# HW 5

Abigail Mabe

03/22/2024

The COMPAS algorithm should not be utilized in decision making processes aimed at granting (or denying) parole. The COMPAS algorithm itself is roughly 65% accurate, with proven statistical biases that disadvantage those within a protected class. Although the algorithm was created as a supplementary tool to aid decision-making, the arbitrary discrimination under which it makes predictions can unconsciously supersede fair judgement. The lack of accountability involved with using supplementary decision-making predictions from a black-boxed, for-profit algorithm is also an issue. Furthermore, the algorithm raises many ethical concerns as a method of "predictive policing". Therefore, due to statistical biases, ethical dilemmas, and issues with accountability, the COMPAS algorithm should be avoided in legal decision-making processes.

Statistical notions of fairness objectively determine whether measures of fairness differ between protected and unprotected classes. In context, it is a mathematical way to determine whether arbitrary discrimination against the protected class is present in the COMPAS algorithm. In this case, the protected class are black defendants, and the unprotected class are white defendants. In the following analyses, it is important to note the metric to determine whether the algorithm passes each test is based off a widely used legal convention $\epsilon = .2$. The first statistical notion of fairness the algorithm fails is disparate impact. Disparate impact measures the predictive fairness in chance of recidivism between black and white defendants; namely, it measures the equal predictive treatment between these two classes.

A similar statistical measure to disparate impact is statistical parity, in which the algorithm also fails. Two additional statistical measures of fairness to which COMPAS has been held are predictive equality and equal opportunity, which comprise the equalized odds of the algorithm. Predictive equality measures the false positive rate (the rate of predicting recidivism where the defendant "survives" or does not recidivate) between black and white defendants. The algorithm did not pass this measure of statistical fairness, meaning it displayed a significantly unequal level of false positives between the two groups. Equal opportunity, on the other hand, measures the true positive rate (the rate of predicting recidivism where the defendant does recidivate) between black and white defendants. The algorithm did pass this test, but since it did not pass the previous, it also did not pass equalized odds as a whole. Then, the algorithm does not pass the statistical measures of fairness described above. This means COMPAS objectively acts in arbitrary discrimination against black defendants, which makes it hard to justify even as a supplementary decision-maker in terms of granting parole.

Although statistical measures of fairness create an objective way to observe bias within COMPAS, it can be said the justice system is already inherently biased. Some might argue the use of an algorithm such as COMPAS could decrease unintentional biases, especially since we are aware of the arbitrary discrimination the algorithm creates. However, the use of COMPAS might instead proliferate the bias within the justice system; even if the bias is made aware by the users, the algorithmic predictions can lead to additional implicit biases. This has the possibility to create a positive feedback loop, where the inherent bias of the system feeds into the bias of the algorithm and vice-versa, although it is hard to say without knowing the workings of the algorithm.

Another issue with the use of the algorithm as a supplement to decision-making in granting parole is the concern of accountability. As a judge, if COMPAS is utilized as supplementary material in the decision to grant (or deny) parole, who would take responsibility for the decision? It is hard to place blame on an algorithm, so the blame would ultimately be placed on the judge themselves. As a judge, it might be hard to

settle with the fact that you are getting predictive information from a black-boxed, for-profit algorithm, and that you may be blamed for utilizing what it tells you. It might also be difficult to separate an algorithmic suggestion from a decision. However, it might be easier for a judge to sit with a decision that they made simply from listening to the case presented, working entirely from their own trained perspective. In this case, it also gives equal opportunity to all those who present a case without predictive bias playing a role; especially when that prediction is disproportionately based on race. It is also important to note although Northpoint Incorporated claims they did not use race in their algorithm, they could have used proxies such as zip code that indicate race and create the arbitrary discrimination shown.

Now, to view the use of COMPAS from a purely ethical standpoint, there are many objections to be made in addition to the ones observed above. First, "predictive policing" often unintentionally creates positive feedback loops, which can be demonstrated in the following example: if a defendant is predicted to recidivate, and gets denied parole, they may be more inclined to act in spite and recidivate. In this situation, the algorithm was proven true, although the defendant may have not recidivated if they were not predicted to, denying them parole. This is obviously an easy cycle to fall into and should be a hesitation when considering to utilize COMPAS in the legal system. In addition, COMPAS seems to turn people and their cases into a singular data point; it takes away the humanity of the system. This directly violates Emmanuel Kant's Categorical Imperative. The categorical imperative in question states an action is justifiable only as far as it at all times treats moral agents as ends never merely as means. In this case, people are viewed merely as means to the algorithm. Then, according to the ethical framework of Deontology, the COMPAS algorithm should not be utilized. In addition, Justice as Fairness, a phrase upheld by John Rawls, is violated by the utilization of this algorithm. According to Rawls' difference principle, where differences exist, resources should be allocated to protect the most vulnerable. Then, if there is the potential for arbitrary discrimination, as there is in COMPAS, Rawls would argue not to use the algorithm; Justice as Fairness is not upheld in COMPAS and should therefore not play even a supplementary role in the justice system.

Although there are many objections to using the COMPAS algorithm, described above, Wisconsin's supreme court upheld the algorithm for use in parole decision-making. The ruling stated that the algorithm must include additional warnings that are expected to be interpreted by a judge. However, many of those warnings may be hard to interpret, especially when presented to those without a statistical background. This renders them void in the context of decision-making, and the realistic fact is that the algorithm could easily be misused or misinterpreted. This makes it even more dangerous as a tool to be used as supplementary decision-making.

Therefore, to ensure equality in deciding cases without arbitrary discrimination, the COMPAS algorithm should not be utilized as a supplement to parole decision-making. It has many accountability, ethical, and statistical issues that would inhibit proper understanding and interpretation in court. Utilizing the algorithm could easily lead to arbitrary discrimination in granting or denying parole, beyond implicit biases that already exist.