Analyzing Resume Bias using Logistic Regression and a Decision Tree Model

Josie Kelley and Abby Kohan

Abstract:

Our study examines biases in resume screening using machine learning approaches applied to Bertrand and Mullainathan's 2004 dataset of nearly 5,000 resumes sent to Boston and Chicago job postings. Through logistic regression and decision tree modeling, we identified factors that significantly influenced callback decisions. Our findings reveal that first names associated with racial identity substantially impact callback rates, with Black-sounding names decreasing callback likelihood by 36.8%. Both models showed explainable biases for factors like special skills (increasing callback odds by 117.9%) and years of experience, while also demonstrating persistent bias based on perceived racial identity from first names. The decision tree confirmed name-based discrimination as the second most influential factor after experience. These results emphasize the need for anonymous application processes and rigorous testing of algorithmic hiring tools to prevent the perpetuation of historical biases in recruitment processes.

Introduction:

As college students, approaching graduation and actively seeking internships and job opportunities, we have become increasingly aware of how hiring processes may be influenced by factors beyond just skills and general qualifications. Many applicants hope they get to the interview stage to demonstrate abilities like conversation skills, critical thinking, and other soft skills that don't necessarily go onto a resume, but how can they show these if they're already taken out of the process because of biases out of their control? This awareness led us to explore potential biases in resume screening, a critical first step in most hiring processes. Our interest was further piqued by the work of Bertrand and Mullainathan (2004), whose field experiment "Are Emily and Greg More Employable than Lakisha and Jamal?" noted significant racial discrimination in the labor market.

The transition from education to employment represents a pivotal moment in many students' lives, and the thought that factors beyond their control could influence this transition is quite unsettling. Building on Bertrand and Mullainathan's research, we sought to examine whether modern computational approaches like machine learning models would reproduce or potentially amplify these biases when trained on historically biased data. Using the dataset from Bertrand and Mullainathan's study, which included nearly 5,000 resumes sent to job postings in Boston and Chicago, we applied two distinct machine learning approaches, logistic regression and decision tree modeling, to analyze callback patterns. Our goal was not only to identify potential biases in hiring algorithms but also to understand which resume characteristics most significantly influence callback decisions.

By exploring this intersection of data analytics and social equity, we hope to contribute to the ongoing conversation about fairness in hiring practices and the responsible implementation of algorithmic decision-making.

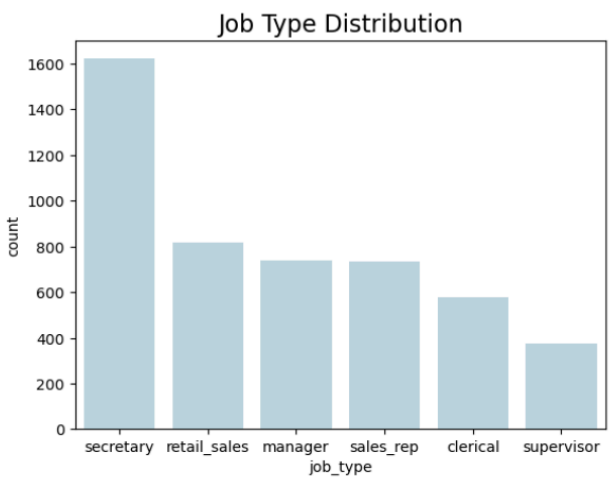
Roles

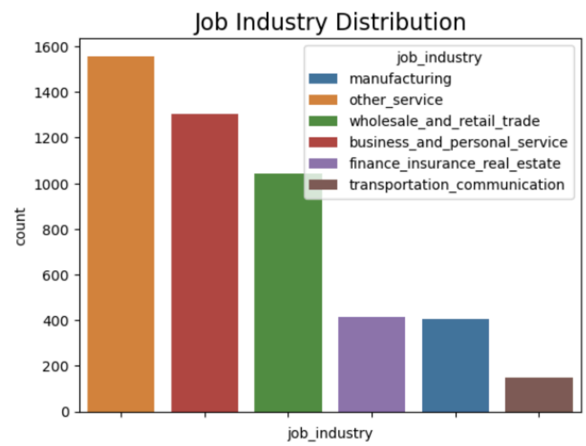
We collaborated on a majority of the project together and met frequently to discuss and work on the proposal, slideshow, and final poster and report. Abby found the source, Bertrand and Mullainathan (2004), which included the data we used and some basis for our project. As for coding, Josie cleaned and sorted the data. Visually, Abby made the odds ratio graphs, and Josie produced the job type, job industry, education required, and experience required distributions.

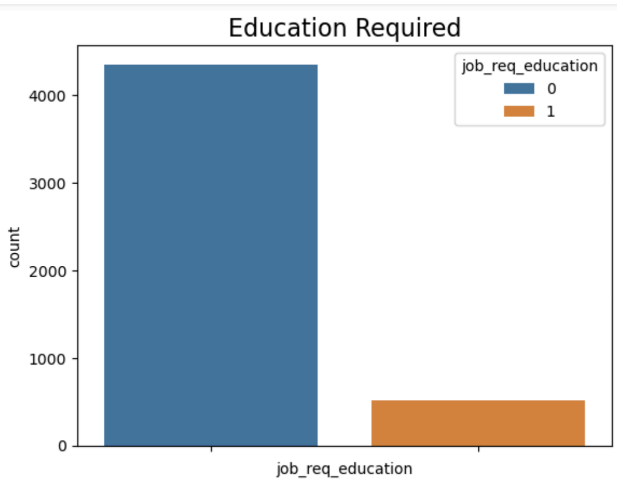
Methods

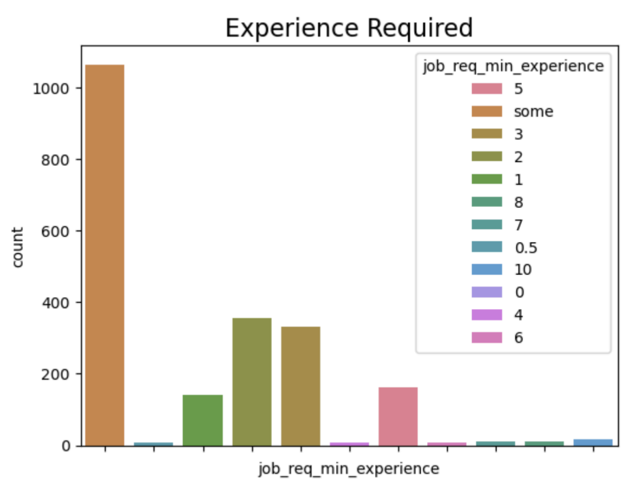
We used a data frame from a study done on callbacks from jobs in Boston and Chicago. Our data frame includes 30 variables and 4870 entries. Each entry refers to a resume sent to a specific job posting. The first name data entries are artificial and associated with a typical name for the recorded race. The variables are job ad id, job city (Boston or Chicago), job industry (manufacturing, other service, wholesale and retail trade, business and personal service, transportation and communication, finance or insurance or real estate), job type (secretary, retail sales, manager, sales representative, clerical, supervisor) if the job is a federal contractor, if the job is an equal opportunity employer, the job ownership (private, nonprofit, public), if the job has any requirements, if the job requires communication, if the job requires education, minimum education requirement, if the job requires computer, if the job requires organization, if the requires school, if a callback was received, firstname, race, gender, years of college, if the applicant has a college degree, if the applicant has honors, if the applicant worked during schools, years experience of the applicant, if the applicant has computer skills, if the applicant has special skills, if the applicant has volunteering, if the applicant is affiliated with the military, if the applicant has employment holes in their resume, if the resume has an email address, and resume quality rating.

For the study, callback was used for the dependent variable. The independent variables used are first name, college degree (binary), worked during school (binary), computer skills (binary), special skills (binary), employment holes (binary), resume quality, years of college, and years of experience. All other variables were dropped. The entries with Nan occur in variables that we are using, so we did not have to drop any rows of the data frame. The first name variable was replaced with a “1” if the name is white sounding, and was replaced with a “2” if the name is Black sounding. We performed a Chi-Squared Test, Logistic Regression, and Decision Tree. We also graphed the distribution of four variables that we did not consider for our study. These variables are job industry, job type, education required, and experience required. We found that most jobs are secretary jobs in other services, business, and personal services, and wholesale and retail trade. We also found that most jobs do not require education and experience. We will talk about how this affected our study in the conclusion section.









Results:

Chi-Squared Test Results: We first ran a Chi-Squared Test of Independence on our variables in order to determine which variables are connected to callback. The variables: gender, years of college, college degree, worked during school, computer skills, volunteer, resume quality, and email address did not have statistically significant p-scores. Originally, we planned to use gender in our name variable (coded 1 or 2 by race), but the Chi-Squared p-score was 0.38, which is not statistically significant. On the other hand, the race p-score was $5.00 \cdot 10^{-5}$, which is significant. Special skills and employment holes had low p-scores equal to $1.41 \cdot 10^{-14}$ and $6.91 \cdot 10^{-7}$, respectively. Honors and first name had low p-scores equal to $9.422 \cdot 10^{-7}$ and .035, respectively. In general, the variables: special skills, employment holes, honors, race, and first name were most connected to determining callback in order of lowest to highest p-score.

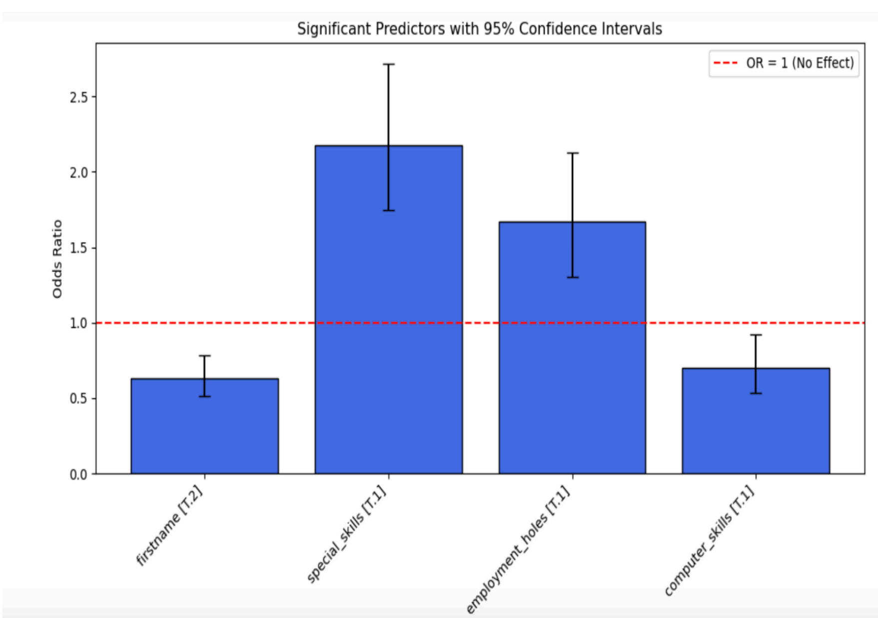
Logistic Regression Results:

The logistic regression analysis reveals several key factors that significantly influence whether an applicant receives a callback. Most notably, first name, special skills, employment holes, and years of experience emerged as the most statistically significant predictors, all with p-values indicating strong significance. Note that these p-values are not exactly zero but very close. Interestingly, traditional credentials such as a college degree (p=0.238) and having worked during school (p=0.162) showed considerably less significance in predicting callbacks. This pattern suggests that hiring managers may be prioritizing practical experience and specific

	P>|z|
Intercept	0.000
C(firstname)[T.2]	0.000
C(college_degree)[T.1]	0.238
C(worked_during_school)[T.1]	0.162
C(computer_skills)[T.1]	0.012
C(special_skills)[T.1]	0.000
C(employment_holes)[T.1]	0.000
C(resume_quality)[T.low]	0.080
years_college	0.667
years_experience	0.000

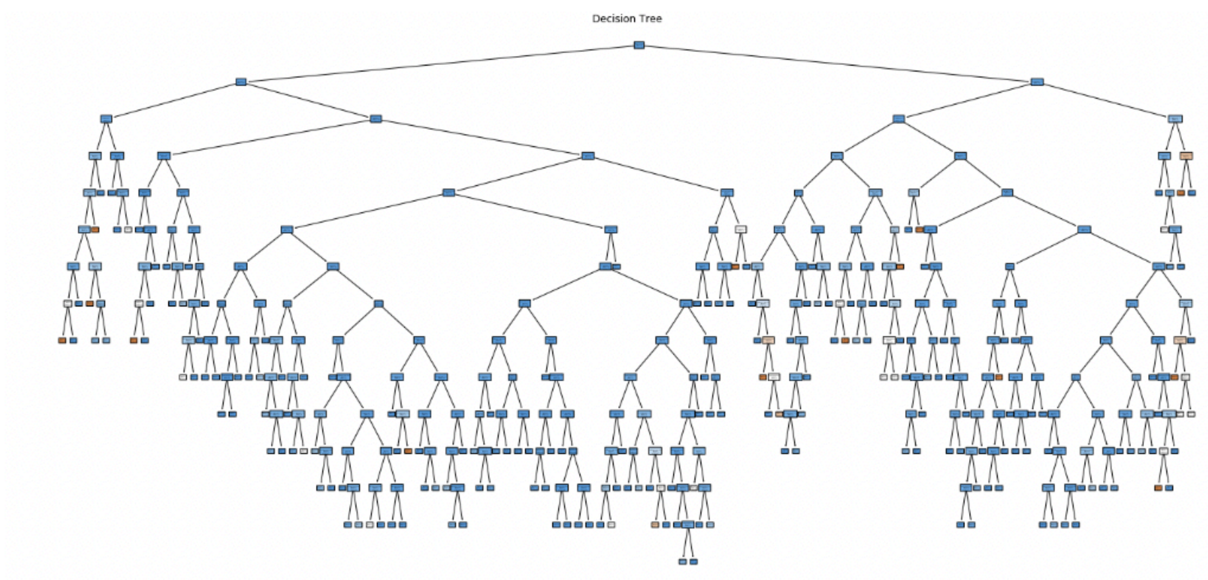
skills over educational credentials in their initial screening process. Computer skills (p=0.012) showed moderate significance, suggesting technical competencies do factor into callback decisions, though not as strongly as other variables. Additionally, years of college showed moderate significance (p=0.667), which might reflect that duration of education matters somewhat, even if degree completion itself isn't as critical for callbacks.

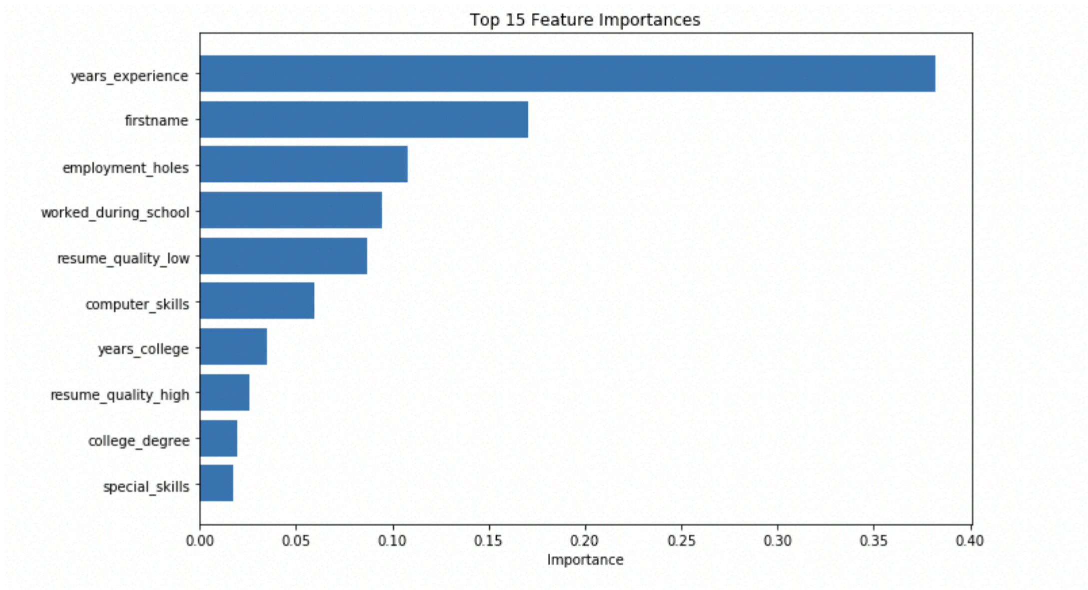
Additionally, we analyzed the odds ratio of each variable, which tells us the difference in odds of getting a callback based on one variable while holding the others constant. The proximity of the odds ratio to 1 tells us how more or less likely a candidate is to receive a callback based on each variable. Shown in the graph below, we visualized the variables with the most significant odds ratios. These results included being 36.8% less likely to receive a callback with a Black-sounding first name, being 29.6% more likely by listing computer skills, being 66.7% more likely when having gaps in employment, and including special skills, which more than doubles (117.9%) the odds of receiving a call back.



Decision Tree Results:

The decision tree model was designed to predict whether a person receives a callback based on features from their resume. After encoding the relevant data and training the model, the most influential factor was found to be years_experience, followed by firstname, employment_holes, and worked_during_school. The high importance of firstname suggests potential bias in the decision-making process, possibly reflecting discrimination based on names. The visualized tree shows how the model splits data based on these features, and the feature importance plot provides insight into which variables drive predictions most strongly. Overall, the analysis highlights both predictive patterns and ethical considerations in automated hiring processes.





Conclusion

Our models both had biases when it came to determining callback. The logistic regression model mostly had special skills and first-name bias when all variables other than the one being measured were held constant. It makes sense to us that there was a bias on special skills because they inform potential experience that an applicant brings to a job. The decision tree had a bias in years of experience and first name. Again, years of experience are an explainable bias. However, both models were biased to white sounding first names when choosing who should get a callback. First name should not be contributing to the callback, and it is not an explainable bias. Most jobs used in our data did not require education or a lot of experience or education.

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 job. Perhaps some job industries or types have hiring managers that are more biased. Additionally, our data was from 2004, and we hope that there have been improvements with the callback rate 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 applying to one specific job.

We would probably see a shift in our data if the first name and race were not included in our data. We recommend that companies assign a unique identification number to each applicant if they are using an algorithmic approach for determining callback. Then, variables that reference race would be scratched, potentially making the model less biased. We also recommend companies that are using algorithms to decide on callbacks 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. The study that we used informed us that there was bias in callback even before algorithmic approaches were considered. Companies must be very careful using past data because it can be very biased. As said before, if they are forced to use biased data, they should perform tests and make changes to improve their models.

In closing, we have determined that algorithms are quite biased when determining if an applicant should get a callback or not. Our algorithms are biased, potentially due to our data being outdated and including jobs that do not have many requirements. Perhaps, we can say that jobs that do not require much are especially biased, but we would need to test other data first.

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)