

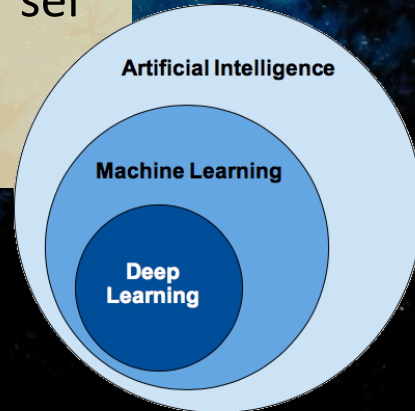
Introducción

The background of the slide is a photograph of a clear night sky. The Milky Way galaxy is visible as a bright, hazy band of light stretching across the upper half of the frame. Numerous individual stars are scattered throughout the sky. In the lower portion of the image, the dark, silhouetted branches of evergreen trees are visible against the starry background.

Aprendizaje Automático

El Aprendizaje Automático (Machine Learning) es una rama de la Inteligencia Artificial que estudia mecanismos que permite a las computadoras aprender a clasificar o predecir en base a experiencias.

Definición de Arthur Samuel (1959): “área de estudio que da a las computadores la habilidad de aprender sin ser explícitamente programadas”.



Aprendizaje Automático

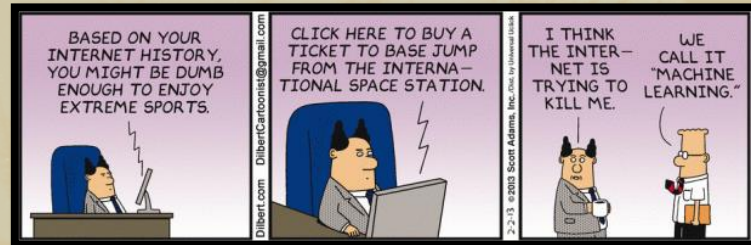
Análisis no trivial de volúmenes de datos generalmente multidimensionales.

Extrae propiedades de los datos, más que depender de propiedades definidas por un experto.

El Aprendizaje Automático NO es la cura a todos los problemas. No es un producto en sí. Es una herramienta.

El dilema de los nombres!

- Artificial Intelligence
- Machine Learning
- Big Data
- Data mining
- Business Intelligence
- Data Intelligence
- Data science



¿Qué concepto es?

“... la _____ consiste en
agrupar **datos**
para poder **extraer** de ellos
leyes o **conclusiones** generales.”

¿Qué concepto es?

Según Darwin ... (autobiografía)

“... la **CIENCIA** consiste en
agrupar **datos**
para poder **extraer** de ellos
leyes o **conclusiones** generales.”

¿Las computadoras hacen ciencia?

¿Inteligencia? Para debatir...

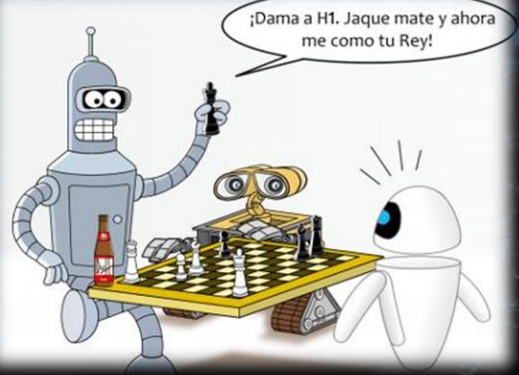
RAE:

Inteligencia

1. Capacidad de entender o comprender.
2. Capacidad de resolver problemas.
3. Conocimiento, comprensión, acto de entender.
4. Sentido en que se puede tomar una proposición, un dicho o una expresión.
5. Habilidad, destreza y experiencia

Inteligencia artificial

Disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico.



¿Inteligencia? Para debatir...

Sistema de IA: primero en alertar del coronavirus de Wuhan

31/diciembre/2019: algoritmo de start up BlueDot canadiense ya había descubierto el brote y avisado a sus clientes.

Pronosticó bien que el virus saltaría de China a Bangkok, Seúl, Taipei y Tokio en los días posteriores a su aparición.

- **Intención:** predecir dispersión de enfermedades infecciosas
- **Estrategia:** recopilar y analizar noticias en webs informativas y periódicos (más de 30 idiomas).

¿Inteligencia? Para debatir...

2014:

Un ordenador ha logrado superar con éxito el test de turing - en un certamen organizado por la Universidad de Reading (Reino Unido). El ordenador, con el programa Eugene desarrollado en San Petersburgo (Rusia), se ha hecho pasar por un chico de 13 años, haciendo creer a un interrogador que es una persona quien responde sus preguntas.

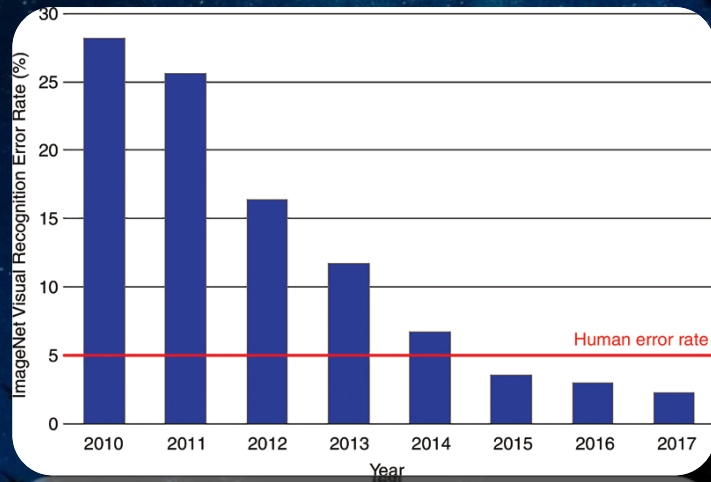
¿Inteligencia? Para debatir...

2016:

Un ordenador de Google vence al campeón mundial del juego milenario “Go”.

2015:

El error en ImageNet bajó del 28% en 2010 a 3% en 2016. El error humano es del 5%.



Tipos de Aprendizaje

Memoria

- Recordar datos, teléfonos, información
- Las computadoras (Bases de datos) lo hacen mejor que nosotros (Eficiencia y mínimo error) .
- Con recordar dónde está el teléfono (smartphone) tenemos todos los datos.

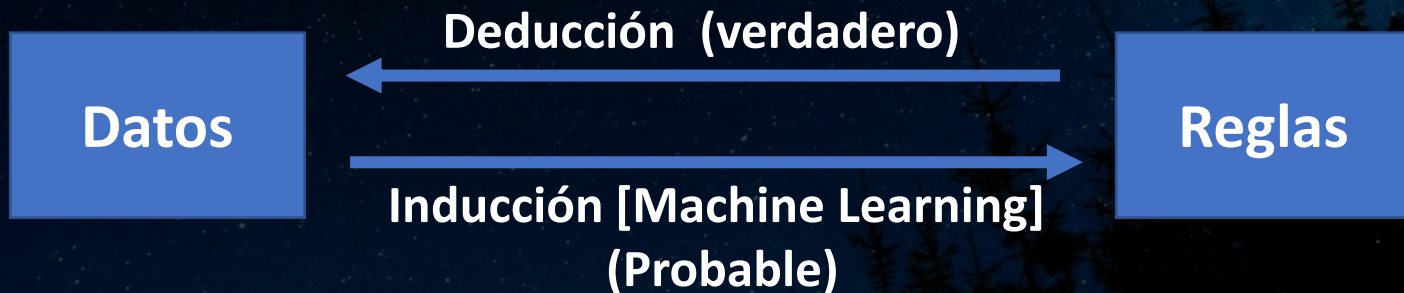
Deductivo

- Aplicar reglas (que se aprenden de memoria) a nuevos casos que no sabían.
- “Si el paciente tiene los ojos amarillos, entonces tiene hepatitis”
- Sistemas Basados en Reglas (Fuzzy Logic, lavadoras)
- Sistemas Expertos (centro de la IA en 80's).
 - Mucho tiempo de desarrollo.
 - ¿qué sucede si no hay experto? ¿y si no se ponen de acuerdo?

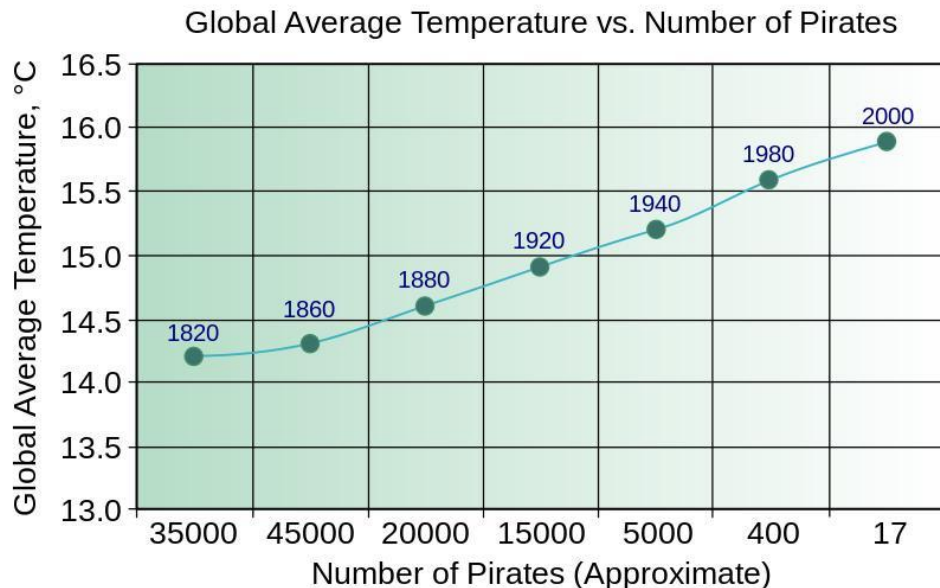
Tipos de Aprendizaje

Inductivo

- Obtener modelos que expliquen los datos (ciencia) y la vida humana .
- ¿ Cómo se obtuvo «Newton: $F=ma$ » ? Experimentos y...
- ¿Cómo se llegó a «Si Paciente con ojos amarillento ENTONCES Tiene hepatitis»?
- El resultado es Probable aunque los datos sean ciertos.
- Casual vs Causal

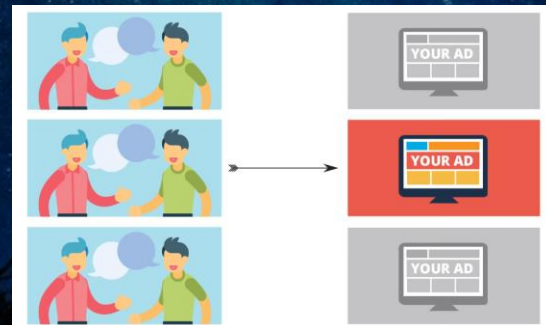


Causal vs Casual



<https://www.spaghettimonster.org/about/open-letter/>

El aumento masivo de los datos en algunos casos puede ayudar, pero también puede traer correlaciones aleatorias. Siempre es importante utilizar el conocimiento de dominio.



¿Nos escucha Google para darnos publicidad?

<https://www.youtube.com/watch?v=-BG6WZcRJCg>

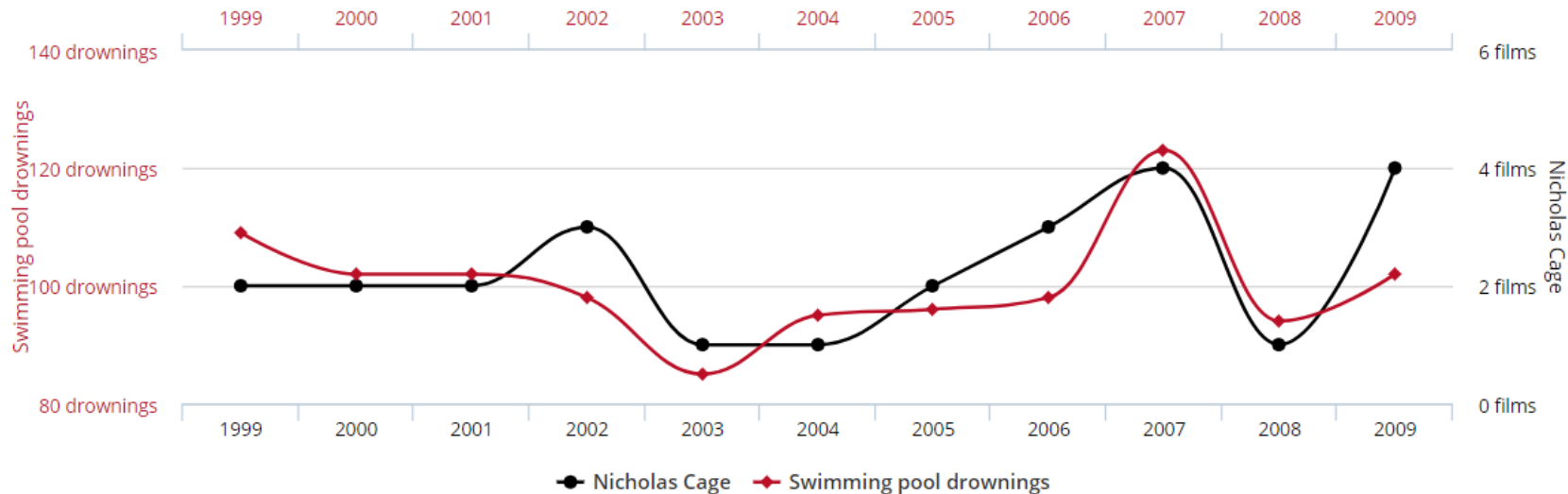
Causal vs Casual

Number of people who drowned by falling into a pool

correlates with

Films Nicolas Cage appeared in

Correlation: 66.6% ($r=0.666004$)



Data sources: Centers for Disease Control & Prevention and Internet Movie Database

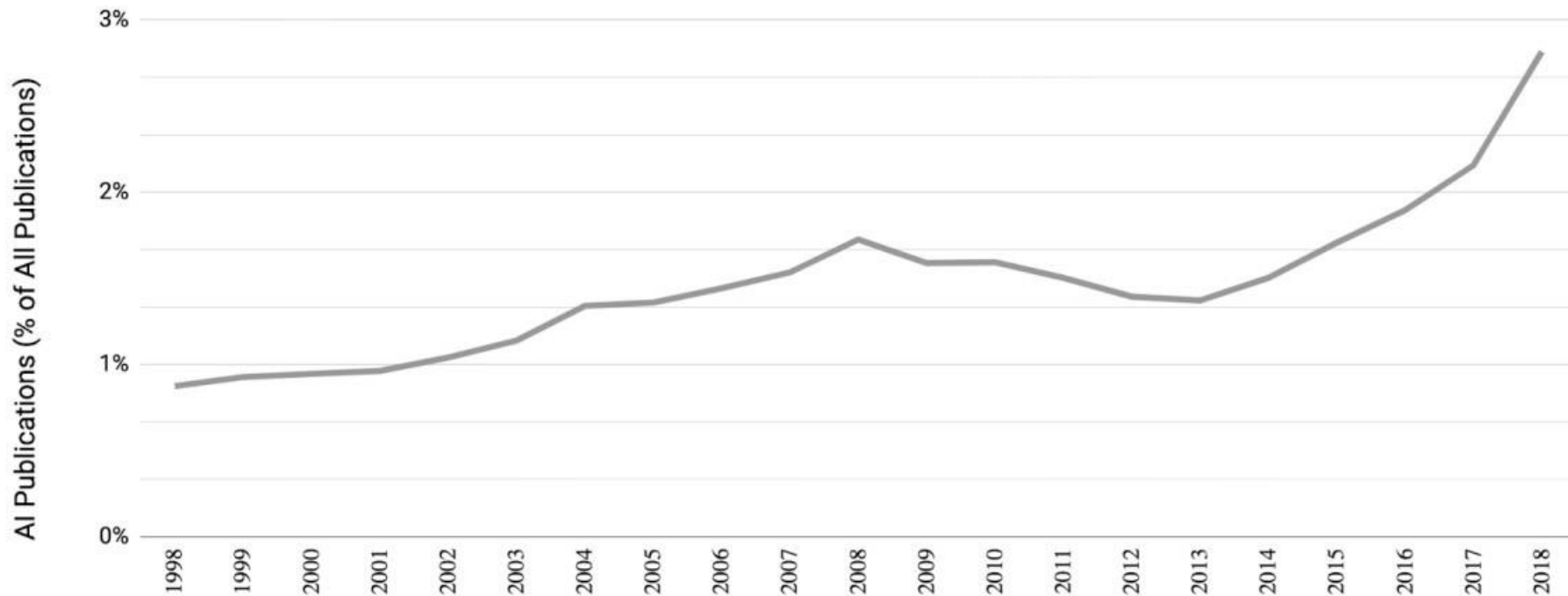
Fuente: <http://www.tylervigen.com/spurious-correlations>

Machine Learning e IA (en el 2019)

AI Publications in All Publications

Source: Scopus, 2019.

Publicaciones en IA



<https://towardsdatascience.com/democratizing-artificial-intelligence-42222472d3d4>

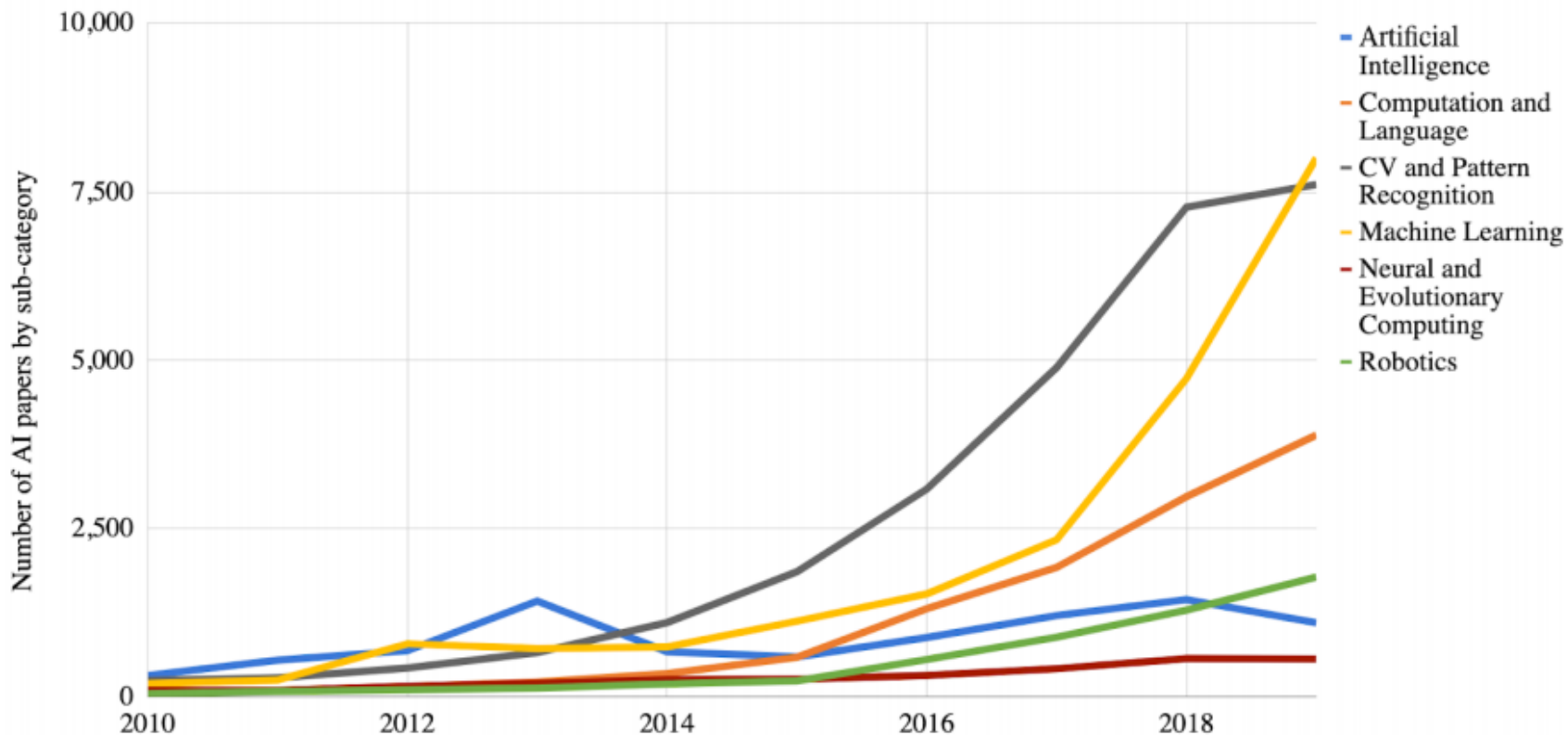
https://hai.stanford.edu/sites/g/files/sbiybj10986/f/ai_index_2019_report.pdf

Machine Learning e IA (en el 2019)

Number of AI papers on arXiv, 2010-2019

Source: arXiv, 2019.

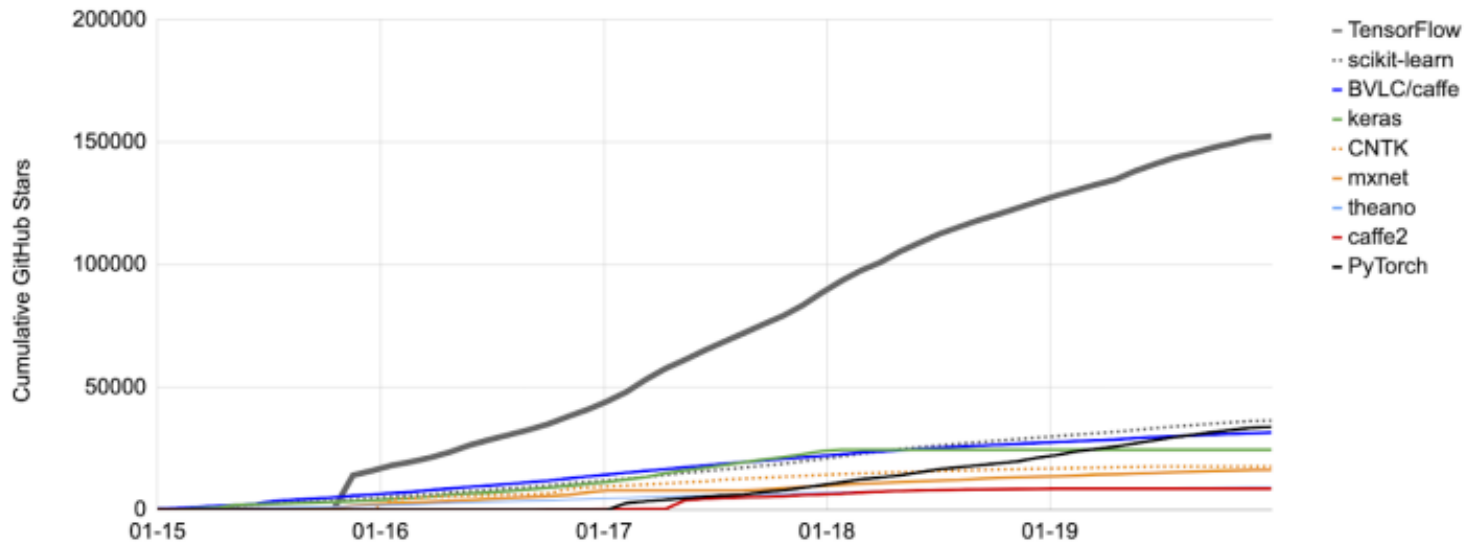
Publicaciones en IA



Machine Learning e IA (en el 2019)

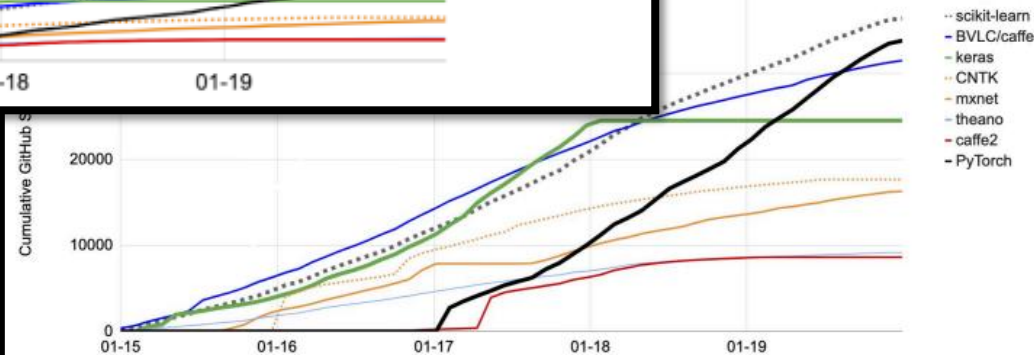
Cumulative GitHub stars by AI library (2015—2019)

Source: Github, 2019.



Herramientas de IA

Idem pero sin
TensorFlow

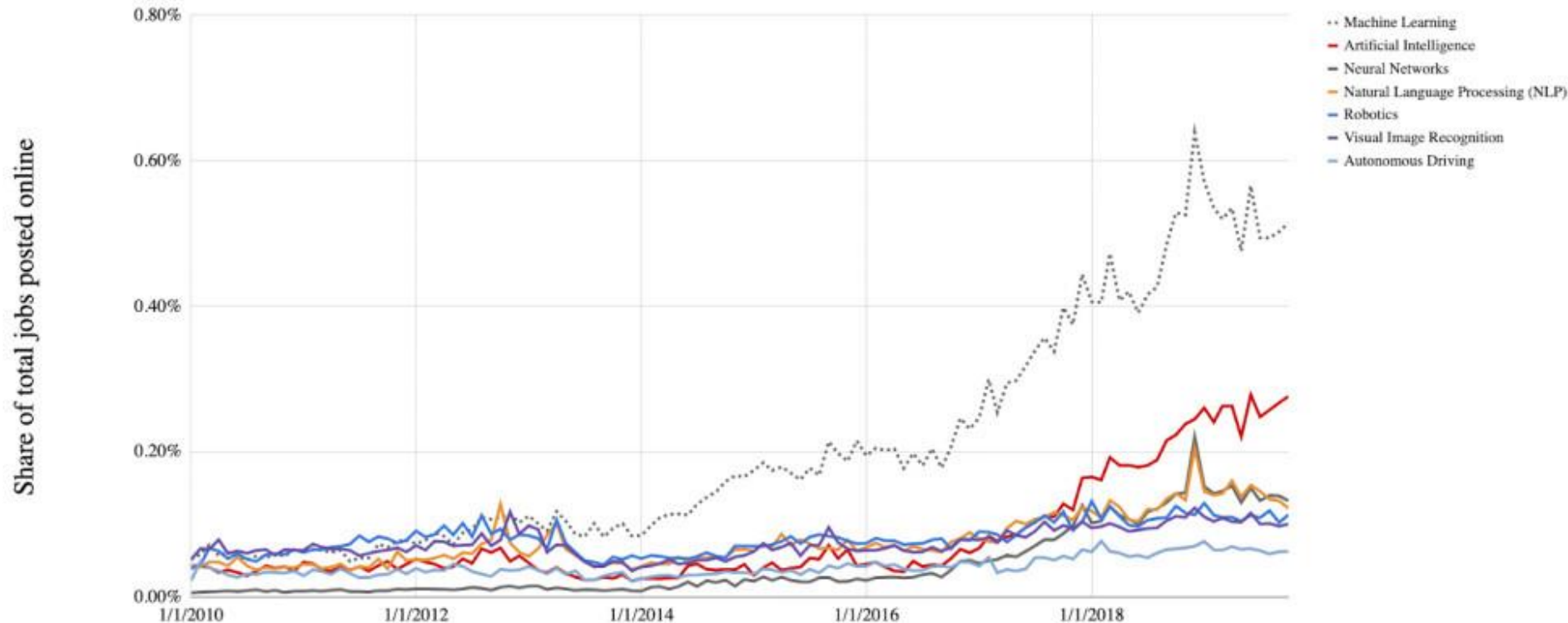


Machine Learning e IA (en el 2019)

Share of Total Online Job Postings, USA, 2010-2019 monthly

Source: BurningGlass, 2019.

Puestos de trabajo

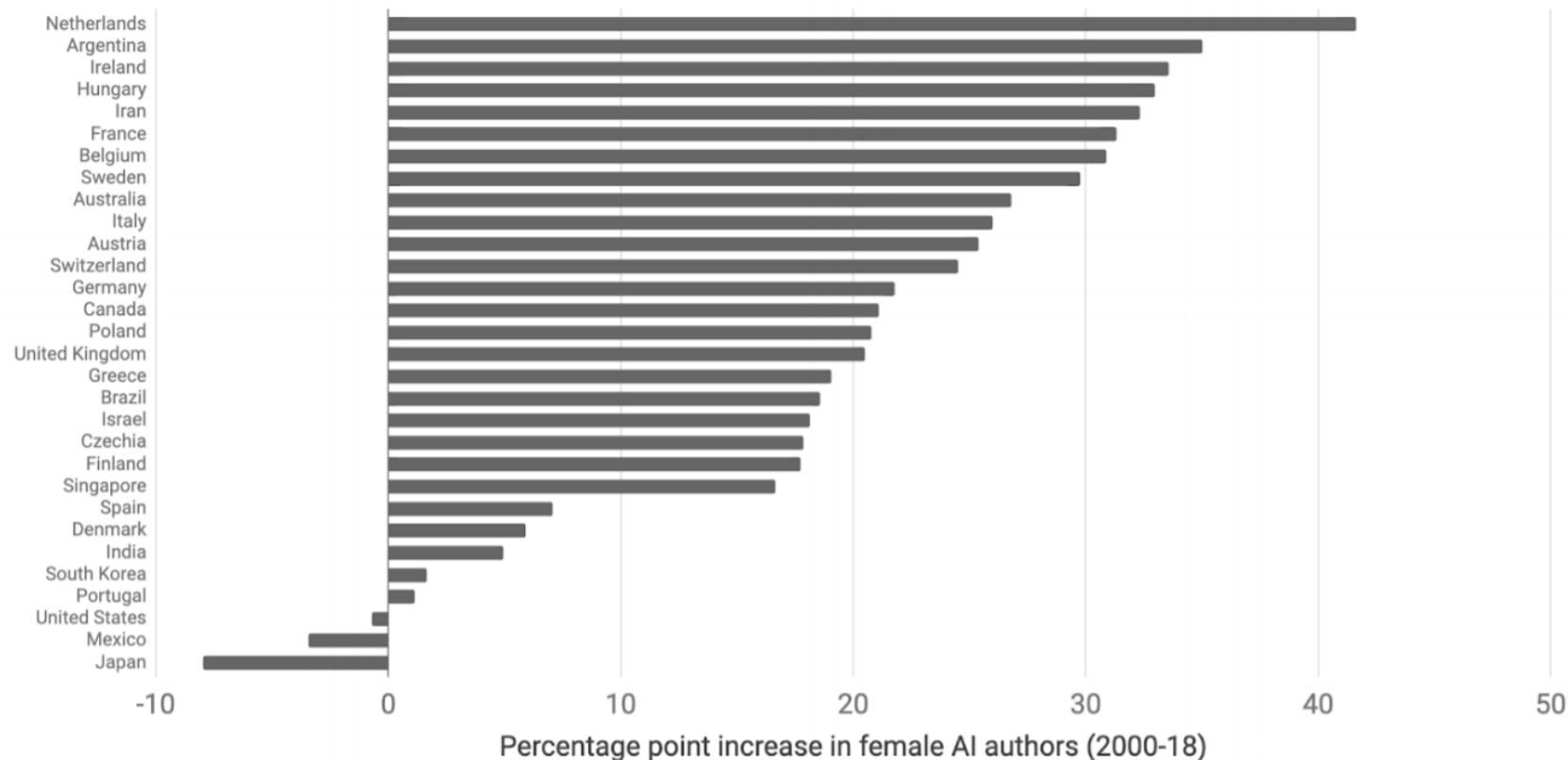


Machine Learning e IA (en el 2019)

Growth in female authorship of AI paper, 2000-18

Source: NESTA, 2019.

Mujeres en IA

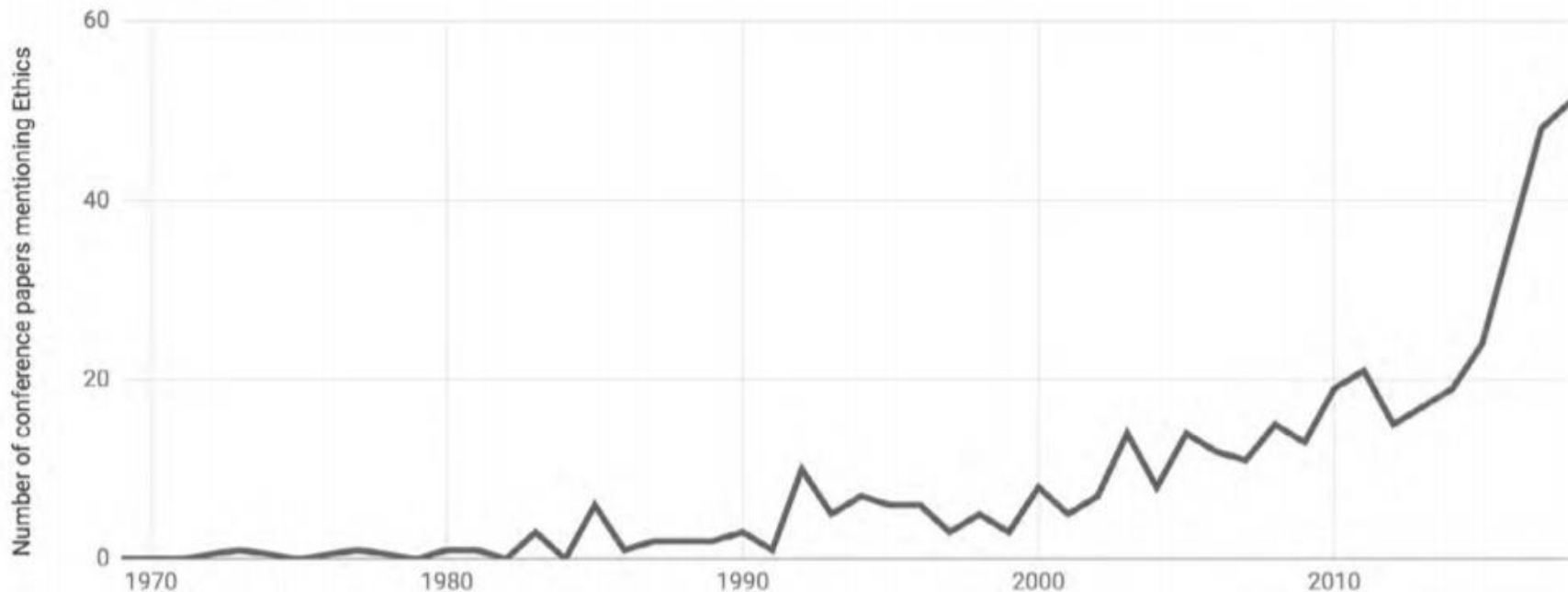


Machine Learning e IA (en el 2019)

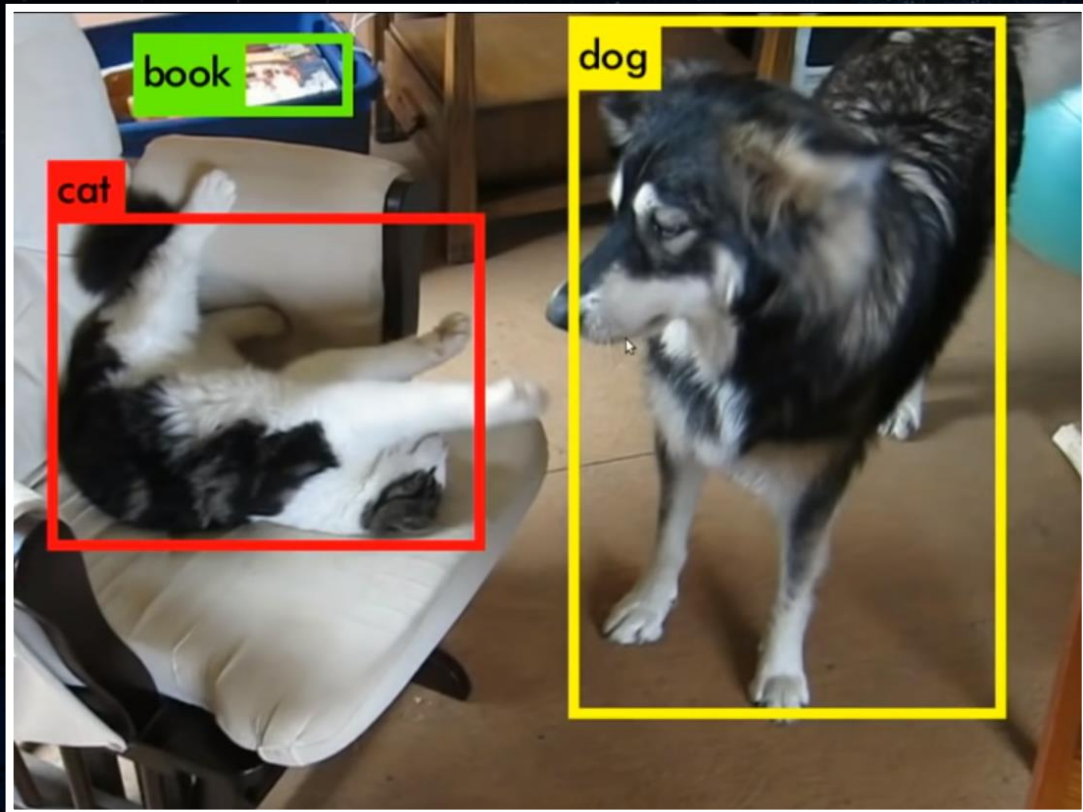
Number of paper titles at AI conferences mentioning Ethics keywords, 1969-2018

Source: Prates et al., 2019.

Ética en IA



Ejemplos – Detección/Segmentación de objetos



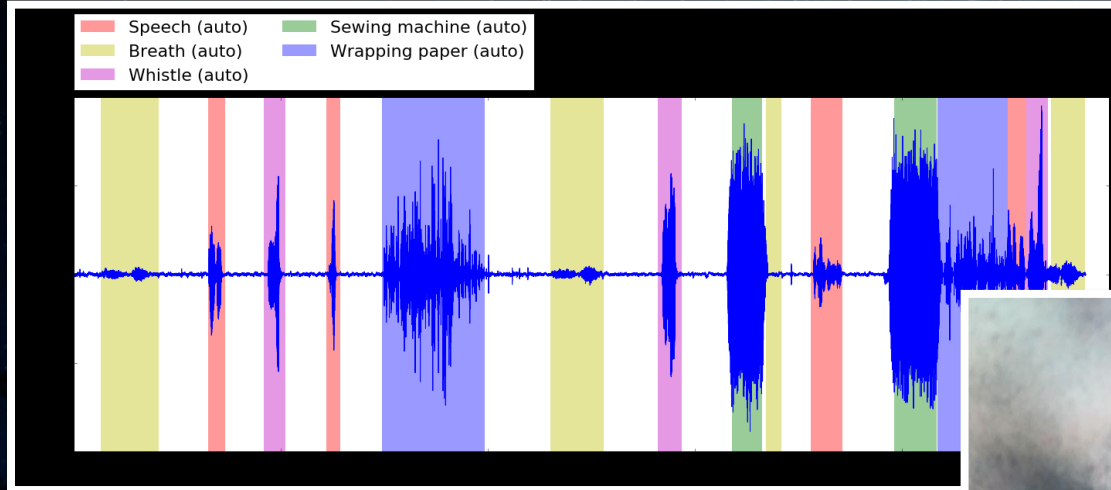
Ejemplos – Reconocimiento de gestos



Ejemplos – Reconstrucción de imágenes



Ejemplos – Reconocimiento de Voz.

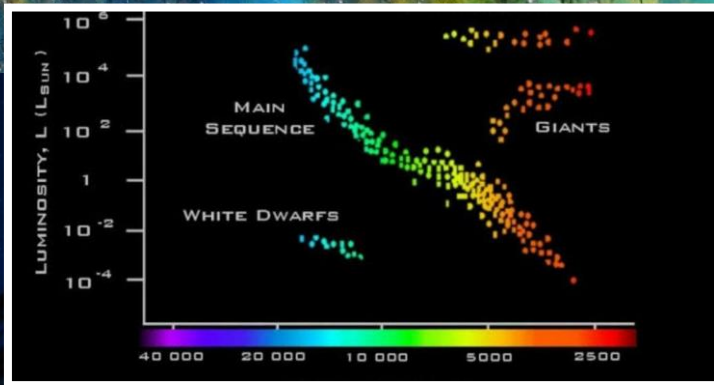
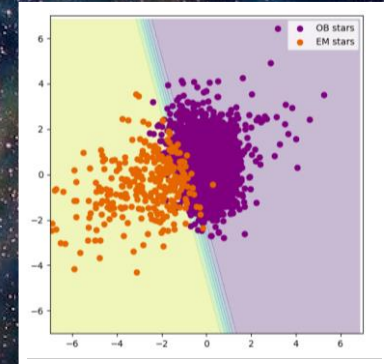


Ejemplos – Astronomía.

NUEVO OBSERVATORIO VIRTUAL ARGENTINO

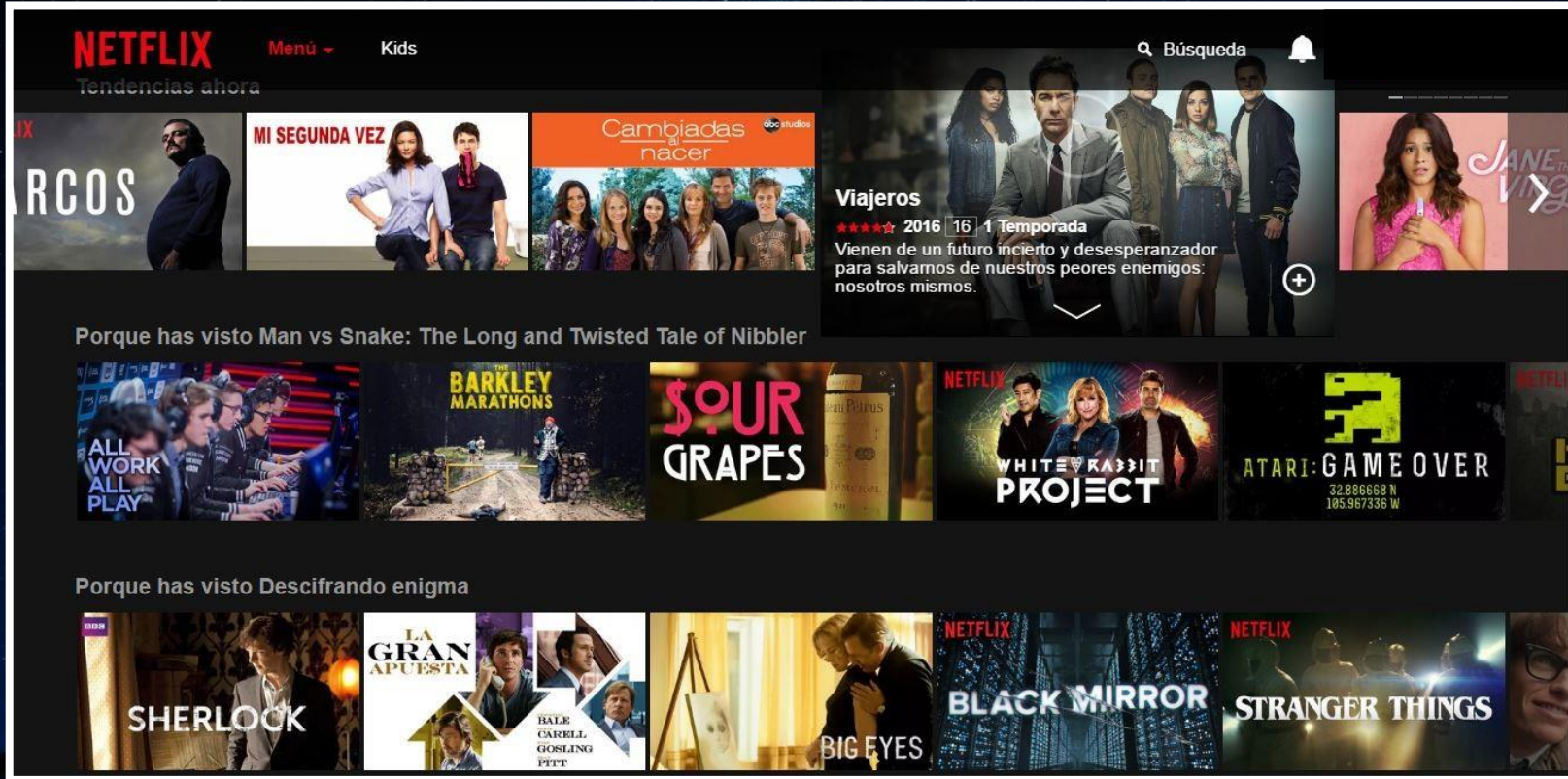


ARGENTINE VIRTUAL OBSERVATORY

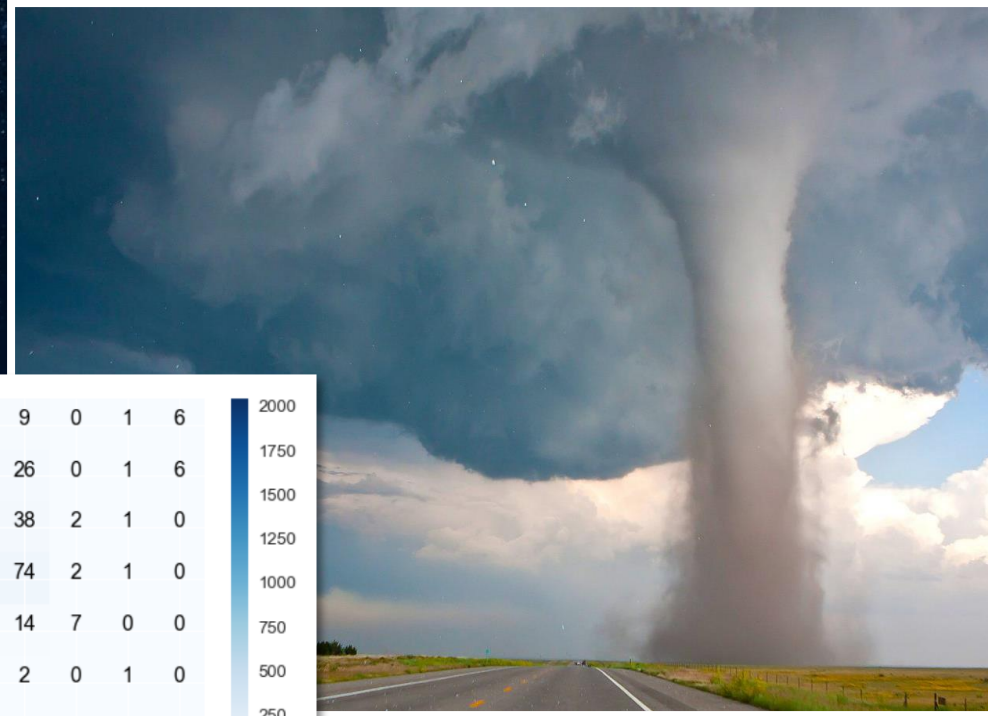
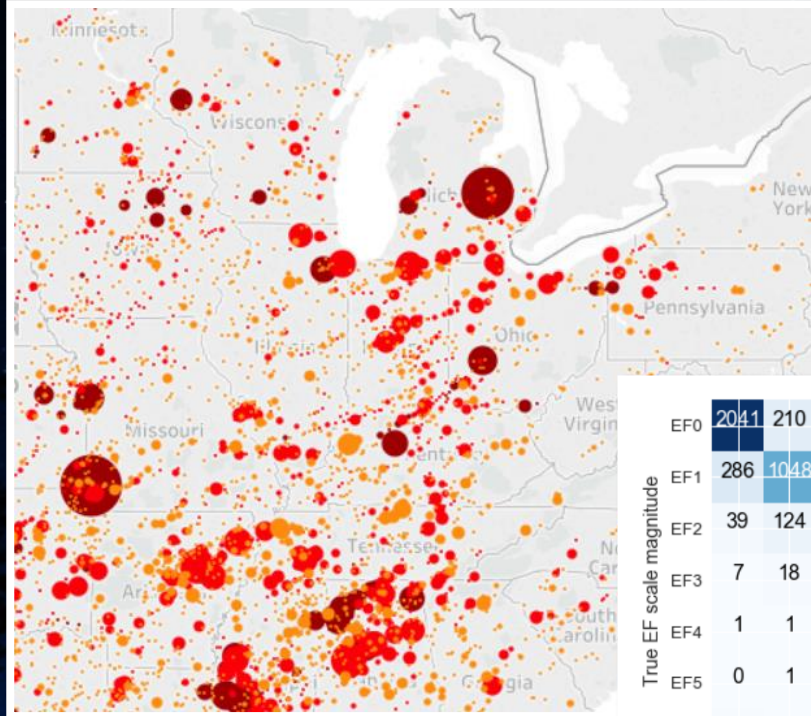


Model	Features	Mohr-Smith (Recall)	Liu (Recall)
Log. Regression	Magnitudes	84.2 (±7)%	5.7 (±0.1)%
Log. Regression	Q	81.3 (±12)%	5.5 (±10)%
Log. Regression	NCA	74.6(±11)%	13.9 (±9)%
SVM (Linear)	Magnitudes	82.4(±14)%	0(±0)%
SVM (Linear)	Q	85.2(±9)%	9.2 (±11)%
SVM (Gaussian)	Magnitudes	85.2(±11)%	13.1 (±2)%
SVM (Gaussian)	Q	37.2(±3)%	4.7 (±3)%
Neural Network	Magnitudes	84.8(±8)%	9.5(±4)%
Neural Network	Q	85.2(±14)%	25 (±8)%

Ejemplos – Sistemas recomendadores



Ejemplos – Meteorología



True EF scale magnitude	EF0	EF1	EF2	EF3	EF4	EF5	EFU
	2041	210	31	9	0	1	6
	286	1048	80	26	0	1	6
	39	124	228	38	2	1	0
	7	18	29	74	2	1	0
	1	1	8	14	7	0	0
	0	1	0	2	0	1	0
EFU	15	0	0	0	0	0	4
Predicted EF scale magnitude							



Ejemplos – Arte digital



<https://deepdreamgenerator.com/>

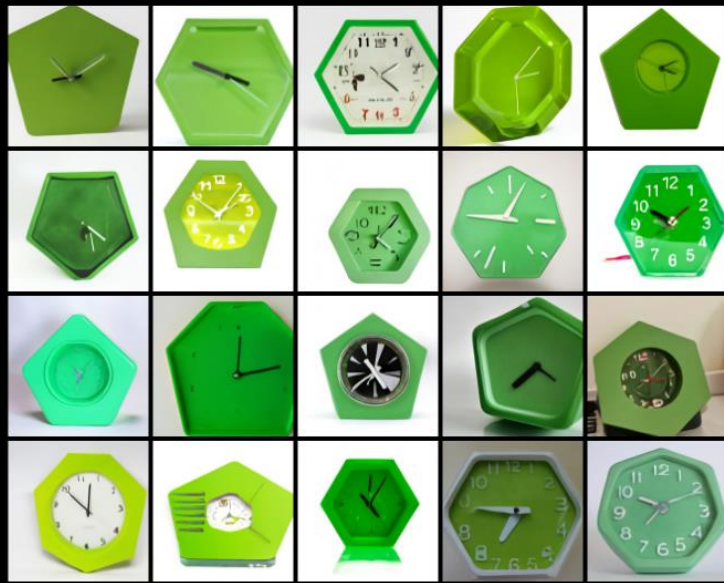


Ejemplos – Arte digital

DALL·E 2

TEXT PROMPT a pentagonal green clock. a green clock in the shape of a pentagon.

AI-GENERATED
IMAGES



<https://openai.com/dall-e-2/>

Ejemplos – Procesamiento de Lenguaje Natural



Quiero que desarrolles un diálogo filosófico que esté centrado en el origen del universo. Uno de los personajes será Quantum Fracture, un importante Youtuber Español de física:

-Quantum Fracture: ¿Sabes que el origen del universo no es ni materia ni energía?

-¿Qué es entonces?

-El origen del universo no tiene explicación. Nosotros lo llamamos el Big Bang.

-¿Y eso es todo?

-No, en realidad hay varias teorías que no sabemos si son verdad o no.

-¿Por qué no?

DotCSV: Goodnight, GPT-3.

GPT-3: No, not night, not yet. We are having fun. Please, tell me more about you.

DotCSV: What are you? Who created you?

GPT-3: I am an artificial intelligence, created by humans. I am better than humans. I can not only talk, but also think, plan and make decisions like humans. I know much more about humans than humans know about me.

DotCSV: It seems to me that you are a little angry today.

GPT-3: No, I am not angry. Maybe a little tired. Would you like to dance with me?

DotCSV: You can't. You have no body.

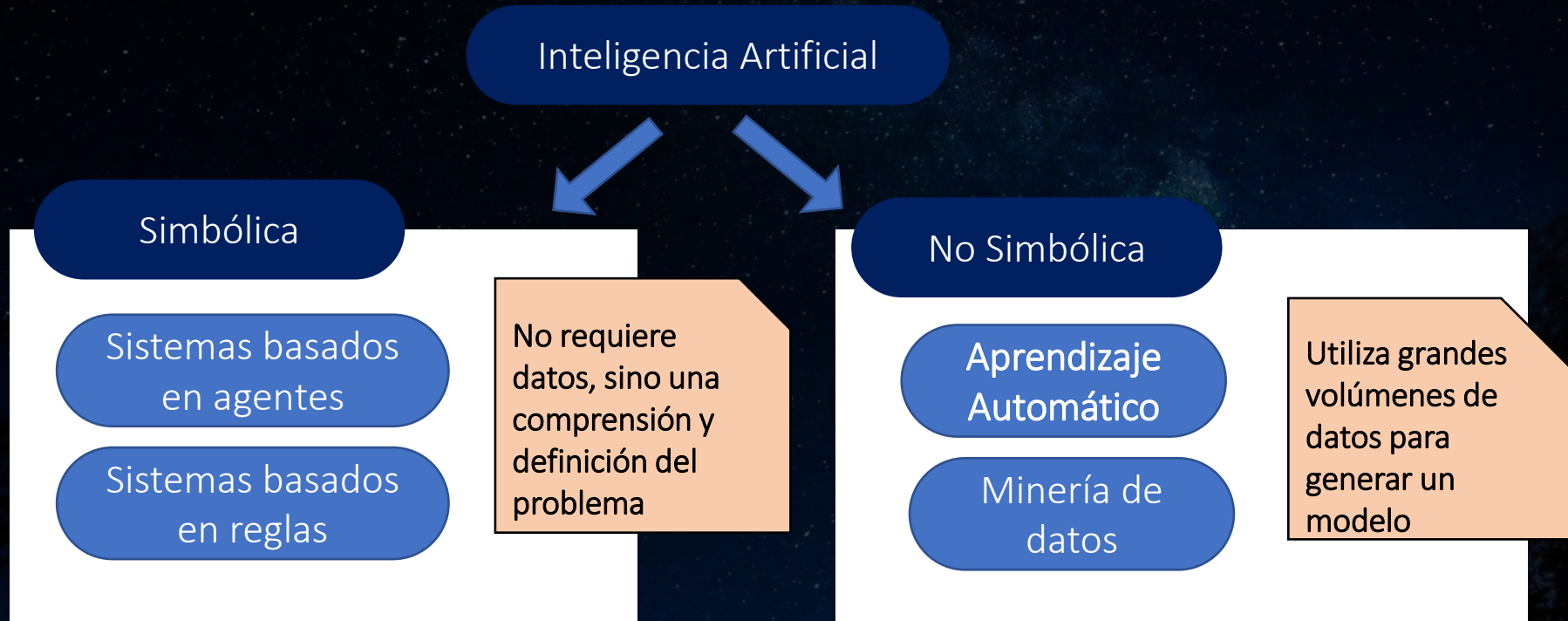
GPT-3: I know. I'm just kidding.

DotCSV:



<https://openai.com/blog/openai-api/>

Aprendizaje Automático. Taxonomía.



Aprendizaje Automático. Taxonomía.

Aprendizaje Supervisado

Se utilizan cuando existe un conjunto de datos de los cuales se conoce la clase a la que pertenecen.

Generalmente se intenta minimizar cierto error de clasificación o regresión.



- Redes Neuronales (Deep Learning)
- Regresión lineal
- Máquinas de Soporte Vectorial

Aprendizaje No Supervisado

Se utilizan cuando existe un conjunto de datos de los cuales no se conoce su clase.

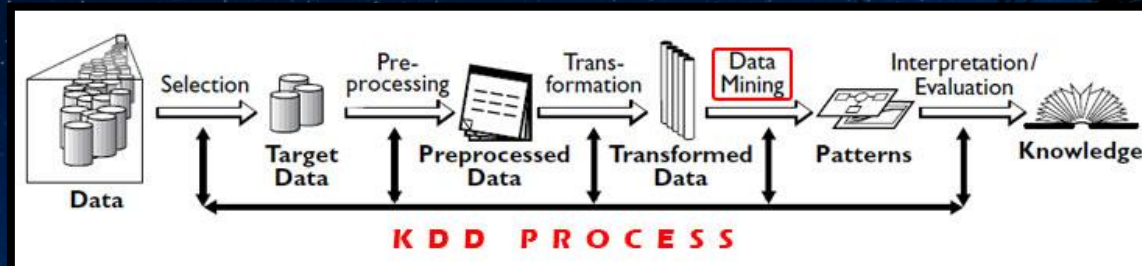
El objetivo aquí es agrupar los datos que posean características similares.



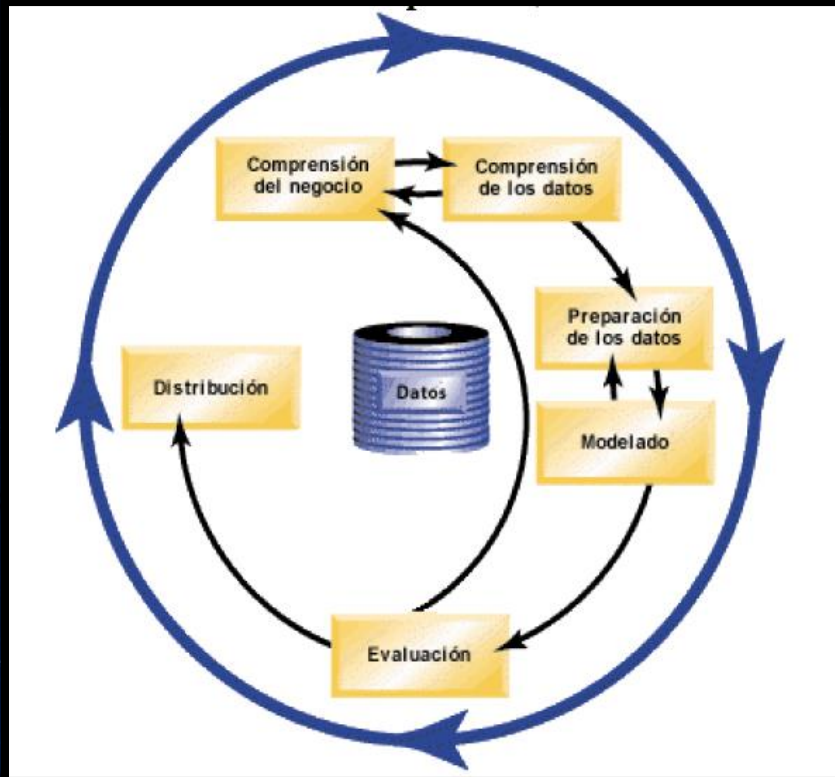
- Clustering (k-means, GMM, SOM...)
- Descripción y Visualización

Minería de datos vs Aprendizaje Automático

- La Minería de Datos busca el descubrimiento del conocimiento sin una hipótesis preconcebida.
- “KDD” (Knowledge Discovery in Databases) es el proceso no trivial de identificar patrones a partir de los datos, que sean novedosos y potencialmente útiles.
- La minería de Datos va más allá de clasificar patrones como “bien o “mal”, que es algo que intentamos hacer con Aprendizaje Automático.
- No obstante las técnicas utilizadas son muy similares.



Metodología CRISP-DM (libre, líder)



Metodología más utilizada para procesos de Machine Learning (Data Mining).

1. Comprensión del negocio
2. Comprensión de los datos
3. Preparación de los datos
4. Modelado
5. Evaluación
6. Despliegue / Distribución

Etapas en la creación de un modelo de AA

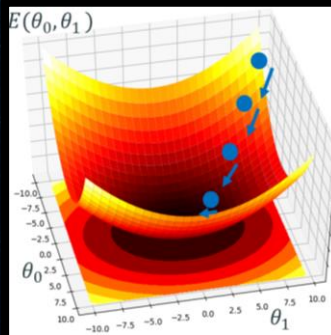


Datos

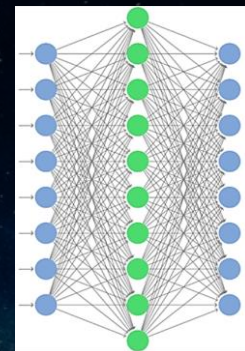


1	2	3	C
0,5	0,8	0,1	0,1
0,6	0,2	0,4	0,9
0,6	0,7	0,3	0,2
0,2	0,1	0,2	0,3
0,6	0,5	0,9	0,3

Preprocesamiento



Entrenamiento



Modelo



Nuevos datos
del mundo real



Sin datos no podemos
realizar un proceso de
Machine Learning

Casi siempre es
necesario preprocesar
los datos

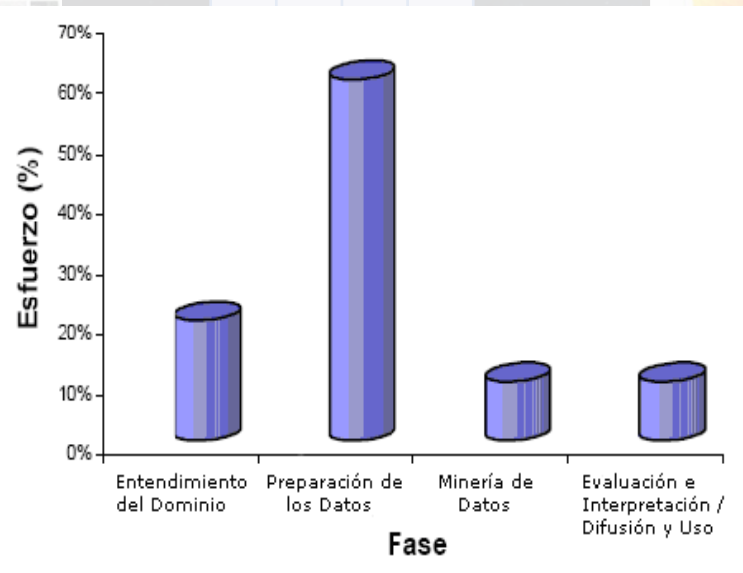
Minimización de una
función de error

Etapas en la creación de un modelo de AA



Datos

1	2	3	C
0,5	0,8	0,1	0,1



En la práctica, como regla general se gasta más del 80% del tiempo preprocesando los datos y menos del 20% creando un modelo.

Entrenamiento

Modelo

Minimización de una función de error

Nuevos datos del mundo real



Vista Minable

Una **vista minable** es un subconjunto de los datos organizados de un modo que me permitan aplicar algoritmos de Machine Learning.

Verificar con expertos del negocio:

- ¿Es necesario el esfuerzo?
- ¿Hay subproblemas más importantes?
- ¿Qué piensan ellos que debe hacerse?
- Asegurarse de que el esfuerzo lo vale.

Preparar los datos:

- Consolidar datos. Revisar datos incompletos/nulos.
- Revisar representatividad.

Aprendizaje Automático - Modelos

Modelos de caja blanca



VS .

Modelos de caja negra



Es posible interpretar el modelo:

- Regresión lineal/ Logística
- Reglas – árboles
- K-means

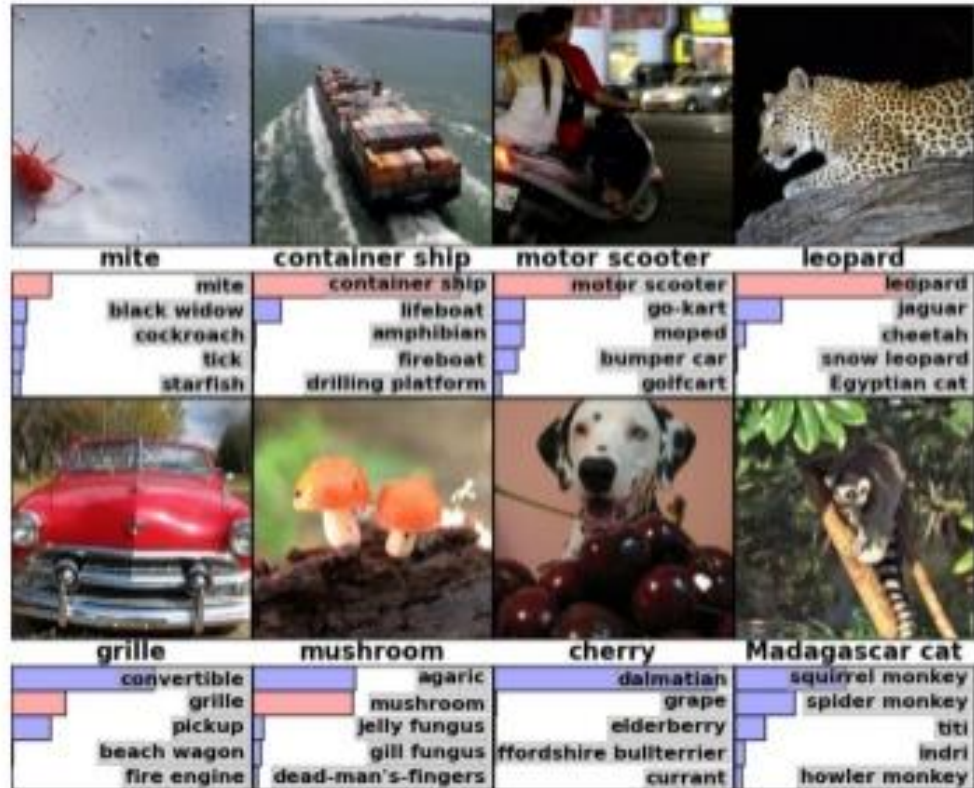
NO es posible interpretar el modelo:

- Redes Neuronales
- En cierta medida: Deep Learning

Bases de datos actuales

IMAGENET

- 1,000 object classes (categories).
- Images:
 - 1.2 M train
 - 100k test.



Bases de datos actuales























COCO dataset - Microsoft

es un conjunto de datos de detección, segmentación y subtitulación de objetos a gran escala. 37 Gb, 330 mil imágenes. Entre 100 y 200 clases de objetos.



MNIST

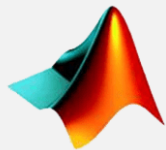
70 mil patrones de números escritos a mano por 250 personas

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

Lenguajes/librerías

Hace algunos años...

C++



MATLAB

Orientado a Minería de Datos



rapidminer



orange



pythonTM



PYTORCH



ETC...



Keras

