HW 5

Maria Rubio

11/08/2023

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 philosohpical evidence cited.

Dear Judge, I am writing to you in order to determine whether it is the best decision to use the COMPAS algorithm as a decision-making tool. Despite its growing popularity, I can state with absolute certainty that the use of this algorithm in the judicial field is not appropriate due to its various statistical and ethical limitations. Its lack of accuracy, coupled with issues of transparency and racial bias, raise concerns that cannot be ignored. Allowing COMPAS to influence your decisions could compromise the principles of fairness and justice that should govern its decisions.

In order to justify my refusal to use this algorithm in the legal field, it is necessary to take into account several problems.

First, one of the main problems lies in the lack of transparency and audibility of the model. COMPAS is a "black box model", which means that its internal calculations and decision criteria are not accessible for review or audit. This limits control over its accuracy and creates uncertainty about the origin of possible biases in its results. In keeping with the concepts of "risk" and "loss" in the notes, COMPAS not only misassesses the risk of certain individuals, but by operating without transparency, it does not allow the individual who has been sentenced to appeal or recourse.

The empirical risk function that could guide a fair model would be:

$$L(\theta, \hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \ell(\theta_i, \hat{\theta}_i),$$

In COMPAS this function is minimized without adequately considering specific subgroups. By applying this quadratic loss minimization, the model is optimized for overall accuracy, but does not guarantee a fair distribution of errors across different racial groups. It is not entirely adequate since fairness is required rather than simply statistical accuracy, and therefore becomes problematic in court decision contexts.

Not only is the statistical perspective important, but using the COMPAS algorithm for sentencing purposes can be controversial from an ethical standpoint as well.

From a deontological perspective, the use of COMPAS in parole decisions is questionable because it violates the Kantian principle of treating individuals as ends in themselves and not as means to an end. Kant

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

[2] knit to a pdf to ensure page count

argues that an action is moral if it can be universalized without contradiction and if it treats each person with dignity, without instrumentalizing them. Using an algorithm that discriminates against certain groups turns these individuals into means to a supposed end of "security," ignoring their value as persons. This, in turn, undermines the judicial system's commitment to equality and impartiality, as it prioritizes a statistical prediction over the fair treatment of each person.

On the other hand, from a utilitarian approach, one could argue that COMPAS maximizes social welfare by attempting to reduce recidivism, which would benefit society as a whole. However, the social and psychological harms that COMPAS inflicts on specific communities outweigh its potential benefits. Utilitarianism, which promotes maximizing well-being and minimizing suffering, would not justify an algorithm that generates disproportionate harm to certain populations if this does not result in a clear and justified benefit to society. The inequalities amplified by COMPAS generate resentment and distrust in the system, eroding the sense of justice that the judicial system must preserve.

Finally, from virtue ethics, justice is an essential virtue that should guide every judicial decision. If the use of COMPAS amplifies prejudices or discriminates against certain individuals, its implementation in the judicial system would be inconsistent with this virtue of justice. Virtue ethics implies that judges should act according to fair and balanced values, evaluating each case without relying on a tool that reinforces inequalities.

In conclusion, due to the demonstrated problems of accuracy, transparency, and ethics, I recommend that COMPAS not be used in parole decisions. Reliance on a tool that perpetuates racial bias and whose lack of transparency prevents review and audit of its decisions poses a risk to the fundamental principles of justice and fairness in the judicial system. Instead of relying on COMPAS, judges should base their decisions on individualized assessments and clear ethical principles that respect the dignity of each person and promote a truly fair and impartial system.