Homework 5

IZ Raad

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 [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.

As a statistical consultant, I would advise judges against utilizing the COMPAS algorithm when deciding if a defendant should be granted parole. In short, this is an unfair algorithm that is biased against non white defendants, and thus actively contributes to injustices that already exist within our court system.

Before exploring the consequences of a biased algorithm, I must first prove that the algorithm is indeed unfair. At first glance, COMPAS appears to be quite the opposite, as its accuracy among black and white defendants is roughly the same (63 and 59 percent, respectively). However, scrutinizing the composition of the cases that COMPAS correctly and incorrectly classifies reveals discrepancies in the algorithm's decision-making. Indeed, black defendants were twice as likely to be misclassified with a high risk of recidivism than their white counterparts. Along similar lines, white defendants where misclassified as being low risk about twice as often as their black peers. With that, we claim that COMPAS fails to meet the predictive equality metric of fairness. We see that while COMPAS makes mistakes in both racial categories, its mistakes are often in favor of white defendants and penalize black defendants. The fact that black defendants are more likely to bear the burden of the algorithm's mistakes lead me to label the algorithm as unjust. Moreover, the black box nature of this algorithm makes it hard to understand proxies for race that are being used to classify defendants. That is, because we do not understand COMPAS's decision making process, we cannot account for variables that may be used as stand-ins for race, such as zip code.

So we've decided that COMPAS is unfair. What are the moral implications of this? This algorithm is trained on a system that already weighted against black individuals. The racism in America's court systems is well known. Using an algorithm to disproportionally dole out detrimental decisions to an already disadvantaged group will not allow us to make any progress. In order to make strides in the direction of legal fairness, we must do all we can to mitigate chance of inequity. This includes choosing to not use the COMPAS algorithm. Additionally, using a biased algorithm may perpetuate a positive feedback loop in our justice system. Defendants denied parole may become discouraged and feel failed by the court system. This may make them less privy to rehabilitation, making our COMPAS algorithm a self-fulfilling prophecy that damages groups groups prone to false misclassification as high risk. Increased denial of parole to black defendants also has the potential to reinforce societal biases and exacerbate differential treatment.

COMPAS's moral implications stretch beyond considerations of its fairness. Using an algorithm like COMPAS removes humanity from the decision making process. Treating people as data points and feeding

---

them through an algorithm will never be able to truly capture situational complexities on a case-by-case basis. Defendants are robbed of the thoughtful consideration that their cases deserve. Indeed, this very treating of humans as data points treats them as a means to an end– something that deontologists like Kant have cautioned against. Each individual is not merely data to feed through an algorithm to efficiently gain a verdict. Instead, they are people whose cases deserve to be treated with respect.

Utilitarians may counter my decision, claiming that an algorithm like COMPAS makes the decision process more efficient, thus allowing more defendants to receive parole decisions. However, consequentialist logic can can also be applied against COMPAS. The algorithm correctly predicts recidivism about 61 percent of a time. This is marginally better than a coin flip. However, COMPAS's marginal benefit is drastically outweighed by all its moral implications, outlined in previous paragraphs. Using cost-benefit like logic, COMPAS produces more net harm than good, and thus should not be used. We may also consider COMPAS from a virtue ethics standpoint. This moral framework guides people to act in alignment of certain virtues like fairness and justice. An algorithm disproportionately impacting one group is not in accordance with the virtue of fairness, so it should not be involved in the process of granting parole.