Student Model Fairness Audit Checklist

Notes for SHS · Data Science · Topic 62 · Original synthetic demonstration

1. Frame the decision and harm

Define intended use, who may be affected, what positive/negative predictions mean, who is accountable, and whether automation is justified.

2. Inspect data provenance

Confirm lawful access, consent, data quality, sample coverage, label validity, missing values, proxies and participation of relevant stakeholders.

3. Separate training, validation and test

Fit transformers on training data only; choose models and mitigations on validation folds or validation data; keep final held-out evaluation locked.

4. Calculate group-based metrics

For each adequately sized group report n, actual positives, predicted positives, TP, TN, FP, FN, selection rate, accuracy, TPR, FNR and FPR. Show uncertainty where appropriate.

5. Compare fairness definitions

Demographic parity compares selection rates. Equal opportunity compares TPRs. Equalized odds compares TPRs and FPRs. Report tradeoffs and say why a chosen measure fits the application.

6. Investigate—not just optimize

Examine measurement errors, sample bias, label artifacts, pipeline assumptions, model thresholds and organizational consequences. Avoid pretending one score is a legal ruling.

7. Check intervention consequences

Record the changed feature/pipeline, validation selection, untouched test performance, new burdens, costs, privacy risks and remaining disparities.

8. Document human oversight

Include appeal or correction procedures, monitoring owners, incident reporting, periodic reviews and a decision to pause or reject automation if harms remain.

Simulation disclaimer

The downloadable dataset contains only 2,400 artificially generated records. The two groups are deliberately fictional and have no demographic interpretation.