Justice in Numbers:Unveiling Bias in UCI Communities and Crime Data

Aswathy Gopalakrishnan, Jini Saroja, Sneha Mathummal Parapurath, Yonge Li

March 14, 2025

1 Motivation

Artificial intelligence is rapidly making its mark in all sectors of the society. But the safety and ethical concerns of these models are still questionable. While some can enhance efficiency, the context in which the model is used greatly determines its impact, especially in sensitive areas like criminal justice. According to the EU AI act, risks associated are categorized into four tiers, and assessing criminal risks comes under the top tier of unacceptable risks. So, is it really worth using AI for criminal justice?

This assignment aims to analyse the UCI communities and crime dataset to understand the underlying biases. By unveiling these biases, we aim to evaluate the ethical viability of the dataset, and also to propose strategies to enhance fairness, equity, and transparency in the model.

2 Implementation

2.1 Data Analysis

In this phase, the features were analysed and categorised. A correlation analysis was done to gain a deeper understanding of their relevance, interplay between them, and their dependencies on the target attribute, 'ViolentCrimesPerPop'. Both the positive and negative correlations were included to understand how crimes rates are affected.

2.2 Data Cleaning

Before proceeding with detailed analysis, data cleaning was done to ensure dataset's integrity. The columns that were found to be irrelevant from the correlation analysis were removed. Also, any columns, that had more than 50% missing data, were also eliminated.

2.3 Bias Identification and Analysis

In the analysis for bias identification, the aim was to uncover potential biases in various categories that can lead to violent crime rates. The racial features, age features, employment conditions, socioeconomic influences, age trends, divorce rates, and family structures were analysed to see their influence on violent crime rates.

Key biases identified are:

- Selection bias: The dataset is not representative of the population.
- Sample size issues: Underrepresentation of some racial groups.

- Geographical bias: High crime rates associated with specific racial demographics.
- Optimist bias: Data reaffirming an existing belief.[1]

NOTE: The visualizations along with detailed descriptions can be found in the code. They are not included here due to page constraints.

2.4 Modeling

XGBoost regressor[2] was used for modelling along with RandomizedSearchCV for hyperparameter optimization. This approach is helpful to find best model with good combination of parameters. In order to reduce the dimensionality of the dataset and to capture maximum variance in the features, Principal Component Analysis (PCA) technique was implemented. Once the training is completed, then the model was evaluated using standard performance metrics such as RMSE [3]and R2. Model is then tested with the test set.

2.5 Fairness metrics

Various biases were identified during data analysis phase, but for fairness evaluation only racial bias was considered. Using AIF360 library[4], fairness metric for each sensitive attribute is calculated with Disparate Impact, mean difference and statistical parity difference. Results are analysed and visualized to identify potential bias in the model. It was observed that the model strongly favours racePctWhite group while racepctblack, racePctHisp group experience notable disadvantages.

2.6 Bias Mitigation and Comparison

Fairness metric shows significant disparities across racial groups. To mitigate this disparity, Reweighing technique from AIF360[4] is used. It is a preprocessing model which helps to mitigate the bias before model training. This will adjust weights to balance privileged and underprivileged groups, to reduce the effects of biased dataset, ensuring that different racial groups contribute equally to model training, thereby improving fairness. Fairness metrics is recalculated after training model with the reweighed dataset and compared with the initial fairness metrics. The analysis showed a great reduction in disparities, confirming the effectiveness of reweighing technique. Model performance remained largely unchanged, indicating that fairness was improved without compromising on accuracy.

NOTE: For coding, GitHub[2][4] code is used as a reference. ChatGPT was used as an aid to support difficult parts of coding.

3 Conclusion

Biases were identified and strategies were implemented to enhance model's fairness. However, we have concluded that certain biases, such as optimistic bias, cannot be fully mitigated. The population of white remains higher than other races which creates a limitation in data. Removing this as in the mitigation strategy may reduce the bias in the model. But this actually decreased the model's accuracy. This indicates that some biases are necessary. These are the biases that reflect deeper societal issues where long-lasting problems like racism, poverty, lack of education, low income, and unemployment continue to prevail among the different races[6]. Therefore, even if we remove the biases in model, the societal bias remains. This highlights another crucial point. A model may be unbiased, but the users of the system may bring their own biases, causing similar issues. We wish it were straightforward to eliminate societal bias as it is in the model. It isn't and hence we conclude that these datasets are indeed unacceptable to use for crime analysis.

References

[1] C. W. Kramer, "Systemic Racism in Crime: Do Blacks Commit More Crimes than Whites?," Liberty Matters, Liberty Fund, Feb. 2024
https://oll.libertyfund.org/publications/liberty-matters/2024-02-13-systemic-racism-in-crime-do-blacks-commit-more-crimes-than-whites, (accessed Feb. 2025).

[2] VerbalKint95, "Fairness Project README," GitHub,
https://github.com/VerbalKint95/fairness/blob/main/README.md, (accessed Mar. 2025).

[3] HolisticAI Team, "Mitigating Bias in Regression: U.S. Crime Example," HolisticAI Documentation,
https://holisticai.readthedocs.io/en/latest/gallery/tutorials/bias/mitigating_bias/regression/examples/example_us_crime.html, (accessed Mar. 2025).

[4] Trusted AI Team, "AIF360 Examples Repository," GitHub,
https://github.com/Trusted-AI/AIF360/tree/main/examples, (accessed Mar. 2025).

[5] "African-American History," Wikipedia,
https://en.wikipedia.org/wiki/African-American_history, (accessed Feb. 2025).

[6] "African-American History," Wikipedia,
https://en.wikipedia.org/wiki/African-American_history, (accessed Feb. 2025).