

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

Ingeniería en Desarrollo y Gestión de Software



Extracción de Conocimiento en Bases de Datos

I.3. Reporte de investigación de los tipos de aplicaciones, procesamiento y herramientas para inteligencia artificial, machine learning, data mining y big data

IDGS91N

PRESENTA:

Giselle Cantú Chávez

NOMBRE DEL DOCENTE:

Ing. Luis Enrique Mascote Cano

Chihuahua, Chih., 25 de septiembre de 2025

ÍNDICE

INTRODUCCIÓN	3
DESARROLLO	5
Inteligencia Artificial	5
<i>Tipos de Aplicaciones</i>	5
<i>Procesamiento en IA: Batch vs. Streaming</i>	6
<i>Herramientas y Tecnologías de la IA</i>	8
Machine Learning	9
<i>Tipos de aplicaciones</i>	10
<i>Procesamiento en ML</i>	10
<i>Herramientas y tecnologías</i>	11
Data Mining (DM).....	11
<i>Tipos de aplicaciones</i>	11
<i>Procesamiento en DM</i>	12
<i>Herramientas y Tecnologías</i>	12
Big Data	13
<i>Tipos de aplicaciones</i>	13
<i>Procesamiento en Big Data</i>	13
<i>Herramientas y tecnologías</i>	14
CONCLUSIONES Y PRÓXIMOS PASOS	16
REFERENCIAS Y FUENTES CONSULTADAS	18

ÍNDICE DE FIGURAS

Ilustración 1	16
----------------------------	-----------

INTRODUCCIÓN

Cuando escuchamos hablar de inteligencia artificial, machine learning, minería de datos y big data, pareciera que son términos distintos pero que al mismo tiempo se entrelazan de manera inevitable. El tema principal de este reporte es precisamente explorar cómo se conectan y diferencian estos cuatro dominios, identificando aplicaciones concretas, modalidades de procesamiento y las herramientas que más se utilizan en la práctica. Me interesa analizar no solo la teoría detrás de cada uno, sino también el impacto real que tienen en distintos sectores, desde la educación hasta la salud y la industria, porque hoy en día estas tecnologías ya no son opcionales, sino parte del día a día de la transformación digital.

Realizar este reporte me resulta pertinente porque me permite entender con mayor claridad en qué contextos se usa cada tecnología y cómo puedo aprovechar ese conocimiento en proyectos futuros. No basta con saber que existe la inteligencia artificial o el big data: lo verdaderamente importante es comprender para qué se emplea cada enfoque, cuándo conviene trabajar con procesamiento batch o streaming, y qué tipo de herramientas me ofrecen más ventajas según el problema que busque resolver. Esta reflexión es clave para no quedarme en la superficie, sino para desarrollar criterio propio en la toma de decisiones técnicas.

En lo personal, considero que este proyecto contribuye a mi formación profesional de manera directa, porque me obliga a trabajar con fuentes académicas y técnicas confiables, a organizar la información de forma crítica y a familiarizarme con herramientas que muy probablemente tendré que dominar en el campo laboral. Al analizar casos de uso reales y tecnologías que ya son estándar en la industria — como TensorFlow, PyTorch, Apache Spark o RapidMiner— voy construyendo un puente entre lo que aprendo en la universidad y lo que se exige en el mundo profesional. Este proceso también me ayuda a fortalecer mis habilidades de investigación y redacción, que son indispensables en el desarrollo de software y en cualquier área de TI.

El contenido de este reporte está organizado en cuatro capítulos principales. En el primero, me enfoco en la inteligencia artificial, revisando sus aplicaciones, los modos de procesamiento que la sustentan y las herramientas más utilizadas. En el segundo, abordo el machine learning, resaltando su relación con la IA pero también su identidad propia como disciplina enfocada en el aprendizaje de los algoritmos. El tercer capítulo lo dedico a la minería de datos, una práctica que ha evolucionado de los métodos estadísticos clásicos hacia enfoques mucho más automatizados y predictivos. Por último, en el cuarto capítulo, analizo el big data, sus aplicaciones en entornos donde el volumen y la velocidad de la información son críticos, y las tecnologías que lo hacen posible. La conclusión integra una reflexión comparativa de los cuatro dominios y la importancia de elegir la herramienta y el modo de procesamiento adecuados.

DESARROLLO

Inteligencia Artificial

La inteligencia artificial (IA) la entiendo como la rama de la informática que busca que las máquinas realicen tareas que normalmente requieren de la inteligencia humana, como el razonamiento, el aprendizaje, la percepción o la toma de decisiones. No se trata únicamente de automatizar procesos, sino de diseñar sistemas capaces de adaptarse y mejorar con la experiencia, tal como lo mencionan Menacho et al. (2024), quienes destacan su papel como herramienta de apoyo al aprendizaje autónomo. En la práctica, la IA se convierte en una tecnología transversal que permea sectores tan distintos como la salud, la educación, la industria o el comercio.

Tipos de Aplicaciones

- ✓ **Visión por computador.** La visión por computador permite que las máquinas analicen e interpreten imágenes y videos de manera automática. Este tipo de aplicación se utiliza para identificar patrones visuales, detectar objetos o incluso diagnosticar enfermedades a partir de estudios médicos. Un caso real es el uso de IA en radiología para la detección temprana de tumores, donde algoritmos de visión superan en precisión a evaluaciones humanas en ciertas condiciones (Pickering, 2023). Empresas como Google también aplican esta tecnología en Google Photos, donde se reconoce y clasifica automáticamente a las personas y objetos en las imágenes.
- ✓ **Procesamiento del lenguaje natural (PLN).** El procesamiento del lenguaje natural busca que los sistemas entiendan, interpreten y generen lenguaje humano. Esto se refleja en asistentes virtuales, chatbots o traductores automáticos. Carrillo Fernández, Delgado Casana y Castro Martín (2025) resaltan que los modelos de lenguaje generativo se están utilizando incluso

en contextos médicos, como apoyo en la resolución de emergencias. En la vida cotidiana, lo vemos en aplicaciones como Siri o Alexa, que nos permiten interactuar con dispositivos mediante comandos de voz.

- ✓ **Sistemas de recomendación.** Los sistemas de recomendación son otra aplicación fundamental de la IA, ya que personalizan la experiencia del usuario analizando su comportamiento y preferencias. Empresas como Netflix o Spotify utilizan estos sistemas para sugerir películas, series o canciones basadas en el historial de cada persona. Yaranga Vite, Rodríguez y Delgado (2025) destacan que la integración de IA con big data fortalece precisamente este tipo de aplicaciones, ya que la recomendación depende del análisis de grandes volúmenes de información en tiempo real.

Procesamiento en IA: Batch vs. Streaming

Cuando pienso en cómo se procesan los datos en inteligencia artificial, me doy cuenta de que no siempre se hace de la misma manera. En muchos casos se trabaja en **modo batch**, que significa procesar grandes volúmenes de información en bloques, de forma acumulada y no en tiempo real. Este enfoque resulta muy útil cuando se entrenan modelos de machine learning o redes neuronales profundas, porque se necesita trabajar con conjuntos de datos masivos que se organizan y analizan en etapas. Así lo reflejan Hernández Leal, Ponce Hernández y Sánchez Tovar (2017), quienes destacan que el big data y la IA se apoyan en este tipo de procesamiento para generar conocimiento a partir de enormes volúmenes de información estructurada y no estructurada.

En cambio, el **procesamiento en streaming** se emplea cuando la prioridad es obtener resultados inmediatos. Aquí, los datos se analizan en el momento en que llegan, sin esperar a acumularlos en lotes. Apache Spark, por ejemplo, ofrece Structured Streaming para que los mismos algoritmos puedan trabajar tanto en

batch como en tiempo real sin duplicar la lógica de programación (Apache Software Foundation, 2025). Esta capacidad me parece esencial en aplicaciones como la detección de fraudes en transacciones financieras, donde una alerta debe emitirse en cuestión de segundos.

Algo similar ocurre con **Apache Kafka**, que se ha convertido en la base de muchas arquitecturas de IA en tiempo real gracias a su capacidad de manejar flujos de mensajes distribuidos a gran escala (Apache Software Foundation, 2025). Empresas de comercio electrónico lo usan para procesar millones de eventos de usuarios en sus plataformas y alimentar sistemas de recomendación en vivo.

Además, tecnologías como **Apache Flink** fueron creadas específicamente para ejecutar algoritmos sobre flujos continuos de datos. Su fortaleza está en la baja latencia, lo que lo hace ideal para monitoreo de sensores IoT o análisis de logs en sistemas críticos (Apache Software Foundation y Proyecto Flink, s. f.).

En mi caso, me queda claro que elegir entre batch o streaming no es una cuestión trivial. Para entrenar un modelo de visión por computador, como los usados en radiología (Pickering, 2023), conviene un esquema batch que permita trabajar con miles de imágenes en cada ciclo. Pero si hablamos de un sistema de recomendación que debe actualizarse en el mismo instante en que cambio de preferencia musical, se necesita streaming con herramientas como Kafka o Flink. Incluso, como señalan Yaranga Vite, Rodríguez y Delgado (2025), la integración de IA con big data depende de que ambos enfoques se combinen estratégicamente, porque muchas soluciones necesitan tanto análisis histórico en batch como respuestas inmediatas en streaming.

Herramientas y Tecnologías de la IA

Al revisar las tecnologías que sostienen a la inteligencia artificial, me doy cuenta de que existe un ecosistema muy amplio, pero hay algunas herramientas que se han convertido en referentes por su popularidad y capacidad de adaptación.

1. TensorFlow

Una de las más conocidas es TensorFlow, desarrollada por Google. Su fortaleza está en permitir construir y entrenar redes neuronales profundas a gran escala, tanto en CPU como en GPU. Según el propio equipo de TensorFlow (2023), la plataforma está diseñada para aplicaciones de visión por computador, procesamiento de lenguaje natural y sistemas de recomendación. En la práctica, empresas como Airbnb la utilizan para predecir patrones de demanda y mejorar su sistema de precios dinámicos.

2. PyTorch

Otra herramienta que se ha vuelto esencial es PyTorch, impulsada por la comunidad de desarrolladores y por Meta. La ventaja que le encuentro es su flexibilidad: ofrece un entorno más dinámico para prototipar modelos, lo que facilita a los investigadores probar rápidamente nuevas arquitecturas. The Linux Foundation y los colaboradores de PyTorch (2025) explican que esta librería se usa ampliamente en visión por computador y NLP. Un ejemplo es Tesla, que ha empleado PyTorch para entrenar sus sistemas de conducción autónoma.

3. Scikit-learn

Aunque no está enfocada en deep learning, scikit-learn me parece una librería fundamental en IA porque ofrece algoritmos clásicos de machine learning: clasificación, regresión, clustering y reducción de dimensionalidad. Los

desarrolladores de scikit-learn (2025) destacan su facilidad de integración con Python y otras librerías como NumPy o Pandas. Se utiliza en ámbitos como el análisis predictivo de datos de salud, por ejemplo, para anticipar el riesgo de enfermedades crónicas a partir de historiales médicos.

4. RapidMiner

Además de librerías de código abierto, hay plataformas como RapidMiner, que integran minería de datos, machine learning y despliegue de modelos en un entorno visual. Altair (s. f.) documenta que RapidMiner permite diseñar flujos de trabajo completos sin necesidad de programar en detalle, lo cual es útil para equipos multidisciplinarios. Lo he visto aplicado en proyectos de retail para analizar patrones de compra y segmentar clientes.

5. Weka

Finalmente, una herramienta con mucha tradición en investigación es Weka, desarrollada por la Universidad de Waikato. Según su manual oficial (University of Waikato, 2018), Weka reúne algoritmos de machine learning que se pueden aplicar fácilmente sobre conjuntos de datos para tareas como clasificación y clustering. En educación, López Zambrano, Lara Torralbo y Romero (2021) destacan cómo Weka ha sido usado en revisiones sistemáticas de predicción de rendimiento estudiantil.

Machine Learning

El machine learning es una rama de la IA que permite a las computadoras aprender de los datos y mejorar su rendimiento sin ser programadas explícitamente. Pickering (2023) describe cómo los algoritmos de ML han cambiado la toma de decisiones en cardiología, mostrando su relevancia en sectores críticos.

Tipos de aplicaciones

1. Clasificación y regresión

En ML, la clasificación se usa para asignar etiquetas y la regresión para predecir valores continuos. Scikit-learn facilita estos procesos y se ha usado en predicciones médicas de enfermedades crónicas (Scikit-learn developers, 2025).

2. Aprendizaje supervisado en educación

López Zambrano, Lara Torralbo y Romero (2021) explican que los algoritmos supervisados permiten predecir el rendimiento estudiantil a partir de datos previos, contribuyendo a prevenir la deserción escolar.

3. Aprendizaje profundo

El deep learning, como subcampo del ML, es la base de tecnologías como el reconocimiento facial. TensorFlow se usa ampliamente en este tipo de modelos (TensorFlow Team, 2023).

Procesamiento en ML

El entrenamiento de modelos suele hacerse en **batch**, ya que se necesitan volúmenes grandes de datos organizados. Hernández Leal, Ponce Hernández y Sánchez Tovar (2017) señalan que los conjuntos masivos procesados por lotes son esenciales para construir predicciones robustas.

En contraste, el **streaming** se aplica cuando un modelo debe actualizarse en tiempo real. PyTorch es una herramienta flexible que permite adaptarse a este enfoque, especialmente en vehículos autónomos donde las decisiones deben ser instantáneas (The Linux Foundation y PyTorch Contributors, 2025).

Herramientas y tecnologías

- **Scikit-learn**: para algoritmos básicos y medianos, muy usado en análisis predictivo (Scikit-learn developers, 2025).
- **TensorFlow**: en aplicaciones de deep learning para visión e interpretación de lenguaje (TensorFlow Team, 2023).
- **PyTorch**: popular en investigación por su dinamismo en la construcción de modelos (The Linux Foundation y PyTorch Contributors, 2025).
- **Weka**: sigue siendo referencia en investigación académica para la enseñanza de algoritmos (University of Waikato, s. f.).
- **RapidMiner**: empleado en negocios para análisis predictivo de ventas y segmentación (Altair, s. f.).

Data Mining (DM)

La minería de datos consiste en descubrir patrones, relaciones y conocimiento útil a partir de grandes conjuntos de datos. Riquelme, Ruiz y Gilbert (2006) describen este campo como un puente entre estadísticas tradicionales y nuevas técnicas automatizadas.

Tipos de aplicaciones

1. Segmentación de clientes

RapidMiner facilita identificar patrones de compra y agrupar clientes por perfiles (Altair, s. f.).

2. Predicción en educación

López Zambrano, Lara Torralbo y Romero (2021) muestran cómo se aplica la minería de datos para anticipar el desempeño académico.

3. Detección de anomalías

Weka permite aplicar clustering para detectar irregularidades en datos financieros (University of Waikato, 2018).

Procesamiento en DM

El análisis en **batch** es común porque muchos proyectos de minería de datos requieren limpiar, transformar y procesar información histórica (Riquelme, Ruiz y Gilbert, 2006).

Sin embargo, con la llegada de big data, el **streaming** también se integra. Apache Spark facilita la aplicación de algoritmos de minería directamente sobre flujos de datos, como ocurre en monitoreo de sensores industriales (Apache Software Foundation, 2025).

Herramientas y Tecnologías

- **RapidMiner:** enfocado en minería visual y despliegue de modelos (Altair, s. f.).
- **Weka:** muy utilizado en investigación y educación (University of Waikato, s. f.).
- **Scikit-learn:** provee algoritmos de clustering y reducción de dimensionalidad (Scikit-learn developers, 2025).
- **Apache Spark:** ejecuta minería de datos sobre flujos masivos (Apache Software Foundation, 2025).
- **Hadoop:** almacena y organiza la información que luego se explota con minería (Apache Software Foundation, s. f.).

Big Data

El big data se refiere al manejo de volúmenes de información tan grandes, rápidos y variados que requieren tecnologías especiales para ser procesados. Hernández Leal, Ponce Hernández y Sánchez Tovar (2017) destacan que este campo implica tres características principales: volumen, velocidad y variedad.

Tipos de aplicaciones

1. **Análisis en tiempo real de redes sociales**

Kafka permite procesar millones de mensajes y publicaciones al instante, lo que ayuda a empresas a conocer tendencias y opiniones (Apache Software Foundation, 2025).

2. **Predicción de demanda energética**

Spark se usa en proyectos de smart cities para analizar el consumo eléctrico y ajustar la producción en tiempo real (Apache Software Foundation, 2025).

3. **Optimización en logística**

El big data analiza rutas y tiempos de entrega, permitiendo que empresas de transporte reduzcan costos. Yaranga Vite, Rodríguez y Delgado (2025) enfatizan que la combinación de IA y big data potencia la toma de decisiones estratégicas.

Procesamiento en Big Data

El **batch** se utiliza para análisis históricos, como ocurre en Hadoop, donde los datos se procesan en bloques almacenados (Apache Software Foundation, s. f.).

El **streaming**, en cambio, es esencial cuando los datos se generan sin pausa. Flink se centra en estos escenarios y ha sido adoptado en proyectos de monitoreo

de sensores IoT por su baja latencia (Apache Software Foundation y Proyecto Flink, s. f.).

Herramientas y tecnologías

- **Apache Hadoop:** sistema para almacenar y procesar grandes volúmenes de datos (Apache Software Foundation, s. f.).
- **Apache Spark:** combina batch y streaming en un mismo entorno (Apache Software Foundation, 2025).
- **Apache Kafka:** especializado en el manejo de flujos de eventos distribuidos (Apache Software Foundation, 2025).
- **Apache Flink:** diseñado para análisis en tiempo real con baja latencia (Apache Software Foundation y Proyecto Flink, s. f.).
- **RapidMiner:** integra big data en su flujo de minería y aprendizaje automático (Altair, s. f.).

Ilustración 1

Ejemplo de código



CONCLUSIONES Y PRÓXIMOS PASOS

Al finalizar este reporte, confirmo que los cuatro dominios estudiados —inteligencia artificial, machine learning, minería de datos y big data— comparten una base común pero también presentan diferencias muy marcadas en cuanto a aplicaciones, modalidades de procesamiento y herramientas. La investigación me permitió responder a la pregunta central del trabajo: ¿cómo se aplican y con qué recursos técnicos se desarrollan estas tecnologías en la práctica? La respuesta es clara: cada una cumple una función específica, pero todas se complementan y, en conjunto, forman el núcleo de la transformación digital actual.

Pude comprobar que la inteligencia artificial tiene un papel transversal en múltiples sectores, con aplicaciones que van desde la visión por computador hasta los sistemas de recomendación. El machine learning, aunque forma parte de la IA, se distingue porque centra sus esfuerzos en que los algoritmos aprendan a partir de datos y mejoren progresivamente. La minería de datos, por su parte, representa la evolución del análisis estadístico hacia un enfoque más automatizado que busca patrones y conocimientos útiles en grandes volúmenes de información. Finalmente, el big data demuestra ser el cimiento que hace posible manejar tanto datos históricos en batch como flujos de información en streaming, integrando tecnologías que aseguran la escalabilidad y el rendimiento.

Otro hallazgo importante fue reconocer que no existe un único tipo de procesamiento que resuelva todas las necesidades. Aprendí que el batch es ideal para entrenamientos, análisis históricos y escenarios donde se prioriza la precisión sobre la inmediatez, mientras que el streaming resulta indispensable en entornos donde las decisiones deben tomarse en segundos, como en la banca o el monitoreo de sensores. Comprender estas diferencias me ayuda a visualizar cuándo y cómo aplicar cada enfoque.

En lo personal, elaborar este trabajo representó una oportunidad para fortalecer mis habilidades de investigación y análisis crítico. No solo me acerqué a la teoría, sino que también entendí mejor cómo las herramientas oficiales como

TensorFlow, PyTorch, Spark o RapidMiner son utilizadas en casos reales por empresas y proyectos. Este aprendizaje me deja claro que mi formación profesional requiere de una visión amplia y a la vez práctica, porque en el futuro tendré que elegir las tecnologías más adecuadas según las necesidades de cada proyecto.

Por último, considero que este reporte me permitió integrar la teoría con la práctica de una forma mucho más clara. Al estudiar ejemplos concretos de aplicaciones y contrastarlos con los conceptos de batch y streaming, me di cuenta de que puedo llevar este conocimiento a mis propios proyectos académicos y profesionales. Esa es la ganancia más grande: tener la confianza de que puedo analizar tecnologías complejas, entender su propósito y decidir cómo aprovecharlas estratégicamente en mi carrera.

REFERENCIAS Y FUENTES CONSULTADAS

- 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>
- 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
- Pickering, J. W. (2023). Aprendizaje automático para la toma de decisiones en cardiología: Una revisión narrativa para ayudar a recorrer el nuevo panorama. *Revista Española de Cardiología*.
<https://doi.org/10.1016/j.recesp.2023.01.006>
- Riquelme, J. C., Ruiz, R., y Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. *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>