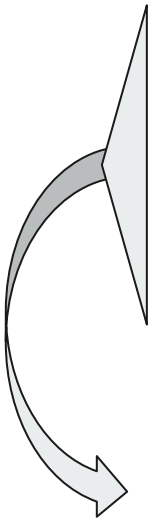
Testing Algorithmic Fairness

How to create a standard for Machine Learning Audits

Variety of Problems



- Multitude of Datasets and Models
- Each model is trained to solve a different problem
- Each problem has unique use cases, contexts, and consequences
- All fairness metrics can never be perfectly satisfied
    - (Impossibility Theorem)

How can we regulate ML models with all of this variability?

There should be some uniformed way of deciding metric thresholds

Stakeholders

Auditors

It is important to be confident that a model is fair and therefore, it is important to have a robust standards for audits

Policy Makers

In order to improve regulation for ML, policy makers must understand best practices for audits and create a standards for performance

Developers

Developers might want to know the best practices and standards for algorithmic fairness so that they can create the best product possible.

Case Study: Heart Disease Predictor

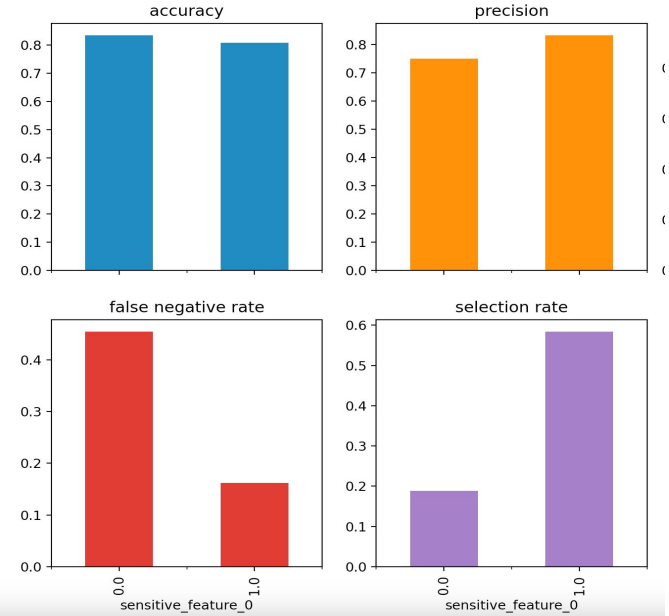
Heart Disease Health Indicators Dataset

253,680 survey responses from cleaned BRFSS 2015 - binary classification

Last Updated: 2 years ago (Version 1)

- Dataset containing information on medical measurements, diet, abilities and habits
- “Sensitive” information: Age, **Sex,** Income, Education
- Performed Logistic Regression on the data (using sklearn library)
- Analysed fairness metrics using fairlearn library
    - Focused on Sex as the sensitive attribute for assessing fairness

Results



- This demonstrates how different metrics convey different ideas of fairness
- Women (sensitive_feature = 0) seem to have better accuracy than men (sensitive_feature = 1)
- However, looking at precision or FNR, it's clear that **the algorithm is much less fair**
- It is impossible to equalize both of these metrics.

Next Steps:

1. **Compare** Logistic Regression Results with Neural Network Trained on the same data
    a. Logistic Regression models have much more interpretability as compared to Neural Networks
2. Look at common **“quick fixes”** to algorithmic fairness and see how they actually alter fairness
    a. Leveling Dataset
    b. Removing Sensitive Features (Unaware Approach)
    c. Additional Model Learning, to equalize fairness metrics (Aware Approach)
3. Meet with experts to understand how fairness assessment is done in practice, and see what regulatory auditings might look like.

Experts

Zhicheng Jiao

Medical AI researcher at Brown's Medical School ~Developer~

Curious About:

- Existing requirements for demonstrating fairness
- Process for demonstrating fairness
- Ideal/Self imposed assessment for fairness

Suresh Venkatasubramanian

Brown Computer Science Professor and Co-author of Blueprint of an AI Bill of Rights ~Policy Maker~

Curious About:

- How regulations can be consistent given variety of applications
- Can a general policy exist
- Ideal Vision for ML Regulation

Goals for Report

Systematic Way of Deciding Which Fairness Metrics to Prioritize for Each Case

After testing models and interviewing experts, I am hoping to develop a sense of what an ideal fairness auditing system might look. The goal of this report is to create a set of auditing guidelines that can be applied to all scenarios in order to determine most critical metrics that should be used. By doing this, the hope is to have an auditing system that it could be used in a regulatory way.

Challenges

How to decide which metrics to prioritize over others:

- Considering severity of consequences “is a false positive better than a false negative?”
- What IS fairness for each situation?
- How to create standards when it is so ambiguous and context dependent?