Big Data - Tarea #3

18/12/2023

Estudiante: Gabriel Valentine

# Datos. Diabetes

El dataset escogido se centra en el diagnóstico de la diabetes mediante la clasificación binaria. Cada atributo en el conjunto de datos representa información médica relevante para un paciente, y el objetivo es predecir si el paciente tiene diabetes o no, basándose en las características proporcionadas.

**Atributos en el Conjunto de Datos:**

1. Number of times pregnant: Número de veces que una paciente ha estado embarazada.
2. Plasma glucose concentration a 2 hours: Concentración de glucosa en plasma después de un test de tolerancia a la glucosa de 2 horas.
3. Diastolic blood pressure: Presión arterial diastólica (mm Hg).
4. Triceps skin fold thickness: Grosor del pliegue cutáneo del tríceps (mm).
5. 2-Hour serum insulin: Nivel de insulina sérica después de 2 horas (mu U/ml).
6. Body mass index: Índice de masa corporal (peso en kg / (altura en m)^2).
7. Diabetes pedigree function: Función de pedigrí de diabetes, que proporciona información sobre la herencia genética.
8. Age: Edad del paciente en años.
9. Class variable: Variable binaria que indica la presencia (1) o ausencia (0) de diabetes.

**Variable de Predicción:**

La variable de predicción es la "Class variable", que es binaria y se utiliza para determinar si un paciente tiene diabetes (1) o no (0).