

HW 5

Sydney Mason

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 philosophical evidence cited.

The Correctional Offender Management Profiling for Alternative Sanctions or COMPAS algorithm for recidivism, or a formerly incarcerated person's likelihood of violating their parole, has repeatedly come under fire for demonstrating racial bias in predicting which offenders would violate their parole. Although many have condemned it, some, including notably the Wisconsin Supreme Court, have said that it does not violate due process. The algorithm was developed in order to supplement a judge's decision by providing a risk score for general risk of recidivism and risk of future violent behavior on a scale of 1-10, 1 being the lowest and 10 being the highest. These risk scores were designed to give the judge an indicator, based on the 137 questions that an offender had answered, what their likelihood of future crime was. The catch with COMPAS? The specific calculations that are used to transform those 137 answers into risk scores are not available to the judge, offender, or their legal team, making COMPAS a black box algorithm. There is no way to determine how it comes to its decision.

In my opinion, the use of COMPAS is not justifiable as a tool to supplement the judgement in decisions about recidivism. Ethically and statistically, it doesn't stand up to intense scrutiny despite the decisions like the Wisconsin Supreme Court's.

Firstly, the overall accuracy of COMPAS as determined in a 2009 study was only 68%. Although data isn't readily available regarding the accuracy of judges when making similar decisions, considering the impact that COMPAS has, a less than 3 in 4 chance of successfully predicting if someone will violate their parole is too large of an accuracy gap to ignore. It is difficult in any scenario to create a threshold at which something can be determined to be accurate enough, but I believe that COMPAS should be approaching anywhere between 75 and 80% accuracy to be used, even if it is only supplementing the decisions of judges. The study also noted that there was a small gap in accurately predicting the likelihood of recidivism between black men and white men, 67% and 69% respectively. Therefore, the already low accuracy is less accurate for black men than white men, demonstrating a difference in how the algorithm makes predictions based on race.

Something important to note is that COMPAS does not specifically have data about the racial background of offenders, but the dataset it is operating off of is rich enough that other variables included in the dataset can be used as proxies. One such proxy is the zipcode of the original arrest for the offender, which due to redlining and other underlying variables, was often indicative of the racial background. So although the questions COMPAS operates off of do not specifically imply race, there is enough information that COMPAS is able to extrapolate.

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

²knit to a pdf to ensure page count

COMPAS violates 2 out of 3 of the statistical measures of fairness, the two being independence and separation. Independence, with a sub metric of disparate impact, examines if one group is greater impacted than another with no acknowledgement of an underlying ground truth. Separation, with a sub metric of equalized odds, examines the rates of false positives and false negatives for the marginalized and non-marginalized groups and confirms that they occur at similar rates. In our own analysis based on available COMPAS data, false positives occurred at a higher rate for black offenders and COMPAS was more likely overall to find them likely to violate parole. In terms of the third, sufficiency, COMPAS does technically satisfy it. Northpointe has argued that since it is only possible for all three of the fairness metrics to be met when there is a perfectly predicting classifier and the underlying ground truth label is split equally amongst the protected variable that the fact that it meets the sufficiency metric is enough to say that it is a fair algorithm. However, since it specifically violated equalized odds which demonstrates that black offenders were more likely to be accused of potential recidivism than white offenders.

The statistical evidence as to why COMPAS shouldn't be used is overwhelming, and goes beyond what I've shared here. But morally and ethically there are issues with it as well. In considering philosophical definitions of fairness, specifically fairness as equality, COMPAS goes against the principle. Fairness as equality states that each person gets the same quantity and quality of good. With the bias that COMPAS demonstrates, the quality of the judgement that a black offender gets is lower than that of a white offender, and is in fact biased towards white offenders getting parole easier. Statistically, and ethically, COMPAS is not a fair algorithm to use.

When considering moral frameworks such as consequentialism, which asks the person to consider primarily the consequences behind a decision, the consequences of using an inaccurate and biased algorithm and making the wrong decision about someone's likelihood of recidivism is far greater than the consequences of not using it. Judges have been making decisions for years without COMPAS, and should continue to do so as COMPAS cannot be justified statistically, so morally and ethically, the consequences of using it are not worth the risk.

Angwin, Julia, et al. "Machine Bias." ProPublica, 23 May 2016, www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.