Transformación de Datos

Preparando los datos para poder predecir

CLAVES PARA TRABAJAR CON DATOS Compartir los resultados Creación de dashboards interactivos que visualizan datos de manera clara y accesible para la toma de decisiones. **Definir el Problema** Obtención de Datos Definir cuál es problema que queremos resolver con los Los Datos se pueden obtener datos que tenemos. de distintas fuentes de datos, web scraping, bases de datos, archivos, etc. Transformación de Datos Proceso de modificar datos para prepararlos para análisis, incluyendo limpieza, normalización, codificación, etc Exploración de Datos (EDA) Es donde vamos a comprender los datos para poder resolver nuestras preguntas de forma

eficaz.

Hack(io)



Principales Fases:

- Imputación de Nulos
- Imputación de Outliers
- Selección de Variables
- Codificación de variables categóricas (Encoding)
- Estandarización

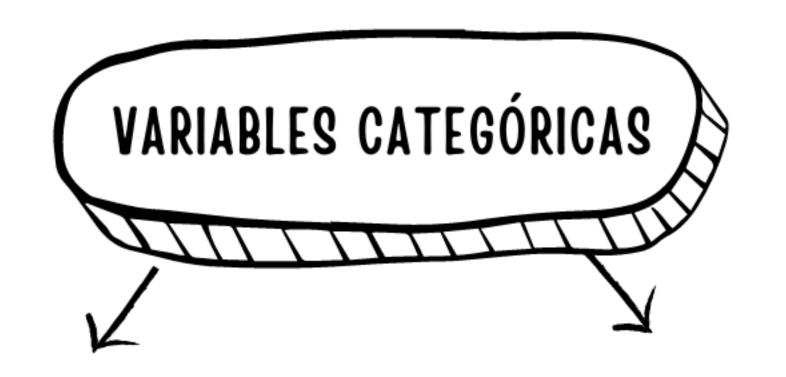
Codificación VC

El encoding es crucial en ML porque los algoritmos requieren datos numéricos, mientras que muchas características son categóricas.

En esta fase vamos a transformar las columnas categóricas a formatos numéricos comprensibles, permitiendo que los modelos aprendan relaciones y patrones precisos para tomar decisiones efectivas.

Para esto primero tendremos que tipo de variables categóricas tenemos en el conjunto de datos.

Tipos Variables Categóricas (basado en si tienen orden o no)



ORDINALES

Son aquellas variables que tienen un orden, como niveles de educación o clasificaciones de opiniones.

NOMINALES

Son categorías sin un orden inherente, como colores o tipos de animales.

$Z = \frac{x_i - \bar{X}}{\sigma}$

Donde:

- x_i : cada valor de la columna
- x̄: es la media de la columna
- σ : la desviación estándar

Estandarización VN

La estandarización es el proceso de transformar los datos para que tengan una media de cero y una desviación estándar de uno.

Se realiza para igualar la escala de las variables predictoras/independientes, lo que facilita la comparación y mejora el rendimiento de los algoritmos.

Web para hacer el modelo