

Métodos de evaluación en clasificación binaria

En **clasificación binaria** (cuando quieres clasificar ejemplos en dos categorías, como *positivo* vs *negativo*), hay **métodos de evaluación** específicos que miden qué tan bien trabaja el modelo. Los más importantes son:

1. Accuracy (Exactitud)

- **Definición:** Porcentaje de predicciones correctas sobre el total de ejemplos.
- **Fórmula:**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Problema:** No es fiable si hay clases desbalanceadas (por ejemplo, 95% negativos).

2. Precision (Precisión)

- **Definición:** De todas las veces que el modelo predijo *positivo*, ¿cuántas veces acertó?

- **Fórmula:** $\text{Precision} = \frac{TP}{TP + FP}$

- **Importancia:** Alta precisión es crucial cuando **el costo de un falso positivo es alto**. Ejemplos:

- En sistemas de detección de spam marcar como spam un correo importante.
- En detección de fraudes bancarios bloquearle la tarjeta a un cliente honesto.
- En sistemas de reconocimiento facial en seguridad identificar erróneamente a una persona como sospechosa.

3. Recall (Sensibilidad o Exhaustividad)

- **Definición:** De todos los ejemplos *positivos* reales, ¿cuántos encontró el modelo?

- **Fórmula:** $\text{Recall} = \frac{TP}{TP + FN}$

- **Importancia:** Alta recall es crucial cuando **es muy costoso perder un positivo**. Ejemplos:
 - En detección de fraudes financieros no detectar a un verdadero fraude.
 - En seguridad informática (detección de malware o intrusiones) no detectar un ataque real.
 - En sistemas de alerta temprana (por ejemplo, incendios forestales, tsunamis, fallas de maquinaria crítica) no emitir una alerta cuando realmente había peligro.

4. F1-Score

- **Definición:** Media armónica entre precisión y recall.
- **Fórmula:**
$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$
- **Importancia:** Útil si quieres un balance entre precisión y recall.

5. ROC Curve y AUC (Área bajo la curva)

- **ROC Curve:** Muestra la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) a diferentes umbrales de decisión.
- **AUC:** Área bajo esa curva.
 - 1 = modelo perfecto
 - 0.5 = modelo al azar
 - < 0.5 = peor que azar

6. Matriz de Confusión

- **Descripción:** Una tabla que resume los resultados de las predicciones:

	Predicho Positivo	Predicho Negativo
Real Positivo	TP	FN
Real Negativo	FP	TN

- Permite ver *exactamente* dónde falla el modelo.

7. Otros más avanzados:

- **PR Curve (Precision-Recall Curve):** Más útil que ROC cuando las clases están muy desbalanceadas.
- **Balanced Accuracy:** Promedio de recall entre las clases.
- **Matthews Correlation Coefficient (MCC):** Medida robusta incluso en clases muy desbalanceadas.

¿Qué es una Curva ROC?

La curva ROC (Receiver Operating Characteristic) **muestra** cómo cambian los errores del modelo cuando **ajustas el umbral** de decisión.

En vez de fijar un umbral fijo (por ejemplo, 0.5 para decidir si algo es positivo o negativo), la curva ROC **explora todos los umbrales posibles** (desde 0 a 1) y **grafica**:

- **Eje Y: TPR** (*True Positive Rate*) → qué tan bien captas los positivos reales.
- **Eje X: FPR** (*False Positive Rate*) → cuántos negativos reales estás confundiendo como positivos.

¿Qué es TPR? (True Positive Rate)

- También llamado **Recall** o **Sensibilidad**.
- Mide el **porcentaje de positivos reales que detectas** correctamente.

- **Fórmula:**
$$TPR = \frac{TP}{TP + FN}$$

Interpretación:

Un TPR alto significa que **casi todos los positivos verdaderos** están siendo capturados.

¿Qué es FPR? (False Positive Rate)

- Mide el **porcentaje de negativos reales que el modelo clasifica erróneamente como positivos**.

- **Fórmula:**
$$FPR = \frac{FP}{FP + TN}$$

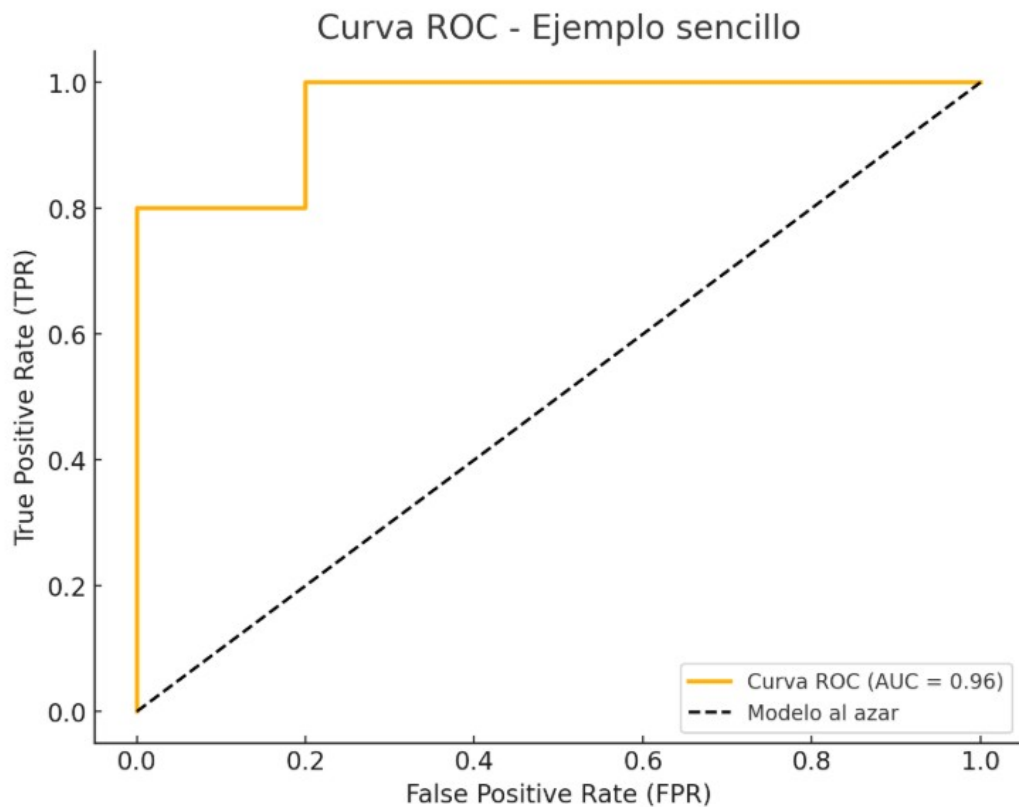
Interpretación:

Un FPR bajo significa que **casi ningún negativo** está siendo confundido como positivo.

Intuición visual:

- **Punto (0,1)** en la curva ROC = **modelo perfecto**: 0% falsos positivos, 100% verdaderos positivos.
- **Diagonal 45°** (de (0,0) a (1,1)) = **modelo al azar**: tan bueno como tirar una moneda.
- **Área bajo la curva (AUC ROC)** mide la calidad general del modelo.
 - Más cerca de 1 = mejor modelo.

Curva ROC - Ejemplo Sencillo



Observa cómo la curva se forma **evaluando distintos umbrales** de decisión sobre las predicciones de probabilidad.

- La línea punteada representa un modelo aleatorio (sin poder de discriminación).
- Nuestra curva real está **por encima** de esa línea, indicando que el modelo es **mejor que el azar**.

Aquí puedes ver cómo **cambian el TPR y el FPR** a medida que variamos el **umbral de decisión** del modelo:

Evolución De TPR Y FPR Según El Umbral

	Threshold	True Positive Rate (TPR)	False Positive Rate (FPR)
1	1.9	0.0	0.0
2	0.9	0.2	0.0
3	0.7	0.8	0.0
4	0.4	0.8	0.2
5	0.35	1.0	0.2
6	0.05	1.0	1.0

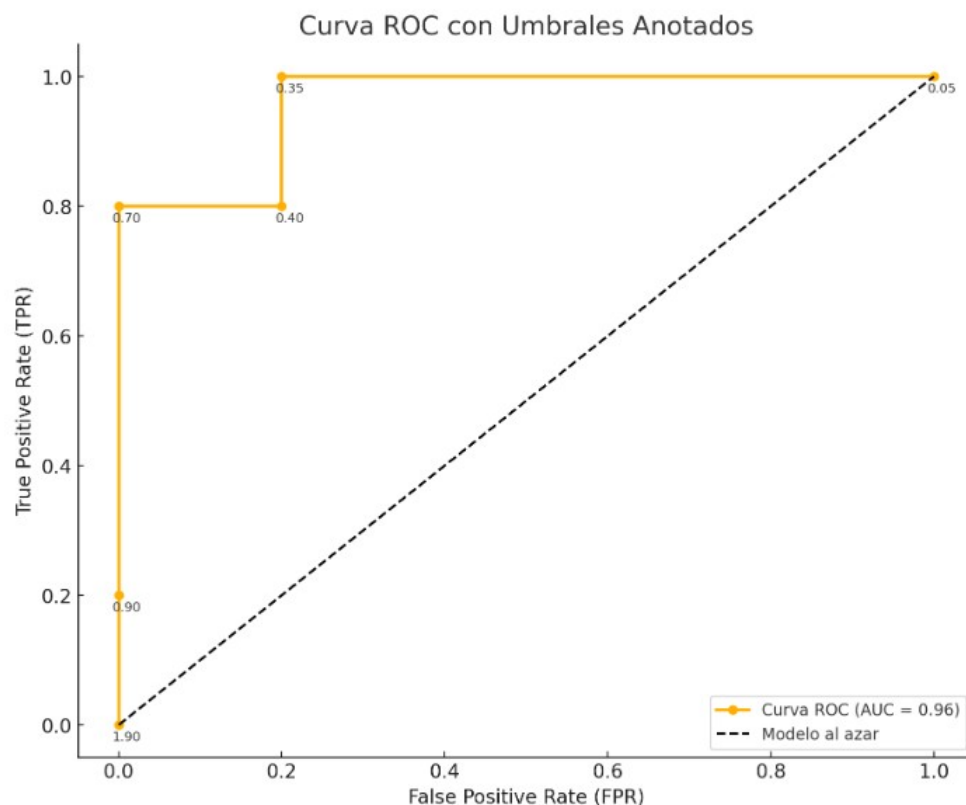
Cuando bajamos el umbral (es decir, somos más permisivos para decir "positivo"), sube el TPR pero también sube el FPR.

En una gráfica ROC:

- Cada punto de la curva corresponde a un umbral diferente.
- Los umbrales no se dibujan explícitamente en la gráfica.
- Lo que ves es **cómo el modelo se comporta** (en términos de **TPR** y **FPR**) a medida que el **umbral varía de 1 a 0**.
- Cuando el umbral es **muy alto** (por ejemplo, 0.9):
 - Solo predices como "positivo" los casos de altísima probabilidad.
 - TPR es bajo (detectas pocos positivos) y FPR es muy bajo.
- Cuando el umbral es **bajo** (por ejemplo, 0.3):
 - Aceptas muchas predicciones como positivas.
 - TPR es alto (detectas casi todos los positivos), pero también el FPR sube.

Así, **al cambiar el umbral**, se traza la curva moviéndote de la esquina inferior izquierda (0,0) hacia la esquina superior derecha (1,1).

Curva ROC Con Umbrales Anotados



¿Cómo elegir el umbral óptimo?

Depende mucho de **qué te importa más** en tu problema:

1. Si quieres más precisión (menos falsos positivos):

- **Meta:** Cada vez que el modelo diga "positivo", quieres que casi siempre acierte.
- **Acción:**
 - **Sube el umbral** (por ejemplo, de 0.5 a 0.7 o 0.8).
 - Así solo etiquetas como positivos los casos con alta confianza.
- **Cuándo aplicarlo:** Cuando un falso positivo es muy costoso (ej., bloquear tarjetas sin razón).

2. Si quieres más recall (menos falsos negativos):

- **Meta:** Capturar todos los verdaderos positivos, aunque cometes más errores.
- **Acción:**
 - **Baja el umbral** (por ejemplo, de 0.5 a 0.3 o 0.2).
 - Así predices más casos como positivos.
- **Cuándo aplicarlo:** Cuando perder un positivo es muy costoso (ej., no detectar un fraude real).

3. Si quieres un balance entre precisión y recall:

- **Meta:** Buen equilibrio entre no cometer muchos errores y no perder positivos.
- **Acción:**
 - Busca el umbral que **maximice el F1-score**.
 - El F1-score es la media armónica de precisión y recall.
- **Método:**
 - Puedes calcular el F1-score para distintos umbrales y elegir el máximo.

4. Método gráfico simple para elegir umbral óptimo (ROC):

- Busca el punto más cercano a la esquina superior izquierda (0,1).
- Ese es el punto donde tienes **máximo TPR** y **mínimo FPR** al mismo tiempo.

Esto a veces se llama "maximizar la sensibilidad y especificidad conjunta".

Resumen rápido:

Objetivo	Acción sobre umbral	Qué se optimiza
Minimizar falsos positivos	Subir el umbral	Mayor precisión
Minimizar falsos negativos	Bajar el umbral	Mayor recall
Balancear ambos	Maximizar F1-score	Equilibrio entre precision y recall

Objetivo	Acción sobre umbral	Qué se optimiza
Mejor compromiso ROC	Punto más cercano a (0,1)	Buen TPR y bajo FPR