



UNIVERSIDAD  
**SAN IGNACIO  
DE LOYOLA**

**FACULTAD DE INGENIERÍA**  
**Ingeniería Informática y de Sistemas**

Proyecto Desarrollo Basado en Plataformas

**Curso:**  
**Desarrollo Basado en Plataformas**

**Alumnos:**  
JARA PEÑA, ANGIE JESSENIA  
ORTIZ VALDIVIA, AARON EDUARDO

**Docente:**  
Ing. Diaz Arenas, Daniel Jesus

**Lima – Perú**  
**2023-1**

## Ciencia de Datos - Big Data

Hoy en día, con la globalización de las tecnologías y su uso indispensable para cualquier actividad, la mayoría de las personas cuentan con redes sociales, blogs, cuentas en distintos sitios web, etc; todos estos medios en donde existe un usuario, detrás hay una “avalancha” de datos la cual muchas empresas buscan obtener con la finalidad de poder mejorar sus tomas de decisiones y movimientos de negocio estratégicos. Debido al auge de las tecnologías, estos datos tuvieron un crecimiento exponencial y analizar estos, presenta un desafío, ya que son voluminosos y en su mayoría son datos no estructurados, por lo que, los datos importantes son difíciles de analizar. Frente a este desafío que se presenta con el análisis de datos, es que surge la Ciencia de Datos y Big Data.

Frente al problema que presenta el análisis de datos, un estudio realizado por Vedant Agarwal (2023), relaciona el concepto de múltiples universos en paralelo con cómo es que funciona Big Data en distintos softwares. Agarwal, relaciona la teoría de Schrodinger y Deutsch de múltiples universos paralelos, y el funcionamiento del procesamiento de datos en Big Data; un software que tiene esta lógica es MapReduce. Ambos conceptos toman provecho del paralelismo para procesar grandes volúmenes de datos de manera eficiente, dividiendo el trabajo en tareas independientes, por lo que, esto otorga una ejecución simultánea y acelerada, similar a la existencia de universos paralelos trabajando en conjunto. Al considerar el enfoque del "multiverso de datos", donde diferentes transformaciones y variables generan resultados estadísticos únicos, se puede abordar de manera más completa el análisis de datos en busca de soluciones más robustas e insights más precisos.

Si bien, Big Data es una tecnología actual, no muchas PYMES y MYPES cuentan con los recursos necesarios para desarrollar o emplear softwares orientados al análisis de datos masivos, ya que estos suelen ser costosos y requieren de muchos recursos para mantenerlos. Es por eso, que Song Chen, para superar estas dificultades, propone como idea el diseño de un sistema de procesamiento de big data basado en algoritmos genéticos. Además plantea, que si seguimos con el desarrollo de tecnologías informáticas, se puede llegar a superar esta dificultad, debido a estas, ofrecen características como la expansión elástica, la agrupación de recursos y el pago por uso, lo que genera un fácil acceso a recursos informáticos de gran escala. También, se presenta la idea de emplear tecnologías de equilibrio de carga para ofrecer una mayor escalabilidad del cluster de servicios, lo que permitiría manejar grandes volúmenes de datos fragmentados de distintas fuentes.

Por otro lado, la ciencia de datos también ha logrado jugar un papel importante en el campo de la nutrición precisa en la integración y análisis de la información necesaria para desarrollar recomendaciones nutricionales individuales. Al combinar bases de datos de fuentes con tecnologías "combinadas" y otra información relevante, la ciencia de datos permite obtener información significativa. Utilizando técnicas avanzadas de análisis de datos, como el aprendizaje automático y la extracción de datos, los investigadores pueden analizar grandes cantidades de datos y descubrir patrones complejos, correlaciones y relaciones entre la

composición de los alimentos, los perfiles genómicos y otros factores de salud. Esto ayuda a comprender mejor cómo interactúan los nutrientes con los genes y las características individuales, y cómo estas interacciones pueden afectar la respuesta de una persona a los alimentos.

Los avances en las tecnologías de la información y la comunicación (TIC) han facilitado la integración de la ciencia de datos en la nutrición de precisión. En la actualidad, constantemente se crean y almacenan grandes volúmenes de datos en diversos sectores sociales, dando origen al término “big data” (BD). Además de su tamaño, BD se caracteriza por su rapidez, heterogeneidad y diversidad de información. La ciencia de datos, con sus enfoques y herramientas especializados, permite la gestión y el análisis efectivos de estos enormes datos y recopila información valiosa para una alimentación precisa. Así, la intersección de bases de datos y bases de datos nutricionales se convierte en un factor clave para personalizar las recomendaciones dietéticas y comprender las complejidades de la relación entre alimentación y salud.

Las tecnologías moleculares y genómicas avanzadas, como la secuenciación de próxima generación y la metabolómica, han revelado una fascinante diversidad microbiana y sus interacciones en los ecosistemas. Estas técnicas nos permiten comprender las estructuras de la comunidad microbiana y los patrones fundamentales de su composición y dinámica a escala global. Además, los enfoques fisiológicos como la microespectroscopia Raman ofrecen oportunidades para el análisis no destructivo de la composición química de las células microbianas. Estos avances demostraron el papel central de los microorganismos en las funciones y servicios ecológicos de los ecosistemas.

La validación de modelos predictivos y el análisis de datos experimentales son esenciales en este campo, la ciencia de datos aplicada en la ciencia del clima. Se propone una metodología sistemática y automatizada que utiliza el término "ecosistema de datos" para analizar tanto datos como modelos. Este método se puede aplicar en varios campos, incluida la ingeniería química, donde los modelos predictivos se validan utilizando datos experimentales. Se considera como caso de estudio la evaluación de un modelo cinético que describe la pirólisis y combustión de hidrocarburos.

## Bibliografía

- *Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad.* (s/f).

Powerdata.Es. Recuperado el 12 de junio de 2023, de

<https://www.powerdata.es/big-data>

- Agarwal, V., Kumar, K. P., CyrusManoj, K. P., & Prathap, B. R. (2023). Comprehensive study of the relationship between multiverse and big data. *Measurement. Sensors*, 27(100763), 100763 .  
<https://doi.org/10.1016/j.measen.2023.100763>
- Chen S. Design of computer big data processing system based on genetic algorithm. Research Square; 2023. DOI: 10.21203/rs.3.rs-2555410/v1.
- Wong-González, E. (2010). ¿ Después de un análisis de variancia qué? Ejemplos en ciencia de alimentos. *Agronomía mesoamericana*, 21(2), 349-356.  
[https://www.scielo.sa.cr/scielo.php?pid=S1659-13212010000200016&script=sci\\_arttext&tlng=en](https://www.scielo.sa.cr/scielo.php?pid=S1659-13212010000200016&script=sci_arttext&tlng=en)
- Puyol Moreno, J. (2014). Una aproximación a Big Data= Una aproximación a Big Data.  
<http://e-spacio.uned.es/fez/eserv/bibliuned:rduned-2014-14-7150/Documento.pdf>
- Aguilar, LJ (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones* . Editor Alfaomega Grupo.  
[https://books.google.es/books?hl=es&lr=&id=1GywDAAAQBAJ&oi=fnd&pg=PT6&dq=big+data&ots=\\_XVaU71cZL&sig=ThymvxMw8ea8Fao0yHpRhr\\_P2Fs#v=onepage&q=big%20data&f=false](https://books.google.es/books?hl=es&lr=&id=1GywDAAAQBAJ&oi=fnd&pg=PT6&dq=big+data&ots=_XVaU71cZL&sig=ThymvxMw8ea8Fao0yHpRhr_P2Fs#v=onepage&q=big%20data&f=false)
- Hernández-Leal, EJ, Duque-Méndez, ND, & Moreno-Cadavid, J. (2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *TecnoLógicas* , 20 (39), 17-24.  
[http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0123-77992017000200002](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-77992017000200002)