

RETO 1: Clasificación de Muestras Biológicas

Objetivo:

Codificar las variables categóricas en un conjunto de datos que contiene información sobre muestras biológicas recolectadas para un estudio genético.

Dataset simulado:

- `muestra_id`: Identificador único
- `especie`: ["ratón", "rata", "conejo"]
- `sexo`: ["macho", "hembra"]
- `hábitat`: ["laboratorio", "campo"]
- `peso_g`: valor numérico

Actividades:

1. Aplicar **Label Encoding** a la columna `sexo` y `hábitat`.
 2. Aplicar **One-Hot Encoding** a la columna `especie`.
 3. Crear variables dummy con Pandas (`get_dummies`) para comparar los métodos.
 4. Comparar las formas resultantes (`shape`) y reflexionar sobre cuándo conviene usar cada técnica.
-



RETO 2: Codificación de Datos Meteorológicos

Objetivo:

Preparar variables categóricas para modelar predicción del clima.

Dataset simulado:

- `día`: 1 a 10
- `tipo_clima`: ["soleado", "nublado", "lluvioso", "tormentoso"]
- `región`: ["norte", "sur", "centro"]
- `temperatura_max`: temperatura en °C

Actividades:

1. Aplicar Label Encoding a `tipo_clima`.
 2. Aplicar One-Hot Encoding a `región` usando `ColumnTransformer`.
 3. Aplicar `get_dummies` a ambas columnas para comparar.
 4. Mostrar visualización (ej. gráfico de barras) de los valores codificados.
-



RETO 3: Análisis de Ensayos Clínicos

Objetivo:

Codificar variables de un ensayo clínico para análisis estadístico.

Dataset simulado:

- paciente_id
- grupo_tratamiento: ["placebo", "experimental", "control"]
- sexo: ["masculino", "femenino"]
- resultado: ["mejoró", "sin cambio", "empeoró"]
- nivel_glucosa: numérico

Actividades:

1. Usar LabelEncoder para resultado.
 2. Aplicar One-Hot Encoding a grupo_tratamiento y sexo.
 3. Usar get_dummies con drop_first=True para evitar multicolinealidad.
 4. Mostrar las diferencias de forma entre codificación ordinal y categórica.
-



RETO 4: Análisis de Muestras de Agua

Objetivo:

Codificar los tipos de contaminantes y fuentes de muestra para análisis de clasificación.

Dataset simulado:

- muestra_id
- fuente: ["río", "pozo", "lluvia", "canal"]
- tipo_contaminante: ["orgánico", "metálico", "biológico"]
- nivel_contaminación: numérico
- zona: ["urbana", "rural"]

Actividades:

1. Aplicar Label Encoding a tipo_contaminante.
2. One-Hot Encoding a fuente y zona.
3. Crear variables dummy con pandas.get_dummies().
4. Comparar los resultados codificados y explicar cuál sería más adecuado para un modelo lineal.