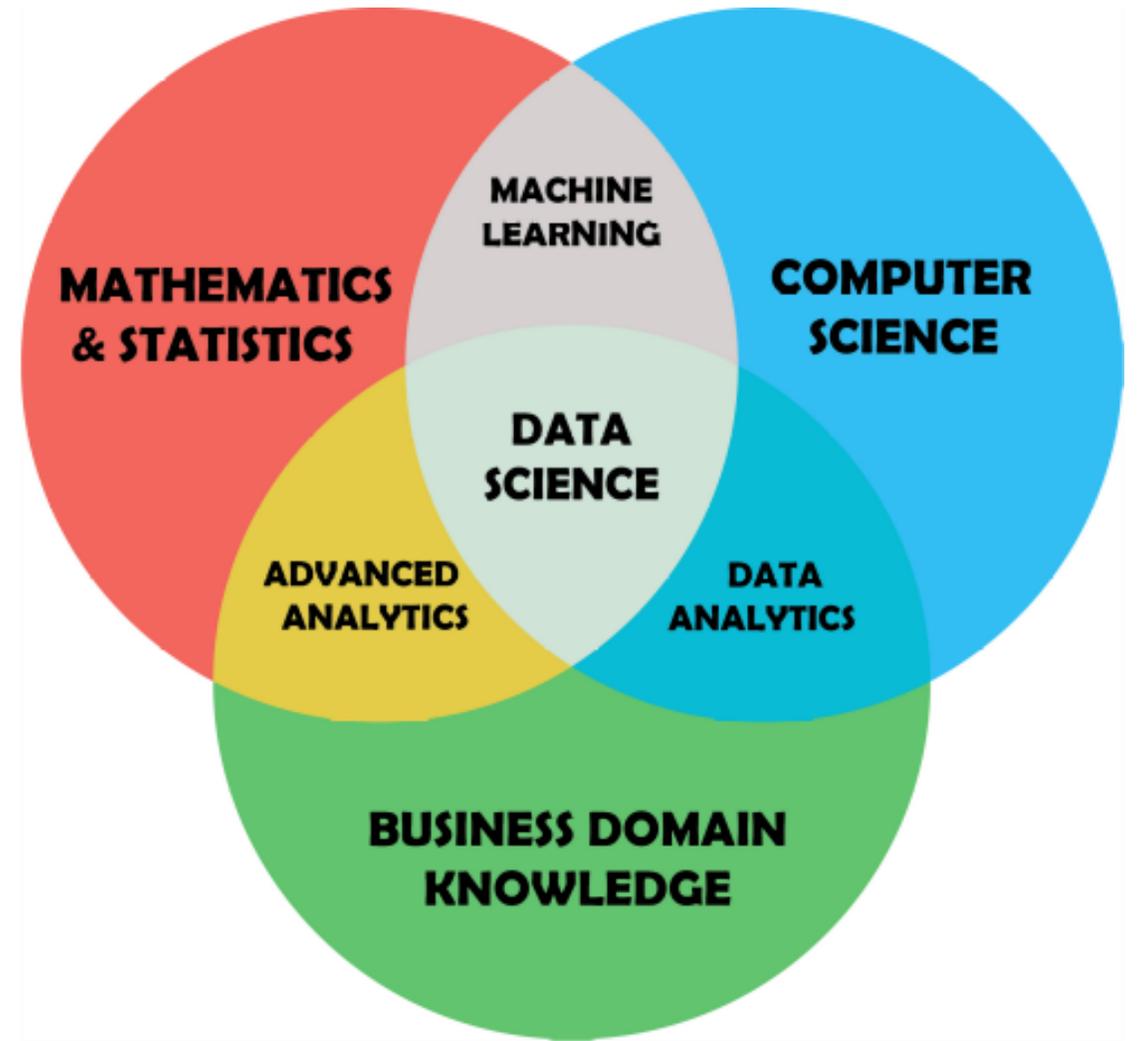
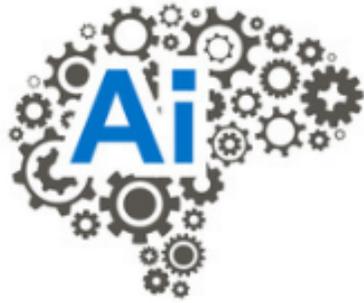
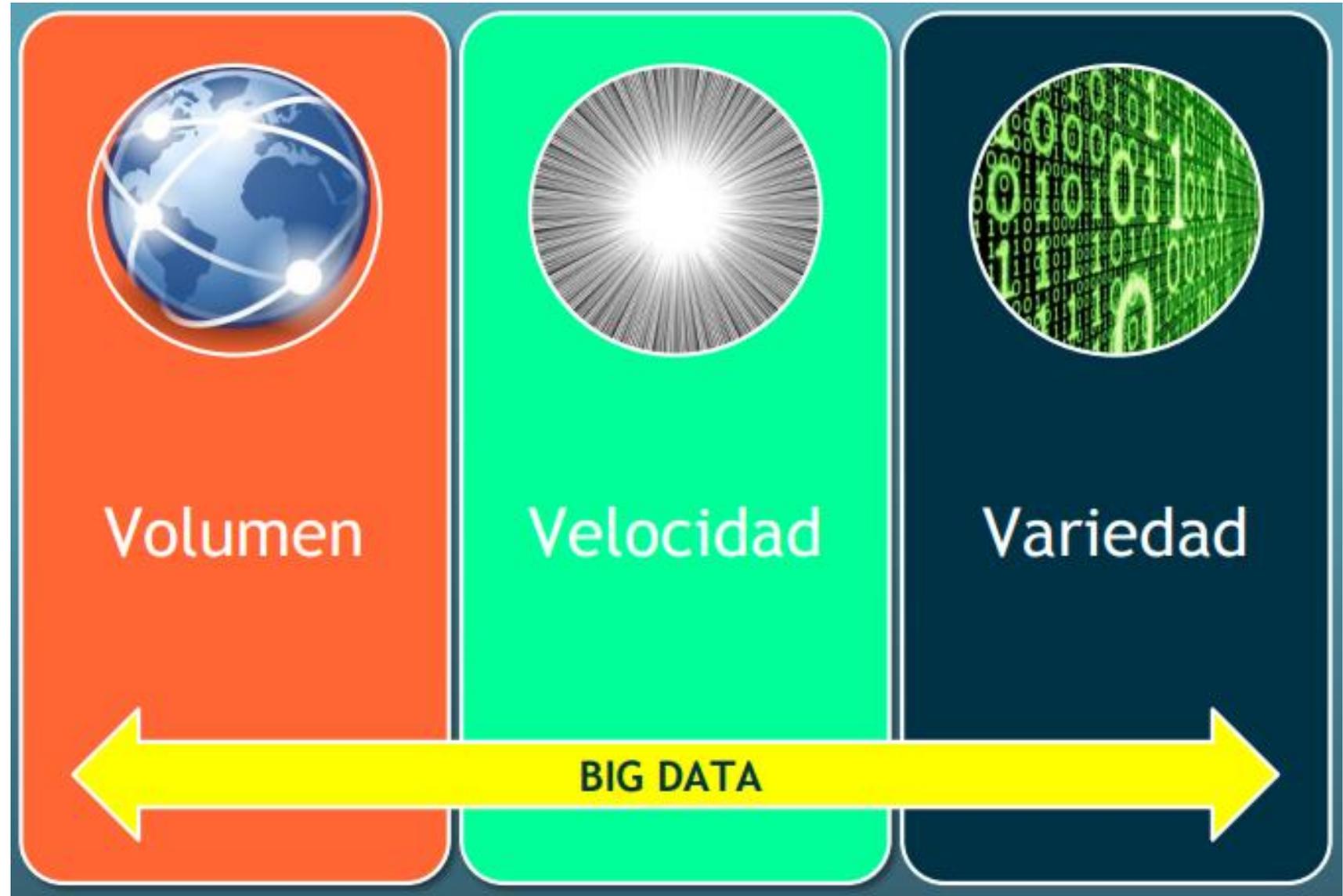


# Conceptos iniciales



Battle of the Data Science Venn Diagrams, David Taylor

# Conceptos iniciales

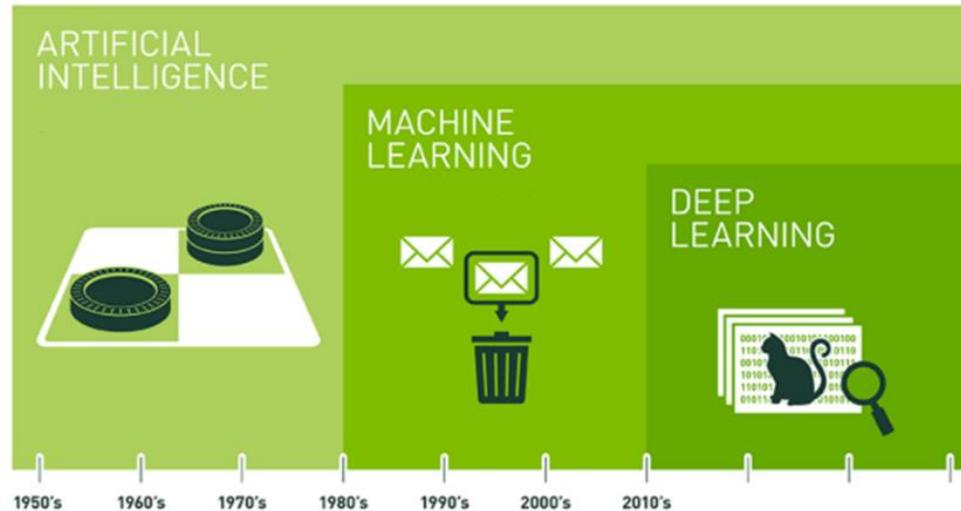


# Machine Learning (Aprendizaje de Máquina)

- **Arthur Samuel, 1959:** “El machine learning (ML) es el campo de estudio que le da a los computadores la habilidad de aprender sin haber sido explícitamente programados”.
- **Tom Mitchell, 1997:** Se dice que un programa de computadora aprende de experiencia E, con respecto a alguna tarea T y alguna medida de desempeño P, mejora con experiencia E, si su desempeño en T, como fue medido con P, mejora con experiencia E.
- **Gavin Hackling, 2014:** el ML es el diseño y estudio de artefactos de programas computacionales que usan experiencia pasada para realizar decisiones futuras

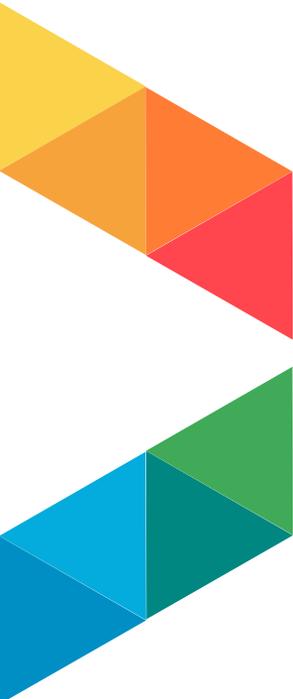


# Machine Learning



Primeros algoritmos de Inteligencia Artificial: Perceptron, 1958

El ML es un subconjunto de la inteligencia artificial (IA) que se centra en desarrollar sistemas que aprenden, o mejoran el rendimiento, en función de los datos que consumen. IA es un término amplio que se refiere a sistemas o máquinas que imitan la inteligencia humana. **Se suele mencionar al ML y a la IA en las mismas conversaciones, y los términos a veces se usan indistintamente, pero no significan lo mismo.** Un aspecto importante a destacar es que aunque todo modelo de ML es IA, no toda IA es ML.

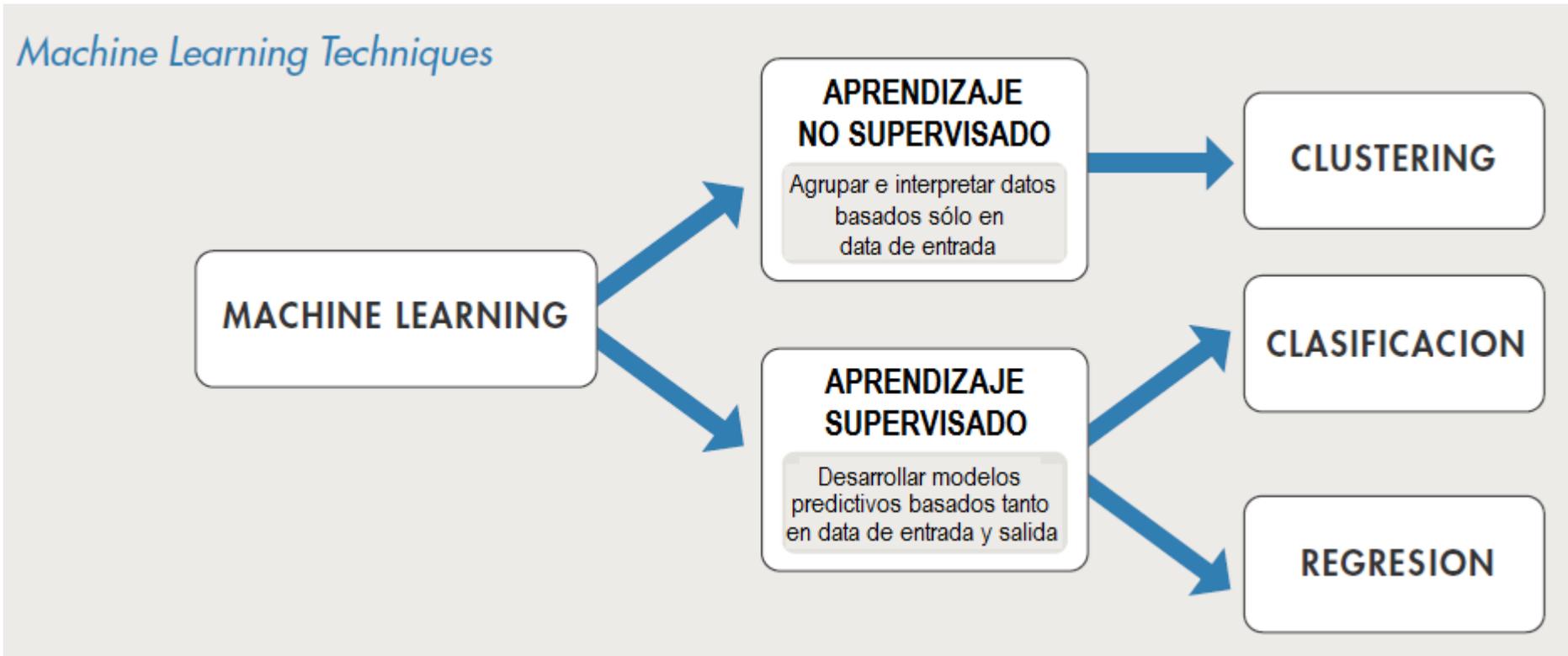


# *Machine Learning*

Dentro del ML encontramos 2 enfoques, el aprendizaje supervisado y el aprendizaje no supervisado.

- El **aprendizaje supervisado** consiste en predecir los valores de un conjunto de datos de salida, a partir de un conjunto de datos de entrada. Se le llama supervisado porque conforme el modelo predice las salidas para datos de prueba, se calcula el error entre lo que predijo el algoritmo y el valor real. El objetivo es minimizar el error, ajustando la función de densidad de probabilidad que relaciona las entradas con las salidas.
- En el **aprendizaje no supervisado**, solo se tienen conjuntos de datos de entrada, sin conocer su relación con variables de salida, por lo que no tiene como verificar fácilmente si el desempeño del modelo es el adecuado. En esta categoría, los modelos de ML realizan principalmente agrupaciones o clasificaciones de los datos, según las variables de entrada y su diferenciación. Son modelos con métricas no muy precisas y que en ocasiones dependen de ciertas heurísticas para validar sus resultados.

# Machine Learning

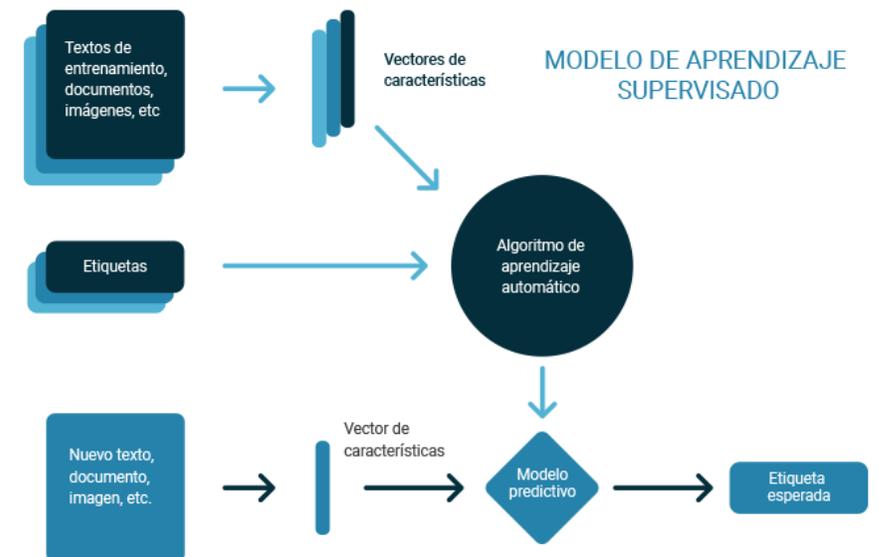


# Aprendizaje Supervisado

El objetivo de las técnicas de aprendizaje supervisado es construir modelos que realizan predicciones basadas en evidencia en presencia de incertidumbre.

Estos algoritmos toman un conjunto conocido de datos de entrada y respuestas conocidas a los datos (salidas) y entrenan modelos para generar predicciones razonables para la respuesta a nuevos datos.

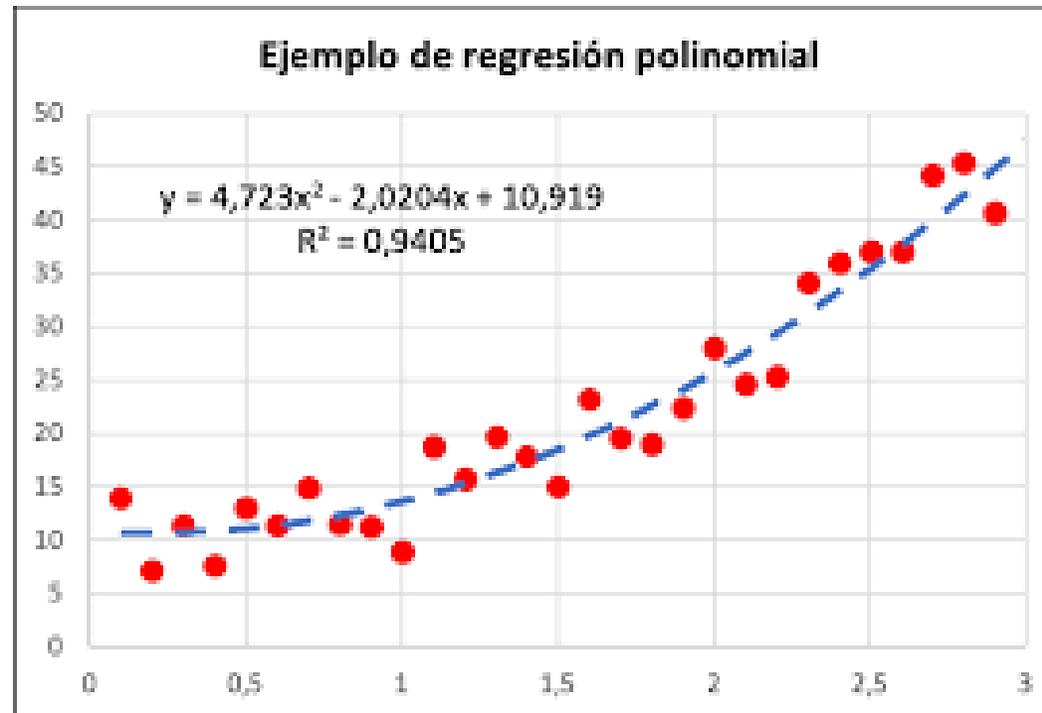
- Los **modelos de clasificación** predicen respuestas discretas, asociando los datos de entrada en categorías.
- Los **modelos de regresión** predicen respuestas continuas.



# Aprendizaje Supervisado: Modelo de Regresión

Se utiliza para predecir el valor de un atributo continuo. Consiste en encontrar la ecuación que ajusta de forma óptima un conjunto de puntos (n-dimensiones).

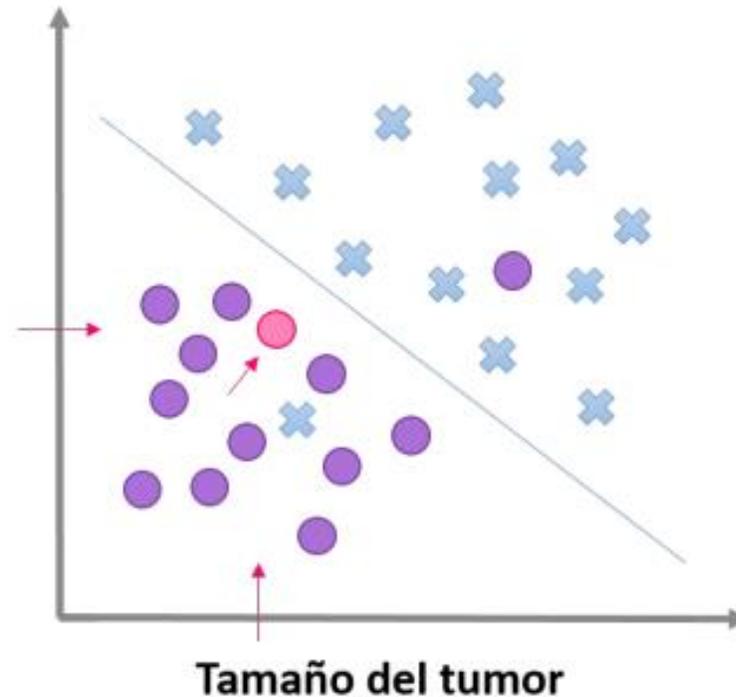
Se utiliza cuando la precisión no es crítica y el número de variables es pequeño.



# Aprendizaje Supervisado: Modelo de Clasificación

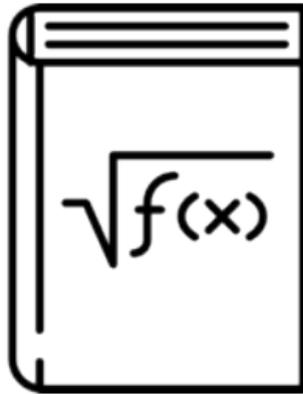
Se utiliza para predecir el resultado de un atributo con valor discreto (a, b, c, ...) dadas unas características ( $X_0, X_1, X_2, X_3, \dots, X_n$ ).

El método simple de clasificación es el binario, donde se clasifica un registro de variables de entrada en 1 o 0. La clasificación múltiple es una extensión de la clasificación binaria.

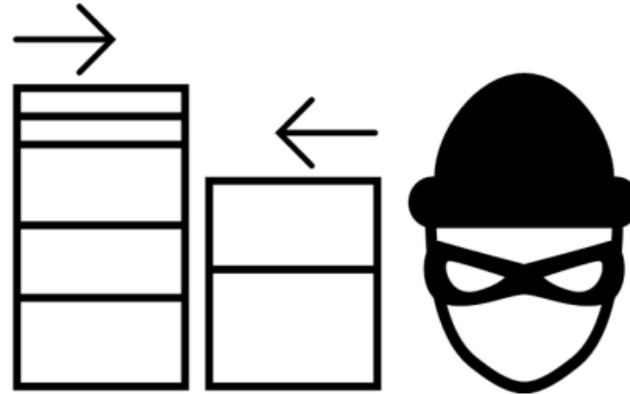


## ¿Cuándo Deberíamos Usar ML?

Cuando tenga una tarea compleja o problema que involucra una gran cantidad de datos o muchas variables, pero ninguna fórmula o ecuación existente que pueda modelar el problema.



Reglas y ecuaciones demasiado complejas para ser modeladas



Las reglas de una tarea o el patrón de fraude cambian periódicamente



La naturaleza de los datos cambia o el programa necesita adaptarse

# Más Datos, Más Preguntas, Mejores Respuestas...

Los algoritmos de ML encuentran patrones naturales en los datos que generan ideas y ayudan a tomar mejores decisiones y predicciones.

Se usan todos los días para tomar decisiones críticas:

- Diagnóstico médico
- Comercio de valores
- Previsión energética
- Vigilancia
- Recomendaciones
- Comportamiento de compra
- Marketing online
- Entre otros.



# Algoritmos Aprendizaje Supervisado

Algorithm	Prediction Speed	Training Speed	Memory Usage	Required Tuning	General Assessment
Logistic Regression (and Linear SVM)	Fast	Fast	Small	Minimal	Good for small problems with linear decision boundaries
Decision Trees	Fast	Fast	Small	Some	Good generalist, but prone to overfitting
(Nonlinear) SVM (and Logistic Regression)	Slow	Slow	Medium	Some	Good for many binary problems, and handles high-dimensional data well
Nearest Neighbor	Moderate	Minimal	Medium	Minimal	Lower accuracy, but easy to use and interpret
Naïve Bayes	Fast	Fast	Medium	Some	Widely used for text, including spam filtering
Ensembles	Moderate	Slow	Varies	Some	High accuracy and good performance for small- to medium-sized datasets
Neural Network	Moderate	Slow	Medium to Large	Lots	Popular for classification, compression, recognition, and forecasting