

Caruzulca Tiylla Alex Eli

Evaluación de un modelo de clasificación con un filtro de spam

Supongamos que queremos clasificar nuestros correos en spam (correo no deseado) y no spam (correo deseado).

La matriz de confusión calculada es la siguiente:

	Predicted negative (non-spam)	Predicted positive (spam)
Negative class (non-spam)	264	14
Positive class (spam)	22	158

Análisis

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + TP + FN}$$
$$= \frac{158 + 264}{158 + 264 + 14 + 22} = \frac{422}{458} = 0.922$$

$$\text{Precisión} = \frac{TP}{TP + FP}$$
$$= \frac{158}{158 + 14} = \frac{158}{172} = 0.918$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad \Delta \quad \text{Sensibilidad}$$
$$= \frac{158}{158 + 22} = \frac{158}{180} = 0.878$$

$$F1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$= 2 \times \frac{0,918 \times 0,878}{0,918 + 0,878} = 0,898$$

True positive Especificidad

$$FPR = \frac{FP}{FP + TN}$$

$$= \frac{264}{264 + 14} = 0,949$$

Comparación Sensibilidad y Especificidad.

0,878      0,949

Conclusión

Al superar el umbral, podemos concluir que el modelo tiene un buen rendimiento con alta precisión.