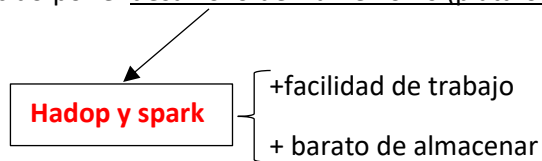


Tema 1	2
1. Introducción	2
2. IA → Inteligencia Artificial	2
3. ¿Qué tipo de tareas podemos realizar?	3
4. Tipos de datos	3
5. Fuentes de Datos	4
6. Plataforma y programabilidad: Tecnología	4
7. Características de Big Data	4
8. Visión estratégica: las 5 P del Big Data y los proyectos de datos	4
9. Definición de Big Data	5

Tema 1

1. Introducción

El crecimiento de Big data se ha visto favorecido por el desarrollo de frameworks (plataformas de código abierto).



Con la aparición de Internet of things (IoT), más objetos y dispositivos están conectados a internet, recopilando datos sobre los patrones de uso de los clientes y además la aparición del aprendizaje automático ha permitido el aprovechamiento de estos datos.

La computación en la nube ha ampliado la utilidad de Big Data.

Big Data se fundamenta en dos ideas → almacenamiento y análisis de datos.

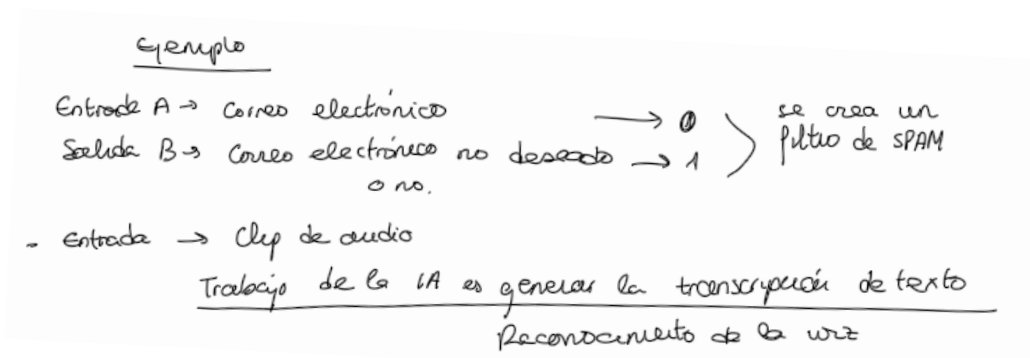
No se puede entender sin los nuevos desarrollos y algoritmos para entender los datos

Así surgen términos como; Aprendizaje estadístico, inteligencia artificial, machine learning y Deep learning

Debido a la aparición del aprendizaje automático y otras disciplinas el aprendizaje estadístico ha surgido como un nuevo subcampo en estadística centrado en el modelo y la predicción supervisada y no supervisada.

¿Cuáles son las fuerzas que han fomentado la era BIG DATA? Son 3 fuerzas;

- 1) Volumen y variedad de datos.
- 2) Computación en la nube.
- 3) Aplicación y desarrollo de algoritmos aplicables a situaciones reales para crear nuevo valor.



2. IA → Inteligencia Artificial.

Su aumento ha sido impulsado por una herramienta; **el aprendizaje automático**.

- El tipo más usado de aprendizaje automático es un tipo de IA que aprende de A o B, podemos definirlo como **Aprendizaje supervisado**.

La IA ha despegado recientemente debido al aumento de las redes neuronales y gracias al **aprendizaje profundo**.

¿Cómo alcanzamos un nivel óptimo de rendimiento? → Necesitamos dos cosas; Tener gran cantidad de datos y ser capaz de entrenar una red neuronal grande.

Muchas compañías pueden entrenar grandes redes neuronales con una cantidad suficientemente grande de datos, obteniendo un rendimiento bastante bueno gracias al crecimiento de la velocidad de computación y de los procesadores especializados (uds de procesamiento de gráficos o las GPU).

3. ¿Qué tipo de tareas podemos realizar?

Los tipos de tareas que pueden realizar los algoritmos teniendo los suficientes datos disponibles son:

Aprendizaje
supervisado

- **Descripción** → cuando se combinan diferentes conjuntos de datos en un solo lugar o se muestran en un mapa, se puede obtener una mejor comprensión al verlo desde varias perspectivas. Esto puede ayudar a tomar mejores decisiones. Siempre hay nuevas formas de mejorar la forma en que se presentan los datos para mejorar la toma de decisiones.
- **Inferencias** → conocer las causas subyacentes de un fenómeno y la influencia de cada factor en el resultado es esencial para poder tomar medidas efectivas para cambiarlo.
- **Predicciones** → con los algoritmos, se pueden hacer predicciones.
- **Clasificaciones** → clasifica los elementos en los grupos a los que pertenecen. Esta tarea es especialmente adecuada

Aprendizaje
no supervisado

- **Agrupamientos/Clustering** → Clustering es una técnica que consiste en agrupar objetos similares juntos en un grupo, de tal manera que los objetos en un mismo grupo sean más parecidos entre sí que con los objetos en otros grupos. Esta técnica es útil en marketing para segmentar a los clientes y entender mejor sus necesidades y comportamientos.
- **Recomendaciones** → son como una predicción, recomienda algo que quieres. Hay muchas técnicas; como el filtrado colaborativo o algoritmos de aprendizaje automático, el análisis de redes para crear motores de búsqueda como Google™.
- **Sistemas cognitivos** → un sistema cognitivo es un enfoque más avanzado y complejo para la automatización y la toma de decisiones en situaciones complejas.

4. Tipos de datos

¿Por qué existe Big Data? Existe gracias a la medida actual en la que la información puede generarse y ponerse a disposición.

Diferencia entre digitalización y dataficación

- Digitalización son datos en formato digital, y su siguiente paso es la data, que consiste en describir o enunciar un fenómeno en un formato cuantificado para que pueda ser tabulado y analizado.

Conclusión y diferencia fundamental

- La digitalización permite que la información analógica se transfiera y almacene en un formato digital más conveniente.
- La dataficación garantiza la versión digitalizada de las señales analógicas para generar información que no se había inferido con estas señales en su forma original.

El rendimiento depende del volumen y la variedad de datos.

- SQL → Elementos que podemos almacenar en 1 tabla.
- BIG DATA → Dato estructurado (tabla) o semiestructurado.

5. Fuentes de Datos

- 1) Máquinas
- 2) Gente
- 3) Corporaciones

6. Plataforma y programabilidad: Tecnología

El tamaño de los datos y la complejidad del procesamiento requieren un almacenamiento y rendimiento computacional adecuado. Según Google Trends, "**Hadoop**" es la consulta más asociada con "Big Data" y es la tecnología más destacada en este campo. Hadoop es un framework de código abierto que permite el procesamiento distribuido de grandes cantidades de datos utilizando un grupo de máquinas y modelos específicos de programación informática.

Principales componentes de Hadoop →

- **El sistema de archivos HDFS de Hadoop** facilita el acceso y la gestión de datos que están distribuidos en un entorno de almacenamiento complejo y disperso.
- **MapReduce** es un modelo de programación que se ha diseñado específicamente para implementar algoritmos distribuidos y paralelos de manera eficiente. Sin embargo, hay ciertas tareas que no pueden ser realizadas únicamente utilizando el algoritmo de computación MapReduce. Por ello, en los últimos años, se han agregado muchos otros complementos y módulos al ambiente de Hadoop para ampliar sus funcionalidades y mejorar su capacidad de procesamiento de datos.

7. Características de Big Data

Definición de Big Data → Conjunto de datos más grandes y complejos, especialmente de nuevas fuentes de datos. Tienen tanto volumen que el software tradicional no puede administrarlo.

¿Cuáles son las Vs del Big Data?

- 1) **Volumen** de los datos.
- 2) **Velocidad** a la que se reciben los datos y se actúa.
- 3) **Variedad** de los datos disponibles.
- 4) **Veracidad** (que cantidad de precisión puede ser un conjunto de datos)
- 5) **Valencia** → es la conexión entre los elementos de datos. Los gráficos de redes pueden ser útiles para visualizar y analizar estas conexiones de datos.
- 6) **Valor** → aspecto fundamental en Big Data. Todo debe ir enfocado a crear valor para el negocio, personas y/o sociedad.

8. Visión estratégica: las 5 P del Big Data y los proyectos de datos.

Distinguimos 5 elementos que determinan la creación de valor en una visión estratégica de un proyecto de datos.

- 1) **Propósito.**
- 2) **Personas.**
- 3) **Procesos** → captura el valor de las fuentes de datos y lo transforma en valor, hay dos tipos de procesos.
 - Procesos técnicos
 - Procesos organizativos.
- 4) **Plataformas** (arquitectura del sistema, el diseño e implementación de hardware...)
- 5) **Programabilidad** (herramientas y lenguajes de programación).

9. Definición de Big Data

Se refiere a **4 temas clave**; información, tecnologías, métodos e impacto o creación de valor.

La generación de valor mediante una propuesta que comienza con la recolección de datos provenientes de diversas fuentes, incluyendo máquinas, personas y corporaciones. Estos datos son voluminosos, variados, veloces, veraces y conectados, y son analizados utilizando una variedad de algoritmos y técnicas para describirlos, predecirlos, clasificarlos, agruparlos y hacer recomendaciones. La plataforma en la que se ejecutan estos algoritmos cuenta con un entorno de programación que culmina en una visualización, métrica, ranking, recomendación o alerta, que sirve para facilitar el proceso de toma de decisiones o para una acción automática.