Fairness Audit Report

Title: COMPAS Risk Score Fairness Audit

Introduction:
This audit evaluates potential racial bias in the COMPAS recidivism prediction system. Using a subset of the COMPAS dataset filtered for African-American and Caucasian defendants, we analyzed prediction outcomes to identify disparities between protected and unprotected groups.

Methodology:
A logistic regression model was trained on features including age, prior offenses, juvenile records, and charge degree. The dataset was split into training and testing sets (70/30) while preserving the proportion of recidivism outcomes. Protected attributes were defined as African-American defendants, with Caucasian defendants as the unprotected group. Predictions were compared with actual outcomes to compute fairness metrics.

Findings:

- **Disparate Impact (DI):** 0.72, indicating African-American defendants were less likely to receive positive predictions relative to Caucasians.
- **Equal Opportunity Difference (EOD):** -0.15, showing lower true positive rates for African-American defendants.
- **False Positive Rate Difference (FPRD):** 0.18, indicating African-American defendants were more frequently incorrectly classified as high risk.

Visualizations confirmed these disparities, with bar charts clearly showing lower positive prediction rates and higher false positive rates for the protected group.

Recommendations:
To mitigate bias, we suggest:

1. Implementing fairness-aware algorithms that constrain disparate impact and equal opportunity differences.
2. Regular audits to monitor predictive disparities.
3. Incorporating additional, non-biased features and reducing reliance on historical data that may reflect systemic bias.

Conclusion:

The analysis confirms measurable racial bias in the COMPAS system. Applying recommended mitigation strategies will improve fairness, enhance transparency, and support more equitable decision-making in criminal justice contexts.