

Ciencia de Datos

- Módulo 2

Introducción al Aprendizaje automático



Inteligencia Artificial



Fuente:

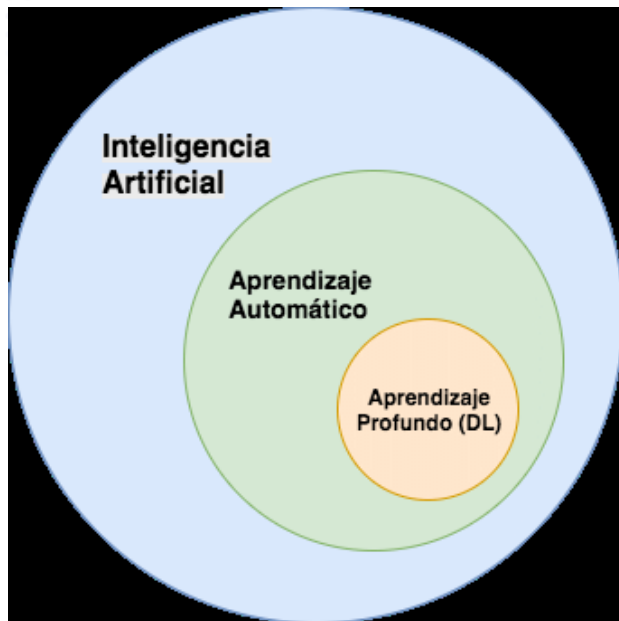
<https://www.pandasecurity.com/es/mediacenter/mobile-news/inteligencia-artificial/>

Inteligencia Artificial es un campo de estudio que combina matemática, estadística e informática para intentar desarrollar tecnología que pueda mostrar inteligencia humana.

“Crear programas de ordenador o máquinas capaces de conductas que consideraríamos inteligentes si las efectuaran seres humanos”
(Kaplan, Jerry: *Inteligencia artificial. Lo que todo el mundo debe saber*, 2017)



Aprendizaje Automático (Machine learning)



Fuente: <https://www.researchgate.net/>

Dentro del campo de Inteligencia Artificial existen diversos subcampos:

- Planificación: técnicas para realizar determinadas acciones a través de código dándole instrucciones precisas a las máquinas.
- Machine Learning: conjunto de herramientas y técnicas que aprenden de las experiencias y extraen patrones de los datos.



Aprendizaje Automático

“Campo de estudio que brinda a las computadoras la capacidad de aprender sin ser programadas explícitamente” Arthur samuel, 1959

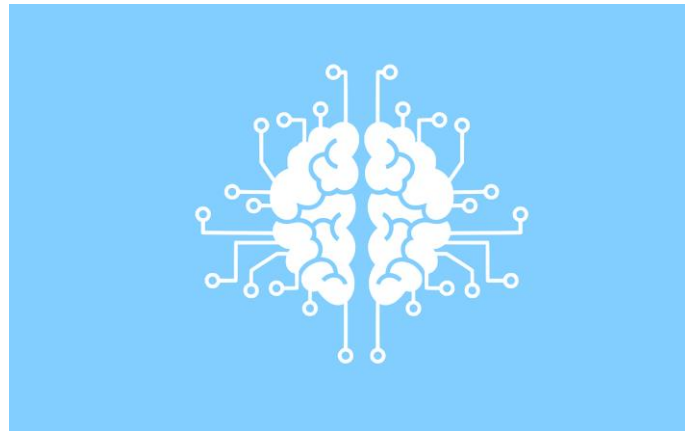


Fuente: <https://www.xatakaciencia.com/computacio/>



Aprendizaje Automático

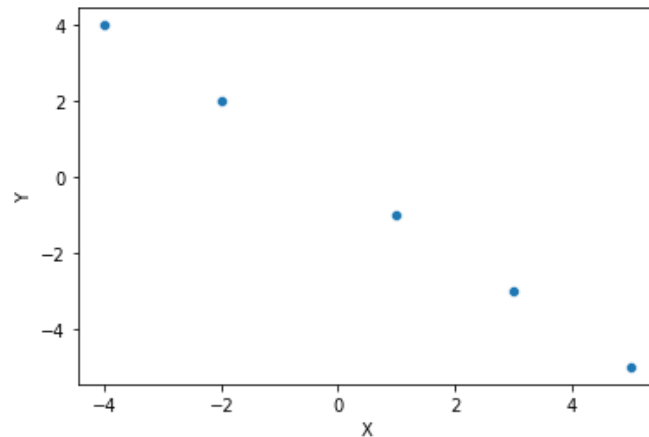
Cuando nos referimos a aprendizaje automático o ML (Machine Learning), estamos hablando de una sub rama de la inteligencia artificial, son algoritmos que buscan patrones en un grupo de datos, luego, al recibir un dato nuevo pueden estimar el valor de una característica específica de ese dato nuevo.



Aprendizaje Automático

Por ejemplo, supongamos que tenemos el siguiente grupo de datos. Si recibimos un nuevo dato de $X = 2$ ¿Qué valor tendría en Y ? ¿Cómo podemos saberlo?

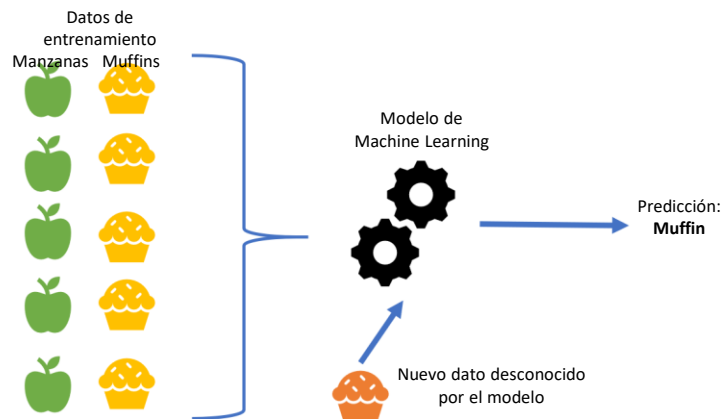
X	Y
1	-1
-2	2
3	-3
5	-5
-4	4



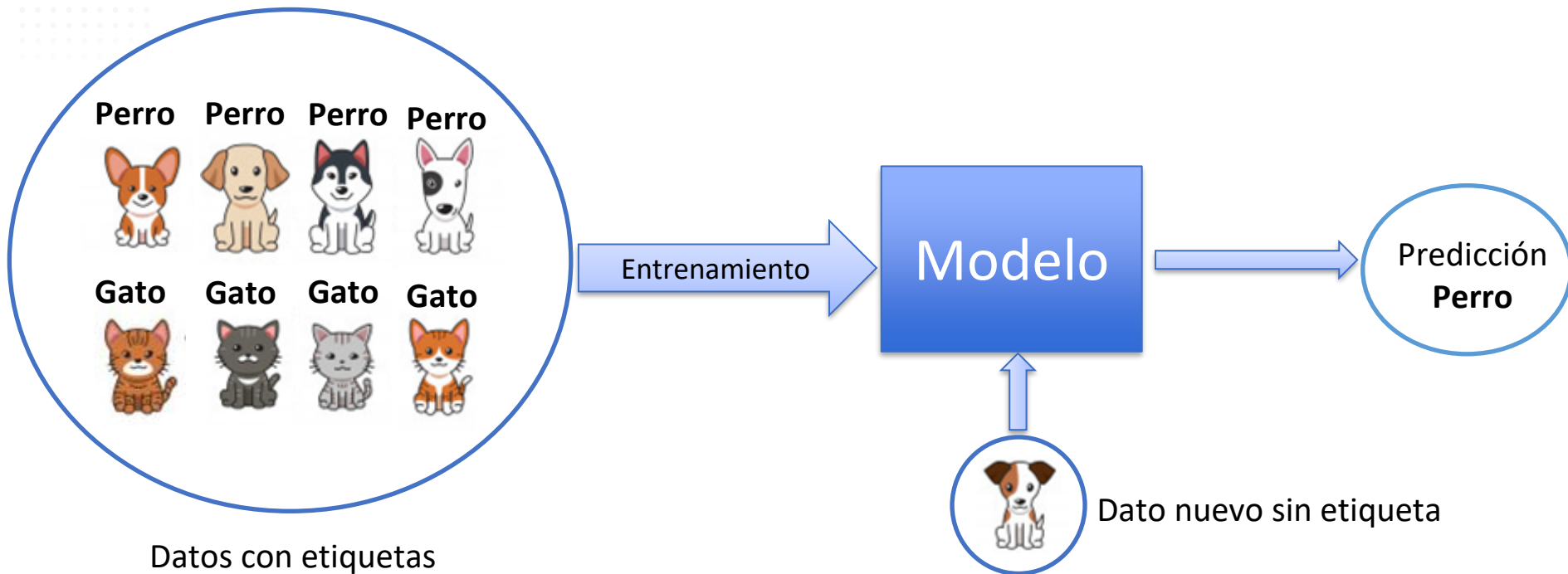
Aprendizaje Automático - Tipos

Supervisado: Se dice que un modelo de ML es supervisado cuando, los datos con los que se entrena el modelo tienen los valores que queremos predecir, a estos se le dice que están etiquetados.

Por ejemplo, si queremos predecir cuántos goles marcará nuestro equipo en base a datos de partidos anteriores, estos datos deben tener cuántos goles marcó nuestro equipo en cada uno de ellos.



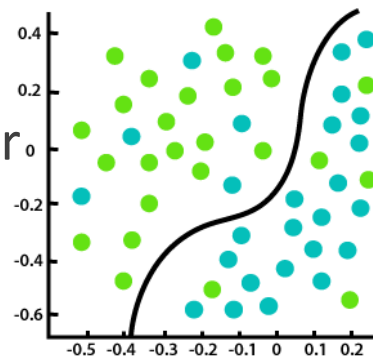
Aprendizaje supervisado



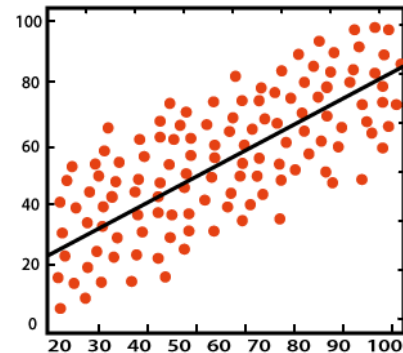
Aprendizaje Automático - Tipos

Los principales usos de los modelos supervisados son:

- Clasificación: Cuando la variable a predecir es una clase, por ejemplo: Enfermo o no enfermo, la raza de la foto de un animal, etc.
- Regresión: Cuando la variable a predecir es un valor, por ejemplo: Precio de un objeto, nota de un estudiante, probabilidad de lluvia, etc.



Clasificación



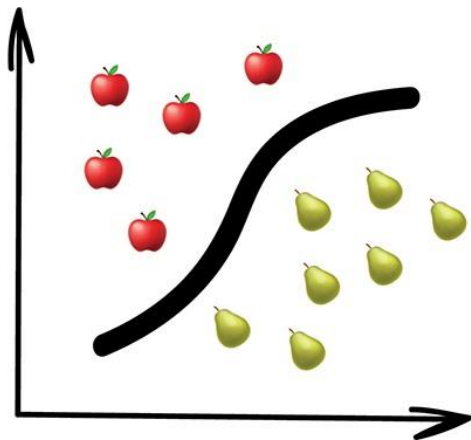
Regresión

Fuente: <https://www.javatpoint.com/regression-vs-classification-in-machine-learning>



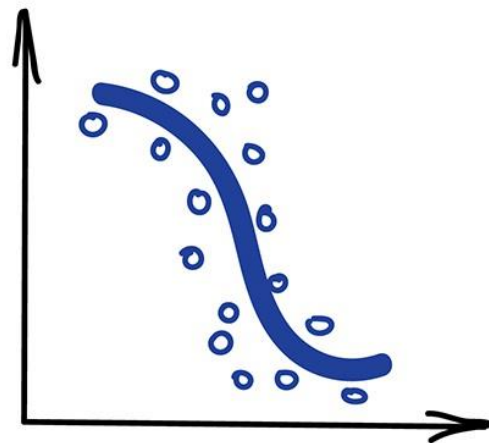
Clasificación - Regresión

Etiqueta cualitativa



Classification

Etiqueta cuantitativa



Regression

Fuente: <https://www.juanbarrios.com/inteligencia-artificial-y-machine-learning-para-todos>



Aprendizaje Automático - Tipos

La mayoría de los modelos de aprendizaje automático supervisados pueden aplicarse tanto para problemas de clasificación como de regresión, sin embargo cada modelo suele ser mejor en determinados problemas.

Es importante saber que no existe el modelo perfecto que sirve para todos los casos, solo con la experiencia podremos saber que modelos aplicar a cada caso.

Algunos modelos de ML supervisados son:

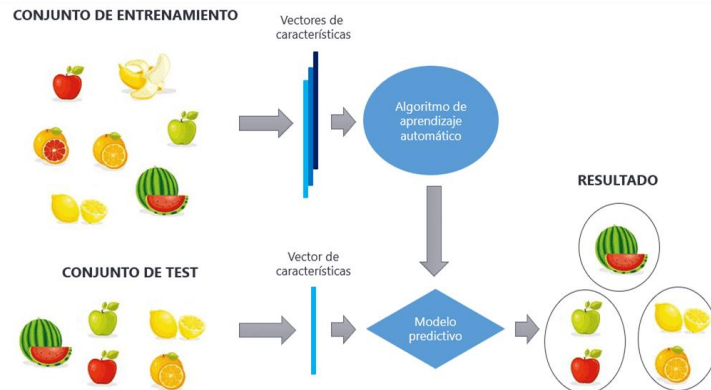
- Árboles de decisión.
- Regresión Lineal.
- Regresión Logística.
- Máquinas de vector soporte.
- Vecinos más cercanos.
- Redes Neuronales.



Aprendizaje Automático - Tipos

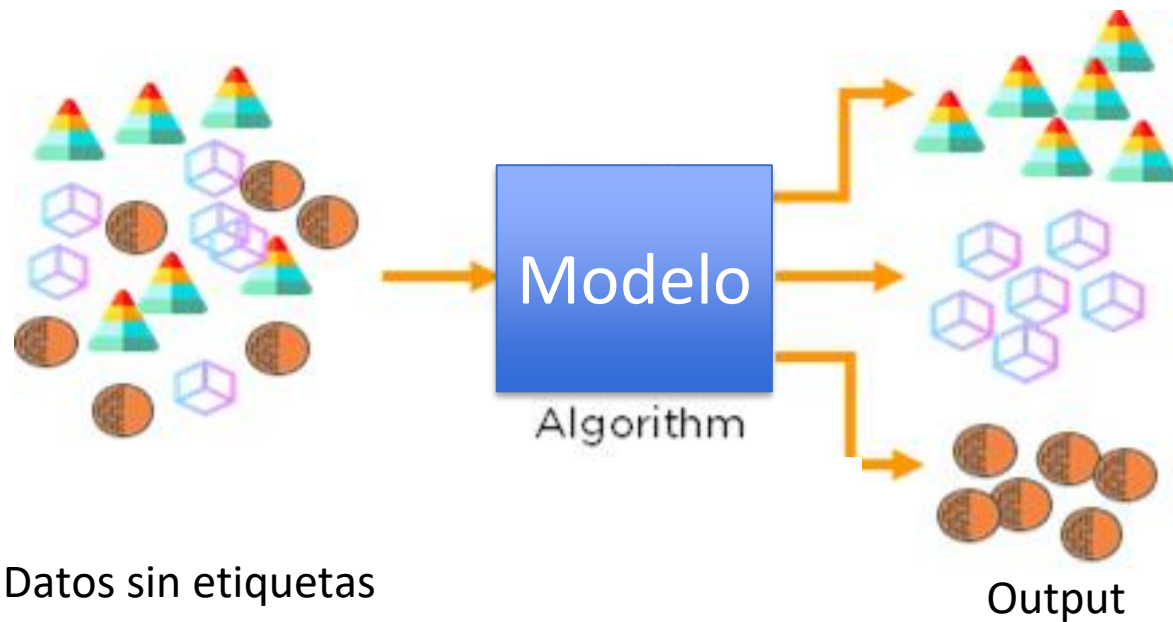
No supervisado: Estos modelos se entrenan sin información sobre el atributo que se quiere predecir, por lo que la evaluación de su desempeño es más compleja. Sueles usarse para realizar agrupaciones o clusters y como métodos de reducción de dimensionalidad.

Por ejemplo, cuando una compañía quiere hacer grupos de clientes según sus gastos mensuales, cantidades de compras, visitas a la tienda, etc.



Fuente: <https://www.diegocalvo.es/aprendizaje-no-supervisado/>

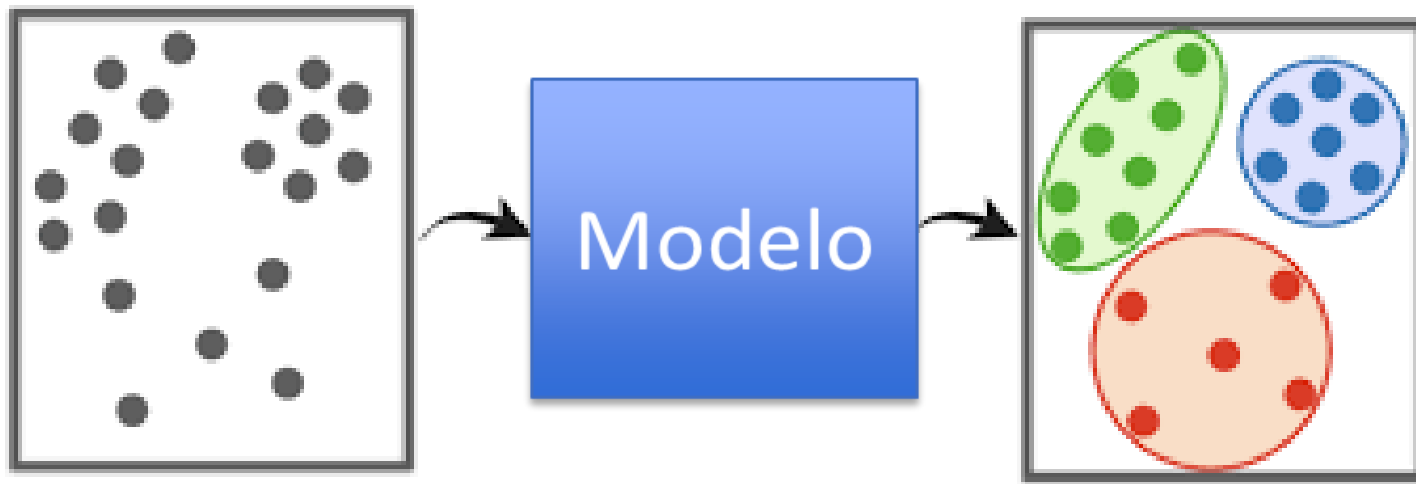
Aprendizaje No supervisado



Fuente: <https://es.clariba.com/machine-learning-for-business>



Aprendizaje No supervisado



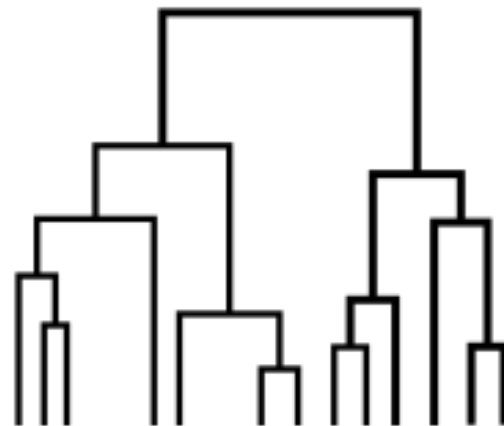
Fuente: <https://es.clariba.com/machine-learning-for-business>



Aprendizaje Automático - Tipos

Los modelos de aprendizaje no supervisado más comunes son:

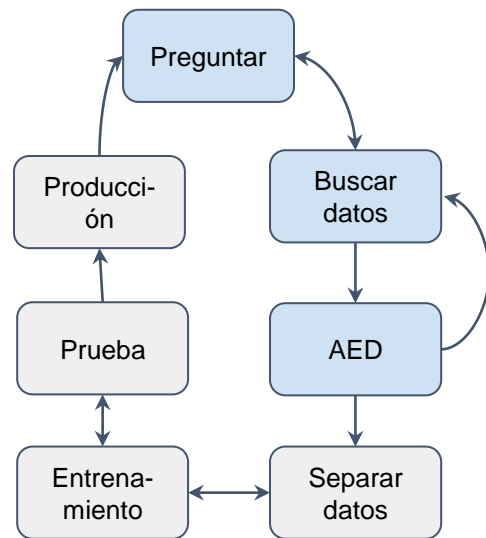
- K-Means.
- PCA (Análisis de componentes principales)
- Propagación de afinidad.
- Clustering aglomerativo.



Aprendizaje Automático - Pasos

En todo modelo de Machine Learning hay algunos **pasos comunes** que nos sirven de guía:

- **Definición del problema:** Siempre que se va a realizar un proyecto de ML es para resolver un problema específico, por ende, primero debemos de realizarnos una pregunta la cual buscaremos resolver.
- **Buscar los datos:** Por lo general debemos definir qué datos necesitamos y de dónde los podemos encontrar. Hoy en día hay muchos datos libres, sean generados por organismos gubernamentales u ONG, o en algunos casos incluso empresas privadas.
- **Análisis exploratorio de datos:** Durante esta etapa se *limpian* los datos y se realiza una exploración de los mismos, buscando características y relaciones que nos ayuden a seleccionar qué variables usar o que modelo usar.



Aprendizaje Automático - Pasos

- **Dividir los datos de entrenamiento y prueba.** Esta división es de vital importancia para poder definir si el modelo tiene un buen rendimiento o no.
- **Entrenamiento del modelo:** En esta etapa se **entrenan** diferentes de modelos y se **prueban** tomando en cuenta alguna métrica previamente definida, seleccionando el modelo que mejor desempeño tenga.
- **Puesta en producción:** Finalmente se realiza la puesta en producción, este paso depende de qué clase de proyecto sea, puede que en algunos casos, colocar en producción solo signifique generar las predicciones para alguna colección de datos, o puede ser implementarlo en un sistema que reciba datos periódicamente.

