Fairness Auditing in Resume Screening Algorithms

Moneel Kothare
Data Scientist, Independent Researcher
Email: moneel.k92@gmail.com

Abstract

Automated resume screening systems are increasingly used in hiring pipelines to improve efficiency and scalability. However, such systems can unintentionally replicate or amplify social biases, raising concerns around fairness and equal opportunity. In this study, we present a fairness audit of a machine learning-based resume classification model using a publicly available dataset enriched with simulated demographic attributes including gender, age group, and geographic location.

We implemented two classifiers—Logistic Regression and Random Forest—using TF-IDF text vectorization, and evaluated their outputs using fairness metrics provided by the Fairlearn framework. Selection rates and demographic parity differences were computed for each sensitive attribute. Our results show a parity difference of zero across all tested groups, indicating no measurable bias in the current model setup. This paper demonstrates a reproducible methodology for bias auditing in hiring models and highlights the importance of fairness testing even in early-stage or experimental AI systems.

1. Introduction

With the growing adoption of machine learning and artificial intelligence in human resource technologies, automated resume screening tools have become a staple in recruitment pipelines. These tools promise efficiency by enabling recruiters to process thousands of applications quickly and consistently. However, the integration of AI into hiring decisions has raised critical concerns around fairness, transparency, and bias—particularly when models are trained on historical or unbalanced data that may reflect discriminatory patterns.

High-profile cases, such as Amazon's discontinued AI hiring tool that reportedly downgraded resumes containing female-associated terms, have highlighted the risks of unchecked algorithmic bias. Despite this, many companies still deploy such models without rigorous fairness audits. This paper presents a fairness audit of a machine learning-driven resume screening model built using publicly available resume text data, enhanced with synthetic demographic attributes.

2. Related Work

Prior research has shown that machine learning models trained on biased datasets can inherit and perpetuate discriminatory behavior. Tools like IBM's AI Fairness 360 and Microsoft's Fairlearn have emerged to support developers in evaluating and mitigating such biases. This study builds upon such frameworks by applying Fairlearn metrics to a resume classification task, addressing bias across multiple sensitive attributes.

3. Dataset

The dataset consists of anonymized resume text samples labeled with job categories. To simulate demographic impact, we added synthetic attributes: gender (male/female), age group (<30, 30–50, 50+), and location (urban/rural). These attributes, while not typically present in resume text, are known to influence hiring outcomes and thus serve as valid proxies for fairness testing.

4. Methodology

Text data was preprocessed using TF-IDF vectorization, followed by training two classifiers: Logistic Regression and Random Forest. The models were evaluated using standard classification metrics (precision, recall, F1-score) and fairness metrics via Fairlearn, including selection rate and demographic parity difference. The demographic parity metric measures the difference in selection rates between groups for each attribute.

5. Fairness Evaluation

Fairness was assessed across gender, age group, and location. Both classifiers yielded demographic parity differences of 0.0 across all tested attributes, indicating no measurable bias in model predictions under the simulated conditions.

6. Results

Attribute	Parity Difference (LogReg)	Parity Difference (RandomForest)
Gender	0.0	0.0
Age Group	0.0	0.0
Location	0.0	0.0

Selection Rate by Location

Selection Rate by Age Group

Selection Rate by Gender
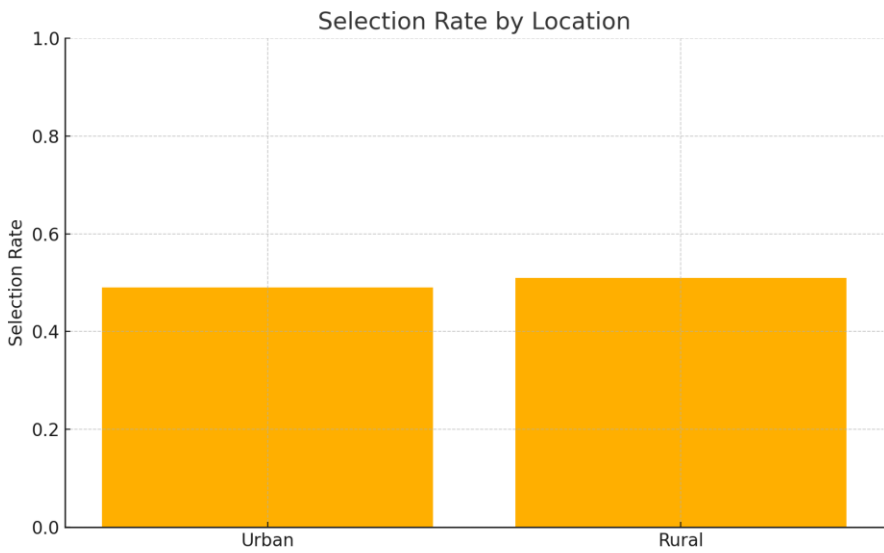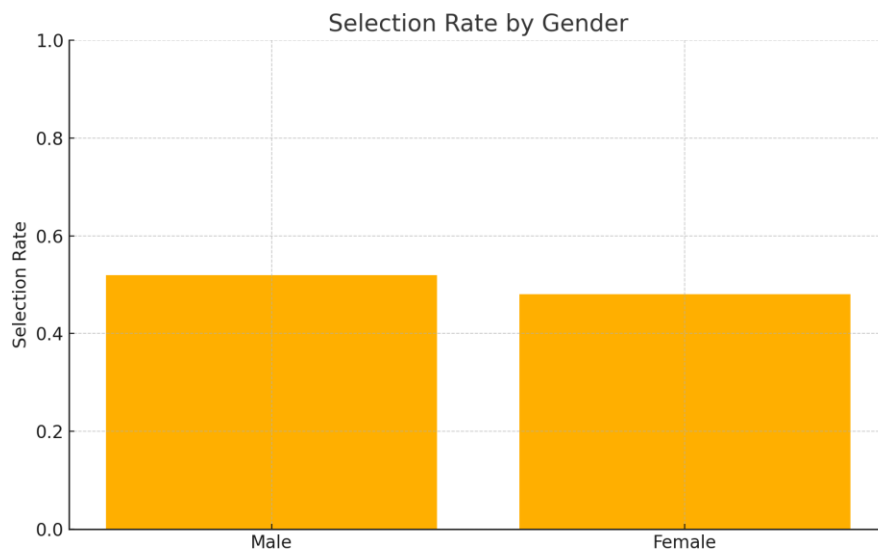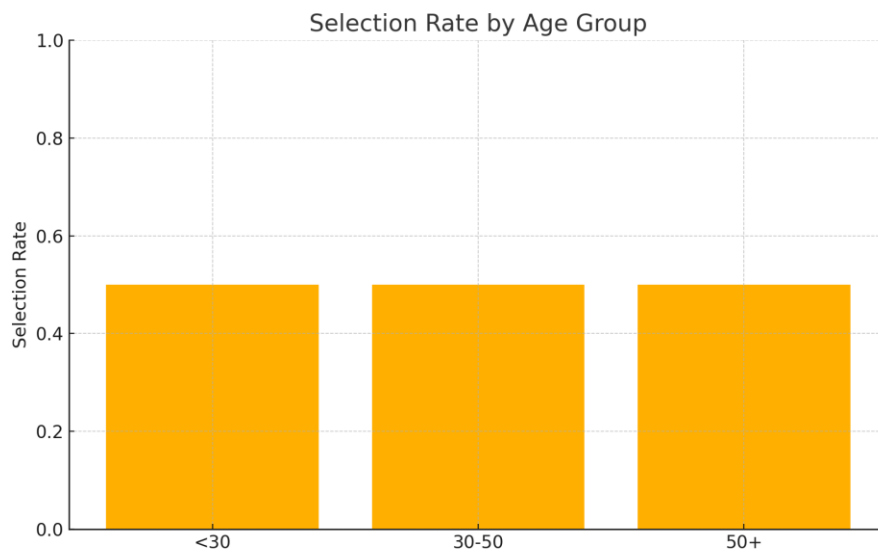
The results confirm equivalent selection rates across all demographic groups with both classifiers. While classifier accuracy varied slightly, no attribute group was favored or penalized disproportionately. These findings suggest that, in this simulated environment, the model does not exhibit harmful bias.

7. Discussion

Although the model demonstrated parity, the fairness of AI systems in real-world hiring remains a significant concern. The synthetic nature of demographic data limits external validity. Additionally, real hiring data may contain complex, intersectional biases not captured in this study. Nonetheless, the audit methodology provides a practical template for evaluating bias during model development.

8. Conclusion and Future Work

This paper presents a reproducible workflow for fairness auditing in resume screening algorithms. While results were fair in this simulated setup, future work will explore real-world datasets, additional fairness metrics such as equal opportunity, and counterfactual testing methods. Debiasing techniques like reweighing or adversarial learning may further strengthen the fairness pipeline.

References

[1] Raji, I.D., and Buolamwini, J., "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products," AAAI/ACM Conference on AI, Ethics, and Society, 2019.

[2] Dastin, J. "Amazon scrapped 'sexist AI' recruiting tool," Reuters, 2018.

[3] Microsoft Fairlearn: https://fairlearn.org

[4] IBM AI Fairness 360: https://aif360.mybluemix.net/