

Guía de Métricas de Clasificación en Machine Learning

Cuando se entrena un modelo de clasificación (como regresión logística, KNN, árbol de decisión, etc.), es fundamental evaluar su rendimiento utilizando métricas específicas. Estas métricas ayudan a entender qué tan bien el modelo está prediciendo las clases del problema.

1. Precisión (Precision)

La precisión indica qué proporción de las predicciones positivas fueron realmente correctas.
Fórmula: $\text{Precision} = \text{Verdaderos Positivos} / (\text{Verdaderos Positivos} + \text{Falsos Positivos})$

Ejemplo:

Si el modelo predice que 100 personas tienen experiencia alta, pero solo 80 realmente la tienen, la precisión sería $80 / 100 = 0.80$ (80%).

2. Exhaustividad (Recall o Sensibilidad)

El recall mide qué proporción de los verdaderos positivos fueron detectados por el modelo.
Fórmula: $\text{Recall} = \text{Verdaderos Positivos} / (\text{Verdaderos Positivos} + \text{Falsos Negativos})$

Ejemplo:

Si existen 120 personas con experiencia alta y el modelo detectó correctamente a 80 de ellas, el recall sería $80 / 120 = 0.67$ (67%).

3. F1-score

El F1-score es la media armónica entre precisión y recall. Es útil cuando se quiere un balance entre ambos.

Fórmula: $F1 = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$

Ejemplo:

Si el modelo tiene una precisión de 0.80 y un recall de 0.67, entonces el F1-score sería:
 $F1 = 2 * (0.80 * 0.67) / (0.80 + 0.67) \approx 0.73$

4. Exactitud (Accuracy)

La exactitud indica el porcentaje total de predicciones correctas.

Fórmula: $\text{Accuracy} = (\text{Verdaderos Positivos} + \text{Verdaderos Negativos}) / \text{Total de predicciones}$

Ejemplo: Si el modelo acertó 90 de 122 casos en total, entonces la exactitud es $90 / 122 \approx 0.74$ (74%).

5. Support

Support representa cuántos ejemplos reales hay de cada clase en el conjunto de prueba. No es una métrica de rendimiento, pero ayuda a entender el contexto de las demás métricas.

Ejemplo:

Si hay 63 ejemplos de clase 0 y 59 ejemplos de clase 1, el support para cada clase será 63 y 59 respectivamente.

6. Macro Average (Promedio Macro)

Es el **promedio simple** de una métrica (como precisión, recall o F1-Score) calculado por igual para **cada clase, sin importar cuántos ejemplos tenga cada clase**.

Sirve para evaluar el rendimiento del modelo **tratando a todas las clases por igual**, lo cual es útil cuando **las clases están balanceadas** o cuando **quieres dar la misma importancia a todas las clases**.

Supongamos que tienes 3 clases y sus F1-Scores individuales son:

- **Clase A: 0.90** Clase B: 0.60 **Clase C: 0.30**

Entonces: $\text{Macro F1} = (0.90 + 0.60 + 0.30) / 3 = 0.6$

7 . Weighted Average (Promedio Ponderado)

Es el promedio **ponderado por el número de ejemplos (support)** de cada clase.

Se usa cuando **las clases están desbalanceadas**, ya que da más peso a las métricas de las clases que tienen **más ejemplos**.

Usando el mismo ejemplo anterior con los siguientes supports (número de ejemplos reales por clase):

- Clase A: F1 = 0.90, soporte = 100
- Clase B: F1 = 0.60, soporte = 50
- Clase C: F1 = 0.30, soporte = 50

$\text{Weighted F1} = (100 \cdot 0.90) + (50 \cdot 0.60) + (50 \cdot 0.30) / 100 + 50 + 50 = 90 + 30 + 15 / 200 = 0.675$