

Precision, Recall, F1, Accuracy en clasificación

[Jose Martinez Heras](#)

Términos es Español

Estas métricas también tienen su correspondiente nombre en español, pero es importante que sepas su nombre en inglés porque muchas librerías ([scikit-learn](#)), las tienen ya implementadas. En esta tabla puedes encontrar la correspondencia.

Inglés

Español

Precision

Precisión

Recall

Exhaustividad

F1-score

Valor-F

Accuracy

Exactitud

Confusion Matrix

Matriz de Confusión

True Positive

Positivos Verdaderos

True Negative

Negativos Verdaderos

False Positive

Positivos Falsos

False Negative

Negativos Falsos

Ejemplo de Marketing

Vamos a utilizar un ejemplo de marketing para entender mejor qué es lo que miden cada una de estas métricas y su significado en el ámbito de los negocios.

Imagínate que estamos llevando la campaña de marketing para un banco. Al banco le interesa vender un fondo de inversión a sus

clientes, porque así pueden ganar dinero por la comisión de gestión.

Podríamos contactar con todos los clientes del banco y ofrecerles el fondo de inversión. Esto es bastante ineficiente porque la mayoría de los clientes no estarán interesados. Sería más eficiente contactar con unos pocos, recoger datos, hacer machine learning y predecir qué otros clientes tienen más probabilidad de aceptar la oferta del banco.

Sin embargo, (casi) ningún modelo de Machine Learning es perfecto. Esto quiere decir que:

- habrá clientes con los que contactaremos porque el modelo ha predicho que aceptarían y en realidad no lo hacen (*False Positive [FP], Positivos Falsos*).
- habrá también clientes con los que no contactaremos porque el modelo ha predicho que no aceptarían que en realidad si lo hubieran hecho (*False Negative [FN], Negativos Falsos*).

El modelo de Machine Learning también acertará (esperemos que mucho). A efectos prácticos esto significa que:

- habrá clientes con los que contactaremos porque el modelo ha predicho que aceptarían y en realidad sí que lo hacen (*True Positive [TP], Positivos Verdaderos*).
- habrá clientes que no contactaremos porque el modelo ha predicho que no aceptarían la oferta y en realidad no lo hacen (*True Negative [TN], Negativos Verdaderos*).

Confusion Matrix (Matriz de Confusión)

Vamos a explicar como funciona la matriz de confusión con un ejemplo hipotético de marketing. En este ejemplo, contactamos a 100 clientes y 80 de ellos nos dicen que no están interesados y 20 de ellos que sí.

Nuestro modelo (en el ejemplo) no es muy bueno, aunque dependiendo de qué métrica usemos podría parecer que es mejor de lo que es.

Hemos utilizado como valores de la clasificación binaria:

- 0: no está interesado
- 1: sí está interesado

		predicción	
		0	1
realidad	0	70	10
	1	15	5

		predicción	
		0	1
realidad	0	TN	FP
	1	FN	TP

Matriz de Confusión con un ejemplo de marketing

En la matriz de confusión de la izquierda podéis ver los valores para este ejemplo. En la matriz de confusión de la derecha, los nombres genéricos cuando usamos la nomenclatura inglesa: *True Negative [TN]*, *True Positive [TP]*, *False Positive [FP]*, *False Negative [FN]*.

Truco: para recordar fácilmente la matriz de confusión:

- *Positivo (Positive) o Negativo (Negative)*: se refiere a la predicción. Si el modelo predice 1 entonces será *positivo*, y se predice 0 será *negativo*.
- *Verdadero (True) o Falso (False)*: se refiere si la predicción es correcta o no.

Precision (Precisión)

Con la métrica de precisión podemos medir la **calidad** del modelo de machine learning en tareas de clasificación. En el ejemplo, se refiere a que la precisión es la respuesta a la pregunta ¿qué porcentaje de los clientes que contactemos estarán interesados?

Para calcular la precisión usaremos la siguiente fórmula:

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

		predicción	
		0	1
realidad	0	TN	FP
	1	FN	TP

Precisión (precision)

En el ejemplo de marketing, siguiendo los datos de la matriz de confusión, tenemos que:

$$precision = \frac{TP}{TP + FP} = \frac{5}{5 + 10} = 0.33$$

Es decir, que sólo un 33% de los clientes a los que contactemos estarán realmente interesados. Esto significa que el modelo del

ejemplo se equivocará un 66% de las veces cuando prediga que un cliente va a estar interesado.

Recall (Exhaustividad)

La métrica de exhaustividad nos va a informar sobre la **cantidad** que el modelo de machine learning es capaz de identificar. En el ejemplo, se refiere a que la exhaustividad (recall) es la respuesta a la pregunta ¿qué porcentaje de los clientes están interesados somos capaces de identificar?

Para calcular la exhaustividad (recall) usaremos la siguiente fórmula:

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

		predicción	
		0	1
realidad	0	TN	FP
	1	FN	TP

Exhaustividad (recall)

En el ejemplo de marketing, siguiendo los datos de la matriz de confusión, tenemos que:

$$recall = \frac{TP}{TP + FN} = \frac{5}{5 + 15} = 0.25$$

Es decir, el modelo sólo es capaz de identificar un 25% de los clientes que estarían interesados en adquirir el producto. Esto significa que el modelo del ejemplo sólo es capaz de identificar 1 de cada 4 de los clientes que sí aceptarían la oferta.

F1

El valor F1 se utiliza para combinar las medidas de precision y recall en un sólo valor. Esto es práctico porque hace más fácil el poder comparar el rendimiento combinado de la precisión y la exhaustividad entre varias soluciones.

F1 se calcula haciendo la media armónica entre la precisión y la exhaustividad:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

En el ejemplo de marketing, combinando precision y recall en F1 nos quedaría:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} = 2 \cdot \frac{0.33 \cdot 0.25}{0.33 + 0.25} = 0.28$$

El valor F1 asume que nos importa de igual forma la precisión y la exhaustividad. Esto no tiene que ser así en todos los problemas. Por ejemplo, cuando necesitamos predecir si hay riesgo de que un trozo de **basura espacial** se choque con un satélite, podemos valorar más la exhaustividad a riesgo de tener una peor precisión. Por eso elegimos F2 en lugar de F1 para esa competición de machine learning.

En este caso podemos usar F2, que es la fórmula genérica del valor F para beta = 2. La fórmula genérica de F es la siguiente:

$$F_{\beta} = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}$$

Accuracy (Exactitud)

La exactitud (accuracy) mide el porcentaje de casos que el modelo ha acertado. Esta es una de las métricas más usadas y favoritas ... que te recomiendo evitar! El problema con la exactitud es que nos puede llevar al engaño, es decir, puede hacer que un modelo malo (como el del ejemplo) parezca que es mucho mejor de lo que es.