

# Actividad 1: Conceptos Básicos de Ciencia de Datos

---

Renzo Zagni  
Matricula: A01795457  
25 de Septiembre de 2023

En este informe, se abordarán de manera concisa los conceptos clave de Aprendizaje Automático, Minería de Datos, Big Data y Analítica de Datos, así como la interrelación entre cada uno de estos elementos en el contexto de la Ciencia de Datos.

## **Aprendizaje Automatico**

Aprendizaje Automático, comúnmente conocido como Machine Learning (ML), constituye el campo de estudio dedicado al desarrollo de algoritmos computacionales con el propósito de identificar y extraer patrones en los datos.

Este proceso se divide en dos fases fundamentales: la fase de modelado, en la cual el algoritmo de aprendizaje de máquina se aplica al conjunto de datos para identificar patrones útiles y la fase de análisis, donde se emplea el modelo previamente creado para realizar un análisis más profundo de los datos, lo que puede revelar atributos o patrones importantes en el dominio..

## **Minería de Datos**

La Minería de Datos se refiere al proceso de descubrir patrones, conocimientos y relaciones dentro de los conjuntos de datos. Aunque a menudo es confundida con el Aprendizaje Automático, la Minería de Datos se centra principalmente en el análisis de datos estructurados, con un enfoque orientado a aplicaciones comerciales

## **Big Data**

El termino Big Data se refiere a conjuntos de información extremadamente grandes y complejos que los métodos tradicionales de procesamiento de datos tienen dificultades para manejar. Big Data abarca volúmenes masivos de datos estructurados y no estructurados, que a menudo requieren tecnologías avanzadas para su almacenamiento, procesamiento y análisis.

Big Data se rige por tres características principales conocidas como las tres "V":

1. **Volumen:** describe la alta cantidad de datos a procesar.
2. **Velocidad:** describe la alta velocidad con que la data es generada y procesada.
3. **Variedad:** esta característica se refiere a los muchos tipos de datos presentes los que pueden ser no estructurados o semi estructurados

## **Analitica de Datos**

La Analítica de Datos se refiere al proceso o los procesos de examinar conjuntos de datos con el fin de descubrir patrones, tendencias y relaciones significativas. A medida que la data se ha vuelto más extensa y compleja, este proceso ha evolucionado gracias a mejoras en algoritmos y recursos computacionales.

La evolución de la analítica de datos se puede resumir en 6 fases principales:

1. **Estatística Descriptiva:** Incluye medidas de tendencia central y variabilidad.
2. **Aprendizaje Estadístico:** Implica la aplicación de técnicas como regresión lineal y logística basadas en el método de mínimos cuadrados de Gauss.
3. **Análisis Discriminatorio:** Se centra en clasificación y reconocimiento de patrones basados en la Discriminatoria Lineal de Fisher.
4. **Aprendizaje Automático:** Involucra la implementación de modelos computacionales para aprender de los datos.
5. **Ensamblados:** Consiste en el uso de modelos combinados para obtener predicciones más precisas.

6. **Redes Neuronales de Aprendizaje Profundo:** Basadas en redes neuronales artificiales con un enfoque en aprendizaje de representación profunda. El termino "profundo" se refiere al uso de múltiples capas en la red neuronal.

### **Relación entre Aprendizaje Automático, Minería de Datos, Big Data y Analítica de Datos con la Ciencia de Datos**

La Ciencia de Datos se define como un ecosistema multidisciplinario enfocado al estudio y aplicación de técnicas, métodos y algoritmos para extraer conocimiento y comprensión a partir de grandes conjuntos de datos.

Entre estas disciplinas tenemos al Aprendizaje Automático, la Minería de Datos, Big Data y la Analítica de Datos. Estas además de formar parte del ecosistema de la Ciencia de Datos también forman parte de la evolución de esta misma sumada a grandes cambios en la data (volumen, velocidad, variedad), las herramientas disponibles y los avances en las tecnologías computacionales. La Analítica de Datos, por su parte, se divide en dos áreas: la recolección y el análisis de datos.

Desde el punto de vista de la recolección de datos encontramos a Big Data y las tecnologías de manejo de Big Data como MapReduce o las bases de datos NoSQL que nos permiten capturar y manipular esta data en preparación para las fases de análisis de la Ciencia de Datos.

Desde el punto de vista del Análisis de Datos, la Minería de Datos es probablemente el proceso mas conocido en los ambiente empresariales aunque también juega in papel importante en la Ciencia de Datos debido a su empleo del Aprendizaje Automático y la facilitación de este ultimo al proceso de la Minería de Datos.

La Minería de Datos y el Aprendizaje Automático forman las bases que soportan procesos mas complejos como el Aprendizaje Profundo y las Redes Neuronales Profundas.

Finalmente, la Analítica de Datos y Ciencia de Datos están intrínsecamente ligadas siendo la Analítica de Datos parte integral de la Ciencia de Datos. Mientras que la Ciencia de Datos abarca todo el ciclo de vida de los datos, desde la recopilación hasta la interpretación y comunicación de los resultados, la Analítica de Datos se enfoca específicamente en la etapa de extracción de conocimiento y comprensión a través del análisis de datos

---

### **Referencias:**

Kelleher, J. D., & Tierney, B. (2018). Chapter 2: What are Data and What is a Data Set. Data science, The MIT Press

Kelleher, J. D., & Tierney, B. (2018). Chapter 4: Machine Learning 101. Data science, The MIT Press  
Big data. In Wikipedia, The Free Encyclopedia. Retrieved 16:23, September 20, 2023, from [https://en.wikipedia.org/w/index.php?title=Big\\_data&oldid=1175637937](https://en.wikipedia.org/w/index.php?title=Big_data&oldid=1175637937)

Analysis Discriminante Lineal In Wikipedia, The Free Encyclopedia. Retrieved 18:32, September 21, 2023, from [https://es.wikipedia.org/w/index.php?title=An%C3%A1lisis\\_discriminante\\_lineal&oldid=153688711](https://es.wikipedia.org/w/index.php?title=An%C3%A1lisis_discriminante_lineal&oldid=153688711)