

Apuntes de clase - Semana 6 - 14 de Marzo

María José Porras Maroto
Tecnológico de Costa Rica, Escuela de Ingeniería en Computación
Inteligencia Artificial - IC6200 - Grupo 2
Email: marijopm27@estudiantec.cr

I. MÉTRICAS

Las métricas son las medidas utilizadas para indicar el rendimiento de un modelo predictivo. Además, las métricas resultan ser la forma más objetiva de evaluar y comparar un modelo.

A. Matriz de Confusión

La matriz de confusión, también llamada matriz de error, es una tabla resumida empleada para evaluar la eficacia de un modelo de clasificación. Se utiliza para visualizar y analizar el rendimiento del modelo al predecir las clases de un conjunto de datos, permitiendo identificar los aciertos y errores en las predicciones.

- Problema de clasificación binaria donde tenemos un evento positivo (1) y un evento negativo (0).
- El modelo predice la clase $P(1)$ y efectivamente eran de la clase $P(1)$.
- El modelo predice la clase $N(0)$ y efectivamente eran de la clase $N(0)$.
- El modelo predice la clase $P(1)$ cuando tenía que predecir la clase $N(0)$ - En pruebas de hipótesis se le conoce como error tipo 1.
- El modelo predice la clase $N(0)$ cuando tenía que predecir la clase $P(1)$ - En pruebas de hipótesis se le conoce como error tipo 2.

Target class	P	N
Predicted class		
P	TP	FP (Type I)
N	FN (Type II)	TN

Fig. 1. Matriz de Confusión

B. Accuracy

Nivel de clasificación correcta entre todos los intentos realizados. Su utilidad se ve reflejada cuando los errores por clase son igual de importantes, pero puede no ser una métrica de evaluación suficiente.

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

C. Precision

Mide los errores tipo 1. La tasa de predicciones positivas correctas entre todas las predicciones positivas

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

D. Recall

Mide los errores tipo II. También puede definirse como la tasa de predicciones correctas entre todos los ejemplos positivos del dataset.

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

E. F1-Score

Esta métrica es comúnmente usada en problemas de clasificación, especialmente cuando tenemos desequilibrio de clases.

$$\frac{2 * precision * recall}{precision + recall}$$

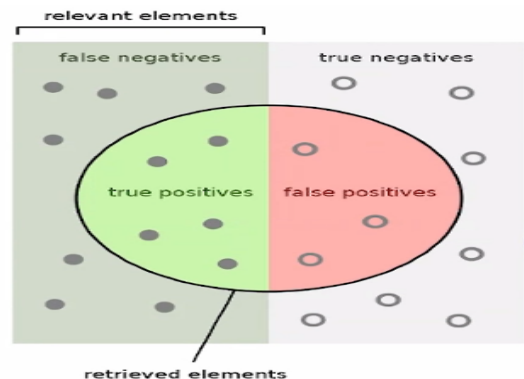


Fig. 2. F1 Score

II. CASO DE ESTUDIO

Escenario de estudio donde de un dataset de 1000 pacientes, se busca generar un diagnóstico para identificar si tienen o no cáncer.

- Total de pacientes: 1000
- Pacientes con cáncer (Clase Positiva): 30
- Pacientes sin cáncer (Clase Negativa): 970

Los valores de salida obtenidos:

- TP: 25 pacientes con cáncer correctamente identificados
- FN: 5 pacientes con cáncer que el modelo clasificó incorrectamente como no cancerosos
- TN: 950 pacientes sin cáncer correctamente identificados.
- FP: 20 pacientes sin cáncer que el modelo clasificó incorrectamente como cáncer.

Métricas obtenidas

1) *Matriz de confusión*: A continuación, se presenta la matriz de confusión correspondiente a la predicción de cáncer sobre el ejercicio de caso de uso:

TABLE I
MATRIZ DE CONFUSIÓN PARA LA PREDICCIÓN DE CÁNCER

	Cáncer	No cáncer
Predicción Cáncer	25	5
Predicción No cáncer	20	950

2) *Accuracy*:

$$\frac{TP + TN}{TOTALPACIENTES} = \frac{25 + 950}{1000} = 0,975 = 97.5$$

A pesar de ser un buen modelo, considerando que su objetivo

principal resulta el diagnóstico de pacientes con cáncer, cómo lo hizo?

3) *Recall*:

$$\frac{TP}{TP + FN} = \frac{25}{25 + 5} = 0,833 = 83.3$$

En este caso se tiene una visión específica del modelo

identificando la clase cáncer. En este punto es en el que se puede observar que a pesar de tener alto accuracy, la métrica de recall indica que existe un espacio para la mejora.

4) *Precision*:

$$\frac{TP}{TP + FP} = \frac{25}{25 + 20} = 0,55 = 55$$

Se puede notar sobre esta métrica que el nivel de precision

con el que se cuenta es realmente bajo y supone otro punto de mejora sobre el modelo.

5) *F1- Score*:

$$\frac{2 * precision - recall}{precision + recall} = \frac{2 * 0,55 * 0,833}{0,55 + 0,833} = \frac{0,9163}{1,383} = 0,662 = 66,2$$

El F1-Score permite explorar que realmente la capacidad del modelo para clasificar de manera correcta las clases minoritarias no es buena y que debe ser mejorada.

A. ROC y AUC

Receiver Operating Characteristic : La curva ROC se emplea para evaluar el desempeño de algoritmos de clasificación binaria, donde se distinguen entre dos clases o categorías (por ejemplo, 1 o 0, Verdadero o Falso, etc.). En lugar de ofrecer un valor único como la mayoría de las métricas, la curva ROC ofrece una representación visual que muestra la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos.

Area Under the Curve: permite evaluar la capacidad de discriminación del modelo a la hora de hacer clasificaciones.

Si se piensa, por ejemplo, en una regresión logística, en esta se evalúa la tasa de falsos positivos y verdaderos positivos en relación con un umbral. La premisa es que existe una relación entre el umbral y estas dos tasas; a medida que el umbral varía, también lo hacen los falsos positivos. El objetivo es maximizar el área bajo la curva (AUC) para que se acerque a 1, lo que indicaría que el modelo está identificando correctamente los casos positivos y minimizando los errores.

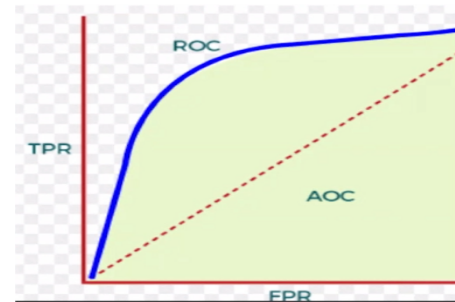


Fig. 3. ROC - AUC

REFERENCES

- [1] Cañadas, R. (2021) Curvas ROC — Usos de ROC-AUC en machine learning. Abdatum. <https://abdatum.com/ciencia/curvas-roc>
- [2] Lee, A. (2020) Comprensión de la Matriz de Confusión y Cómo Implementarla en Python. DataSource.ai. <https://www.datasource.ai/es/data-science-articles/comprension-de-la-matriz-de-confusion-y-como-implementarla-en-python>
- [3] Sanahuja, P.M. (2021).Entendiendo la curva ROC y el AUC: Dos medidas del rendimiento de un clasificador binario que van de la mano. <https://polmartisanahuja.com/entendiendo-la-curva-roc-y-el-auc-dos-medidas-del-rendimiento-de-un-clasificador-binario-que-van-de-la-mano/>