



Aplicaciones, procesamientos y herramientas de análisis de datos

Extracción de Conocimiento en Bases de Datos

Luis Eduardo Aguilar Sarabia

IDGS91N

Docente: Luis Enrique Mascote Cano

Introducción

La Inteligencia Artificial, el Machine Learning, la Minería de Datos y el Big Data son disciplinas que forman parte del análisis de datos y de su aplicación práctica en distintos entornos. Cada una de ellas cumple un rol particular: desde la simulación de procesos cognitivos y el aprendizaje automático, hasta la detección de patrones y el manejo de grandes volúmenes de información. Describir sus aplicaciones, modalidades de procesamiento y herramientas permite entender sus diferencias y puntos de encuentro, así como su utilidad en la solución de problemas específicos. Esta visión conjunta facilita reconocer su valor como apoyo en la innovación y en la toma de decisiones estratégicas.

Sección 1: Inteligencia Artificial (IA)

Definición

La IA es la disciplina que busca desarrollar sistemas capaces de imitar la inteligencia humana, resolviendo problemas mediante percepción, razonamiento, aprendizaje y toma de decisiones.

Tipos de aplicaciones

1. Visión por computador

- *Descripción:* Permite que las máquinas interpreten imágenes y videos.
- *Ejemplo:* Los sistemas de reconocimiento facial de Facebook/Meta.

2. Procesamiento de lenguaje natural (PLN)

- *Descripción:* Comprensión y generación de lenguaje humano.
- *Ejemplo:* ChatGPT de OpenAI.

3. Sistemas expertos

- *Descripción:* Programas que simulan la toma de decisiones de expertos en dominios específicos.

- *Ejemplo:* MYCIN en medicina para diagnósticos.

Procesamiento

- **Batch:** Usado para entrenar modelos con grandes volúmenes de datos históricos (ej. entrenar redes neuronales profundas).
- **Streaming:** Empleado en tareas que requieren respuesta inmediata, como la detección de objetos en cámaras de seguridad en tiempo real.

Herramientas y tecnologías

1. **TensorFlow:** Framework para crear y entrenar modelos de IA. Popular por su escalabilidad en proyectos de investigación y producción.
2. **PyTorch:** Biblioteca de código abierto, muy usada en investigación por su flexibilidad.
3. **OpenCV:** Librería enfocada en visión por computador.
4. **IBM Watson:** Plataforma que ofrece soluciones de IA listas para empresas.

Sección 2: Machine Learning (ML)

Definición

El ML es un subcampo de la IA enfocado en algoritmos que aprenden de los datos sin ser programados explícitamente.

Tipos de aplicaciones

1. **Modelos de predicción**
 - *Ejemplo:* Predicción de riesgo crediticio en bancos.
2. **Clasificación de datos**
 - *Ejemplo:* Detección de spam en correos electrónicos.
3. **Sistemas de recomendación**
 - *Ejemplo:* Netflix recomendando películas según hábitos de usuarios.

Procesamiento

- **Batch:** Ideal para entrenamiento con datasets grandes y estáticos.
- **Streaming:** Se usa en la actualización continua de modelos con datos en tiempo real (ej. fraudes en transacciones financieras).

Herramientas y tecnologías

1. **Scikit-learn:** Biblioteca en Python para tareas de clasificación, regresión y clustering.
2. **XGBoost:** Muy utilizada en competiciones de ML por su eficiencia en boosting.
3. **Keras:** API de alto nivel para redes neuronales.
4. **Mlib (Apache Spark):** Librería distribuida para ML en entornos Big Data.

Sección 3: Data Mining (DM)

Definición

La minería de datos es el proceso de descubrir patrones, relaciones y tendencias útiles en grandes volúmenes de datos mediante técnicas estadísticas, matemáticas y de ML.

Tipos de aplicaciones

1. **Segmentación de clientes**
 - *Ejemplo:* Campañas de marketing dirigidas en Amazon.
2. **Análisis de patrones de compra**
 - *Ejemplo:* Reglas de asociación como “quien compra pan también compra leche”.
3. **Detección de anomalías**
 - *Ejemplo:* Identificación de transacciones bancarias sospechosas.

Procesamiento

- **Batch:** Común en análisis exploratorios de grandes datasets históricos.
- **Streaming:** Usado en aplicaciones de detección de anomalías en tiempo real.

Herramientas y tecnologías

1. **RapidMiner:** Plataforma gráfica para análisis de datos.
2. **Weka:** Herramienta de minería de datos con múltiples algoritmos integrados.
3. **Orange:** Software interactivo para análisis visual y minería de datos.
4. **KNIME:** Plataforma de integración para análisis de datos y minería.

Sección 4: Big Data

Definición

Big Data se refiere al manejo y análisis de volúmenes masivos de datos que superan la capacidad de las herramientas tradicionales.

Tipos de aplicaciones

1. **Análisis de redes sociales**
 - *Ejemplo:* Twitter analizando tendencias en tiempo real.
2. **Monitoreo de IoT**
 - *Ejemplo:* Sensores urbanos en Smart Cities.
3. **Analítica predictiva empresarial**
 - *Ejemplo:* Walmart analizando patrones de compra globales.

Procesamiento

- **Batch:** Usado en análisis de logs históricos a gran escala.
- **Streaming:** Crítico en sistemas de monitoreo en tiempo real, como seguridad informática.

Herramientas y tecnologías

1. **Hadoop**: Framework para procesamiento distribuido.
2. **Apache Spark**: Procesamiento rápido en memoria.
3. **Kafka**: Plataforma para ingestión de datos en streaming.
4. **Google BigQuery**: Servicio en la nube para análisis de datos a gran escala.
5. **Amazon EMR**: Servicio de AWS para análisis con Hadoop y Spark.

Conclusión

Los cuatro dominios analizados presentan interconexiones claras pero también propósitos diferenciados. La IA busca simular inteligencia humana, el Machine Learning aporta algoritmos que permiten el aprendizaje autónomo, la Minería de Datos se centra en descubrir patrones ocultos y el Big Data ofrece la infraestructura para gestionar información masiva. Aunque comparten herramientas y técnicas, cada uno se adapta a diferentes necesidades: desde la toma de decisiones estratégicas hasta el monitoreo en tiempo real.

Referencias

<https://www.ibm.com/mx-es/think/topics/artificial-intelligence>

<https://www.ibm.com/mx-es/think/topics/machine-learning>

<https://aws.amazon.com/es/what-is/data-mining/>

<https://www.oracle.com/mx/big-data/what-is-big-data/>