

## ACTIVIDAD EN EQUIPO:

### Introducción General

El aprendizaje de máquina es una disciplina central en la ciencia de datos, y su correcta aplicación implica una combinación de fundamentos teóricos, comprensión del flujo de trabajo de proyectos, tipos de algoritmos, tratamiento de datos y evaluación de desempeño. Esta actividad está pensada para que los equipos investiguen estos elementos clave desde la visión de un **científico de datos profesional o académico**, y los comuniquen con infografías visuales y presentaciones orales.

### Objetivo General

Que los estudiantes comprendan **cómo un científico de datos aplica el aprendizaje automático** (machine learning) para resolver problemas reales, investigando:

### Objetivos

- Investigar conceptos esenciales del aprendizaje de máquina.
- Conectar teoría con prácticas reales del trabajo de un científico de datos.
- Comprender la lógica detrás del uso de distintos algoritmos, validaciones y métricas.
- Comunicar visualmente lo aprendido mediante una infografía clara y técnica.

### Entregables:

- **Infografía profesional (digital, Canva o PDF)**
  - **Presentación oral de 7-10 minutos**
  - **Citación de al menos 3 fuentes confiables**
- 

## Equipo 1: Ciclo de Vida del ML + Clasificación + Métricas de Desempeño

### Investigar:

- ¿Qué es el ciclo de vida de un proyecto de ML (CRISP-DM u OSEMN)?
- ¿Qué es una tarea de clasificación?
- Algoritmos comunes: k-NN, Árboles, SVM, Regresión Logística.
- Métricas para clasificación: exactitud, precisión, recall, F1-score, matriz de confusión.

### Visualizaciones sugeridas:

- Diagrama del flujo CRISP-DM.
  - Comparativa visual de algoritmos de clasificación.
  - Ejemplo con matriz de confusión (visual).
- 

## Equipo 2: Preprocesamiento + Problemas de Regresión + Evaluación de Modelos

### **Investigar:**

- ¿Qué es el preprocesamiento de datos? ¿Qué tareas incluye?
- ¿Qué es una tarea de regresión? ¿Qué problemas aborda?
- Tipos de errores: sobreajuste, subajuste, generalización.
- Métricas de evaluación de regresión: MAE, RMSE,  $R^2$ .

### **Visualizaciones sugeridas:**

- Ejemplo visual de datos antes/después del preprocesamiento.
  - Gráfica de predicción de regresión vs valores reales.
  - Curvas de error (entrenamiento vs validación).
- 

## **Equipo 3: Tipos de Aprendizaje + Validación Cruzada + Bias y Varianza**

### **Investigar:**

- Diferencias entre aprendizaje supervisado, no supervisado y por refuerzo.
- ¿Qué es la validación cruzada (k-fold, leave-one-out)?
- ¿Qué es el trade-off entre bias y varianza?
- ¿Cómo se evita el sobreajuste usando validación adecuada?

### **Visualizaciones sugeridas:**

- Tabla comparativa de tipos de aprendizaje.
  - Esquema de validación cruzada.
  - Gráfico de bias vs varianza con explicaciones.
- 

## **Equipo 4: Científico de Datos + Clasificación Multiclase y Multietiqueta + Métricas de Evaluación Avanzadas**

### **Investigar:**

- ¿Qué hace un científico de datos? ¿Cómo se prepara para problemas de ML?
- Clasificación multiclase vs multietiqueta.
- Métricas adicionales: AUC-ROC, precisión macro vs micro.
- ¿Cómo se visualiza el desempeño en problemas complejos?

### **Visualizaciones sugeridas:**

- Mapa de habilidades del científico de datos.
- Esquema de clasificación multiclase vs multietiqueta.
- Gráfico AUC-ROC con interpretación.