HW 5

Samuel Reid

12/29/2023

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 philosophical evidence cited.

STUDENT RESPONSE

As crime continues to plague the nation, many states have looked for ways to improve outcomes both for the victim and the perpetrator, and one element of that is assessing the perpetrator's potential to reoffend. Simultaneously, the rise of machine learning has allowed far more advanced models than could be reasonably used before to enter the courtroom, as in all other spheres. While the COMPAS algorithm was widely used to accomplish this purpose, it has unconquerable flaws, both taken individually and as an element of a wider algorithmic movement, and should not be used under any circumstances.

First, we will examine statistical problems borne out by this specific algorithm. The most commonly touted problem is that of disparate impact based on race. While overall accuracy is similar across races, white defendants are, by large, classified as less risky than they actual are, and black defendants are classified as more risky (with misclassification rates of 23% vs 45% and 48% vs 28% respectively). Aside from being racist, this problem is also self-fulfilling, as prolonged stays in prison may be conducive to increased criminal behavior, due to associating with others criminals and institutionalization that leads to learned helplessness in the outside world. This problem may be inherent in the structure of the classifier itself. If you are attempting to maximize accuracy, it may be slight demographic trends in the population are magnified and interpolated onto every case of that demographic in the testing data. Propublica also found that the prediction of risk scores is far more uniform for black defendants than the right-skewed distribution for whites, which could indicate a small training sample size, or a lack of high-quality/representative data that would be able to capture the full spectrum of experiences. Whatever the cause, near-uniform distributions are themselves a reason to question the efficacy of a model, as they by definition give us little information about the output of an input. While I wasn't able to find a conclusive answer for this, it appears (from looking at Wikipedia) the algorithm itself follows a regression-type model, in which elements are chosen by the researchers to be included in the model, and the output is a very direct function of those factors. It may produce better outcomes in terms of accuracy and fairness if the data was left unstructured, and a neural-network or similar tool with hidden elements was used to extract more complex interaction relationships from the data, rather than just finding variable coefficients. Most damning of all, we can look toward the work Dressel and Farid [3], who found that COMPAS is not substantially better (and is often worse) at predicting outcomes than even an untrained person when asked to provide a personal judgement for a given

[1] https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis
[2] knit to a pdf to ensure page count
[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5777393/

defendant. This should be enough to discard the algorithm immediately, all the drawbacks presented are only justifiable if the increase in accuracy is significant. They also found that COMPAS was less accurate than a logistic regression using only two features, indicating that the algorithm is seriously over-fitted, and includes extraneous variables that will only muddy the classification process on test data. While all of these model limitations could theoretically be explained to a judge with no statistical background, it would be time-consuming, likely insufficient, and obscures that the main problem is the overall small net benefit and large amount of drawbacks. This utilitarian consideration leads into a philosophical discussion.

Usage of the COMPAS algorithm is not permissible under any of the philosophical frameworks we have discussed. First, from a consequentialist perspective, the most immediate consequences are obvious, but no less important. If we keep those who have reformed their ways in prison, as we are depriving them of basic rights and freedoms unnecessarily. If we let out offenders, we are depriving the general populace and specific victims of the right to not be terrorized. All the statistical issues discussed are themselves negative consequences of the algorithms use. These consequences exist within the existing system, and ultimately the purpose of COMPAS is to address these issues specifically, however it introduces problems at the same time. First, who is paying for this software? Northpointe isn't giving the software away. As it is being used for a public state-run service, it is most likely being pushed onto the average taxpayer. Though that consequence might be small for an average individual, the fact that it could be prevented by simply not buying/subscribing to a unnecessary product makes it abhorrent. Additionally, the consequence of reducing accountability in judges and parole boards is that they will become more detached from the process, and less willing to hear out a given defendant in any case when they could just ignore the emotionally effecting process and leave the decision up to an algorithm, such that any amount of algorithmic involvement will ultimately dominate the process.

The implications raised by this accountability issue are more applicable when looking through a deontological lens. On the surface, the intention of using COMPAS is benevolent. Protecting society from violent repeat offenders is an especially popular talking point for both parties in the current climate, and harnessing the power of statistics to overcome human judgement errors and biases makes a lot of sense. However, one may ask whether the action is actually conducive to human betterment, or instead taking the easy way out. Deontology is often defined as a study of duty. It is the judge and parole board's duty to accurately assess a defendant and make a human judgement as to their risk of offending. If COMPAS is weighted heavily in the decisions they make, they are derilecting their duty as human actors and reducing their moral responsibility. Taking on this responsibility is part of the job. A lot of people would have problems if neurosurgeons were paid the seven figures they are now, yet did all the work of pushing a button to start up a surgery-bot, no matter how good the machines were.

Virtue ethics is less enlightening, outside of the previous examples in which protecting human rights, public safety, and taking accountability are all virtues worth instantiating. On an a more abstract level, I posit this: knowing that your continued stay in prison is the result of an unchangeable algorithm, rather than a person you might have the opportunity to connect with and win over, is destructive to the human spirit and blunts the emotions of both parties involved to an unacceptable extent. To some degree this is intended, in order to remove bias from the process. However I question the idea that, in the age of increased social detachment in all facets of life, instrumentalizing yet another venue for potent social communication for little to no benefit is worthwhile.

In conclusion, use of the COMPAS algorithm or any similar algorithm is unjustifiable both statisically and morally, and the courtroom should be kept a place in which the human elements are used to influence the process of the law, rather than the other way around.