

CIENCIA DE DATOS

GRADO EN BIOMEDICINA

DR. ARIEL CARIAGA-MARTÍNEZ

CIENCIA DE DATOS: Organización de los trabajos

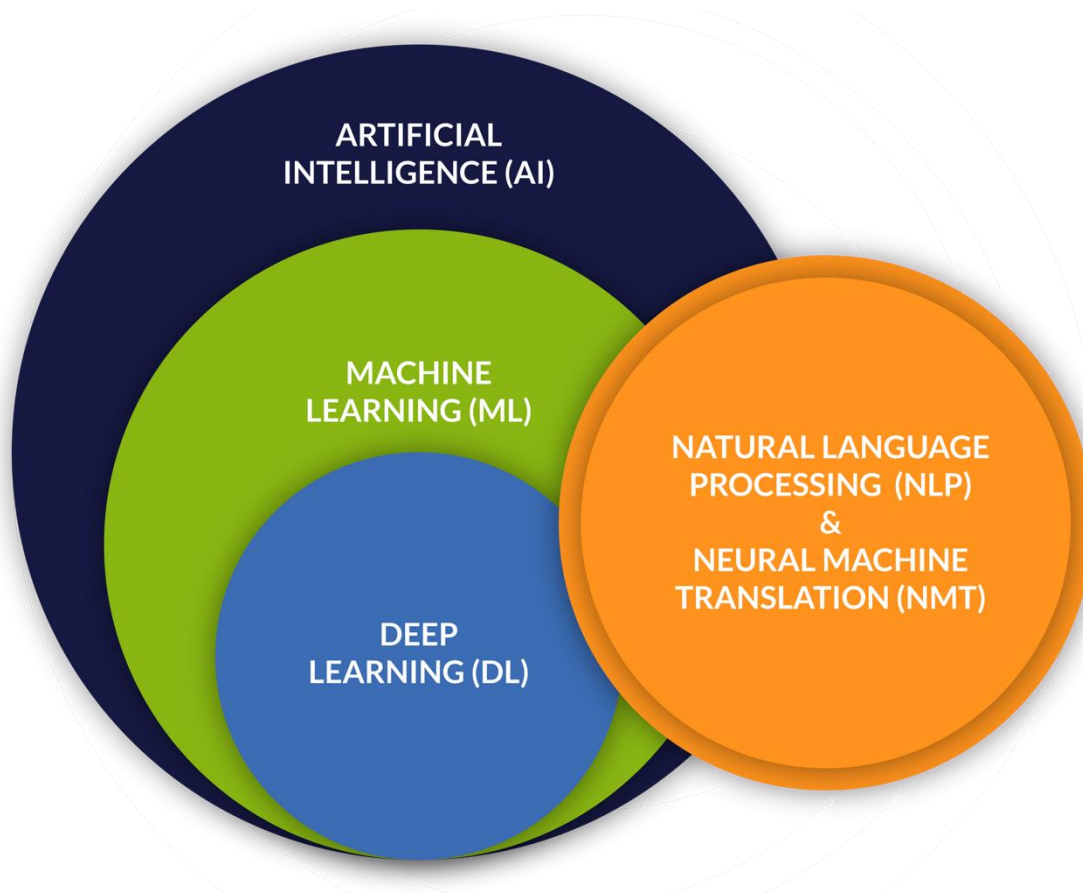
Aplicada – Técnica vs tecnología –
Clave: Bases (presentaciones)
AVANCES EN EL PROYECTO FINAL

- A) Porción teórica con seguimiento.
- B) Porción de trabajos: co-crearemos el conocimiento.

Objetivo final:

- Ser capaces de realizar y defender un proyecto básico de ciencia de datos.
- No hay que “entregar” nada de los trabajos: se pueden enviar cuando haya dudas.
- Repositorio GIT → https://github.com/cariagamartinez/uax_ds.git

¿Qué es la Ciencia de datos?



- Ciencia.
- Datos.

¿Cómo la abordamos?

<https://www.ibm.com/topics/data-science>

DATOS Y FORMATOS

(RAE) Información necesaria y que sirve de fundamento para llegar al conocimiento exacto de algo.

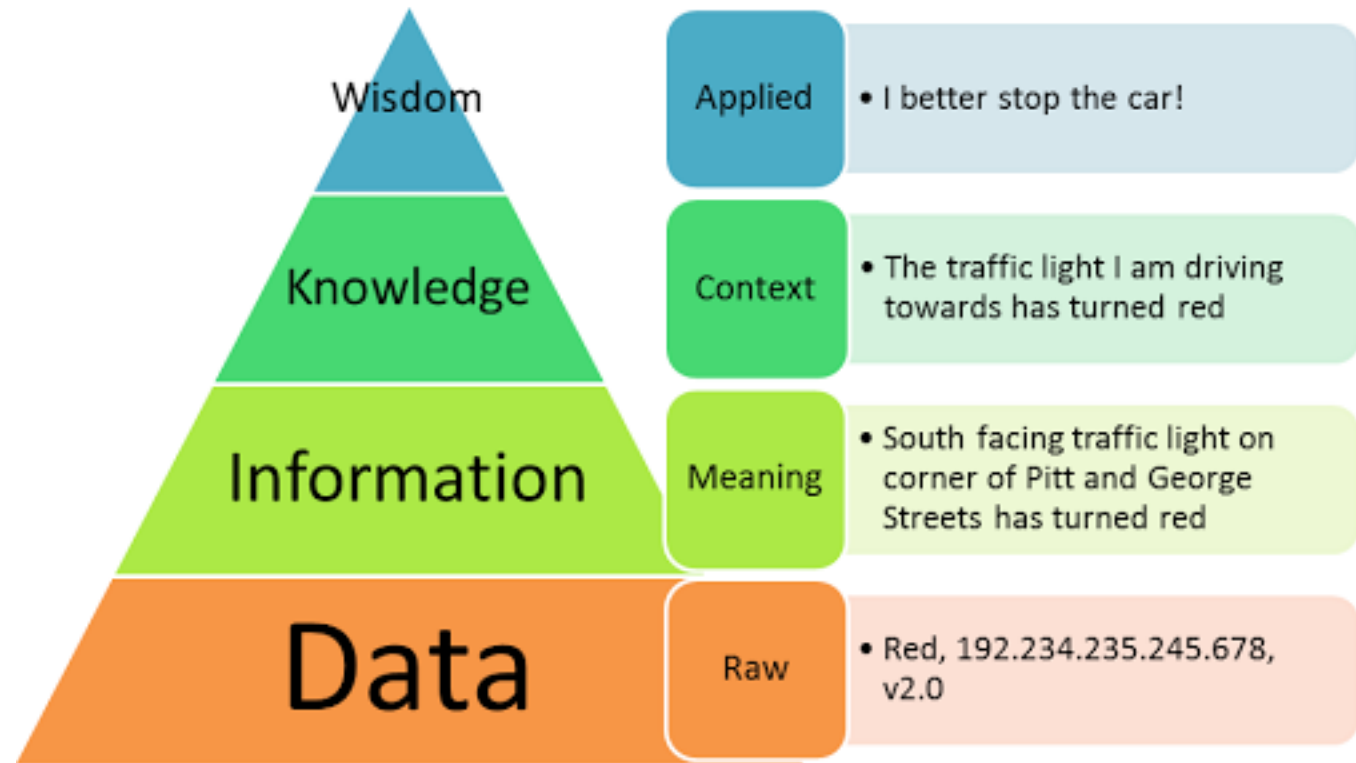
Un dato es un conjunto discreto, de factores objetivos, sobre un hecho real.

Representación simbólica de una transacción.

Formato (tipo) vs Formato (informático) → estructura, no estructura, semi-estructurado, grafos, lenguaje natural, generado por máquina, audio, video, “streaming”.

<https://manoli-iborra.github.io/BigData/index.html>

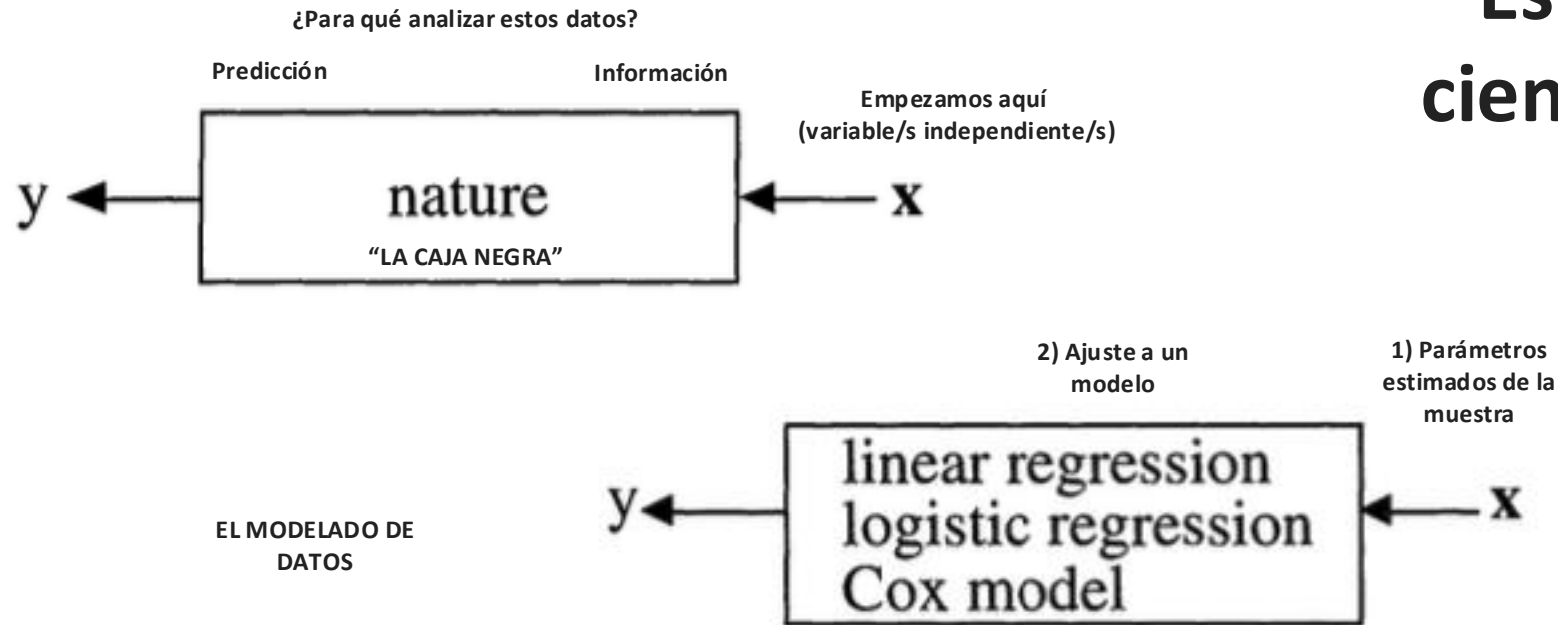
¿Qué es la Ciencia de datos?



- No es programación: [a buena parte de la] programación “no le importan los datos”: “leyes para ordenadores”.
- No es estadística: agrega partes como big-data (específicos) y aplicaciones dirigidas.
- No es BI: (retrospección) es también prospección y se aplica a más áreas que los negocios.
- No es ML (solamente): ciencia de datos incluye todo lo necesario para la ingesta de datos y su presentación.

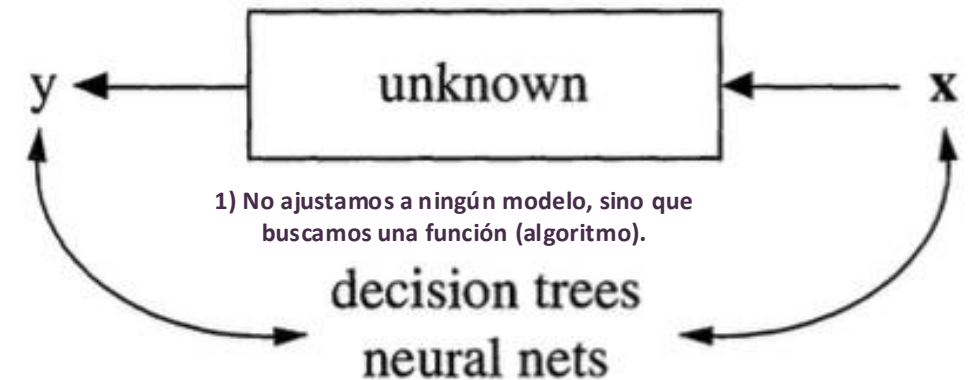
La ciencia de datos es interdisciplinar e implica extraer conocimiento a partir de datos.

Estadística y ciencia de datos



EL MODELADO
ALGORITMICO

2) Si la predicción "es buena", el "modelo" es bueno.



Statistical Science
2001, Vol. 16, No. 3, 199-231

Statistical Modeling: The Two Cultures

Leo Breiman

ALGUNAS DIFERENCIAS

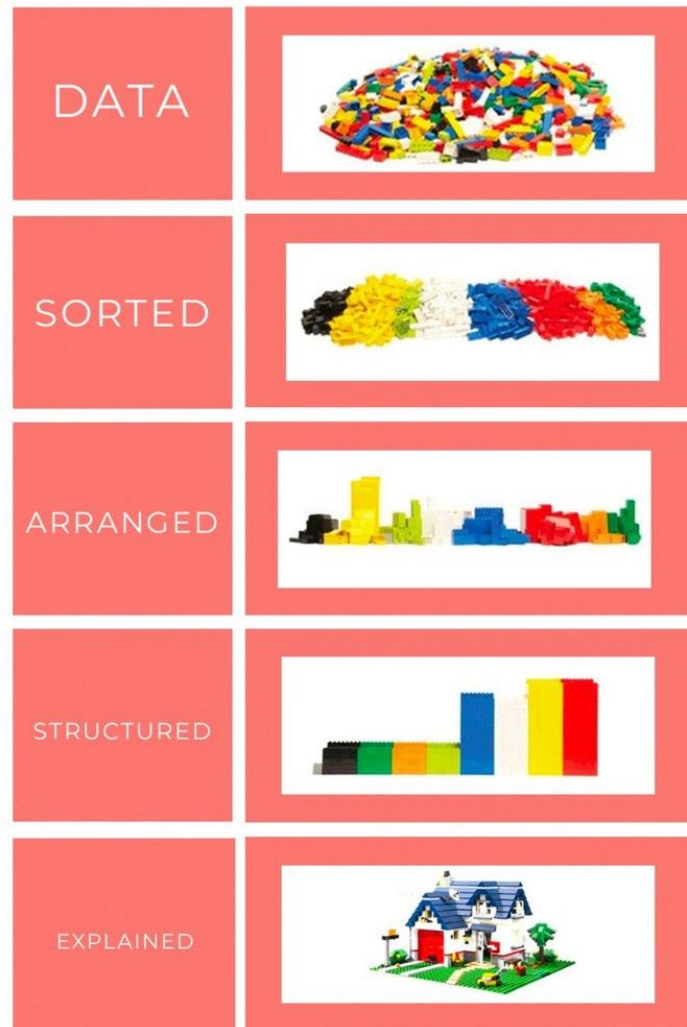
Analista de datos: perfil técnico (de datos a información)

Científico de datos: modelos (información a conocimiento): matemática más estadística

Ingeniero de datos: diseño y mantenimiento del soporte/almacenamiento/carga, etc. Puede ser un perfil técnico/universitario.

Arquitecto de datos: genera la infraestructura necesaria para el tratamiento de (grandes cantidades de) datos. Seguridad, gobernanza, escalabilidad, etc.

Fases y herramientas en Ciencia de Datos



¿De qué forma recogemos los datos?

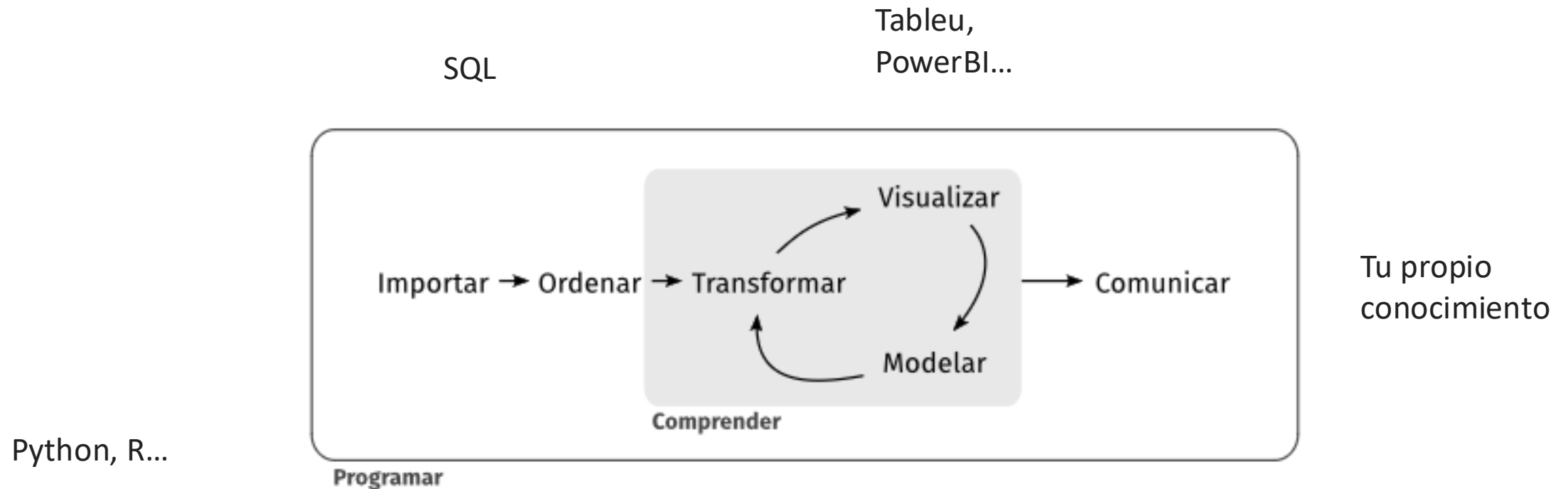
¿Cómo los podríamos organizar?

¿Cómo los podríamos transformar?

¿Cómo los podríamos modelar?

¿QUÉ HISTORIA (CONVINCENTE) PODEMOS CONTAR?

Fases y herramientas informáticas en Ciencia de Datos.

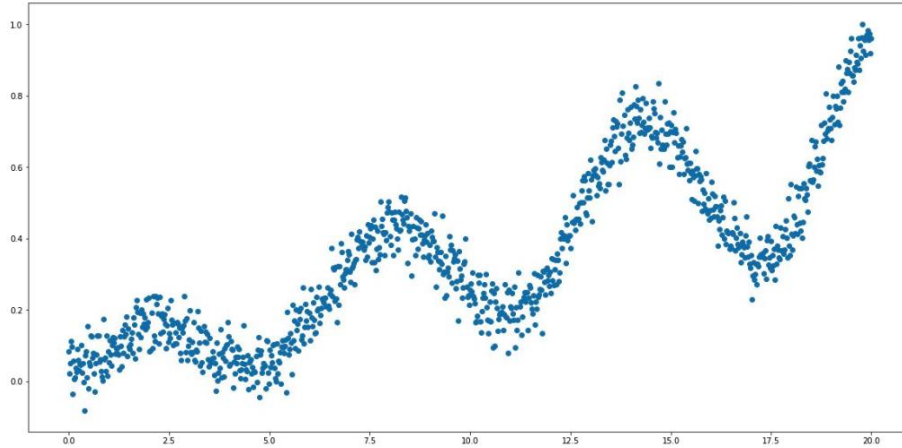


ETL: extract, transform, load

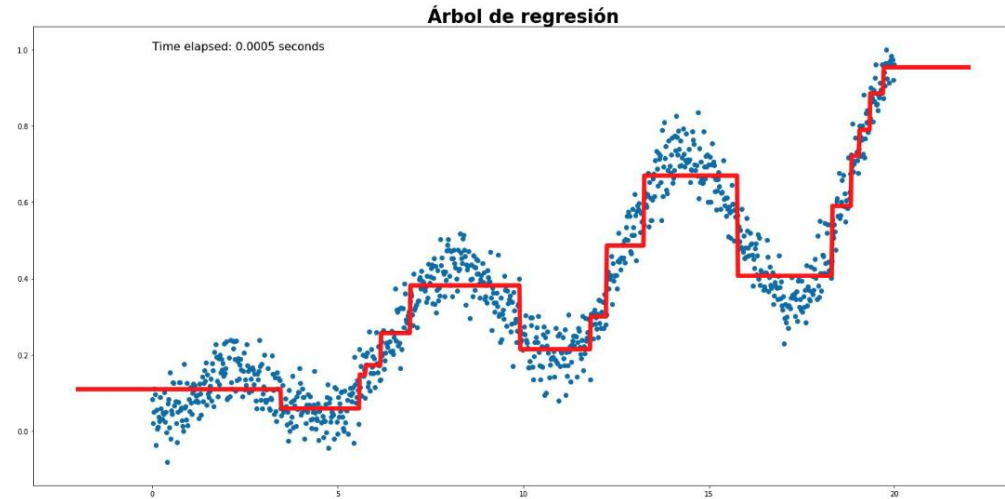
<https://es.r4ds.hadley.nz>

“Aprendizaje estadístico”: Un ejemplo sencillo de uso de ciencia de datos

PROBLEMA: Predecir valores siguientes.

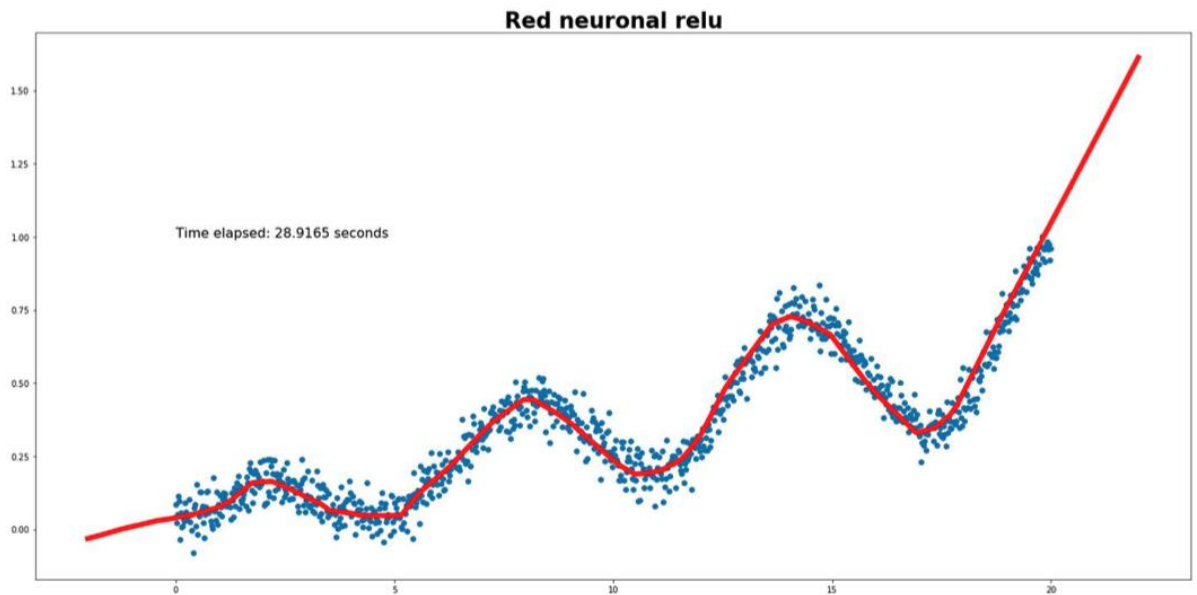
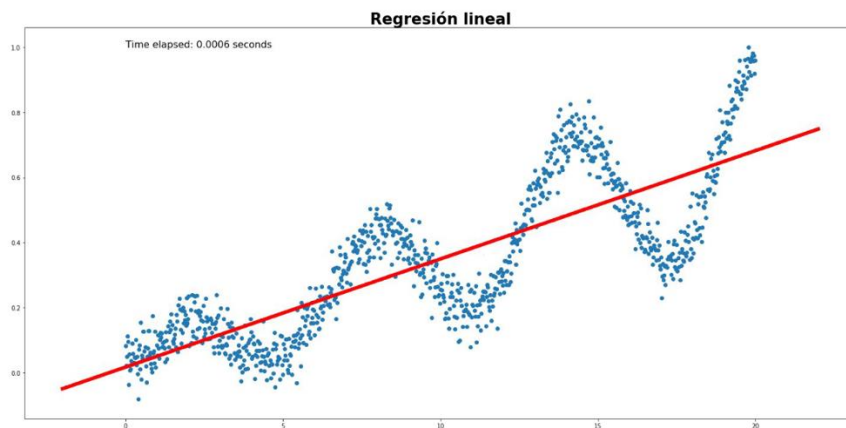


Un algoritmo de ML típico (más o menos “poderoso”)



Una red neuronal sencilla

Un estudiante de bachillerato (¿¿grado??)



Aplicaciones

10 minutos = Aplicaciones en la industria, la banca, salud y la ciencia en general.

¿Qué es? Analítica descriptiva, analítica explicativa/diagnóstica; analítica predictiva y analítica prescriptiva/proyectiva

GRACIAS

DR. ARIEL CARIAGA-MARTINEZ

CIENCIA DE DATOS

ACARIMAR@UAX.ES

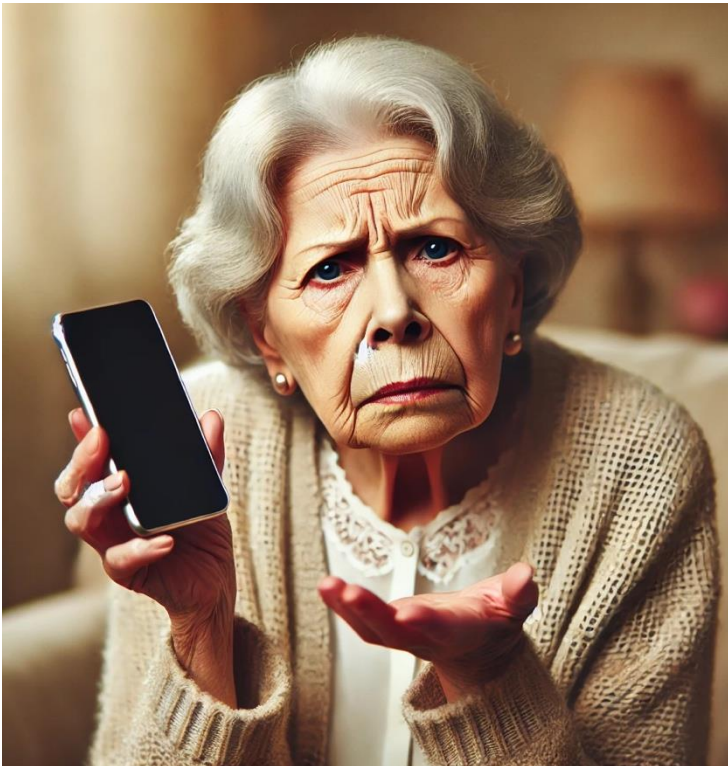
TRABAJO

GRADO EN BIOMEDICINA

DR. ARIEL CARIAGA-MARTÍNEZ

UN ANÁLISIS INTUITIVO

Un primer análisis



Vía Dall-E

Vamos a realizar un análisis intuitivo:

- Formar equipos (2-4 personas)
- Bájate el dataset del trabajo 1.
- Debatir, organizar y realizar un “análisis de datos”: ¿qué debería incluir?
- POR AHORA puedes usar R, Python, Excel, Numbers...
- Presentar los datos en forma de conclusiones: cuanto más específicas y visuales, mejor...

La restricción...deberíamos poder explicar cada fase del studio realizado (o propuesto) a la señora de la fotografía...