

# Métricas de clasificación

Cuando creamos distintos modelos de clasificación (Regresión Logística, RandomForestClassifier, etc) nos interesa conocer si el modelo está clasificando correctamente lo que queremos.

Hay distintas métricas que miden cosas distintas que pueden ayudarnos en algunos casos pero en otros pueden confundirnos, por ende, es muy importante tener bien claro qué es lo que estamos midiendo.

- TP: Cantidad de clasificaciones positivas correctas
- TN: Cantidad de clasificaciones negativas correctas
- FN: Cantidad de clasificaciones negativas incorrectas
- FP: Cantidad de clasificaciones positivas incorrectas

**Exactitud** (accuracy)

$$\frac{TP+TN}{TP+TN+FP+FN}$$

Es la fracción de predicciones que el modelo realizó correctamente sobre el total. La exactitud sola no cuenta la historia completa cuando se trabaja con un **conjunto de datos desequilibrados** en el que hay una disparidad significativa entre el número de etiquetas positivas y negativas.

**Precisión** (precision)

$$\frac{TP}{TP+FP}$$

¿Qué proporción de identificaciones positivas fue correcta? De todas las que predije como positivas, cuántas acerté.

**Recall**

$$\frac{TP}{TP+FN}$$

¿Qué proporción de positivos reales se identificó en forma correcta? De todas las que en realidad pasaron, cuántas acerté.

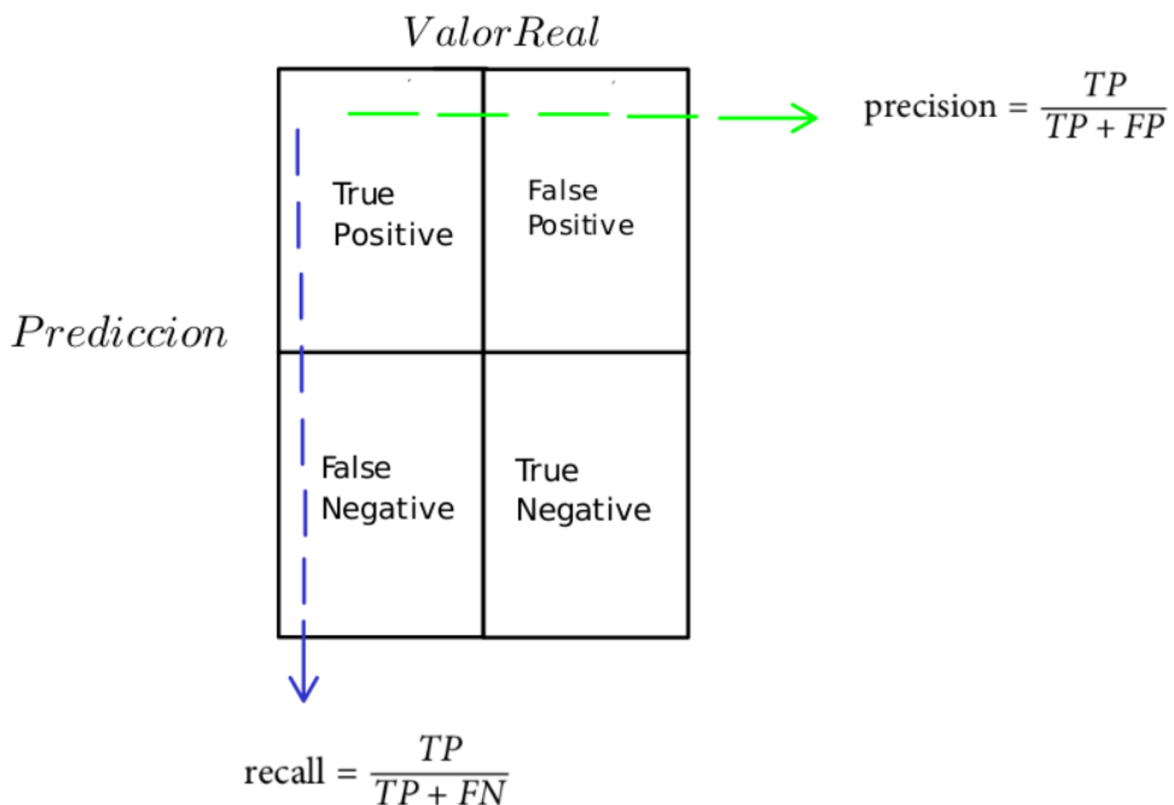
TPR es sinónimo de Recall - (True Positive Rate) ¿Cuál es la proporción de casos acertados sobre el total de casos? o **¿Cuál es la proporción de casos positivos**

**que clasifique como positivos?** Es decir, sobre el total de casos predecidos por el modelo, ¿cuántos fueron TP (acertó correctamente la clasificación)?

**FPR** - (False Positive Rate)

$$\frac{FP}{FP+TN}$$

De todas las que en realidad no pasaron, cuántas no acerté.



### F1-Score




El valor F1 se utiliza para combinar las medidas de precisión y recall en un sólo valor.

$$F1 = 2 * \frac{\text{precisión} \cdot \text{recall}}{\text{precisión} + \text{recall}}$$

F1-score en clasificaciones multiclase: Adoptamos métodos de promediación para el cálculo de la puntuación F1, lo que da como resultado un conjunto de diferentes puntuaciones medias (macro, ponderada, micro) en el informe de clasificación.




- Macro: Calcula métricas para cada etiqueta y encuentre su media no ponderada. Esto no tiene en cuenta el desequilibrio de etiquetas, trata a todas las clases por igual.

Se utiliza cuando tenemos un dataset desbalanceado y queremos tratar a todas las clases de igual manera.

Label	Per-Class F1 Score	Macro-Averaged F1 Score
 <b>Airplane</b>	0.67	$\frac{0.67 + 0.40 + 0.67}{3}$ <b>= 0.58</b>
 <b>Boat</b>	0.40	
 <b>Car</b>	0.67	

- Micro: Calcula métricas globalmente contando el total de verdaderos positivos, falsos negativos y falsos positivos.




Se usa cuando tenemos un dataset balanceado y queremos una métrica fácilmente entendible sobre la performance global sin importar la clase.

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Micro-Averaged F1 Score
 <b>Airplane</b>	2	1	1	$\frac{TP}{TP + \frac{1}{2}(FP + FN)} = \frac{6}{6 + \frac{1}{2}(4 + 4)}$ <b>= 0.60</b>
 <b>Boat</b>	1	3	0	
 <b>Car</b>	3	0	3	
<b>TOTAL</b>	<b>6</b>	<b>4</b>	<b>4</b>	

En el reporte de clasificación, “micro” figura como “accuracy”. Esto es porque micro computa la proporción de las observaciones correctamente clasificadas de todas las observaciones → esta definición es la que usamos para accuracy.

- Weighted (o ponderada): Calcula métricas para cada etiqueta y encuentra su promedio ponderado por soporte (la cantidad de instancias verdaderas para cada etiqueta). Con el promedio ponderado, el promedio de salida habría representado la contribución de cada clase ponderada por el número de ejemplos de esa clase dada.

Se utiliza cuando tenemos un dataset desbalanceado pero queremos darle más peso a las clases con más ejemplos.

Label	Per-Class F1 Score	Support	Support Proportion	Weighted Average F1 Score
 <b>Airplane</b>	0.67	3	0.3	$(0.67 * 0.3) + (0.40 * 0.1) + (0.67 * 0.6)$ <b>= 0.64</b>
 <b>Boat</b>	0.40	1	0.1	
 <b>Car</b>	0.67	6	0.6	
<b>Total</b>	-	10	1.0	

Support refiere a que sólo había esa cantidad de observaciones de esa clase en el data set.