight", "Normal_Weight", "Overweight_Level_I", "Overweight_Level_II")) {
    return(0)
  } else if (categoria %in% c("Obesity_Type_I", "Obesity_Type_II", "Obesity_Type_III")) {
    return(1)
  } else {
    return(NA)
  }
}
```

```
data$NObeyesdad <- sapply(data$NObeyesdad, reemplazar_categorias)

write.csv(data, "./data/transformacion.csv", row.names = FALSE)
```

/indent Luego de hacer la transformacion volvemos a votar para verificar que este balanceado la cantidad de unos y ceros.

```
data <- read.csv("./data/transformacion.csv")

columna_categorias <- "NObeyesdad"

tabla_frecuencias <- table(data[, columna_categorias])
tabla_frecuencias
```

```
##
##      0      1
## 1139  972
```