

1. INTRODUCCIÓN AL APRENDIZAJE AUTOMÁTICO

DEFINICIÓN 1: Inteligencia Artificial.

La Inteligencia Artificial (IA) es una rama de la informática que se enfoca en la creación de sistemas capaces de realizar tareas que requieren «inteligencia humana», tales como el reconocimiento de voz, la toma de decisiones, el procesamiento del lenguaje natural, y la resolución de problemas.

DEFINICIÓN 2: Aprendizaje Automático.

El Aprendizaje Automático (Machine Learning) es una subdisciplina de la Inteligencia Artificial que se centra en el desarrollo de algoritmos y modelos que permiten a las máquinas aprender a partir de datos sin ser explícitamente programadas para cada tarea específica.

DEFINICIÓN 3: Minería de datos.

La minería de datos es el proceso de extraer conocimiento de grandes volúmenes de datos, identificando patrones y reglas que permitan entender y predecir eventos que generaron dichos datos.

DEFINICIÓN 4: Metodología CRISP-DM.

CRISP-DM es una metodología estándar en minería de datos que sigue una secuencia de fases:

- **Comprensión del negocio:** Definir los objetivos de negocio y plantear metas de minería de datos.
- **Comprensión de los datos:** Analizar la calidad y características de los datos.
- **Preparación de los datos:** Seleccionar, limpiar y preparar los datos para el modelo.
- **Modelado:** Aplicar técnicas de minería de datos para construir modelos predictivos o descriptivos.
- **Evaluación:** Validar los modelos para asegurar que cumplen los objetivos del negocio.
- **Despliegue:** Implementar los resultados y modelos en el entorno de la organización.

1.1 Tipos de aprendizaje

DEFINICIÓN 5: Aprendizaje Supervisado.

Dado un conjunto de datos de entrenamiento

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\},$$

donde $x_i \in \mathbb{R}^n$ son las características de las instancias y $y_i \in Y$ son las etiquetas asociadas (variable objetivo), el **aprendizaje supervisado** busca determinar una función

$$f: \mathbb{R}^n \rightarrow Y$$

que aproxime la etiqueta y de una instancia $x \in \mathbb{R}^n$.

DEFINICIÓN 6: Aprendizaje No Supervisado.

Dado un conjunto de datos de entrenamiento

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\},$$

donde $x_i \in \mathbb{R}^n$ son las características de las instancias que no posee etiquetas asociadas, el **aprendizaje no supervisado** busca determinar una función

$$h: \mathbb{R}^d \rightarrow \{C_1, C_2, \dots, C_k\}$$

que asigne cada instancia x_i a un grupo (clúster) C_j bajo algún criterio.

1.2 Tipos de tareas

DEFINICIÓN 7: Clasificación.

La clasificación es una tarea de aprendizaje supervisado donde el conjunto de etiquetas es finito o discreto. Un caso típico es la clasificación binaria donde $Y = \{0, 1\}$.

DEFINICIÓN 8: Regresión.

La regresión es una tarea de aprendizaje supervisado que donde el conjunto de etiquetas se entiende como un conjunto continuo.

DEFINICIÓN 9: Agrupamiento.

El agrupamiento (clustering) es una tarea de aprendizaje no supervisado que consiste en particionar un conjunto de datos $\{x_1, x_2, \dots, x_n\}$, donde $x_i \in \mathbb{R}^d$, en k grupos o clústeres $\{C_1, C_2, \dots, C_k\}$.