

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

Desarrollo y Gestión de Software



Extracción de Conocimiento en Bases de Datos

I.1. Reporte de investigación de los conceptos, beneficios, restricciones y retos de inteligencia artificial, machine learning, data mining y big data.

IDGS91N

PRESENTA:

Hugo Chaparro

DOCENTE:

Enrique Mascote

Chihuahua, Chih., 26 sept 2025

Introducción.

Los términos Inteligencia Artificial (IA), Machine Learning (ML), Minería de Datos (DM) y Big Data se refieren a dominios tecnológicos estrechamente relacionados, pero con enfoques y objetivos distintos. La IA es un campo amplio que busca crear sistemas capaces de emular la inteligencia. El aprendizaje automático (ML) es una subárea de la IA centrada en algoritmos que “aprenden” de datos. La minería de datos (DM) estudia métodos para extraer patrones y conocimiento de grandes conjuntos de datos. Big Data describe el manejo de volúmenes masivos de datos complejos (gran volumen, velocidad y variedad de datos). Este reporte analiza y compara cada dominio según definición, origen histórico, beneficios, restricciones y retos actuales.

Definición de cada dominio

- **Inteligencia Artificial (IA):** se define como *“la simulación de procesos de inteligencia humana por parte de máquinas, especialmente sistemas informáticos”*. En esencia, la IA engloba sistemas que razonan, aprenden y resuelven problemas como humanos.

Origen: El concepto surgió en 1950 con Alan Turing, quien planteó si las máquinas podían “pensar”, y se formalizó en la conferencia de Dartmouth (1956) bajo el término “Inteligencia Artificial”. Desde entonces ha pasado por olas de entusiasmo y recesos, evolucionando de sistemas expertos y lógicos clásicos a modelos de IA actuales basados en ML y redes neuronales profundas.

- **Machine Learning (ML):** es la disciplina que estudia *“el desarrollo de algoritmos y modelos estadísticos que permiten a los sistemas de cómputo realizar tareas sin instrucciones explícitas, basándose en patrones e inferencias”*. En otras palabras, los sistemas de ML aprenden y mejoran su rendimiento analizando datos.

Origen: El término *aprendizaje automático* fue introducido en 1959 por Arthur Samuel en el contexto de juegos y redes neuronales tempranas. A partir de entonces, ML ha crecido enormemente con avances como el perceptrón (años 60), los algoritmos de *backpropagation* (años 80) y el auge del *deep learning* desde la década de 2010.

- **Minería de Datos (DM):** es el proceso de *“recopilar, limpiar, procesar, analizar y obtener conclusiones útiles de los datos”*. Básicamente, DM se ocupa de descubrir patrones, relaciones o tendencias en conjuntos de datos grandes.

Origen: El término se asocia al campo KDD (“Knowledge Discovery in Databases”) de finales de los 80. Se popularizó en la década de 1990 con el crecimiento de bases de datos empresariales. Desde entonces, incorpora técnicas estadísticas, de ML y bases de datos para extraer conocimiento de los datos (por ejemplo, reglas de asociación y clasificación).

- **Big Data:** se refiere a entornos con conjuntos de datos de volumen, velocidad y

variedad extremadamente altos. Según ISO/IEC 20546:2019, “*Big Data implica conjuntos de datos extensos en volumen, velocidad, variedad y/o variabilidad*”. Big Data no es solo cantidad de datos, sino el uso de arquitecturas distribuidas para almacenarlos y procesarlos.

Origen: El término ganó tracción en los años 2000 cuando las fuentes de datos (internet, sensores, redes sociales) crecieron exponencialmente. Doug Laney introdujo el modelo “3Vs” (volumen, velocidad, variedad) en 2001 para describir este fenómeno. Desde entonces, Big Data ha evolucionado con tecnologías como Hadoop, Spark y bases de datos NoSQL.

Tabla comparativa de beneficios

Domini o	Beneficio 1	Beneficio 2	Beneficio 3
IA	Automatización de tareas complejas (p. ej., diagnóstico médico)	Mejora de la toma de decisiones con modelos predictivos (personalización)	Innovación en productos y servicios (por ej. asistentes virtuales)
ML	Precisión en predicciones y clasificaciones (p. ej., detección de fraude)	Aprendizaje y adaptación continua al analizar nuevos datos	Eficiencia en el análisis de grandes datos (automatiza análisis estadísticos)
DM	Descubrimiento de patrones ocultos en datos (mejor conocimiento del cliente)	Soporte a decisiones estratégicas basado en datos históricos	Generación de modelos de riesgo, segmentación y optimización de recursos

Big Data **Análisis** de enormes volúmenes de datos para insights (p. ej., tendencias de mercado) **Procesamiento** en tiempo real heterogéneos de datos a gran velocidad (reacción inmediata) Integración de datos **estructurados y no estructurados)**

Restricciones y retos actuales

● Inteligencia Artificial (IA):

- *Restricciones:* Dependencia de grandes volúmenes de datos de calidad y

de recursos computacionales avanzados. Falta de estándares unificados o marcos regulatorios estables para su implementación segura.

- *Retos:* Éticos y sociales: sesgos en los datos y toma de decisiones «caja negra», lo cual afecta la confianza pública. Regulación y privacidad: proteger datos personales y derechos humanos ante sistemas automatizados. Alineación de la IA con valores humanos (explicabilidad y transparencia en algoritmos).

- **Machine Learning (ML):**

- *Restricciones:* Necesidad de datos etiquetados o representativos; modelo puede sobre ajustarse a datos de entrenamiento (sobreajuste) o fallar al generalizar. Requiere selección cuidadosa de algoritmos y parámetros, y conocimiento experto.
- *Retos:* Interpretabilidad de modelos complejos (p. ej., redes neuronales profundas). Mitigación de sesgos algorítmicos y equidad. Escalabilidad: entrenar modelos con ultragrandes conjuntos de datos implica demanda computacional y energética. Cumplimiento de privacidad (ML difusos).

- **Minería de Datos (DM):**

- *Restricciones:* Preocupaciones de privacidad y ética en la extracción de información personal. Calidad y heterogeneidad de los datos (datos incompletos, ruidosos o sesgados) reducen la fiabilidad de los patrones extraídos.
- *Retos:* Volumen de datos muy grande (convergencia con Big Data): la minería de datos debe escalar a entornos de Big Data. Defensa contra ataques o manipulaciones de datos (p. ej., en redes sociales). Adaptación a datos en tiempo real y a la “minería” de datos no estructurados (texto, imagen, audio).

- **Big Data:**

- *Restricciones:* Requiere infraestructuras complejas y costosas (almacenamiento distribuido, procesamiento paralelo). Limitaciones técnicas en velocidad de red y procesamiento al escalar aún más los datos. Restricciones legales sobre datos transfronterizos y derechos de privacidad.
- *Retos:* Gobernanza y seguridad: asegurar la calidad y privacidad de datos masivos. Integración de datos de diversa naturaleza (multiformato, alta variabilidad). Mantener actualizadas las tecnologías a ritmos de crecimiento exponencial. Formar talento especializado (escasez de profesionales de Big Data e IA).

Conclusión

La IA, el ML, la Minería de Datos y el Big Data constituyen un ecosistema tecnológico interrelacionado: la IA (campo amplio) se nutre de ML (métodos de aprendizaje) y de Big Data (datos masivos) para extraer conocimiento mediante minería de datos. Los beneficios de estas tecnologías son evidentes en eficiencia, innovación y mejoras en la toma de decisiones. Sin embargo, deben abordarse restricciones técnicas (cómputo, calidad de datos) y desafíos éticos, de privacidad y equidad. Se esperan avances en explicabilidad de los modelos, normas internacionales (e.g. ISO, OCDE) y formación de talento, lo cual impulsará su adopción responsable. A futuro, la convergencia de IA con Big Data probablemente genere sistemas más potentes de análisis predictivo y automatización, siempre y cuando se gestione adecuadamente la ética y regulación.

Referencias

- Organización para la Cooperación y el Desarrollo Económico (OCDE). (2019). *Recomendación del Consejo sobre Inteligencia Artificial: Principios para una IA confiable*. OECD.
- AWS (Amazon Web Services). (s.f.). *¿Qué es el machine learning?* Recuperado de <https://aws.amazon.com/es/what-is/machine-learning/>
- Craig, L., Tucci, L. y Laskowski, N. (2018). *¿Qué es la IA? Una explicación a fondo sobre la inteligencia artificial*. ComputerWeekly.
- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer.
- ISO/IEC. (2019). *Big data — Overview and vocabulary* (ISO/IEC 20546:2019). International Organization for Standardization.
- Fundación Aquae. (2018). *Alan Turing, el padre de la inteligencia artificial*. Aquae Wiki.