

Maximizing Fairness with Synthetic data: Access to Emergency Fund in Sub-Saharan Africa

Charavee Basnet Chettri

and

Vivian Wei

and

Betty Pu

and

Ziyue Yang *

Department of Statistical and Data Sciences, Smith College, Northampton MA

March 5, 2024

Abstract

Financial inclusion is paramount for economic stability and resilience, particularly in diverse regions like Sub-Saharan Africa, spanning low to high-income countries and encompassing both resource-intensive and non-resource-intensive economies. This study focuses on a crucial aspect of financial resilience: the accessibility of emergency funds, defined as having access to 1/20 of Gross National Income (GNI) per capita in local currency within 30 days. Leveraging previous colleagues' exploratory work on the Global Financial Inclusion Database 2021, our objective is to mitigate inherent gender biases in the dataset by rebalancing it with synthetic data, thereby enhancing fairness in predicting emergency fund accessibility. Through predictive machine learning modeling, we aim to contribute to the economic empowerment of individuals in Sub-Saharan Africa, ultimately fostering resilience and reducing disparities in access to essential financial resources.

Keywords: Fairness; predictive machine learning; emergency funds; financial inclusion; ethics; gender bias; debiasing; synthetic data; Sub-Saharan Africa; fairness metrics; Global Financial Inclusion Database

*We are grateful to the Women at the Table team and Sofia Kypraiou for the project inspiration and suggesting the path forward

1 Introduction

In the realm of financial inclusion, the accessibility of emergency funds plays a pivotal role in determining an individual's financial stability and resilience, especially in developing countries.[1] In this project, our goal is to predict the possibility for people in Sub-Saharan African countries to come up with emergency funds, defined as 1/20 of GNI per capita in local currency, within a 30-day period[2]. This prediction serves as a crucial factor for establishing future public financial policies, such as determining the eligibility of individuals for loans and financial assistance. The significance of this problem lies within its direct impact on the economic well-being and empowerment of individuals in developing regions. According to the Global Financial Inclusion (Global Findex) Database 2021 published by the World Bank, only a little over half of people over 15 years of age in developing economies could access extra funds within 30 days if faced with unexpected expenses [2]. Therefore, there is a pressing need to understand the factors influencing this accessibility and eliminate inherent bias in the dataset. By delving into this issue, we not only contribute to enhancing financial inclusion but also aid in mitigating the different effects of financial shocks on vulnerable populations.

Defining fairness is essential since the concept itself is relative among different people. In the data science discourse, fairness encompasses three key aspects: individual fairness, group fairness, and causal fairness[4]. Individual fairness focuses on preventing discrimination against individuals with similar relevant characteristics. This means ensuring that individuals in similar situations receive similar outcomes from the model, regardless of irrelevant factors[3]. Group fairness aims to prevent disparities in outcomes for different groups. This ensures equal opportunities for all groups, regardless of their membership[8]. Causal fairness goes beyond simply observing disparities and delves into understanding

their underlying causes. It seeks to mitigate these root causes to achieve fair outcomes within and across groups[6].

Our previous colleagues conducted analysis using the <AI & Equality> Human Rights Toolbox and used a Decision Tree Classifier machine learning model implemented via Python to predict access to emergency funds with 68% accuracy. Their work laid a solid foundation by exploring demographic and financial variables within the dataset, and assessed the fairness of the decision tree classifier, particularly concerning gender bias, and then applied various processing techniques to enhance the fairness of the model[7].

Based on their work, our group aims to incorporate synthetic data to rebalance the dataset, ensuring equitable representations amongst both genders in the dataset. Our approach aims to consider a broader selection of machine learning algorithms and mitigate the disparities the previous colleagues found in the decision tree’s classifier’s predictions and enhance fairness with synthetic data, ultimately better predict access to emergency funds in south-Saharan Africa countries[5].

References

- [1] Rashmi Umesh Arora. 222Links between financial inclusion and financial stability: A study of BRICS. In *Handbook of BRICS and Emerging Economies*. Oxford University Press, 12 2020.
- [2] Asli Demirgüç-Kunt, Leora Klapper, Dorothe Singer, and Saniya Ansar. *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19*. World Bank, Washington, DC, 2022.
- [3] Will Fleisher. What’s fair about individual fairness? In *Proceedings of the 2021*

AAAI/ACM Conference on AI, Ethics, and Society, AIES '21, page 480–490, New York, NY, USA, 2021. Association for Computing Machinery.

- [4] Sofia Kypraiou. What is Fairness? *Feminist AI*, sep 13 2021. <https://feministai.pubpub.org/pub/what-is-fairness->.
- [5] Kypraiou, Sofia. Synthetic data for AI equality. <https://aiequalitytoolbox.com/wp-content/uploads/2023/10/Copy-of-Synthetic-Data-for-AI-Equality-website.pdf>, 2023. Accessed on January 29, 2024.
- [6] Drago Plecko and Elias Bareinboim. Causal fairness analysis, 2022.
- [7] Rose Porta, Alejandra Munoz Garcia, Margaret Bassney, and Aushanae Haller. Analysis of access to emergency funds in sub-saharan countries– a human rights-based approach. Manuscript under preparation. Department of Statistical and Data Sciences, Smith College, Northampton MA, 2022.
- [8] Tim R  z. Group fairness: Independence revisited. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 129–137, New York, NY, USA, 2021. Association for Computing Machinery.