HW 5

Mackie Jackson

11/08/2024

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.

I believe that judges should exercise the wisdom earned from years of legal experience and forego the use of the COMPAS algorithm's assistance. In a previous court case (homework assignment), I argued that utilitarian thought proposes that the correct action is the one that maximizes pleasure, and so the use of COMPAS is correct if it does so as well. I also proposed that in order to determine if a judge's use of COMPAS is morally justifiable, we might simply look at the ratio of correct decisions made with and without algorithmic supplementation. This would measure if the algorithm is maximizing societal pleasure by preventing recidivism more than a judge would unassisted. While there is no empirical evidence of judges being more or less accurate when utilizing COMPAS, there is an existing study that compares the accuracy of the algorithmic predictions to those made by a test group of individuals with no legal expertise. COMPAS and survey respondents had the same accuracy rate of around 65 percent (Dressel & Farid 2018). This leads us to question the utilitarian value of implementing a complex algorithm whose predictive intricacies are black boxed from public view if a human with no criminal justice knowledge is just as accurate.

We statisticians know that any real classifier is unable to adhere to more than one of the three fairness metrics, especially because Black defendants have an overall higher recidivism

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

[2] knit to a pdf to ensure page count

rate in this case. However, the violation of two other metrics is significant when we apply an algorithm to diverse and potentially vulnerable groups. In COMPAS's defense, Northpointe has argued that it partially satisfies predictive parity because the risk of recidivism is the same across protected and non-protected classes (in high risk cases). I point out a significant equalized odds violation: Black defendants were falsely predicted to reoffend at a rate of 44.9 percent, approximately two times higher than white defendants at 23.5 percent. White defendants who reoffended were falsely predicted to not do so 47.7 percent of the time, approximately twice as high as their Black counterparts at 28 percent (Dressel & Farid 2018).

Philosopher John Rawls proposes that we should make decisions that are blind to features about an individual such as race or class, instead opting to determine what the best outcome for any person might be. Even though COMPAS doesn't incorporate race as a variable, we have seen that proxy variables like class and location drive disparate predicitons based on race. As a judge, you are a person whose decisions shape the very lives of individuals in the criminal justice system. Rather than opting to use an algorithm whose accuracy is similar to someone with no legal experience, and that disparately falsely predicts recidivism for Black defendants and non-recidivism for white defendants, I suggest that you rely on the evidence presented to you and your own abilities.

References

Dressel, Julia, and Hany Farid. 2018. "The Accuracy, Fairness, and Limits of Predicting Recidivism." Science Advances 4 (1): eaao5580. https://doi.org/10.1126/sciadv.aao5580.