

---

---

# Big Data y Machine Learning con Python

Proyecto de fin de ciclo: **Juan Medina**

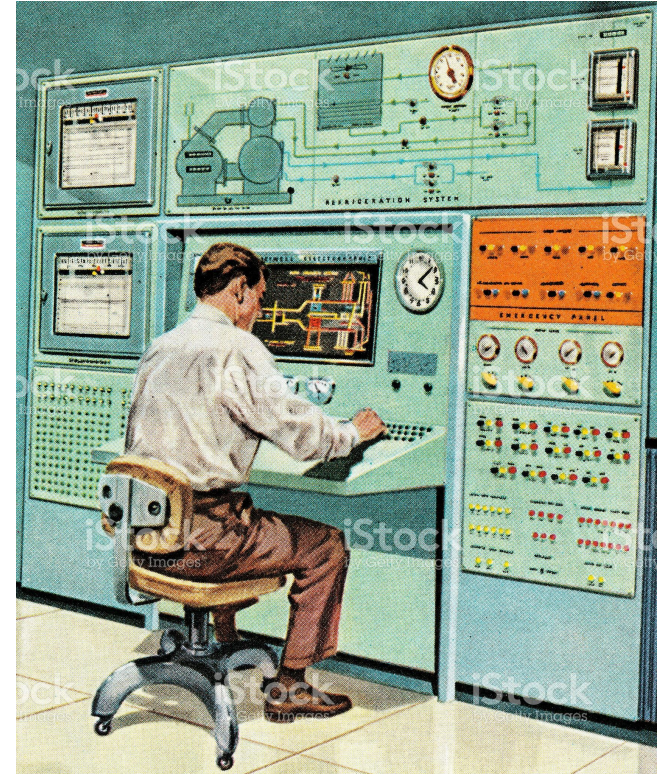
---

---

# ¿Qué es Big Data? - De dónde vienen los datos

Al principio

- Introducción manual de datos
- Pocos operarios
- Poca capacidad de almacenamiento



# ¿Qué es Big Data? - De dónde vienen los datos

## Dispositivos

- Conexión a internet de alta velocidad
- Cada vez más potentes
- Proporcionan información de localización

- Más grandes
- Más rápidos
- Más asequibles

### Dispositivos Móviles



### Sensores

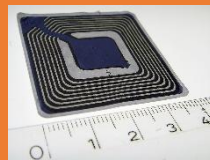


- Medidores inteligentes de energía
- Dispositivos médicos
- Vehículos
- Edificios

### Discos Duros



### RFID



- Etiquetas de radiofrecuencia

# ¿Qué es Big Data? - De dónde vienen los datos

## Aplicaciones

### Open Data



Datos para su uso libre:  
meteorológicos, socio  
económicos, de energía,  
servicios públicos,....

### Redes Sociales



Millones de publicaciones  
cada día: texto, localización  
geográfica, fotos, vídeo,...

### Smart Cities



Aplicaciones y sistemas para  
la mejora de la ciudad

### Internet of the Things



Redes de dispositivos que se  
comunican de forma  
autónoma: smartphones,  
electrodomésticos,  
vehículos,...

# ¿Qué es Big Data? - De dónde vienen los datos

## Sitios Open Data

- España: <http://datos.gob.es/>
- Irlanda: <https://data.gov.ie/>
- EEUU: <https://www.data.gov/>
- Francia: <https://www.data.gouv.fr/>

# ¿Qué es Big Data?

## Sistema de clasificación

- Una estrella: Datos accesibles en cualquier tipo de formato
- Dos estrellas: Datos accesibles organizados en un formato propietario (Datos en Excel)
- Tres estrellas: Datos accesibles organizados en un formato no propietario (csv en lugar de Excel)
- Cuatro estrellas: Los datos además poseen una URL única y los datos se estructuran bajo el estándar W3C
- Cinco estrellas: Los datos además están enlazados con otros datos son una estructura similar.



LOD

Conjunto de datos abiertos de fuentes distintas enlazados bajo estándares similares



RDF

Datos abiertos estructurados en URIs bajo estándares W3C



CSV

Datos abiertos estructurados en software libre



XLS

Datos abiertos estructurados en un software propietario



PDF

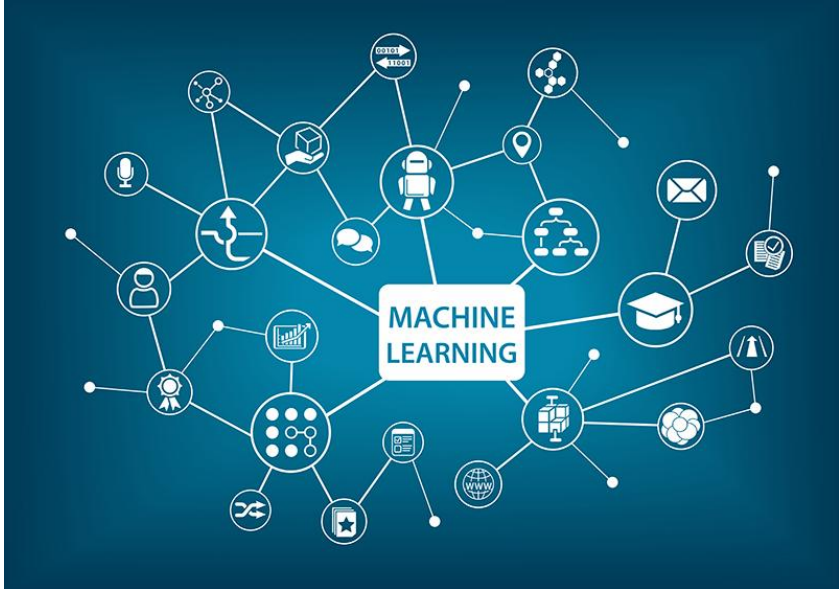
Datos abiertos disponibles en la red, pero sin estructurar y sin formato

# ¿Qué es Big Data?



# ¿Qué es Machine Learning?

El Machine Learning, o aprendizaje automático, se refiere a la ciencia que permite a los ordenadores aprender sin que hayan sido explícitamente programados para algo, sino que son capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos, de este modo son capaces de utilizar su experiencia para tomar decisiones futuras más acertadas





# ¿Qué es Machine Learning? - Tipos de aprendizaje

- **Aprendizaje supervisado:** Se le proporciona al programa una serie de datos que incluyen la solución correcta.
- **Aprendizaje no supervisado:** Solo se le proporciona una serie de datos al programa que el mismo debe clasificar en grupos encontrando patrones en los mismos.
- **Aprendizaje por refuerzo:** Esta clase de entrenamiento se basa en la técnica de ensayo/error. El programa predice una respuesta y se le entrega la solución, cuantas más veces se ejecute más posibilidad de acierto tendrá.

# Kaggle (<https://www.kaggle.com/>)

- **Desarrollado tras el concurso de Netflix**
- **Existen plataformas similares**
  - TopCoder (<https://www.topcoder.com/>),
- **Adquirida por Google**