



¿Qué buscas?

English version

[Click here](#) for an automatically translated version

Lo último

[Machine Learning en Operaciones Espaciales](#)

[Segmentación de Imágenes con Redes Convolucionales](#)

[Algoritmos Genéticos y Memoria Visual](#)

[TorchServe para servir modelos de PyTorch](#)

[Detección de anomalías en espacio](#)

[Inicio](#) > [machine learning](#) > Precision, Recall, F1, Accuracy en clasificación

# Precision, Recall, F1, Accuracy en clasificación

Actualizado 09/10/2020 | por [Jose Martinez Heras](#)

Cuando necesitamos evaluar el rendimiento en [clasificación](#), podemos usar las métricas de precision, recall, F1, accuracy y la matriz de confusión. Vamos a explicar cada uno de ellos y ver su utilidad práctica con un ejemplo.

## Índice

1. Términos es Español
2. Ejemplo de Marketing
3. Confusion Matrix (Matriz de Confusión)
4. Precision (Precisión)
5. Recall (Exhaustividad)
6. F1
7. Accuracy (Exactitud)
8. Calculando precision, recall, F1, accuracy en python con scikit-learn
9. Resúmenes
10. Recursos

## 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

## Contenido relacionado



[15 Librerías de Python para Machine Learning](#)

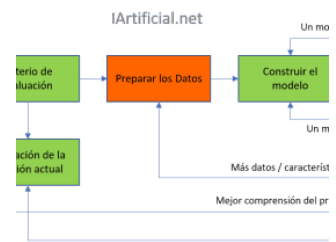


[Análisis de Errores en Machine Learning](#)

True Positive	Positivos Verdaderos
True Negative	Negativos Verdaderos
False Positive	Positivos Falsos
False Negative	Negativos Falsos



Error Cuadrático Medio para Regresión



Las 7 Fases del Proceso de Machine Learning

## 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 sí 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

realidad	0	1
	70	10
0	15	5
1		

realidad	0	1
	TN	FP
0	FN	TP
1		

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 equivocaré 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:

		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.

El accuracy (exactitud) se calcula con la siguiente fórmula:

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

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{5 + 70}{5 + 70 + 10 + 15} = 0.75$$

Es decir, el modelo acierta el 75% de las veces. Como ves, la exactitud es una métrica muy engañosa. De hecho, si tuviésemos un modelo que siempre predijera que el cliente nunca va a estar interesado, su accuracy sería del 80%.

**Peligro:** la métrica accuracy (exactitud) no funciona bien cuando las clases están desbalanceadas como es en este caso. La mayoría de los clientes no están interesados en la oferta, así que es muy fácil acertar diciendo que no lo van a estar. Para problemas con clases desbalanceadas es mucho mejor usar precision, recall y F1. Estas métricas dan una mejor idea de la calidad del modelo.

## Calculando precision, recall, F1, accuracy en python con scikit-learn

La librería de python [scikit-learn](#) implementa todas estas métricas. Para usarlas sólo tienes que seguir sus instrucciones:

- `sklearn.metrics.confusion_matrix`
- `sklearn.metrics.precision_score`
- `sklearn.metrics.recall_score`
- `sklearn.metrics.f1_score`
- `sklearn.metrics.accuracy_score`

Precision, Recall, F1, Accuracy y la Matriz de Confusión son métricas de **clasificación**. El artículo [Error Cuadrático Medio para Regresión](#) explica las métricas de **regresión**.

## Resumen

En este artículo hemos visto cuáles son las métricas más extendidas para evaluar el rendimiento de modelo supervisado en tareas de clasificación.

Hemos destacado que la métrica **accuracy** (exactitud) es engañosa cuando las clases están desbalanceadas. Cuando decimos que es engañosa nos referimos a que nos hace creer que el modelo es mejor de lo que en realidad es. De hecho, he puesto un ejemplo de un modelo relativamente malo para poner de manifiesto este ejemplo.

Las medidas de precision, recall y F1 son mucho más representativas y funcionan tanto si las clases están balanceadas como si no:

- **Precision** nos da la calidad de la predicción: ¿qué porcentaje de los que hemos dicho que son la clase positiva, en realidad lo son?
- **Recall** nos da la cantidad: ¿qué porcentaje de la clase positiva hemos sido capaces de identificar?
- **F1** combina Precision y Recall en una sola medida
- La **Matriz de Confusión** indica qué tipos de errores se cometen

## Recursos

- [Métricas](#) en scikit-learn que incluyen las descritas y muchas más
- [Vídeo](#) (en inglés) donde explico estas métricas
- [Artículo](#) en la wikipedia

## Relacionado

1. [Inteligencia Artificial aplicada a meneame.net](#) En este artículo, vamos a aplicar Inteligencia Artificial a todas las noticias de portada de meneame.net en 2018. Empezaremos realizando un análisis estadístico y visualización de datos. Después usaremos Procesamiento...
2. [Clustering \(Agrupamiento\), K-Means con ejemplos en python](#) El clustering consiste en la agrupación automática de datos. Es un tipo de aprendizaje automático no-supervisado. En castellano se denomina agrupamiento. Vamos a ver en más detalle en qué consiste...
3. [Análisis de Errores en Machine Learning](#) El análisis de errores es una de las fases del proceso de machine learning más importantes. El análisis de errores nos va a permitir saber qué hacer para mejorar el...
4. [Aguathon: mi solución al primer Hackathon del Agua](#) En este artículo explico mi solución al Aguathon: el primer Hackathon del Agua, organizado por ITAINNOVA. Introducción al Aguathon: el Hackathon del Agua de ITAINNOVA El Instituto Tecnológico de Aragón...

📁 [machine learning](#)

🔗 [accuracy](#), [aprendizaje supervisado](#), [clasificación](#), [confusion matrix](#), [exhaustividad](#), [f1](#), [matriz de confusión](#), [precision](#), [recall](#)

◀ [Basura Espacial: competición con machine learning](#)

▶ [Clustering \(Agrupamiento\), K-Means con ejemplos en python](#)

## 8 comentarios en «Precision, Recall, F1, Accuracy en clasificación»



**Lita**

03/06/2020 a las 03:59

Hola, muy bueno y didáctico, una pregunta . Porque no usas la curva AUC Roc Como medida también de rendimiento?, gracias de antemano

[Responder](#)

**Jose Martinez Heras**

03/06/2020 a las 10:24

Hola Lita, pensaba hablar de la curva ROC (Receiver Operating Characteristic) y AUC (Area Under the Curve) en otro post ... aunque ahora que lo comentas, también podría venir bien aquí. Seguramente sea un nuevo post y lo enlace desde aquí.

[Responder](#)

**gabriel**

13/08/2020 a las 11:26

gracias, se entiende muy bien

[Responder](#)

21/09/2020 a las 00:52

Estimado

y que pasa en un sistema de recomendación, es recomendable tomar en cuenta la exactitud?

[Responder](#)

**Manuel**

04/04/2021 a las 12:12

Hola

Muy didáctico el artículo.

No me ha quedado claro cuando comentas de evitar el uso de accuracy, es siempre o solo cuando los datos estan desbalanceados?

O se debería de usar el accuracy junto a las demas?

Gracias

[Responder](#)

**Jose Martinez Heras**

05/04/2021 a las 20:55

Cuanto más desbalanceadas estén las clases, más problemático será el uso del accuracy. Para clasificación binaria, si tienes 50% – 50% funcionaría bien. Sin embargo, en la práctica las clases nunca están balanceadas.

Además, cuando comuniques cual es el accuracy de una solución, tienes el inconveniente añadido que tienes que especificar la distribución de las clases para que el que te escuche (o lea) pueda entender realmente lo bien que funciona.

Yo prefiero usar precision y recall, incluso más que sólo F1, porque tengo mayor visibilidad de cuales son las fortalezas y debilidades de la solución.

[Responder](#)

**Ruben**

24/06/2022 a las 21:49

muy buena explicacion, todas las funciones de sklearn como «sklearn.metrics.precision\_score()» tienen un parametro average{'micro', 'macro', 'samples', 'weighted', 'binary'} or None, default='binary' que significa ese parametro no acabo de entender la explicacion oficial.

[Responder](#)

29/10/2022 a las 21:02

Hola

Si queremos aplicar un modelo de clasificación a datos desbalanceados se suele recomendar:

- Utilizar pesos distintos para cada clase o utilizar una matriz de costes.
- No utilizar la Accuracy sino métricas más robustas, como la F1 o mejor aún la AUC ROC o mejor aún la AUC PR.

Mi pregunta es... ¿debemos utilizar sólo una de estas recomendaciones o ambas simultáneamente?

Es decir, si utilizamos la AUC\_PR... ¿tenemos que seguir aplicando pesos a la entrada?

y viceversa,

Si estamos utilizando pesos ¿sigue siendo desaconsejable usar la Accuracy?

[Responder](#)

## Deja un comentario



Publicar comentario

### Más Información

[Inicio](#)

[Guía rápida](#)

[Sobre mí](#)

[Contacto](#)

### Legal

[Términos y condiciones](#)

[Política de cookies](#)

[Política de privacidad](#)