

MINERIA DE DATOS Y BIG DATA

Minería de datos y big data cara a cara

Para poder hablar de minería de datos o data mining es preciso que exista el big data, pero no son lo mismo. El data mining también se conoce como exploración de datos y hace referencia a las técnicas que permiten analizar el grueso de la información para detectar patrones de conducta y tendencias que ayuden a las empresas a tomar decisiones en base a esa información que permanecía escondida.

Comprendiendo la minería de datos

Por lo tanto, descubrir pautas de comportamiento que se encuentran ocultas entre enormes volúmenes de información es el principal objetivo de la minería de datos.

Los proyectos de data mining se ejecutan en las siguientes fases:

Definición del objetivo

El primer paso es identificar cuál es el problema que se quiere resolver a través del minado. El propósito puede ser conseguir más clientes, crear estrategias para retener a los clientes actuales o tratar de reducir costes en algún proceso de la producción.

Identificación de los datos necesarios

Se trataría de definir qué datos será necesario recopilar para cumplir el objetivo planteado.

Formulación y pretratamiento

Aquí es cuándo comienza realmente el proceso de minería, ya que es el momento de seleccionar la información que resultará útil y proceder a su limpieza.

Modelado

En esta etapa se implantan los algoritmos matemáticos que definen la minería de datos a través de la elección de las herramientas más adecuadas en cada caso. Es el momento de configurar la información de manera que pueda resultar útil para el propósito definido en la primera fase.

Entrenamiento y evaluación

En este punto ya se ha construido un modelo matemático del que habrá que comprobar su eficacia en diferentes conjuntos de datos.

Verificación y explotación

En este último nivel es necesario revisar el modelo final e informar de los hallazgos que se produzcan. Además, es el momento de implantar el sistema de minado.

El papel del big data

Por otra parte, el big data es el término genérico utilizado para referirse al trabajo con ingentes cantidades de información que proviene de muy diversas fuentes. La tecnología big data es capaz de localizar, capturar, almacenar y gestionar los datos, de manera que se pueda sacar un rendimiento de ellos.

Los analistas de datos indican que el big data cumple con las llamadas cinco uves entre sus características:

La inmensa cantidad de datos a analizar es la principal cualidad del big data.

La información se genera a gran velocidad, lo que propicia que muchos datos se queden desfasados en poco tiempo.

La proveniencia de la información es muy heterogénea, pudiendo llegar desde redes sociales, dispositivos electrónicos, bases de datos, etc.

Este apartado es crucial, ya que mucha de la información obtenida no será verdadera. Por ello, es importante contar con herramientas de verificación competentes.

El último punto hace referencia al rendimiento que se les pueda sacar a todos esos datos, es decir, la capacidad de transformarlos en algo productivo.

Principales diferencias entre minería de datos y big data

Las diferencias principales entre big data y data mining podrían señalarse en los siguientes puntos:

El data mining implica una vista cercana sobre los datos, mientras que en el big data se ocupa de una exploración general de esos datos.

La minería se focaliza solamente en un tipo de datos, por ejemplo los estructurados, mientras que el big data puede trabajar con todo tipo de datos a la vez.

Data mining es una herramienta de vital importancia para desenterrar información que se encuentra oculta en grandes cantidades de datos. Los datos que se manejan en la minería pueden ser tanto grandes como pequeños, sin embargo, el big data se relaciona más con los procesos de tratamiento voluminoso de datos.

Minería de datos y big data utilizan diferentes procedimientos para manejar la información que analizan, pero son procesos complementarios más que antagónicos. Ambos tienen como objetivo transformar una enorme cantidad de datos en información relevante y útil para servir a los intereses para los que se ha decidido implementar esta tecnología.