HW 5

Alyson Longworth

3/27/2024

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.*

The results of the COMPAS algorithm should not be included in the decision making process for granting parole in order to prevent any potentially biased influence to judges whatsoever. For instance, not only does this algorithm typically predict with low accuracy rates, but the inaccuracies additionally tend to skew against black defendants. In the article "How We Analyzed the COMPAS Recidivism Algorithm," Larson, Mattu, Kirchner, and Angwin discuss how the Cox model was used to compare actual events to the predictions made by the COMPAS algorithm and it was discovered that this algorithm accurately scored for risk of recidivism only 63.6 percent of the time. Not only was this lower than that of which Northpointe reported, it was additionally lower than the typical 'rule of thumb' of decent predictive accuracy, which is noted at 70 percent or above. When testing the odds of receiving a higher score from a logistic regression model, race was additionally found to be a significant predictor of higher scores, with black defendants being 45 percent more likely to get a higher score for risk of recidivism and 77.3 percent more likely to get a higher score for risk of violent recidivism. Nonetheless, I argue that although both the scores for the risk of recidivism and the risk of violent recidivism for black and white defendants were proven to be both skewed, there is a stronger argument for racial bias in the imbalances observed in the false negative and false positive rates between these two groups.

Although the rates to which this algorithm misclassified black defendants and white defendants in general were found to be roughly similar, the COMPAS algorithm misclassified black defendants as higher risk of both recidivism and violent recidivism nearly twice as often as white defendants, but correctly classified them to recidivate only slightly more than white defendants. This was additionally observed when testing the opposite and it was discovered that it was 63 percent more likely to misclassify white defendants with low risk of recidivism than black defendants. To my understanding, it is a weak argument to claim that based on the prediction imbalances between black and white defendants alone, the COMPAS algorithm is biased. However, because repeatedly higher false negative rates were observed toward black defendants, and false positive rates were repeatedly skewed toward white defendants, it can be proven that the COMPAS algorithm demonstrates some sort of underlying racial bias against black defendants. Not only does this algorithm have relatively low accuracy rates to begin with, but the inaccuracies of this algorithm appear to be imbalanced and biased against black defendants. With an accuracy rate of only 63.6 percent, it can be argued that this algorithm may not be very much more accurate than simply flipping a coin. With that being said, it can even be argued that flipping a coin for these decisions would result in not only potentially similar

[1] https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis
[2] knit to a pdf to ensure page count

accuracy rates, but even more balanced false negative and false positive rates between white defendants and black defendants.

Viewed from a deontological and virtue ethics perspective, the use of this algorithm's biased influence on parole decisions is immoral. From a virtue ethics perspective, influencing future opportunities of a defendant by inaccurate and biased data is not only dishonest, but results in a lack of respect and empathy toward the defendants themselves. Continuous reliance on this algorithm has the potential to result in a situation where judges depend more and more on the accuracy of the algorithm. Furthermore, from a deontological perspective and according to the second categorical imperative, treating humans as a mere means to an end is unethical. By categorizing defendants and depending solely on the algorithm's output rather than treating a defendant as an individual and unique person, a judge will dehumanize them and reduce them to mere data points. Consequently, defendants will have less opportunity to present their unique stories and backgrounds. This risks prioritizing efficiency over morality, at the expense of a defendant's potential future, which proves the immorality of the COMPAS algorithm's influence on parole decisions.

References:

Larson, Jeff, et al. "How We Analyzed the Compas Recidivism Algorithm." ProPublica, 23 May 2016, www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.