

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS
SUPERIORES DE MONTERREY

*Actividad 1. Conceptos básicos de ciencia de
datos*

Daniel Edgardo García Bonilla
A01797180

CIENCIA Y ANALÍTICA DE DATOS

15/01/2025

1. Aprendizaje automático

El aprendizaje automático o machine learning es el proceso de diseño y evaluación de algoritmos computacionales mediante los cuales los ordenadores buscan patrones de comportamiento en una fuente de datos, haciendo uso de estadística y auxiliándose del poder computacional de los ordenadores de hoy en día[1]. El empleo de algoritmos de aprendizaje automático nos ayuda a explicar el comportamiento de una variable con gran precisión a través del estudio de los atributos del objeto estudiado. Esto nos permite el empleo de esta herramienta en tantos campos como podamos imaginar.

Según el profesor Ethem Alpaydin (2009), los tipos de aprendizaje automático son:

1. Aprendizaje supervisado: El modelo aprende de datos etiquetados (entrada y salida conocidas).
2. Aprendizaje no supervisado: El modelo encuentra patrones en datos no etiquetados.
3. Aprendizaje por refuerzo: El modelo aprende a tomar decisiones a través de pruebas y recompensas [2].

2. Minería de datos

El término minería de datos es el campo de estudio que apunta directamente al análisis de grandes cantidades de datos almacenados en las bases de datos en los data warehouses. El punto focal de la minería de datos es el descubrimiento de conocimiento en las bases de datos. Para esto, la minería de datos se beneficia de diferentes campos de estudio o disciplinas que van desde expertos en sistemas, aprendizaje automático y estadística, así como personas que conozcan a profundidad el dominio de la base de datos estudiada para facilitar la extracción de conocimiento[1].

3. Big data

El big data es un término utilizado para describir datos extremadamente grandes y complejos de manejar con las herramientas de bases de datos tradicionales. De manera convencional, se atribuyen 5 características principales para identificar qué es big data. Estas 5 características se conocen como las 5 V del big data. Volumen, velocidad, variedad, veracidad y valor[1].

4. Analítica de datos

La analítica de datos es el estudio de los datos a través de herramientas estadísticas que nos permitan identificar el comportamiento de los mismos. Como herramientas estadísticas nos referimos a estadística descriptiva, patrones de distribución, modelamiento estadístico así como el empleo de visualización de datos. [1].

5. ¿Cómo se relaciona cada uno de estos 4 elementos con la ciencia de datos?

Como se muestra en la figura 3, todo el proceso de Ciencia de Datos comienza con los datos capturados y almacenados. Estos se convierten poco a poco en grandes volúmenes de datos que deben ser procesados mediante técnicas de minería de datos. El aprendizaje automático y las redes neuronales profundas, diseñadas para encontrar patrones de comportamiento de los datos almacenados y así generar predicciones, activan el proceso de minería de datos y aprendizaje profundo. Como resultado, existe una interdependencia entre el aprendizaje automático y la minería de datos.

Por otra parte, la inteligencia artificial, que toma como base estos algoritmos de aprendizaje automático, inspira a los equipos de ciencia de datos que analizan los resultados mediante la analítica de datos, inteligencia de negocios, estadísticas y visualización de datos para generar conocimientos relevantes para las personas u organizaciones que lo requieran.

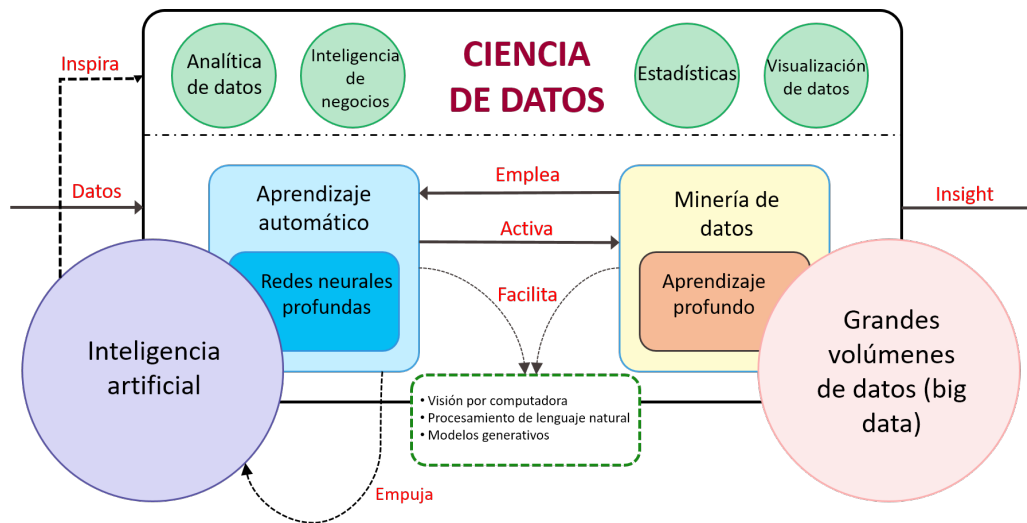


Figura 1: Diagrama de ciencia de datos [3]

Referencias

- [1] John D. Kelleher, Brendan Tierney *Data Science*, The MIT Press Essential Knowledge Series, 2018.
- [2] Ethem Alpaydin, *Introducción al Aprendizaje Automático*, Mit Pr, 2009.
- [3] The Data Science Puzzle, Explained. KDnuggets.
<https://www.kdnuggets.com/2016/03/data-science-puzzle-explained.html> (visitado 15-04-2024)