

Evaluation card

The confusion matrix, the six numbers and how to say each one.

	Missed the visit	Attended
Flagged	True positive	False positive, a wasted visit
Not flagged	False negative, the harm	True negative

Number	Question it answers	Say it as
Base rate	How common is the outcome	About a third of women miss
Sensitivity	Of those who missed, how many were flagged	The list caught seven in ten of the women who missed
Specificity	Of those who attended, how many were left alone	Nine in ten attending women were not visited unnecessarily
Positive predictive value	Of those flagged, how many missed	Six in ten visits reached a woman who would have missed
Area under the curve	How well does it rank overall	Report it, then report the operating point
Calibration	Do the scores mean what they say	Needed before any score is shown to a person

- Evaluate at the capacity of the service, for example precision at forty.
- Compare with the base rate and with the simple rule.
- A decision curve says whether the model changes any decision.

Based on the ERAS Africa Machine Learning in Data Analysis course, 2026, the WHO guidance on ethics and governance of artificial intelligence for health, TRIPOD+AI, PROBAST and the scikit-learn documentation. Free to copy and adapt with attribution.