

Ejercicio: Limpieza de Dataset.

Ejercicio:

- Explica qué problema puede causar cada uno de los dataset al entrenar un modelo.
- Propón una limpieza razonable y muestra el dataset final.

1) Eliminación de duplicados (no tan obvios)

ID_Cliente	Nombre	Edad	Ciudad
101	Ana	20	Madrid
102	Luis	22	Sevilla
101	Ana	20	Madrid
101	Ana	21	Madrid

2) Tratamiento de valores nulos (con elección)

Alumno	Edad	Nota_examen
Ana	18	8
Luis	×	6
Marta	19	×
Juan	45	9

3) Normalización vs estandarización

Persona	Edad	Horas_estudio
Ana	18	2
Luis	25	6
Marta	40	1

4) Codificación de variables categóricas (decisión consciente)

Cliente	Nivel_estudios
Ana	Primaria
Luis	Secundaria
Marta	Universidad

5) Outliers: detectar y decidir

Cliente	Consumo_kWh
A	210
B	230
C	220
D	2.500