

---

## RESULTADO DE APRENDIZAJE

---

### RdA de la asignatura:

- **RdA 1:** Plantear los conceptos fundamentales del aprendizaje automático, incluyendo los principios básicos, técnicas de preprocesado de datos, métodos de evaluación y ajuste de modelos, destacando su importancia en el análisis y resolución de problemas de datos.

### RdA de la actividad:

- Comprender qué es el aprendizaje no supervisado y sus aplicaciones generales.
- Diferenciar entre los tipos de agrupamiento: exclusivo, superposición, jerárquico y probabilístico.
- Identificar preguntas clave y reflexiones iniciales sobre el tema mediante la interacción con un modelo de IA.

---

## INTRODUCCIÓN

---

**Pregunta inicial:** ¿Cómo podríamos encontrar patrones en un conjunto de datos sin etiquetas?

---

## DESARROLLO

---

### Actividad 1: Introducción al Aprendizaje No Supervisado

Se utilizará un video introductorio para familiarizar a los estudiantes con el concepto de aprendizaje no supervisado y sus aplicaciones en problemas de datos sin etiquetas.

### ¿Cómo lo haremos?

- **Proyección de video:** Se presentará un video educativo que explique qué es el aprendizaje no supervisado.

Enlace al video: [Qué es el Aprendizaje No Supervisado](#).

- **Interacción con ChatGPT:** Después del video, cada estudiante formulará preguntas a ChatGPT relacionadas con los términos o conceptos que no entendieron durante la visualización.

**Verificación de aprendizaje:** Se llevará a cabo una discusión grupal donde los estudiantes compartirán:

- Los términos que investigaron.
- Las respuestas proporcionadas por ChatGPT.

Términos guía para las preguntas: K-Means, Clusterización jerárquica, DBSCAN, Modelos de mezclas gaussianas, Ruido en los datos, Detección de anomalías.

## Actividad 2: Definiciones clave

En esta actividad, los estudiantes explorarán las definiciones y características de los principales tipos de agrupamiento en el aprendizaje no supervisado.

### ¿Cómo lo haremos?

- **Presentación de conceptos:** El docente explicará brevemente las definiciones y ejemplos de los principales tipos de agrupamiento en aprendizaje no supervisado:
  - **Agrupamiento exclusivo:** Cada dato pertenece exclusivamente a un único grupo. Ejemplo: k-means.
  - **Agrupamiento con superposición:** Los datos pueden pertenecer a múltiples grupos con grados de pertenencia. Ejemplo: clustering difuso (*fuzzy clustering*).
  - **Agrupamiento jerárquico:** Los datos son organizados en una estructura de árbol jerárquico. Ejemplo: dendrogramas.
  - **Agrupamiento probabilístico:** Los datos se asignan a clústeres basándose en modelos probabilísticos. Ejemplo: Gaussian Mixture Models (GMMs).
- **Exploración guiada:** Los estudiantes realizarán dos tareas principales en grupos pequeños:
  1. **Preguntar a ChatGPT:** Cada grupo formulará las siguientes preguntas al modelo:
    - ¿Cuál es la principal diferencia entre el agrupamiento jerárquico y el agrupamiento exclusivo?
    - ¿Cuándo es preferible usar un modelo probabilístico como GMM frente a k-means?
    - ¿Qué limitaciones tienen los métodos de agrupamiento con superposición en aplicaciones reales?
  2. **Búsqueda de referencias:** Usando la plataforma [Consensus](#), los estudiantes buscarán un artículo científico que respalde o amplíe las respuestas obtenidas de ChatGPT. Deberán identificar la idea principal del artículo y cómo se relaciona con el tema discutido.

**Verificación de aprendizaje:** Cada grupo presentará sus hallazgos de la siguiente manera:

- Resumen de las respuestas obtenidas de ChatGPT.
- Referencia científica encontrada en Consensus y cómo esta valida o amplía las respuestas del modelo.

El docente facilitará la discusión y aclarará dudas adicionales sobre los conceptos.

### **Actividad 3: Efecto de la Normalización en la Clusterización**

En esta actividad, los estudiantes analizarán cómo las diferentes técnicas de normalización pueden afectar el resultado de los algoritmos de clusterización.

#### **¿Cómo lo haremos?**

- **Presentación inicial:** Se presenta un gráfico que muestra cómo diferentes normalizaciones (por ejemplo, *StandardScaler*, *MinMaxScaler*, *RobustScaler*, *Normalizer*) afectan la distribución de los datos en un espacio bidimensional.
- **Exploración individual:** Cada estudiante deberá analizar el gráfico y responder preguntas como:
  - ¿Cuál técnica de normalización parece más adecuada para separar los grupos visualmente?
  - ¿Cómo podrían estas diferencias impactar en algoritmos como k-means o clustering jerárquico?
- **Exploración del cuaderno de Jupyter:** Se proporcionará un cuaderno de Jupyter con ejemplos prácticos sobre el efecto de la normalización y el escalado en el agrupamiento K-Means.

Enlace al cuaderno: [07-Normalizacion](#).

#### **Verificación de aprendizaje:**

- Cada estudiante dará un análisis breve que incluya:
  - Observaciones sobre cómo la normalización afecta los resultados de k-means.
  - Una conclusión sobre qué técnica consideran más adecuada para este conjunto de datos y por qué.

---

### **CIERRE**

---

**Tarea:** Los estudiantes deberán realizar las siguientes actividades de forma individual:

- Buscar en [Kaggle](#) un conjunto de datos que sea adecuado para implementar un algoritmo de clusterización. Ejemplos de conjuntos de datos incluyen: segmentación de clientes, agrupación de películas o canciones, o análisis de patrones en datos demográficos.
- Descargar el conjunto de datos seleccionado e identificar qué características podrían necesitar preprocesamiento (como normalización o manejo de valores faltantes).

**Pregunta de investigación:**

1. ¿Cómo podrían detectarse anomalías en un conjunto de datos usando aprendizaje no supervisado?
2. ¿Qué factores pueden dificultar la identificación de clústeres claros en datos reales, y cómo podrías abordarlos?