

## PROPUESTA DE INVESTIGACIÓN

Del Guidice, Mateo – De Lora, Josip – Nieto, Ian – Rodríguez, René – Sakhrani, Dhruv

### Nombre de la investigación:

Relación entre los hábitos alimenticios, la actividad física y los niveles de obesidad: un análisis predictivo con machine learning

### Descripción de la base de datos a utilizar:

La base de datos contiene registros de personas con distintas características relacionadas con la salud y el estilo de vida. Incluye variables como:

Nombre del campo	Tipo de dato	Descripción en español
Gender	Categorico	Género
Age	Continuo	Edad
Height	Continuo	Estatura
Weight	Continuo	Peso
family_history_with_overweight	Binario	¿Algún familiar sufre o ha sufrido de sobrepeso?
FAVC	Binario	¿Consumes alimentos altos en calorías con frecuencia?
FCVC	Entero	¿Sueles incluir vegetales en tus comidas?
NCP	Continuo	¿Cuántas comidas principales realizas al día?
CAEC	Categorico	¿Consumes alimentos entre comidas?
SMOKE	Binario	¿Fumas?
CH2O	Continuo	¿Cuánta agua bebes diariamente?
SCC	Binario	¿Controlas las calorías que consumes diariamente?

<b>FAF</b>	Continuo	<b>¿Con qué frecuencia realizas actividad física?</b>
<b>TUE</b>	Entero	<b>¿Cuánto tiempo usas dispositivos tecnológicos como celulares, videojuegos, TV, etc.?</b>
<b>CALC</b>	Categorico	<b>¿Con qué frecuencia consumes alcohol?</b>
<b>MTRANS</b>	Categorico	<b>¿Qué medio de transporte utilizas con mayor frecuencia?</b>
<b>NObeyesdad</b>	Categorico	<b>Nivel de obesidad</b>

Esta base es ampliamente utilizada para modelos de clasificación y regresión en temas de salud pública.

#### **Esta base de datos fue seleccionada por las siguientes razones:**

Tiene un conjunto de variables relevante y bien definido\* para estudiar la obesidad.

Permite aplicar técnicas de análisis estadístico y de machine learning\*.

Está disponible en formato abierto, lo que permite su uso sin restricciones éticas o legales.

Representa una problemática actual de salud global con potencial impacto social.

#### **Introducción del contexto:**

La obesidad es uno de los problemas de salud pública más importantes en el mundo.

Factores como la alimentación inadecuada, el sedentarismo, y malos hábitos de vida son determinantes clave en su desarrollo. Entender cómo estos factores interactúan y predicen el nivel de obesidad permite generar mejores políticas de prevención y tratamiento. Esta investigación busca explorar dichos factores usando análisis de datos y modelos predictivos.

#### **Línea de tiempo de investigación**

Fecha	Actividad
19 - 21 junio	Búsqueda y revisión de la base de datos seleccionada. Verificación de calidad y estructura.
24 - 27 junio	Reporte de Avance: entrega con estadísticas descriptivas y visualizaciones iniciales.
28 junio - 2 julio	Análisis descriptivo profundo. Visualización de correlaciones entre variables.
3 - 5 julio	Simulación de eventos discretos.
6 - 8 julio	Aplicación de simulación de Monte Carlo.
9 - 12 julio	Modelado predictivo con machine learning.
13-15 julio	Redacción del reporte final, conclusiones y recomendaciones. Subida del código y base de datos a GitHub.
16 - 17 julio	Preparación de la presentación: enfoque en hallazgos clave y reflexión personal.
18 julio	Entrega final del proyecto + presentación oral del estudio.

#### **Fuente:**

<https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habit+s+and+physical+condition>