# Investigating Bias in Resource Allocation for Homelessness/Houselessness Prevention and Intervention

Abigail Santiago[1] , Andrew Solis[2], Dr. Anne Bowen[2]
[1]Davidson College, [2]The University of Texas at Austin, Texas Advanced Computing Center

TEXAS
The University of Texas at Austin

## Background

Many scholars studying homelessness, as well as the U.S. Department of Housing and Urban Development (HUD), are turning to modeling and machine learning to predict homelessness in order to better understand the problem and better allocate limited resources. This work includes approaches such as creating predictive models for targeting interventions (Kube, Das and Fowler, 2019), using regression models to extract common predictors (Nisar, Vachon, Horseman and Murdoch, 2019), and Allegheny County in Philadelphia's implementation of their Allegheny Housing Assessment, or AHA. However, there is no discussion having to do with possible bias influencing the decisions and the implications of machine learning being applied to homelessness resource allocation, outside of Eticas Consulting's Audit of the AHA (Eticas Research and Consulting, 2020).

## Research Focus

Can I uncover possible areas of bias in chosen machine learning algorithms by examining fairness and looking for signs of underspecification in order to better understand how these models make decisions and speak on the ethical implications of relying on these algorithms for decisions regarding homelessness?

## Methods and Experiments

### Fairness

- Data from UCI Adult dataset (income of either >50k or <50k)
- Train Gradient Booster Classifier model
  - Accuracy: 86%
- Use python's fairlearn library to visualize model's f1 score based on sensitive features (Figures 3-5)
  - F1 score encapsulates precision (exactness) and recall (completeness)
- Sensitive features chose: Gender and Race
  - Broken down into subgroups

### Underspecification

- Data from UCI Adult dataset (income of either >50k or <50k)
- Create Rashomon Sets of LinearSVC and Random Forest Classifier Models
- Apply stress tests to find signs of underspecification
  - Randomly drop 1%, 5%, 10%, and 20% of data
- Observe f1 scores across the Rashomon Set

## Results

***How do f1 scores vary across the Rashomon Sets?***

**LinearSVC**

| | 1% Dropped | 5% Dropped | 10% Dropped | 20% Dropped |
|---|---|---|---|---|
| Range of Precision Across Models | 0.86 – 0.88 | 0.86 – 0.88 | 0.86 – 0.88 | 0.86 – 0.88 |
| Range of Recall Across Models | 0.92 – 0.94 | 0.92 – 0.94 | 0.92 – 0.94 | 0.92 – 0.94 |
| Range of f1 Scores Across Models | 0.90 | 0.90 | 0.90 | 0.90 |

Figure 1: Range of Accuracy, Recall, and f1 Scores Across Stress Tests for the set of LinearSVC Models

**Random Forest Classifier**

| | 1% Dropped | 5% Dropped | 10% Dropped | 20% Dropped |
|---|---|---|---|---|
| Range of Precision Across Models | 0.88 – 0.89 | 0.88 – 0.89 | 0.88 – 0.89 | 0.88 – 0.89 |
| Range of Recall Across Models | 0.92 – 0.93 | 0.92 – 0.93 | 0.92 – 0.93 | 0.92 – 0.93 |
| Range of f1 Scores Across Models | 0.90 – 0.91 | 0.90 – 0.91 | 0.90 – 0.91 | 0.90 – 0.91 |

Figure 2: Range of Accuracy, Recall, and f1 Scores Across Stress Tests for the set of Random Forest Classification Models

***How "fair" is the model on an Individual Scale?***

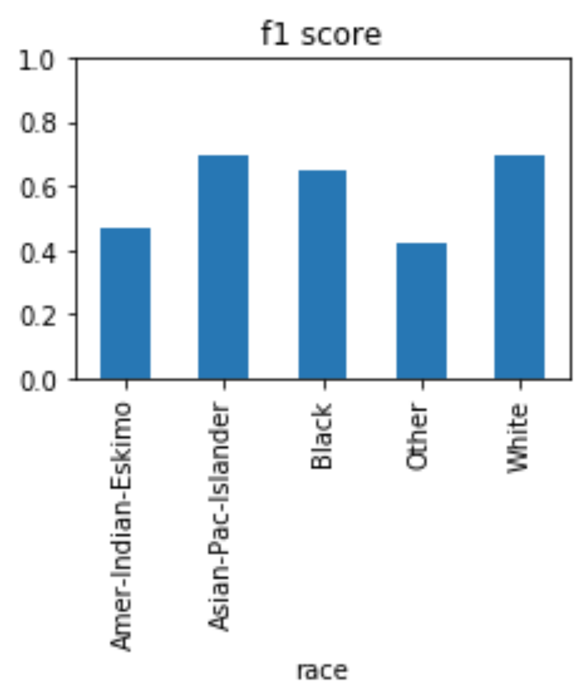


Figure 3: *F1 score of Gradient Booster Classifier Model by race*

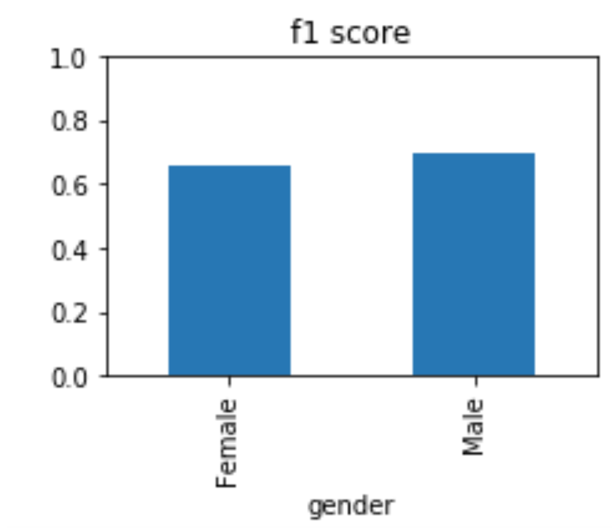


Figure 4: *F1 score of Gradient Booster Classifier Model by gender*

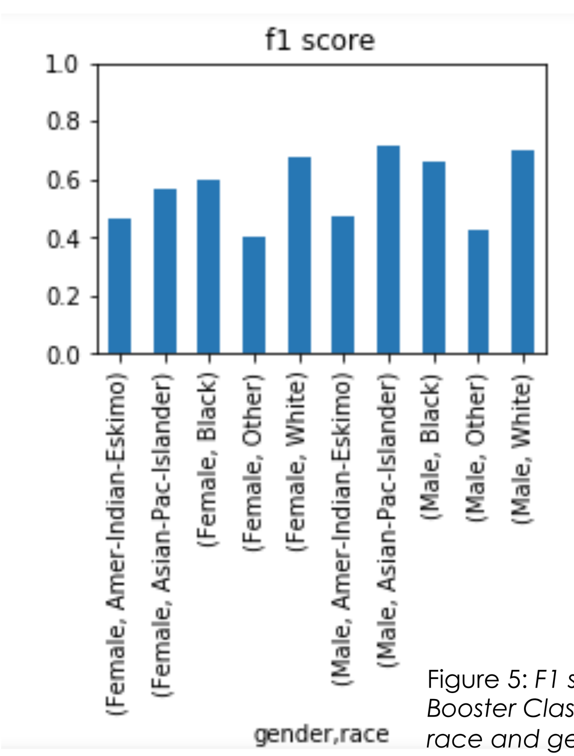


Figure 5: *F1 score of Gradient Booster Classifier Model by race and gender subgroups*

## Conclusion

- Focusing on the output of these models could lead to inequitable allocation of resources
- No signs of underspecification does not mean that this is not a possible issue
- Even if they are perfectly "fair," a greater discussion should be had about bias before relying on them to make decisions about resource allocation for homelessness

## Future Work

- Use data either from the HMIS or an integrated administrative dataset
- Rerun experiments on Rashomon sets with more in depth stress tests
- Also observe the f1 scores in the Rashomon sets across sensitive features

## Acknowledgments

Thank you to Dr. Anne Bowen and Andrew Solis for their guidance throughout this project. Thank you to Dr. Kelly Gaither for her support and help. Thank you to Rosalia Gomez, the Texas Advanced Computing Center (TACC), and the Cyberinfrastructure for Social Change REU, funded by NSF Grant #1852538,.

## References

Eticas Research and Consulting, 2020. *Algorithmic Impact Assessment of the predictive system for risk of homelessness developed for the Allegheny County*. Eticas Research and Consulting.

Kube, A., Das, S. and Fowler, P., 2019. Allocating Interventions Based on Predicted Outcomes: A Case Study on Homelessness Services. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33, pp.622-629.

Nisar, H., Vachon, M., Horseman, C. and Murdoch, J., 2019. *Market Predictors of Homelessness: How Housing and Community Factors Shape Homelessness Rates Within Continuums of Care*. U.S. Department of Housing and Urban Development Office of Policy Development and Research.

TACC
TEXAS ADVANCED COMPUTING CENTER
NSF
STAMPEDE2