

# When Small Decisions Have Big Impact: Fairness Implications of Algorithmic Profiling Schemes

Christoph Kern<sup>1,\*</sup>, Ruben L. Bach<sup>2</sup>, Hannah Mautner<sup>3</sup> and Frauke Kreuter<sup>1,4</sup>

<sup>1</sup>Department of Statistics, LMU Munich, Germany

<sup>2</sup>Mannheim Centre for European Social Research, University of Mannheim, Germany

<sup>3</sup>dmTECH, Karlsruhe, Germany

<sup>4</sup>Joint Program in Survey Methodology, University of Maryland, USA

## Abstract

Algorithmic profiling is increasingly used in the public sector with the hope to allocate limited public resources more effectively and objectively. One example is the prediction-based profiling of job seekers to guide the allocation of support measures by public employment services. However, empirical evaluations of unintended discrimination and fairness concerns are rare in this context. We systematically compare and evaluate statistical models for predicting job seekers' risk of becoming long-term unemployed with respect to subgroup prediction performance, fairness metrics, and vulnerabilities to data analysis decisions using large-scale German administrative data. We show that despite achieving high prediction performance on average, profiling models can be considerably less accurate for vulnerable social subgroups and that different classification policies can have very different fairness implications.

## Keywords

Algorithmic Fairness, Modeling Decisions, Statistical Profiling

## 1. Motivation

The field of fairness in machine learning (fairML) has made considerable progress in proposing fairness notions and metrics to assess biases of prediction models [1, 2]. As the development of fairML methodology is often centered around a limited number of (U.S.-based) benchmark data sets [3], their systematic application in real-world scenarios, however, lags behind. This is particularly the case for high-stakes ADM applications in the public sector as agencies may not disclose detailed documentation of their profiling models and data access is restricted. Nonetheless, ADM approaches such as the AMAS model to classify job seekers in Austria [4] have received considerable public attention due to concerns of algorithmic biases. Following preliminary work on fairness implications of algorithmic profiling of job seekers [5, 6], we set out to conduct a systematic fairness evaluation of profiling models using real-world administrative data with labor market histories of over 300,000 German job seekers.

Facing limited resources, many public employment services (PES) apply profiling to efficiently prevent long-term unemployment (LTU) [7, 8]. Profiling is used at entry into unemployment


---


*EWAF'23: European Workshop on Algorithmic Fairness, June 7–9, 2023, Winterthur, Switzerland*

\*Corresponding author.

✉ christoph.kern@stat.uni-muenchen.de (C. Kern)

ORCID 0000-0001-7363-4299 (C. Kern); 0000-0001-5690-2829 (R. L. Bach); 0000-0002-7339-2645 (F. Kreuter)

 © 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

such that a PES caseworker can intervene early on and, e.g., support individuals at risk of LTU in resuming work through targeted support programs. Implementing an algorithmic profiling system to target job seekers in practice involves a number of critical design decisions, however. Questions that need to be answered include, for example, what type of prediction method should be applied? Which type of information should be used for model training? How should resources be allocated based on a prediction model's outputs? Eventually, such decisions can substantially affect the extent to which different societal groups are targeted by support programs. This especially includes the risk of perpetuating discrimination against historically disadvantaged groups, as debated in the context of the AMAS model [4].

Against this background, we compare and evaluate algorithmic profiling models for predicting job seekers' risk of becoming long-term unemployed with respect to (subgroup) prediction performance, fairness metrics, and vulnerabilities to data analysis decisions in this study. Focusing on Germany as a use case, we evaluate profiling models by utilizing administrative data on job seekers' employment histories that are routinely collected by German public employment services. Our contribution to the literature on algorithmic profiling and fairness in profiling is twofold: (1) We conduct a systematic fairness auditing of different prediction models and report on the implications of implementing algorithmic profiling of job seekers in a European use case under realistic conditions. (2) We evaluate fairness implications of data analysis decisions such as using different classification thresholds and training data histories. This analysis shows how modeling decisions along the prediction pipeline can have group-specific downstream effects with a focus on the eventual allocation of support measures.

## 2. Methods and Results

We use regression and tree ensemble techniques to build profiling models. For each technique, we train multiple sets of prediction models that differ in the time frame and features that are used for model training. For each model, three classification policies for prioritizing job seekers are implemented that focus on very high, high and medium predicted risks of LTU. Next to comparing the profiling models with respect to group-specific prediction performance, we study fairness implications of the models' classifications based on (conditional) statistical parity difference, false negative rate difference and consistency in two evaluation data sets.

We focus on four groups of job seekers: Female, non-German (i.e., foreign-born), female non-German and male non-German individuals. Numerous studies have shown that women and individuals with a migration background are disproportionately affected by unemployment and have lower job prospects [for Germany, see 9, 10, 11]. There is consistent experimental evidence that part of these differences can be attributed to statistical (stereotyping based on assumed group averages) and taste-based (prejudice against minority groups) discrimination in hiring decisions [12]. Our fairness evaluation therefore aims to study whether discrimination against these groups would be learned and eventually perpetuated or mitigated under a given algorithmic profiling scheme.

Our results show that applying a standard machine learning pipeline to administrative labor market data can have detrimental consequences for the individuals that would be affected by the models' predictions. While our profiling models achieve good overall performance

scores that are comparable with results reported in other countries, strong differences in prediction performance across groups emerge. While the models perform similarly well for female job seekers, predictions are less accurate for foreign-born job seekers. This is particularly troubling given the history of discrimination on the labor market based on ethnicity. The drop in performance is consistent across model types, feature sets and training histories and clearly visible for both evaluation data sets.

In the light of group-specific prediction error, choosing between different classification thresholds has considerable fairness implications. Focusing on statistical parity, we observe group differences in the proportions of unemployment episodes that are predicted as LTU that exceed true differences in base rates and are highly sensitive to the classification threshold. Foreign-born (non-German) job seekers may have a higher or lower chance of being eligible for support measures than German job seekers, depending on whether high or medium risk individuals would be targeted by PES. Turning to false negative rates, it becomes evident that the observed parity differences can in part be attributed to systematic prediction error. Compared to German job seekers, true LTU episodes of foreign-born job seekers are often not correctly detected by the profiling models under high risk classification policies. The opposite holds true under a medium risk policy.

We highlight that different thresholds do not only imply different precision-recall trade-offs, but also different amplifications of group-specific biases. That is, the allocation of resources based on predictions may not only be differently (in)efficient, but also discriminatory against social groups to different degrees. As structural differences on the labor market are (over)incorporated into profiling models, their predictions can be used to either mitigate or reinforce group differences, depending on the choice of the intervention regime. Against this background, awareness of the learned group-specific patterns and errors is essential for guiding informed discussions between developers, policy makers and PES stakeholders.

## Acknowledgments

Christoph Kern's and Ruben Bach's work was supported by the Baden-Württemberg Stiftung (grant "FairADM – Fairness in Algorithmic Decision Making" to Ruben Bach, Christoph Kern and Frauke Kreuter). Hannah Mautner worked on the project while she was a graduate student at the Institute for Employment Research (IAB) in Nuremberg, Germany.

## References

- [1] S. Mitchell, E. Potash, S. Barocas, A. D'Amour, K. Lum, Algorithmic fairness: Choices, assumptions, and definitions, *Annual Review of Statistics and Its Application* 8 (2021) 141–163. doi:10.1146/annurev-statistics-042720-125902.
- [2] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, A survey on bias and fairness in machine learning, *ACM Comput. Surv.* 54 (2021). doi:10.1145/3457607.
- [3] A. Fabris, S. Messina, G. Silvello, G. A. Susto, Algorithmic fairness datasets: the story so far, *Data Mining and Knowledge Discovery* 36 (2022) 2074–2152. doi:10.1007/s10618-022-00854-z.

- [4] J. Holl, G. Kernbeiß, M. Wagner-Pinter, Das AMS-Arbeitsmarktchancen-modell, 2018. [https://ams-forschungsnetzwerk.at/downloadpub/arbeitsmarktchancen\\_methode\\_%20dokumentation.pdf](https://ams-forschungsnetzwerk.at/downloadpub/arbeitsmarktchancen_methode_%20dokumentation.pdf).
- [5] S. Desiere, L. Struyven, Using artificial intelligence to classify jobseekers: The accuracy-equity trade-off, *Journal of Social Policy* 50 (2021) 367–385. doi:10.1017/S0047279420000203.
- [6] D. Allhutter, F. Cech, F. Fischer, G. Grill, A. Mager, Algorithmic profiling of job seekers in Austria: How austerity politics are made effective, *Frontiers in Big Data* 3 (2020). URL: <https://www.frontiersin.org/article/10.3389/fdata.2020.00005/full>. doi:10.3389/fdata.2020.00005.
- [7] A. Loxha, M. Morgandi, Profiling the unemployed: a review of OECD experiences and implications for emerging economies, *Social Protection and labor discussion paper SP 1424* (2014).
- [8] J. Körtner, G. Bonoli, Predictive algorithms in the delivery of public employment services, <https://osf.io/j7r8y/download>, 2021. Accessed December 27, 2022.
- [9] I. Kogan, New immigrants—old disadvantage patterns? labour market integration of recent immigrants into germany, *International Migration* 49 (2011) 91–117.
- [10] M. Arntz, R. A. Wilke, Unemployment duration in germany: individual and regional determinants of local job finding, migration and subsidized employment, *Regional Studies* 43 (2009) 43–61.
- [11] M. Jacob, C. Kleinert, Marriage, gender, and class: The effects of partner resources on unemployment exit in germany, *Social Forces* 92 (2014) 839–871.
- [12] D. Neumark, Experimental research on labor market discrimination, *Journal of Economic Literature* 56 (2018) 799–866. doi:10.1257/jel.20161309.