

Model choice card

When a tree, a forest, a boosted model or a regression, and what each costs in explainability.

Family	Tends to win when	Costs
Logistic regression	Small data, additive signal, must be explained or implemented by hand	Misses strong interactions
Decision tree	A readable flowchart is needed for a health post	Unstable alone, overfits without depth limits
Random forest	Interactions matter and there is enough data	Nobody can read it, explanation moves to importance
Gradient boosting	Competitions and large tabular data	Tuning, transparency, and it needs a maintained computer

- Compare on identical folds with a metric matched to the decision, and report the spread.
- A difference smaller than the fold to fold spread is not a difference. Break ties by explainability and fit for the setting.
- Importance is reliance, not cause. Compute it on held out data and watch for shared credit and proxies.

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.