

# Modelo de Predicción de Ventas

José Manuel Villegas Santamaría

Sebastián Zapata Patiño



# El Problema: Ineficiencia de Recursos

## Situación Actual

- Pronósticos subjetivos basados en intuición
- Saturación operativa del equipo técnico
- Datos no estructurados y desaprovechados
- Desconexión con historial de clientes

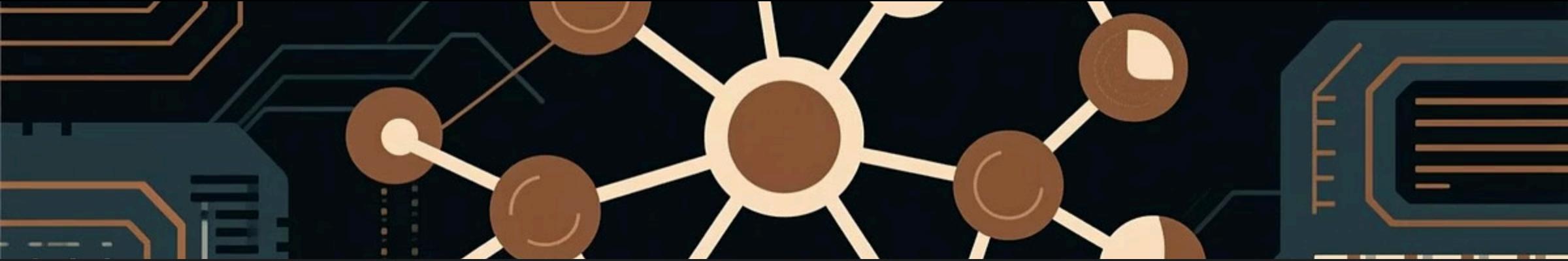
## Impacto en el Negocio

El equipo dedica el mismo tiempo a oportunidades con 5% de probabilidad que a las de 60%, desperdiando recursos críticos de ingeniería.

### Tasa de conversión global: 27.4%

- 7 de cada 10 esfuerzos se pierden.





# Justificación Técnica: ¿Por Qué Regresión Logística?



## Naturaleza Probabilística

El modelo calcula la probabilidad, como un porcentaje (de 0% a 100%), de que cerremos una venta. Esto nos permite decidir cómo actuar basándonos en qué tan seguros estamos de que la venta se hará realidad.



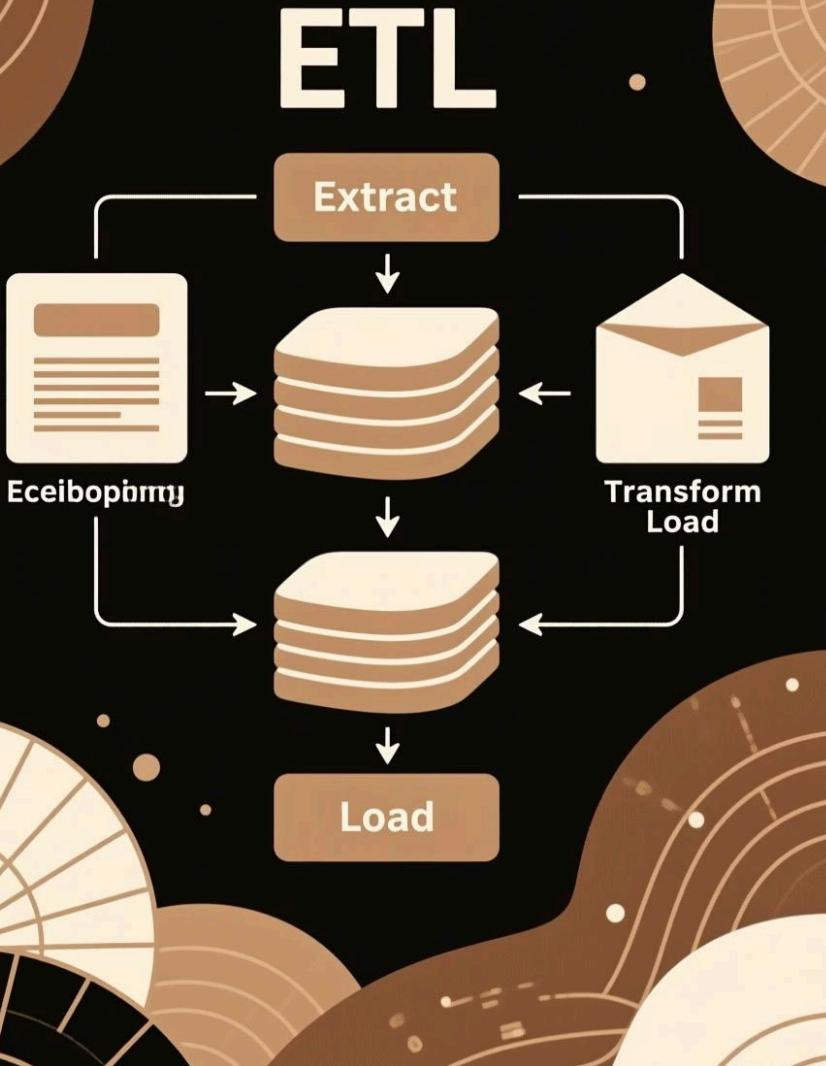
## Explicabilidad (White-Box)

Los coeficientes permiten analizar el impacto de cada variable. Ejemplo: Zona\_Costa negativo indica menor probabilidad de éxito.



## Eficiencia en Datos Tabulares

Generaliza mejor que modelos complejos en datasets medianos, evitando sobreajuste y optimizando recursos computacionales. Además se adapta a el dataset utilizado.



# Ingeniería de Datos y Taxonomía

01

## Limpieza y Normalización

Función extraer\_categoria utiliza diccionario complejo (CATEGORIA\_KEYWORDS) que mapea términos técnicos y los clasifica en 6 categorías macro.

02

## Feature Engineering

Conversión de texto no estructurado a categorías analíticas mediante reglas y diccionarios taxonómicos.

03

## Codificación One-Hot

Variables Cliente, Zona, Vendedor y Categoría transformadas para el modelo predictivo.

04

## Persistencia en SQLite

Base de datos local almacena usuarios e historial de predicciones con encriptación SHA-256.

# Resultados de Evaluación del Modelo

**76.22%**

## Accuracy (Exactitud)

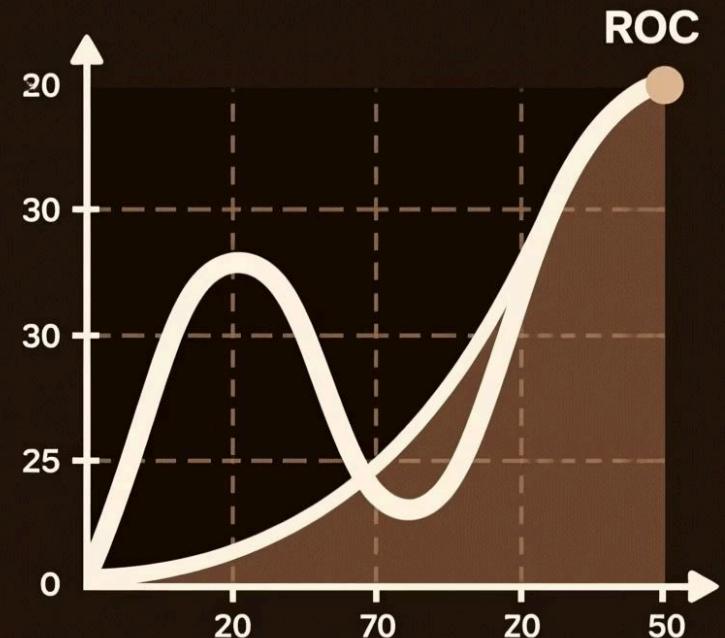
El modelo acierta en 3 de cada 4 casos, superando el umbral del 70% considerado éxito operativo en ventas B2B.

**77.97%**

## AUC-ROC

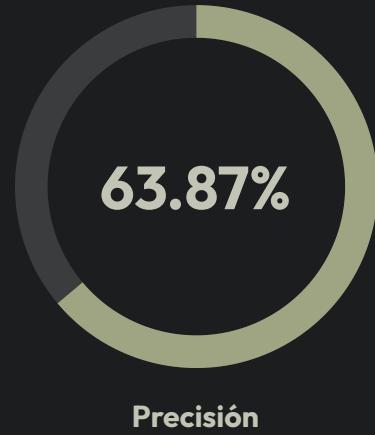
Capacidad robusta para distinguir entre clases. Confirma criterio de discriminación sólido del sistema.

Dataset robusto de **74,311 cotizaciones históricas** con **20,382 ventas generadas** y **5,288 clientes únicos**.



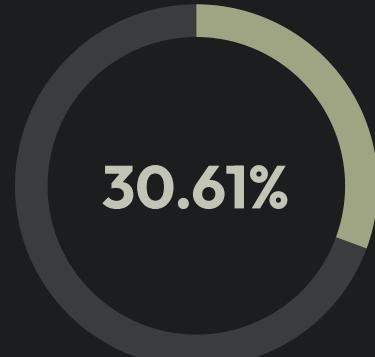
# Análisis Crítico: Modelo Conservador

## Precisión vs. Sensibilidad



De las oportunidades marcadas como "Ganadas",

el 63.87% realmente se ganaron.



El modelo detectó solo el 30.61% de todas las oportunidades reales.

## Interpretación Estratégica

El modelo es altamente selectivo: prefiere "perderse" una oportunidad dudosa antes que desperdiciar tiempo del equipo técnico.

Actúa como **filtro exigente** que optimiza el recurso más costoso: el tiempo del ingeniero experto.





# Matriz de Confusión y Optimización de Recursos

## Verdaderos Negativos (TN)

**50,399 casos** - Excelente identificación de cotizaciones "basura". Mayor ahorro de costos operativos.

## Verdaderos Positivos (TP)

**6,239 casos** - Oportunidades correctamente identificadas como ganadoras.

## Falsos Negativos (FN)

**14,143 casos** - "Dinero dejado en la mesa". Gestión automatizada de bajo costo para recuperarlos.

## Falsos Positivos (FP)

**3,530 casos** - Tasa moderada controlada por la alta precisión del modelo.



# Estrategias según Predicción:

## \* Probabilidad < 50%

### Diagnóstico

Cliente o zona con antecedentes negativos históricos. Costo de adquisición supera retorno potencial.

**Estrategia:**  
Automatización y  
Descarte para  
Ahorro de OPEX

### Acciones Prescriptivas

- Cotización estándar automatizada o lista de precios pública
- Sin visitas técnicas presenciales ni levantamiento en campo
- Email marketing automatizado para capturar valor residual
- Desestimar si no hay presupuesto confirmado

# \* Probabilidad > 50%



## Rango 50-70%: Moderada

Visita de diagnóstico técnico. Ofrecer alternativa Premium y Económica.

**Venta Consultiva:** Indagar puntos de dolor. Usar módulo Cross-Selling.



## Rango >70%: Agresiva

Cliente recurrente + Zona fuerte + Producto líder.

**Proceder a cerrar.**

**Enfoque:** Términos de pago, tiempos de entrega, descuentos por volumen. Bloqueo de competencia.





# Conclusiones y Valor Estratégico

## Filtro de Costos Inteligente

Modelo optimizado para eficiencia operativa: concentra esfuerzos en el 63.87% de predicciones positivas confiables.

## Transformación de Activos

Base de datos histórica se convierte en activo intangible que perdura más allá de la rotación del personal.

## IA Accesible para PYMES

No requiere Big Data masiva ni infraestructura costosa. Dataset estructurado moderado genera alto impacto.

## Calidad sobre Complejidad

80% del éxito recae en ingeniería de características. Limpieza de datos superior a complejidad del algoritmo.