

TP 1

Tecnología Digital VI

Piñera - Saint-Nom - De Diego

2023-08-22

Ejercicio 1

Para este trabajo, elegimos un dataset que contiene los hábitos alimenticios y condiciones físicas de individuos entre los 14 y 61 años de edad, de los países de México, Perú y Colombia. La información contiene 17 atributos y 2111 observaciones. El 77% de los datos fue generado sintéticamente utilizando la herramienta Weka y el filtro SMOTE, mientras que el 23% de los datos fue recopilado directamente de usuarios a través de una plataforma web. El dataset no tiene valores faltantes.

El link para encontrarlo es: <https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition>.

Los **atributos relacionados con los hábitos alimenticios** son:

- Consumo frecuente de alimentos altos en calorías (FAVC),
- Frecuencia de consumo de vegetales (FCVC),
- Número de comidas principales (NCP),
- Consumo de alimentos entre comidas (CAEC),
- Consumo diario de agua (CH20) y
- Consumo de alcohol (CALC).

Los **atributos relacionados con la condición física** son:

- Monitoreo del consumo de calorías (SCC),
- Frecuencia de actividad física (FAF),
- Tiempo de uso de dispositivos tecnológicos (TUE),
- Medio de transporte utilizado (MTRANS).
- Otras variables obtenidas fueron: Género, Edad, Altura y Peso.

Los registros del valor de retorno están etiquetados con la variable de clase NObeyesdad (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.

Para poder clasificar los usuarios en individuos con o sin obesidad, aplicamos una transformación del dataset para convertirlo en una clasificación binaria. Para eso vamos a cambiar las etiquetas de las categorías anteriormente mencionadas en la columna “NObeyesdad” (Insufficient Weight, Normal Weight, Overweight Level

I, Overweight Level II) por 0 indicando que el individuo no tiene obesidad y (Obesity Type I, Obesity Type II y Obesity Type III) por 1 indicando que el individuo tiene obesidad.

Generamos una tabla de frecuencias que muestra cuántas veces aparece cada categoría única en la columna “NObesidad”.

```
##
## 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

El criterio es (Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II) por 0 indicando que el individuo no es obeso y (Obesity Type I, Obesity Type II y Obesity Type III) por 1 indicando que el individuo es obeso.

Luego de hacer la transformación volvemos a comparar la cantidad de ocurrencias para verificar que este balanceada la cantidad de unos (1s) y ceros (0s), para comprobar que exista suficiente información sobre individuos sin obesidad como individuos con obesidad y así poder crear un mejor modelo de predicción.

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