



Toma de decisiones con Aprendizaje de Máquina

EDGAR.SANCHEZ@LOGICSTUDIO.NET
[@EDGARSANCHEZ](https://twitter.com/EDGARSANCHEZ)
[EC.LINKEDIN.COM/IN/EDGARSANCHEZ](https://www.linkedin.com/in/edgarsanchez)

logicstudio
AGILE BUSINESS
SOLUTIONS

Agenda

- ¿Qué es aprendizaje de máquina?
- Definición del problema
- Proceso de modelamiento
- Ejemplo práctico

¿Qué es aprendizaje de máquina?

Se dice que un programa de computadora aprende de la experiencia **E** con respecto a alguna tarea **T** y alguna medida de rendimiento **R**, si su rendimiento en **T**, medido por **R**, mejora con la experiencia **E**.

Tom Mitchell, 1988.

¿T, E, R?

- Tarea: Clasificar transacciones en legales / fraudulentas
- Experiencia: Histórico de transacciones clasificadas
- Rendimiento: Porcentaje de transacciones bien clasificadas

Detalle importante

El sistema **no** contiene lógica que realiza la clasificación.

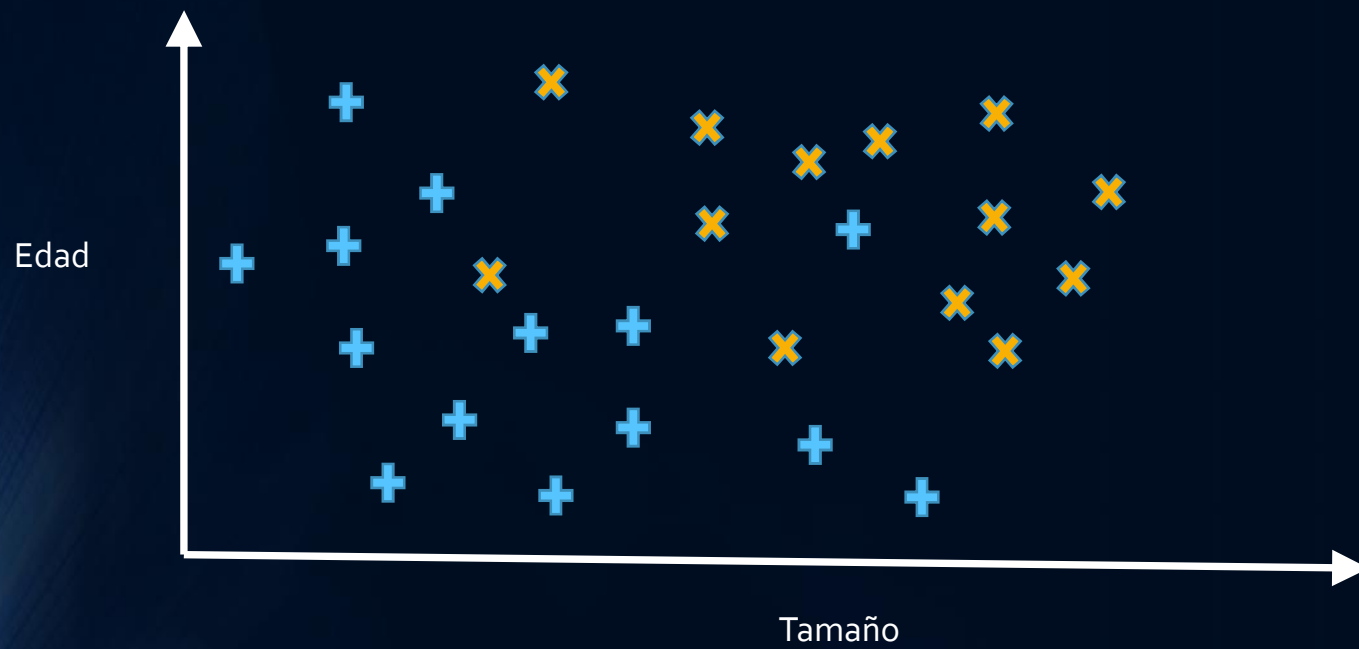
El sistema tiene un algoritmo que:

1. Encuentra un patrón en los elementos clasificados como positivos
2. Usa el patrón para clasificar a nuevos elementos
3. Evalúa la precisión de esta clasificación

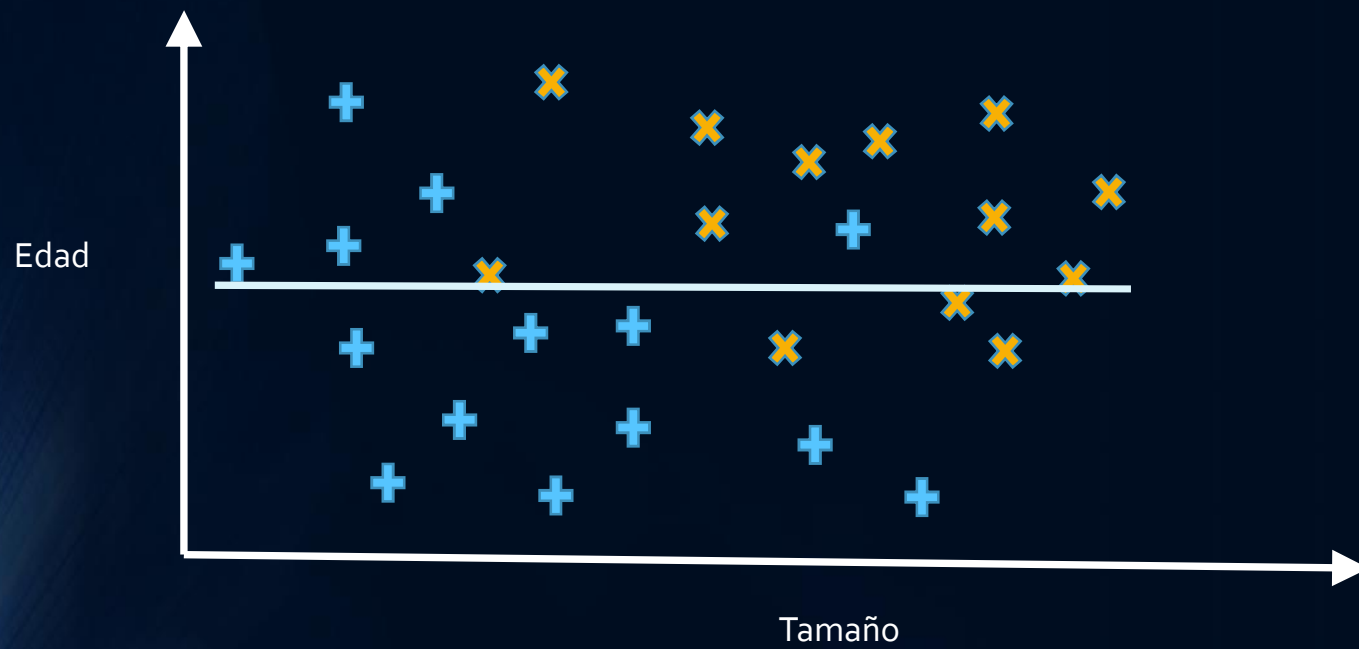
Este proceso se repite, mejorando la precisión paulatinamente.

¡El sistema **aprende** a clasificar elementos cada vez mejor!

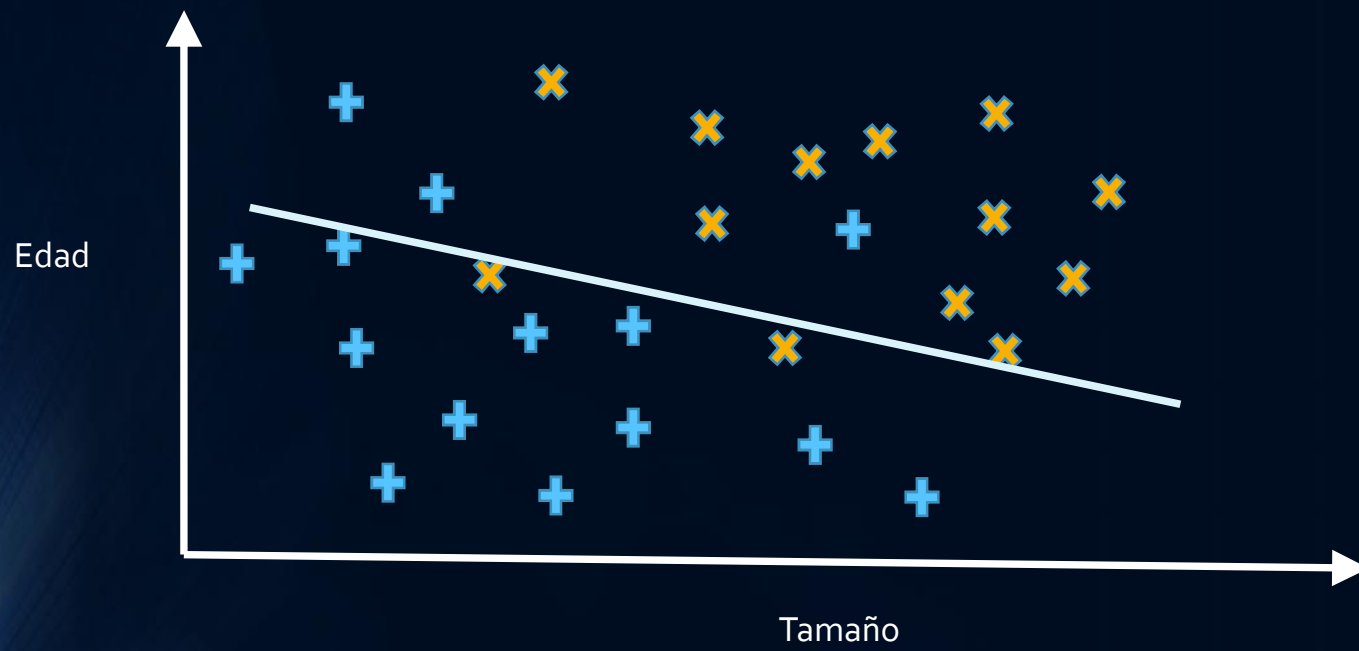
Ejemplo Sencillo



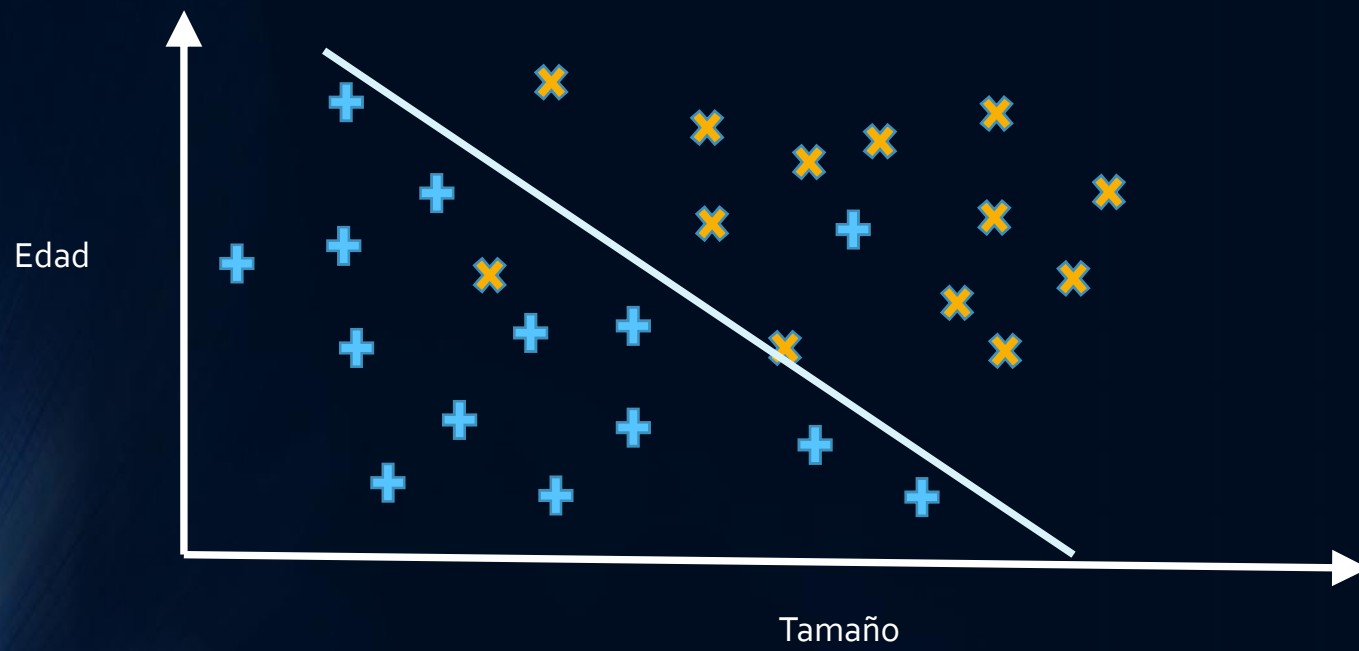
Ejemplo Sencillo



Ejemplo Sencillo



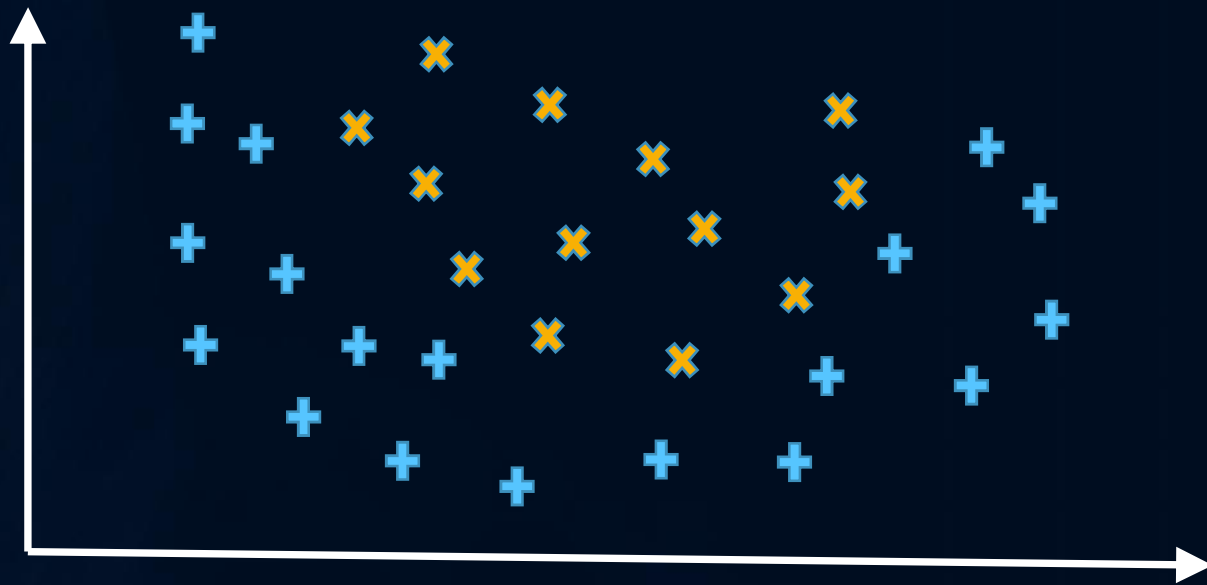
Ejemplo Sencillo



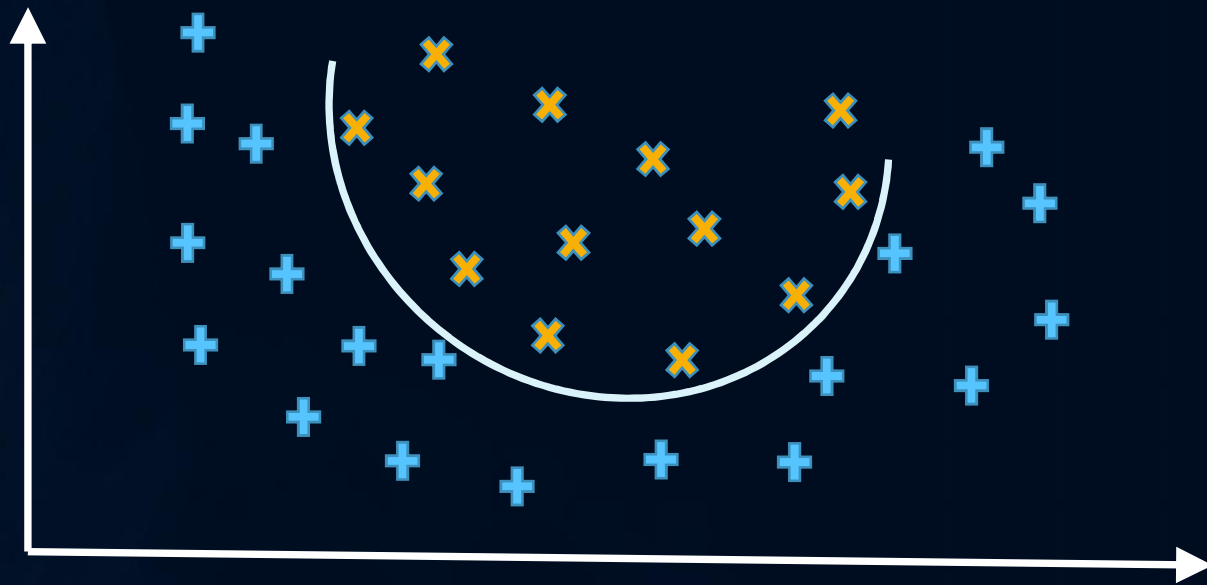
Clasificación de tumores

- Tamaño
- Edad
- Color
- Dureza
- Paredes celulares
- ...

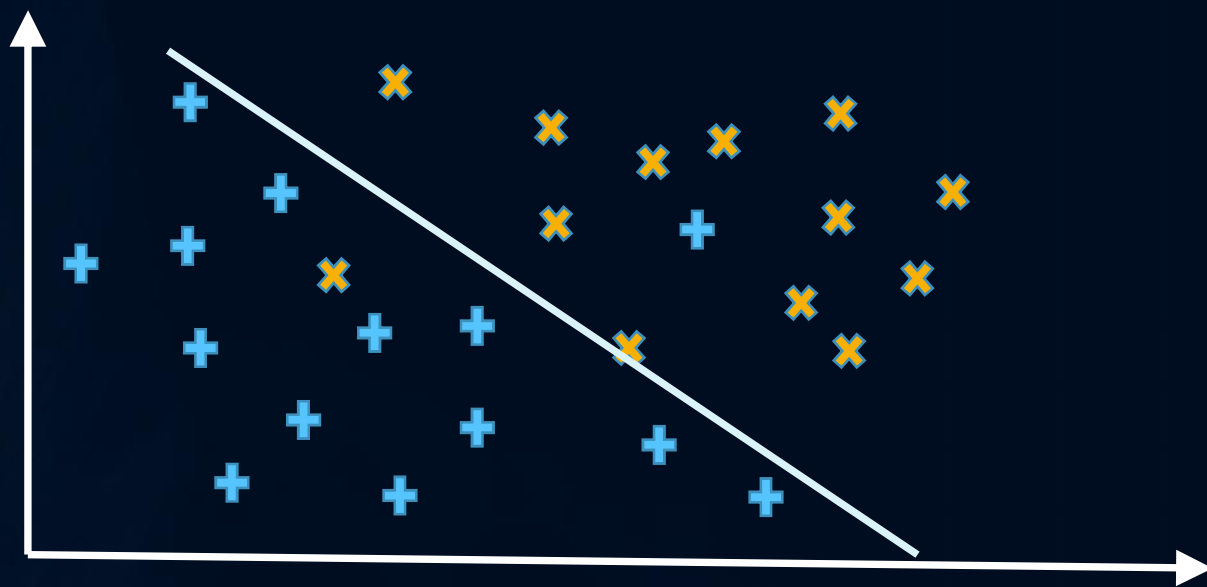
No siempre es tan fácil



No siempre es tan fácil

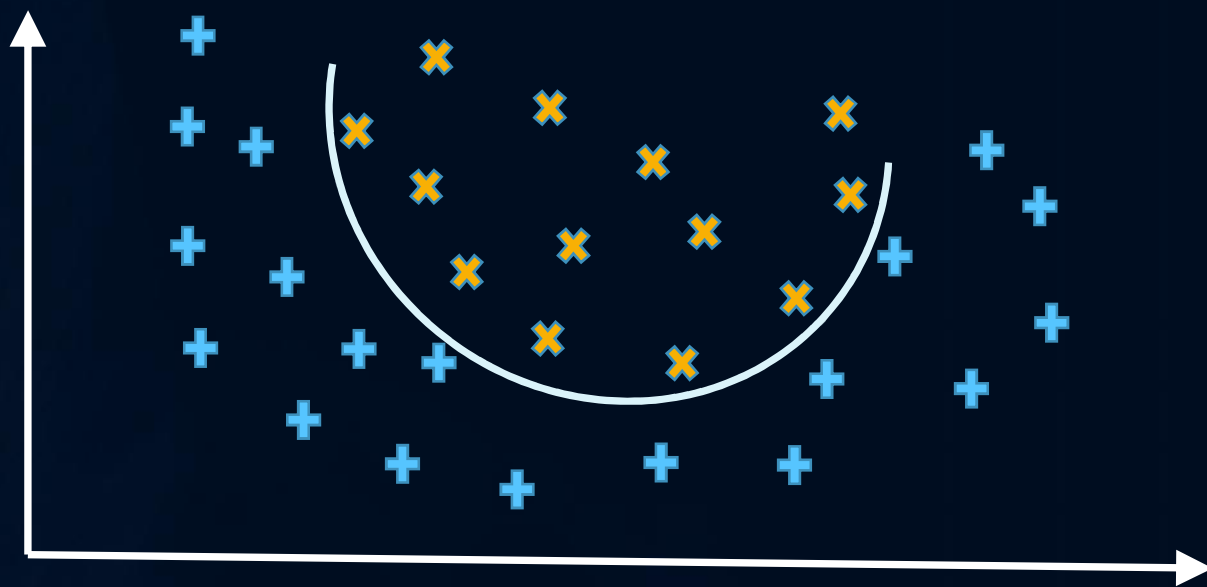


Modelo Lineal



$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Modelo cuadrático



$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

Regresión Logística

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^{(i)}), y^{(i)})$$

$$= -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)})) \right]$$

Problemas prácticos

- ¿Qué parámetros son relevantes?
- ¿De dónde se obtienen los datos iniciales?
- Limpieza, homologación, regularización

¡Estas actividades toman, en promedio, el **70% del esfuerzo!**

MÁS Problemas prácticos

- El volumen de información puede ser alto
 - Usar un producto de Big Data
- El modelo se debe correr múltiples veces sobre esa información
 - El proceso de descubrimiento/afinamiento del modelo es iterativo y de larga duración

Pero la evaluación de nuevos casos puntuales es rápida 😊.

Principales tipos de problema

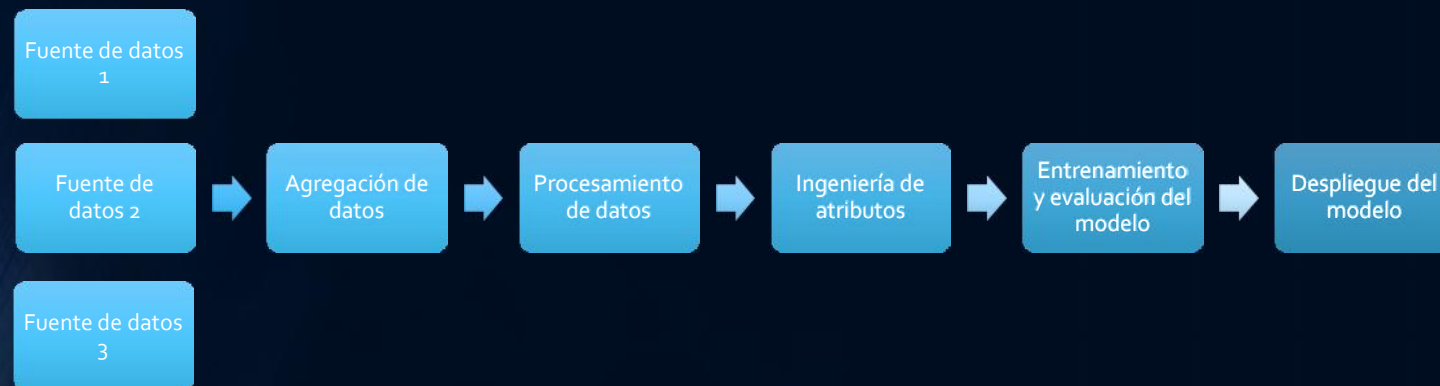
- Detección de anomalías
- Clasificación
- Agrupación (clustering)
- Regresión

Cada problema tiene múltiples modelos candidatos.

Definición del problema específico

- Detección de fraude en transacciones en línea
- Una de las aplicaciones más tempranas del aprendizaje de máquina
- Un problema de clasificación binaria
- Población de las clases (muy) desbalanceada

Proceso de modelamiento



Ejemplo práctico

AZURE ML