

Actividad 2: Rasgos característicos de Big Data

Descripción: Define las cinco características de Big Data: volumen, variedad, velocidad, veracidad y valor. Menciona dos ejemplos representativos de cada característica.

Condiciones de satisfacción: Claridad en la definición de cada característica y dos ejemplos representativos de cada una.

Elabora: Moisés Marín

Volumen - Es la característica que describe el tamaño de los grandes volúmenes de datos, usualmente se cruza la línea que divide los bajos y los grandes volúmenes de datos cuando tienen un tamaño que requiere estrategias especiales para almacenarlos y manipularlos.

Por ejemplo:

1. Google procesa cerca de 4.5 millones de búsquedas cada segundo [1], cada una es una búsqueda en un índice de más de 100,000,000 de gigabytes de tamaño.[2]
2. 55,140 fotos se comparten en Instagram cada minuto [1], si suponemos que cada foto tiene un tamaño de 1 megabyte entonces Instagram recibe 55.14 gigabytes cada minuto, o bien 79.40 terabytes al día.

Variedad - Es la característica que describe el origen y los tipos de datos. Los datos pueden provenir de redes sociales, captura de imágenes o sensores.[3] En cuanto a tipos se clasifican en estructurados y no estructurados. Es la principal característica por la que las empresas recurren a soluciones de grandes volúmenes de datos.[4] Algunos autores utilizan una tercera categoría: datos semi-estructurados.[5]

Por ejemplo:

1. datos en una base de datos con un esquema definido son datos estructurados
2. datos provenientes de sensores, imágenes o correos electrónicos son datos no estructurados.[3]

Velocidad - Es la característica que describe la frecuencia con la que los datos son creados, almacenados y procesados.[6]

Por ejemplo:

1. procesamiento de datos en tiempo real para detección de fraudes en transacciones bancarias
2. monitoreo de eventos en redes sociales

Veracidad - Esta característica describe la fiabilidad de los datos recibidos. Invertir tiempo en asegurar la calidad de los datos, la confiabilidad del origen de los datos así como de su procesamiento son factores importantes para incrementar la veracidad de los grandes volúmenes de datos.[7]

Por ejemplo:

1. retirar la imparcialidad, anormalidades, inconsistencias y duplicación en datos recopilados en redes sociales
2. verificar que el procesamiento de datos se realiza de una manera adecuada para satisfacer necesidades del negocio

Valor - Esta característica describe la capacidad de tomar decisiones basadas en los grandes volúmenes de datos, es decir, qué tan accionables son y cuánto conocimiento puede extraerse de ellos.

Por ejemplo:

1. las tecnologías de procesamiento de lenguaje natural que permiten obtener el sentimiento positivo o negativo de publicaciones en redes sociales
2. la aseguradora MetLife puede estimar con dos meses de anticipación cuando un cliente planea cancelar una póliza o adquirir una nueva.[8]

Referencias

- [1] Karki, D. (2020). Can you guess how much data is generated every day? Febrero 2, 2021, de takeo.ai. Sitio web:
<https://www.takeo.ai/can-you-guess-how-much-data-is-generated-every-day/>
- [2] Google.com (2021). How Search organizes information. Febrero 2, 2021, de google.com. Sitio web:
<https://www.google.com/search/howsearchworks/crawling-indexing/>
- [3] Dumbill, E. (2012). Volume, Velocity, Variety: What You Need to Know About Big Data. Febrero 2, 2021, de Forbes.com. Sitio web:
<https://www.forbes.com/sites/oreillymedia/2012/01/19/volume-velocity-variety-what-you-need-to-know-about-big-data/?sh=6ad17cdf1b6d>
- [4] Bidadi Raju, M. (2020) Data Analytics | Big data | Case study: V's of Big data. Febrero 2, 2021, de towardsdatascience.com. Sitio web:
<https://towardsdatascience.com/data-analytics-big-data-case-study-vs-of-big-data-1d3dc5118759>
- [5] Wolff, R. (2020) What Is Semi-Structured Data & How to Analyze It. Febrero 2, 2021 de Monkey Learn. Sitio Web:
<https://monkeylearn.com/blog/semi-structured-data/>
- [6] Instituto de Ingeniería del conocimiento (2020) Las 7 V del Big data: Características más importantes. Febrero 2, 2021 de Instituto de Ingeniería del Conocimiento. Sitio Web:
<https://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/>
- [7] GutCheck (2019) Veracity: The Most Important “V” of Big Data. Febrero 2, 2021 de gutcheckit.com. Sitio Web:
<https://www.gutcheckit.com/blog/veracity-big-data-v/>
- [8] BBVA (2020) The five V's of big data. Febrero 2, 2021 de bbva.com. Sitio Web:
<https://www.bbva.com/en/five-vs-big-data/>