

**UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA**

**TECNOLOGÍAS DE LA INFORMACIÓN**



**EXTRACCIÓN DE CONOCIMIENTO EN BASES DE DATOS**

**REPORTE DE INVESTIGACIÓN DE LOS TIPOS DE  
APLICACIONES, PROCESAMIENTO Y HERRAMIENTAS  
PARA INTELIGENCIA ARTIFICIAL, MACHINE LEARNING,  
DATA MINING Y BIG DATA.**

***IDGS81N***

**PRESENTA:**

**REGINA CHÁVEZ TAMAYO**

**DOCENTE:**

**LUIS ENRIQUE MASCOTE CANO**

**Chihuahua, Chih., 25 de septiembre de 2025**

## Índice

|                                    |   |
|------------------------------------|---|
| Introducción .....                 | 3 |
| Inteligencia Artificial (IA) ..... | 3 |
| Tipos de aplicaciones .....        | 3 |
| Procesamiento.....                 | 3 |
| Herramientas y tecnologías.....    | 4 |
| Machine Learning (ML) .....        | 4 |
| Tipos de aplicaciones .....        | 4 |
| Procesamiento.....                 | 4 |
| Herramientas y tecnologías.....    | 5 |
| Data Mining (DM).....              | 5 |
| Tipos de aplicaciones .....        | 5 |
| Procesamiento.....                 | 5 |
| Herramientas y tecnologías.....    | 5 |
| Big Data.....                      | 6 |
| Tipos de aplicaciones .....        | 6 |
| Procesamiento.....                 | 6 |
| Herramientas y tecnologías.....    | 6 |
| Conclusión .....                   | 7 |
| Referencias.....                   | 8 |

## Introducción

En la práctica, Inteligencia Artificial, Machine Learning, Data Mining y Big Data suelen convivir en los mismos proyectos, pero no son sinónimos. La IA describe sistemas capaces de percibir, razonar y actuar; ML es el conjunto de técnicas estadísticas/computacionales que aprenden patrones a partir de datos; DM se centra en descubrir conocimiento accionable en conjuntos de datos (a menudo con flujos guiados y análisis exploratorio); y Big Data aborda el procesamiento/almacenamiento distribuido de datos masivos y veloces. Este reporte resume, para cada dominio, aplicaciones frecuentes, cuándo usar procesamiento batch vs streaming, y las herramientas más usadas en la industria.

## Inteligencia Artificial (IA)

La Inteligencia Artificial es una rama de la informática que busca diseñar sistemas capaces de imitar ciertas capacidades humanas, como percibir el entorno, razonar, aprender y tomar decisiones. Su objetivo es resolver problemas complejos de forma autónoma o semiautónoma, aplicando algoritmos que permiten interpretar información y actuar en consecuencia. Un ejemplo claro es el reconocimiento facial, donde el sistema identifica personas a partir de imágenes o videos.

## Tipos de aplicaciones

1. **Visión por computador:** clasificación/detección/segmentación de imágenes y video (por ejemplo, conteo automático de inventario en cámaras de almacén).
2. **Procesamiento de Lenguaje Natural (PLN):** análisis de sentimientos, NER, chatbots (por ejemplo, asistente para responder FAQs de una dependencia pública).
3. **Sistemas conversacionales/Agentes:** flujos de conversación e integración con APIs para realizar tareas (por ejemplo, agente que agenda citas y consulta sistemas internos).
4. **Robótica/Autonomía:** percepción y planificación para movilidad o manipulación (por ejemplo, robots móviles de entrega en campus industriales).

## Procesamiento

- **Batch:** entrenamientos de modelos, evaluación fuera de línea, generación de embeddings a gran escala. Útil cuando los datos se acumulan y el costo es alto (entrenar un modelo de visión/PLN).

- **Streaming:** inferencia en tiempo real (por ejemplo, moderar texto al vuelo, detectar objetos en video en vivo) y actualización continua de señales/colas de eventos.

## Herramientas y tecnologías

- **TensorFlow:** plataforma de ML “end-to-end”, con Keras, despliegue en servidores/móviles y ecosistema maduro. Ideal para prototipos a producción.
- **PyTorch:** librería de tensores y entrenamiento flexible, muy usada en investigación e industria; integra GPU/CPU y ecosistema amplio.
- **Hugging Face Transformers:** modelos SOTA (BERT, GPT-like, ViT, etc.) listos para inferencia/ajuste fino en visión/PLN/audio.
- **OpenCV:** biblioteca líder en visión, con cientos de algoritmos de procesamiento de imagen/video (pre/post-proceso alrededor de redes).
- **ONNX Runtime:** motor de inferencia de alto rendimiento para modelos exportados a ONNX (acelera CPU/GPU y edge). Útil para producción multi-framework.

## Machine Learning (ML)

El Aprendizaje Automático es un subcampo de la IA que se centra en la creación de modelos capaces de aprender patrones a partir de datos históricos, sin ser programados de manera explícita para cada tarea. A través de métodos estadísticos y computacionales, los sistemas de ML pueden predecir, clasificar o detectar comportamientos futuros. Por ejemplo, en plataformas de streaming se utilizan modelos de ML para recomendar contenido personalizado en función de lo que cada usuario consume.

## Tipos de aplicaciones

1. Clasificación/regresión supervisada (por ejemplo, predicción de demanda).
2. Clustering/segmentación (por ejemplo, agrupar ciudadanos por patrones de servicio).
3. Sistemas de recomendación (por ejemplo, contenido o trámites sugeridos).
4. Detección de anomalías (por ejemplo, fraude transaccional o fallas en sensores).

## Procesamiento

- **Batch:** entrenamiento inicial, re-entrenos periódicos, validación y selección de modelos.
- **Streaming:** scoring en tiempo real (p. ej., calcular probabilidad de fraude al recibir una transacción), actualizaciones incrementales de características.

## Herramientas y tecnologías

- scikit-learn, una biblioteca en Python para ML clásico (supervisado/ no supervisado, selección de modelos, pipelines). Excelente para tabulares.
- TensorFlow (Keras) y PyTorch para modelos profundos (visión/PLN, series de tiempo).
- Hugging Face Transformers para transfer learning en PLN/visión.
- ONNX Runtime para inferencia rápida en producción (reduce latencias/costos).

## Data Mining (DM)

La Minería de Datos es el proceso de explorar grandes conjuntos de información con el fin de descubrir patrones, tendencias o relaciones significativas que no son evidentes a simple vista. Se apoya en técnicas de estadística, aprendizaje automático y bases de datos para transformar datos en conocimiento útil. Es ampliamente utilizada en áreas como marketing, donde permite identificar segmentos de clientes y anticipar sus preferencias de compra.

### Tipos de aplicaciones

- Segmentación/Clustering (por ejemplo, perfiles de usuarios para campañas).
- Reglas de asociación (por ejemplo, artículos comprados juntos).
- Minería de texto (tópicos, palabras clave, NER) para bases documentales.

### Procesamiento

- **Batch:** exploración y construcción de flujos (ETL + modelado + validación).
- **Streaming:** menos común en DM tradicional, pero posible con conectores a colas/eventos para refrescar tableros o detectar outliers en continuo.

## Herramientas y tecnologías

- **KNIME Analytics Platform:** plataforma visual open-source con cientos de nodos para acceso/transformación/ML y orquestación reproducible.
- **RapidMiner (Altair):** entorno de minería de datos/ML con operadores y MLOps; útil para acelerar ciclos analíticos.
- **Orange Data Mining:** visual programming + biblioteca Python para minería (widgets de clasificación, visualización y texto).
- **WEKA:** conjunto de algoritmos de ML/DM y experimentación académica; clásico para enseñanza y prototipado.

## Big Data

Big Data hace referencia al manejo de volúmenes masivos de datos que, por su tamaño, velocidad y variedad, no pueden procesarse con herramientas tradicionales. Incluye tanto el almacenamiento distribuido como el análisis de esos datos para generar valor en tiempo real o a gran escala. Su importancia radica en que permite a las organizaciones obtener insights estratégicos, como sucede en el análisis de datos de redes sociales para entender tendencias de opinión pública.

## Tipos de aplicaciones

1. ETL/ELT y data lakes/warehouses a gran escala.
2. Análisis en tiempo real (telemetría, IoT, clickstreams).
3. ML/IA a escala (entrenos distribuidos, scoring sobre streams).

## Procesamiento

- **Batch:** jobs programados de transformación/análisis (por ejemplo, Spark SQL sobre HDFS o cargas diarias a un warehouse).
- **Streaming:** pipelines de eventos de baja latencia (por ejemplo, Kafka + Spark/Flink Structured Streaming) para detección en tiempo real y dashboards vivos.

## Herramientas y tecnologías

- **Apache Spark:** motor unificado de procesamiento distribuido (SQL, MLlib, streaming estructurado, grafos).
- **Apache Hadoop (HDFS/MapReduce/YARN):** base de almacenamiento distribuido y procesamiento batch; aún presente en muchos lagos de datos.
- **Apache Kafka:** plataforma de event streaming para publicar/almacenar/procesar flujos de eventos con alta escalabilidad.
- **Apache Flink:** procesamiento de streams y batch con baja latencia y estado manejado.
- **Google BigQuery:** data warehouse serverless y “AI-ready” para consultas SQL de petabytes y casi en tiempo real.
- **Apache Hive / HBase:** SQL-on-Hadoop para data warehousing; y base NoSQL columnar para lecturas/escrituras aleatorias a gran escala.

## Conclusión

Al analizar los cuatro dominios podemos ver que están muy relacionados, pero cada uno cumple un papel diferente. La Inteligencia Artificial es la meta general, porque busca que las computadoras puedan “pensar” o actuar como lo haría una persona. El Machine Learning es la forma en que esos sistemas aprenden de los datos para mejorar con el tiempo. La Minería de Datos se enfoca más en descubrir información útil y patrones que estaban escondidos. Finalmente, Big Data es el soporte que permite manejar cantidades enormes de datos y procesarlos de manera rápida.

También es importante considerar cómo se procesan los datos, el método por lotes (batch) funciona mejor cuando se trabaja con grandes cantidades que se pueden analizar poco a poco, mientras que el método en streaming es clave cuando se necesita reaccionar al momento, como en la detección de fraudes o el monitoreo de sensores.

En conclusión, conocer estas diferencias ayuda a elegir la herramienta y el tipo de procesamiento adecuados para cada situación. Esto no solo mejora el resultado de los proyectos, sino que también hace que el trabajo sea más eficiente y que los datos realmente se conviertan en conocimiento útil para tomar decisiones.

## Referencias

Apache Spark. (2025). Documentation. Apache Software Foundation.

<https://spark.apache.org/documentation.html>

Google Cloud. (2025). BigQuery documentation. Google Cloud.

<https://cloud.google.com/bigquery/docs>

Hugging Face. (2025). Transformers documentation. Hugging Face.

<https://huggingface.co/docs/transformers>

KNIME. (2025). KNIME Documentation. KNIME. <https://docs.knime.com/>

OpenAI ChatGPT. (2025). Consulta de información sobre IA, ML, DM y Big Data. OpenAI.

PyTorch. (2025). PyTorch documentation. Meta AI.

<https://pytorch.org/docs/stable/index.html>

scikit-learn. (2025). User Guide. scikit-learn developers. [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html)

TensorFlow. (2025). Learn & Guides. Google. <https://www.tensorflow.org/learn>