HW 5

Swagat Adhikary

11/8/2024

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 [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 philosohpical evidence cited.

The use of the COMPAS algorithm in parole decisions has generated significant controversy due to its fairness concerns. While algorithmic tools like COMPAS have the potential to provide consistency and structure in the justice system, their application is only justifiable if they meet certain fairness standards, both statistically and philosophically. Upon evaluating COMPAS, it becomes clear that its current form does not satisfy these standards and, as a result, should not be used to influence parole decisions. A tool that disproportionately impacts certain groups cannot ethically serve as a supplement to human judgment, especially in matters as serious as parole. To ensure fairness, we must critically evaluate both the algorithm's statistical outcomes and its moral implications.

From a statistical perspective, COMPAS fails to meet critical fairness metrics such as equalized odds and independence. Equalized odds require that the true positive rates and false positive rates be equal across all protected groups, such as racial categories. In this context, a true positive would indicate a correct prediction that an individual is high-risk and will reoffend, while a false positive indicates that someone who is unlikely to reoffend has been misclassified as high-risk. Studies have shown that COMPAS produces significantly higher false positive rates for Black defendants compared to White defendants. This means that Black individuals are more frequently labeled as high-risk even when they are not likely to reoffend. On the other hand, White defendants experience higher false negative rates, meaning individuals who are more likely to reoffend are wrongly classified as low-risk. This discrepancy creates an unfair advantage for White defendants and an unjust burden for Black defendants, clearly violating the equalized odds fairness criterion. These racial disparities in prediction error rates not only reflect a technical flaw but also exacerbate existing systemic inequalities in the criminal justice system.

The failure of COMPAS to meet multiple fairness criteria is particularly concerning because, while it is theoretically impossible for a classifier to satisfy all fairness criteria at once when base rates differ across groups, COMPAS does not even satisfy two out of the three main fairness criteria. A key result in fairness theory is that when base rates (such as the actual recidivism rates) differ across groups, it is generally impossible to simultaneously satisfy equalized odds, sufficiency, and independence. This is true of any imperfect classifier. However, the fact that COMPAS not only fails to meet equalized odds but also fails to satisfy independence compounds the issue. Independence, also known as demographic or statistical parity, requires that the likelihood of receiving a positive prediction (being classified as high-risk) is the same across all racial groups, regardless of their actual likelihood of reoffending. COMPAS violates this criterion because

[1] https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis

[2] knit to a pdf to ensure page count

Black defendants are classified as high-risk at disproportionately higher rates than White defendants, even when controlling for similar backgrounds and histories. This failure of independence indicates that COMPAS's predictions are not race-neutral, meaning that race or race-correlated factors (such as socioeconomic status or criminal history) indirectly influence the algorithm's risk assessments. Failing two out of the three fairness criteria is significant because it suggests that COMPAS is not just imperfect; it is systematically unreliable across multiple dimensions of fairness. This failure undermines confidence in its ability to provide fair and just outcomes in parole decisions, especially given the high stakes involved.

Philosophically, the use of COMPAS raises significant ethical concerns, particularly when evaluated through the lens of Rawls's Veil of Ignorance and Kant's Categorical Imperative. Rawls's Veil of Ignorance is a thought experiment that requires individuals to make decisions as though they do not know their own position in society, including their race, gender, or socioeconomic status. The purpose of this framework is to ensure that social and political structures are designed to be fair to all, regardless of individual differences. Applying this to the use of COMPAS, no rational person operating behind the Veil of Ignorance would accept an algorithm that systematically disadvantages certain racial groups—such as Black defendants—by disproportionately labeling them as high-risk. This unequal treatment would not be considered just, as no one would risk placing themselves in a position where they could be unfairly penalized due to factors outside their control, such as race. The algorithm's racial bias directly contradicts the principles of fairness advocated by Rawls, making its use in the parole system ethically indefensible.

Similarly, Kant's Categorical Imperative, which is centered around the idea that moral principles must be universally applicable, also provides strong ethical grounds against the use of COMPAS. According to Kant, any action or rule must be capable of being applied universally, without exception, and must respect the inherent dignity of all individuals. Using COMPAS despite its racial biases would violate this principle because it would accept that some individuals are treated unfairly based on their race, a factor beyond their control. This treatment cannot be justified as a universal moral rule, as it fails to respect the equal worth and autonomy of all individuals. Kantian ethics demands that each person be treated as an end in themselves, rather than as a means to an end, which COMPAS fails to do by subjecting certain groups to unfairly harsh outcomes. Therefore, from a Kantian perspective, the use of COMPAS in parole decisions is morally unacceptable.

In conclusion, while COMPAS may offer the promise of consistency and efficiency in the justice system, its statistical flaws and ethical shortcomings make it unsuitable for use in its current form. The algorithm's failure to meet key fairness criteria—such as equalized odds and independence—demonstrates that it cannot reliably provide equitable assessments across different racial groups. Additionally, the philosophical principles of Rawls and Kant further underscore the moral problems associated with using an algorithm that perpetuates racial disparities. Until significant reforms are made to address these biases and ensure that COMPAS can deliver fair and consistent outcomes for all individuals, it should not be used in the parole decision-making process.