

Práctica 02. Función LOGIT.

Sistemas de Aprendizaje Automático.

Álvaro Martínez Lineros.

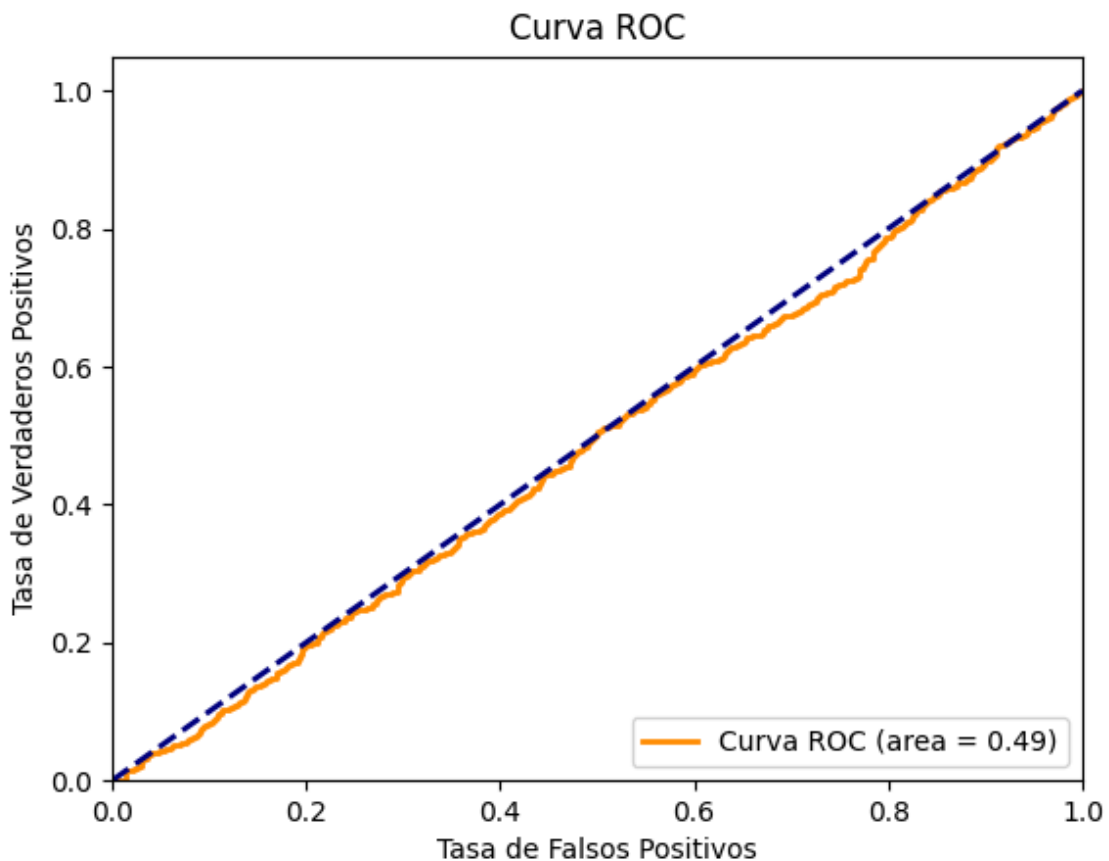
Introducción.

Se desea construir un modelo de clasificación basado en Regresión Logística que ayude a predecir si un cliente es propenso a comprar un producto específico después de recibir una campaña de marketing.

El modelo a utilizar es el Regresión Logística, se valorarán varios hiperparámetros para encontrar la configuración más precisa.

Resultados.

Curva ROC:



Matriz de confusión:

Verdaderos negativos	629	Falsos positivos	348
Falsos negativos	669	Verdaderos positivos	354

Reporte de clasificación:

	precision	recall	f1-score	support
No compra	0.48	0.64	0.55	977
Compra	0.50	0.35	0.41	1023
accuracy			0.49	2000
macro avg	0.49	0.49	0.48	2000
weighted avg	0.49	0.49	0.48	2000

Importancia de las características:

Característica	Importancia
Edad	-0.028720
Ingreso_Anual	-0.018325
Deudas_Activas	-0.022623
Puntaje_Credito	-0.001980

Factores clave.

Tras entrenar el modelo y ejecutar un test se puede sacar la conclusión de que ninguno de los factores son clave para la toma de decisiones.

Limitaciones del modelo.

La baja correlación entre los datos hace que la precisión del modelo sea extremadamente baja. Una posible mejora podría ser añadir otros factores correlacionados con la aprobación del crédito.

Aplicabilidad.

En una empresa este modelo no sería aplicable por su baja precisión y su poca generalización. Haría falta incrementar el f1-score de ambas situaciones a al menos 0.8 para poder considerar su aplicación.

Conclusión.

Este modelo no es aplicable en una situación real. Posibles soluciones serían aumentar el número de factores correlacionados con la aprobación o usar modelos como RandomForest, KNN, etc.