



Bloque 1 • Datos por todas partes

Unidad 1 • Píldora 3

BIG DATA CUANDO LOS DATOS SE VUELVEN INTELIGENTES



BIG DATA: CUANDO LOS DATOS SE VUELVEN INTELIGENTES

Los datos y metadatos que se encuentran en la red no servirían de mucho si no se almacenaran, se clasificaran y se analizaran. La disciplina que se encarga de este proceso recibe el nombre de inteligencia de datos o *Big Data*.

A partir de estos macrodatos podemos generar conocimiento, productos y servicios. Por ejemplo, nos permiten predecir el tiempo, analizar parámetros de salud, mejorar la eficiencia energética o vender más y mejor. Por eso la inteligencia de datos está revolucionando la ciencia, la economía, la política y nuestro estilo de vida.

Se utiliza la inteligencia de datos cuando:



Introducimos mal una palabra de una búsqueda y Google nos la corrige.



Amazon nos muestra productos que podrían interesarnos según lo que hayamos comprado antes.



Snapchat nos descubre usuarios y noticias.



Facebook nos sugiere amistades con gente que es probable que conozcamos.



Spotify elabora una lista semanal de canciones según nuestros gustos.



LAS CUATRO UVES DEL *BIG DATA*

Estas primeras cuatro uves (ya veremos más adelante alguna más) sirven para recordar y describir las características principales del *Big Data*: la gestión de un gran **volumen** de datos, a la mayor **velocidad** posible, almacenados junto a una extensa **variedad** de información, que debe estar **verificada**.



El crecimiento de los datos en la red es constante. Por eso se dice que "los datos nunca duermen". Ya sea de día o de noche, las 24 horas y en cualquier parte del mundo, se generan datos. Esta producción ininterrumpida se dobla cada 40 meses, lo que quiere decir que se generan más datos en un día de los que han existido en los últimos 20 años.



Imprescindible en la creación y análisis de los datos. Dado que estamos generando datos constantemente, necesitamos rapidez para disponer de esa información en tiempo real. La información se procesa tan rápido que podemos conocer datos del presente inmediato y consultar qué está ocurriendo en el mundo ahora mismo. Esto nos permite hacer análisis bastante detallados y complejos que a menudo se integran en otros procesos de trabajo y sistemas.



Hay datos de muchos formatos y tipologías, según su procedencia. Podemos clasificar los macrodatos en:

DATOS PÚBLICOS

Datos que tienen las administraciones públicas (por ejemplo, datos sobre transporte, uso de energía, sanidad, etc.).

DATOS PRIVADOS

Datos derivados de transacciones comerciales, de la navegación web, del uso de la telefonía móvil, etc.



DATOS COMUNITARIOS

Datos producidos principalmente en las redes sociales, contenidos generados por el usuario, etc.

DATOS *QUANTIFIED SELF*

Datos obtenidos y proporcionados por las propias personas que miden y cuantifican sus comportamientos y acciones. Por ejemplo, los datos monitorizados sobre las pulsaciones durante la realización de ejercicio físico que son recogidos por dispositivos móviles.



Los datos obtenidos deben ser fiables, íntegros y auténticos, por lo que es necesario confirmar su veracidad. Y ¿cómo sabemos si los datos son válidos? Esto dependerá de las fuentes y los recursos que hayamos empleado para obtenerlos.

