COMPAS Dataset Fairness Audit – Report

Summary of Findings

The COMPAS recidivism dataset was analysed to examine potential racial and demographic disparities in algorithmic risk scoring. After filtering, the dataset included 6,172 individuals. Demographic analysis showed that African - American defendants represented the largest group (51.44%), followed by Caucasian defendants (34.07%). Males made up a significant majority (80.96%), and the overall two-year recidivism rate was approximately 45.51%.

Exploring decile risk scores revealed clear disparities: African - American defendants were more frequently assigned higher scores (7–10) compared to Caucasian defendants. A logistic regression model predicting "High Score" demonstrated statistically significant predictors. Race remained influential even after controlling for age, priors, and charge degree. The odds ratio indicated that African - American defendants had ~45% higher odds of receiving a "High Score" than Caucasian defendants. Additionally, individuals under 25 were found to have ~2.5 times greater odds of being classified high risk, showing age as another strong factor.

These findings confirm the presence of algorithmic bias in COMPAS risk assessment, with African - American defendants disproportionately flagged as higher risk. Such disparities raise serious ethical concerns regarding fairness, justice, and trust in algorithmic decision-making within the criminal justice system.

Remediation Steps

1. **Fairness-Aware Machine Learning**: Integrate techniques such as reweighing or adversarial debiasing to minimize disparate impact while maintaining predictive performance.
2. **Ongoing Algorithmic Audits**: Establish regular audits using fairness metrics (e.g., disparate impact ratio, equal opportunity difference) to track bias over time.
3. **Feature Scrutiny**: Reassess input features to ensure they are not proxies for race or other protected attributes.
4. **Transparency & Explainability**: Document scoring logic and provide interpretable outputs so stakeholders can understand how risk scores are generated.
5. **Human Oversight**: Use COMPAS as a supplementary tool, not as the sole determinant of judicial decisions, ensuring contextual review in each case.

Conclusion

This audit highlights how bias can persist in widely used AI systems. Addressing these disparities requires a combination of **technical interventions, policy safeguards, and ethical governance** to ensure that risk assessment tools are fair, accountable, and trustworthy.