

Curso de aprendizaje estadístico

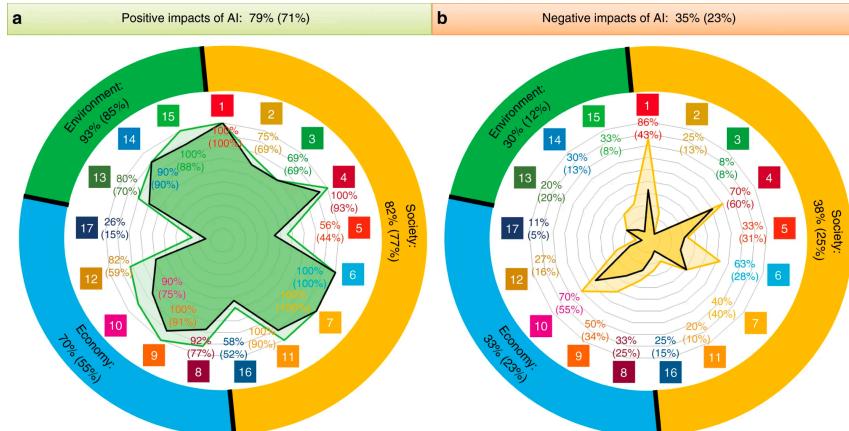
Hernan David Salinas
Universidad de Antioquia

Introducción

Unesco: 17 objetivos de desarrollo sostenible y la agenda 2030



169 targets recognized in the 2030 Agenda for Sustainable Development. [Ref]



nature communications

Explore content ▾ About the journal ▾ Publish with us ▾

nature > nature communications > perspectives > article

Perspective | Open Access | Published: 13 January 2020

The role of artificial intelligence in achieving the Sustainable Development Goals

Ricardo Vinuesa, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia

Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark & Francesco Fuso Nerini

Nature Communications 11, Article number: 233 (2020) | Cite this article

296k Accesses | 595 Citations | 836 Altmetric | Metrics

communications materials

Explore content ▾ About the Journal ▾ Publish with us ▾

nature > communications.materials > perspectives > article

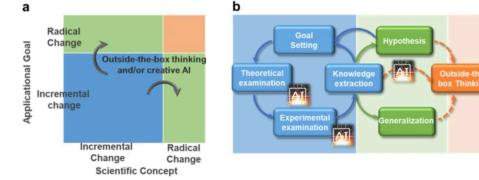
Perspective | Open Access | Published: 30 August 2022

Why big data and compute are not necessarily the path to big materials science

Nachiro Fujimura, Brian DeCost, Jason Harrick-Simpers & Samuel E. Lofland

Communications Materials 3, Article number: 59 (2022) | Cite this article

3477 Accesses | 2 Citations | 10 Altmetric | Metrics



Introducción

Algunas tecnologías de la industrialización



• Bing Chat (2023). Generadas por IA. 18 de julio de 2023



Omnipresente: Medicina, Ciencias, Banca, transporte, chatbot ...etc

"las máquinas ya no son meras herramientas, sino parejas igualitarias en la exploración científica, intercambiando ideas, intuición y comprensión con los humanos" Jim Gray, premio Turing 1998

"De los muchos, diversos y fascinantes desafíos a los que nos enfrentamos hoy en día, el más intenso e importante es cómo entender y dar forma a la nueva revolución tecnológica que implica nada menos que una transformación de la humanidad"
(Schwab, 2017, p. 1, Unesco 2021).

Introducción

Big data

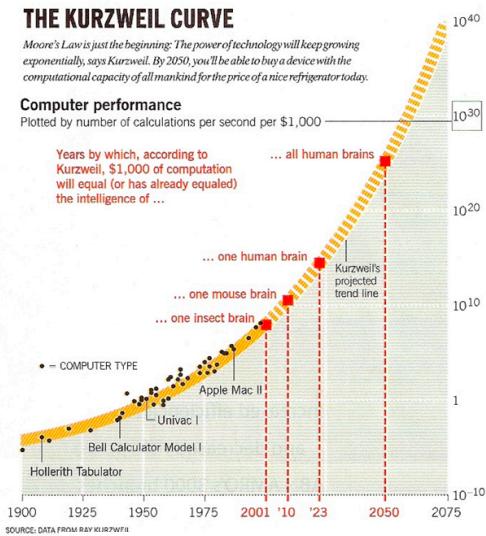
THE KURZWEIL CURVE

Moore's Law is just the beginning: The power of technology will keep growing exponentially, says Kurzweil. By 2050, you'll be able to buy a device with the computational capacity of all mankind for the price of a nice refrigerator today.

Computer performance

Plotted by number of calculations per second per \$1,000

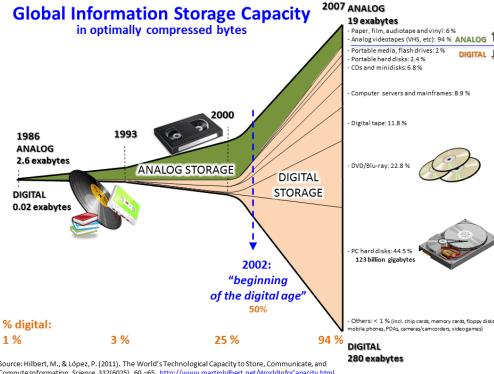
Years by which, according to Kurzweil, \$1,000 of computation will equal (or has already equaled) the intelligence of ...



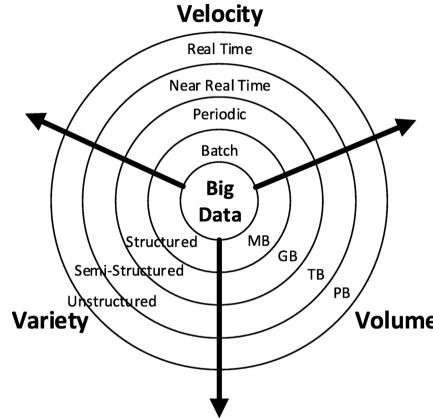
Ley de rendimientos acelerados,
generalización de Moore

- Big Data aspects
- Variety: Data types
 - Velocity: Data production and processing speed
 - Volume: Data size
 - Veracity: Data reliability and trust
 - Value: Worth derived from exploiting Big Data

Global Information Storage Capacity in optimally compressed bytes

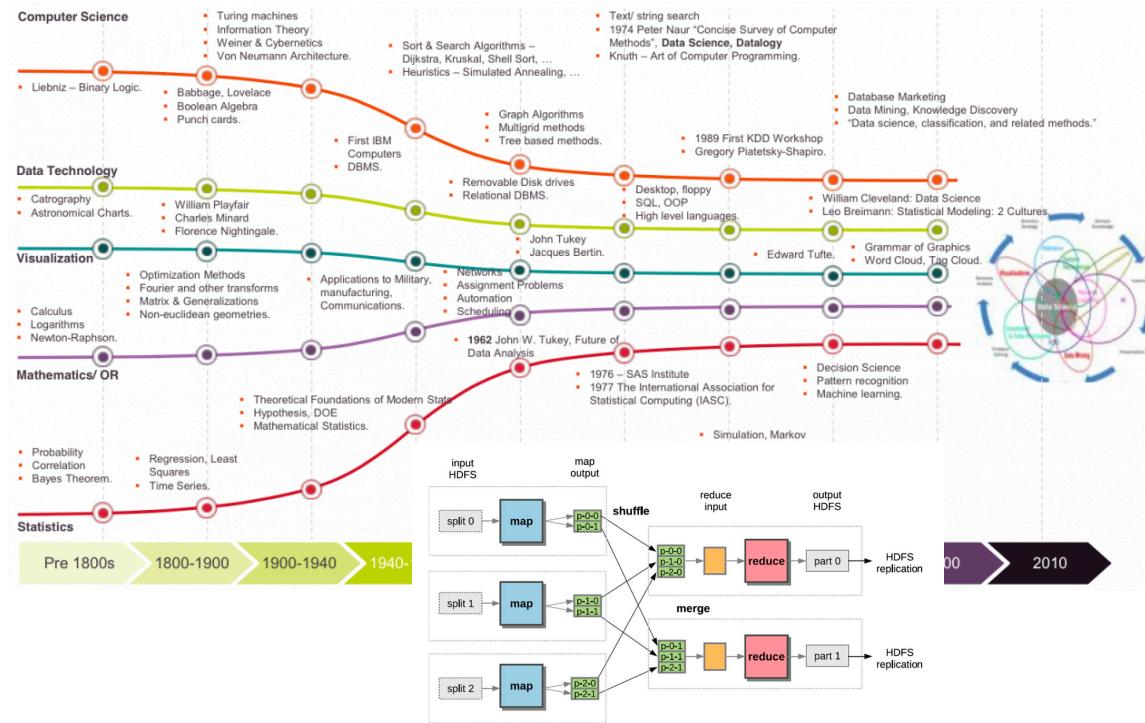


Source: Hilbert, M., & Lopez, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 6-6. <http://www.martinhilbert.net/WestStorageCapacity.html>



Introducción

Evolution diferentes areas



Introducción

Aplicaciones big data en ciencia

Sloan Digital Sky Survey (Encuesta Digital del Cielo de Sloan).



Los conjuntos de datos incluyen imágenes y espectros ópticos e infrarrojos, así como mapas espectroscópicos integrales de campo. Estos datos son muy valiosos para hacer big data y han permitido muchos descubrimientos científico



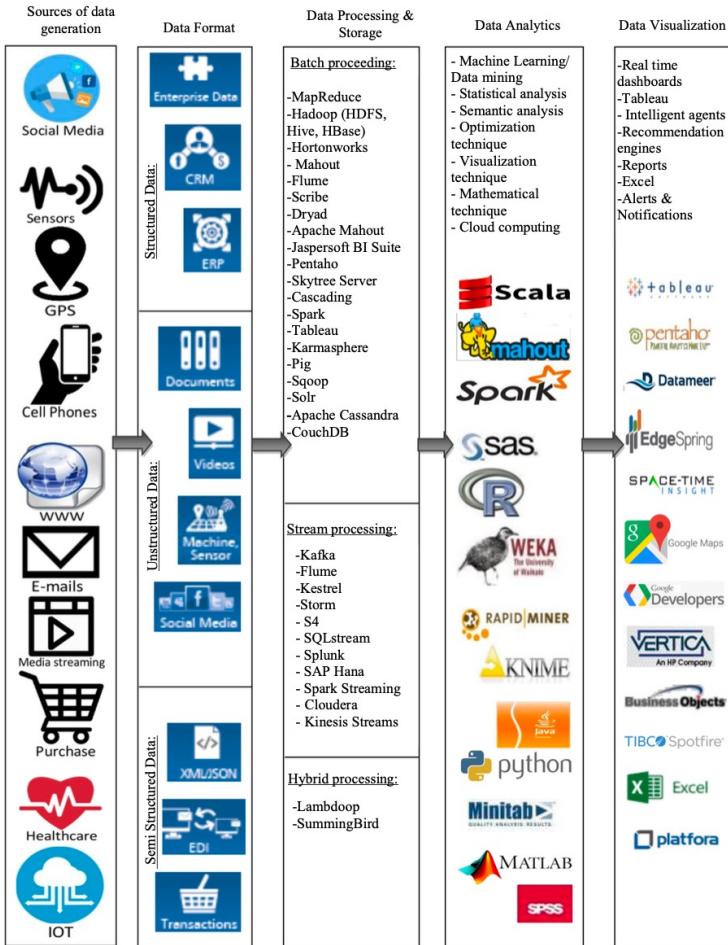


Fig. 3 Classification big data Framework



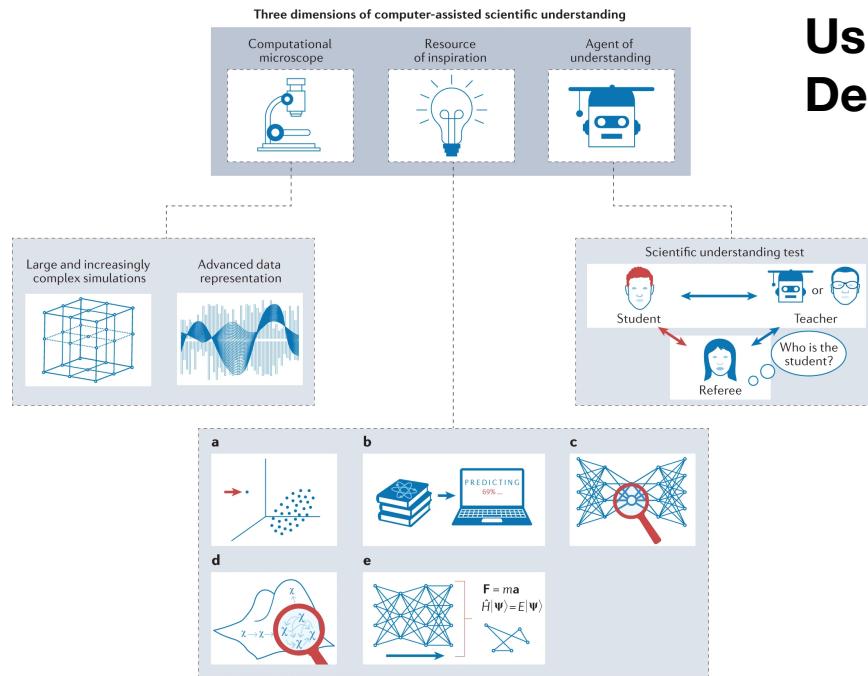
Característica	CERN	Big data empresarial
FUENTE	Los datos se generan en las colisiones de partículas que tienen lugar en los detectores del LHC, como ATLAS, CMS, ALICE o LHCb. Estos detectores registran el paso de cada partícula como una serie de señales electrónicas, que se envían al	Los datos pueden provenir de diversas fuentes internas o externas a la organización, como transacciones, registros, sensores, redes sociales, etc. Estos datos pueden tener diferentes formatos y estructuras, y pueden ser
Tipos	Los datos del CERN son principalmente estructurados, ya que se registran como una serie de señales electrónicas que se digitalizan y se almacenan como eventos de colisión. Estos datos tienen un formato y una organización definidos, y se pueden	El big data empresarial puede ser estructurado, semi-estructurado o no estructurado, dependiendo del grado de organización y formato de los datos. Los datos estructurados tienen un formato y una organización definidos, como las
Objetivos	El big data del CERN tiene como objetivo avanzar en el conocimiento científico, mediante el análisis de los datos para extraer información física relevante y buscar nuevas partículas o fenómenos físicos. Los datos del CERN se usan para	El big data empresarial tiene como objetivo mejorar la toma de decisiones, la eficiencia, la innovación y la competitividad de una organización, mediante el análisis de los datos para extraer información útil y valiosa. Los
Métodos	El big data del CERN se analiza mediante programas y herramientas específicos desarrollados por la comunidad científica del CERN y sus colaboradores, como el sistema de disparo, la Red Mundial de Computación del LHC (WLCG), o el	El big data empresarial se analiza mediante diversas técnicas y herramientas de análisis de datos, como el procesamiento paralelo y distribuido, el aprendizaje automático, la inteligencia artificial, la visualización de datos, etc.



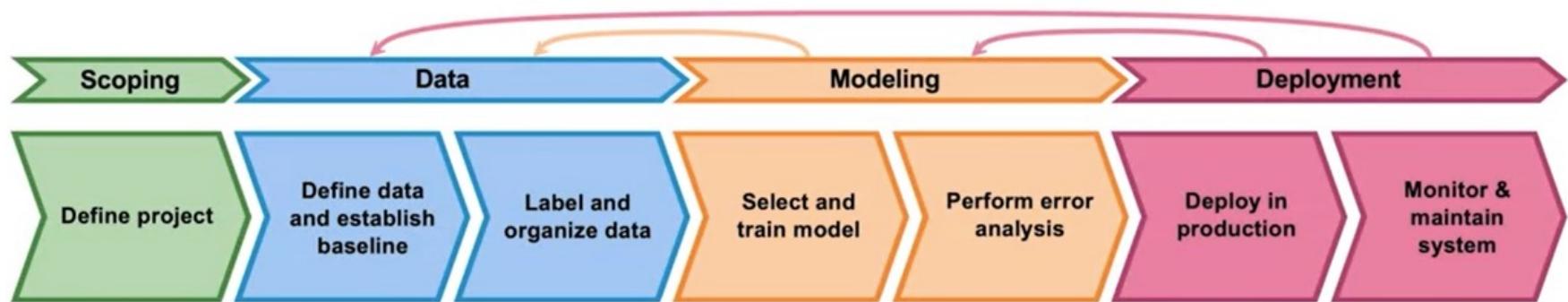
Introducción

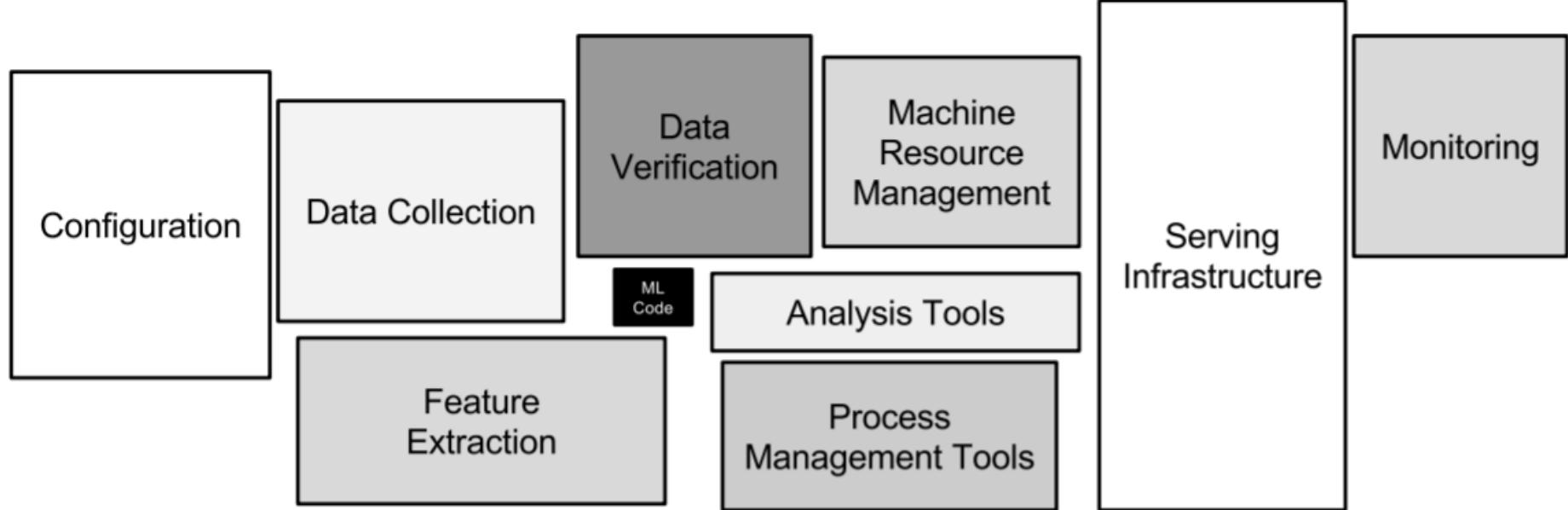
Computación científica

Usuario Desarrollador



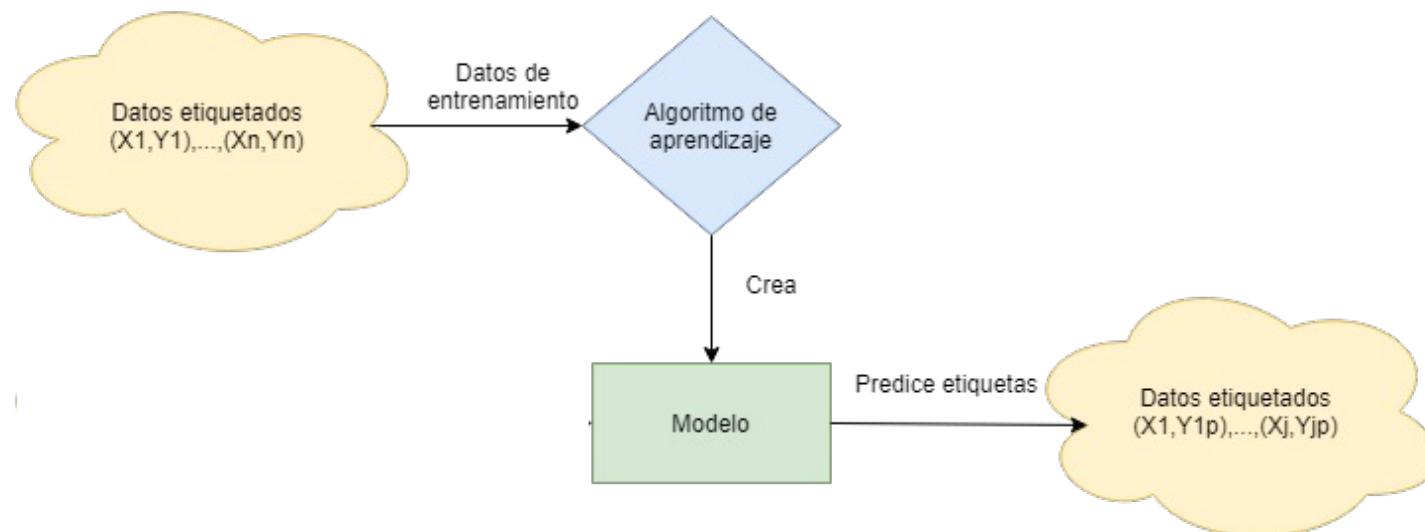
The ML project lifecycle





Aprendizaje supervisado

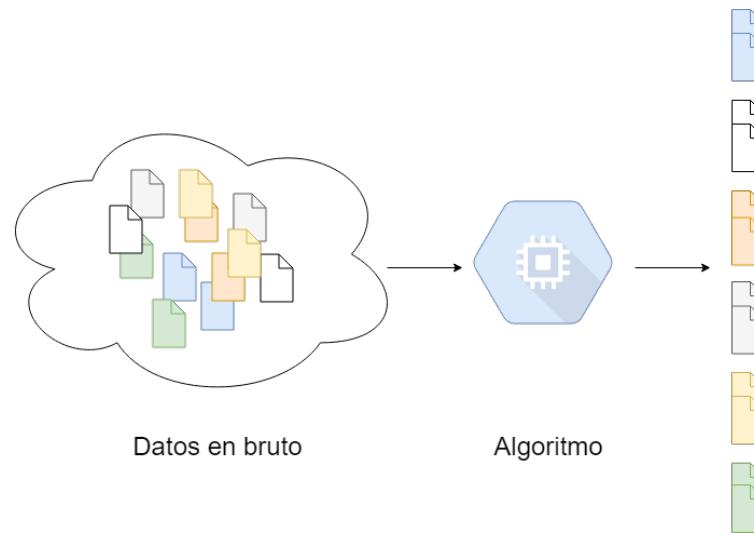
Aprendizaje supervisado: Es el tipo de aprendizaje en el que se proporcionan datos etiquetados con la salida deseada a la computadora. El objetivo es que la computadora aprenda a predecir la salida a partir de los datos de entrada. [Por ejemplo, se puede entrenar a una computadora para reconocer imágenes de gatos y perros, proporcionándole muchas imágenes con la etiqueta correspondiente](#)



Aprendizaje no supervisado

•².

• **Aprendizaje no supervisado:** Es el tipo de aprendizaje en el que no se proporcionan datos etiquetados ni salidas deseadas a la computadora. El objetivo es que la computadora descubra patrones, estructuras o agrupaciones en los datos sin ninguna guía previa. [Por ejemplo, se puede usar el aprendizaje no supervisado para segmentar clientes según sus características o preferencias¹².](#)



Aprendizaje semisupervisado

Es el tipo de aprendizaje que combina el aprendizaje supervisado y el no supervisado. Se utiliza cuando se dispone de una gran cantidad de datos, pero solo una pequeña parte está etiquetada. El objetivo es aprovechar tanto los datos etiquetados como los no etiquetados para mejorar el rendimiento del modelo. [Por ejemplo, se puede usar el aprendizaje semisupervisado para mejorar el reconocimiento de voz, utilizando datos etiquetados y no etiquetados](#)



Aprendizaje por refuerzo

Es el tipo de aprendizaje en el que la computadora aprende mediante la interacción con un entorno y recibiendo recompensas o castigos por sus acciones. El objetivo es que la computadora encuentre la mejor estrategia para maximizar la recompensa acumulada. [Por ejemplo, se puede usar el aprendizaje por refuerzo para enseñar a un robot a caminar, dándole una recompensa por cada paso que da correctamente](#)

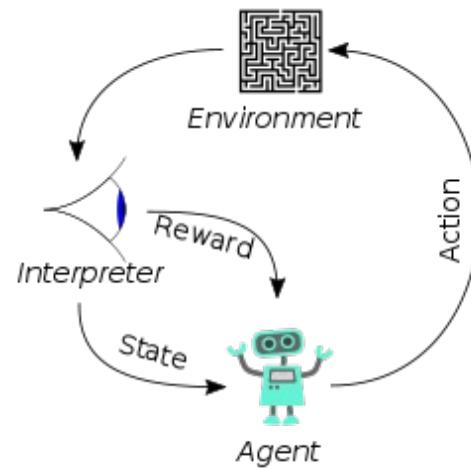


Chart Title			
Aprendizaje supervisado	Aprendizaje no supervisado	Aprendizaje por refuerzo	Aprendizaje semisupervisado
Algoritmos de regresión, algoritmos bayesianos, algoritmos de árbol de decisión, algoritmos de redes neuronales, algoritmos de aprendizaje profundo ¹	Algoritmos de agrupación, algoritmos de reducción de dimensión ¹	Algoritmos que aprenden mediante la interacción con un entorno y recibiendo recompensas o castigos por sus acciones, como los algoritmos Q-learning o los algoritmos SARSA ²	Algoritmos que combinan el aprendizaje supervisado y el no supervisado

Nombre del algoritmo	Tipo de aprendizaje	Aplicación
Clasificación de Naïve Bayes	AS	Clasificación
Bosques aleatorios	AS	Clasificación y regresión
Regresión lineal	AS	Regresión
Regresión logística	AS	Clasificación
Máquinas de vectores de soporte (SVM)	AS	Clasificación y regresión
K vecinos más cercanos (KNN)	AS	Clasificación y regresión
Redes neuronales y aprendizaje profundo	AS	Clasificación y regresión
Algoritmos de agrupación	NS	Clasificación
Algoritmos de reducción de dimensión	NS	Regresión
Algoritmos Q-learning o SARSA	APR	Refuerzo
Árboles de decisión	AS	Clasificación y regresión
Análisis discriminante lineal (LDA)	AS	Clasificación
Análisis discriminante cuadrático (QDA)	AS	Clasificación
Análisis de componentes principales (PCA)	NS	Regresión
Análisis de correlación canónica (CCA)	NS	Regresión
Algoritmo genético (GA)	NS	Optimización o búsqueda
Autoencoder (AE)	NS	Codificación o compresión

