Classification formula reference

Notes for SHS | Data Science | Topic 87 | 25 November 2026

Synthetic example: TP=24, FP=6, FN=8, TN=62

The 100-case confusion matrix is fictional teaching arithmetic, NOT the actual Iris model result.

Metric	Formula	Synthetic result
Accuracy	(TP+TN)/N	86.00%
Precision	TP/(TP+FP)	80.00%
Recall	TP/(TP+FN)	75.00%
Specificity	TN/(TN+FP)	91.18%
False positive rate	FP/(FP+TN)	8.82%
F1	2PR/(P+R)	77.42%
Balanced accuracy	(Recall+Specificity)/2	83.09%
Gini impurity	1 - sum(class_proportion squared)	0.48 for 6:4 node

Use the right context

For multiclass data, define class-specific one-versus-rest precision/recall and explicitly choose macro, micro or support-weighted averages. If a denominator is zero, document the chosen convention rather than silently treating the undefined fraction as a normal measured value.

Probability rules

Sigmoid(z)=1/(1+exp(-z)); softmax_k=exp(z_k)/sum(exp(z_j)). A score is not necessarily calibrated probability. Choose thresholds using training or validation information only.