

25/09/2025



# Características de IA, ML, DM y Big Data

*Extracción de Conocimiento en Bases de Datos*

Luis Eduardo Aguilar Sarabia

***IDGS91N***

Docente: Luis Enrique Mascote Cano

# Introducción

Este documento describe y compara cuatro dominios relacionados: Inteligencia Artificial (IA), Machine Learning (ML), Data Mining (DM) y Big Data. Cada sección incluye: definición oficial, breve origen/evolución (2–3 líneas), tres beneficios clave, dos restricciones y dos retos actuales. Al final verás una tabla comparativa de beneficios, una lista conjunta de restricciones/retos y una conclusión con perspectivas futuras.

## Definiciones y breve historia

### Inteligencia Artificial

#### Definición oficial resumida:

La Inteligencia Artificial es la capacidad de un sistema computacional para realizar tareas que normalmente requieren inteligencia humana (razonar, aprender, planificar y percibir).

#### Origen y evolución:

El término se institucionaliza en la conferencia de Dartmouth (1956). Desde heurísticas simbólicas y sistemas basados en reglas, la IA evolucionó hacia métodos estadísticos y redes neuronales profundas, impulsada por mayores datos y capacidad de cómputo.

#### Beneficios clave:

1. Automatización de tareas cognitivas y toma de decisiones a escala.
2. Mejora de la eficiencia operativa (ej. detección de fraudes, diagnósticos médicos).
3. Nuevas capacidades: visión por computadora, procesamiento de lenguaje natural y agentes autónomos.

#### Restricciones:

- Requerimientos altos de datos y cómputo para modelos complejos.
- Dificultad de explicabilidad y cumplimiento regulatorio en sectores sensibles.

#### Retos actuales:

- Explicabilidad / interpretabilidad de modelos complejos (p. ej. redes profundas).
- Mitigación de sesgos y equidad en decisiones automatizadas.

## Machine Learning

### Definición oficial resumida:

Machine Learning es una subdisciplina de la IA que dota a los sistemas de la capacidad de aprender patrones y mejorar su rendimiento a partir de datos, sin programación explícita de reglas.

### Origen y evolución:

Conceptos surgieron en los años 50 (p. ej. Arthur Samuel). ML creció desde algoritmos estadísticos clásicos (regresión, árboles) hacia aprendizaje profundo y modelos de gran escala impulsados por grandes conjuntos de datos y GPU.

### Beneficios clave:

1. Predicción de tendencias y comportamiento (forecasting).
2. Automatización y reducción de errores manuales en tareas repetitivas.
3. Adaptación y mejora continua con datos nuevos (reentrenamiento).

### Restricciones:

- Riesgo de sobreajuste y necesidad de datos etiquetados de calidad.
- Costes y complejidad para desplegar y mantener modelos en producción.

### Retos actuales:

- Generalización fuera de la distribución de entrenamiento (robustez).
- Privacidad y gobernanza de datos para entrenamiento (p. ej. datos sensibles).

## Data Mining

### Definición oficial resumida:

Data Mining es el proceso de explorar grandes conjuntos de datos mediante técnicas estadísticas y de ML para descubrir patrones, relaciones y conocimiento útil para la toma de decisiones.

### Origen y evolución:

Aparece como disciplina en los 80–90 vinculada a bases de datos y estadística aplicada; evolucionó integrando técnicas de ML, y hoy forma parte de pipelines de análisis y BI (business intelligence).

### Beneficios clave:

1. Descubrimiento de patrones ocultos y segmentación de clientes.
2. Soporte a la toma de decisiones mediante insights accionables.
3. Optimización de procesos (p. ej. detección de anomalías, recomendadores).

### Restricciones:

- Calidad y limpieza de datos (garbage in → garbage out).
- Interpretación y validación de patrones estadísticos (falsos positivos).

### Retos actuales:

- Integración de resultados en flujos de negocio (operacionalización).
- Escalabilidad frente a volúmenes crecientes y heterogeneidad de datos.

## Big Data

### Definición oficial resumida:

Big Data describe conjuntos de datos tan voluminosos, rápidos y/o variados que las técnicas tradicionales no bastan; requiere arquitecturas y herramientas escalables para almacenamiento, procesamiento y análisis. (NIST y otros organismos ofrecen definiciones formales).

### Origen y evolución:

Término popularizado en la década de 2000 por el crecimiento masivo de datos digitales, sensores y redes; evolucionó hacia ecosistemas (Hadoop, Spark, data lakes) que soportan procesamiento distribuido.

### Beneficios clave:

1. Posibilidad de análisis a gran escala para mejorar decisiones estratégicas.
2. Detección en tiempo real y analítica avanzada (streaming).
3. Combinación de fuentes variadas (texto, sensores, logs) para insights más ricos.

### Restricciones:

- Coste de almacenamiento, procesamiento y transferencia de datos.
- Latencia y complejidad para ofrecer análisis en tiempo real.

### Retos actuales:

- Gobernanza, privacidad y cumplimiento (GDPR/ Leyes locales).

- Interoperabilidad y estandarización entre sistemas y formatos.

## Beneficios comparados

<b>Dominio</b>	<b>Beneficio 1</b>	<b>Beneficio 2</b>	<b>Beneficio 3</b>
<b>Inteligencia Artificial</b>	Automatización de decisiones complejas	Mejora en precisión de tareas cognitivas	Nuevas capacidades (NLP, visión)
<b>Machine Learning</b>	Predicción de tendencias	Reducción de errores manuales	Adaptación con datos nuevos
<b>Data Mining</b>	Descubrimiento de patrones	Insights para decisiones	Optimización de procesos
<b>Big Data</b>	Análisis a gran escala	Analítica en tiempo real	Integración de fuentes variadas

## Referencias

<https://www.ibm.com/mx-es/think/topics/artificial-intelligence>

<https://www.ibm.com/mx-es/think/topics/machine-learning>

<https://aws.amazon.com/es/what-is/data-mining/>

<https://www.oracle.com/mx/big-data/what-is-big-data/>