



De mirar el pasado a
anticipar el futuro



Machine Learning con Python para la Toma de Decisiones Empresariales

Jorge Israel Frometa Moya

Árboles de Decisión: La lógica de las reglas de negocio.



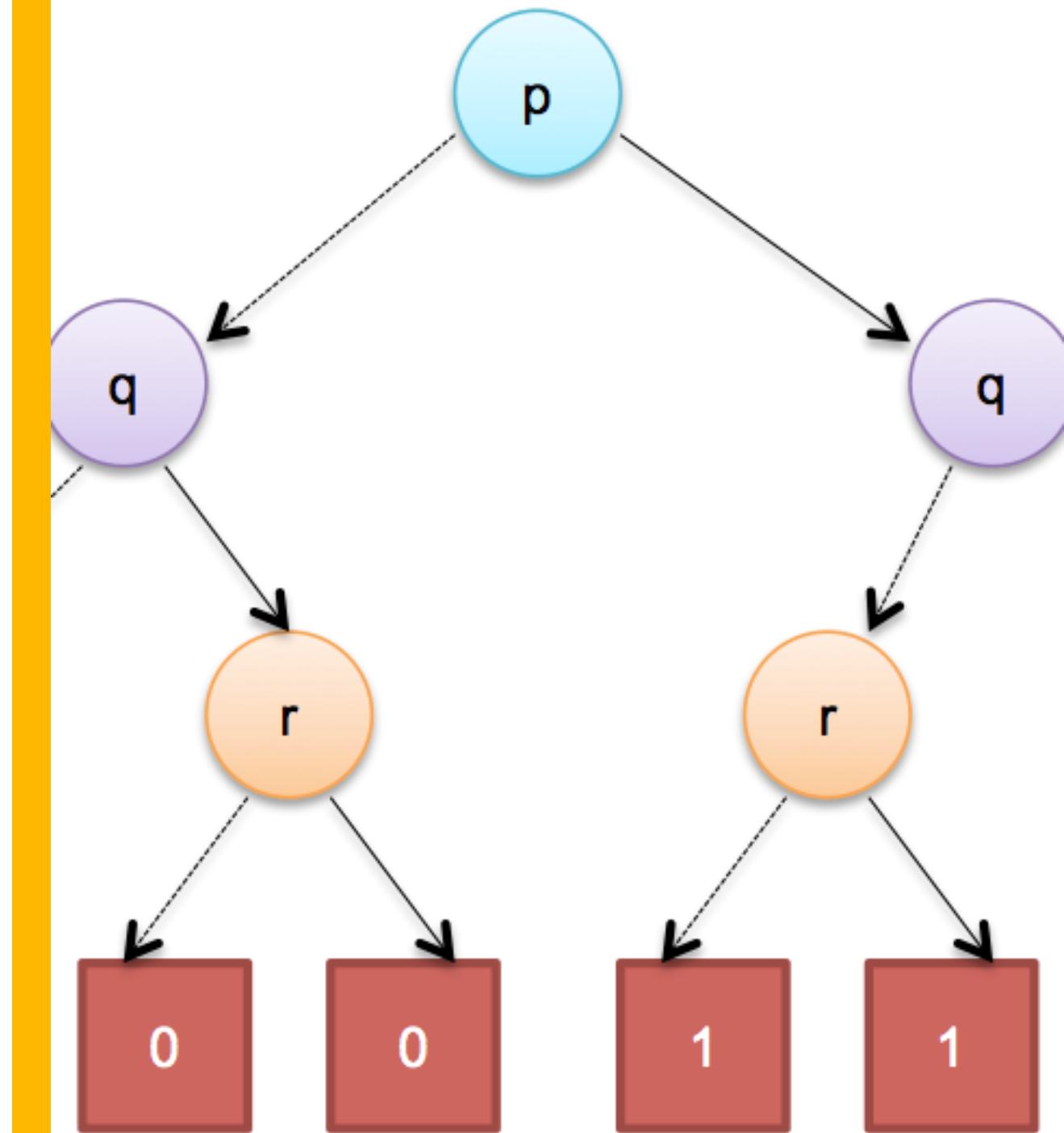
- De cajas negras a modelos transparentes.

No es una fórmula compleja, es un mapa de carreteras para segmentar clientes.

¿Cómo "piensa" un Árbol?

- El Nodo Raíz: La pregunta más importante (la que más separa a los clientes buenos de los malos).
- Las Ramas: Los diferentes caminos según la respuesta.
- Las Hojas: La decisión final (se queda o se va).

Analogía de Negocio: Es como el manual de formación de un comercial: "Primero mira si el cliente es antiguo; si lo es, mira si tiene deudas...".



Ventajas e Inconvenientes



Ventaja (Interpretación): Puedes explicar exactamente por qué el modelo tomó una decisión.

Robustez a outliers: Decisiones basadas en umbrales, no en valores extremos.

→ Un cliente con ingreso \$1M no destruye el modelo.

Riesgo (Underfitting): La simplicidad tiene un costo, puede que el modelo no logre capturar la complejidad de las relaciones.

Riesgo (Overfitting): Si dejamos que el árbol crezca demasiado, memorizará casos aislados (ruido) en lugar de patrones generales. Aquí vuelve el concepto de Parsimonia: "Poda el árbol para que sea útil".

ENTROPÍA: EL TERMÓMETRO DEL DESORDEN



CONCEPTO INTUITIVO:

¿Qué tan "mezclados" están tus datos?

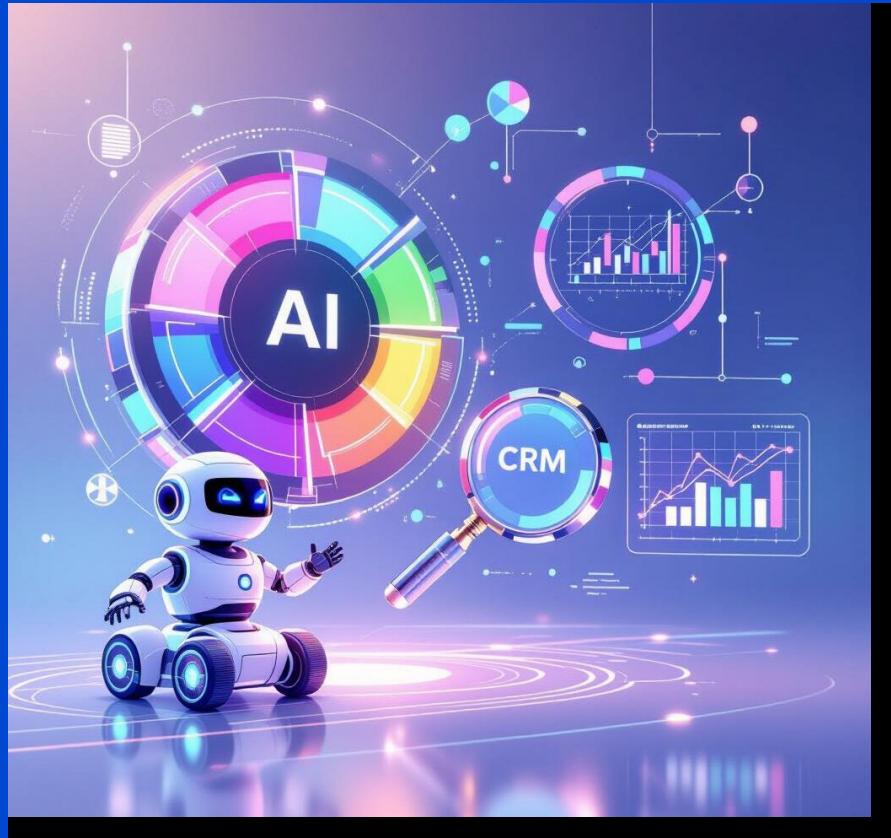
- Entropía ALTA = Caos total (50% sí, 50% no)
- Entropía BAJA = Orden perfecto (100% sí, 0% no)

EJEMPLO PRÁCTICO:

Nodo con 100 clientes:

- 50 compran, 50 no → Entropía = 1.0 (máximo desorden)
- 90 compran, 10 no → Entropía = 0.47 (más orden)
- 100 compran, 0 no → Entropía = 0.0 (orden perfecto)

CHECKLIST DE IMPLEMENTACIÓN



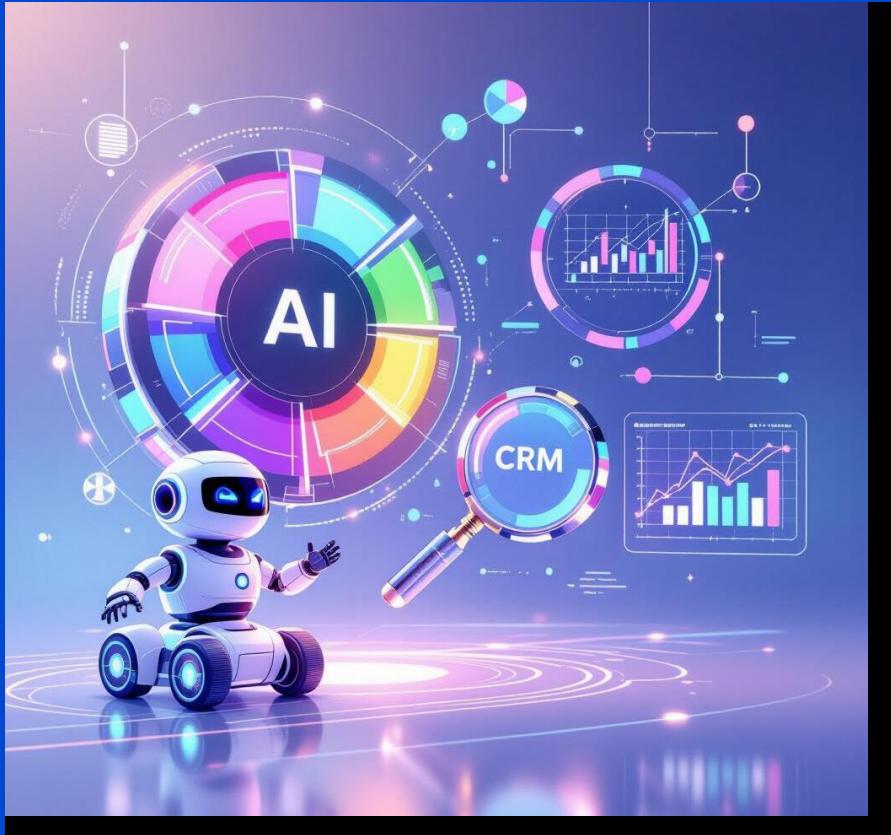
CALIDAD DE DATOS

- Variables categóricas codificadas (LabelEncoder/OneHot)
- Valores missing manejados (árboles los manejan)
- Outliers identificados (aunque árboles son robustos)

ENTRENAMIENTO

- División train/test (80/20)
- Validación cruzada para hiperparámetros (ejemplo: profundidad del árbol)
- Métricas de evaluación (accuracy, recall, precisión)

CHECKLIST DE IMPLEMENTACIÓN



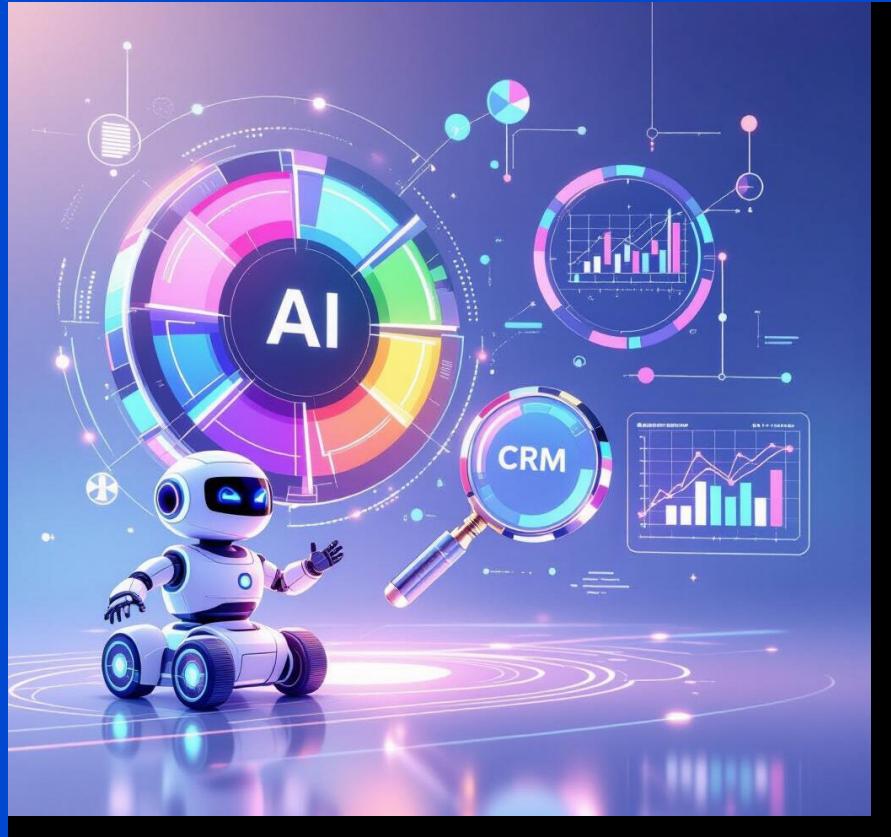
INTERPRETABILIDAD

- Reglas extraídas y documentadas
- Importancia de variables calculada
- Explicación preparada para stakeholders

MONITORIZACIÓN

- Plan para actualización periódica
- Métricas de desempeño en producción
- Sistema para añadir nuevas reglas manualmente
- Auditoría de decisiones controvertidas

CHECKLIST DE IMPLEMENTACIÓN



ÉTICA Y CUMPLIMIENTO

- Variables sensibles excluidas (género, etnia, etc.)
- Sesgos evaluados en diferentes grupos
- Cumplimiento con regulaciones (GDPR, etc.)
- Mecanismo de apelación humana