



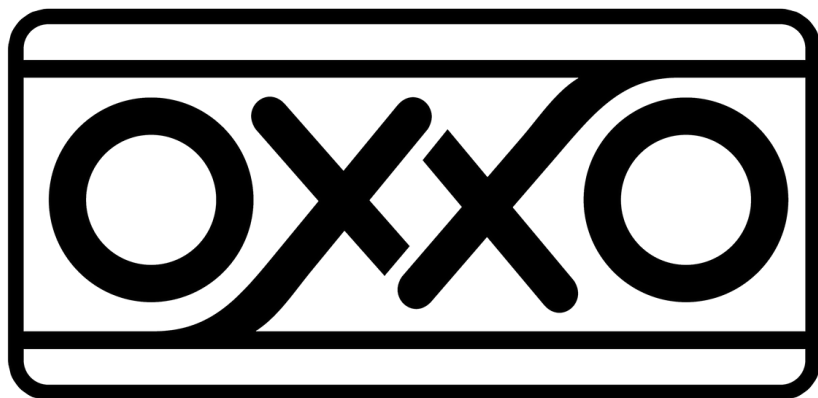
# DESCRUBRIMIENTO DE NUEVOS PUNTOS DE VENTA

Gael Arnulfo Ordaz Zamora  
Jose Davud Banda Rodriguez  
Natalia Sofia Guevara Hernandez  
Alejandro Adriaensens Martinez

# Objetivo

---

Desarrollar un modelo de predicción que a partir de una ubicación (latitud/longitud) determine si una tienda Oxxo tiene alto potencial de éxito, es decir, cumpla con su meta de venta.

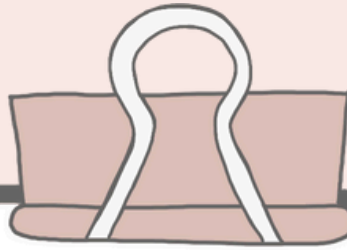




# Limpieza de Datos

---

- Interpolar datos por casos . Se uso una regresion lineal para imputar estos datos
- Una simulacion de MonteCarlo con una variable “normal” para quitar nulos.
- Eliminar Nans de Segmento maestro
- Venta igual se hizo una suavizacion de outliers a traves de Z-Score.

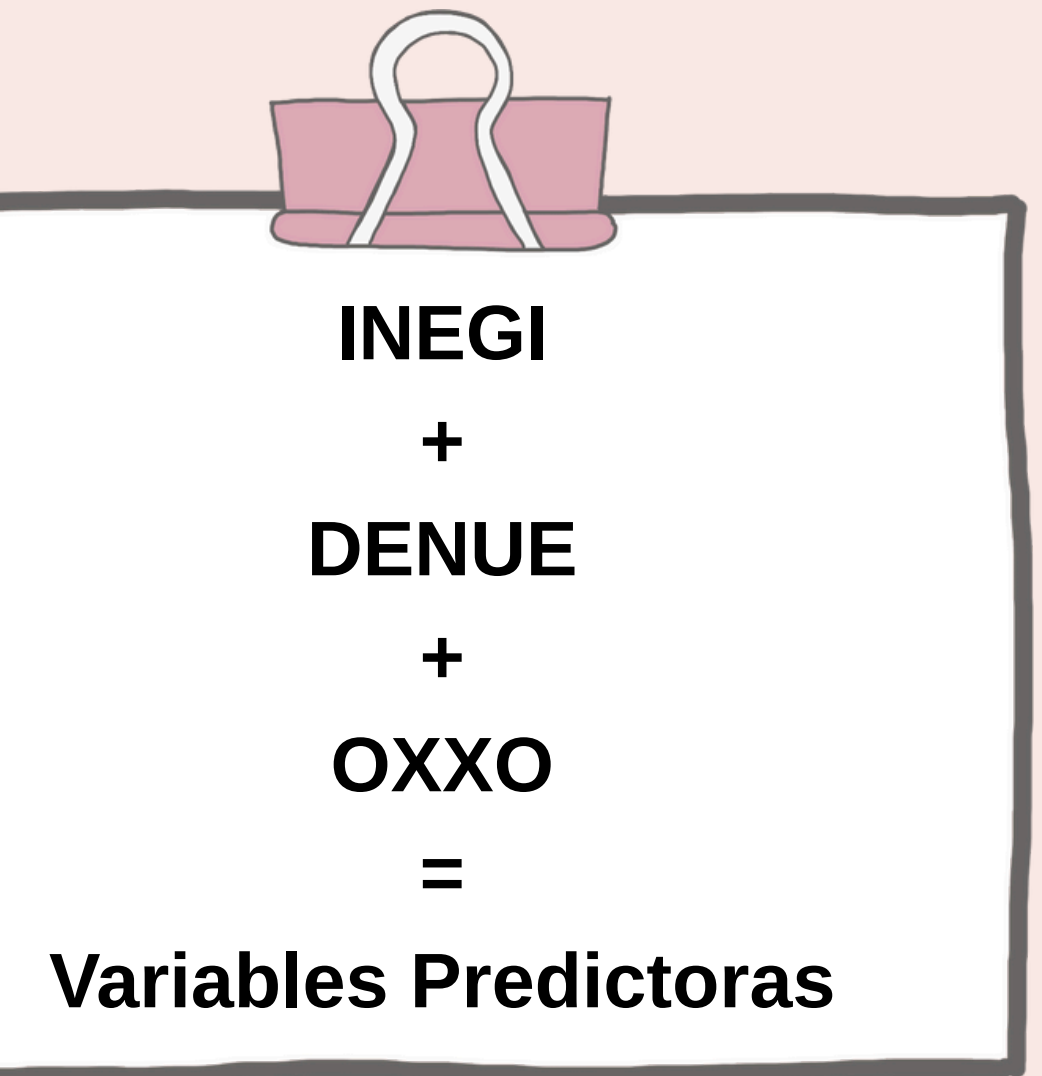
- 
- 1 Entender datos**
  - 2 Interpolar valores**
  - 3 Asegurarse de tener datos limpio**



# Features Generadas

---

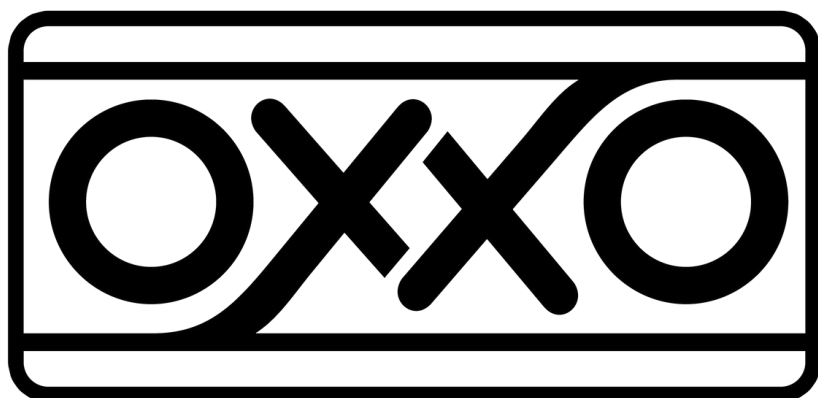
- Integración de variables demográficas externas del INEGI (e.g., población, edad, escolaridad).
- Enriquecimiento del dataset con datos del DENUE sobre unidades económicas cercanas (tipo, cantidad y giro comercial).
- Aplicación de One-Hot Encoding para variables categóricas relevantes.



# Criterio de decisión

---

El criterio de decisión se basa en la métrica denominada porcentaje de cumplimiento, calculada como el promedio de ventas de la tienda dividido entre la venta meta. A partir de este valor, se define la variable objetivo éxito, donde se asigna un 1 si el porcentaje de cumplimiento es mayor a 1.5 (indicando que la tienda es exitosa) y un 0 en caso contrario.



# MODELO PREDICTIVO

## NUEVOS OXXO



## Análisis del balance de clases

- Se identificó un desbalance entre clases (éxito = 1 vs. 0), lo cual es relevante para ajustar métricas y parámetros como `scale_pos_weight` en modelos como XGBoost.

## Análisis del balance de clases

- One-Hot Encoding para variables categóricas.
- SimpleImputer con estrategia mean para variables numéricas faltantes.

## Búsqueda de hiperparámetros

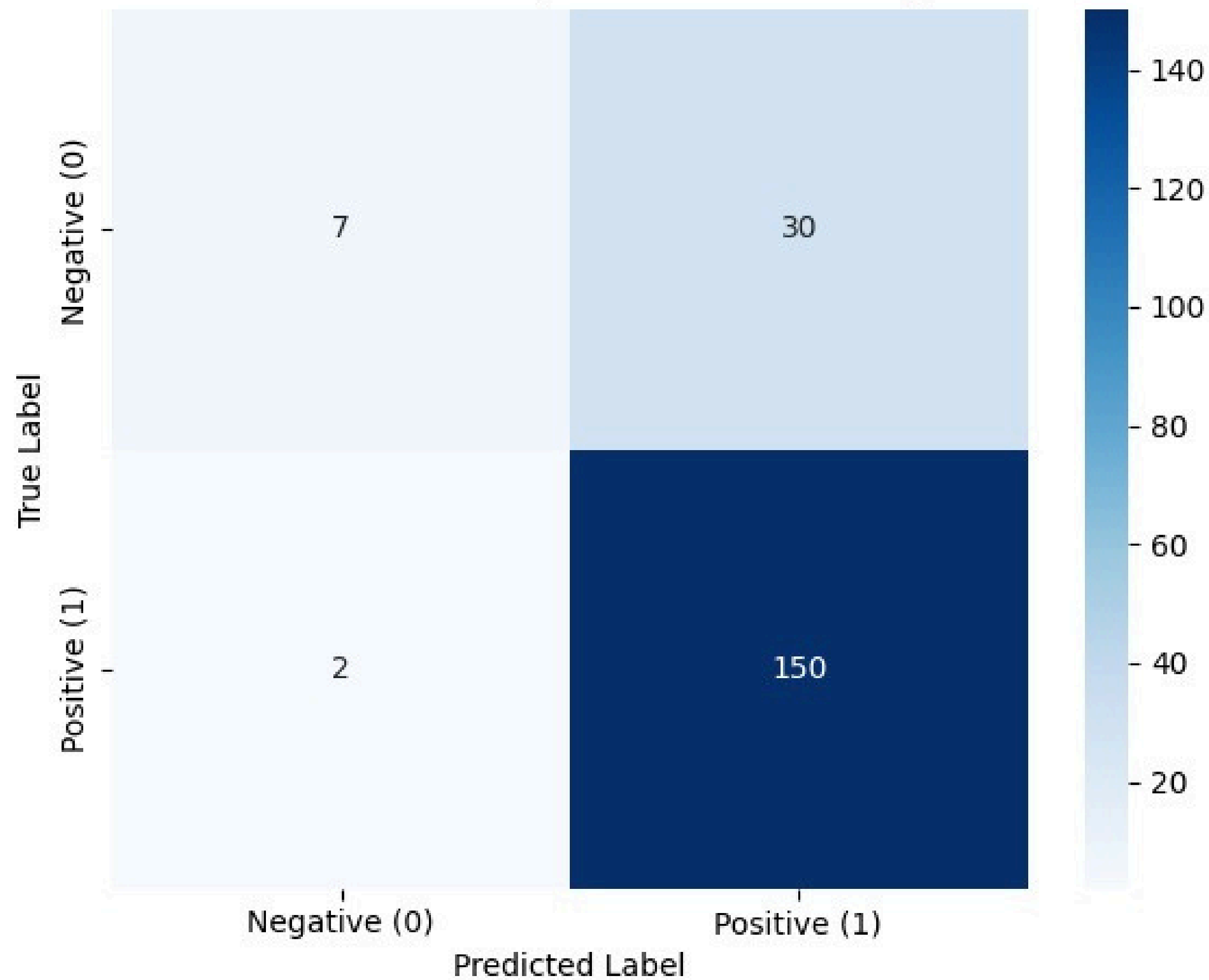
Se utilizó GridSearchCV con validación cruzada de 5 pliegues. El modelo fue optimizado para la métrica F1 de la clase 0, priorizando la reducción de falsos positivos al predecir éxito.

## Evaluación

- Reporte de clasificación (precisión, recall, F1-score por clase).
- Matriz de confusión para observar errores y aciertos del modelo.

M  
O  
D  
E  
L  
O

Confusion Matrix (Best XGBoost Model)





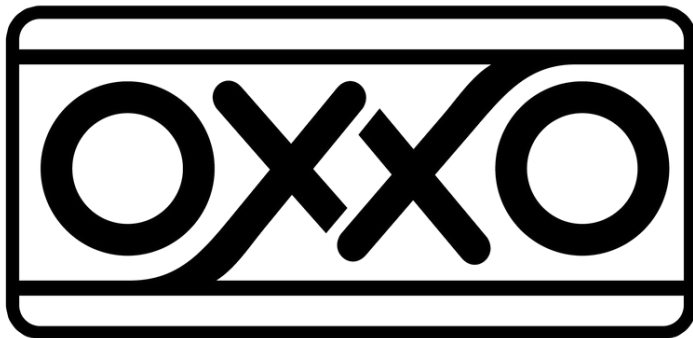
# Resultados

## Buenas oportunidades

Precisión del 83% y un recall del 99% para tiendas exitosas (clase 1), lo cual significa que el modelo acierta casi siempre al identificar tiendas con alto potencial, minimizando el riesgo de dejar pasar buenas ubicaciones.

## Seguridad

El F1-score de 0.90 en esta clase refuerza la solidez del modelo al equilibrar precisión y recall, lo que es clave para decisiones comerciales seguras.



## Competencia

Si, el recall es bajo , sin embargo indica que en algunos casos se etiquetan como exitosas tiendas que no lo serán. Este comportamiento es intencionalmente tolerable ya que mencionaron que aunque la tienda no fuera tan exitosa prefieren abrir una nueva ubicación a perderla con la competencia..

•¿Qué sugerencias o consideraciones podrían hacer para llevar este modelo a otros negocios (tiendas bara, caffenio, farmacias Yza, etc.)?

Escalabilidad

Reproducibilidad

Foco comercial  
realista

Datos abiertos

Clustering

Revelar patrones  
latente

Segmentación  
del territorio

Identificar zonas  
prioritarias





¡GRACIAS!

**FEMSA**