STOR 390: HOMEWORK 5

Riley Harper

March 26, 2024

Abstract: 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 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 (knit to a pdf to ensure page count) 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 graded both on the merits of its persuasive appeal but also the applicability of the statistical and philosophical evidence cited.

In the landscape of judicial decision-making, the integration of algorithmic tools such as the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system presents both groundbreaking potential and significant ethical quandaries. As a statistical consultant, my advice to a presiding judge contemplating the inclusion of COMPAS scores in parole decisions leans towards a cautious approach. This recommendation is not an outright dismissal of an algorithm's utility but a nuanced position advocating for a balanced integration of COMPAS outputs within a broader, multifaceted decision-making framework. Statistically, the COMPAS system has been scrutinized for its predictive validity and potential biases. Originally hailed for it's similar accuracy across racial groups, studies, such as the one conducted by ProPublica, have illuminated concerns over differential false positive (FPR) and true negative (TNR) rates among racial groups, suggesting that the algorithm may disproportionately label minority offenders as high risk for recidivism [Larson et al., 2016]. From a statistical perspective, such disparities raise critical questions about the fairness and reliability of using COMPAS scores as a significant factor in parole decisions. While no predictive model is devoid of error, the ethical implications of systematic biases necessitate a cautious application, especially in contexts as impactful on an individuals future as the criminal justice system.

Philosophically, the debate touches on the core principles of justice, equity, and the rehabilitation ethos of the parole system. Relying heavily on an algorithmic assessment can inadvertently diminish the individualized consideration of a parolee's circumstances, potentially overlooking the multifaceted nature of human behavior and social reintegration prospects by reducing an individual to a mere number. A fundamental philosophical question arises from this, can fairness be quantified, and if so, at what cost? To this I would argue that, whether we like it or not, we, as individuals, label other individuals with these scores of risk every day. Even subconsciously, all individuals have an element of prejudice in their decision-making. This prejudice, developed by a combination of lived experiences and influenced by others, allows us to make decisions in scenarios where right and wrong isn't as clear cut as it often is in mathematical fields. Since it is already being done by us as individuals, one might question if having machines develop the prejudice is any worse. You may have your own opinion on this, as do I which is yes. The reason for my agreeance is that humans are known to be biased and prejudiced, a fact likely not news to you or any other reader. In fact, studies have shown that humans are often less trusting of peers compared to machines, leading to an over reliance on their decisions, even when they're based on immutable characteristics [Ahmad et al., 2023, Schaich Borg et al., 2023]. Allowing these prejudiced decisions to play out over a longer period can lead to a self-fulfilling prophecy where underserved populations are situated in heavily policed areas with a quick-to-act police force that enforces the law first and asks questions second. The algorithm may ultimately be correct, but it was the decision to heavily police and harshly enforce laws in these areas that led to the algorithm's predictions being correct. So, what if humans ignore the machine when it is shown to be making prejudiced decisions? As mentioned, we are also prejudiced. In ”Moral AI: And How we Get There,” by Schaich Borg, Sinnott-Armstrong, and Conitzer, evidence was found that in Kentucky, the introduction of algorithmic predictions of bail violations was associated with judges offering no-bail release to White defendants more often than to Black defendants. This was found to have occurred because judges were more likely to overrule the algorithm’s predictions for moderate-risk defendants if the defendants were Black [Schaich Borg et al., 2023]. Viewing this from a consequentialist perspective, it’s clear that the consequences of these decisions will leave underserved populations in a cycle of battling the officials and law enforcement who are supposed to ”serve and protect.” I argue for either fully allowing artificial intelligence to run a system while humans and our tools audit their performance or vice versa, rather than sharing the task. When working in tandem like this, we offer a checks and balance system to one another that reduces the echo chamber of prejudice seen in the original implementation of COMPAS.

Before implementing COMPAS or any criminal justice algorithm, several prerequisites must be met. Firstly, the accuracy of FPR and TNR between different groups, such as race or gender, or proxy variables that combine several immutable characteristics, must be considered. Using a veil of ignorance, one would oppose an algorithm like COMPAS, which is more likely to predict a false positive and less likely to predict a true negative for recidivism among minority individuals, especially if placed into a disadvantaged group in society. Ideally, an algorithm would perform similarly across these groups in terms of FPR and TNR. However, an algorithm that performs equally poorly across groups is not a solution. Flipping a coin, which would be correct only 50% of the time, is only slightly less accurate than the COMPAS algorithm but without bias. The reliance on an algorithm that cannot decisively outperform a simplistic method such as coin flipping, which boasts a 50% accuracy rate, poses a profound ethical dilemma. The use of such an algorithm, under the guise of efficiency or objectivity, could inadvertently institutionalize bias within judicial processes, perpetuating cycles of inequality. Further, it could mask the subjective nature of decision-making under a veneer of technological impartiality. In light of these considerations, it is crucial that any algorithm deployed in such high-stakes contexts undergoes rigorous, ongoing evaluation not only of its predictive accuracy but also of its fairness and impact on all segments of society. Furthermore, the discussion extends beyond mere statistical validation to encompass the broader implications of algorithmic governance in justice. The integration of such technologies into the judicial system demands transparency, accountability, and a commitment to rectifying identified biases. Stakeholders, particularly those from marginalized communities, must have a voice in how these tools are developed, deployed, and audited. The ultimate goal should not be to replace human judgment with algorithmic determinations but to enhance the former with the latter’s insights, ensuring that technology serves as a tool for justice enhancement rather than an instrument of inequity. In envisioning a future where algorithms like COMPAS are part of the criminal justice landscape, it is essential to establish robust frameworks that include independent oversight, public reporting of algorithmic performance, and mechanisms for addressing disparities when they are identified. By adopting a holistic approach that considers the statistical, ethical, and societal dimensions of algorithmic implementation, the justice system can leverage technology to improve decision-making processes while steadfastly upholding the principles of fairness and equity.

Overall, while COMPAS and similar algorithms used to predict recidivism have room for improvement, they can provide valuable insights into risk factors associated with recidivism when used judiciously. This should be complemented by a robust framework that includes qualitative assessments, personal circumstances, and rehabilitative efforts. The goal is to employ a holistic decision-making process that respects the complexity of human behavior while striving for a just and equitable penal system that does not dehumanize the individual under accusation. After all, we are a country founded on the principle of innocence until proven guilty. The development and advancement of technology should enhance, not undermine, this principle for everyone, regardless of their skin color or gender.

References in the Discussion

[Ahmad et al., 2023] Ahmad, S. F., Han, H., Alam, M. M., et al. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, 10:311.

[Larson et al., 2016] Larson, J., Mattu, S., Kirchner, L., and Angwin, J. (2016). How we analyzed the compas recidivism algorithm. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

[Schaich Borg et al., 2023] Schaich Borg, J., Sinnott-Armstrong, W., and Conitzer, V. (2023). *Moral AI: And How We Get There*. Penguin Books.