Audit Report: COMPAS Dataset Bias Analysis

Introduction

This audit examines the COMPAS dataset, focusing on potential bias in risk score predictions across racial groups. The analysis utilizes established fairness metrics to assess disparities between African-American and Caucasian individuals.

Fairness Metrics Summary

a. **False Positive Rate difference:** *-0.221*

This metric indicates that the false positive rate (FPR) for African-American individuals is 22.1 percentage points higher than for Caucasian individuals. A negative value highlights that African-Americans are more likely to be incorrectly classified as high risk compared to Caucasians.

b. **Average Odds Difference:** *-0.184*

The average odds difference combines disparities in both false positive and false negative rates. The negative value suggests that African-Americans experience less favourable outcomes in risk predictions compared to Caucasians.

c. **Disparate Impact Ratio:** *0.744*

A disparate impact ratio below 1 signal potential adverse impact. Here, African-Americans are less likely to receive favourable predictions (low risk) than Caucasians, with the ratio indicating a substantial disparity.

d. **Statistical Parity difference:** *-0.212*

This metric measures the difference in the probability of receiving a favourable outcome between groups. The negative value shows that African-Americans are 21.2 percentage points less likely to be predicted as low risk compared to Caucasians.

Visualization

The attached bar chart illustrates the False Positive Rate for both groups. African-Americans have a notably higher FPR (0.484) compared to Caucasians (0.705), visually reinforcing the quantitative findings.

Conclusion

The audit reveals significant disparities in COMPAS risk score predictions, with African-American individuals facing higher false positive rates and less favourable outcomes. These findings suggest the presence of bias in the COMPAS algorithm, underscoring the need for careful consideration and potential mitigation strategies when deploying such risk assessment tools in criminal justice settings.