Analyzing Resume Bias Using Logistic Regression and Decision Trees

Josie Kelley and Toby Kohan
Data 300 Final Project

Abstract

This study analyzes bias in resume screening using machine learning models on Bertrand and Mullainathan's dataset of 5,000 resumes. Our logistic regression and decision tree models reveal that Black-sounding names decrease callback likelihood by 36.8%, while factors like special skills increase odds by 117.9%. The decision tree confirms name-based discrimination as the second most influential factor after experience. Results highlight the need for anonymous applications and rigorous testing of hiring algorithms to prevent perpetuating historical biases.

Introduction

Resume screening occurs all the time when applications are sent to a job posting. Hiring teams quickly scan resumes, and sometimes bias can slip through. One example of this bias is assuming some applicants are more qualified for jobs based on the way their name sounds. Hiring teams are beginning to use machine learning to screen resumes. However, this can also cause unintentional bias on the bases of perceived race, gender, skills, and so on if the model are trained on data that shows a bias?

Models

Chi Squared Test:
We first ran a chi-squared table test on our variables in order to determine which ones informed the results the most. We found three independent variables to be the most statistically significant for determining callback: race, first name, and special skills.

Logistic Regression:
We used logistic regression to predict if an applicant should get a callback or not. The p-value for first name, special skills, employment holes, and years experience are much more significant. This means that these variables play a bigger role in determining callback in comparison to college degree and worked during college.

Odds ratio:
Odds ratios tell us the difference in odds of getting a callback based on one variable while holding the others constant.

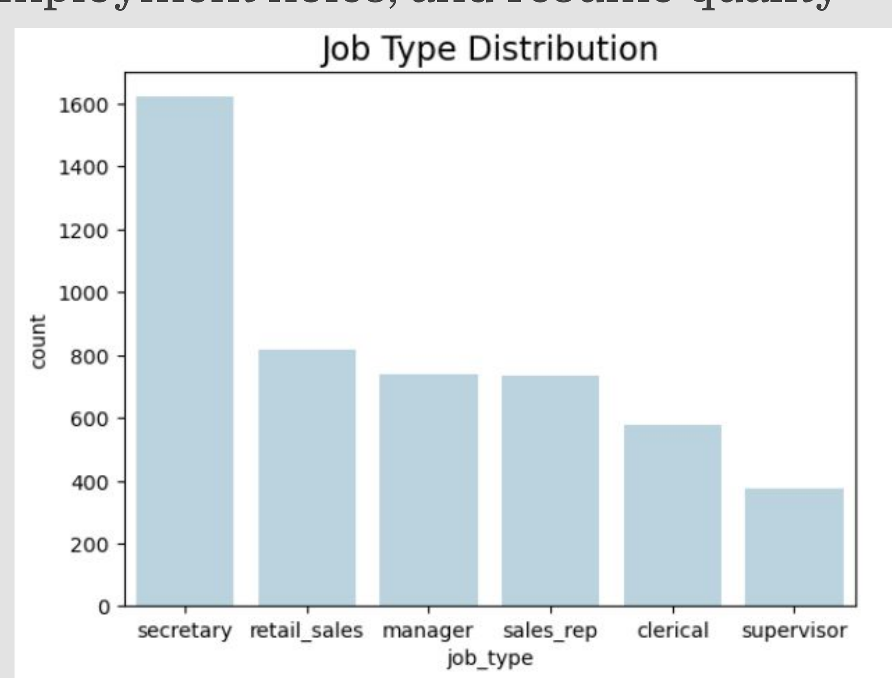
Decision Trees
We used a decision tree model to measure the importances of each feature.

Data and Preprocessing

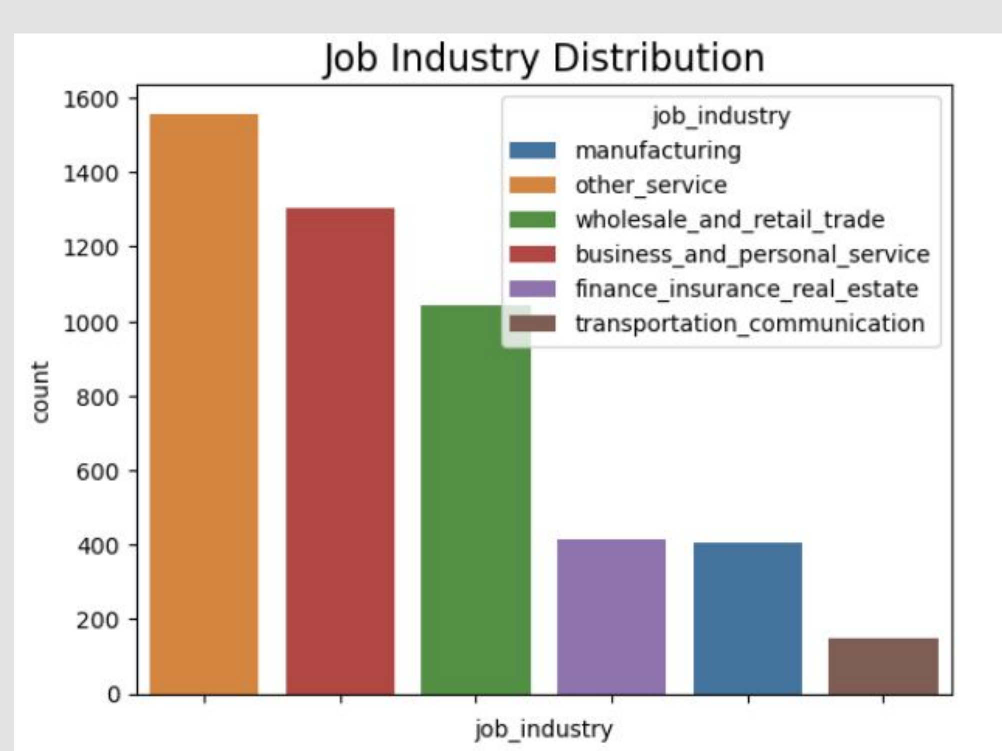
We used a dataframe resume characteristics that were sent to job openings in Boston and Chicago and a label: callback or no callback, this was the data used in a study we found which analyzed a similar resume bias issue.

We cleaned the data and dropped unnecessary variables before running it through our models. We also grouped names based on race and gender.

Our Data:
- 4870 samples
- Dependent variable: resume received callback or not
- Independent variables: first name, years college, worked during school, years experience, computer, skills special skills, employment holes, and resume quality



We graphed the distribution of job type and job industry. Besides the outlier of secretary, most jobs have about the same popularity. However based on job industry, it seems like all the jobs seem to not require a lot of education/outside training. This makes the hiring process even more susceptible to bias because many people could be qualified for the jobs listed in our data.

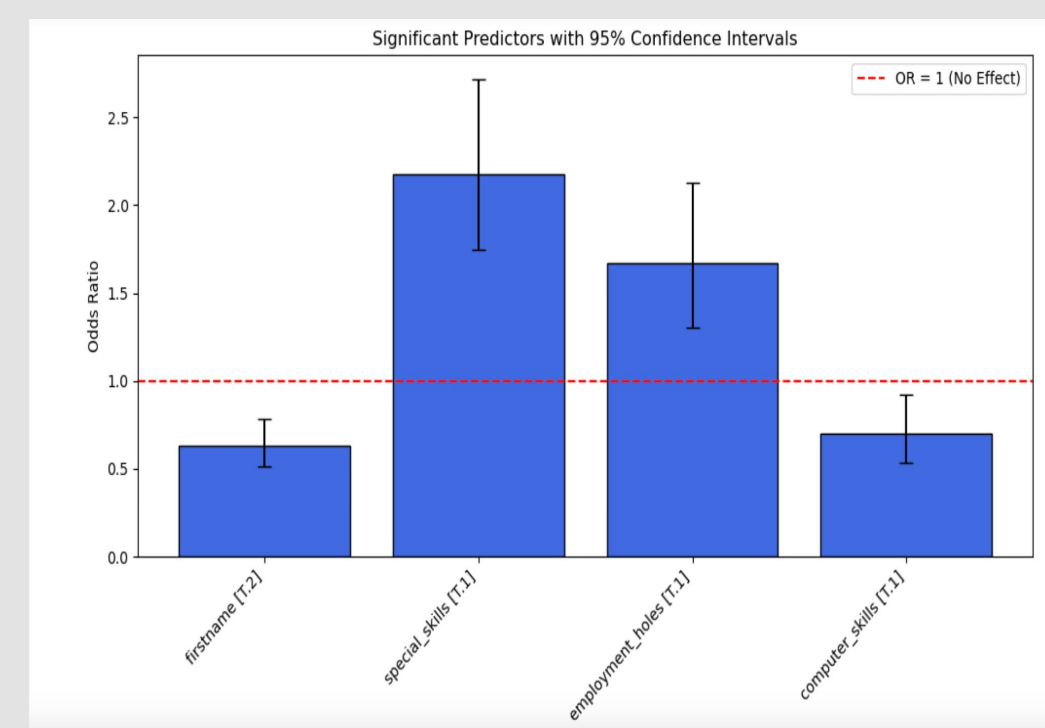


Results

- Chi-Squared Test: Race ($p=5.00\times10^{-5}$), special skills ($p=1.41\times10^{-14}$), employment holes ($p=6.91\times10^{-7}$), and honors ($p=9.42\times10^{-7}$) showed strongest connection to callback decisions.

Logistic Regression Findings

- Applicants with Black-sounding names →36.8% less likely to receive callbacks
- Including special skills → +117.9%
- Computer Skills → increased odds by 29.6%
- Employment Gaps → increased callback by 66.7%



Decision Tree Analysis

- Most Influential Factors (in order):
  1. Years of experience
  2. First name (indicating racial bias)
  3. Employment holes
  4. Worked during school

Both models revealed significant bias based on perceived racial identity from first names, an inexplicable factor that should not influence hiring decisions. This finding highlights the critical need for anonymized application processes and algorithmic fairness testing to prevent the perpetuation of historical discrimination patterns in modern hiring systems.

Conclusion and Limitations

Our models both had bias. The logistic regression model mostly had special skills and first name bias. The decision tree had bias in years experience and first name. Both models were biased to white sounding first names when choosing who should get a callback. However, it should be noted that most jobs used in our data did not require education or a lot of experience. In the future, we would like to repeat our project on different data. We would like to choose data that focuses on a specific industry and job type and has the same requirements for each jobs. Additionally, our data was from 2004, and we hope that there have been improvements with callback data and biases.

Additionally, we only focused on one race in our study, but there are more races to consider besides just Black and White. We would like to repeat our project on different identities.

We recommend companies that are using algorithms to decide on callback to run tests to make sure their models are not biased. If they are biased, then they should make steps to solve this by adding artificial samples or changing job requirement thresholds.

Interpretation/Implementation

- Could we see a shift where companies or screening algorithms assign an ID number or something similar? This could reduce any bias caused by the perceived race of someone's name.
- This data helps candidates decide what is truly important to put on resumes and what employers are actually looking for and what they value.
- It is understandable that special skills are a big contributor to callback as they inform potential experience for the job. However, contributors like race should not be playing a significant role in callback rates.

Source

Bertrand M, Mullainathan S. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination". The American Economic Review 94:4 (991-1013)