

Práctica 03. Función Naive Bayes.

Sistemas de Aprendizaje Automático.

Álvaro Martínez Lineros.

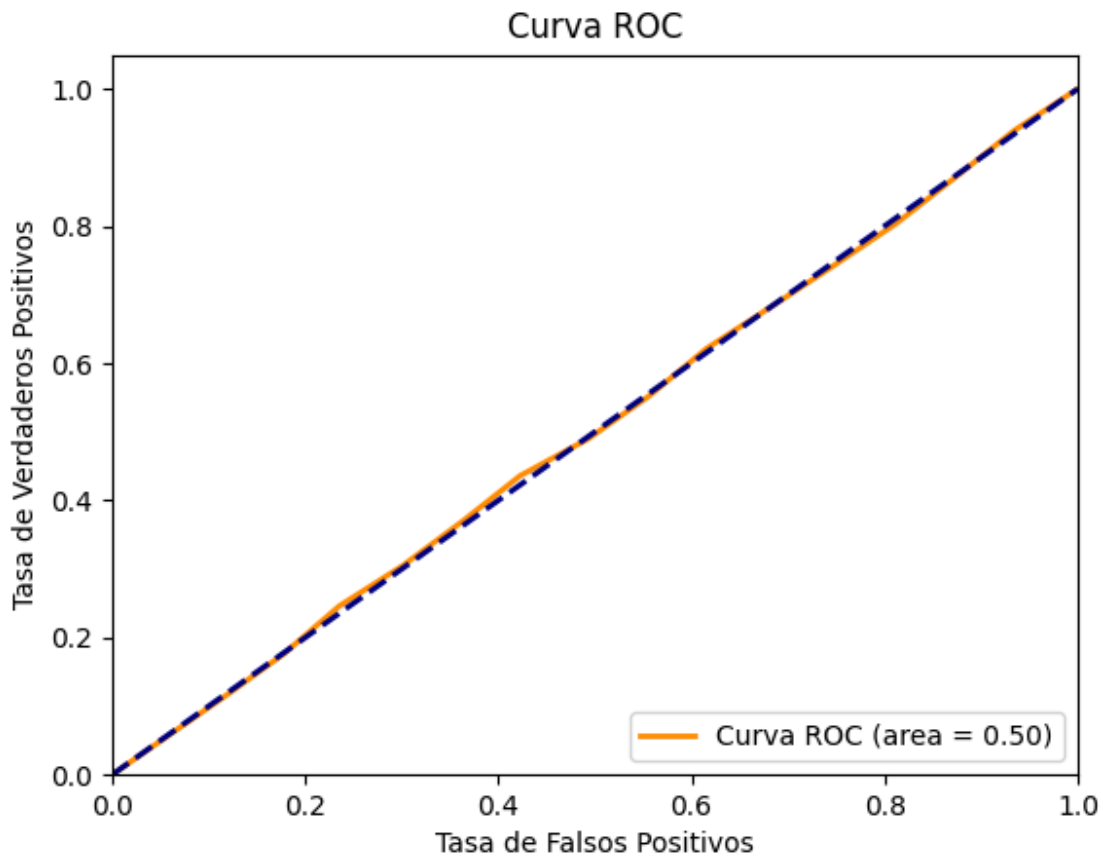
Introducción.

Se desea construir un modelo de clasificación basado en Naive Bayes 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 Bernoulli Naive Bayes porque sirve para datos categóricos (0,1), se valorarán varios hiperparámetros para encontrar la configuración más precisa.

Resultados.

Curva ROC:



Matriz de confusión:

Verdaderos negativos	3505	Falsos positivos	3404
Falsos negativos	1576	Verdaderos positivos	1511

Reporte de clasificación:

	precision	recall	f1-score	support
No compra	0.69	0.51	0.58	6909
Compra	0.31	0.49	0.38	3087
accuracy			0.50	2000
macro avg	0.50	0.50	0.48	2000
weighted avg	0.57	0.50	0.52	2000

Probabilidades condicionales:

No spam

Característica	Probabilidad
Contiene_Oferta	0.496372
Contiene_Enlace	0.492947
Contiene_Adjunto	0.499470
Longitud	0.508194

Spam

Característica	Probabilidad
Contiene_Oferta	0.503628
Contiene_Enlace	0.502568
Contiene_Adjunto	0.493355
Longitud	0.496372

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 detección del spam.

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.