

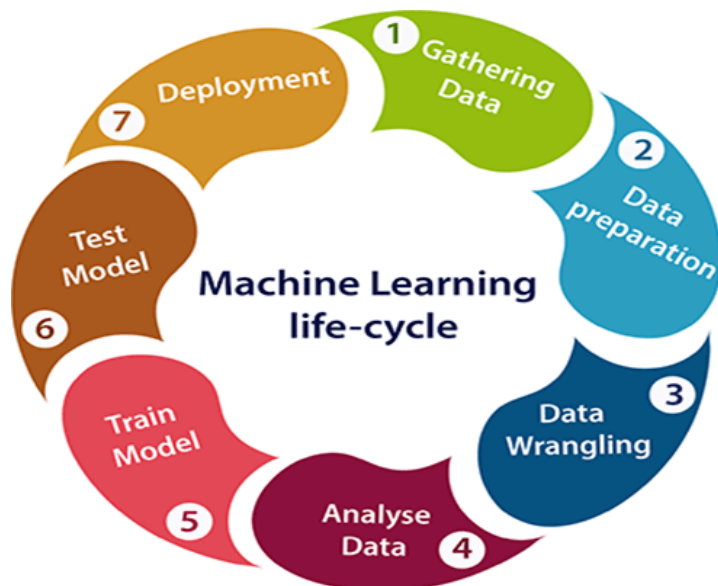
Proyecto de Evaluación de Módulo – *"Aplica lo aprendido en Machine Learning"*

Objetivo:

El alumno deberá elegir uno de los tres tipos de problema (regresión, clasificación o clustering) y desarrollar un mini-proyecto END2END aplicando todo lo aprendido:

- Carga y exploración del dataset
- Preprocesamiento (limpieza, encoding, nulos, escalado, etc.)
- Analítica, distribuciones, relación entre variables, relación con la target... (Visualizaciones)
- Elección y aplicación del modelo adecuado, probar diferentes modelos, diferentes parámetros...
- Evaluación de resultados
- Hacer una predicción de un registro aleatorio

(Obviamente no haremos el deployment)



OPCIONES DE PROYECTO

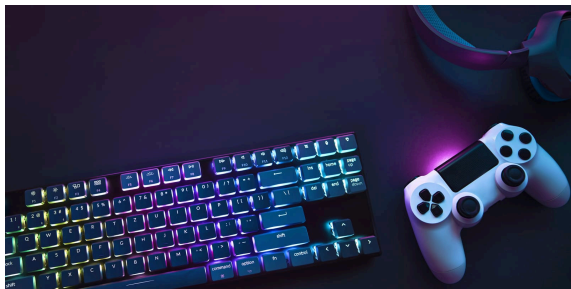
REGRESIÓN - Video Game Sales Dataset

Enlace: [Video Game Sales Dataset en Kaggle](#)

Descripción: Este conjunto de datos contiene información sobre ventas de videojuegos en diferentes plataformas, géneros y regiones, lo que lo hace valioso para diversos análisis y casos de uso empresariales.

- Contiene columnas como: **Name**, **Platform**, **Year**, **Genre**, **Publisher**, **Global_Sales**, etc. 💡 Tiene variables categóricas y numéricas, permite encoding, limpieza de nulos y regresión lineal, SVR o XGBoost.

Objetivo sugerido: Predecir las ventas globales (**Global_Sales**) a partir de variables como género, plataforma y año.



CLASIFICACIÓN - Heart Disease Dataset

Enlace: [Heart Disease Dataset en Kaggle](#)

Descripción: Este conjunto de datos, proveniente del repositorio de UCI, incluye atributos de pacientes utilizados para predecir la presencia de enfermedades cardíacas. Es ampliamente utilizado para proyectos de clasificación en el ámbito de la salud.

- Columnas como: **edad**, **sexo**, **presión arterial**, **colesterol**, **azúcar**, etc. Permite entrenar modelos como KNN, RandomForest, LogisticRegression, usar métricas y matriz de confusión.

Objetivo sugerido: Clasificar si una persona tiene enfermedad cardíaca (**target** binario).

CLUSTERING - Mall Customers Dataset

Enlace: [Mall Customers Dataset en Kaggle](#)

Descripción: Este conjunto de datos proporciona información sobre clientes de un centro comercial, incluyendo datos demográficos, ingresos anuales y hábitos de gasto, ideal para proyectos de segmentación de clientes mediante técnicas de clustering.

- Columnas como: `CustomerID`, `Gender`, `Age`, `Annual Income (k$)`, `Spending Score`. Muy visual, permite estandarización, PCA, método del codo, silhouette score y visualización de clusters.

Objetivo sugerido: Segmentar clientes en grupos usando KMeans



¿Qué debe entregar cada alumno?

- Exploración del dataset (`.describe()`, `.value_counts()`, nulls, etc.)
- Preprocesamiento de datos (encoding, escalado, limpieza)
- Modelo entrenado (según el tipo elegido)
- Métricas y visualizaciones adecuadas
- Análisis final + gráfica final + comentario interpretativo

En un repositorio, con un `README.md` bien explicado todo, un script con el código y una presentación (slides).