HW 5

Holden Wright

12/29/2023

This homework is meant to give you practice in creating and defending a position with both statistical and philosophical evidence. We have now extensively talked about the COMPAS [1] data set, the flaws in applying it but also its potential upside if its shortcomings can be overlooked. We have also spent time in class verbally assessing positions both for an against applying this data set in real life. In no more than two pages [2] take the persona of a statistical consultant advising a judge as to whether they should include the results of the COMPAS algorithm in their decision making process for granting parole. First clearly articulate your position (whether the algorithm should be used or not) and then defend said position using both statistical and philosophical evidence. Your paper will be grade both on the merits of its persuasive appeal but also the applicability of the statistical and philosophical evidence cited.

As a statistical consultant I recommend the court does not employ the use of the COMPAS algorithm in decision making processes when deciding whether or not to grant parole. An argument can be made that an algorithm such as COMPAS that relies solely on numbers should be able to remove some systemic bias from a decision made by humans who has inherent biases. However, in the case of COMPAS the algorithm has been shown to have only mildly accurate predictions, not much better than a coin flip, and the black box nature of the algorithm prevents the public from auditing it and working to remove biases and improve the algorithm. The algorithm's failure to satisfy equalized odds and other metrics of statistical fairness like disparate impact and statistical parity adds more fuel to ethical argument against the use of the COMPAS algorithm. The unfairness of the algorithm manifests itself in the form of racial biases that undermine the idea that everyone is equal under the law in our justice system.

While the COMPAS algorithm does not satisfy independence metrics like disparate impact and statistical parity, I do not believe that this poses as much of a problem as the algorithms failure to solve the separation metric, equalized odds. To satisfy independence criteria the algorithm's prediction must not depend on the protected class, in this case race. However, it does not take into account the ground truth. Separation metrics like equalized odds permit that the algorithm must not have an inequality in false positive rates or true positive rates between the protected classes. If the algorithm were to not satisfy independence criteria but were to satisfy separation criteria this would not be a reflection of the bias of the algorithm but rather a reflection of systemic racism at work in our society. In the case that the algorithm is more likely to predict recidivism in a minority than a non-minority but is equally likely to correctly predict recidivism in a minority and a non-minority then the algorithm is factoring in inherent racial bias in our society and accounts for it in its predictions, but not to an unfair extent. Unfortunately the COMPAS algorithm does not fall into this camp. The COMPAS algorithm does not satisfy separation criteria and has a higher false positive rate for minorities than non-minorities. A higher false positive rate means that the algorithm is more likely to incorrectly predict recidivism for minorities than non-minorities. This is a major issue with the algorithm and should bring any conclusion drawn from it into question.

Judging an algorithm off statistical metrics of fairness is incomplete without bringing in philosophical metrics of fairness. There are three broad frameworks of fairness and justice: need, equality, and merit, but for the purposes of judging the COMPAS algorithm I will focus on need and merit. From a need standpoint the COMPAS algorithm is a clear failure. It has clear inherent biases against minorities and does not put

[1] https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis
[2] knit to a pdf to ensure page count

every case on a level playing field. Looking at COMPAS using John Rawls' thought experiment, the Veil of Ignorance, the algorithm is clearly taking characteristics like race into account and unfairly judging certain individuals based off of that. Looking at COMPAS from a merit standpoint leads us to the same unjust conclusion. While merit equality offers a path to judge classes differently, it does so under certain conditions. The difference in judgement must be based on merit and has to be earned. The difference in false positive rates produced by COMPAS shows that the difference in judgement is not based on merit and is instead based on bias.

Given the statistical biases and philosophical dilemmas associated with the COMPAS algorithm, I recommend that it not be used as a factor in parole determinations. The algorithm's known biases and lack of transparency risk undermining the fairness and moral responsibility that are important to a proper justice system.