

## Práctica 01. Árbol de decisión.

Sistemas de Aprendizaje Automático.

Álvaro Martínez Lineros.

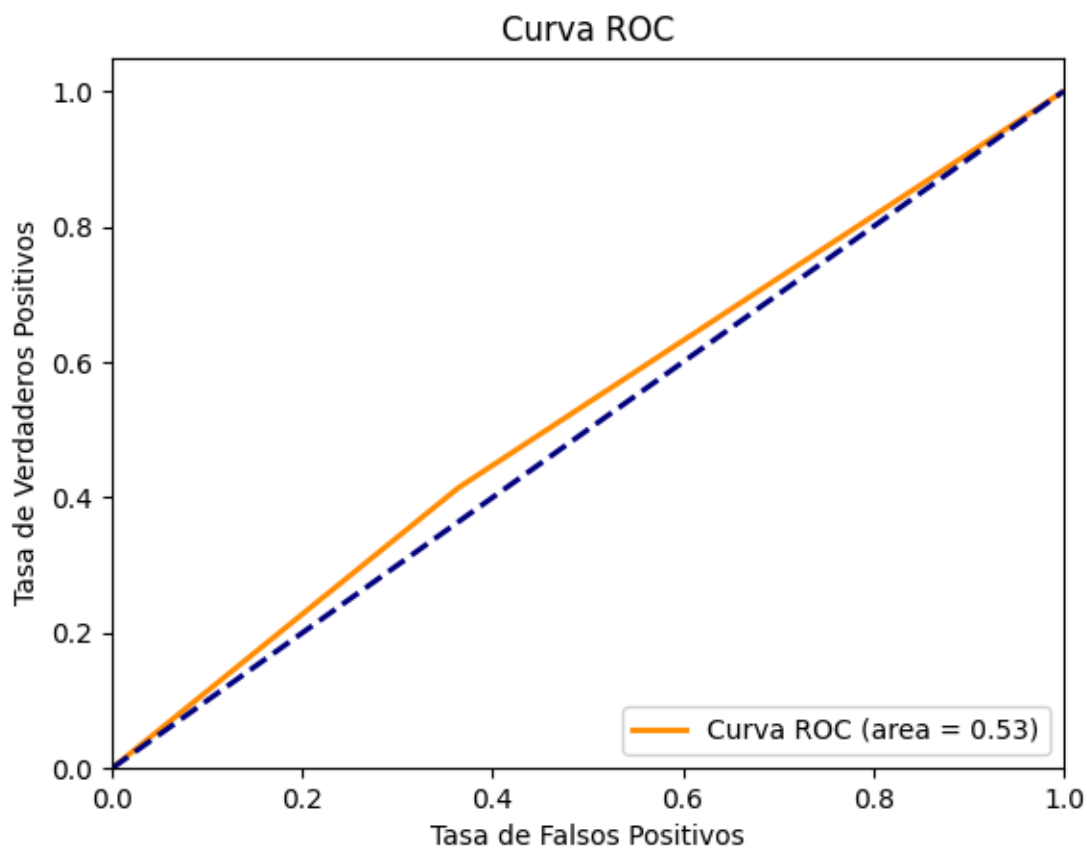
### Introducción.

Se desea construir un modelo de clasificación basado en Árboles de Decisión 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 Árbol de decisión, se valorarán varios hiperparámetros para encontrar la configuración más precisa.

### Resultados.

Curva ROC:



Matriz de confusión:

Verdaderos negativos	87	Falsos positivos	31
Falsos negativos	52	Verdaderos positivos	30

#### Reporte de clasificación:

	precision	recall	f1-score	support
No compra	0.63	0.74	0.68	118
Compra	0.49	0.37	0.42	82
accuracy			0.58	200
macro avg	0.56	0.55	0.55	200
weighted avg	0.57	0.58	0.57	200

#### Importancia de las características:

Característica	Importancia
Edad	0.317133
Ingreso_Anual	0.254751
Frecuencia_Visitas	0.074039
Numero_Compras	0.091661
Tiempo_En_Web	0.262416

#### Factores clave.

Tras entrenar el modelo y ejecutar un test se puede sacar la conclusión de que la edad y el ingreso anual son los factores más determinantes a la hora de decidir si una persona comprará o no.

#### Limitaciones del modelo.

La muestra de datos para el entrenamiento y la prueba del modelo es bastante escasa. Mil muestras no son suficientes para entrenar un modelo robusto y con una alta predicción. El modelo actual cuenta con una precisión total del 58%, siendo del 49% para las compras. Es como aplicar un modelo random de dos opciones.

Además, el modelo se ha hecho muy profundo (15 de profundidad máxima) lo que puede limitar su interpretabilidad y generalización.

#### 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 la muestra para poder entrenar mejor el modelo o cambiar a otro como la regresión logística o el Random Forest.