

## ¿Qué es la Minería de Datos?

Como lo menciona la página de Microsoft “la minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos” (Microsoft, 2019). El objetivo general de la minería de datos consiste en extraer información de un conjunto de datos, trabajarla, pulirla y transformarla en una estructura comprensible para su uso posterior.

La forma de operar de la minería de datos consiste en la generación de un modelo de minería, este proceso se puede definir por medio de seis pasos los cuales son; definir el problema, preparar los datos, explorar los datos, crear modelos, explorar y validar modelos, implementar y actualizar modelos.

Este modelo de minería es un proceso dinámico e iterativo, “Una vez que ha explorado los datos, puede que descubra que resultan insuficientes para crear los modelos de minería de datos adecuados y que, por tanto, debe buscar más datos. O bien, puede generar varios modelos y descubrir entonces que no responden adecuadamente al problema planteado cuando los definió y que, por tanto, debe volver a definir el problema” (Microsoft, 2019).

**Definir el problema:** consiste en analizar los requisitos empresariales, así como definir el ámbito del problema, las métricas y los objetivos concretos del proyecto de minería de datos.

**Preparar los datos:** se tiene que saber con qué datos se trabajara para tener una limpieza de ellos, no solamente implica quitar los datos no válidos sino también buscar correlaciones escondidas en los datos.

**Explorar datos:** se tienen que conocer los datos para tomar las decisiones correctas, las técnicas de exploración que se usan son; calcular los valores mínimos y máximos, calcular la media y desviaciones estándar, y examinar la distribución de los datos.

**Generar modelos:** para llegar a este paso se necesita haber terminado la exploración de datos, aquí se definen las columnas de datos que se van a usar, se tendrá que crear una estructura de minería de datos. Esta estructura de datos se vinculará al origen de datos.

**Explorar y validar los modelos:** en este paso se comprobará su eficacia, antes de subir a producción el modelo de minería se tiene que testear para comprobar si funciona correctamente.

**Implementación y actualización de modelos:** el último paso, consiste en implementar los modelos que funcionan mejor en un entorno de producción.

Las características de la minería de datos son:

- Puede trabajar con grandes cantidades de información.
- El procesamiento es de hardware.
- La búsqueda de datos es mediante patrones.
- El resultado de los datos provee ayuda empresarial para tomar decisiones.

La función de la minería de datos es preparar, sondear y explorar datos para sacar información oculta en ellos.

En las tendencias encontramos:

- Distributed data mining: es la respuesta tecnológica a las necesidades que plantean las bases de datos distribuidas. Los algoritmos se centran en estas particulares tareas, prestando especial atención a todo lo relacionado con su análisis y modelado.
- Minería de datos en tiempo real: la extracción de conocimiento a partir de información para la toma de decisiones empresarial ha de ser capaz de aportar el dinamismo que los mercados requieren hoy día. Los datos se procesan en tiempo real y también deben poder ser analizados a este ritmo.
- Exploración de aplicaciones: explota las posibilidades de la prospección en los datos para aportar soluciones punteras a cada sector, contribuyendo a la innovación y garantizando un mejor servicio al cliente. Los campos de las finanzas, las telecomunicaciones o la investigación médica son algunos de los más favorecidos, aunque no los únicos.

## Referencias

Microsoft. (2019, enero 9). *Conceptos de minería de datos*. Microsoft. Consultado el 17 de enero de 2021. <https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>

Microsoft. (2018, mayo 5). Selección de características (minería de datos). Microsoft. Consultado el 17 de enero de 2021. <https://docs.microsoft.com/es-es/analysis-services/data-mining/feature-selection-data-mining?view=asallproducts-allversions>