Optimization Final Project: Fair Candidate Selection

Nidhish Nerur[1] and Sandra You[2]

[1]MIT nerur325@mit.edu
[2]MIT sayou@mit.edu

December 20, 2024

1 Candidate Hiring Problem

Our project aims to address the challenge of hiring candidates with high performance potential while ensuring diversity. Specifically, we perform subset selection to identify 50 candidates for a hypothetical job opening. Companies struggle to select the most qualified candidates who satisfy job description requirements such as experience level and technical skills while balancing equitable representation. The objective of our subset selection is to maximize the performance potential of the selected group and ensure representation across gender, race, age, and other factors.

2 Motivation

Research has explored the trade-off between expected mean performance and minority hiring within multistage selection strategies (Finch *et al.*, 2009). We aim to study whether optimization techniques can be applied to balance this trade-off between performance and fairness. Relying only on the performance metric during hiring could lead to an imbalance in the types of individuals hired, potentially favoring certain demographic groups over others. Diversity is crucial for organizations not only because it promotes fairness but also because it has been shown to drive business success. According to a McKinsey Report in 2018 (Hunt *et al.*, 2018), companies with diverse leadership teams, especially in terms of gender and ethnicity, are more likely to outperform their peers in profitability and long-term value creation. In addition to these insights, it is important to ensure that individuals who are seeking to switch divisions or career paths are not unfairly excluded. We hope to use advanced optimization techniques to create more equitable representation, thus adding unique perspectives to the workplace.

3 Dataset Descriptions

We used both synthetic job listings and candidate information to perform our fair candidate selection. The Job Dataset contains specific hypothetical job opening(s) with relevant descriptors that the company is seeking in a particular candidate (Rana, 2023b). The features for the job dataset are discussed here:

- **Min_Exp**: Minimum years of experience required for the job.
- **Max_Exp**: Maximum years of experience allowed for the job.
- **QualificationMapped**: The education level for the job in standardized categories.
- **Min_Salary**: The minimum salary offered for the job.
- **Max_Salary**: The maximum salary offered for the job.
- **WorkMapped**: The type of work status, such as full-time, contract, part-time.
- **PythonCode**: Whether Python proficiency is required for the job.
- **JavaScriptCode**: Whether JavaScript proficiency is required for the job.

- **SocialMediaSkill**: Whether social media management is expected for the role.
- **CADSoftwareSkill**: Whether expertise in CAD (Computer-Aided Design) software is necessary for the job.
- **NetworkDesignSkill**: Whether skills in network design are needed for the job.

The Employee Dataset contains 3000 employee records and provides key features about the candidates' current role and level of expertise (Rana, 2023a). Both datasets are synthetic but they are meant to mirror real-world candidate and job details. The final employee dataset used includes these features:

- **YearsWorked**: Number of years the candidate has worked.
- **TitleMapped**: Mapped job title, standardized for various roles.
- **EducationMapped**: The standardized education level of the candidate.
- **EmployeeTypeMapped**: Employment type, such as full-time, part-time, or contract.
- **Age**: Age of the candidate to analyze trends related to demographics.
- **GenderMapped**: The gender of the candidate, mapped into categories for diversity analysis.
- **RaceMapped**: The standardized race of the candidate.
- **PerformanceScoreMapped**: The candidate's performance score, mapped to categories such as exceeds expectations, meets expectations, or needs improvement.
- **PythonCode**: Inferred skill level in Python programming.
- **JavaScriptCode**: Inferred skill level in JavaScript programming.
- **SocialMediaSkill**: The candidate's proficiency in social media management.
- **NetworkDesign**: Inferred expertise in designing and managing computer networks.
- **CADSoftwareSkill**: Proficiency in using CAD (Computer-Aided Design) software.

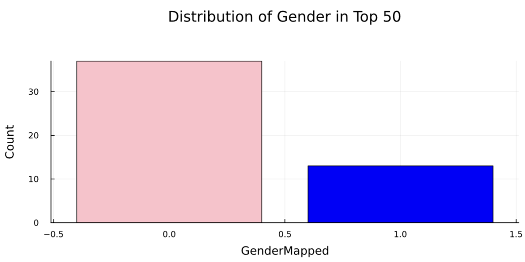
The `PerformanceScoreMapped` column is mapped to random integers within specific ranges based on the category values: `Exceeds` (70-100), `Fully Meets` (50-70), `Needs Improvement` (20-50), and `PIP` (0-20). The `PIP` category indicates the employee is likely to be fired unless he or she significantly improves performance. In subsequent sections, we discuss how we utilized this data in the fair candidate selection process.

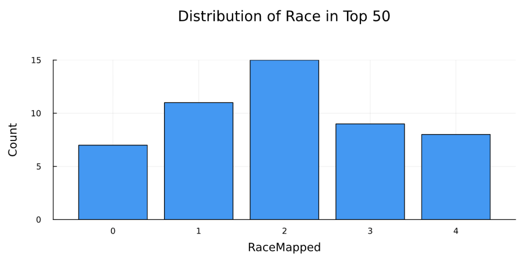
4 Optimization Approach

To approach this problem, we created benchmark models, defined a fairness metric, and developed a performance-fairness optimization model. We discuss each of these steps in greater detail below:

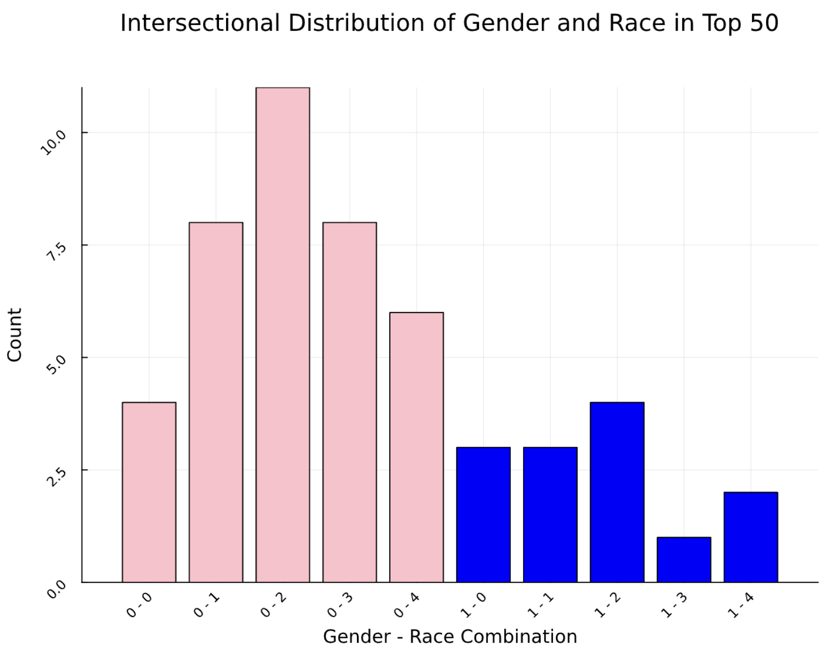
4.1 Performance Baseline Model

The simplest way for a hiring manager to select a subset from a list of candidates is to filter solely by their prior performance evaluation scores with no other considerations. We expect this model to be biased against certain gender, racial, or age groups and disproportionately select people from particular demographic backgrounds. To confirm our belief, we selected the top 50 candidates by their performance score and visualized the relevant distributions of gender (0: Female, 1: Male) and race (0: Asian, 1: Black, 2: White, 3: Other, 4: Hispanic), where gender is colored pink for female and blue for male:

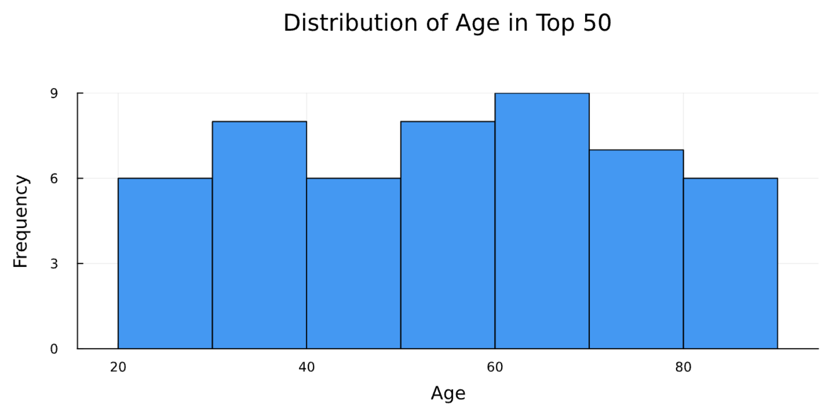




There is a clear gender divide, with 37 women and 13 men selected. In addition, white individuals represent the majority, followed by blacks, and there appear to be discrepancies between the racial groups. The overall dataset has roughly even proportions of gender and racial groups, so we can directly inspect differences in the raw counts. To further our results, we created a graph to look at the intersection between race and gender:



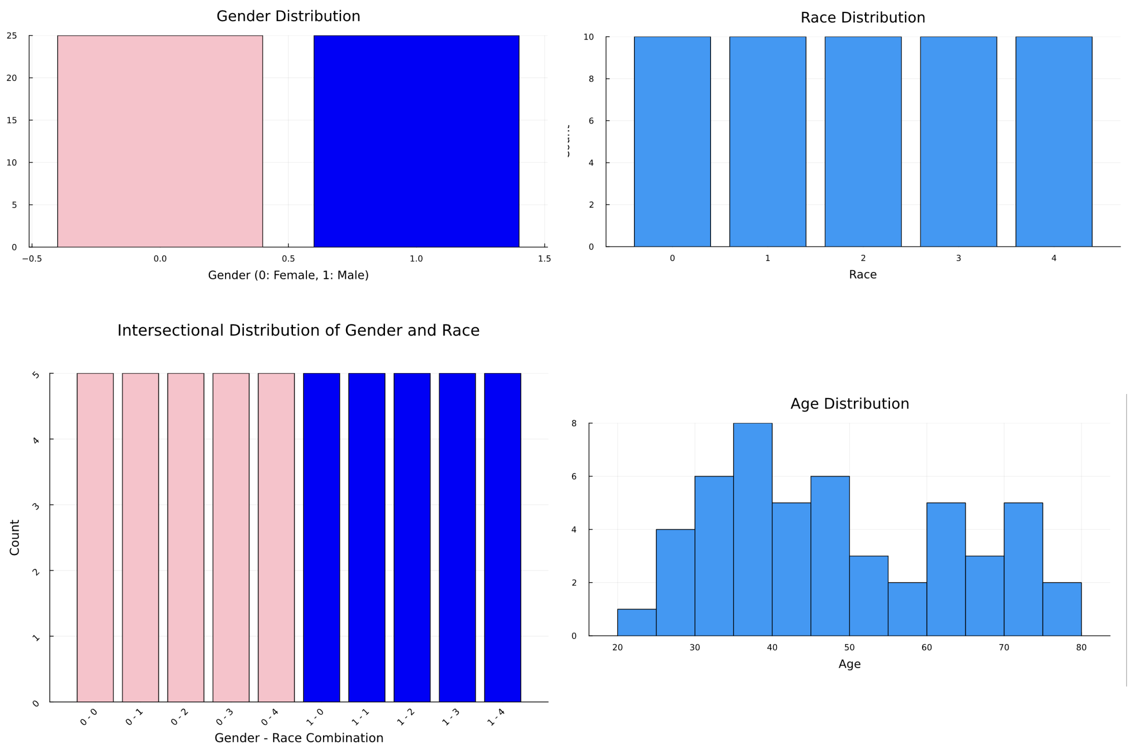
In the above graph, we see a stark gap across all the female-race and male-race pairs, where females in each racial group outnumber males of the corresponding race. This is quite surprising but perhaps this is because our dataset only contains 3,000 candidates and the data was synthetically created. Additionally, we see that white females make up the largest proportion relative to all intersections, followed by black females, then females in the Other racial group. We further sought to look at the age distribution to assess whether the baseline model selects people across age groups:



The benchmark seems to perform quite well in selecting people somewhat evenly across the age groups. This model achieves the highest average performance score of roughly 97.46, yet does not take into account equality and fairness between demographic groups.

4.2 Fairness Baseline Model

Now, we wanted to select 50 candidates such that we would have perfect fairness between gender, race, the intersectionality of gender and race, and age. We showcase the demographic distributions for the fairness benchmark model:



Although this model perfectly balanced the gender and racial groups, it struggled to create a completely even selection across age groups. The average performance score of these 50 selected individuals is only 58.68, much lower than the prior baseline performance model. Consequently, we believe there is a trade-off between performance and fairness. Given this insight, we wanted to quantify discrepancies in demographic and age parity for a given subset of 50 selected candidates. Hence, we developed a fairness metric discussed in the next section.

4.3 Fairness Metric

The formulation for our fairness metric F is shown here:

$$F = 1 - w_1 D_r - w_2 D_g - w_3 D_{rg} - w_4 A$$

Components

1. **Race Discrepancy Penalty** D_r:

$$D_r = \sum_r |p_r - q_r|$$

- p_r: The proportion of individuals of race r in the overall population.
- q_r: The proportion of individuals of race r in the selected group.
- The D_r penalty measures the absolute difference in the proportion of each gender in the overall population relative to the selected group.

2. **Gender Discrepancy Penalty** D_g:

$$D_g = \sum_g |p_g - q_g|$$

- p_g: The proportion of individuals of gender g in the overall population.
- q_g: The proportion of individuals of gender g in the selected group.
- The D_g penalty measures the absolute difference in the proportion of each gender in the overall population relative to the selected group.

3. **Intersectional Discrepancy Penalty** D_{rg}:

$$D_{rg} = \sum_g \sum_r |p_{rg} - q_{rg}|$$

- p_{rg}: The proportion of individuals of race r and gender g in the overall population.
- q_{rg}: The proportion of individuals of race r and gender g in the selected group.
- The D_{rg} penalty measures the absolute difference in the proportion of each race-gender combination in the overall population relative to the selected group.

4. **Age Discrepancy Penalty** A:

$$A = \frac{\left|\mu_{\text{age}}^{\text{overall}} - \mu_{\text{age}}^{\text{selected}}\right|}{\sigma_{\text{age}}^{\text{overall}}}$$

- $\mu_{\text{age}}^{\text{overall}}$: The mean age of individuals in the overall population.
- $\mu_{\text{age}}^{\text{selected}}$: The mean age of individuals in the selected group.
- $\sigma_{\text{age}}^{\text{overall}}$: The standard deviation of ages in the overall population.
- The A penalty measures the difference in the average age between the selected group and the overall population over the standard deviation of the population group's age.

Weights

The weights w_1, w_2, w_3, w_4 are non-negative and satisfy:

$$w_1 + w_2 + w_3 + w_4 = 1$$

The fairness metric F calculates how closely the selected subset of candidates aligns with the overall population of candidates with respect to race, gender, intersectionality of race and gender, and age distributions. From a business perspective, this metric helps ensure that hiring, interview, and admission decisions are more equitable and representative of the population. This metric has a maximum value of 1 which indicates perfect fairness, while lower scores suggest greater levels of disparity. As a result, we hope to consider both performance and fairness when deciding which subset of 50 candidates to select for an interview.

4.4 Optimization Formulation

We decided to keep our objective as maximizing the performance score for the 50 selected candidates, while adding group fairness measures to the constraints. Additionally, we had specific qualifications and preferences outlined in the job listings dataset, so we added constraints to ensure some of the candidates would satisfy certain education or experience level requirements. We also inferred skills based on the candidate's current role and ensured at least a certain proportion of candidates had the relevant skills detailed in the job description. The full formulation for n potential candidates and descriptions of the constraints are highlighted here:

Objective Function

$$\max \quad \sum_{i=1}^{n} P_i \cdot x_i$$

Subject to Constraints

1. **Experience Constraint:**

$$\text{min\_exp} - 3 \leq \text{years\_worked}_i \cdot x_i \leq \text{max\_exp} + 3, \quad \forall i \in \{1, \ldots, n\}$$

- min_exp: Minimum preferred years of experience.
- max_exp: Maximum preferred years of experience.
- years_worked_i: Number of years candidate i has worked.
- This ensures the candidate's current number of years of work experience is within the job's desired minimum and maximum experience level, with a buffer of 3 years. We had to make the buffer quite large to have enough feasible optimization solutions.

2. **Education Match Constraint:**

$$\sum_{i=1}^{n} e_i \cdot x_i \geq 0.3 \cdot \sum_{i=1}^{n} x_i$$

- e_i: Binary indicator ($e_i = 1$ if candidate i's education matches the job requirements, 0 otherwise). There are 3 education types: Bachelors, Masters, and PhD.
- x_i: Binary decision variable ($x_i = 1$ if candidate i is selected, 0 otherwise).
- The education matching constraint guarantees that at least 30% of selected candidates will have the education background sought after by the particular job.

3. **Employee Type Match Constraint:**

$$\sum_{i=1}^{n} \text{ETM}_i \cdot x_i \geq 0.3 \cdot \sum_{i=1}^{n} x_i$$

- ETM_i: Binary indicator ($\text{ETM}_i = 1$ if candidate i's employee type matches the job's requirements, 0 otherwise). ETM represents Full-Time, Contract, and Part-Time status.
- We ensure that at least 30% of candidates called for the interview currently have a work status that is the same as the one the job seeks.

4. **Age Group Constraints:**

$$\sum_{i \in \{\text{Age}_i \in (22,35]\}} x_i \geq \rho_1 \cdot \sum_{i=1}^{n} x_i$$

$$\sum_{i \in \{\text{Age}_i \in (35,50]\}} x_i \geq \rho_2 \cdot \sum_{i=1}^{n} x_i$$

$$\sum_{i \in \{\text{Age}_i \in (50,80]\}} x_i \geq \rho_3 \cdot \sum_{i=1}^{n} x_i$$

$$\text{Age}_i \cdot x_i \leq 80, \quad \forall i \in \{1, \ldots, n\}$$

- Age_i: Age of candidate i.
- ρ_1, ρ_2, ρ_3: Proportions ensuring representation from different age groups.
- The age constraints ensure that we select a certain proportion across 3 age groups, namely, (22, 35], (35, 50), and (50, 80]. Initially, we limited age to the United States retirement age of about 67, but we did not have enough feasible solution subsets for subsequent analysis. Thus, we ensure we do not hire anyone over 80.

5. **Gender Constraint:**

$$\sum_{i=1}^{n} G_i \cdot x_i \geq \rho_4 \cdot \sum_{i=1}^{n} x_i$$

- G_i: Gender indicator for candidate i ($G_i = 0$ for female, $G_i = 1$ for male).
- ρ_4: Weight controlling the minimum proportion of male candidates.
- The gender constraint guarantees that we select at least ρ_4 proportion of male candidates.

6. **Race Group Constraints:**

$$\sum_{i=1}^{n} R_{ki} \cdot x_i \geq \rho_5 \cdot \sum_{i=1}^{n} x_i, \quad \forall k \in \{0, 1, 2, 3, 4\}$$

- R_{ki}: Binary indicator ($R_{ki} = 1$ if candidate i belongs to race group k, 0 otherwise).
- ρ_5: Weight ensuring minimum representation of each racial group.
- Similar to the gender constraint, we ensure at least ρ_5 proportion for each racial group.

7. **Skill Match Constraints:**

$$\sum_{i=1}^{n}(\text{Python}_i \cdot \text{Python}_{\text{job}}) \cdot x_i \geq 0.2 \cdot \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n}(\text{JavaScript}_i \cdot \text{JavaScript}_{\text{job}}) \cdot x_i \geq 0.2 \cdot \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n}(\text{SocialMedia}_i \cdot \text{SocialMedia}_{\text{job}}) \cdot x_i \geq 0.2 \cdot \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n}(\text{CADSoftware}_i \cdot \text{CADSoftware}_{\text{job}}) \cdot x_i \geq 0.2 \cdot \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n}(\text{NetworkDesign}_i \cdot \text{NetworkDesign}_{\text{job}}) \cdot x_i \geq 0.2 \cdot \sum_{i=1}^{n} x_i$$

- Skill_i: Binary indicator for a specific skill (e.g., Python, JavaScript, SocialMedia, CADSoftware, NetworkDesign).
- $\text{Skill}_{\text{job}}$: Job requirement indicator for a specific skill.
- We made sure that if the job wants a particular skill like Python coding then at least 20% of selected candidates need to have Python experience.

8. **Subset Selection Constraint:**

$$\sum_{i=1}^{n} x_i = 50$$

- x_i: Binary decision variable indicating whether candidate i is selected.
- 50: Total number of candidates to be selected.
- This constraint ensures we select only 50 candidates, given that Decision Analytics, Inc. has limited time and resources to interview people.

Now that we have the fairness metric and optimization formulation, we wanted to assess whether there truly is a trade-off between performance and fairness. We modify the ρ weights in the above formulation and plot changes in the fairness metric. Furthermore, we validate our model outcomes across multiple jobs to determine robustness of our results.

5 Key Findings

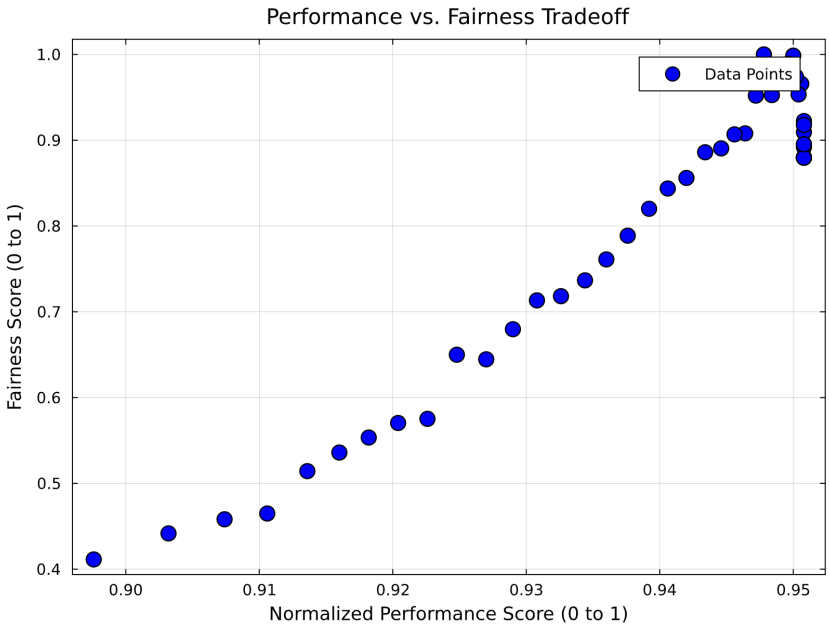
For our initial results, we selected a Web Developer job posting, which identifies candidates with strong JavaScript skills, a couple of years of work experience, and a Bachelors' education level preference. We show how the performance and fairness scores change as we re-weight the sensitive attribute constraints (i.e., Age, Gender, and Race constraints). Then, we proceed to perform a sensitivity analysis, displaying how the other fairness measures change as we optimize over a particular equality consideration. Finally, we discuss the validity of the performance and fairness scores generated by our model across 10 different jobs.

5.1 Performance-Fairness Trade-off

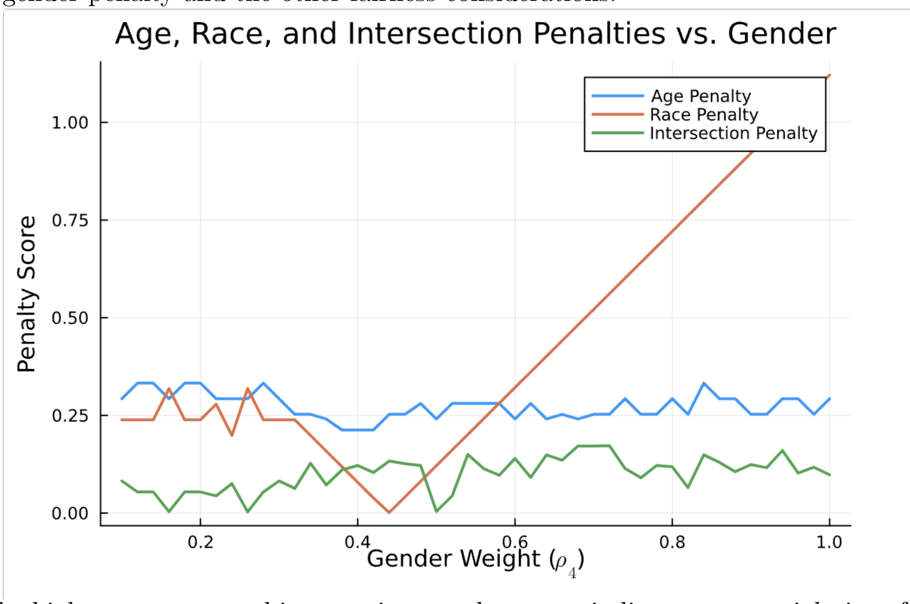
We use the objective value of our optimization formulation to assess the total performance score of the 50 candidates. Then, we compute the fairness metric for the specific subset of candidates. To compare the optimization models across both performance and fairness, we set all the weights in the fairness metric of $w1$, $w2$, $w3$, and $w4$ to be 0.25, which equally weights the race, gender, intersectionality, and age penalties. We initially set the ρ proportions in the optimization problem to 0.05, which ensures at least 5% representation across each fairness guideline. Then, we vary each of the ρ values independently to determine the impact of changing one fairness constraint on both the performance and overall fairness metrics.

5.2 *Fairness Sensitivity Analysis*

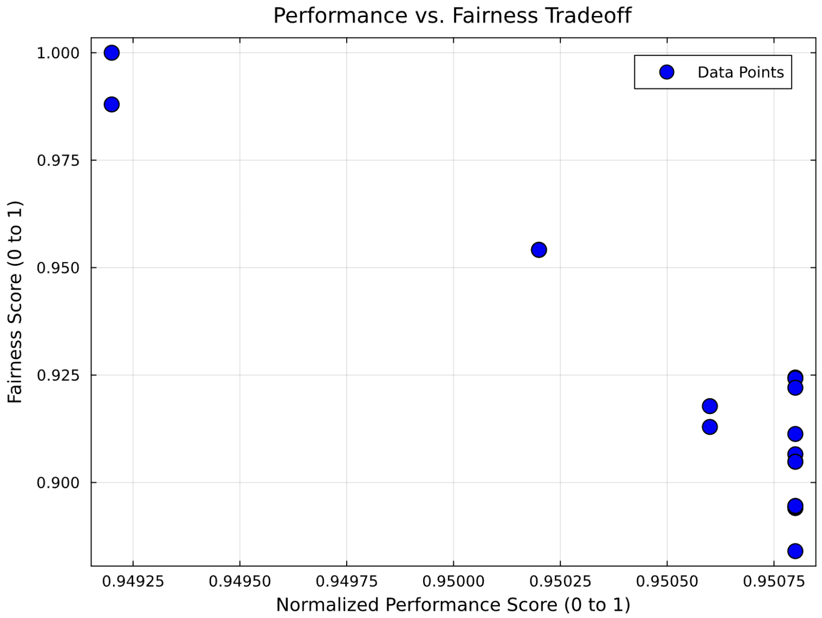
First, we experiment with adjusting only the ρ_4 weight for the gender constraint, which represents the minimum proportion of male candidates the model is enforced to select. We tested ρ_4 values ranging from 0.1 to 1, and we continuously re-ran the optimization formulation to find different subsets of 50 candidates. For the results, we only retained the performance and fairness scores if we found a feasible subset of 50 candidates satisfying all our given conditions. Then, we show the relationship between performance score and fairness across the different gender weights, where each datapoint is for a specific ρ_4 value:



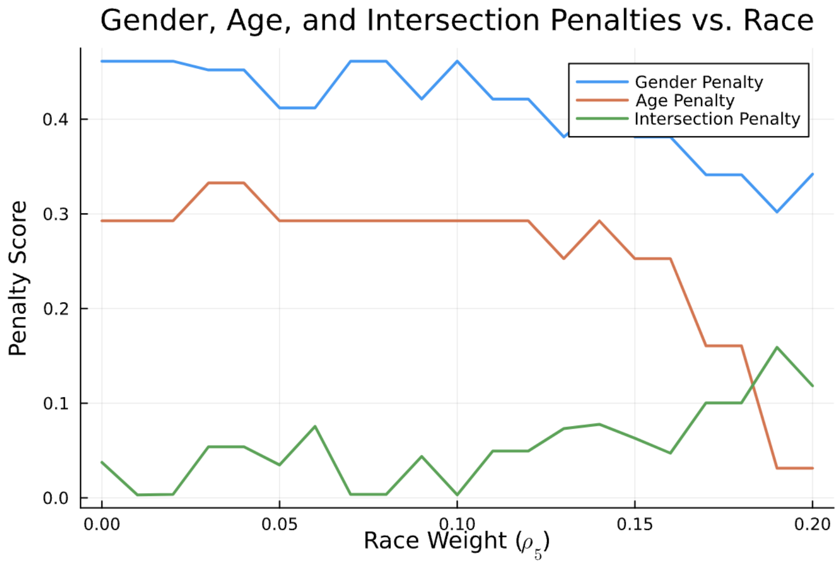
Contrary to our initial expectations, we see that generally as the performance score increases, the fairness metric tends to rise as we adjust the gender proportion. The graph indicates the gender proportion can have a net improvement in both fairness and performance. However, we see that after a certain performance level, the fairness metric begins to decline, perhaps suggesting a trade-off at higher performance levels. Subsequently, we wanted to further investigate the interplay between the gender penalty and the other fairness considerations:



In context, the higher age, race, and intersection penalty scores indicate greater violations for that particular fairness consideration. As we increase the gender weight on the x-axis, the model is forced to select more male candidates, and we see the associated changes in the other penalties. In the graph we see that the age and intersectionality of race and gender stay somewhat stable, but the race penalty continues to increase as we require a greater proportion of males to be selected. The extreme case is when ρ_4 is 1, meaning we have to select only male candidates. In this case, we see a significant penalty in the race category. Next, we repeat this procedure by modifying the race ρ_5 weight independently and plotting the associated performance-fairness curve:



In this case, there is a clear trade-off because as we increase the performance score, fairness is somewhat lower with different racial proportions. The ρ_5 weight indicates we need at least ρ_5 proportion across all 5 of the racial groups represented. The perfect scenario matches the population distribution of about 20% in each racial group. The graph indicates that accounting for racial discrepancies may result in greater penalties across the other fairness considerations, so we look at the changes across the age, gender, and intersection scores:



As the race weight increases to create more equal groups, the gender and age penalties seem to somewhat decline while the intersection penalty rises. Even though the overall gender penalty declined, ensuring fairness at the intersection of race and gender is more challenging to uphold. As a result, we see trade-offs in optimizing fairness across multiple dimensions. We wanted to further confirm whether the performance-fairness results hold for different hypothetical job postings, which is explored in the next section.

5.3 Model Validation with Multiple Jobs

We show the performance and fairness scores for 10 different jobs using fixed ρ weights of 0.05 and w weights of 0.25 in our optimization and fairness formulations, respectively:

Table I: Job Performance and Fairness Scores

Job Index	Job Title	Performance Score	Fairness Score
2	Operations Manager	4754	0.7555
3	Web Developer	4766	0.7095
23	Software Engineer	4766	0.7239
32	Psychologist	4607	0.8034
35	Data Analyst	4808	0.8024
65	Customer Service Manager	4607	0.8066
503	UX/UI Designer	4607	0.8034
679	Sales Associate	4580	0.8398
826	Investment Banker	4808	0.8028
1001	Financial Advisor	4613	0.7881

Our formulation has variability across the given 10 positions, but the performance and fairness scores are reasonably high. This reflects our model is performing well in balancing these two objectives given our objective function and fairness constraints. We expect our formulation to provide greater information for data and software positions, as we added fields to indicate when a candidate possesses skills in Python, JavaScript, and Network Design that match the corresponding job requirements. In contrast, we did not add any specific skills for Psychology or Finance, so our model may not be as generalizable to those jobs. Overall, we see our model has some applicability in different industries, but we would need to add more skill-match constraints to ensure its efficacy.

6 Impact

The impact of our work is that we have developed a framework that maintains a strong balance between performance and fairness across different jobs, while including key domain-specific features. With our rigorous optimization formulation and fairness considerations, our model aims to achieve demographic parity across race, gender, intersectionality, and age. Accounting for these fairness measures is vital to fostering diversity and enhancing organizational inclusivity. We hope to minimize bias in the interview selection and hiring processes, providing equitable access to all.

Naturally, it is quite challenging to perfectly optimize for fairness across all dimensions while attaining the best performing candidates. This was evident in our findings as we adjusted the gender and race weights to see how our model adapts. Our work can be applied across industries to support data-driven, fair decision-making practices in human resources.

7 Future Direction

In the future, we would refine our model by adding more specific skill match constraints for job categories like Psychology and Finance to optimize its application across industries. We could also further investigate how the fairness metrics, such as gender, race, and age, vary as we optimize over one specific metric to understand the trade-offs in adjusting fairness priorities. We would also like to gather more real candidate information online to enhance the model's generalizability. To further validate the robustness of our model results, we would compute the performance and fairness metrics across a wider range of jobs. Fair candidate selection is a challenging task as we have to prioritize multiple conflicting objectives, and our model formulation is just the first step to create equitable hiring practices.

References

Finch, David M., Edwards, Bryan D., and Wallace, J. Craig (2009). "Multistage selection strategies: simulating the effects on adverse impact and expected performance for various predictor combinations". *National Library of Medicine.*

Hunt, Dame Vivian *et al.*, (2018). *Delivering through Diversity.* Tech. rep. McKinsey & Company.

Rana, Ravender Singh (2023a). "Comprehensive Candidate Dataset". *Kaggle.*

— (2023b). "Comprehensive Job Dataset". *Kaggle.*