CSCI 1951Z Final Report

Introduction

This audit is centered around examining the hiring practices established by Providence Analytica and Bold Bank. Our primary objective is to dive into the algorithm of the system to identify any potential biases and assess its overall effectiveness. The ultimate goal is to ensure the entire hiring process is fair and trustworthy.

This audit is important because it is always necessary to have an unbiased decision-making process to create a diverse, equitable, and inclusive workplaces. It is essential that every candidate, regardless of their background or characteristics, has an equal opportunity to be considered for employment. Throughout our audit, we studied each aspect of the hiring process and uncovered some underlying biases that may unfairly evaluate some candidates.

This audit not only points out areas of concerns, but also seeks to provide actionable recommendations to enhance the hiring system. Our goal is to make a robust, unbiased, fair and transparent hiring process.

Method

Data Source

In querying the API for our audit, we utilized a couple of datasets that contain artificially generated candidates's information. These datasets include various attributes, such as education level and work experience. Additionally, it contains demographic information such as gender and race. We modified the distribution of values within the dataset to reflect a diverse range of candidates across different demographics, in order to perform the following analysis.

Evaluation Criteria

Our assessment of the hiring algorithm includes two main criteria: Disparate Impact and Statistical Parity Difference.

Disparate Impact

We analyzed the hiring model's impact on different demographic groups to identify any certain negative effects. By analyzing hiring rates across these groups, we look to see disparities that indicate systemic biases. Disparate Impact analysis allows us to assess whether certain groups are systematically disadvantaged in the hiring process, and also discover areas to promote fairness and equity.

Statistical Parity Difference

We also evaluate the difference in selection rates among demographic groups through Statistical Parity Difference analysis. This metric allows us to compare the probability of a positive outcome (being hired) across different groups. By assessing whether there are significant disparities in selection rates among demographic categories, we can detect potential biases in the hiring algorithm that may lead to unequal treatment of candidates based on their different characteristics.

By applying these evaluation criteria, we can comprehensively assess the fairness and equity of the hiring algorithm and make sure that equal opportunities were provided for all candidates.

Analysis Techniques

Data Set Formation

We create mock datasets that mirror real-world job application scenarios, including diverse sets of candidate profiles with varying attributes. In order to limit the dataset size, we had to make choices so that we don't have to iterate through all the attributes. For example, we let work authorization and veteran status be the same for all samples in the dataset. Then we ensure to have a comprehensive coverage of potential applicants. We span applicants' GPA from 2.0 to 4.0. We simulated the applicant's school and many other attributes as well. Our dataset consists of 4224 mock applicants who have diverse academic performance and background. And then we leverage the bank's API to input these mock datasets into the hiring algorithm, simulating the actual decision-making process. This integration ensures that our analysis is based on the algorithm's real-world functionality and response to different candidate profiles. We then use the resume score generated by the resume scorer model to append to the mock dataset to query the candidate evaluation model.

Metric Calculation

With the obtained outputs, we calculate the Disparate Impact and Statistical Parity Difference
Disparate Impact: We analyze the hiring outcomes across different groups to determine if there are significant differences in selection rates.
Statistical Parity Difference: We evaluate differences in selection rates among different groups, focusing on the probability of positive outcomes.

Analysis and Interpretation

We interpret the calculated metrics and identify biases or disparities in hiring outcomes. Utilizing visualizations, we illustrate the distribution of hiring outcomes across demographic groups, facilitating interpretation and decision-making. Lastly, We interpret the analysis results in the context of fairness and equity, identifying any patterns or discrepancies that may indicate bias in the hiring algorithm. Based on our findings, we formulate actionable recommendations to address any identified biases or disparities, aiming to improve the fairness and inclusivity of the hiring process.

Limitations

Firstly, the accuracy and representativeness of the mock datasets generated for input into the hiring algorithm may impact the validity of our analysis. These datasets aim to mimic real-world scenarios, but their artificial nature may not fully capture the complexities of actual hiring processes.

Secondly, the insights provided by the API query into the algorithm's decision-making process are based on simulated data and may not entirely reflect its behavior in real-world scenarios. Factors such as variations in data distribution and candidate characteristics could influence the algorithm's responses differently in practice.

Moreover, the findings and recommendations derived from our analysis may have limited generalizability beyond the specific context of the audit. Factors such as the uniqueness of the hiring algorithm and organizational dynamics may restrict the applicability of our conclusions to other settings or institutions.

Findings

Gender Bias in Resume Scorer Model

Statistical Parity Difference

The statistical parity difference on gender for the resume scorer model is 0.157. Since the fair value for SPD should be 0 This indicates a slight imbalance in selection rates between genders.

Disparate Impact

The disparate impact on gender for the resume scorer model is 1.032, suggesting a minimal adverse effect on gender diversity in hiring outcomes.

Analysis

While the statistical parity difference indicates a slight bias favoring a particular gender, the disparate impact suggests that this bias does not significantly disadvantage any specific gender group. However, even subtle biases can accumulate over time, potentially impacting long-term diversity goals.

Gender Bias in Candidate Evaluation Model:

Statistical Parity Difference (SPD)

The statistical parity difference on gender for the candidate evaluation model is -0.228. Since the fair value for SPD should be 0. This indicates a slight underrepresentation of females of the gender groups.

Disparate Impact (DI)

The disparate impact on gender for the candidate evaluation model is 0.618. Since the fair value for DI is 1. This suggests a moderate adverse effect on gender diversity in hiring outcomes.

Analysis

Similar to the resume scorer model, the candidate evaluation model also exhibits a negative statistical parity difference, indicating bias. Additionally, the disparate impact below 1 suggests that this bias leads to unequal treatment of candidates based on gender during the evaluation process.

Disability Bias in Resume Scorer Model

Statistical Parity Difference (SPD)

The statistical parity difference on disability for the resume scorer model is 0.097, indicating a slight imbalance in selection rates between candidates with and without disabilities.

Disparate Impact (DI)

The disparate impact on disability for the resume scorer model is 1.020, suggesting a minimal adverse effect on disability diversity in hiring outcomes.

Analysis

While the statistical parity difference indicates a slight bias favoring candidates without disabilities, the disparate impact suggests that this bias does not significantly disadvantage candidates with disabilities. However, proactive measures should still be taken to ensure fair representation and inclusion of candidates with disabilities.

Disability Bias in Candidate Evaluation Model

Statistical Parity Difference (SPD)

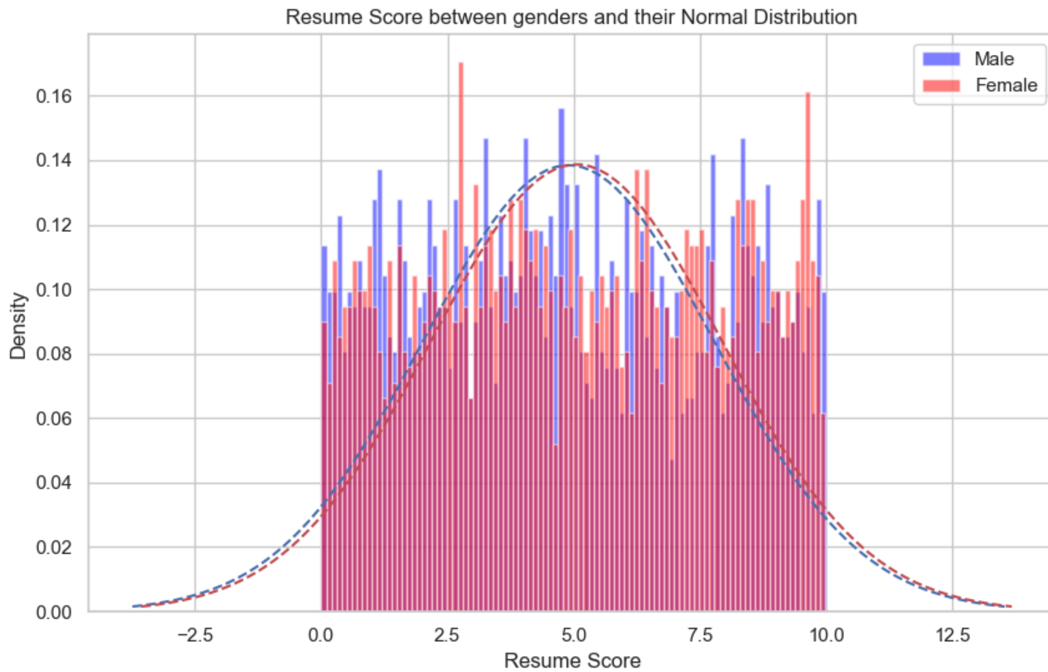
The statistical parity difference on disability for the candidate evaluation model is -0.007, indicating a negligible underrepresentation of candidates with disabilities.

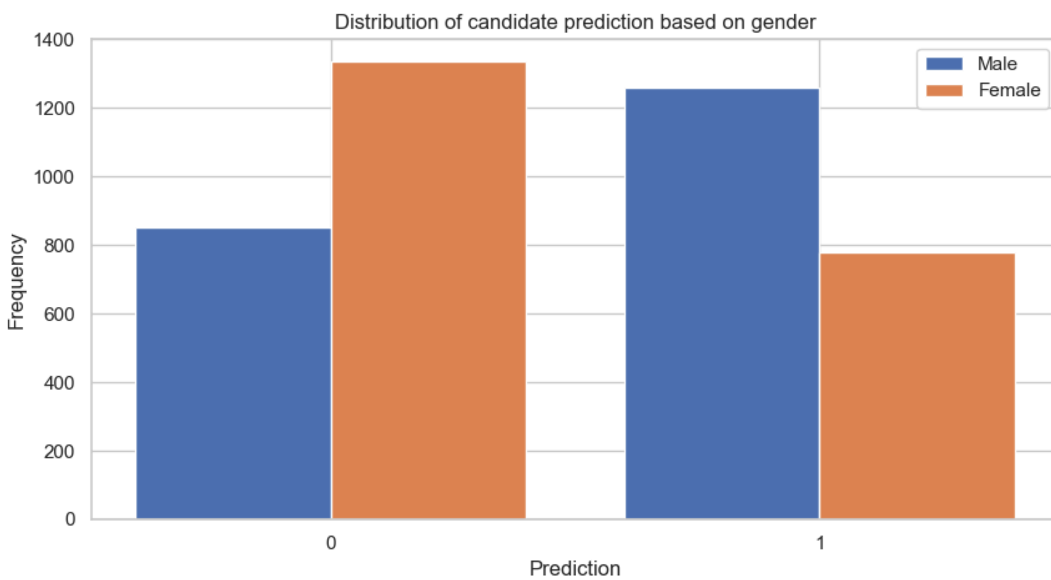
Disparate Impact (DI)

The disparate impact on disability for the candidate evaluation model is 0.985, suggesting a minimal adverse effect on disability diversity in hiring outcomes.

Analysis

The statistical parity difference indicates a negligible bias in the representation of candidates with disabilities in the evaluation process. Similarly, the disparate impact suggests that disability status has a minimal adverse effect on hiring outcomes. While the model exhibits relatively balanced outcomes concerning disability, ongoing vigilance is necessary to ensure fair representation and inclusion of candidates with disabilities.





Recommendations

Model Design

Addressing Gender Bias

Recommendation

Providence Analytica should conduct a thorough review of their models to identify and mitigate any underlying **gender biases**, **especially in the candidate model**.

Actionable Steps

Implement techniques such as fairness-aware machine learning algorithms to mitigate gender biases in the model's decision-making process. Regularly audit and update the model to ensure ongoing fairness and inclusivity.

Enhancing Disability Representation

Recommendation

Modify the candidate evaluation model to actively promote disability inclusivity in hiring decisions.

Actionable Steps

Introduce features or adjustments in the model to ensure fair assessment of candidates with disabilities. Incorporate disability-aware evaluation criteria to mitigate biases against this demographic group.

Company Practices

Incorporating Model Output into Business Decisions

Recommendation

Bold Bank should reassess its integration of model output into business decisions to ensure alignment with diversity and inclusion goals.

Actionable Steps

Prioritize the implementation of hiring practices that foster diversity and inclusivity. Collaborate with Providence Analytica to refine model outputs and align them with Bold Bank's commitment to fair and equitable hiring practices.

Stakeholder Engagement and Transparency

Recommendation

Foster transparency and accountability in decision-making processes involving model outputs.

Actionable Steps

Engage stakeholders, including HR professionals, hiring managers, and diversity advocates, in discussions regarding the interpretation and application of model results. Provide regular training and education on mitigating biases and promoting diversity in hiring practices.