

Priorizar usuarios de alto crecimiento en una plataforma de delivery

Ciencia de Datos Aplicada 2025-2

Juan David Valencia - 201728857
Juan Esteban Cuellar - 202014258

Contexto y objetivo

La empresa es una plataforma de delivery de comida con dos equipos clave:

Growth: acompaña al usuario hasta su 4ta orden.

Engagement: toma el relevo desde la 4ta orden.

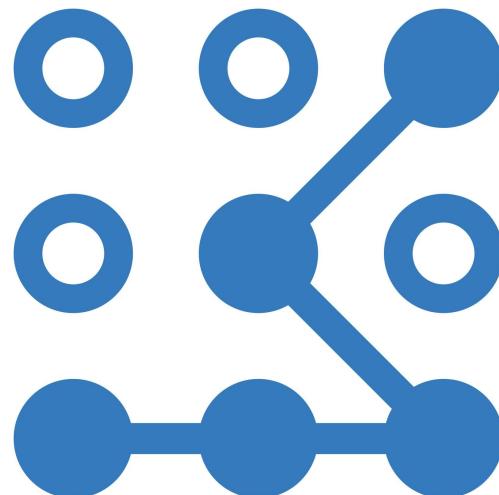
Problema: No hay un esquema claro para priorizar qué nuevos usuarios de Engagement tienen mayor potencial de crecimiento.



Problema de negocio

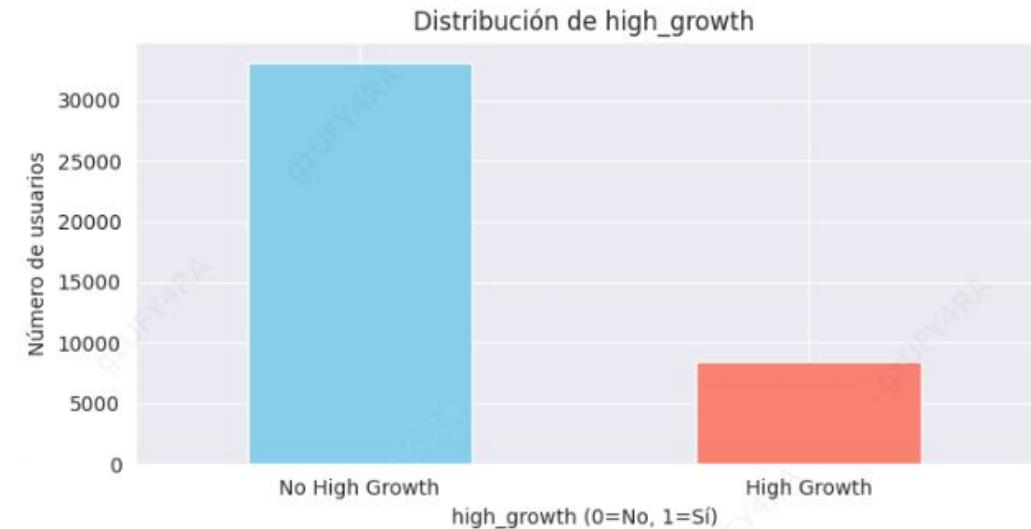
Muchos incentivos, poca priorización

- Hoy se envían bonos y descuentos de forma bastante homogénea.
- No se diferencia bien qué usuarios tienen más potencial de crecimiento.
- Queremos priorizar a quién vale más la pena invertirle.



Cohorte y definición de high_growth

- Nos enfocamos en usuarios que llegan a la 4^a orden.
- Miramos sus órdenes en los 3 meses siguientes.
- Definimos delta_orders y high_growth = top ~20 %.



Preparación: ejemplo de transformaciones

Volumen y actividad

total_orders

log_total_orders

orders_per_day

Velocidad y recencia

efo_to_four_scaled

days_since_first_order

recencia_num

categoria_recencia

Diversidad y lealtad

shop_diversity

category_diversity

shops_per_order

categories_per_order

is_multi_shop

is_multi_category

dominant_category

brand001_ratio

Segmentación y contexto

r_segment

city_token

Variables objetivo

delta_orders

high_growth

Diseño del modelo

Tarea: predecir high_growth (0/1).

Modelos probados: Random Forest, XGBoost, LightGBM.

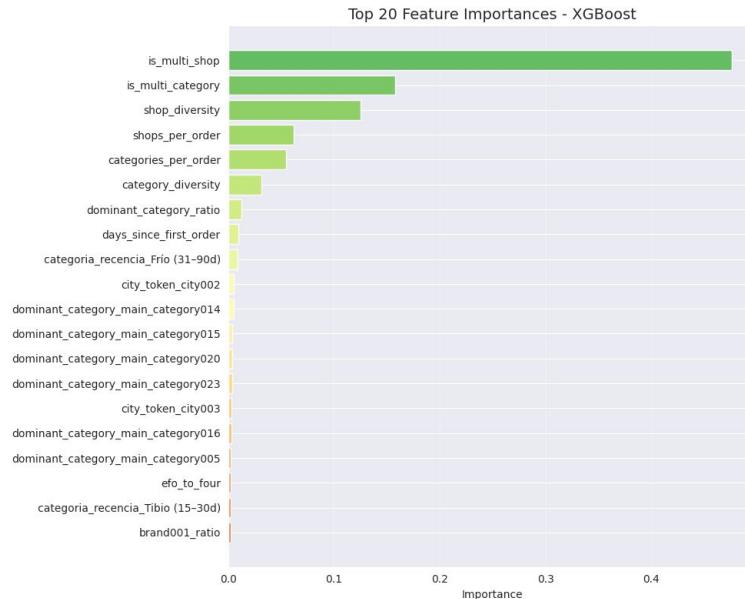
Validación: Train / Val / Test + 5-fold CV, métrica principal AUC-ROC, complemento F1 y Precision@20 %.

 TABLA COMPARATIVA:

Métrica	Random Forest	XGBoost	Mejor
AUC-ROC	0.9982	0.9998	XGB
Precision@20%	0.9670	0.9934	XGB
F1-Score	0.9614	0.9868	XGB
Precision	0.9324	0.9808	XGB
Recall	0.9923	0.9929	XGB
Accuracy	0.9838	0.9946	XGB

 Resumen de victorias: RF=0 | XGB=6

Qué aprendió el modelo (variables importantes)

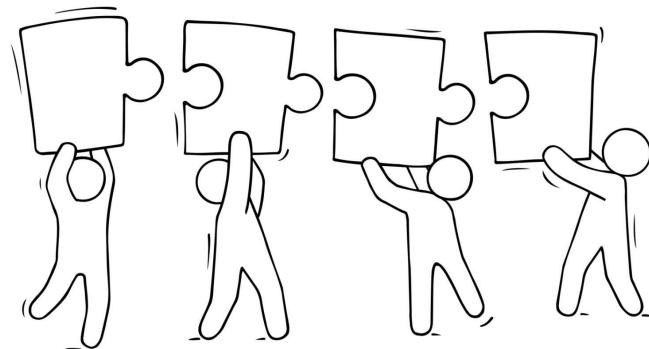


Top: is_multi_shop, is_multi_category, diversidad de tiendas/categorías, algo de recencia y ciudad.

Producto de datos

Ejemplos de segmentos y acciones

- Top-20 % + multitienda + multicategoría → beneficios de lealtad.
- Top-20 % + multi tienda pero monocategoría → cross-selling de nuevas categorías.
- Top-20 % + monotienda → campañas co-brandeadas con esa tienda.
- Score moderado + recencia 31–90 días → campañas de reactivación.



Uso responsable y límites

Solo usamos comportamiento agregado, sin datos personales visibles.

El modelo prioriza beneficios, no restringe servicio.

Hay que revisarlo en el tiempo y validar con pruebas A/B.



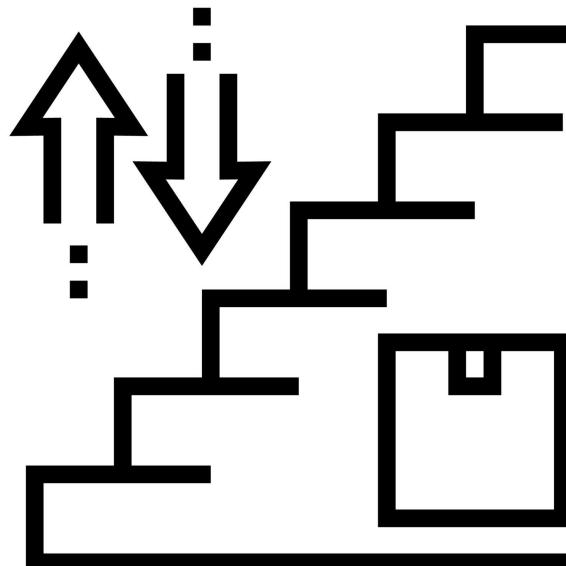
Aprendizajes y siguiente paso

Aprendizajes

- Importancia de definir bien recencia, delta_orders y high_growth.
- Los usuarios de alto crecimiento son multitienda/multicategoría.
- Tener un pipeline claro ayuda mucho a iterar.

Próximos pasos

- Probar con cohortes más recientes.
- Medir impacto con A/B tests.
- Explorar modelos complementarios (regresión de delta_orders, etc.).



Gracias