



De mirar el pasado a
anticipar el futuro



Machine Learning con Python para la Toma de Decisiones Empresariales

Jorge Israel Frometa Moya

Clasificación: La lógica del "Sí" o el "No".



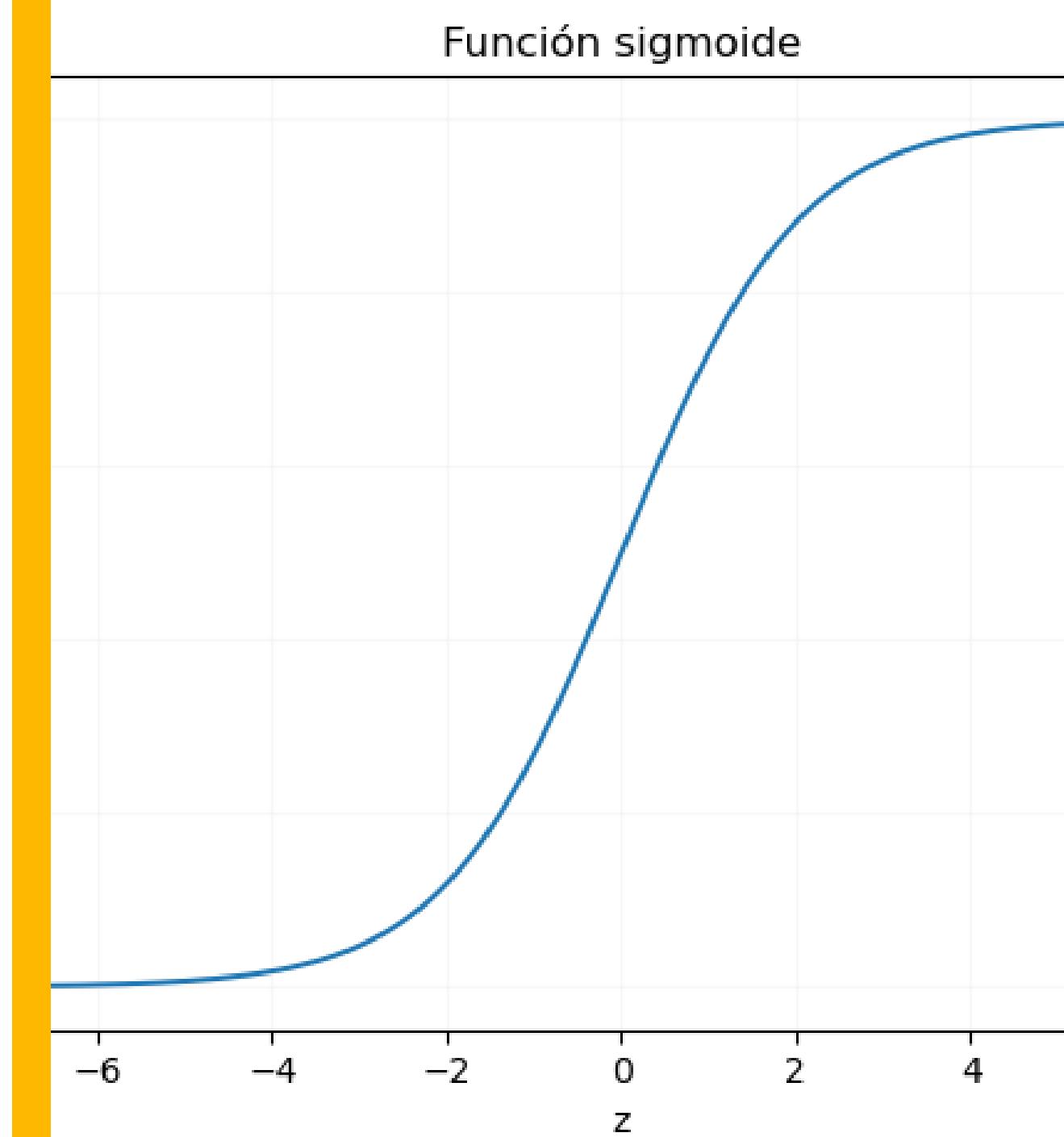
- Prediciendo el comportamiento del cliente con Regresión Logística.

No predecimos un valor,
predecimos una **probabilidad**.

Función sigmoide

¿Por qué no usar Regresión Lineal aquí?

- El problema: Si usamos una línea para predecir si alguien compra (1) o no (0), la línea podría darnos valores como 1.5 o -0.2, lo cual no tiene sentido en negocio.
- La solución: La Función Sigmoide. Esta función "aplasta" cualquier número para que esté entre 0 y 1.
- Interpretación: "Hay un 85% de probabilidad de que este cliente sea Premium".



¿CONFIAR O NO CONFIAR EN EL MODELO?



La Frontera de Decisión (Threshold)

- Concepto: Por defecto, si la probabilidad es > 0.5 : "SÍ".

AJUSTE ESTRATÉGICO POR INDUSTRIA:

BANCA (Créditos):

Umbral: 0.7

Filosofía: "Mejor rechazar bueno que aprobar malo"

Costo Falso Positivo: Pérdida por impago (\$10K)

Costo Falso Negativo: Interés no cobrado (\$500)

¿CONFIAR O NO CONFIAR EN EL MODELO?



AJUSTE ESTRATÉGICO POR INDUSTRIA:

E-COMMERCE (Marketing):

Umbral: 0.3

Filosofía: "Mejor contactar de más que perder venta"

Costo Falso Positivo: Email no abierto (\$0.01)

Costo Falso Negativo: Venta perdida (\$50)

SALUD (Diagnóstico):

Umbral: 0.1

Filosofía: "Mejor falso positivo que muerte"

Costo Falso Positivo: Prueba adicional (\$100)

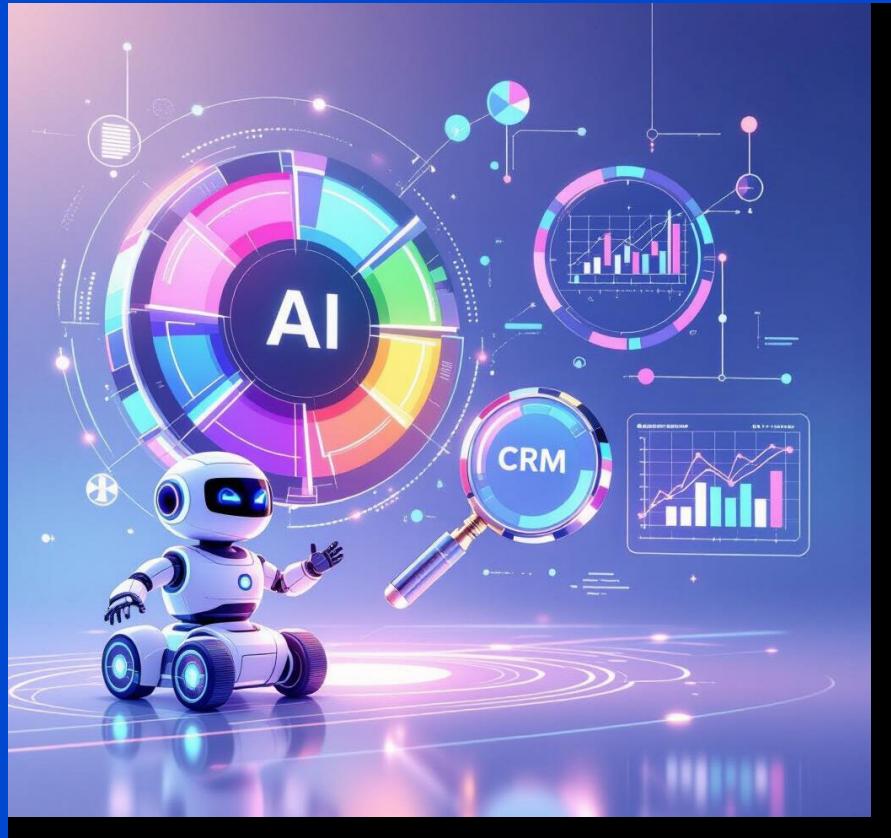
Costo Falso Negativo: Vida en riesgo (incalculable)

Introducción a la Matriz de Confusión

- No todos los errores cuestan lo mismo.
- Falso Positivo: Predecir que alguien compra y no lo hace (gastamos marketing en vano).
- Falso Negativo: Predecir que no compra y sí lo hacía (perdemos una venta).

VALORES PREDICCIÓN	Verdaderos positivos	Falsos Positivos
VALORES REALES	Falsos Negativos	Verdaderos Negativos

7 PECADOS CAPITALES DE LA CLASIFICACIÓN



1. RIESGO DE UMBRAL FIJO

"Usamos 0.5 porque es el default"

→ REALIDAD: El negocio cambia, el umbral no

2. RIESGO DE DESBALANCE

"99% no compran, 1% compran"

→ Modelo aprende: "Decir 'NO' siempre = 99% de acierto"

→ SOLUCIÓN: Oversampling, undersampling, pesos de clase

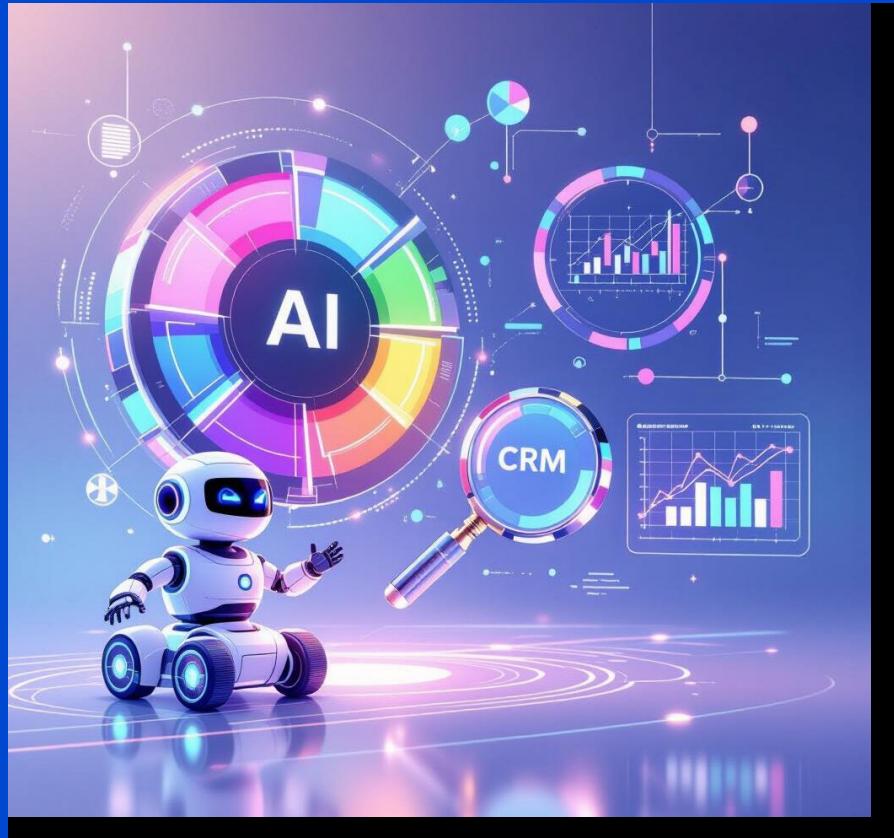
3. RIESGO DE SOBREOPTIMIZAR ACCURACY

"¡Logramos 95% de accuracy!"

→ Pero solo detecta el 10% de los fraudes reales

→ MEJOR MÉTRICA: F1-score, Precision-Recall

7 PECADOS CAPITALES DE LA CLASIFICACIÓN



4. RIESGO DE NO ACTUALIZAR

"El modelo de 2020 aún funciona"

→ REALIDAD: Comportamientos cambian post-pandemia

→ SOLUCIÓN: Monitorización continua, retraining

5. RIESGO DE EXPLICABILIDAD

"La IA dice que no le des crédito"

→ "¿Por qué?" → "No sé, la caja negra lo dice"

→ SOLUCIÓN: Modelos explicables como árboles de decisión

6. RIESGO DE AUTOMATIZACIÓN CIEGA (RIESGO ÉTICO-LEGAL)

"El sistema rechaza automáticamente"

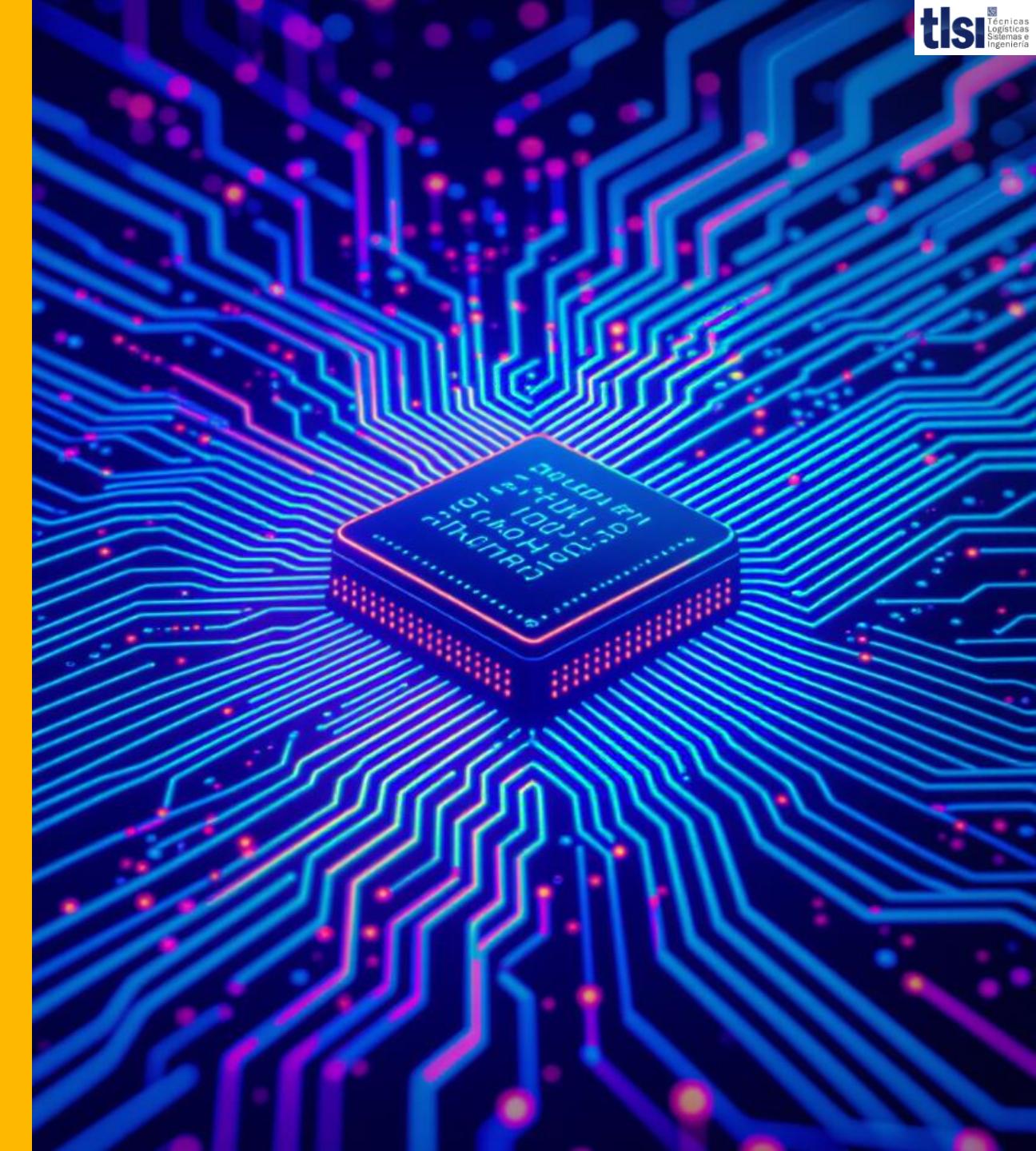
→ Sinapelación humana

→ CASO REAL: Amazon hiring algorithm biased against women

ENTENDIENDO LA “PROBABILIDAD” - NO ES ADIVINACIÓN

NIVELES DE CONFIANZA:

- 0.95-1.00: "Casi seguro" → Actuar inmediatamente
- 0.75-0.94: "Muy probable" → Incluir en campaña
- 0.55-0.74: "Posible" → Monitorear, bajo costo
- 0.45-0.54: "Inciso" → No actuar, recolectar más datos
- 0.00-0.44: "Improbable" → Excluir de acciones



INTERPRETABILIDAD

- Coeficientes interpretados en términos de negocio
- Variables importantes identificadas
- Explicación lista para stakeholders

