

## HW 5

Sam Pell

03/27/2024

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 <sup>1</sup> 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 and against applying this data set in real life. In no more than two pages <sup>2</sup> 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.

### *STUDENT RESPONSE*

The COMPAS algorithm's use in legal proceedings is inherently immoral. Whether as a final decision or as an aid to the decision of a judge, it should not be allowed to impact the decision of a court, and should be left out of the legal system. The COMPAS Algorithm is a black box system that unfairly discriminates against black defendants, and does so using statistically unsound methods, namely, the use of proxy variables. The algorithm then compiles the data it has to predict how likely a person is to commit another crime in the next two years, if released. The result is a score from 0 to 10, arranged into 'Low,' 'Medium,' and 'High' risk, which is then presented to a judge. While there is certainly merit to data-aided decision-making, there are significant ethical issues with not only this tool, but also the process it is being used for, criminal justice. For one, the tool is not particularly effective in its prediction rates. It correctly classified defendants just sixty-five percent of the time. While that may seem encouraging, it is best to remember that in classification tasks like this, a minimum rate of fifty percent is expected simply by choosing randomly. Meaning this algorithm only outperformed a coin flip by about fifteen percent. Furthermore, its statistical bias was shifted toward white defendants. The white defendants tended to have lower false positive rates, as well as higher false negative rates (false positives being where COMPAS predicted their recidivism incorrectly such that they would offend, and false negatives being where COMPAS predicted them to be safe or non-repeat offenders, yet they were). Meanwhile, black defendants saw much the opposite of

---

<sup>1</sup> <https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis>

<sup>2</sup> knit to a pdf to ensure page count

this trend. This resulted in a biased system, where black people were being recommended against systematically, and in a way that was arbitrary. This is inherently unfair and unjust, as no person should have to endure arbitrary discrimination, least of all when trying to escape the American prison system. The cause of the bias was almost certainly the inclusion of proxy variables, those being variables that may represent something innocuous, like ZIP code, but which are extremely divisive along the lines of some protected class boundary. In the case of this algorithm, ZIP and other variables were known to have been used to sort the decision rule out. This means that, whether race was in the algorithm or not, it effectively was. To continue the critique, it is essential to mention that the algorithm is intended to be used as a tool, only a tool, and not the judge's entire decision rule. It is accompanied by multiple warnings, and can easily be overruled. In fact, in many cases, there is evidence that the COMPAS decision has been overruled, as it sorts people into Low, Medium, and High risk categories, but the judges often let High risk prisoners out, or force Low risk prisoners to remain. However, for some judges, this is much more of a part of their decision making process. The concern really is the slippery slope here. Some judges may be able to mitigate the effects of the algorithm's predictions, but others may allow it to dominate their decisions. This is already unfair to the defendants, as it reasserts the aforementioned arbitrary discrimination, but it also treats them as data points. Means to an end. Simple, non-human court cases. At that point, why have a judge at all? This algorithm cannot be allowed to take over the decision of a judge in this way, and people must have a chance at their day in court, at mercy. And so, in summary, while the COMPAS algorithm may have benefits for some judges, who find it cruelly efficient, it should not be used in any fair court. The slippery slope to impersonal decisions, the proxy variables, and the reaffirmations of cultural biases and arbitrary means of discrimination should be enough to convince any judge to wean off of using Northpointe's COMPAS algorithm.