**Why Is My Classifier Discriminatory?**

1. Very briefly, what is the paper generally about?
   * We give a procedure for analyzing discrimination in predictive models with respect to cost-based definitions of group fairness, emphasizing the impact of data collection.
2. Background and motivation.
   * Recent attempts to achieve fairness in predictive models focus on the balance between fairness and accuracy. In sensitive applications such as healthcare or criminal justice, this trade-off is often undesirable as any increase in prediction error could have devastating consequences.
   * In this work, we argue that the fairness of predictions should be evaluated in context of the data, and that unfairness induced by inadequate samples sizes or unmeasured predictive variables should be addressed through data collection, rather than by constraining the model.
3. What is proposed or claimed in the paper?
   * propose the use of bias-variance-noise decompositions for separating sources of discrimination.
   * suggest procedures for estimating the value of collecting additional training samples.
   * propose the use of clustering for identifying subpopulations that are discriminated against to guide additional variable collection.
4. What supporting evidence is provided?
   * Case studies on three real-world datasets
5. Why does the proposal/claim matter?
   * We identify that existing approaches for reducing discrimination induced by prediction errors may be unethical or impractical to apply in settings where predictive accuracy is critical, such as in healthcare or criminal justice. As an alternative, we propose a procedure for analyzing the different sources contributing to discrimination.
   * collection of additional samples is often sufficient to improve fairness, and that existing post-hoc methods for reducing discrimination may unnecessarily sacrifice predictive accuracy when other solutions are available.
   * identifying clusters or subpopulations with high predictive disparity would allow for more targeted ways to reduce discrimination.

Notes:

The term “discrimination" to refer to specific kinds of differences in the predictive power of models when applied to different protected groups.