



PERSONALIZED FASHION RECOMENDATIONS

PARADIGMAS Y PROGRAMACIÓN PARA CIENCIA DE DATOS

ISAAC MONROY VILLEGAS
ÁNGEL NOEL PÉREZ MARTÍNEZ
VALENTINA AYELEN CRUZ MENDOZA
MILENA FERNANDA RIVERA HERNÁNDEZ



iiimas

Objetivo

PODER PREDECIR HASTA 12 ARTÍCULOS QUE CADA CLIENTE COMPRARÁ EN EL SIGUIENTE PERÍODO DE 7 DÍAS. Y DESARROLLAR UN PRODUCTO DE RECOMENDACIONES BASADO EN LAS COMPRAS PREVIAS (**3 MESES**) DEL CLIENTE Y LOS PRODUCTOS OFRECIDOS POR LA EMPRESA H&M.



Recordatorios

- Transacciones (customer_id, article_id, timestamp),
- Articulo (category, department, product_type, description, price, image)
- Cliente (age, postal code, entre otros).

dataset_recomendacion_final

customer_id	article_id	label	total_purchase	last_purchase_date	recency_days	purchase_count	average_price
dda4b507ba647aec8fe0c10767133f05cce8b847c0e1fbd9092a085724b32356	850917001	1	11	2020-09-20	2	3530	0.02470393719690780
db48cdc0f7cedaea840f50dc2e7b52dd84e0d67c48ac820316cbb6190098408c	641804002	1	57	2020-09-20	2	38	0.03302185548617300
087e522ecf2fe0bf21eb9df8f1d93d0159e7aacad094bce505deefddc556e75c	850241001	1	14	2020-08-29	24	394	0.016158048696549900
7efc9f365f8905b3b7112097822e6c1acaba16bc991972431794028be310d944	862482009	1	32	2020-09-14	8	652	0.031092882395757400

Resultados



Compras pasadas del cliente

“JANE DOE”

ÚLTIMOS 3 MESES:

¿QUÉ TENDENCIAS OBSERVAMOS DE JANE?



Logistic Regression

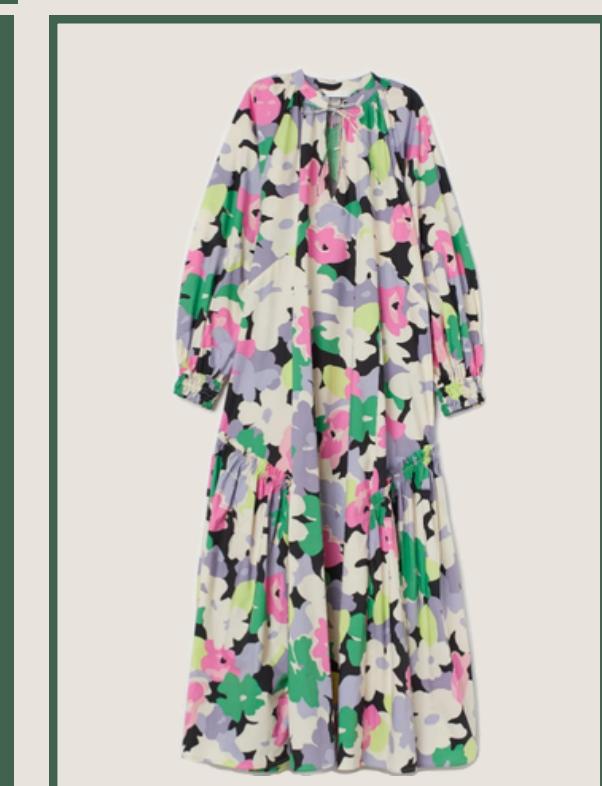
Aplicamos Regresion Logistica debido a que es un modelo de clasificacion que predice la probabilidad de que un evento ocurra, en este caso, que un cliente compre un articulo.

Accuracy (Exactitud)	: 0.7872
Precision	: 0.7997
Recall (Sensibilidad)	: 0.5728
F1 Score	: 0.6675
ROC AUC Score	: 0.8662
Log Loss	: 0.4757

Classification Report :				
	precision	recall	f1-score	support
0	0.78	0.91	0.84	269025
1	0.80	0.57	0.67	160000
accuracy			0.79	429025
macro avg	0.79	0.74	0.76	429025
weighted avg	0.79	0.79	0.78	429025

Logistic Regression

Customer_id	Prediction
0001d44dbe7f6c4b35200abdb052c77a87596fe1bdcc37e011580a479e80aa94	850244001, 817361007, 815434001, 811899002, 909869004, 913688001, 812363005, 851802001, 853474004, 823518002, 888010002, 869424001



Naive Bayes

Accuracy (Exactitud) :	0.7056
Precision	: 0.8616
Recall (Sensibilidad) :	0.2505
F1 Score	: 0.3881
ROC AUC Score	: 0.8138
Log Loss	: 0.5767

- Velocidad y Escalabilidad
- Moda Cambia → Entrenamiento Diario
- Si tenemos 50 colores y 20 secciones, Naive Bayes trata cada característica de forma independiente.
- ¡Te da la proba de que SI compro!

Classification Report:		precision	recall	f1-score	support
No Compra	0.69	0.98	0.81	269098	
Compra	0.86	0.25	0.39	159927	
accuracy			0.71	429025	
macro avg	0.77	0.61	0.60	429025	
weighted avg	0.75	0.71	0.65	429025	

Gaussiano

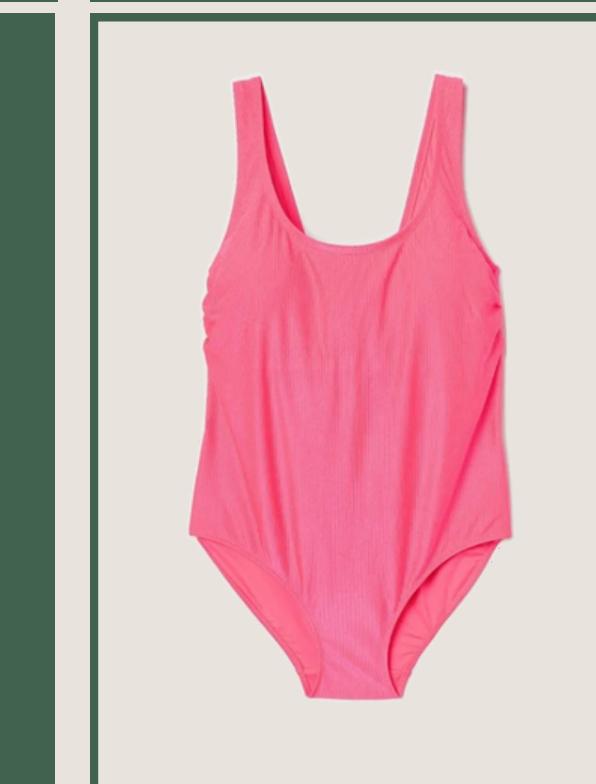
Naive Bayes

Customer_id

0001d44dbe7f6c4b35200abdb052c77a87596fe1bdcc37e011580a479e80aa94

Prediction

0850244001 | 0817361007 | 0815434001 | 0909869004 | 0811899002 | 0913688001 | 0812363005 |
0851802001 | 0853474004 | 0823518002 | 0900157002 | 0806964002 |



Random Forest

- Previene naturalmente el sobre-entrenamiento
- Permite identificar claramente que la frecuencia de compra (popularidad de prendas) es más importante
- Similar en rendimiento a XGBoost pero menos complejidad de tuning

Accuracy	: 0.8126
Precision	: 0.7553
Recall	: 0.7356
F1 Score	: 0.7453
ROC AUC	: 0.8854
Log Loss	: 0.4096

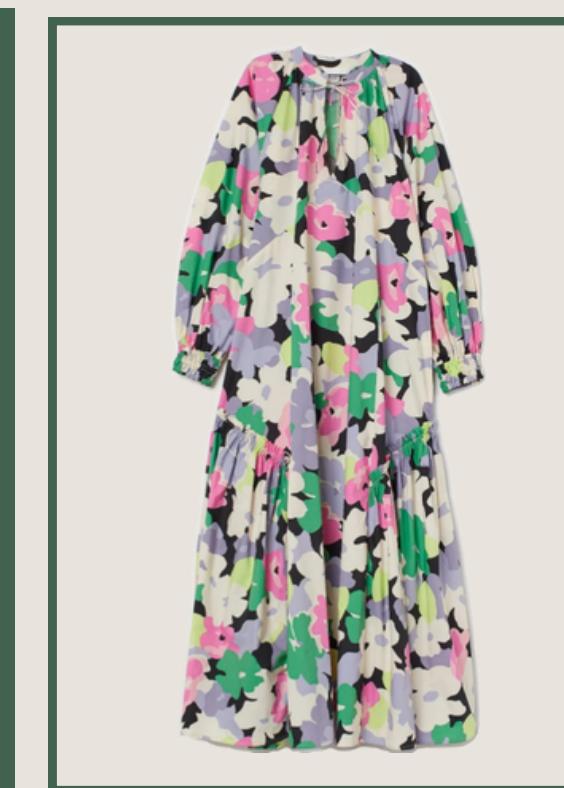
RANDOM FOREST

Customer_id

0001d44dbe7f6c4b35200abdb052c77a87596fe1bdcc37e011580a479e80aa94

Prediction

0850244001 | 0817361007 | 0815434001 | 0913688001 | 0812363005 | 0909869004 | 0851802001 |
0811899002 | 0853474004 | 0869424001 | 0888010002 | 0876461001 |



XGBOOST

Accuracy (Exactitud) : **0.8129**
Precision : **0.7550**
Recall (Sensibilidad): **0.7377**
F1 Score : **0.7463**
ROC AUC Score : **0.8867**
Log Loss : **0.4075**

Classification Report:

	precision	recall	f1-score	support
No Compra	0.85	0.86	0.85	269025
Compra	0.76	0.74	0.75	160000
accuracy			0.81	429025
macro avg	0.80	0.80	0.80	429025
weighted avg	0.81	0.81	0.81	429025

- RENDIMIENTO GENERAL: UN F1-SCORE DE 0.746 Y UNA EXACTITUD DEL 81.3%, VALIDANDO UN MODELO ESTABLE Y CONFIABLE.
- CAPACIDAD DE RANKING: ROC AUC DE 0.8867, DEMOSTRANDO QUE EL MODELO ORDENA EFICAZMENTE LAS PREFERENCIAS DEL USUARIO.
- PRECISIÓN DE COMPRA: 75.5%, LO QUE INDICA QUE 3 DE CADA 4 RECOMENDACIONES DE COMPRA SON ACERTADAS.
- COBERTURA (RECALL): 73.8%, CAPTURANDO LA GRAN MAYORÍA DE LAS OPORTUNIDADES DE VENTA REALES.

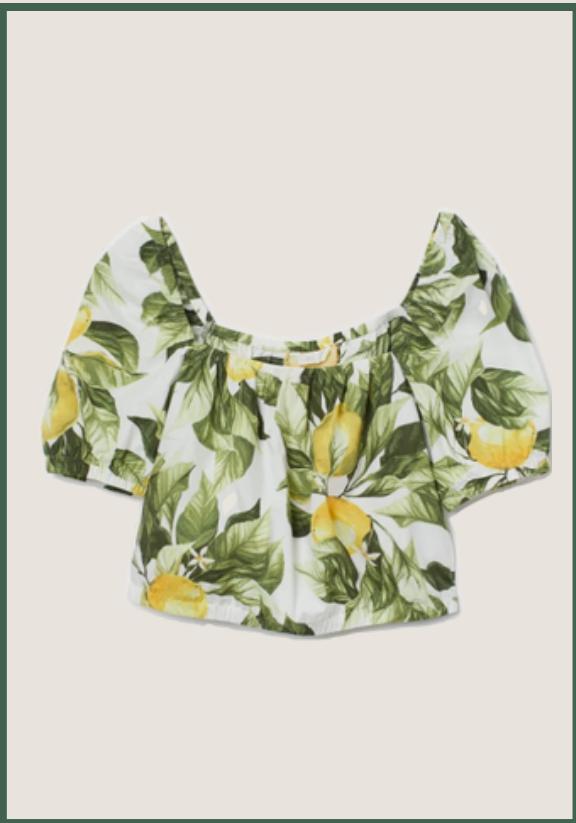
XGBOOST

Customer_id

00006413d8573cd20ed7128e53b7b13819fe5fcf2d801fe7fc0f26dd8d65a85a

Prediction

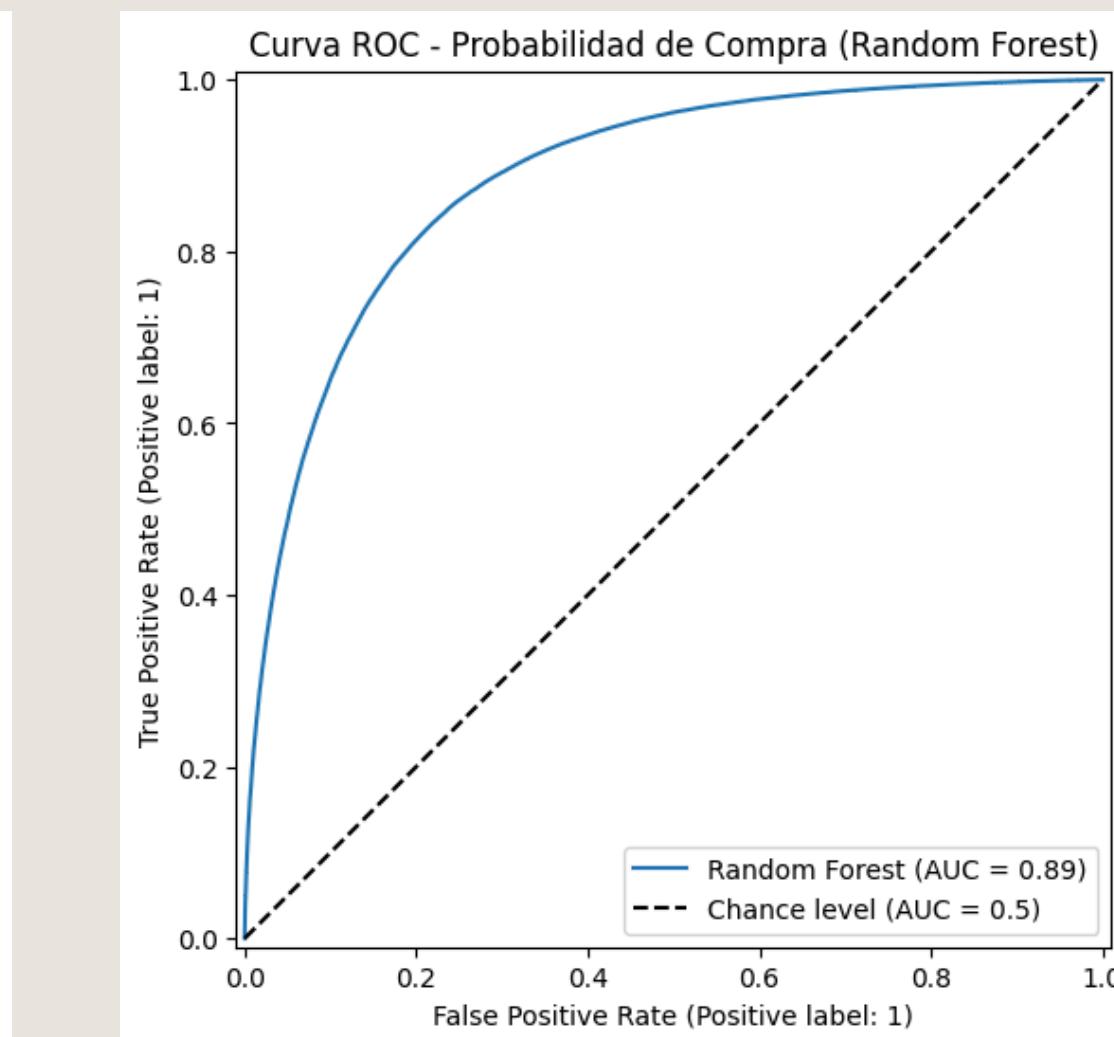
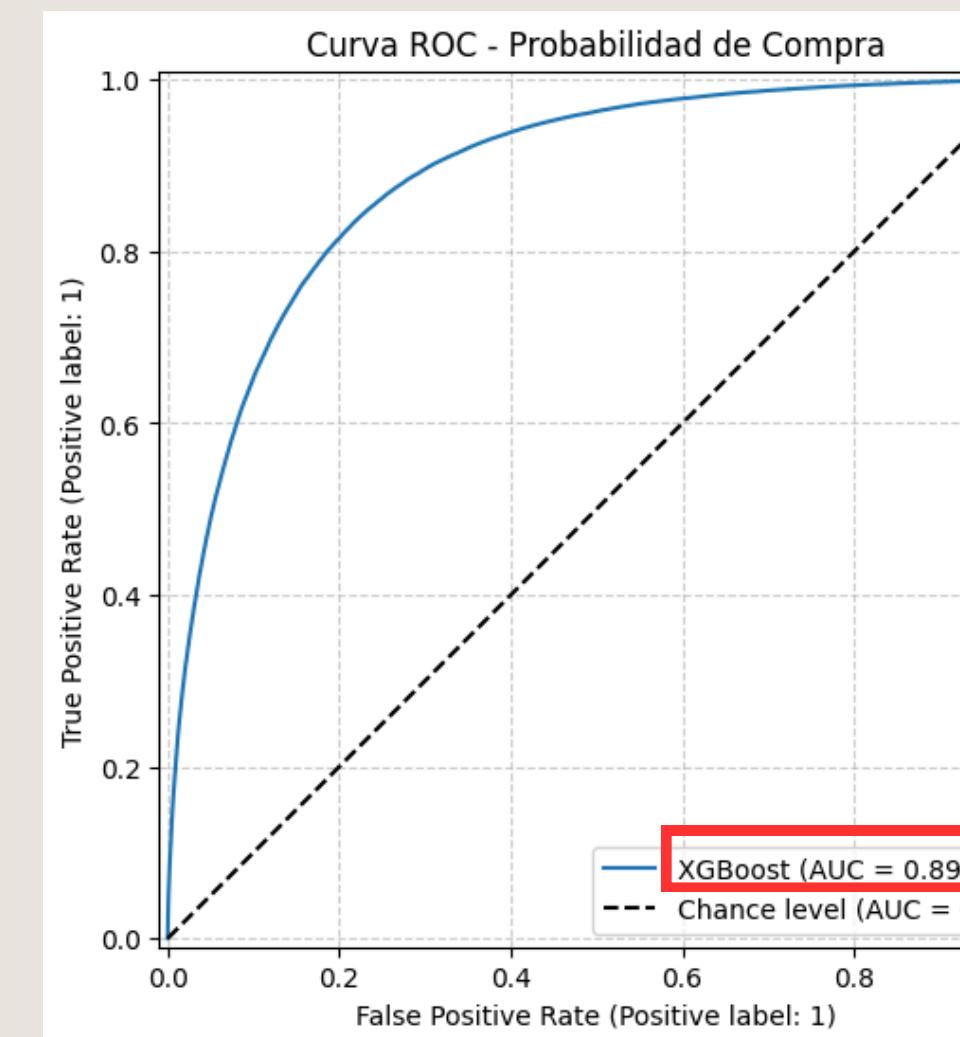
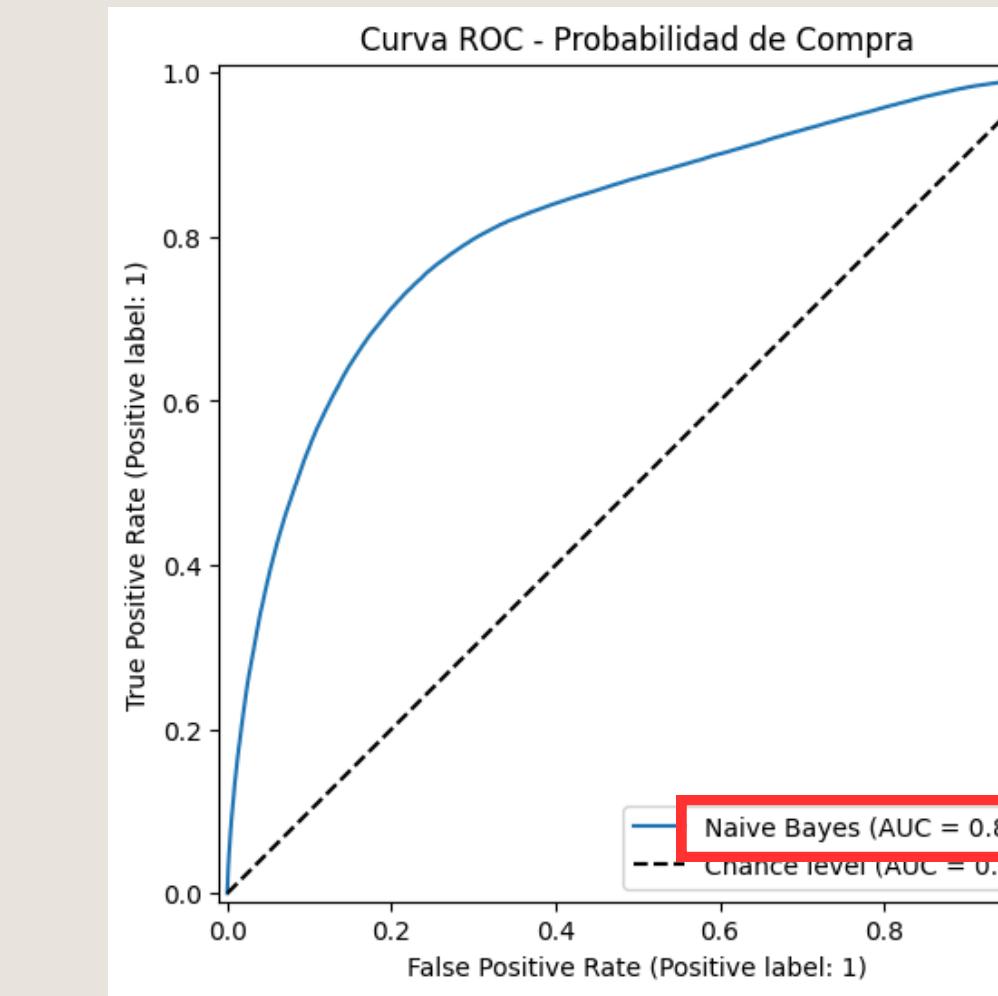
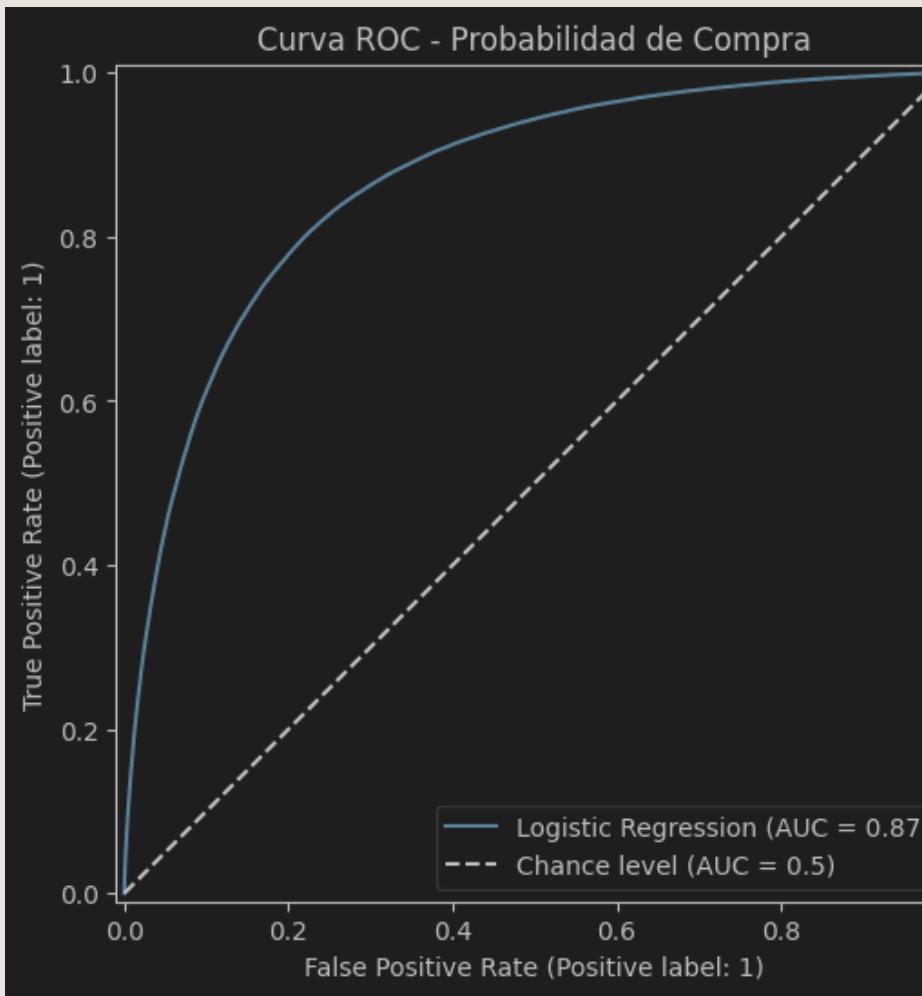
0850244001 | 0817361007 | 0815434001 | 0909869004 | 0811899002 | 0812363005 |
0851802001 | 0913688001 | 0853474004 | 0888010002 | 0823518002 | 0869424001 |



**¿Qué podemos
concluir con estos
resultados?**

1. RENDIMIENTO DE MODELOS (MÉTRICAS)

Curvas ROC



2. Aplicabilidad e Impacto del Modelo

3. Estrategia de Implementación

4. Impacto Estratégico para H&M





¡Gracias!