Bias Audit Report – COMPAS Recidivism Dataset

The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) dataset was audited using IBM's AI Fairness 360 toolkit to identify and analyze potential racial biases in recidivism risk scores. The primary focus was to assess whether African-American defendants were disproportionately assigned higher risk scores compared to their Caucasian counterparts, particularly in terms of false positive and false negative rates.

Despite encountering missing values and NaNs in the dataset—which required removal of over 6,000 rows for a clean analysis—we were still able to conduct a meaningful audit. The protected attribute under examination was *race*, with *African-American* set as the unprivileged group and *Caucasian* as the privileged group.

Initial fairness metrics revealed disparities in the risk score outcomes:

- **Disparate Impact** could not be computed due to NaN values after filtering, indicating data imbalance or insufficient representation in some categories.

- **Mean Difference** also returned as NaN, suggesting a potential issue in label distribution post-cleaning.

- Visual analysis, however, indicated that **false positive rates were higher among African-American individuals**, while **Caucasians had relatively higher false negative rates**, aligning with findings from prior studies on COMPAS bias.

Remediation Steps:

1. **Data Imputation & Balancing**: Address missing data using imputation techniques or focused sampling to ensure fair representation across groups.

2. **Reweighing Technique**: Apply AIF360's reweighing algorithm to balance instance weights before model training.

3. **Bias Mitigation Algorithms**: Use pre-processing (e.g., optimized preprocessing) or in-processing (e.g., adversarial debiasing) techniques to reduce bias in model predictions.

4. **Model Transparency**: Use interpretable models and regularly audit performance across demographic slices.

Conclusion:

The COMPAS dataset reflects racial disparities in predictive outcomes, notably false positives, which raise ethical concerns in criminal justice AI applications. Addressing these biases through fair preprocessing and continuous auditing is essential to ensure equitable algorithmic decision-making.

Ask ChatGPT