

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

Ingeniería en 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:

Giselle Cantú Chávez

NOMBRE DEL DOCENTE:

Ing. Luis Enrique Mascote Cano

Chihuahua, Chih., 26 de septiembre de 2025

ÍNDICE

INTRODUCCIÓN	3
DESARROLLO	5
Definición de Cada Dominio	5
Inteligencia Artificial (IA)	5
Machine Learning (ML)	5
Minería de Datos (DM)	6
Big Data	7
Beneficios Comparados	7
Restricciones y Retos	9
Inteligencia Artificial	9
Machine Learning	9
Minería de Datos	10
Big Data	10
CONCLUSIONES Y PRÓXIMOS PASOS	12
REFERENCIAS Y FUENTES CONSULTADAS	15

ÍNDICE DE FIGURAS

Ilustración 1	11
----------------------------	-----------

INTRODUCCIÓN

En este reporte hablo de cuatro conceptos que son parte del trabajo con información en el mundo actual: Inteligencia Artificial, Machine Learning, Minería de Datos y Big Data. Estos temas están presentes en diferentes campos y ya forman parte de muchas de las soluciones que usamos a diario, aunque a veces no nos damos cuenta. Al estudiarlos de manera organizada puedo entender mejor qué son, cómo surgieron y hacia dónde se dirigen, lo cual es importante dentro de mi carrera.

Decidí enfocarme en estos temas porque hoy en día cualquier profesional en el área de software o de datos se los va a topar en algún momento. No es lo mismo escuchar los términos en clase o en una noticia que sentarse a revisar qué significan de verdad y cómo se aplican. Este trabajo me ayuda a ver con más claridad qué aportan en distintos contextos, qué limitaciones tienen y cuáles son los retos que todavía están pendientes. De esa manera, no me quedo con una idea superficial y puedo dar un punto de vista con más bases cuando me toque hablar de ellos en proyectos o evaluaciones.

El proceso de investigación también me sirve en lo personal porque me obliga a leer con cuidado, a comparar fuentes y a organizar la información de una forma que sea entendible para otros. Sé que en el futuro voy a necesitar explicar ideas técnicas frente a personas que no necesariamente son expertas, y este tipo de prácticas me va preparando. Aquí no solo recopilo información, sino que reviso qué puntos coinciden entre autores, dónde hay diferencias y cómo puedo resumirlo sin perder lo importante. Ese ejercicio fortalece mis habilidades de análisis y de comunicación.

La estructura de este reporte es sencilla. Primero incluyo la definición de cada dominio junto con un breve repaso de su origen y evolución histórica. Después presento los beneficios más importantes de cada uno y los comparo en una tabla que permite verlos de forma rápida. Más adelante hablo de las restricciones que pueden frenar la adopción de estas tecnologías y de los retos actuales que enfrentan en la práctica. Para terminar, redacto unas conclusiones en las que

resumo lo más relevante y menciono el aprendizaje que obtuve al elaborar este trabajo.

Con esta organización busco que el documento sea claro y fácil de seguir. Cada sección cumple un papel distinto: las definiciones ayudan a sentar las bases, los beneficios muestran su utilidad, las restricciones y retos ponen en contexto sus límites, y las conclusiones cierran con una visión general. Así, el reporte cumple con el propósito de la materia y al mismo tiempo se convierte en una experiencia de aprendizaje que me ayuda a crecer como estudiante y como futura profesional en el área de tecnologías de la información.

Otra cosa que rescato de este trabajo es que me obliga a salir de lo fácil. Podría quedarme con definiciones rápidas de internet, pero no tendría sentido. Al buscar fuentes confiables, comparar lo que dicen y armar mis propias conclusiones, aprendo a no conformarme con lo superficial. Eso también me da confianza para hablar de estos temas sin sentir que estoy repitiendo lo que otros dijeron, sino que realmente entendí de qué se trata.

Y, siendo honesta, este tipo de actividades también me ayuda a probar mi disciplina. No siempre es sencillo leer artículos largos o documentos técnicos, pero hacerlo me entrena a no rendirme a la primera y a organizar mi tiempo.

DESARROLLO

Definición de Cada Dominio

Inteligencia Artificial (IA)

La Inteligencia Artificial se entiende como el área que busca crear sistemas capaces de hacer tareas que normalmente requieren de la inteligencia humana, como analizar información, razonar o reconocer patrones. Para Alfonso (2025), la IA es un conjunto de técnicas y modelos que ya influyen en la vida social y económica actual. La **ISO/IEC 22989** la formaliza como sistemas diseñados para producir resultados que reflejen comportamientos inteligentes dentro de un rango definido (ISO, 2022). Dicho en simple, no son máquinas que piensan por sí solas; son programas que logran imitar algunas capacidades que asociamos al pensamiento humano.

El origen de la IA se ubica en 1956, en la Conferencia de Dartmouth, donde investigadores como John McCarthy plantearon por primera vez el término y las primeras preguntas sobre si una máquina podía aprender (Carrillo et al., 2025). Al inicio se usaron reglas lógicas y heurísticas que tenían muchas limitaciones. Con el tiempo, se fueron integrando modelos estadísticos y, en la última década, la explosión de datos y la capacidad de cómputo impulsaron las redes neuronales y el aprendizaje profundo. Hoy ya hablamos de IA generativa, con modelos que crean texto, imágenes y hasta música.

Machine Learning (ML)

El Machine Learning es, en palabras simples, un conjunto de métodos que hacen que un sistema “aprenda” de datos para mejorar su desempeño en una tarea específica. Mitchell (1997) lo define como la capacidad de un programa para aprender de la experiencia. Pickering (2023) agrega que es la base para sistemas

que apoyan diagnósticos en cardiología y otros sectores sensibles, porque son capaces de encontrar patrones que los humanos no detectamos con facilidad. Herramientas como TensorFlow (TensorFlow Team, 2023), PyTorch (The Linux Foundation y PyTorch Contributors, 2025) y Scikit-learn (Scikit-learn developers, 2025) son ejemplos actuales de cómo se materializa el ML en la práctica.

El concepto no es nuevo. En los años cincuenta ya había algoritmos para reconocer patrones simples. En los noventa, con el auge de la Minería de Datos, el ML se consolidó como herramienta útil para negocios y ciencia (Riquelme et al., 2006). La evolución más reciente viene del software libre y las comunidades que crean frameworks abiertos, lo que ha hecho posible que estudiantes, investigadores y empresas pequeñas trabajen con modelos de ML sin tener que construirlos desde cero.

Minería de Datos (DM)

La Minería de Datos se centra en encontrar patrones, tendencias y conocimiento útil dentro de grandes volúmenes de información. Riquelme et al. (2006) la describen como el proceso de descubrir conocimiento en bases de datos. López y Romero (2021) la relacionan con la predicción temprana del rendimiento académico, lo que muestra cómo se aplica fuera del ámbito empresarial. También Menacho et al. (2024) explican que puede reforzar el aprendizaje autónomo en educación superior.

Su origen está en los años ochenta, cuando se buscaba ir más allá del análisis estadístico tradicional y aprovechar las bases de datos crecientes. Herramientas como Weka, de la Universidad de Waikato, ayudaron a estandarizar técnicas y a difundirlas en ámbitos académicos (University of Waikato, 2018). Con el tiempo, la DM se consolidó como parte del ciclo de conocimiento: primero se recopilan datos, luego se procesan y finalmente se transforman en información comprensible y aplicable.

Big Data

El Big Data se refiere al procesamiento de volúmenes de datos que superan las herramientas tradicionales. Hernández et al. (2017) lo definen con las “3V”: volumen, velocidad y variedad. Yaranga et al. (2025) lo destacan por su integración con IA, que potencia la toma de decisiones en sectores clave. No se trata solo de guardar información, sino de manejar flujos masivos en tiempo real para que tengan sentido.

El concepto apareció con fuerza en los 2000, cuando internet, sensores y redes sociales comenzaron a generar cantidades inmensas de datos. Plataformas como Hadoop (Apache Software Foundation, s. f.) y Spark (Apache Software Foundation, 2025) se diseñaron para resolver esos problemas de escalabilidad. Hoy el Big Data se asocia a la Industria 4.0 y se combina con el Business Intelligence y la Inteligencia Artificial, como muestran Pallarés y Acevedo (2025), al hablar de su impacto en la administración estratégica.

Beneficios Comparados

Al revisar los beneficios, se observa que cada dominio aporta cosas distintas, pero todas están relacionadas con el uso más eficiente de la información.

Dominio	Primer beneficio	Segundo beneficio	Tercer beneficio
Inteligencia Artificial	Automatización de procesos complejos (Gómez-Sanchis y Blasco, 2024)	Apoyo en diagnóstico clínico (Carrillo et al., 2025)	Mejora de la experiencia en servicios digitales (Alfonso, 2025)
Machine Learning	Predicción de tendencias (Pickering, 2023)	Reducción de errores manuales (Scikit-learn developers, 2025)	Adaptación constante a nuevos datos (TensorFlow Team, 2023)
Minería de Datos	Identificación de patrones ocultos (Riquelme et al., 2006)	Predicción del rendimiento académico (López y Romero, 2021)	Refuerzo del aprendizaje autónomo (Menacho et al., 2024)
Big Data	Procesamiento masivo de datos (Hernández et al., 2017)	Integración con IA para decisiones estratégicas (Yaranga et al., 2025)	Soporte en la Industria 4.0 (Pallarés y Acevedo, 2025)

Más allá de la tabla, lo interesante es cómo estos beneficios se complementan. La IA brilla en la interacción con el usuario, el ML en aprender patrones, la DM en encontrar conocimiento oculto y el Big Data en darle soporte a todo el ecosistema con infraestructura.

Restricciones y Retos

Inteligencia Artificial

- **Restricciones:**
 - Dependencia de grandes volúmenes de datos de calidad (ISO, 2022).
 - Costos de implementación que no siempre son accesibles (Alfonso, 2025).
- **Retos:**
 - Explicabilidad de modelos en sectores regulados como salud o finanzas (Carrillo et al., 2025).
 - Minimizar sesgos que afectan resultados (Menacho et al., 2024).

Machine Learning

- **Restricciones:**
 - Necesidad de personal con alta preparación técnica (Mitchell, 1997).
 - Resultados sensibles a datos incompletos o sesgados (Pickering, 2023).
- **Retos:**
 - Escalabilidad de modelos en entornos reales (TensorFlow Team, 2023).
 - Reducción del consumo energético en entrenamientos (The Linux Foundation y PyTorch Contributors, 2025).

Minería de Datos

- **Restricciones:**

- Resultados difíciles de interpretar para no expertos (Riquelme et al., 2006).
- Complejidad al integrarse con otros sistemas empresariales (University of Waikato, s. f.).

- **Retos:**

- Asegurar privacidad y ética en el uso de datos (López y Romero, 2021).
- Aplicaciones en educación con recursos limitados (Menacho et al., 2024).

Big Data

- **Restricciones:**

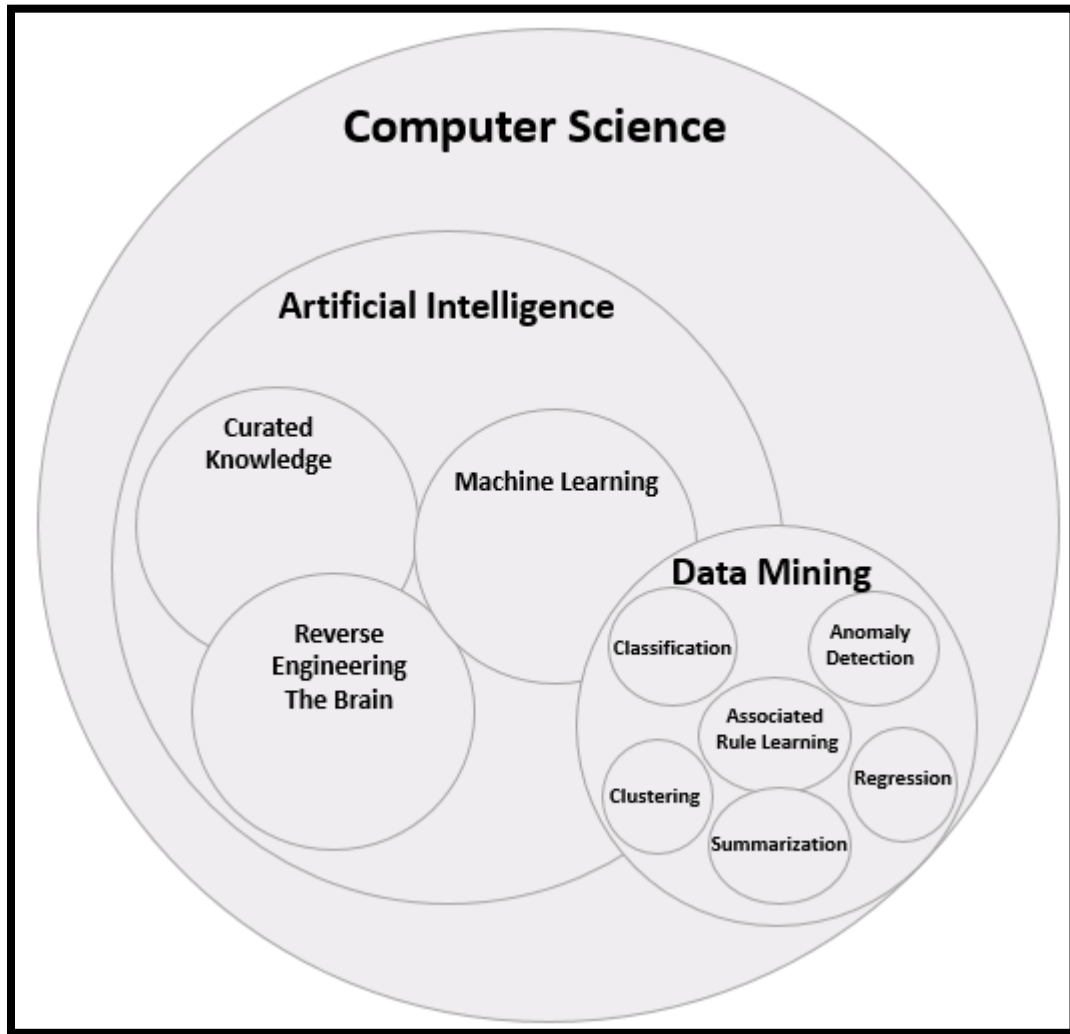
- Costos altos en almacenamiento y procesamiento (Hernández et al., 2017).
- Problemas de latencia en procesamiento en tiempo real (Apache Software Foundation, 2025).

- **Retos:**

- Integración con IA para análisis avanzados en tiempo real (Yaranga et al., 2025).
- Necesidad de gobernanza y estandarización en empresas grandes (Pallarés y Acevedo, 2025).

Ilustración 1

Conceptos dentro de las ciencias computacionales



CONCLUSIONES Y PRÓXIMOS PASOS

El propósito de este reporte fue revisar y comparar cuatro conceptos que ya forman parte de cualquier conversación sobre tecnología: Inteligencia Artificial, Machine Learning, Minería de Datos y Big Data. Desde el inicio la idea era no quedarme solo con definiciones generales, sino analizarlos en conjunto, ver sus ventajas, sus limitaciones y los problemas que todavía tienen pendientes. Después de haber trabajado en cada sección, puedo decir que el planteamiento inicial se cumplió, porque ahora entiendo mejor no solo qué significa cada uno, sino también cómo se relacionan y qué papel juegan en distintos contextos.

Lo primero que puedo concluir es que los cuatro temas no se deben estudiar como si fueran islas separadas. La Inteligencia Artificial es la que da el marco general: busca que las máquinas hagan tareas que asociamos con la mente humana. Dentro de ese marco, el Machine Learning es la vía más usada para lograr que los sistemas aprendan de la experiencia. La Minería de Datos se encarga de descubrir patrones útiles en bases de información, y el Big Data es la base técnica que hace posible trabajar con volúmenes enormes de datos. Ninguno funciona solo. Si pienso en cualquier aplicación moderna, desde un asistente de voz hasta una plataforma de streaming, en realidad hay un poco de cada uno de estos elementos trabajando juntos.

En cuanto a beneficios, no queda duda de que la IA ha marcado un antes y un después en muchos sectores. La automatización de procesos, el apoyo en diagnósticos clínicos y la personalización de experiencias son logros que ya se ven todos los días. El ML aporta la capacidad de predecir tendencias y adaptarse cuando cambian los datos, algo que lo hace ideal para negocios o proyectos donde la información nunca se detiene. La Minería de Datos destaca porque convierte información dispersa en conocimiento que sí sirve para tomar decisiones, desde predecir el rendimiento académico hasta entender mejor a los clientes de una empresa. El Big Data, por su parte, garantiza que todo este ecosistema pueda

manejar volúmenes gigantes de datos en tiempo real. Sin estas bases, ninguna de las otras áreas tendría el mismo alcance.

Pero no todo son ventajas, y esa es otra de las conclusiones importantes. La IA todavía depende demasiado de contar con datos de calidad y requiere infraestructura costosa que no cualquiera puede pagar. El ML necesita equipos de personas con alta preparación técnica y es vulnerable a datos incompletos o sesgados. La Minería de Datos puede ser complicada de interpretar para quienes no son expertos, lo que a veces limita su utilidad. Y el Big Data es caro de mantener y presenta problemas de latencia en tiempo real. Estas limitaciones muestran que, aunque las tecnologías son muy poderosas, no son mágicas: requieren inversión, planeación y un contexto adecuado para que funcionen de verdad.

Los retos actuales también son claros. En IA, el problema de la explicabilidad y el sesgo sigue sin resolverse del todo. En ML, la escalabilidad y la eficiencia energética son asuntos que marcan la agenda. En la Minería de Datos, la privacidad y la ética son temas que no se pueden dejar de lado en un mundo donde cada clic genera información. Y en Big Data, la gobernanza y la estandarización se vuelven urgentes cuando se trata de integrar datos de múltiples fuentes. Estos retos me dejan ver que la tecnología no está terminada, que siempre habrá ajustes y mejoras, y que quienes trabajemos en este campo tenemos mucho que aportar.

Una parte importante de esta conclusión es reconocer que el aprendizaje va más allá de lo técnico. Sí, ahora tengo más claras las definiciones, los beneficios y las limitaciones, pero también aprendí a investigar de manera más crítica. No me conformé con lo primero que aparece en internet; busqué artículos, tesis, manuales y documentos que me dieran información confiable. Eso me obligó a leer con calma, a comparar lo que decía cada fuente y a organizar las ideas de una manera que se entendiera bien. Esa práctica es parte del aprendizaje que me llevo: investigar, analizar y comunicar con orden.

Otra cosa que me queda es la importancia de ver estos temas con responsabilidad. Hablar de datos no es algo neutro: detrás hay personas, decisiones y consecuencias. Si un modelo tiene sesgo, puede discriminar. Si un sistema de Big

Data se usa mal, puede poner en riesgo la privacidad. Estas conclusiones me recuerdan que no basta con saber programar o entrenar modelos; también tengo que pensar en el impacto que tendrán en la vida real. Esa mirada crítica es lo que va a marcar la diferencia en mi formación como profesional.

Al final, este trabajo me hizo crecer en dos sentidos. Por un lado, en lo académico, porque ahora puedo explicar con mayor seguridad qué es cada concepto y cómo se aplican. Por otro lado, en lo personal, porque reforcé la disciplina de leer, resumir y escribir con claridad. Sé que este tipo de habilidades son igual de valiosas que aprender un nuevo lenguaje de programación o manejar un software.

REFERENCIAS Y FUENTES CONSULTADAS

- Alfonso Pomares, F. A., Reyes Baños, R. L., y Cruz Cruz, Y. (2025). La Inteligencia Artificial y su impacto en la sociedad actual. *Universidad & Ciencia*, 14(1), 228–242. Universidad Autónoma de Campeche.
<https://revistavirtual.ucam.mx/index.php/ciencia/article/view/8668>
- Altair. (s. f.). *Altair RapidMiner: Documentación del producto*. Recuperado el 25 de septiembre de 2025 de <https://docs.rapidminer.com/>
- Apache Software Foundation. (s. f.). *Apache Hadoop*. Recuperado el 25 de septiembre de 2025 de <https://hadoop.apache.org/>
- Apache Software Foundation. (2025). *Apache Kafka: Introduction*. Recuperado el 25 de septiembre de 2025 de <https://kafka.apache.org/intro>
- Apache Software Foundation. (2025). *Apache Spark: Structured Streaming programming guide (v3.5.1)*. Recuperado el 25 de septiembre de 2025 de <https://spark.apache.org/docs/3.5.1/structured-streaming-programming-guide.html>
- Apache Software Foundation y Proyecto Flink. (s. f.). *Apache Flink: Learn Flink (overview)*. Recuperado el 25 de septiembre de 2025 de <https://nightlies.apache.org/flink/flink-docs-stable/docs/learn-flink/overview/>
- Carrillo Fernández, A. M., Delgado Casana, J., y Castro Martín, A. (2025). Language models and generative artificial intelligence for teaching resolutions in traumatic dental emergencies. *Revista Médica de Salud Escolar y Universitaria*, 3, e-RMS14022025.
<https://ve.scielo.org/pdf/erms/v3/2960-2467-erms-3-e-RMS14022025.pdf>
- Gómez-Sanchis, J., y Blasco, J. (2024). Inteligencia Artificial en la Agricultura: Retos, oportunidades y estrategias para su adopción. En *Guía práctica de inteligencia artificial. Aplicaciones de vanguardia en sectores estratégicos*

(pp. 25–33). Fundació Parc Científic Universitat de València. ISBN: 978-84-09-67015-4

- Hernández Leal, E. J., Ponce Hernández, F. J., y Sánchez Tovar, Y. (2017). Big Data: una exploración de investigaciones, tecnologías y aplicaciones. *Revista Iberoamericana de las Ciencias Computacionales e Informática*, 6(11). <https://www.redalyc.org/journal/3442/344251476001/html/>
- López Zambrano, J., Lara Torralbo, J. A., y Romero, C. (2021). Early prediction of student learning performance through data mining: A systematic review. *Psicothema*, 33(3), 456–465. <https://doi.org/10.7334/psicothema2021.62>
- Menacho Ángeles, M. R., Cámara Vargas, J. J., y Huamán León, R. J. (2024). Inteligencia artificial como herramienta en el aprendizaje autónomo en educación superior. *Revista de Investigación en Competencias*, 5(2). https://ve.scielo.org/scielo.php?pid=S2739-00632024000200158&script=sci_arttext
- Pallarés Vega, L. A., y Acevedo Durán, D. Y. (2025). Integración de Business Intelligence, Machine Learning y Big Data en la Industria 4.0: Impacto en la administración estratégica y la toma de decisiones empresariales [Tesis de especialización, Universidad Nacional Abierta y a Distancia UNAD]. Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/73478>
- Riquelme, J. C., Ruiz, R., y Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 10(29), 11–18. <https://www.redalyc.org/pdf/925/92502902.pdf>
- Scikit-learn developers. (2025). *User Guide — scikit-learn 1.7.2 documentation*. Recuperado el 25 de septiembre de 2025 de https://scikit-learn.org/stable/user_guide.html
- TensorFlow Team. (2023). *TensorFlow Core: Guide*. Google. Recuperado el 25 de septiembre de 2025 de <https://www.tensorflow.org/guide>

The Linux Foundation y PyTorch Contributors. (2025). *PyTorch: Documentation & Tutorials*. Recuperado el 25 de septiembre de 2025 de <https://docs.pytorch.org/>

University of Waikato. (2018). *Weka Manual for version 3-8-3*. Waikato Environment for Knowledge Analysis. <https://user.eng.umd.edu/~austin/ence688p.d/handouts/WekaManual2018.pdf>

University of Waikato. (s. f.). *Weka 3 — Data Mining with Open Source Machine Learning Software*. Recuperado el 25 de septiembre de 2025 de <https://www.cs.waikato.ac.nz/ml/weka/>

Yaranga Vite, I. P., Rodríguez, F. F., y Delgado, M. (2025). Integración de la inteligencia artificial con Big Data para la mejora de la toma de decisiones en sectores clave: Una revisión sistemática. *Revista de Investigación en Competencias*, 5(4), e504047. <https://ve.scielo.org/pdf/ric/v5n4/2739-0063-ric-5-04-e504047.pdf>