Evaluating and Modifying an Open-Source Applicant Tracking System

Diagnosing Causes of Discrimination Bias within Machine Learning Hiring Models

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Kyle Pecos

October 27, 2023

ADVISORS

MC Forelle, Department of Engineering and Society
Roseanne Vrugtman, Department of Computer Science

Introduction

With the sudden rise and rapid development of artificial intelligence (AI) technologies in the past few years, many companies have begun to heavily invest in this field (Cooban, 2023). These investments encompass a wide variety of fields including automation processes that can efficiently increase business workflow. In the wake of the recent COVID-19 pandemic, many job positions switched to being remote as a means of preventing infection and remain as such. Though these positions are widely sought out, the current demand for these remote positions far exceeds the current number of available positions in the workforce (Bhattarai, 2022). As such, companies have been utilizing AI-related technologies as part of the hiring process to more efficiently scan resumes and reduce the need for human involvement, saving company resources (Schumann et al., 2020).

These models are trained using machine learning, a subset of the field of artificial intelligence, which attempts to teach a computer program how to learn from given data without being explicitly programmed (Mahesh, 2020). While there are many different types of algorithms used for training, each with its own strengths and weaknesses, primarily it is the quality and size of the dataset fed into the model that will determine its overall accuracy and performance. In the case that a dataset is either not large or diverse enough, the model being trained might inherit a form of bias which will then affect its decisions and outcomes. Though many organizations claim to strive for better equity, inclusion, and diversity in their workforce, if precautions are not taken to address and prevent biases that may arise, these machine learning hiring models can adversely discriminate against certain minority and marginalized groups.

Despite machine learning models becoming increasingly common as part of the hiring process for companies, very few are open about how their algorithms are built, validated, and tested to prevent systemic bias in these models (Raghavan et al., 2020). Machine learning algorithms have been shown in the past to discriminate against both women (Dastin, 2018) and people of color (Schumann et al., 2020) likely as a result of biased data being used to train these models. In order to better comprehend and understand how such biases can occur in hiring algorithms as well as prevent them from occurring, I will undertake two separate projects. The first will be a technical project on testing and modifying open-source applicant tracking systems using publicly available resume datasets to determine flaws and potential modifications utilizing ensemble learning to reduce bias. The second will be an STS project focusing on analyzing and synthesizing possible reasons why biases occur in machine learning hiring algorithms and identifying methods that can help reduce and alleviate such issues.

Technical Topic

One of the main challenges that arise when attempting to analyze pre-existing hiring systems is the lack of transparency from many companies regarding the methods and datasets, they use to train their models (Raghavan et al., 2020). We cannot simply rely on promises from companies as often, they are not aware of the bias that forms within their own models. Companies evaluate models using a defined performance measure known as accuracy which can vary depending on the model's overall goal (Géron, 2019). Companies are more likely to prioritize the accuracy of a model than ensure the model remains equitable and fair in its decision-making as doing so would cost additional resources. Though larger companies are more secretive about their models, smaller organizations often do not have the resources to develop such complex systems and often use pre-existing applicant tracking software (Schumann, 2020). Some of this software is a paid service but others are open source allowing anyone to freely view, use, and modify the existing program to suit their own needs. As such, an open-source applicant tracking software will be used, analyzed, and modified for the technical project.

When it comes to the types of algorithms machine learning models can use there are two main categories: supervised and unsupervised learning. Supervised learning algorithms map inputs to outputs based on training data that has already been labeled. Datasets are divided into a training set which will teach the algorithm how to classify and a testing set to compare the accuracy of results. Unsupervised learning algorithms are not given prelabeled information that is “correct,” but are instead left on their own to draw conclusions about features in a dataset. When new data points are added, the algorithm uses previously identified features to classify the newly introduced data (Mahesh, 2020). Most hiring models only utilize supervised learning algorithms and train themselves using past applications as unsupervised learning outputs would not be as easily applicable to the hiring process.

To what extent does a skewed dataset affect the equity of decision-making in machine learning models and by what means can these models be improved to mitigate bias? In order to further research this question, I will train and test the performance of a pre-existing open-source applicant tracking system using custom-tailored datasets. Depending on the type of algorithm the applicant tracking system uses, I will add and test additional supervised methods to apply ensemble learning, a methodology that utilizes several types of algorithms to achieve an overall better performance (Géron, 2019). In 2023, despite women making up 51% of the total population in the U.S. the STEM workforce is only 35% female. Additionally, even combined, Hispanic, Black, American Indian, or Alaska Native workers only make up 24% of the STEM workforce (Grieco, 2023). To mimic these statistics, I plan on creating a dataset of resumes that roughly utilize similar percentages as those found in the STEM workforce. Following this, I will construct additional datasets with increased representation of women and minority groups. In order to gather the necessary resumes to build these datasets, I will utilize open-source data repositories such as Kaggle to gather these training data points.

To establish a baseline for comparison, I will train the unaltered applicant tracking system model using the dataset mimicking the current percentages of the STEM workforce. As this model will be trained on a skewed dataset, the model is likely to develop possible biases that can affect its decision-making on resumes that come from groups not fully represented in the training set. Afterward, I will train additional models of the unaltered system on the more diverse datasets which would probabilistically be less biased than the baseline. The same testing set will consist of resumes solely from minority groups and will be evaluated on every model. The score metrics each model produces for the test set will then be compared to one another. It is through this experiment that I hope to demonstrate how using diversified datasets can lessen the chances for biases to develop in machine learning hiring models. Subsequently, I will test the added supervised methods to see if a specific combination of algorithms can produce a model that makes more equitable decisions when compared to the baseline model that does not have any modifications. I wish to show how utilizing ensemble learning as part of developing a machine learning hiring model can assist in mitigating bias in the case it is not possible to gather more diverse datasets.

STS Topic

What historical, economic, and political factors have contributed to companies using biased, inequitable datasets as part of training machine learning models for their hiring process?

Machine learning hiring models are a black box model to most individuals who simply recognize the results without fully understanding the inner workings of the system. If effort is not put forward into understanding the design of a model and its possible downfalls, it is highly probable that the model will have bias, be it intentional or not. In a case study involving eighteen different vendors who utilized machine learning models as part of the hiring process, it was noted that only seven vendors made concrete claims regarding combating bias within their systems. Even fewer vendors disclosed if their models were properly validated for equity and whether the use of these algorithms could be justified over human recruiters. It is in a company's best interest to identify possible avenues where bias and attempt to mitigate them as under many countries' laws, employers are legally responsible if their hiring practices are shown to be discriminatory (Zschirnt 2016). Failing to address such biases can lead to a lack of differing perspectives and a homogenization of viewpoints. Diversity in the workplace has been shown to increase creativity, innovation, and financial performance of companies and thus should be strived for (Roberge & Van Dick, 2010).

In the case that training datasets are not properly validated for accuracy, quality, and size, it is possible for the model to either underfit or overfit the data. Overfitting occurs when an algorithm performs well on the training data but not on other sets of data, whereas underfitting arises when a model is too simple to make accurate predictions (Géron, 2019). Possible types of biases that might occur in a hiring model if precautions are not taken include both record and statistical bias (Costa et al., 2019). Record bias occurs when training data is incomplete or incorrect due to errors such as misspellings. Missing or incorrect data can either be omitted or replaced with a constant value for each highlighted data point, usually the mean or median of a dataset (Soofi & Awan, 2017). However, this often can lead to lower accuracy predictions for underrepresented minority groups as their entries are either removed or replaced with data that does not accurately represent them. Statistical bias occurs when the dataset is unrepresentative of a larger population which leads to a model overfitting the data (Costa et al., 2019). In the case that a dataset is not diverse enough, the produced model has a higher chance of discriminating against groups that made up the minority in the dataset as the model has fewer data points to train its decisions on.

In my STS project, I plan to examine the historical and sociopolitical aspects that would cause datasets to be skewed towards producing bias in machine learning models that are utilized as part of a company's hiring process and use the framework of machine ethics to develop solutions to help combat this issue. Machine ethics emphasizes moral decision-making in algorithmic programming for the technical, social, and political fields (Allen et. al., 2006). Machine learning results should be viewed with an agnostic lens when interpreting the social sciences as no one method or model can properly encompass all of the intricacies and interpretations of this field (Grimmer et al., 2021). First, I will analyze publicly disclosed information from companies on the details regarding their hiring models and whether they attempt to justify the use of these systems. I hope to discover and research what background factors would possibly lead companies to ignore the issue of bias arising in their systems. Next, I will research how record and structural bias arose in Amazon's machine learning model which was shown to have discriminated against female applicants (Dastin, 2018). This model was trained on recognizing patterns from resumes that were submitted over a ten-year timeframe with most of these applications coming from men. Examining the contrasting viewpoints of the developers and higher-ups as well as Amazon's history regarding their hiring practices can give further insight to explain why training sets in certain industries are often skewed and for what reasons biased datasets are permitted to be used. Finally, I will synthesize research demonstrating possible solutions to mitigating bias within models from machine learning experts through a literature review. Though not all solutions might be feasible, it is hopeful that some will be applicable and practical for use in modern-day companies to eliminate bias in the workplace.

Conclusion

By researching and developing these two projects, it is my hope that the produced results will inspire both developers and lawmakers to push for the development of more transparent and equitable hiring models for all. I wish to demonstrate the importance of utilizing diverse training datasets and how ensemble learning can assist in mitigating discrimination through my technical project which will produce quantifiable prediction scores and a modified applicant tracking system. My STS project will investigate why certain companies do not make eliminating hiring bias a high priority, analyze a case study involving Amazon’s discriminatory candidate rating algorithm, and synthesize methods to mitigate bias within machine learning hiring models. I wish to not only gain a more comprehensive understanding of how these biases are formed and proliferated but also to find practical methods to diminish the discrimination these biased models produce.

References

Allen, C., Wallach, W., & Smit, I. (2006). Why machine ethics?. *IEEE Intelligent Systems*, *21*(4), 12-17. https://doi.org/10.1109/MIS.2006.83

Bhattarai, A. (2022). The great mismatch: Remote Jobs are in demand, but positions are drying up. *The Washington Post*. https://www.washingtonpost.com/business/2022/11/27/remote-jobs-economy/

Cooban, A. (2023). AI investment is booming. how much is hype? *CNN Business*. https://www.cnn.com/2023/07/23/business/ai-vc-investment-dot-com-bubble/

Costa, A., Cheung, C. & Langenkamp, M. (2019). Technical Background. *Hiring Fairly in the Age of Algorithms*, 9-18. http://dx.doi.org/10.2139/ssrn.3723046

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. (2nd ed.). O'Reilly Media, Inc.

Grieco, E., Deitz, S. (2023). Report. National Center for Science and Engineering Statistics. https://ncses.nsf.gov/pubs/nsf23315/report

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine Learning for Social Science: An Agnostic Approach. *Annual Review of Political Science*, *24*(1), 395-419. https://doi.org/10.1146/annurev-polisci-053119-015921

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*, *9*(1), 381-386. https://www.ijsr.net/archive/v9i1/ART20203995.pdf

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 469-481. https://doi.org/10.1145/3351095.3372828

Roberge, M. É., & Van Dick, R. (2010). Recognizing the benefits of diversity: When and how does diversity increase group performance?. *Human Resource Management Review*, *20*(4), 295-308. https://doi.org/10.1016/j.hrmr.2009.09.002

Schumann, C., Foster, J. S., Mattei, N., & Dickerson, J. P. (2020). We Need Fairness and Explainability in Algorithmic Hiring. *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, 1716–1720. https://par.nsf.gov/servlets/purl/10214110

Soofi, A. A., & Awan, A. (2017). Classification techniques in machine learning: applications and issues. *Journal of Basic & Applied Sciences*, *13*(1), 459-465. https://pdfs.semanticscholar.org/2678/e213cec548d278879ceaf01582ee8913cc3f.pdf

Zschirnt, E. (2016). *Measuring Hiring Discrimination - A History of Field Experiments in Discrimination Research* (NCCR Working Paper No. 7). National Center of Competence in Research. http://dx.doi.org/10.2139/ssrn.2780398