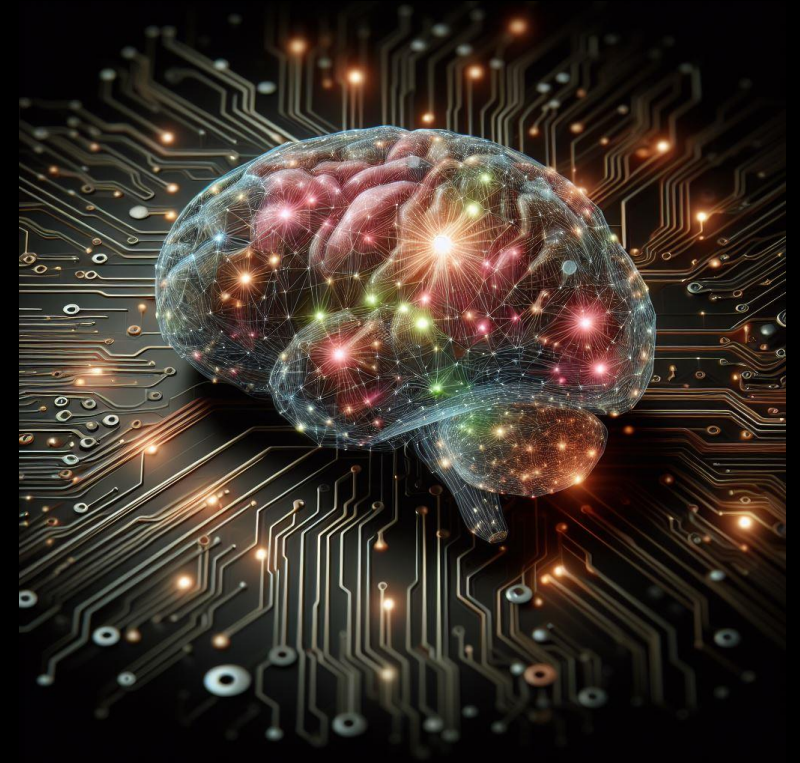
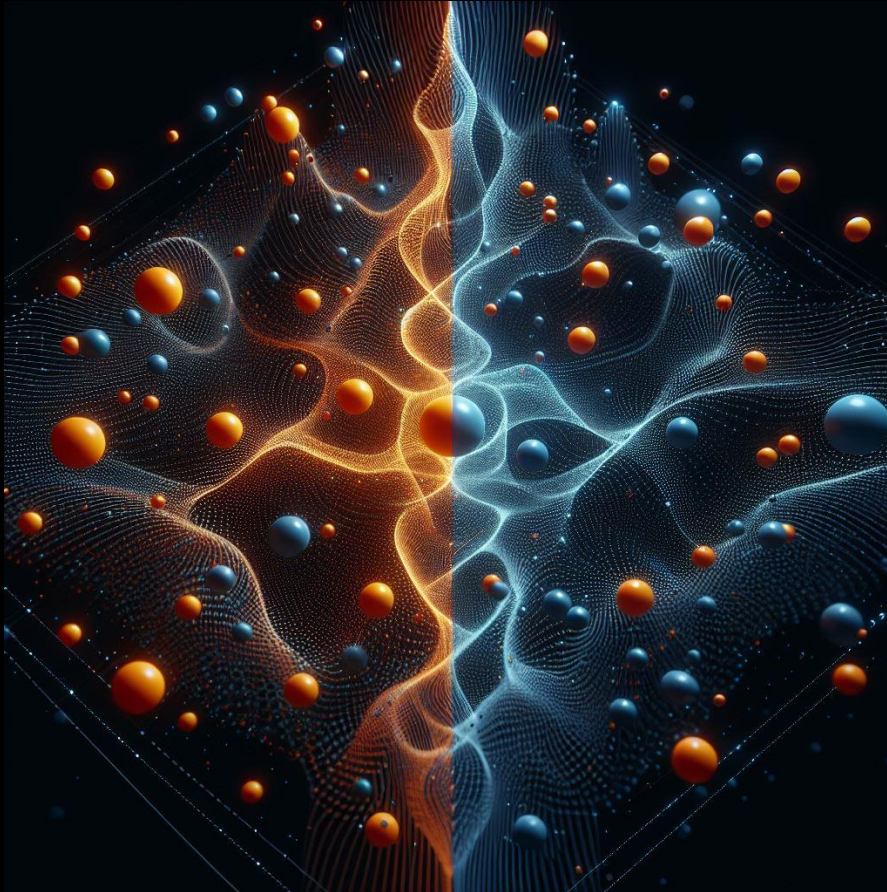


Clasificación Binaria

*Índices de Desempeño o
Métricas. Matriz de confusión.*



DeepData

¿Para qué las métricas de los clasificadores?



- Ayudan a comprender cómo se desempeña un clasificador.
- Útiles para comparar distintos tipos de clasificadores.
- Selección de hiperparámetros.
- Análisis del comportamiento en distintos conjuntos de datos.



Métricas para clasificadores binarios



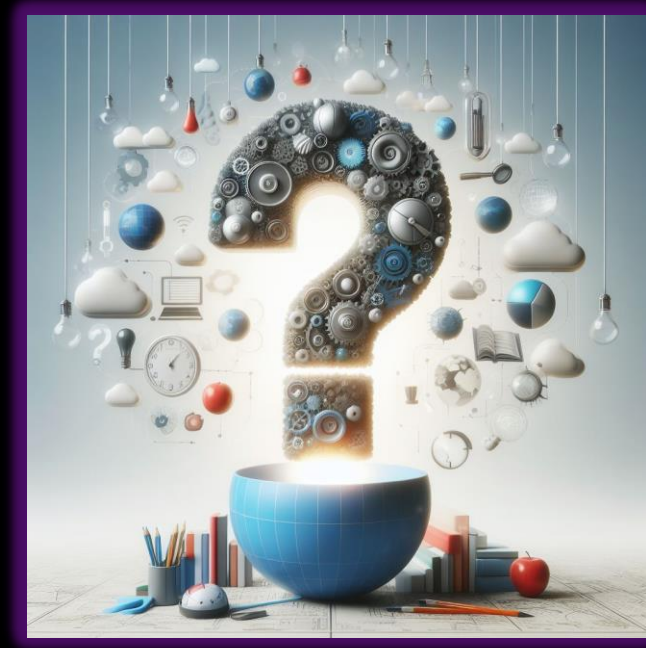
- *Accuracy*
 - *Precision*
 - *Recall* o *Sensitivity*
 - F1
-
- Matriz de confusión
 - Curva ROC y AUC
 - Curva PR y AUPRC



¿Qué métricas utilizar?



- ¿Qué representa la métrica?
- ¿Qué características del clasificador son invisibles a la métrica?
- Selección de métrica en base a la aplicación del clasificador.
- Uso de más de una métrica.
- Es la manera de presentar tu clasificador.



Verdaderos Positivos (TP), Falsos Positivos (FP), Verdaderos Negativos (TN) y Falsos Negativos (FN)

- TP (True Positive): Cantidad de instancias positivas que son clasificadas (correctamente) como positivas.
- FP (False Positive): Cantidad de instancias negativas que son clasificadas (incorrectamente) como positivas.
- TN (True Negative): Cantidad de instancias negativas que son clasificadas (correctamente) como negativas.
- FN (False Negative): Cantidad de instancias positivas que son clasificadas (incorrectamente) como negativas.



Verdaderos Positivos (TP), Falsos Positivos (FP), Verdaderos Negativos (TN) y Falsos Negativos (FN)

- TP (True Positive): Cantidad de instancias positivas que son clasificadas (correctamente) como positivas.
- FP (False Positive): Cantidad de instancias negativas que son clasificadas (incorrectamente) como positivas.
- TN (True Negative): Cantidad de instancias negativas que son clasificadas (correctamente) como negativas.
- FN (False Negative): Cantidad de instancias positivas que son clasificadas (incorrectamente) como negativas.

Practiquemos esto con un ejemplo en Python!



Matriz de Confusión



		Predicho	
		+	-
Verdadero	+	TP	FN
	-	FP	TN
		<i>predichos positivos</i> = $TP + FP$	<i>predichos negativo</i> = $FN + TN$

$$\text{conocidos positivos} = TP + FN$$

$$\text{conocidos negativos} = FP + TN$$

Ejemplo



- Se armará la matriz de confusión desde cero con los valores de TP, FP, TN y FN previamente obtenidos.
- Se comparan los resultados obtenidos aplicando los conocimientos aprendidos con funciones implementadas en scikit-learn.
- Vamos a Python!



Conclusiones y continuación



- Conocemos la importancia de las métricas de clasificación.
- Pudimos computar los valores de TP, FP, TN y FN y armar la matriz de confusión.
- Estamos en condiciones de definir e interpretar el resto de las métricas.

