Liwen Peng
Mentor: Sarah Teichman
SPA DRP Spring 2021

Ethics of Algorithm Decision Making - Fairness

At the start of this quarter, I was introduced to some broad topics within the discussion of algorithm decision-making ethics, including privacy, freedom, independence, and fairness. This was the first time I tried to think about the implications of an algorithm, a set of rules that precisely define a sequence of operations, for individuals. It was valuable since discussions about whether people were using algorithms correctly were important.

Among a bunch of concerns people had, I mainly studied fairness in the past 10 weeks. My mentor and I covered various definitions of fairness, why there was bias, and how people could reduce bias. There were many definitions for fairness. For example, fairness might be defined as equal accuracy for all groups because we want, for example, a medical diagnostic tool to be equally accurate for people of all groups. Fairness could also be defined as equal false positive rates and false negative rates based on the idea that people with the same outcome should be treated the same for all groups. Other definitions may include equal treatment to people with the same score. These were all reasonable but were not mathematically compatible. Thus, people argued about whether an algorithm was fair.

Interestingly, even though algorithms didn't take sensitive group indicators (such as race) as covariates to generate models, their predictions might still show discrimination to sensitive groups of people. The reason might be that data was inherently biased. In this case, it was hard to reduce bias by collecting data another way. So instead, people considered decreasing discrimination at three stages: pre-processing to prepare data before model training, in-processing to modify algorithms during the training phase, and post-processing to reduce bias by manipulating predictions from the model. Specifically, for pre-processing, techniques such as massaging, reweighing, and sampling are some useful solutions.

In the last few weeks, I used one COMPAS data set to evaluate fairness and see how sampling would reduce discrimination. COMPAS is a tool provided by Northpointe to assess the risks of defendants to re-offend in the future. The results indicated that COMPAS decisions showed fairness when looking at equal accuracy but showed discrimination when considering factors like false positive rates and false negative rates. I also applied uniform sampling and preferential sampling for the data to see whether there was a reduction in bias. My results manifested both techniques had little effect, but I might improve my second technique application with a better rank score.

It was fascinating to discuss ethical problems mathematically and apply R to analyze a data set.