Consideraciones sobre los datasets Framingham clean y Cardio Train

Fuentes: Kaggle

Comparando ambos conjuntos de datos:

1. **Primer Dataset (Framingham)**
   * Incluye variables específicas como education, BPMeds, y prevalentStroke que son relevantes para un análisis detallado de los factores de riesgo cardiovascular. Además, tiene una variable objetivo TenYearCHD que es útil para estudios prospectivos de riesgo a 10 años.
2. **Segundo Dataset (Cardiotrain)**
   * Tiene una estructura más directa y puede ser más accesible para manejar, especialmente con variables básicas y claras que son fundamentales en estudios de riesgo cardiovascular como height, weight, ap\_hi, ap\_lo, cholesterol, y gluc. La variable cardio como objetivo permite realizar análisis de clasificación directa para detectar la presencia de enfermedades cardiovasculares.

Ambos datasets son útiles, pero su elección depende del enfoque específico:

* Si el objetivo es realizar un análisis predictivo detallado con un enfoque en la intervención y gestión de factores a largo plazo, el **primer dataset** (Framingham) podría ser más adecuado debido a su enfoque en factores longitudinales y la inclusión de variables médicas específicas adicionales.
* Si se prefiere centrarse en el análisis de datos y la implementación de modelos de aprendizaje de máquina más directos con menos necesidad de preprocesamiento y un enfoque en la identificación inmediata de factores de riesgo, el **segundo dataset** puede ser más práctico y accesible.