

# Matriz de confusión

Evaluación de modelos.

# Matriz de confusión

Para poder utilizar este método de evaluación de un modelo de clasificación necesitamos separar nuestra data de entrenamiento en dos datasets.

**Train (80%)**

**Test (20%)**

Lo que hacemos es entrenar el modelo utilizando las observaciones que están en el dataset train y luego evaluaremos el modelo utilizando las observaciones del dataset test.

Esto nos ayudará a medir cómo se comporta nuestro modelo cuando se lo aplicamos a data nueva. La matriz tiene la siguiente estructura.

## Matriz de confusión

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

**VP** es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo.

**VN** es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo.

**FN** es la cantidad de positivos que fueron clasificados incorrectamente como negativos.

**FP** es la cantidad de negativos que fueron clasificados incorrectamente como positivos.

# Matriz de confusión

## Exactitud (Accuracy)

En general, qué porcentaje de la data clasifica correctamente?

$$\text{Exactitud} = \text{VP} + \text{VN} / \text{Total}$$

## Tasa de error (Misclassification Rate)

En general, qué porcentaje de la data clasifica incorrectamente?

$$\text{Tasa de error} = \text{FP} + \text{FN} / \text{Total}$$

# Matriz de confusión

## Sensibilidad, exhaustividad, Tasa de verdaderos positivos

Traducción al inglés,

- Recall
- Sensitivity
- True Positive Rate

Cuando la clase es positiva, qué porcentaje logra clasificar?

$$\text{Sensibilidad} = \text{VP} / \text{Total Positivos}$$

## Especificidad, tasa de verdaderos negativos

Traducción al inglés,

- Specificity
- True Negative Rate

Cuando la clase es negativa, qué porcentaje logra clasificar?

$$\text{Especificidad} = \text{VN} / \text{Total Negativos}$$

# Matriz de confusión

## Precisión

Cuando predice positivos, que porcentaje clasifica correctamente?

$$\text{Precisión} = \text{VP} / \text{Total clasificados positivos}$$

## Valor de predicción negativo

Cuando predice negativo, que porcentaje clasifica correctamente?

$$\text{VPN} = \text{VN} / \text{Total clasificados negativos}$$

## Matriz de confusión

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	