

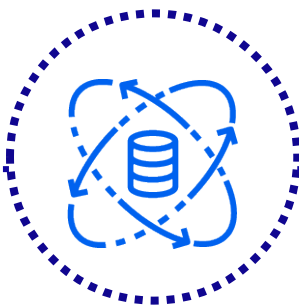
# Introducción a Inteligencia Artificial y Machine Learning



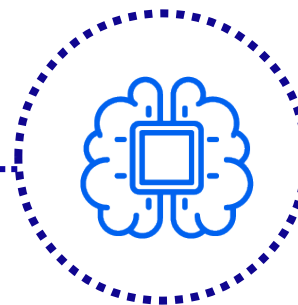
# Agenda de hoy (30min)



Introducción



Conceptos principales  
del Data Science



Usos del Machine  
Learning

# LA ERA DE LA INFORMACIÓN

# Un poco de ciencia ficción:

## Al servicio del avance científico – Julio Verne



El submarino

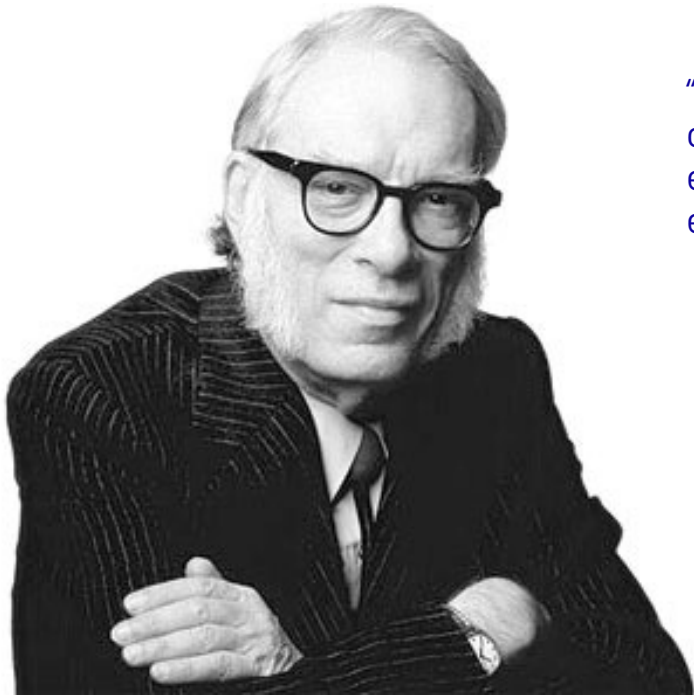


Las videoconferencias



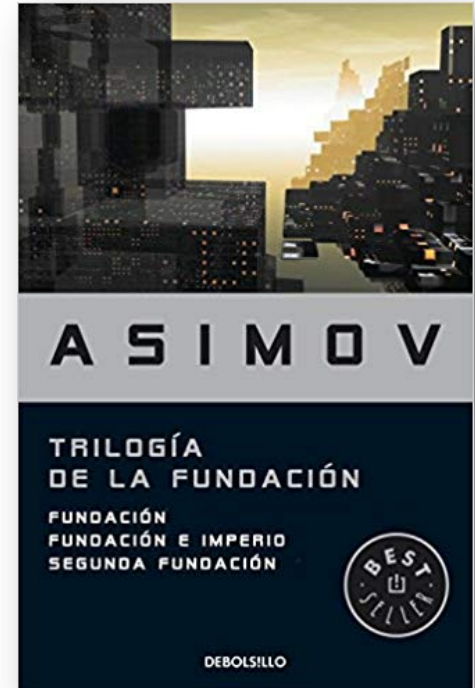
Las transmisiones de radio

# Isaac Asimov y la La Psicohistoria



"Es una ciencia que permite  
calcular el comportamiento  
estadístico de poblaciones  
extremadamente grandes"

- 1942

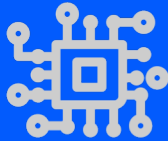


Y si miramos el  
**PRESENTE**

# La explosión de la información digital

Todo, todo deja datos...

2.5 quintillones  
de bytes por  
día



Los datos nos  
hablan, hay  
que saber  
**escucharlos**



# Existen datos...

1 1 1 1 5 9 6 5 2 5 5 2 5 3  
1 2 1 6 5 1 2 5 1 6 3  
7 1 0 4 1 2 3 9 5 1 3 7  
0 1 3 1 0 6 1 2 1 7 4 7  
5 4 2 1 2 3 3 1 7 8  
3 8 1 7 5 1 4  
1 5 3 8  
1 1 1  
1 3  
1

# Se ordenan...

3	2	1	1	1	1	5	1	5	1	2	1	2	5	6
0	1	4	3	1	3	1	3	6	5	5	2	5	1	1
5	0	8	2	1	1	2	1	4	0	8	5	3	3	6
7	1	1	1	1	2	7	1	2	2	4	7	8	6	4
3	3	6	5	1	8	6	3	9	7	5	8	7	9	7
1	1	5	1	1	3	4	8	3	5	0	5	2	4	5

# Los categorizamos

3	2	1	1	1	1	5	1	5	1	2	1	2	5	6
0	1	4	3	1	3	1	3	6	5	5	2	5	1	1
5	0	8	2	1	1	2	1	4	0	8	5	3	3	6
7	1	1	1	1	2	7	1	2	2	4	7	8	6	4
3	3	6	5	1	8	6	3	9	7	5	8	7	9	7
1	1	5	1	1	3	4	8	3	5	0	5	2	4	5

# Extraemos información

## 91 Números

- 59 Son impares
- 32 Son pares
- El número 1 es el más repetido
- La suma de pares es 120

2	2	2	6	2	0	4	6	8	4	
0	6	4	4	8	2	4	2	8	6	2
6	0	8	2	0	2	2	8	6	4	0

3	5	1	1	1	1	5	1	5	1	5	1	7	5	9
7	1	5	3	1	3	1	3	9	5	5	7	5	1	1
5	1	1	3	1	1	3	1	3	7	5	5	3	3	5
7	1	1	1	1	8	7	1	1	1	5	1	3	3	

# La información representa la realidad.

Y la ciencia (no solo la ficción) nos demuestra que entenderla nos permitirá **predecir el futuro!**

## Nace el concepto de **data science**

- Big Data
- Data Mining
- Data Scrapping
- Machine Learning
- Inteligencia Artificial
- Deep Learning
- Business Intelligence
- Data warehouse

¡A la orden para la toma de decisiones!

# Conceptos principales de DATA SCIENCE

# Machine Learning vs Inteligencia Artificial

Un concepto importante de diferenciar



## Inteligencia Artificial

Permite la toma de decisiones de forma autónoma en base a la información que se tiene.



*"Lo más probable es que llueva.  
Llevaré un paragua y tomaré un  
paracetamol."*



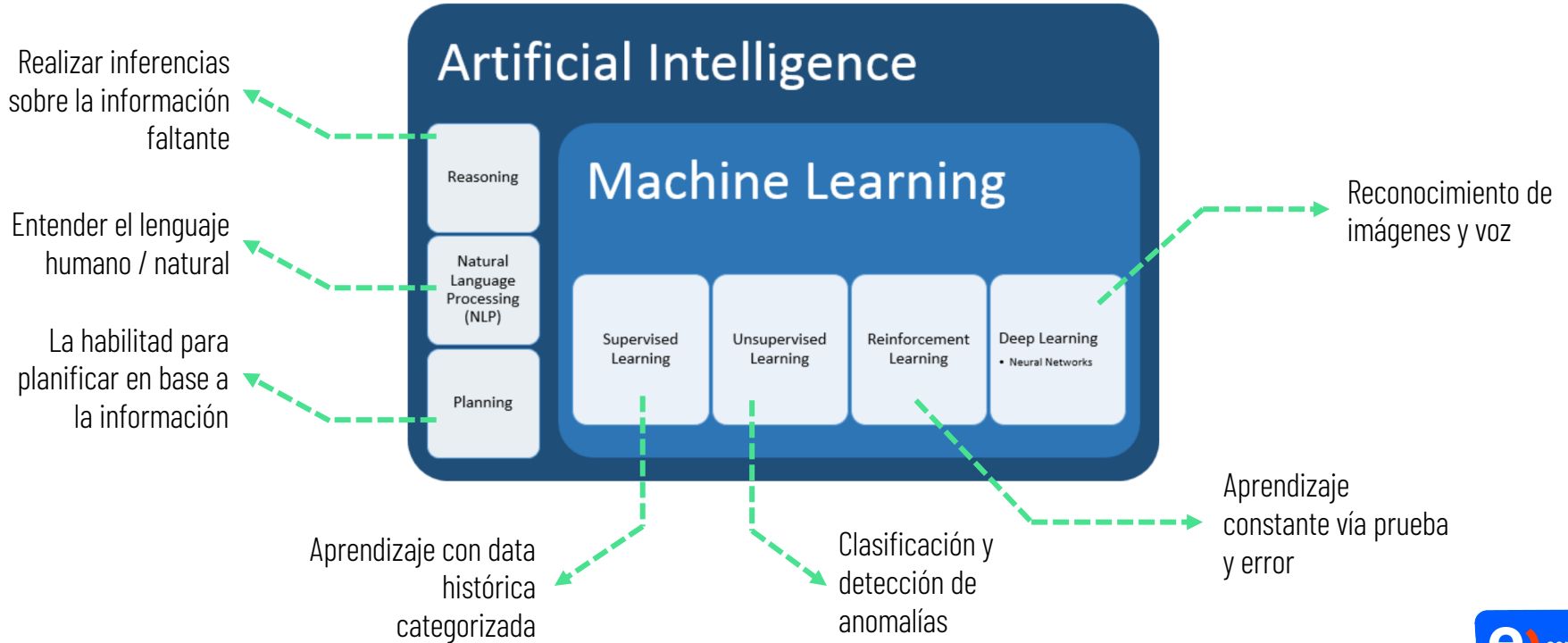
## Machine Learning

Permite aprender, mejorar y predecir de forma automática desde la información histórica recolectada.



*"Está nublado, hace frío, me duele  
una rodilla... Siempre que ocurre eso  
llueve... probablemente lloverá."*

# Tipos de IA/ML





# Tipos de Aprendizaje

## Supervisado

Es la forma de aprendizaje basada en tener pleno conocimiento de las variables y su efecto, sin embargo no siempre su relación.

**Objetivo:** dado un input de datos, predecir cual será el output.

## No supervisado

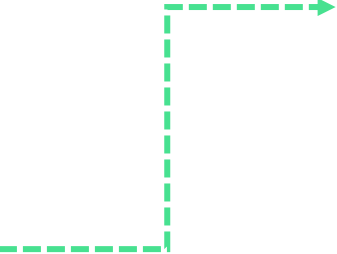
Es la forma de aprendizaje basada en tener un conjunto de datos sin aparente relación entre sí, y busca inferir relación entre ellos.

**Objetivo:** Buscar relaciones entre entidades en un grupo de datos de forma implícita.

*Generar un modelo que permita clasificar o predecir información.*

# Aprendizaje Supervisado (1/2)

Medición	Estado del Cielo	¿Llovió?
28-05-2012	Nublado	Si
29-05-2012	Nublado	Si
30-05-2012	Nublado	Si
31-05-2012	Nublado	Si
01-06-2012	Nublado	Si
02-06-2012	Chubascos	No
03-06-2012	Chubascos	No
04-06-2012	Soleado	No
05-06-2012	Soleado	No
06-06-2012	Chubascos	No
07-06-2012	Chubascos	No
08-06-2012	Nublado	Si
09-06-2012	Nublado	Si
10-06-2012	Chubascos	No
11-06-2012	?????	???

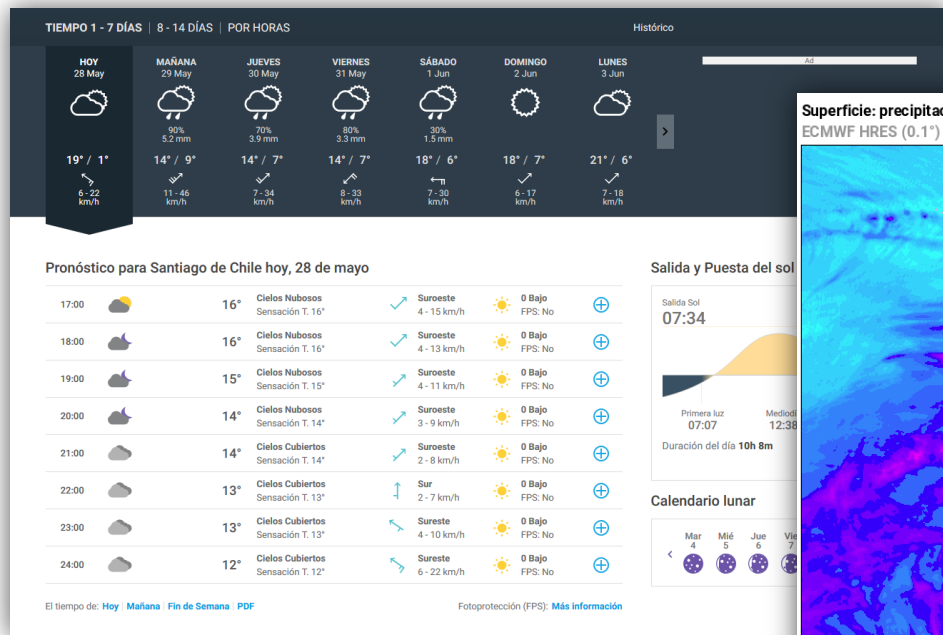


¿Qué podemos concluir que pasará el 11 de Junio?

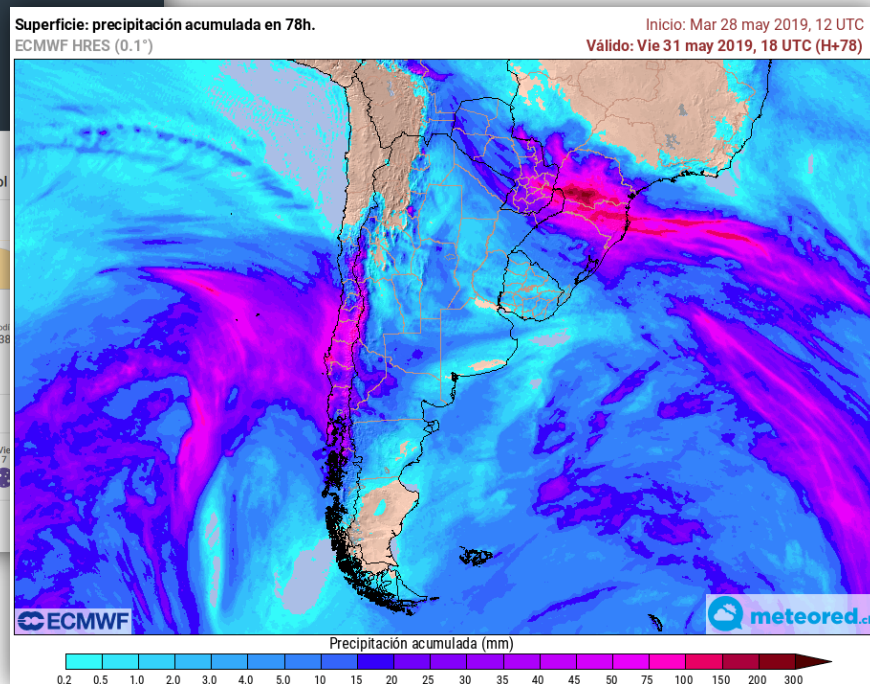
## Aprendizaje Supervisado (2/3)

Medición	Estado del Cielo	Temperatura Mínima	Temperatura Máxima	Presión Atmosférica	Lugar del País	Velocidad Viento (km/h)	Precipitación Acumulada	...	¿Llueve?
28-05-2012	Nublado	12°	20°	--	Santiago	3,5	---	...	Si
29-05-2012	Nublado	17°	19°	--	Santiago	9,4	---	...	Si
30-05-2012	Nublado	8°	20°	--	Santiago	5,4	---	...	Si
31-05-2012	Nublado	13°	20°	--	Santiago	5,4	---	...	Si
01-06-2012	Nublado	8°	20°	--	Santiago	9,4	---	...	Si
02-06-2012	Chubascos	9°	21°	--	Santiago	9,4	---	...	No
03-06-2012	Chubascos	10°	14°	--	Santiago	9,4	---	...	No
04-06-2012	Soleado	2°	14°	--	Santiago	5,4	---	...	No
05-06-2012	Soleado	1°	16°	--	Santiago	9,4	---	...	No
06-06-2012	Chubascos	-1°	18°	--	Santiago	5,4	---	...	No
07-06-2012	Chubascos	3°	23°	--	Santiago	5,4	---	...	No
08-06-2012	Nublado	6°	23°	--	Santiago	7,6	---	...	Si
09-06-2012	Nublado	4°	8°	--	Santiago	7,6	---	...	Si
10-06-2012	Chubascos	2°	17°	--	Santiago	3,5	---	...	No

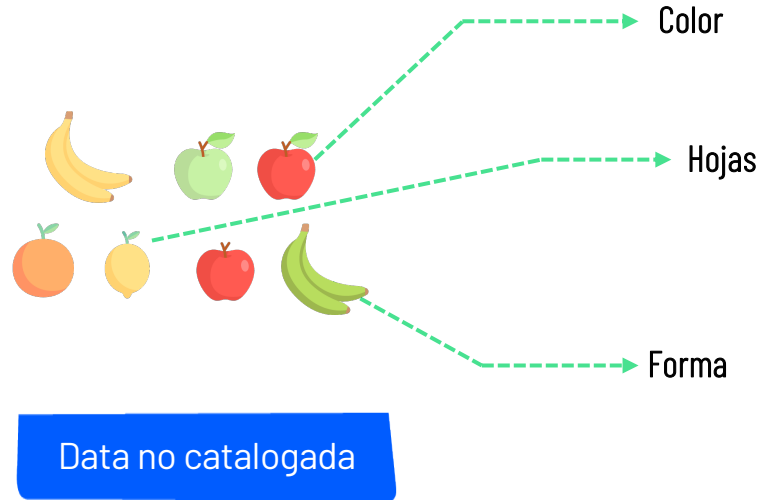
# Creación de Modelos



<https://www.meteored.cl/>



# Aprendizaje no supervisado



# Aprendizaje no supervisado



< Nuevo elemento



Grupo 1  
**Manzanas**



Grupo 2  
**Naranjas**



Grupo 3  
**Platanos**

# Aprendizaje no supervisado



Grupo 1  
**Manzanas**



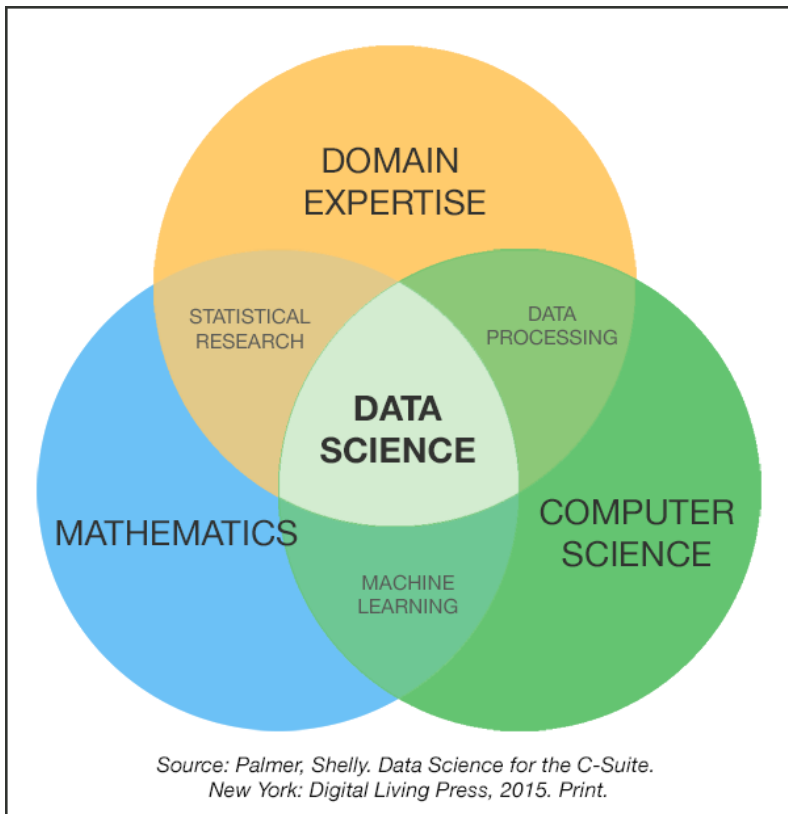
Grupo 2  
**Naranjas**

< Es clasificado



Grupo 3  
**Platanos**

# Perfil de un Data Analyst



- Expertiz en un dominio en particular.
- Conocimiento avanzado en estadísticas, análisis de información y matemáticas.
- Conocimientos en computación e informática. (hacking skills)



# Usos de **Machine Learning**

## Utilizaciones cotidianas de ML



- Recomendaciones en Spotify, YouTube, Netflix.
- Predicción de la rotación de Clientes en una compañía.
- Asistentes Virtuales(Siri, Alexa, Google Now).
- Reconocimiento de imágenes.
- Detección de Fraude.
- Detección de enfermedades en pacientes.
- Mejoras en sistemas de navegación.
- Campañas publicitarias(\*).

# Utilizaciones en Ciberseguridad



- Visibilidad de anomalías en la red.
- Detección de usuarios sospechosos en la red.
- Comportamiento anormal dentro de las aplicaciones.
- Comportamiento anómalo dentro de equipos (durante un incidente).
- Clasificación de llamadas Web anormales.
- Detección temprana de campañas de Phishing.
- Detección de malware dentro de la red.
- Es la base de las hipótesis de Threat Hunting.

# Cierre

# Consideraciones antes de hacer ML

## No es una bala de plata

No es la solución a nuestros problemas, es una herramienta que nos permite ser más eficientes y concentrar esfuerzos en lo esencial.

## Tener analistas capaces de trabajar con la información

Machine Learning es una herramienta de trabajo para los analistas. No un reemplazo.

## Tener procesos maduros para responder a lo detectado

No solo basta con detectar las anomalías, es necesario saber explicarlas y darles solución en caso de que sean necesarios.

## Declarar el problema a resolver

Es crucial definir preguntas a ser respondidas por nuestros algoritmos, si no sabemos que buscar no encontraremos nada.

## Conocer la información disponible

Saber que tan estructurada y que tipo de información existe para el problema a resolver

## Consultar a un experto para levantar y preparar información

Si bien existen soluciones que aplican algoritmos de machine learning, se necesita un data scientist que apoye a estos procesos. No son automáticos

## Palabras finales...



Las tecnologías analíticas y predictivas **no son algo que reemplazarán a nuestros profesionales** ni protegerán automáticamente a las empresas.



Son un **potenciador a las capacidades** de nuestros analistas!



Necesitan una estrategia y estar adaptadas al negocio, por lo tanto, es **necesario entender como utilizarlas** y cuándo.

