



¿Qué es el Aprendizaje Automático?

Las máquinas aprenden patrones de los datos para **predecir** y **clasificar** sin reglas explícitas

El Aprendizaje Automático en Acción



Recomendaciones Personalizadas

Netflix analiza tu historial para sugerir el próximo contenido que amarás



Protección Financiera

Los bancos predicen comportamientos de pago y detectan fraudes en tiempo real



Tres Paradigmas de Aprendizaje



El Niño Aprendiz Obediente (Supervisado)

Aquí, el modelo aprende como un niño al que le das ejemplos claros y corregimos si se equivoca. Le enseñamos: "Esto es una vaca, esto es un perro".

Ejemplo: Como cuando analizamos qué clientes pagaron un préstamo y cuáles no, para predecir si un nuevo cliente será "responsable" o "travieso" con sus pagos.



El Pequeño Explorador Curioso (No Supervisado)

Dejamos al modelo solo con sus juguetes (datos sin etiquetas) y ¡sorpresa! Empieza a agruparlos por colores o formas por sí mismo, sin que nadie le diga cómo. Es el descubrimiento puro.

Ejemplo: Imagina que el modelo descubre "grupos" de clientes que adoran los mismos productos, como si fueran pandillas de amiguitos con gustos similares. ¡Ideal para segmentar el mercado sin que se lo pidamos!

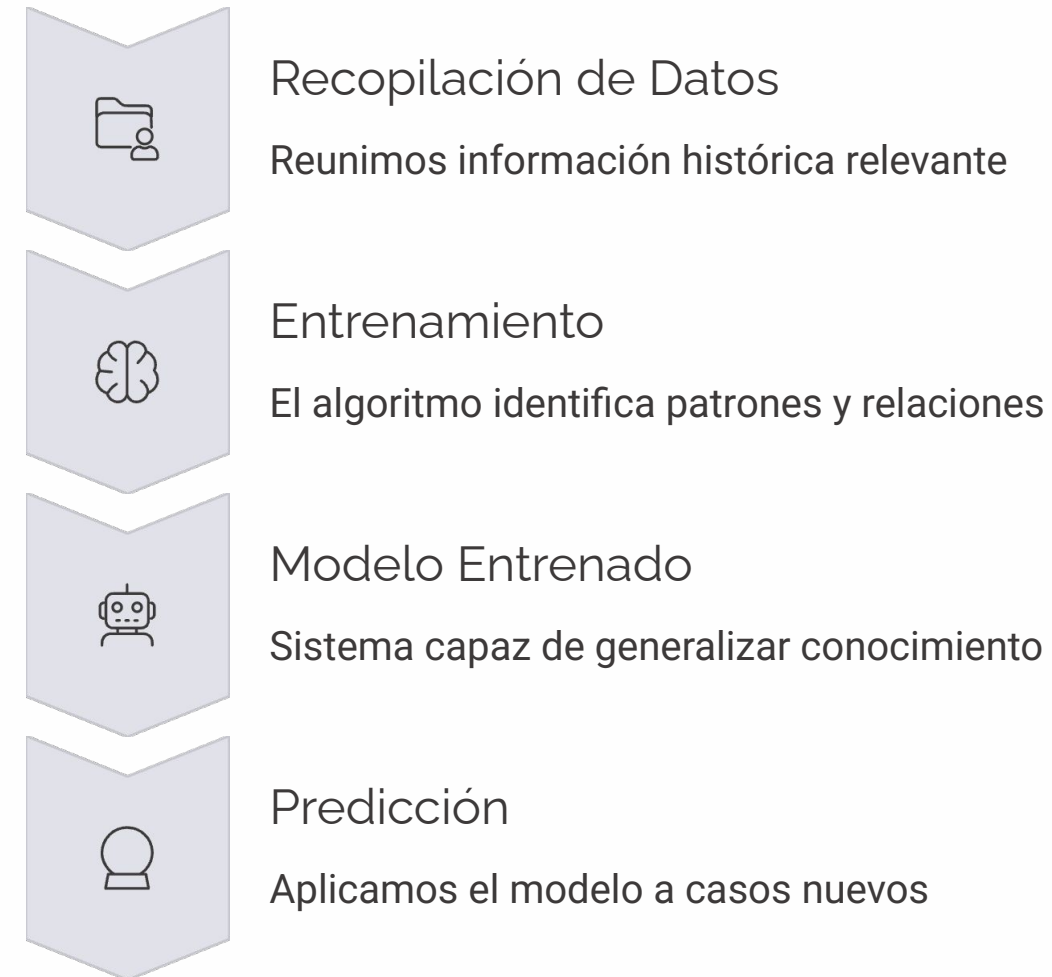


El Genio de la Prueba y Error (Por Refuerzo)

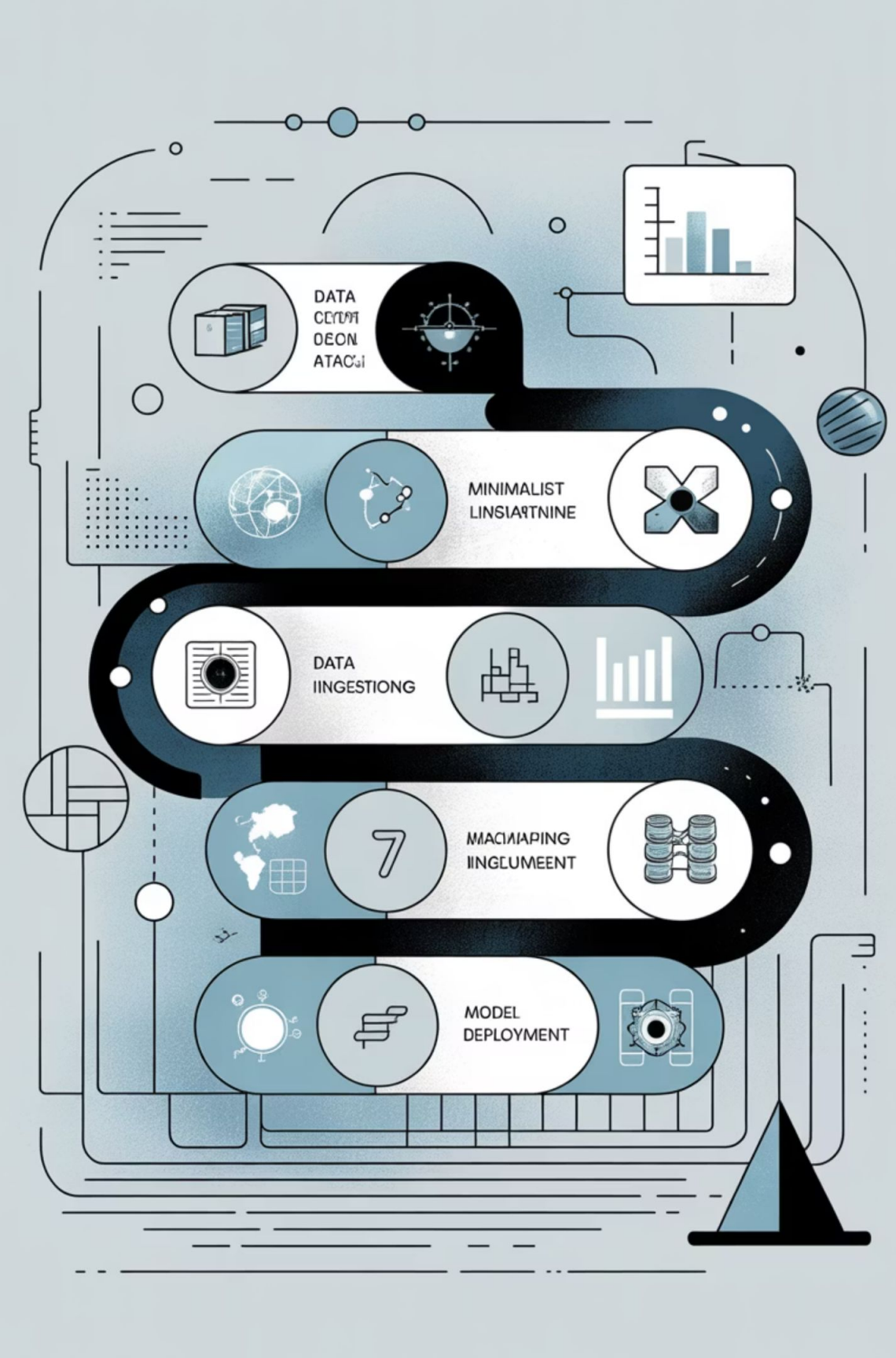
Piensa en el modelo como un niño aprendiendo a montar en bicicleta: se cae (recibe una "penalización"), se levanta y cuando avanza un poquito recibe un "¡bien hecho!" (una recompensa). Con cada intento, mejora su equilibrio.

Ejemplo: Similar a cómo los vehículos autónomos "aprenden" a conducir. Cada decisión correcta es una "recompensa" y cada error (¡sin accidentes, por favor!) es una lección para mejorar con la experiencia.

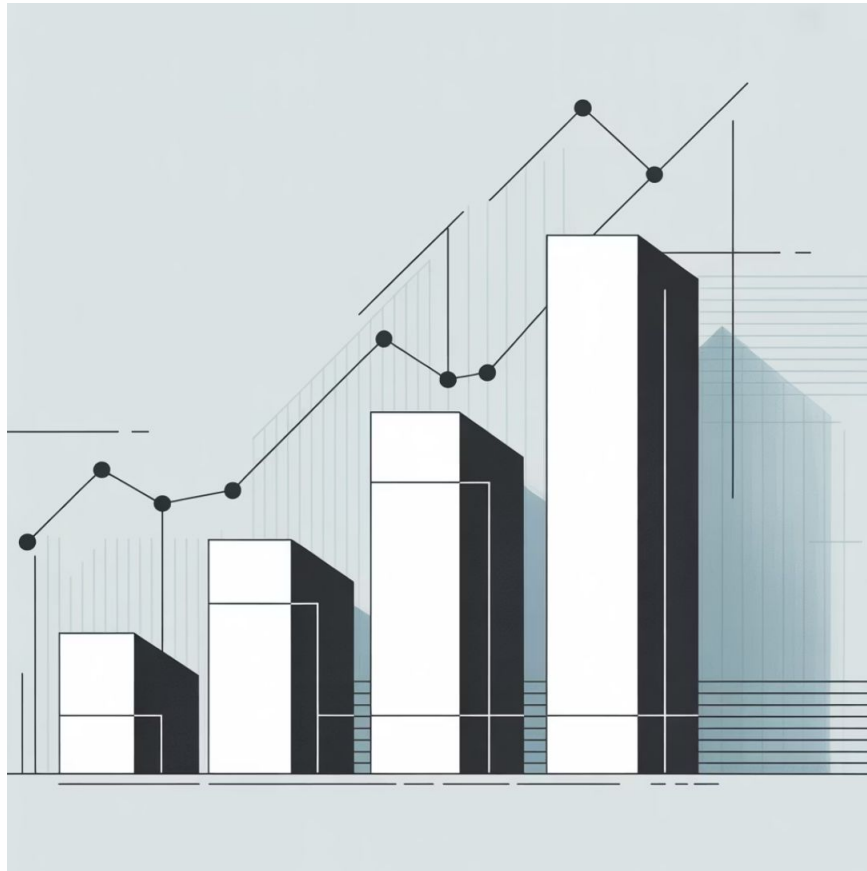
El Proceso de Construcción de un Modelo



Le mostramos casos históricos de clientes y el modelo aprende a predecir el comportamiento de nuevos solicitantes



Dos Tipos de Problemas Fundamentales

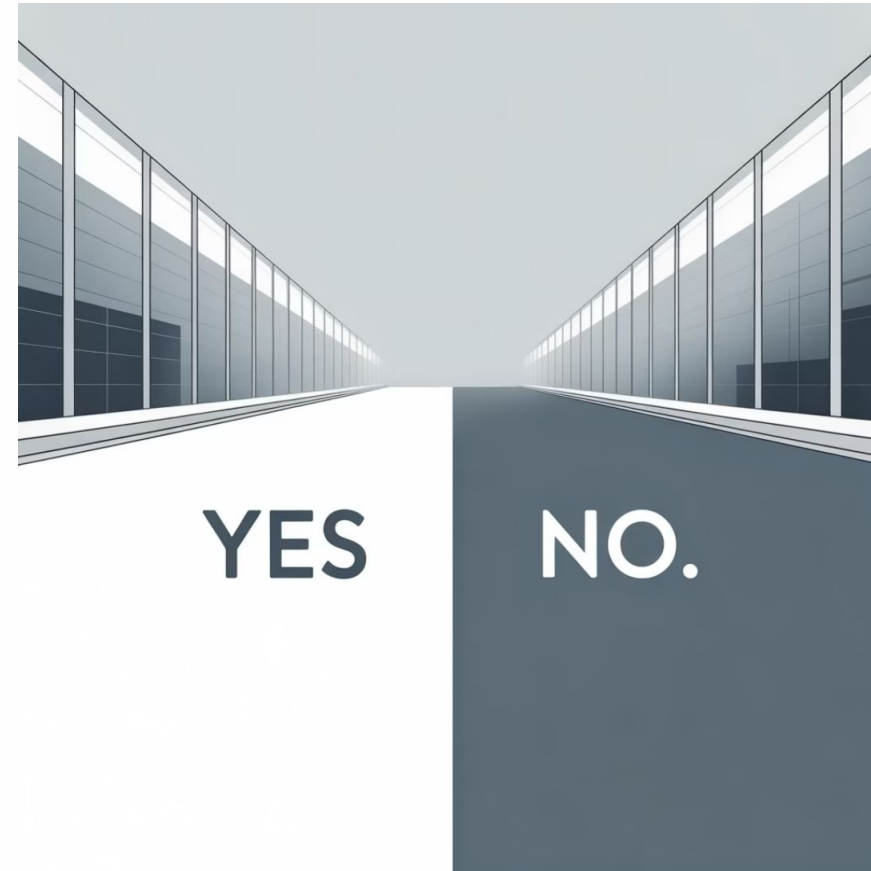


Regresión

Predice **valores continuos**

- Precio de una vivienda
- Temperatura mañana
- Ventas del próximo mes

 Resultado: números en un rango

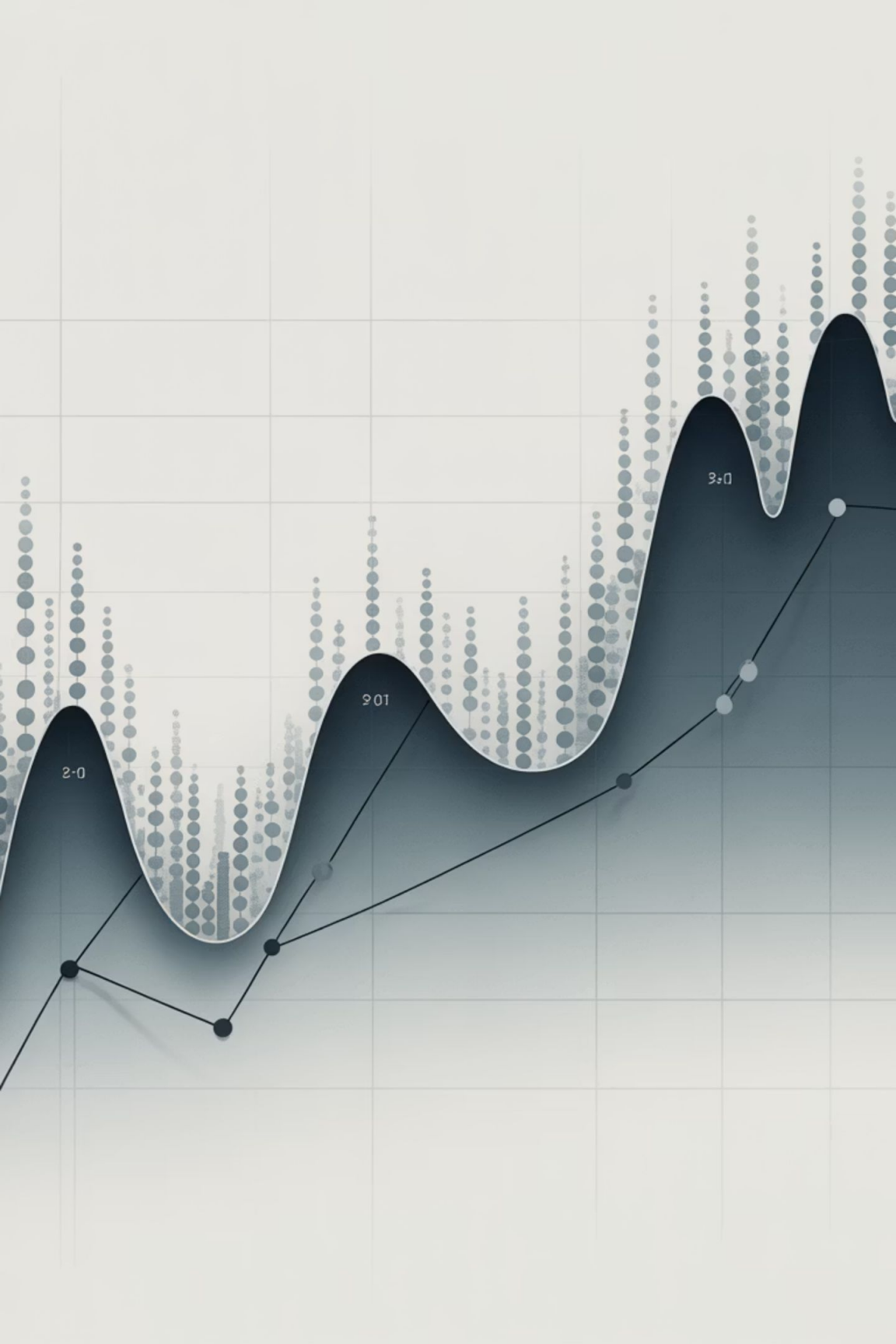


Clasificación

Predice **categorías discretas**

- Pagará / No pagará
- Spam / No spam
- Tipo de cliente (A, B, C)

 Resultado: etiquetas o clases



Regresión Lineal: Encontrando la Relación

Dibuja la **línea recta** que mejor describe la relación entre variables

La Intuición

A mayor ingreso del solicitante, mayor probabilidad de aprobación del préstamo

El Objetivo

Minimizar la distancia entre las predicciones y los valores reales observados

Regresión Lineal: Fundamentos Matemáticos

La Ecuación

$$y = \beta_0 + \beta_1 x + \epsilon$$

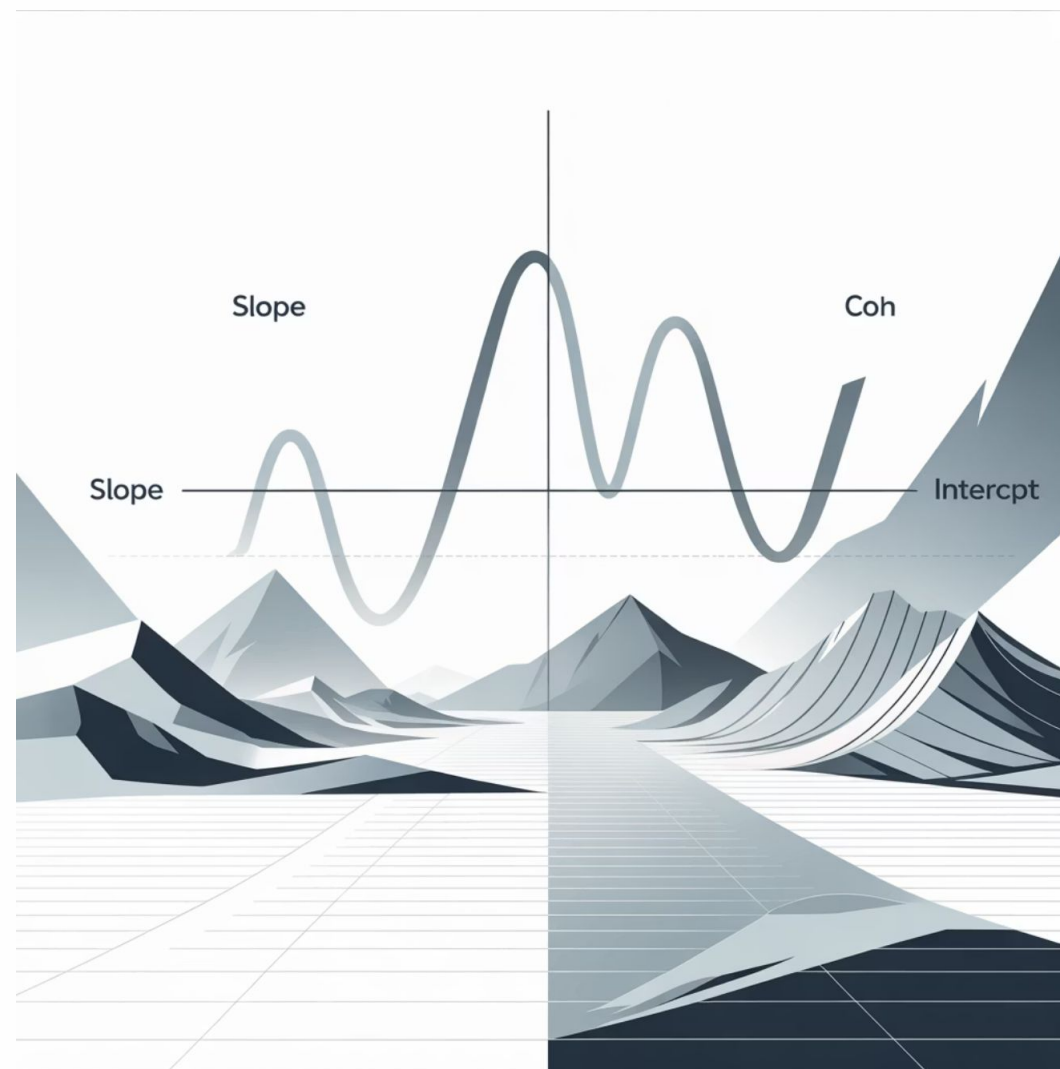
y: variable dependiente (objetivo)

x: variable independiente (predictor)

β_0 : intercepto (valor cuando $x=0$)

β_1 : pendiente (cambio en y por unidad de x)

ϵ : error residual



Interpretación de Coeficientes

Ejemplo: Predicción de salario basado en años de experiencia

Si $\beta_1 = 3.500\text{€}$, cada año adicional de experiencia aumenta el salario esperado en 3.500€

Minimizando el Error: La Línea de Mejor Ajuste



La meta de la Regresión Lineal es encontrar la línea que **minimice la suma de los errores al cuadrado**

¿Qué es el Error?

Es la **distancia vertical** entre un punto de datos real y la línea de predicción. Cuanto menor el error, mejor el ajuste.

Error Cuadrático

Elevamos cada error al cuadrado antes de sumarlos. Esto penaliza los errores grandes y asegura que errores positivos y negativos no se cancelen.

La "Mejor" Línea

El algoritmo prueba diferentes líneas y elige la que resulta en la menor **Suma de Errores Cuadráticos (SSE)** o **Error Cuadrático Medio (MSE)**.

Interpretando tu Modelo de Regresión

Una vez que el modelo "dibuja" su línea, necesitamos entender qué tan bueno es y si cada factor contribuye de forma significativa.



Coeficiente de Determinación (R^2)

Indica qué porcentaje de la variación en tu resultado (ej. el precio de una casa) es explicado por tu modelo. Un 80% significa que el 80% de los cambios en el precio se deben a las variables que usaste, y el 20% a otros factores no incluidos.



Error Estándar (de cada β)

Mide la precisión con la que tu modelo estima la influencia de cada factor (ej. cuánto impactan los metros cuadrados en el precio). Un error pequeño significa que la estimación es muy fiable y cercana al valor real.



Valor P (P-value)

Nos dice si la relación entre un factor y el resultado es estadísticamente significativa. Si el P-value es muy bajo (normalmente menos de 0.05), podemos confiar en que ese factor realmente importa y no es fruto del azar.



Regresión Logística: De Líneas a Probabilidades

Transforma valores continuos en **probabilidades** entre 0 y 1

0

1 Combinación Lineal

Calculamos un valor z a partir de las variables predictoras

0

2 Función Sigmoid

Convertimos z en probabilidad usando σ
 $\sigma(z) = 1/(1+e^{-z})$

0

3 Clasificación

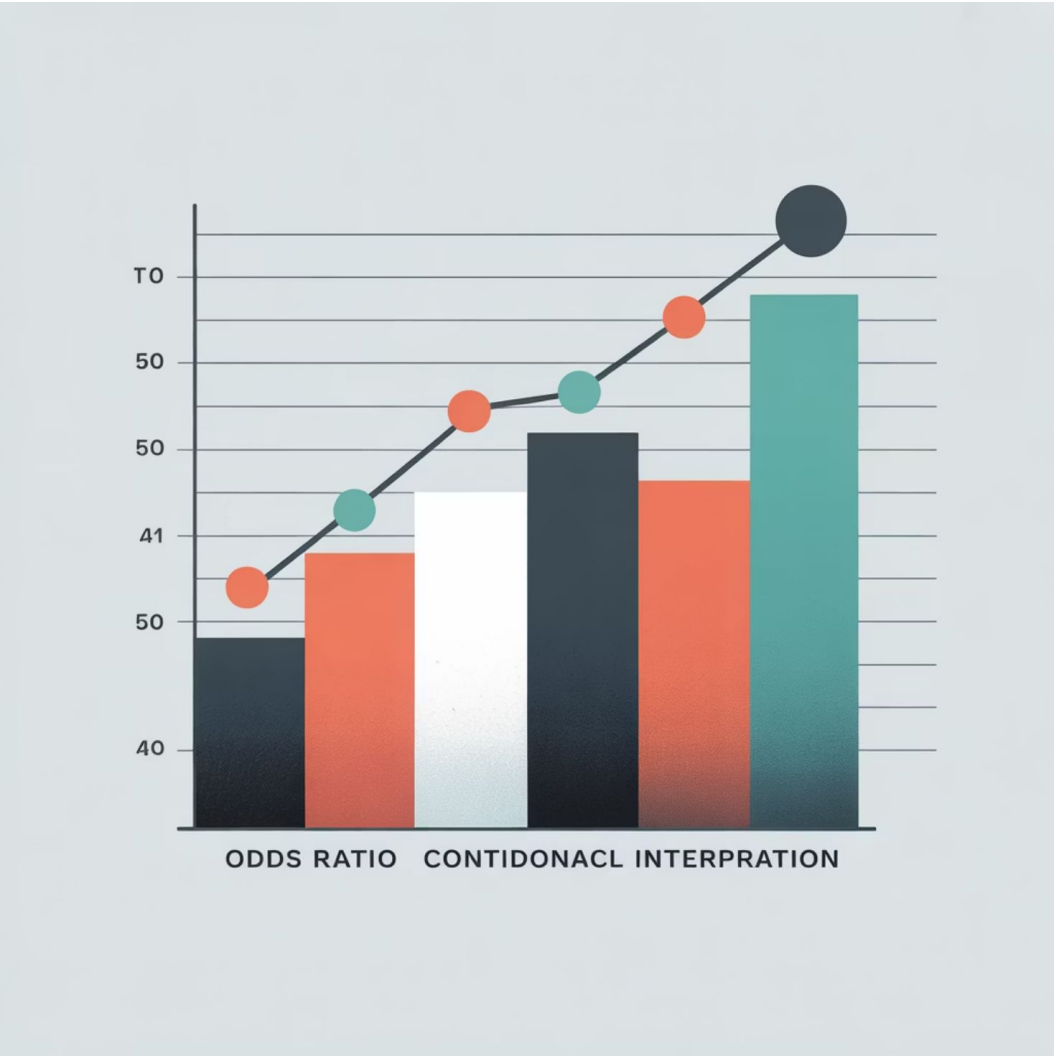
Si $P(\text{pago}) = 0.85 \rightarrow$ clasificamos como 'Pagador'

Interpretación Estadística de la Regresión Logística

Odds Ratio

$$Odds = \frac{P}{1 - P}$$

Representa cuántas veces más probable es el evento positivo vs. el negativo



Coeficientes como Log-Odds

Un coeficiente $\beta_1 = 0.5$ significa que por cada unidad de aumento en x_1 , el log-odds aumenta en 0.5

En términos prácticos: $\exp(0.5) = 1.65 \rightarrow$ las probabilidades se multiplican por 1.65

Pruebas de Significancia

Test de Wald y valores p nos indican qué variables son realmente predictivas

Intervalos de confianza para cuantificar la incertidumbre de nuestras estimaciones

❑ **Poder del análisis:** No solo predecimos, también entendemos el *impacto cuantificado* de cada variable en la probabilidad del resultado

Ejemplo Práctico: Aprobación de Préstamos

Aplicaremos la Regresión Logística para predecir la probabilidad de que un cliente apruebe un préstamo basándonos en algunas variables.

Variables y Coeficientes (Betas)

Supongamos que nuestro modelo ha calculado los siguientes coeficientes:

Intercepto (β_0): -2.5

Ingreso Mensual (β_1): 0.0001

Puntuación Crediticia (β_2): 0.02

Edad (β_3): 0.05

La ecuación subyacente para el log-odds sería:

$$\text{log-odds} = -2.5 + 0.0001 \text{ Ingreso} + 0.02 \text{ Puntuación} + 0.05 \text{ Edad}$$

Odds Ratios e Interpretación

Transformamos los coeficientes para entender su impacto directo en las probabilidades:

Por cada euro adicional de ingreso mensual, las probabilidades de aprobación del préstamo aumentan en un 0.01%.

Un aumento de un punto en la puntuación crediticia multiplica las probabilidades de aprobación por 1.02 (un aumento del 2%).

Por cada año adicional de edad, las probabilidades de aprobación del préstamo se multiplican por 1.05 (un aumento del 5%).

Ingreso Mensual: Odds Ratio = $\exp(0.0001) \approx 1.0001$

Puntuación Crediticia: Odds Ratio = $\exp(0.02) \approx 1.02$

Edad: Odds Ratio = $\exp(0.05) \approx 1.05$

Estos Odds Ratios nos permiten cuantificar el efecto de cada variable en las probabilidades de que el evento (aprobación del préstamo) ocurra, manteniendo otras variables constantes.



K-Nearest Neighbours: La Sabiduría de los Vecinos

¿Cómo Funciona?

1. Elegimos un valor de k (ej: k=5)
2. Identificamos los 5 vecinos más cercanos
3. Clasificamos por mayoría de votos

Distancia más común: Euclidiana

✓ Ventajas

- Extremadamente simple
- No requiere entrenamiento
- Efectivo en muchos casos

✗ Limitaciones

- Lento con grandes volúmenes
- Sensible a la escala
- Necesita normalización

KNN vs. K-Means: Aclaración Importante

K-Nearest Neighbors (KNN)

Es un algoritmo de **clasificación** (o regresión), no de agrupamiento.

Es un algoritmo "**perezoso**" (lazy learner) porque no tiene una fase de entrenamiento explícita.

No entrena un modelo; simplemente almacena todos los datos de entrenamiento.

Para clasificar un nuevo punto, busca a sus **K vecinos más cercanos** y asigna la clase más frecuente entre ellos (por "voto de mayoría").

No utiliza centroides que se mueven o redefinen para la clasificación.

K-Means

Es un algoritmo de **agrupamiento** (clustering), no de clasificación.

Es un algoritmo "**ansioso**" (eager learner) porque sí tiene una fase de entrenamiento.

Crea K grupos en los datos sin etiquetas preexistentes.

Funciona **iterativamente moviendo centroides** hasta que cada punto pertenece al grupo cuyo centroide es el más cercano.

- Su objetivo es encontrar los grupos óptimos y sus centros.



Naive Bayes: El Poder de las Probabilidades

Calcula probabilidades condicionales asumiendo **independencia** entre características

Teorema de Bayes

$$P(C | X) = \frac{P(X | C) P(C)}{P(X)}$$

- $P(C|X)$: probabilidad de clase dado X
- $P(X|C)$: verosimilitud
- $P(C)$: probabilidad a priori

Ejemplo: Filtro de Spam

Correo contiene: "gratis" + "click aquí"

$P(\text{spam} | \text{"gratis"}) \times P(\text{spam} | \text{"click"}) \rightarrow$ **Alta probabilidad de spam**



Especialmente efectivo para clasificación de texto y diagnósticos rápidos

Comparativa de Algoritmos

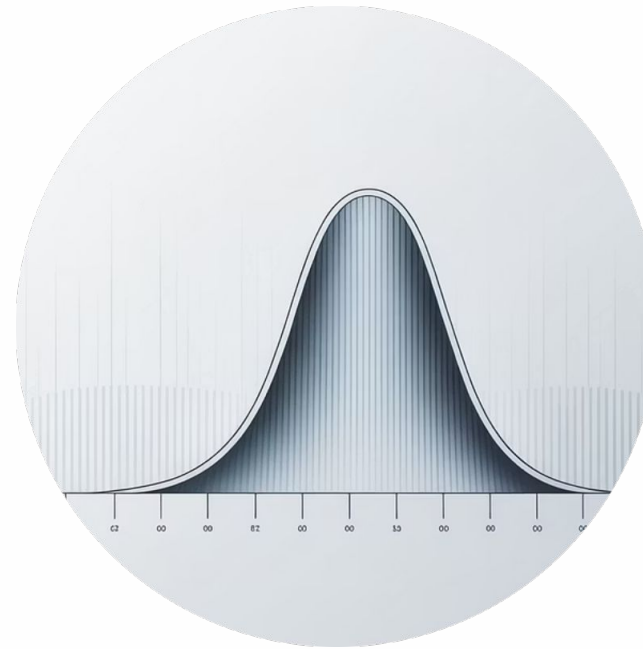


Regresión Lineal

Tipo: Regresión

Ideal para: Relaciones lineales simples y continuas

Ejemplo: Predicción de precios

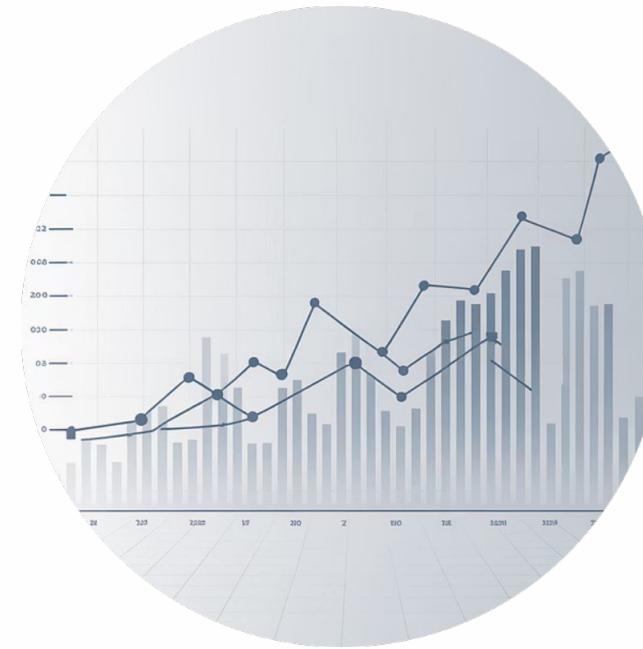


Regresión Logística

Tipo: Clasificación binaria

Ideal para: Probabilidades e interpretabilidad

Ejemplo: Riesgo crediticio



K-Nearest Neighbours

Tipo: Clasificación/Regresión

Ideal para: Patrones locales de vecindad

Ejemplo: Sistemas de recomendación



Naive Bayes

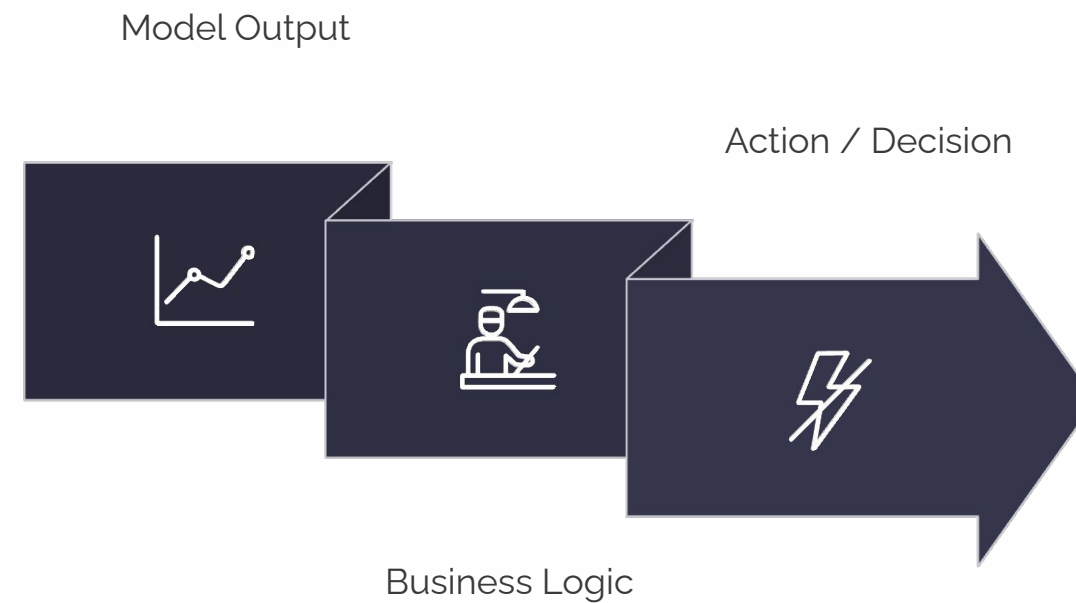
Tipo: Clasificación

Ideal para: Texto y datos de alta dimensión

Ejemplo: Filtrado de spam

Del Modelo a la Acción: La Lógica de Negocio

Un modelo de Machine Learning puede predecir una probabilidad, pero sin una estrategia clara, este número por sí solo carece de valor. Es crucial aplicar la lógica de negocio para transformar predicciones en decisiones accionables y estratégicas.



Esta integración de la analítica avanzada con el conocimiento estratégico de la empresa es lo que realmente impulsa el impacto y la toma de decisiones informadas, y es donde el valor real del Machine Learning se materializa.



Mini Práctica: Dataset de Clientes

Variables del Dataset

Edad del cliente

Ingreso mensual

Historial crediticio

Resultado: Pagó / No pagó

Objetivo del Ejercicio

Entrenar y comparar los 4 modelos presentados

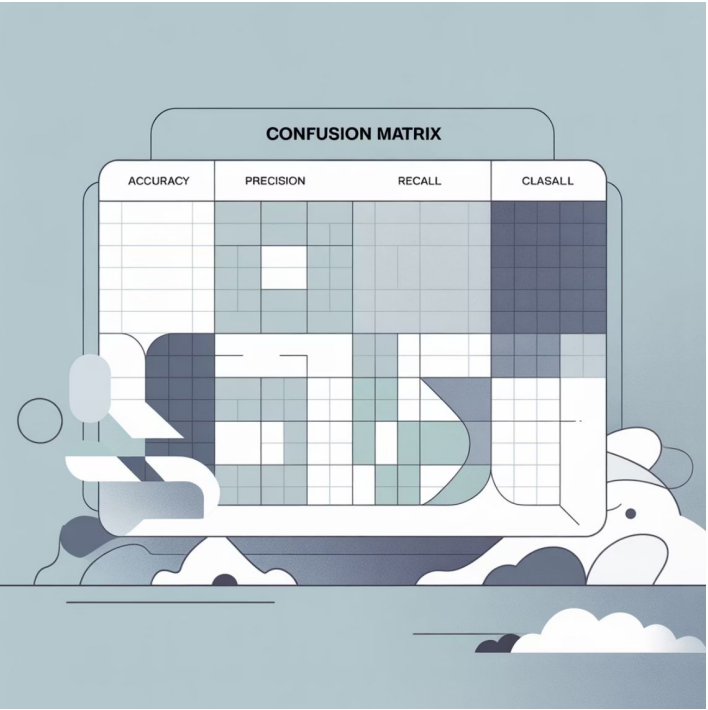
Evaluar cuál funciona mejor para este caso específico

Interpretar los resultados y coeficientes



Próximamente: Implementación práctica con código Python y análisis completo de resultados

Evaluación de Modelos: Midiendo el Éxito



Métricas para Regresión

1

MAE
Error Absoluto Medio: promedio de errores en unidades originales

2

RMSE
Raíz del Error Cuadrático Medio: penaliza errores grandes

3

R^2
Coeficiente de determinación: % de varianza explicada

Métricas para Clasificación

1

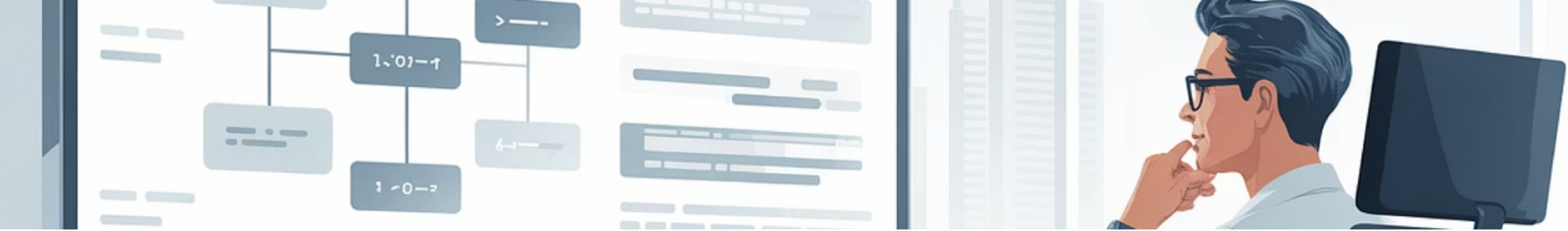
Accuracy
Porcentaje de predicciones correctas

2

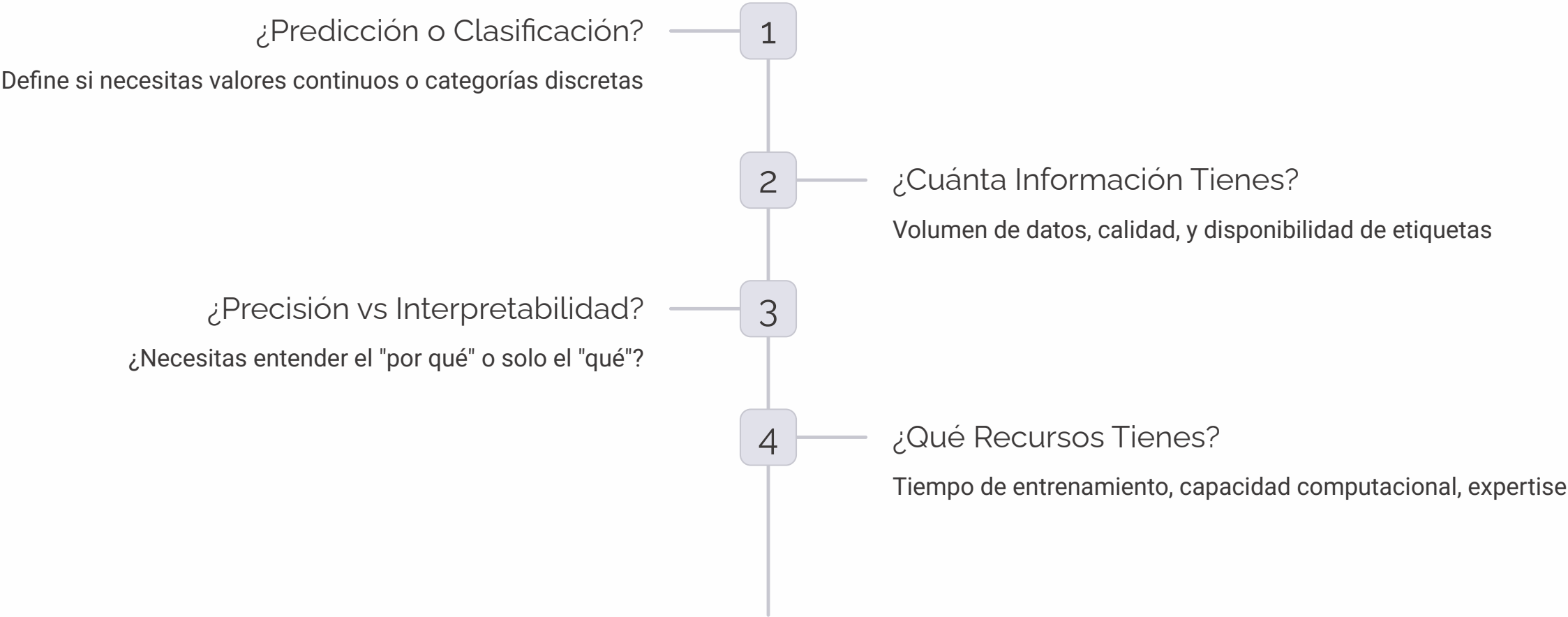
Matriz de Confusión
VP, VN, FP, FN: análisis detallado de errores

3

AUC-ROC
Área bajo la curva: capacidad discriminativa del modelo



Guía para Elegir tu Modelo



¿Preguntas?

Ahora es el momento de resolver cualquier duda que pueda tener.

¡Gracias por su Atención!

Para más información o consultas, no dude en contactarnos:

diego.tauziet@gmail.com

