

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

TECNOLOGÍAS DE LA INFORMACIÓN



EXTRACCIÓN DE CONOCIMIENTOS 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:
SEBASTIÁN ACOSTA ORTIZ

DOCENTE:
**LUIS ENRIQUE MASCOTE
CANO**

Chihuahua, Chih., 25 de septiembre de 2025

Introducción

Estamos rodeados de tecnología que aprende, analiza datos y toma decisiones de manera casi automática. Todo eso es posible gracias a conceptos que quizá hemos escuchado muchas veces, como la Inteligencia Artificial, el Machine Learning, la Minería de Datos y el Big Data. Estos cuatro dominios no solo están cambiando la forma en que las empresas operan, sino también cómo interactuamos con el mundo a través de aplicaciones, redes sociales, plataformas de streaming, servicios en la nube y más. Decidimos investigar estos temas porque, aunque a veces se usan como sinónimos, cada uno tiene sus propias características, funciones y herramientas. A lo largo del trabajo vamos a explicar para qué sirve cada uno, qué tipo de aplicaciones son las más comunes, cómo se procesan los datos en cada caso (ya sea por lotes o en tiempo real) y qué tecnologías se usan normalmente. El objetivo principal de este reporte es entender mejor cómo se relacionan estos conceptos entre sí, en qué se diferencian y por qué es importante saber elegir la herramienta adecuada dependiendo del tipo de datos o problema que queramos resolver. Creemos que, como estudiantes y futuros profesionistas del área tecnológica, tener claras estas bases nos va a ayudar a tomar mejores decisiones en el desarrollo de proyectos reales.

Inteligencia Artificial (IA)

1. Definición breve

La inteligencia artificial (IA) es una tecnología que permite a las computadoras y máquinas simular el aprendizaje humano, la comprensión, la resolución de problemas, la toma de decisiones, la creatividad y la autonomía.

2. Tipos de aplicaciones

Visión por computadora/Machine

La visión por computadora es un tipo de inteligencia artificial que entrena a las computadoras para imitar la forma en que los humanos ven, comprenden lo que observan y actúan en función de esa información procesada y analizada. Tesla

Natural Language Processing

Subcampo de la informática y la inteligencia artificial que emplea el machine learning para permitir que las computadoras comprendan y se comuniquen con el lenguaje humano. ChatGPT

Agente de IA

Un AI agent o agente de inteligencia artificial se refiere a un sistema o programa que es capaz de realizar tareas de forma autónoma en nombre de un usuario o de otro sistema mediante el diseño de su flujo de trabajo y el uso de las herramientas disponibles. DeepMind

3. Procesamiento

En batch, se procesan grandes bloques de datos acumulados en momentos programados, se emplea en entrenamiento de modelos, reentrenamientos periódicos, análisis a gran escala. Como ventajas tiene la eficiencia para tareas pesadas, permite optimización, reproducibilidad

En stream, se procesan datos a medida que llegan, con baja latencia se emplean en inferencia en línea (por ejemplo, recomendaciones en tiempo real, detección de anomalías, control de robots).

Comparación: en IA el entrenamiento muchas veces se hace en batch, mientras la inferencia puede ejecutarse con streaming para reaccionar rápido

4. Herramientas y tecnologías

- Desarrollo y modernización de aplicaciones

Las herramientas de generación de código de IA generativa y las herramientas de automatización pueden agilizar las tareas de programación repetitivas asociadas con el desarrollo de aplicaciones y acelerar la migración y modernización (reformato y reestructuración) de aplicaciones heredadas a escala. Estas herramientas pueden acelerar las tareas, ayudar a garantizar la coherencia del código y reducir los errores.

- Experiencia del cliente, servicio y soporte

Las compañías pueden implementar chatbots y asistentes virtuales impulsados por IA para manejar las consultas de los clientes, los tickets de soporte y más. Estas herramientas emplean el procesamiento de lenguaje natural (PLN) y las capacidades de IA generativa para comprender y responder a las preguntas de los clientes sobre el estado de los pedidos, los detalles del producto y las políticas de devolución. Los chatbots y los asistentes virtuales permiten una asistencia permanente, ofrecen respuestas más rápidas a las preguntas más frecuentes (FAQ), liberan a los agentes humanos para que puedan centrar en tareas de mayor nivel y ofrecen a los clientes un servicio más rápido y coherente

Machine Learning (ML)

1. Definición breve

Directamente debajo de la IA, tenemos machine learning, que consiste en crear modelos entrenando a un algoritmo para hacer predicciones o tomar decisiones basadas en datos. Abarca una amplia gama de técnicas que permiten a las computadoras aprender y hacer inferencias basadas en datos sin estar programados explícitamente para tareas específicas.

2. Tipos de aplicaciones

PCA

- El análisis de componentes principales (PCA) reduce el número de dimensiones de grandes conjuntos de datos a componentes principales que conservan la mayor parte de la información original. Esto se logra transformando variables potencialmente correlacionadas en un conjunto más pequeño de variables, llamadas componentes principales.

t-SNE

- Es una técnica no lineal de reducción de dimensionalidad diseñada específicamente para visualizar datos de alta dimensión en espacios de 2D o 3D. Funciona modelando las similitudes por pares entre los puntos de datos en el espacio de alta dimensión y optimizando su representación en un espacio de menor dimensión para preservar dichas similitudes.

3. Procesamiento

En batch, los modelos se entrena con grandes conjuntos de datos almacenados, de forma offline. Entrenamiento inicial o reentrenamiento al final de un periodo

En streaming, se usan técnicas de aprendizaje incremental o en línea (online learning), donde el modelo se actualiza continuamente con nuevos datos. modelos que reciben nuevos datos continuamente (p. ej. predicción en tiempo real, modelos adaptativos)

En escenarios con alto volumen, el ML debe coexistir con arquitecturas de procesamiento de datos (big data). En el artículo "Exploring the Intersection of Machine Learning and Big Data" se discuten los retos de aplicar ML en entornos de datos grandes y de alta velocidad.

4. Herramientas y tecnologías

Automatización robótica de procesos (RPA)

También conocida como robótica de software, la RPA utiliza tecnologías de automatización inteligentes para realizar tareas manuales repetitivas.

1. Negociación de acciones automatizada: diseñadas para optimizar las carteras de acciones, las plataformas de negociación de alta frecuencia impulsadas por IA realizan miles o incluso millones de operaciones por día sin intervención humana.
2. Detección de fraudes: Los bancos y otras instituciones financieras pueden utilizar el machine learning para detectar transacciones sospechosas. El aprendizaje supervisado puede entrenar un modelo utilizando información sobre transacciones fraudulentas conocidas. La detección de anomalías puede identificar transacciones que parecen atípicas y merecen una investigación más profunda.

Scikit-learn.

Librería en Python para algoritmos clásicos (regresión, clustering, árboles)

Data Mining (DM)

1. Definición breve

La minería de datos es el uso del machine learning y el análisis estadístico para descubrir patrones y otra información valiosa de grandes conjuntos de datos.

2. Tipos de aplicaciones más frecuentes

1. Reglas de asociación / análisis de canasta de mercado

Descubrir relaciones entre ítems (por ejemplo, si un cliente compra A, también compra B) Ejemplo: supermercados utilizan para promociones cruzadas

2. Procesamiento de secuencias / series temporales

Descripción: encontrar patrones en datos ordenados por tiempo Ejemplo: predicción de demanda, análisis de comportamiento temporal

3. Detección de patrones / clustering / reglas de clasificación

Descripción: identificar estructuras ocultas (clusters) o construir reglas que expliquen categorías. Ejemplo: segmentación de clientes, perfiles de usuarios

3. Procesamiento: batch vs streaming en DM

En minería de datos tradicional se trabaja en **batch**, ya que se analizan conjuntos completos de datos. En escenarios donde los datos llegan continuamente, hay variantes de minería incremental o minería sobre flujos (stream mining) para adaptar modelos en línea. En Big Data, la minería de datos debe integrarse con arquitecturas que permitan procesar datos masivos y en tiempo real (ver sección Big Data).

4. Herramientas y tecnologías

Algunas de las herramientas destacadas:

1. **WEKA**: plataforma popular para minería de datos (algoritmos de clasificación, clustering, reglas de asociación)
2. **RapidMiner**: entorno de minería de datos con interfaz visual, integración con ML
3. **KNIME**: (mencionado antes) permite flujos visuales de minería de datos y análisis
4. **ELKI**: framework para algoritmos de clustering, detección de anomalías, índices espaciales [Wikipedia](#)
5. **Orange**: entorno visual para análisis de datos y minería
6. **R** (paquetes como arules, caret, tm)
7. **SAS Enterprise Miner**

Big Data

1. Definición breve

Big data se refiere a conjuntos de datos masivos y complejos que los sistemas tradicionales de gestión de datos no pueden manejar. Cuando se recopilan, gestionan y analizan adecuadamente, los big data pueden ayudar a las organizaciones a descubrir nuevos insights y tomar mejores decisiones empresariales.

2. Tipos de aplicaciones más frecuentes

1. Análisis predictivo / analítica avanzada sobre datos masivos.

- Usar ML a escala para predecir tendencias, fallos, demanda, etc.

2. Procesamiento y análisis de flujos de eventos (*streaming analytics*)

- Ejemplos: monitoreo en tiempo real, detección de anomalías en redes IoT, análisis de clics web.

3. Almacenamiento y consulta de grandes volúmenes de datos / Data lakes / warehouses escalables

- Infraestructura para consolidar y consultar datos: logs de sistemas, históricos, datos estructurados y no estructurados.

3. Procesamiento: batch vs streaming en Big Data

- En Big Data, con frecuencia se combinan ambos enfoques mediante arquitecturas como *Lambda Architecture*, que usan una capa batch para procesamiento global y una capa de streaming para respuestas rápidas. [Wikipedia](#)
- El procesamiento **batch** es apropiado para análisis completos, reconstrucción histórica, cálculos agregados pesados.
- El procesamiento **streaming** se usa cuando es necesario actuar en tiempo real o mantener actualizaciones continuas (por ejemplo, dashboards, alertas).

- Muchos sistemas modernos unifican ambos enfoques (por ejemplo, Apache Flink permite procesamiento unificado de streaming y batch).
- En el artículo “Batch vs Stream Processing: How to Choose” se distingue cuándo usar cada enfoque y en qué escenarios.
- En “Confluent Unites Batch and Stream Processing” se menciona cómo plataformas actuales combinan ambos enfoques para sistemas de IA más reactivos.

4. Herramientas y tecnologías

Aquí algunas herramientas clave del ecosistema Big Data:

- **Apache Hadoop (HDFS + MapReduce / YARN)**: para almacenamiento distribuido y procesamiento en batch
- **Apache Spark**: motor de procesamiento que soporta batch, micro-batch y streaming (con Spark Streaming)
- **Apache Flink**: motor que soporta procesamiento unificado (batch + streaming) con baja latencia. [Wikipedia](#)
- **Apache Kafka**: sistema de mensajería distribuida para flujos de eventos (pub/sub)
- **Apache Storm / Apache Samza**: motores para procesamiento en tiempo real (streaming) (Samza trabaja junto con Kafka). [Wikipedia](#)
- **NoSQL / bases de datos escalables**: MongoDB, Cassandra, HBase, Elasticsearch
- **Data warehouses y lagos de datos**: Delta Lake, Apache Hive, Iceberg
- **Herramientas de consulta y análisis**: Presto / Trino, Apache Drill
- **Plataformas en la nube**: AWS EMR / Kinesis / Glue, Google BigQuery / Dataflow, Azure Synapse / Azure Stream Analytics

Conclusión

Después de analizar los cuatro dominios —Inteligencia Artificial, Machine Learning, Minería de Datos y Big Data— nos damos cuenta de que, aunque tienen muchas conexiones entre sí, cada uno cumple un papel distinto dentro del procesamiento y análisis de datos. Entender estas diferencias es importante porque no todas las soluciones requieren IA o Big Data. A veces, un problema se puede resolver con modelos simples y procesamiento por lotes; otras veces, se necesita procesamiento en tiempo real o infraestructura distribuida. Como mencionan en IBM, elegir bien las tecnologías y los enfoques puede marcar la diferencia entre un sistema que solo analiza datos y uno que genera verdadero valor. La inteligencia artificial es el enfoque más amplio, donde el objetivo es lograr que las máquinas imiten ciertas capacidades humanas como razonar o aprender. Dentro de la IA, el machine learning actúa como una técnica clave que permite que los sistemas mejoren automáticamente a partir de los datos, sin necesidad de estar programados en cada paso.

Referencias

- Badman, A., & Kosinski, M. (2025, February 26). ¿Qué es big data? *ibm.com*. <https://www.ibm.com/mx-es/think/topics/big-data>
- ¿Qué es la minería de datos? (2024, September 30). *ibm.com*. <https://www.ibm.com/mx-es/think/topics/data-mining>
- ¿Qué es machine learning? (2025, June 27). *ibm.com*. <https://www.ibm.com/mx-es/think/topics/machine-learning>
- Stryker, C., & Holdsworth, J. (2025, April 23). ¿Qué es el procesamiento de lenguaje natural (PLN)? *ibm.com*. <https://www.ibm.com/mx-es/think/topics/natural-language-processing>
- Stryker, C., & Kavlakoglu, E. (2025, June 24). ¿Qué es la inteligencia artificial o IA? *ibm.com*. <https://www.ibm.com/mx-es/think/topics/artificial-intelligence>
- What is computer vision?* (n.d.). Intel. Retrieved September 26, 2025, from <https://www.intel.com/content/www/us/en/learn/what-is-computer-vision.html>
- (N.d.). Mdpi.com. Retrieved September 26, 2025, from https://www.mdpi.com/2504-4990/7/1/13?utm_source=chatgpt.com