

Tecnológico de Monterrey

Conceptos básicos de ciencia de datos

Martín Alejandro Molinero Morales  
A01204818  
TC24039 Ciencia y analítica de datos  
Maestría en inteligencia artificial aplicada

¿Qué es aprendizaje automático?

El aprendizaje automático o machine learning es una disciplina dentro del campo de la inteligencia artificial que se centra en el diseño y evaluación de algoritmos que se centran en extraer patrones de un conjunto de datos. El machine learning se compone de técnicas computacionales y estadísticas avanzadas. La aplicación del aprendizaje automático permite que un agente aprenda, es decir que mejore su desempeño al hacer una tarea. Normalmente se busca que estas medidas de desempeño sean iguales o mejores a las de un humano realizando la misma tarea.

¿Qué es minería de datos?

La minería de datos es el proceso mediante el cual se aplica aprendizaje automático para analizar datos estructurados y encontrar patrones en ellos que le permitan obtener o generar conclusiones valiosas de tal manera que se puedan utilizar en un aspecto comercial. El término se podría considerar altamente equivalente en la comunidad académica al término descubrimiento de conocimiento en las bases de datos.

¿Qué es *big data*?

El big data es un término que se utiliza para referirse a la manera en la cual se recibe, almacena y analiza un gran número de datos, el cual fluye y crece exponencialmente a través del tiempo. De acuerdo a Kelleher y Tierney, se caracteriza en términos de las 3 v's: Volumen extremo de datos, Variedad de los tipos de datos obtenidos y Velocidad con la que los datos deben de ser procesados. Al final de la década de los 90 's se empieza a dar el auge del big data, esto debido a la creciente disponibilidad de conectividad que se obtuvo mediante el internet y la gran cantidad de datos generados por los dispositivos electrónicos. Para lograr el procesamiento de grandes volúmenes de datos, se han desarrollado herramientas especializadas como Hadoop.

¿Qué es la analítica de datos?

La analítica de datos es una disciplina que comprende diferentes procesos, en los cuales se busca obtener, transformar e inspeccionar un conjunto de datos de tal manera que se puedan obtener conclusiones valiosas de los mismos que puedan abonar en favor de tomar decisiones positivas para lograr un objetivo.

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

La ciencia de datos es una disciplina que abarca, usa o comprende todas las herramientas o ramas mencionadas anteriormente y otras para obtener conocimiento desde los datos, para poder así, tomar mejores decisiones. Considero que, la relación de cada rama mencionada se podría explicar mejor ejemplificando un proceso común de consumo de datos, que podría verse de la siguiente manera. Se pueden obtener y procesar un gran volumen de datos gracias a las herramientas derivadas del *big data*. Una vez obtenidos, organizados, transformados e inspeccionados mediante técnicas de

*análisis de datos*, pueden ser procesados mediante algoritmos de *aprendizaje automático* y *estadística*, los cuales ayudan a obtener patrones que permitan explotarlos a nivel de negocio mediante el *data mining*. Una vez obtenidos los patrones y el conocimiento de estos datos, se pueden representar mediante técnicas y herramientas de visualización de datos, que a su vez pueden ser presentados por los equipos de business intelligence dentro de la compañía para poder así, tomar decisiones informadas para maximizar la ganancia en la métrica a mejorar para un problema.

## Referencias

- EdX. (2023). Aprende sobre análisis de datos con cursos online. Recuperado el 18 de enero del 2023 desde <https://www.edx.org/es/aprende/analisis-de-datos>
- Kelleher, J. D., Tierney, B. (2018). Data Science. The MIT Press.
- Mayo, M. (2016). The data science puzzle, explained. KDnuggets. Recuperado el 18 de enero del 2023 desde <https://www.kdnuggets.com/2016/03/data-science-puzzle-explained.html/2>
- Shekhar, P. (2022). Key Players/Jobs in the Data Ecosystem. Recuperado el 18 de enero del 2023 desde <https://prashantshekhar.com/blog/key-players-jobs-in-the-data-ecosystem/>
- Vaughan, D. (2020). Analytical Skills for AI and Data Science. O'Reilly Media. Recuperado el 18 de enero del 2023 desde <https://learning.oreilly.com/library/view/analytical-skills-for/9781492060932/ch02.html#idm46388894678408>