

INTELIGENCIA ARTIFICIAL CON DEEP LEARNING

ING. JORGE ALBERTO CASTELLANOS



Universidad de
La Sabana

FACULTAD DE INGENIERÍA

Semana 2

Diplomado en Inteligencia Artificial con Deep Learning

Configuración Experimental

Configuración experimental

Qué es una configuración experimental

Una configuración experimental se refiere al diseño y disposición de los elementos y condiciones bajo las cuales se lleva a cabo un experimento.

En el contexto de **Machine Learning**, una configuración experimental se refiere a cómo se organiza y ejecuta un experimento para entrenar y evaluar modelos de aprendizaje automático.

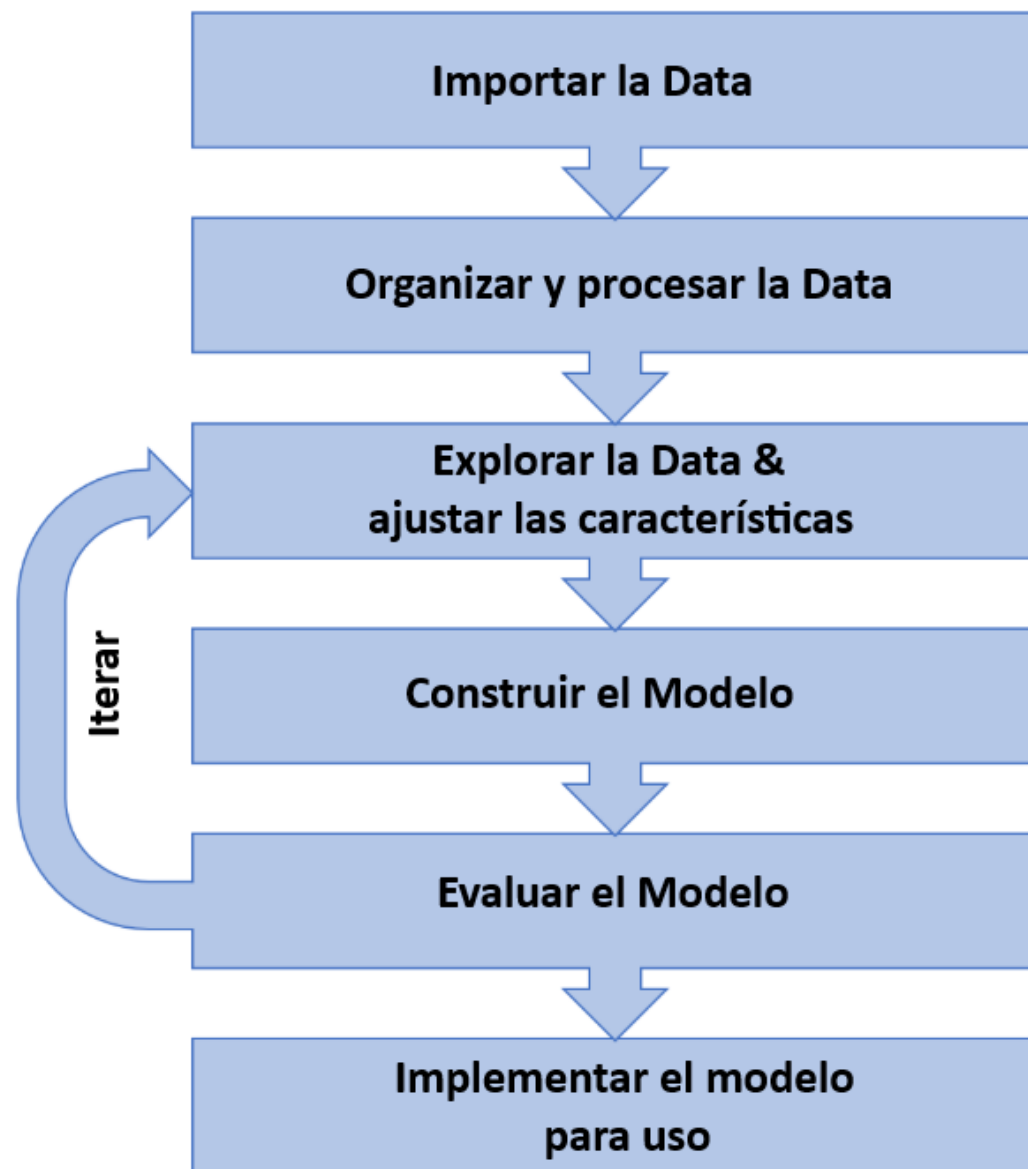
Objetivo

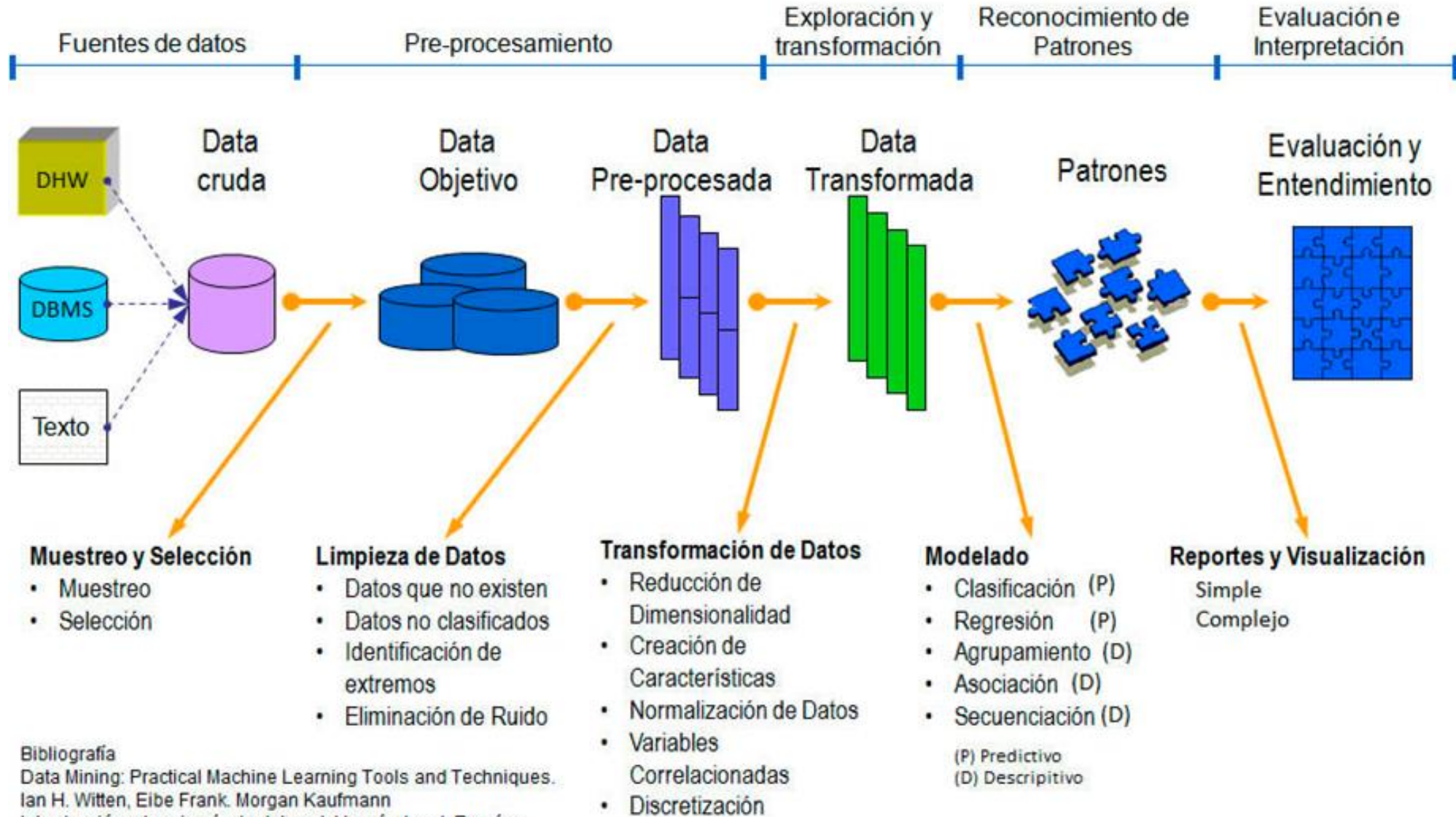
Garantizar que los resultados sean reproducibles, confiables y generalizables.

Revisión

- Configuración Experimental:
 - [Link 1](#)
- Ejercicio Propuesto
 - [Link 2](#)

Visión general del flujo de trabajo



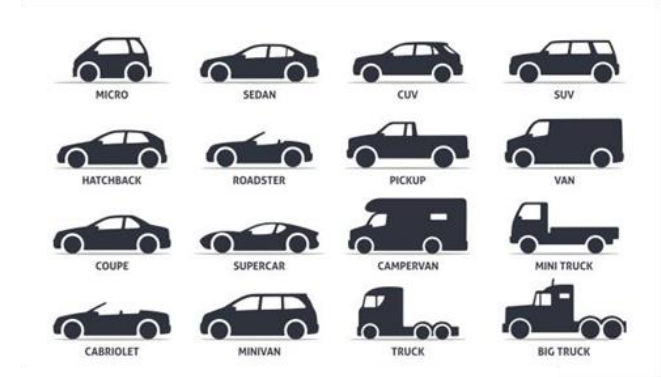


Componentes claves

Datos: Selección y preparación de los conjuntos de datos de entrenamiento, validación y prueba.



3 7 9
1 5
2 0 8
4 6



Preprocesamiento de Datos

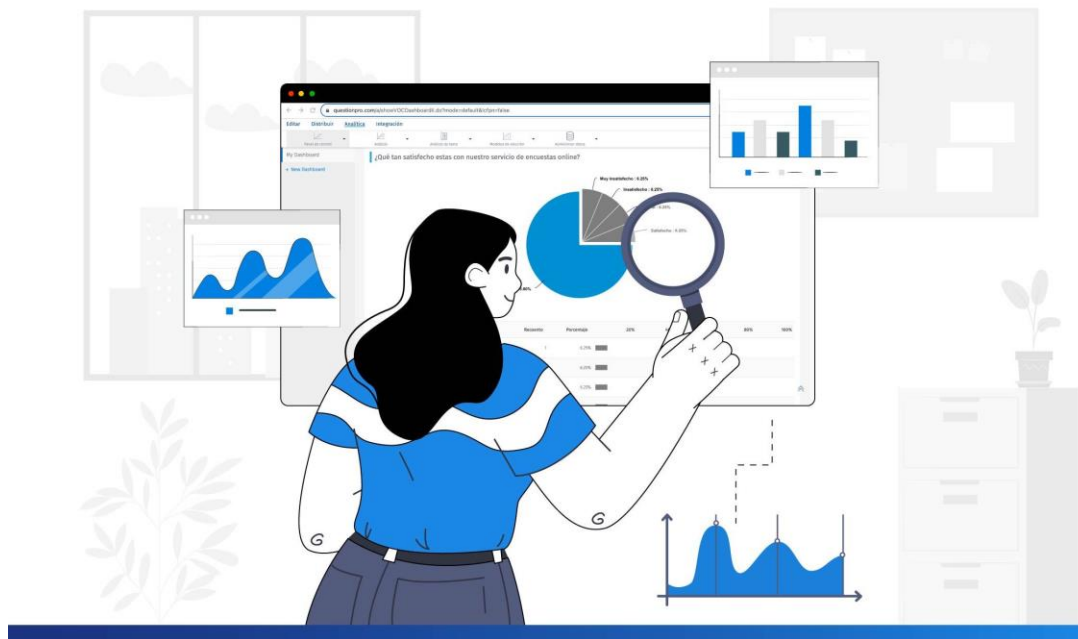
Selección y Muestreo



- Identificar y seleccionar las variables relevantes en los datos.
- Variables de valor
- La muestra permite inferir las propiedades o características de toda la población con un **error** medible y acotable

Preprocesamiento de Datos

Exploración



- Técnicas de **análisis exploratorio** para identificar valores **inusuales**, valores **extremos**, valores **desaparecidos**, discontinuidades entre otras.

Preprocesamiento de Datos

- **Limpieza**

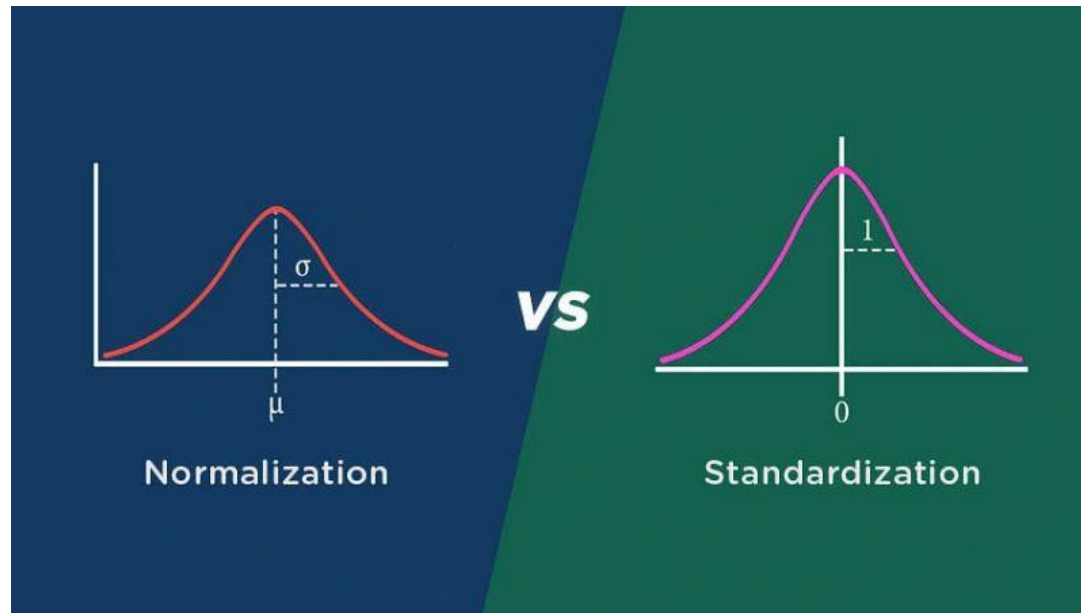
- Es más fácil después de la fase exploratoria



- Manejo de valores faltantes.
- Eliminación de duplicados.
- Corrección de errores de formato.
- Detección de outliers.

Preprocesamiento de Datos

Normalización y Estandarización



- Normalización: Escalar datos a un rango específico.
- Estandarización: Media 0 y desviación estándar 1.

Preprocesamiento de Datos

Codificación de Variables Categóricas

id	color
1	red
2	blue
3	green
4	blue



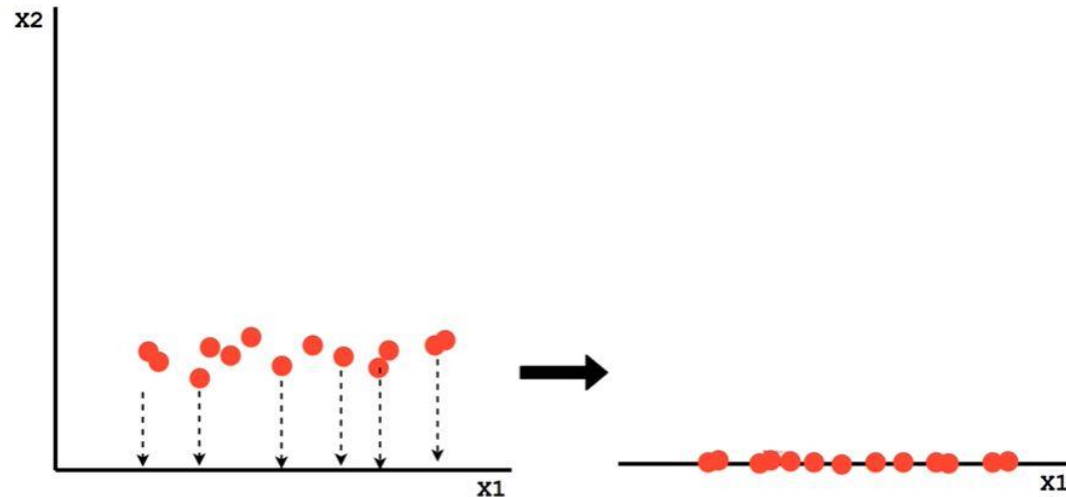
id	color_red	color_blue	color_green
1	1	0	0
2	0	1	0
3	0	0	1
4	0	1	0

- One-Hot Encoding.
- Label Encoding.
- Embeddings.

Preprocesamiento de Datos

Reducción de Dimensionalidad

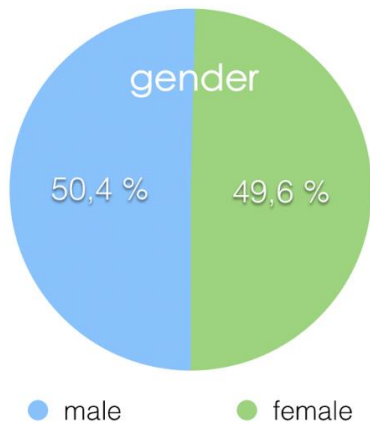
- Selección de características.
- Técnicas como PCA y t-SNE.



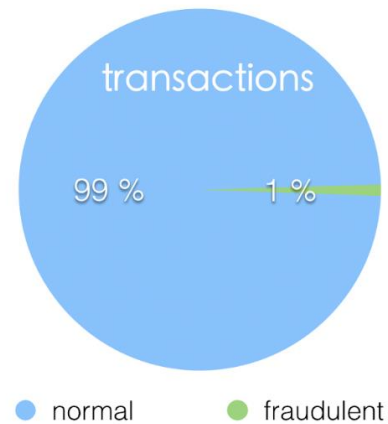
Preprocesamiento de Datos

Manejo de Datos Desbalanceados

Balanced Dataset



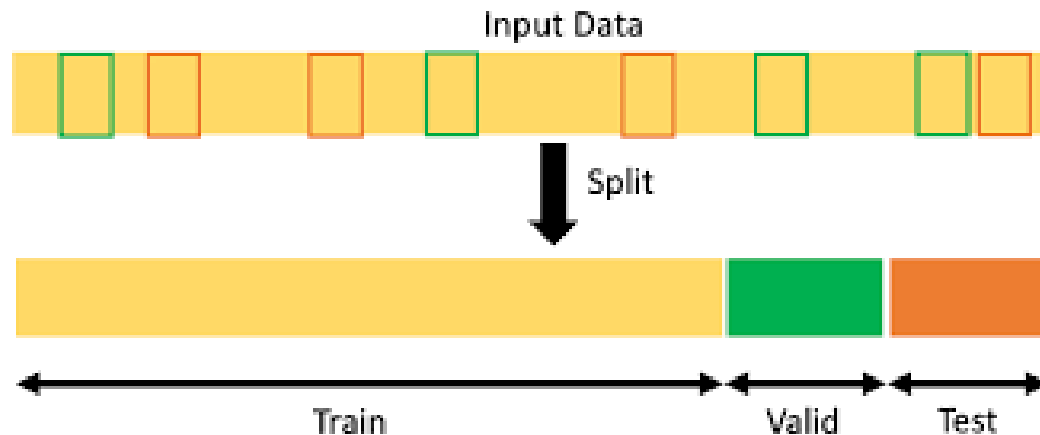
Unbalanced Dataset



- Oversampling (SMOTE).
- Undersampling.
- Pesos en el modelo.

Preprocesamiento de Datos

División de Datos



- Train/Test Split.
- Validación cruzada (k-fold).

Preprocesamiento de Datos

Ingeniería de Características



- Creación de nuevas variables.
- Combinación de variables.

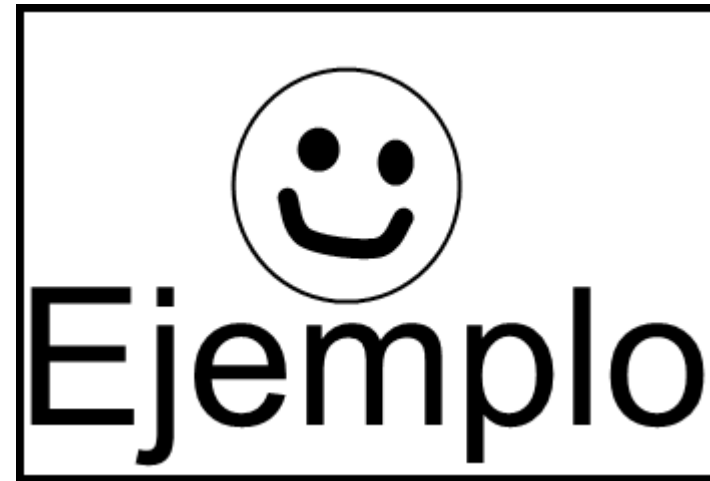
Preprocesamiento de Datos

Conclusiones:

- El preprocesamiento es esencial para obtener resultados confiables.
- Debe adaptarse al problema y los datos.
- Mejora la calidad y el rendimiento de los modelos.

Ejemplo

- Preprocesamiento

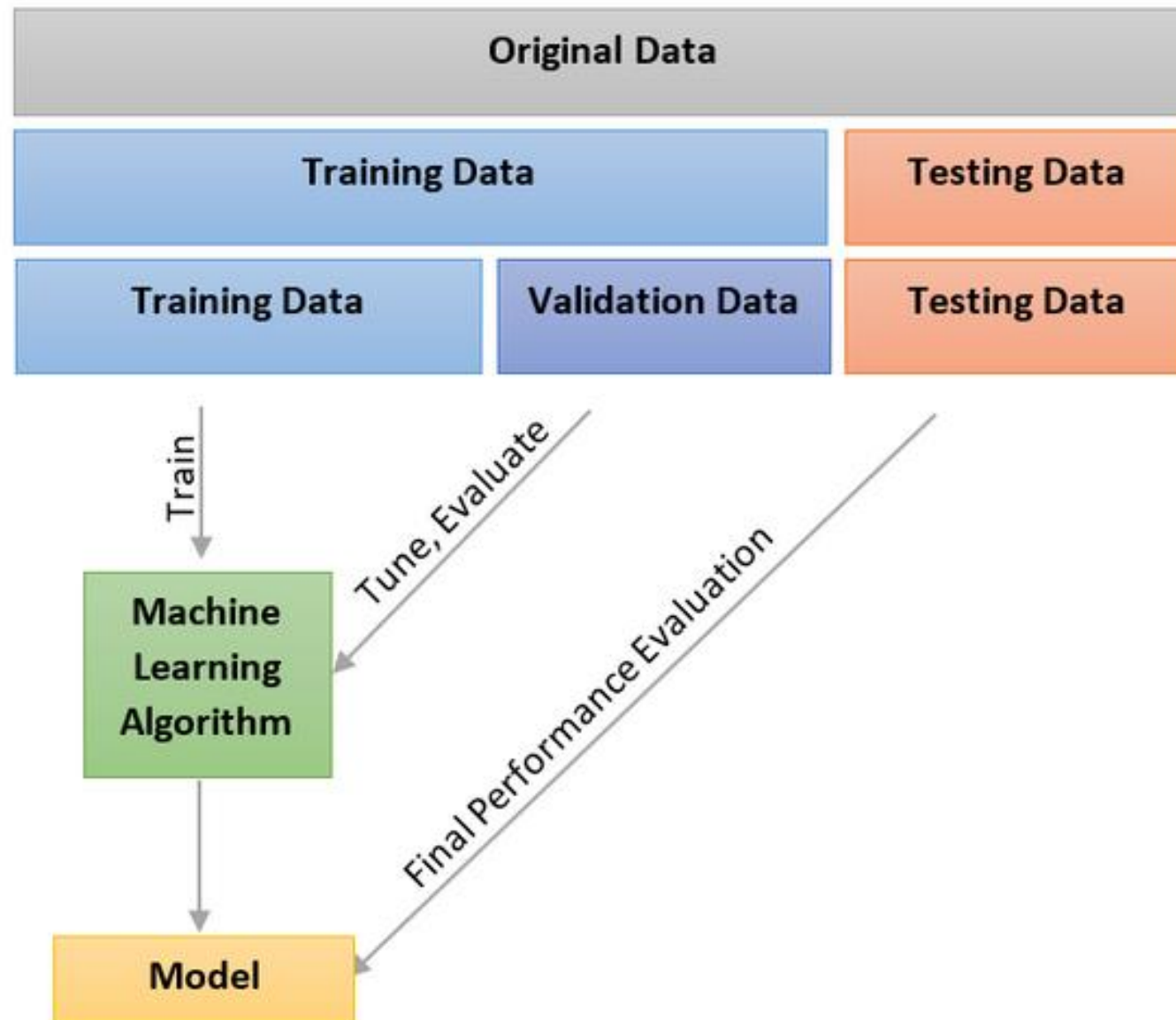


Ejercicio

- Se tiene el siguiente dataset:
- [The Boston Housing Dataset](#)

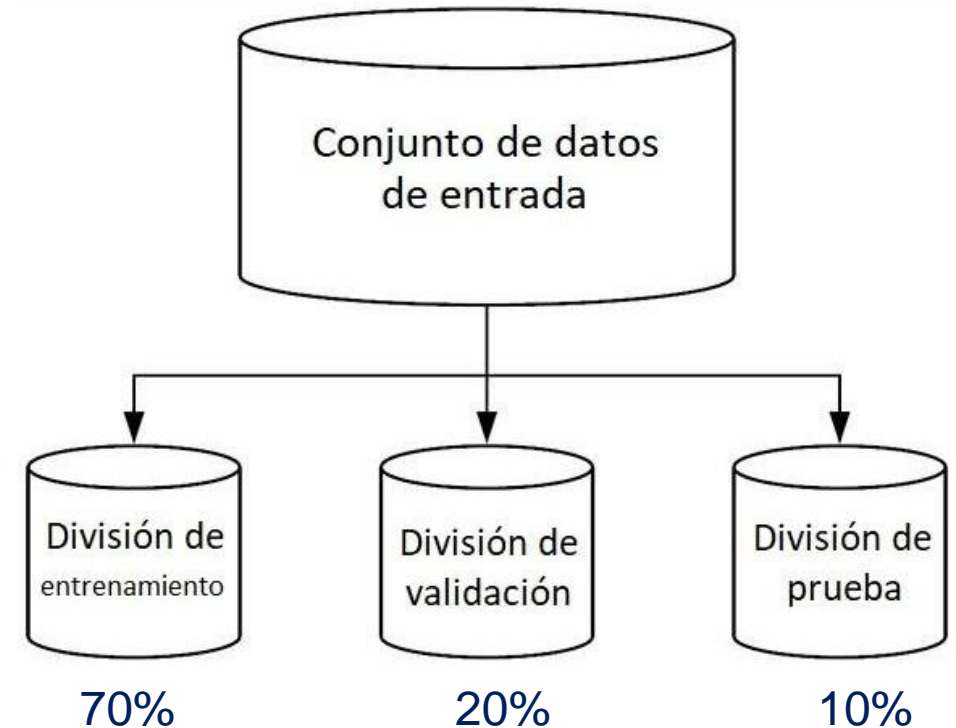
Explore el Dataset.

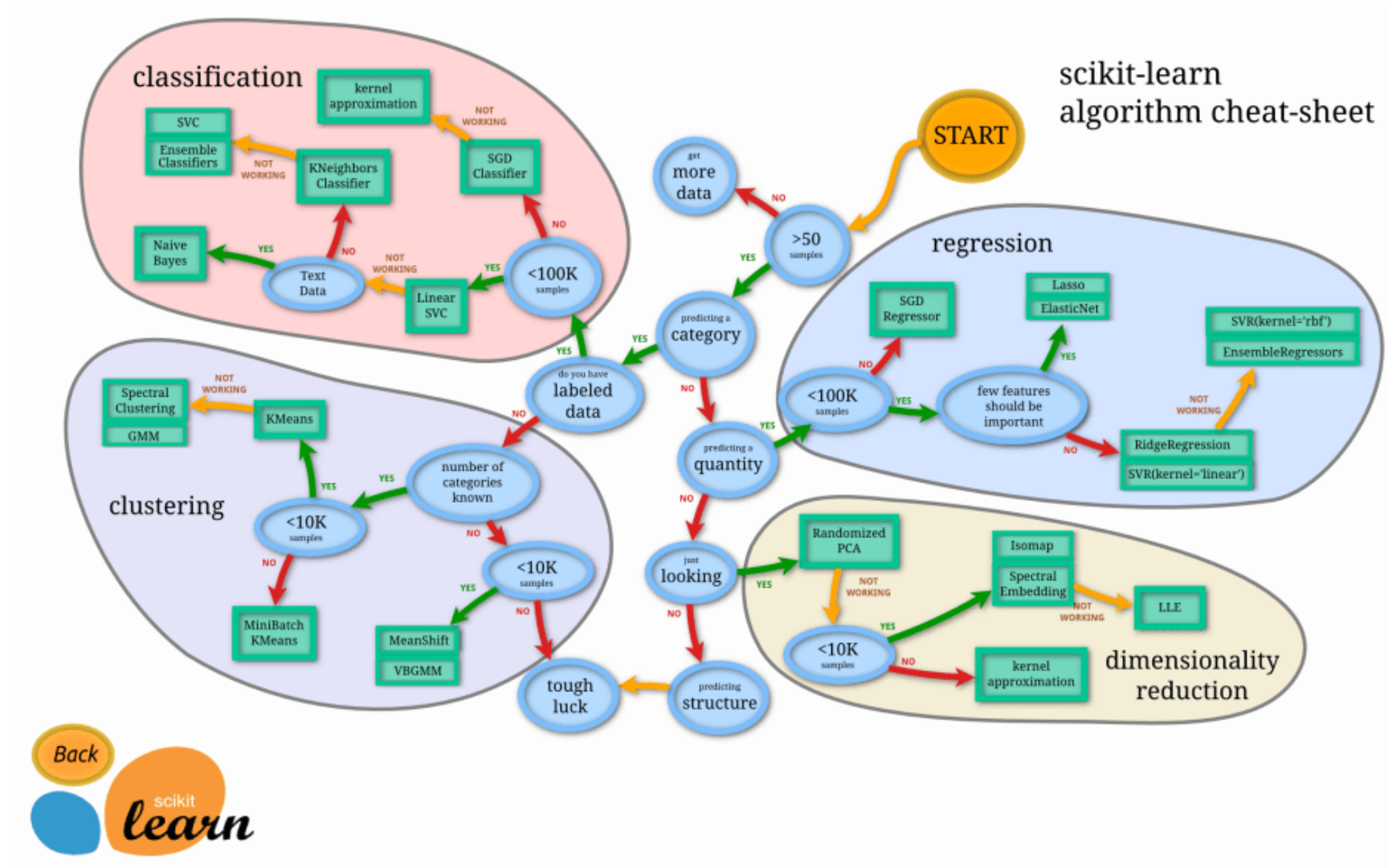
- Identifique las columnas y filas.
- Cuantas características tiene el Dataset.
- Que objetivo puedo obtener al entrenar con este Dataset.
- Realice preprocesamiento a los datos
 - Valores faltante (Mediana, modo)
 - Valores atípicos
 - Normalización(MinMaxScaler,)



Componentes de la configuración experimental

- **Datos:** División en conjuntos de entrenamiento, validación y prueba.
- **Modelos:** Selección de algoritmos y ajuste de hiperparámetros.
- **Métricas:** Elección de métricas de evaluación (exactitud, precisión, recall, f1-Score.).
- **Infraestructura:** Herramientas y entornos de trabajo (Google Colab, Jupyter Notebook, etc.).





Taller Semana

- Realizar el preprocesamiento de los Dataset que se encuentran en el siguiente enlace.
- https://scikit-learn.org/stable/datasets/toy_dataset.html