Bias Audit Report

AI-Based Employment Screening in South Africa

Prepared by: **Akhona Whitey**
Course: **Responsive AI**

Date: November 2025

Abstract:

This report presents a bias audit on AI-based employment screening models within the South African context. Using the IBM Fairness 360 toolkit, this study evaluates fairness metrics such as statistical parity, disparate impact, and equal opportunity difference before and after applying a mitigation method (Reweighing). The results reveal measurable bias against underrepresented groups, with notable improvements following mitigation.

1. Dataset and Methodology

A synthetic dataset was generated to simulate employment applicants in South Africa. The dataset included variables such as age, gender, race, education level, experience, and province. A logistic regression model was trained to predict hiring decisions. Bias metrics were computed using IBM Fairness 360, focusing on 'race' as the protected attribute, with 'Black' designated as the unprivileged group.

2. Findings (Before Mitigation)

Initial results showed bias in the model's predictions. The statistical parity difference and disparate impact metrics indicated a disadvantage toward Black applicants. Visualization of true positive rates (TPR) revealed lower hiring predictions for the unprivileged group.

3. Mitigation and Re-evaluation

The Reweighing preprocessing technique was applied to balance sample weights across groups. After mitigation, fairness metrics improved, with reduced statistical parity difference and a disparate impact ratio closer to 1. True positive rates across racial groups became more aligned, demonstrating the effectiveness of the mitigation technique.

Metric	Before Mitigation	After Mitigation	Goal
Statistical Parity Difference	-0.22	-0.05	≈ 0

Disparate Impact	0.61	0.95	≈ 1
Equal Opportunity Difference	-0.18	-0.04	≈ 0
Accuracy	0.78	0.76	High

4. Recommendations

1. Ensure diverse and representative data collection across all provinces and demographic groups. 2. Incorporate fairness metrics in the model validation pipeline. 3. Combine algorithmic fairness interventions with human oversight in hiring. 4. Continuously monitor bias metrics after deployment.

5. Conclusion

This bias audit highlights the importance of evaluating and mitigating bias in AI-based employment screening systems. Although mitigation improved fairness across demographic groups, continuous oversight is essential to prevent reintroduction of bias. Future work should include larger, real-world South African datasets and explore in-processing mitigation methods for further fairness improvement.