

**UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA**

**DESARROLLO Y GESTIÓN DE SOFTWARE**



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

**EXTRACCIÓN DE CONOCIMIENTO EN BASES DE DATOS**

**PRESENTA:**

KARLA ALEJANDRA DE LA CRUZ ZEA

**DOCENTE:**

ING. LUIS ENRIQUE MASCOTE CANO

22 de septiembre de 2025

## **Contenido**

Introducción .....	2
Sección 1: Inteligencia Artificial (IA) .....	2
Definición .....	2
Sección 2: Machine Learning (ML).....	3
Definición .....	3
Sección 3: Data Mining (DM) .....	5
Definición .....	5
Sección 4: Big Data .....	6
Definición .....	6
Conclusión.....	7

# Introducción

En la actualidad, las organizaciones generan y almacenan enormes volúmenes de datos y requieren técnicas avanzadas para extraer valor de ellos. La Inteligencia Artificial, el Machine Learning, la Data Mining y el Big Data son dominios relacionados, pero con enfoques distintos, es decir, la IA busca sistemas que simulen capacidades humanas, el ML se enfoca en algoritmos que aprenden de los datos, el DM descubre patrones ocultos y Big Data ofrece infraestructura y métodos para manejar datos masivos.

El hecho de conocer los tipos de aplicaciones, modalidades de procesamiento y herramientas más usadas en cada dominio permite seleccionar la tecnología adecuada para cada proyecto, optimizando recursos y resultados.

## Sección 1: Inteligencia Artificial (IA)

### Definición

La IA es el campo que desarrolla sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el razonamiento, la percepción o la toma de decisiones.

### Tipos de aplicaciones

#### Visión por computador

- **Descripción:** Algoritmos que reconocen, clasifican o interpretan imágenes y videos.
- **Ejemplo real:** Reconocimiento facial en aeropuertos (Clear, USA).



#### Procesamiento de lenguaje natural (PLN)

- **Descripción:** Análisis y generación de texto o voz humana.
- **Ejemplo real:** Chatbots como ChatGPT y asistentes como Alexa.

#### Sistemas expertos y de recomendación

- **Descripción:** Motores que sugieren productos, diagnósticos o decisiones basadas en reglas o datos.
- **Ejemplo real:** Recomendaciones de Netflix.

### Procesamiento (batch vs streaming)

- **Batch:** Entrenamiento de modelos grandes en servidores o la nube (p.ej., entrenar un modelo de visión con millones de imágenes).
- **Streaming:** Inferencia en tiempo real (p.ej., identificar objetos en un video en vivo).

### Herramientas y tecnologías

- **TensorFlow:** Framework de código abierto para redes neuronales profundas; usado por Google.
- **PyTorch:** Biblioteca muy popular en investigación y prototipos de deep learning.
- **OpenCV:** Librería para visión por computador.
- **IBM Watson:** Plataforma IA en la nube para PLN y análisis cognitivo.
- **Microsoft Azure AI Services:** Servicios listos para análisis de texto, imágenes y voz en streaming.

## Sección 2: Machine Learning (ML)

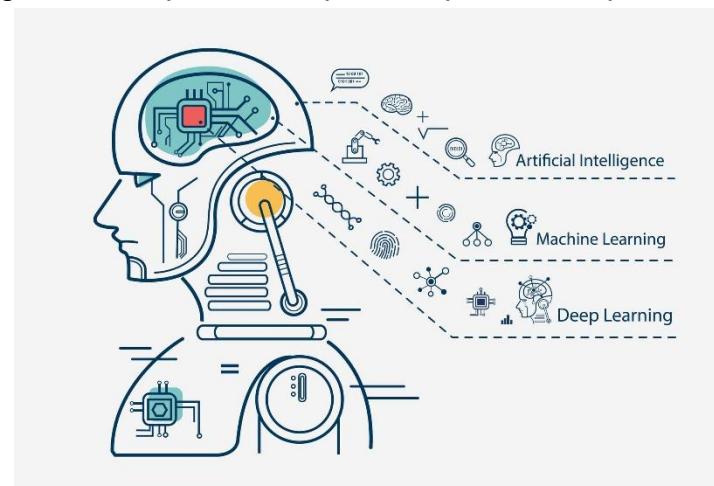
### Definición

Subcampo de la IA que desarrolla algoritmos capaces de aprender patrones a partir de datos sin ser programados explícitamente.

### Tipos de aplicaciones

#### Clasificación y predicción

- **Descripción:** Predecir clases o valores futuros.



- **Ejemplo real:** Modelos de predicción de fraude en tarjetas de crédito.
- **Clustering o segmentación**
  - **Descripción:** Agrupar clientes según comportamiento.
  - **Ejemplo real:** Segmentación de usuarios para campañas de marketing.
- **Sistemas de recomendación basados en ML**
  - **Descripción:** Motores que aprenden de preferencias pasadas.
  - **Ejemplo real:** Recomendaciones de Spotify.

### Procesamiento (batch vs streaming)

- **Batch:** Entrenamiento periódico de modelos con grandes datasets.
- **Streaming:** Actualización de modelos o scoring en tiempo real (p.ej., clasificar transacciones bancarias al momento).

### Herramientas y tecnologías

- **scikit-learn:** Biblioteca Python para ML clásico.
- **XGBoost/LightGBM:** Algoritmos de boosting muy usados en competencias y producción.
- **MLlib (Apache Spark):** ML distribuido en clusters.
- **TensorFlow Extended (TFX):** Pipeline completo para producción de ML.
- **Amazon SageMaker:** Servicio cloud para entrenar y desplegar modelos.

## Sección 3: Data Mining (DM)

### Definición

Proceso de explorar grandes conjuntos de datos para descubrir patrones, tendencias o relaciones útiles.



### Tipos de aplicaciones

- **Detección de fraude**
  - **Descripción:** Identificar comportamientos anómalos en datos transaccionales.
  - **Ejemplo real:** Detección de fraude en telecomunicaciones.
- **Análisis de canastas de mercado**
  - **Descripción:** Encontrar productos que se compran juntos.
  - **Ejemplo real:** Promociones cruzadas en supermercados.
- **Análisis de abandono de clientes (churn)**
  - **Descripción:** Identificar clientes propensos a abandonar un servicio.
  - **Ejemplo real:** Compañías de telefonía móvil.

### Procesamiento (batch vs streaming)

- Tradicionalmente en **batch** (minería sobre históricos).

- Nuevas implementaciones en **streaming** permiten detección de patrones en flujos continuos, útil para fraude en tiempo real.

## Herramientas y tecnologías

- **Weka:** Plataforma educativa para experimentos de minería de datos.
- **RapidMiner:** Entorno visual para minería y análisis predictivo.
- **KNIME:** Plataforma de integración y minería de datos con flujos gráficos.
- **Orange:** Herramienta visual para análisis y minería interactiva.
- **SQL/BigQuery:** Consultas avanzadas para minería en almacenes de datos.

## Sección 4: Big Data

### Definición

Conjunto de técnicas, arquitecturas y herramientas para almacenar y procesar grandes volúmenes y variedad de datos a alta velocidad.

### Tipos de aplicaciones

- **Análisis de redes sociales**
  - **Descripción:** Procesar millones de publicaciones para detectar tendencias.
  - **Ejemplo real:** Monitoreo de opinión pública en Twitter.
- **Procesamiento de logs de sensores IoT**
  - **Descripción:** Recibir y analizar datos de miles de dispositivos.
  - **Ejemplo real:** Ciudades inteligentes (CitySense).
- **Analítica en tiempo real para e-commerce**
  - **Descripción:** Personalizar ofertas según comportamiento en el sitio.
  - **Ejemplo real:** Amazon, Alibaba.



## Procesamiento (batch vs streaming)

- **Batch:** Procesamiento periódico de grandes volúmenes (ETL nocturnas en data lakes).
- **Streaming:** Procesamiento en tiempo real para detectar eventos o alertas instantáneamente.

## Herramientas y tecnologías

- **Apache Hadoop (HDFS, MapReduce):** Almacenamiento y procesamiento distribuido por lotes.
- **Apache Spark:** Procesamiento in-memory para batch y streaming.
- **Kafka:** Sistema de mensajería para ingestión de datos en streaming.
- **Flink / Storm:** Procesamiento de flujos en tiempo real.
- **Google BigQuery / Snowflake:** Almacenes analíticos serverless para Big Data.

## Conclusión

La IA engloba técnicas avanzadas para simular capacidades humanas; el ML se centra en algoritmos que aprenden patrones; el DM busca descubrir información oculta en datos ya existentes; y Big Data provee la infraestructura para procesar datos masivos y variados.

Entonces la elección del modo de procesamiento es crucial. El **batch** es apropiado cuando se tienen grandes volúmenes y no importa esperar. El **streaming** es esencial cuando se requieren resultados inmediatos.

Elegir la herramienta correcta también marca la diferencia: frameworks como TensorFlow o PyTorch son idóneos para IA y ML; Spark y Kafka para Big Data; y plataformas visuales como RapidMiner para DM. Conocer las fortalezas y limitaciones de cada dominio y tecnología ayuda a diseñar soluciones más eficientes, escalables y ajustadas a las necesidades de cada proyecto.

## Referencias bibliográficas (APA)

Apache Software Foundation. (s.f.). *Apache Spark*. Recuperado de <https://spark.apache.org/>

Google Cloud. (s.f.). *BigQuery documentation*. Recuperado de <https://cloud.google.com/bigquery/docs>

KNIME. (s.f.). *KNIME Analytics Platform*. Recuperado de <https://www.knime.com/>

Orange Data Mining. (s.f.). *Orange*. Recuperado de <https://orangedatamining.com/>

RapidMiner. (s.f.). *RapidMiner platform*. Recuperado de <https://rapidminer.com/>

Scikit-learn developers. (s.f.). *Scikit-learn*. Recuperado de <https://scikit-learn.org/>

TensorFlow developers. (s.f.). *TensorFlow*. Recuperado de <https://www.tensorflow.org/>

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier.