EECS 499 – Detecting & Correcting Bias in Algorithmic Decision Making
Professor H V Jagadish (Section 127)
Pranav Mayuram

During the Fall 2016 semester, I will be working on a project in Data Ethics under the guidance of Professor H V Jagadish. This project will attack problems of data bias in different algorithms that are commonly used for critical decision-making. Our challenge will entail looking at biases in the COMPAS system, which analyzes the risk of a criminal being arrested after their release, and is commonly used in criminal justice spaces.

Our goal will be separated into multiple parts. First will be a preliminary analysis of the data, correcting bias between specific traits, such as race or gender. This will check for a baseline error and bias, and then correct each population for their bias. Once this is done, we will attempt to compose these errors, and try to generate a model, which could mathematically describe the biases of the system. Finally, we plan to create a programmatic tool, which has knowledge of these biases, and can therefore correct them. Since the COMPAS algorithm itself is a black box, we don't know why/how its bias exists, our correcting tool would take output from COMPAS, and return corrected scores by using the aforementioned model.

I will be using my knowledge of algorithms, and their inner workings from EECS 281, as well as my experience in tackling non-concrete problems from my past research in EECS 399 under professor Jagadish. This should allow me to better understand how bias could even be introduced into the program. In terms of my math and data analysis knowledge, I will use my work from STATS 250, and potentially, agent based modeling techniques from CMPLXSYS 270. Professor Jagadish and I plan to meet approximately weekly, but that will be adjusted each week, depending on the progress necessary before our next meeting. Once the project is finished, I will be judged based on my data analysis work, particularly the accuracy, and thoroughness with which I complete my analysis. There will be some code written to perform my analysis, but the quality of this is not the basis on which I will be judged. Finally, there will be a write-up detailing the specifics of my work with techniques used, etc.