HW 5

Seth Hall

11/07/2023

Contents

The criminal justice system is supposed to be about fairness, equality, and giving everyone a fair shot. But as technology has become a bigger part of our lives, it's also started to creep into places like the courtroom, where judges are using algorithms to help make decisions about things like parole and sentencing. One of the most well-known of these tools is called COMPAS, which stands for Correctional Offender Management Profiling for Alternative Sanctions. It's designed to predict the likelihood that someone who's been arrested will commit another crime. In theory, this sounds like a good idea because it could help reduce bias and make the system more efficient. But in reality, things aren't so simple. In recent years, there's been a lot of debate about whether algorithms like COMPAS are fair or whether they actually end up making the criminal justice system even more biased. A 2016 investigation by ProPublica brought a lot of these concerns into the spotlight, revealing that COMPAS seemed to be biased against African Americans. This has led to a big question of whether judges should use COMPAS scores when making decisions about parole. As a statistical consultant advising a judge, my advice would be that the COMPAS algorithm should not be used in parole decisions. The evidence shows that it's biased, it's not fully transparent, and it raises a lot of ethical issues. The idea behind COMPAS is to take a bunch of information about a person like their criminal history, age, employment status, and even answers to a personality questionnaire, and use that data to spit out a risk score. This score is supposed to tell judges how likely someone is to reoffend. A higher score means a higher chance of committing another crime, which could lead to a harsher sentence or denial of parole. The issue is that even though COMPAS is supposed to be an objective, data-driven tool, it doesn't always work fairly. In fact, studies have shown that COMPAS is often biased against African American defendants. This means that Black defendants are more likely to be labeled as "high risk" than White defendants, even when they're not any more likely to reoffend. This kind of bias isn't just unfair—it can have serious consequences for people's lives. There are several ways to measure fairness. These metrics are supposed to help us figure out if the tool is treating everyone equally, regardless of their race, gender, or other characteristics. Let's take a closer look at some of these key metrics: statistical parity, equalized odds, predictive parity, and disparate impact. The first measure of fairness is called statistical parity, which is a fancy way of saying that people from different groups should have the same chance of being labeled as high risk. In mathematical terms, it looks like this: $P(\text{Prediction} = \text{High Risk} \mid \text{Race} = A) = P(\text{Prediction} = \text{High Risk} \mid \text{Race} = B)$ This formula means that African American and white defendants should have an equal likelihood of being classified as "high risk" by COMPAS. But according to the ProPublica study, COMPAS fails this test. African Americans were labeled as high risk almost twice as often as White defendants, even though they didn't actually reoffend at a higher rate. This clearly shows that COMPAS is not achieving statistical parity, which is a big red flag if we care about fairness. The consequences of failing to meet statistical parity are pretty serious. If African Americans are more likely to be labeled as high risk, they're more likely to face harsher punishments, like being denied parole or getting a longer sentence. This can have a ripple effect, making it harder for them to reintegrate into society, find jobs, and support their families. It's a cycle that keeps communities of color trapped in the criminal justice system. Another important fairness measure is called equalized odds. This metric focuses on whether an algorithm is equally accurate for different groups. In other words, it's about making sure that the algorithm is just as good at predicting who will reoffend and who won't, no matter their race. The formulas for equalized odds are: $|P(\hat{Y} = 1 \mid (S = 1 \quad Y = 0)) - P(\hat{Y} = 1 \mid (S \neq 1$

Y = 0))| > or |P (ˆY = 1|(S = 1 Y = 1)) − P (ˆY = 1|(S 6 = 1 Y = 1))| > . These formulas mean that COMPAS should have the same rates of true positives (correctly identifying reoffenders) and false positives (wrongly labeling non-reoffenders as high risk) for both African Americans and White defendants. But again, COMPAS falls short. ProPublica found that COMPAS has a higher false positive rate for African Americans, meaning they were more likely to be wrongly labeled as high risk, and a higher false negative rate for White defendants, meaning they were more likely to be wrongly labeled as low risk. This unequal treatment can lead to unjust outcomes where African Americans are unfairly kept in prison longer, while White defendants who are actually high risk get released. This not only impacts the individuals involved but also shakes the public's trust in the fairness of the justice system. Statistical parity is all about whether the algorithm's risk scores are equally reliable for everyone. It checks if a high-risk score is equally likely to be accurate for people of different races. The formula looks like this: |P (ˆY = 1|S 6 = 1) − P (ˆY = 1|S = 1)| > . While COMPAS has been found to meet this criterion to some extent, this doesn't outweigh its failures in other areas. For example, even if the risk predictions are somewhat accurate, the fact that COMPAS has higher false positive rates for African Americans makes it unreliable and unfair in practice. Disparate impact is when a policy or tool, even if it's not intended to be discriminatory, ends up having a negative effect on a particular group. This is often measured by comparing how different groups are affected by a decision or policy. The formula for this is given by P (ˆY = 1|S 6 = 1) / P (ˆY = 1|S = 1) < 1 − . According to ProPublica's findings, COMPAS has a clear disparate impact on African Americans, labeling them as high risk at much higher rates than White defendants who are similarly situated. This kind of bias is not just a technical problem—it has real-life consequences that can ruin lives and perpetuate racial inequality. The fairness impossibility theorem says that it's mathematically impossible for an algorithm to satisfy all fairness criteria at the same time if different groups have different base rates of reoffending. No matter how the algorithm is tweaked, it will always be unfair in some way if there are underlying differences between groups. This is a huge problem because it shows that relying on algorithms like COMPAS to make fair decisions is fundamentally flawed. If we can't make these tools fair for everyone, should we really be using them to make decisions that could change someone's life forever? Deontological ethics are the morals and principles that we adhere to. Somewhat similar to having a moral compass that guides you towards right and wrong irrespective of the results. One of the big names in this kind of ethics that has been discussed in STOR 390 is the philosopher Immanuel Kant, who thought everyone should be treated with respect and dignity. And if you look at COMPAS through the lens of deontology, you can see its shortcomings very clearly. This is an unfair situation, given that the algorithm does have this bias against African Americans. COMPAS violates the ethical principle of treating like cases alike, since it treats people differently based on their race. From this perspective, it is simply wrong for parole decisions to be made using COMPAS because it takes away the right of defendants to be considered impartially. Contrarily, Utilitarianism is focused on the most good for the most great number. It prioritizes consequences over process, so if using COMPAS could somehow decrease total crime, a utilitarian might support it. True, but the situation is a little more complex than that. But if COMPAS can indeed reduce crime, that comes at the price of discriminating against African Americans, which raises the question about whether those benefits might be worth it. Rule Utilitarianism tells us to adhere to those rules whose overall observance produces the greatest good. In all other cases, however, if the rule is to simply apply a biased algorithm with no consideration of social stratification or economic mobility, and so forth, then the results will not be better; they will reinforce existing inequalities. Moreover, even the studies haven't shown that COMPAS is really effective in reducing recidivism which removes any arguments for its utilitarian benefit. One of the biggest ethical problems with COMPAS is that it's not transparent. The company behind it, Northpointe (now known as Equivant), has refused to reveal the inner workings of the algorithm, claiming it's proprietary information. This means that neither the defendants nor their lawyers can challenge the risk scores or understand how they're being calculated. This lack of transparency violates the principle of due process, which is a fundamental right in the justice system. People have the right to know the evidence against them and to challenge it in court. If people are being denied parole based on secret algorithms, that's not justice—it's a violation of their rights. Given all these issues, I strongly recommend that judges do not use COMPAS scores when making decisions about parole. There are too many risks involved in relying on a biased, opaque algorithm to make decisions that could change someone's life. Instead, the justice system should focus on more transparent, human-centered approaches to risk assessment that take into account the complexities of each individual case. As technology continues to advance, it's crucial that the tools that are used to ensure that they are not only effective but

also just are critically examined. The use of algorithms like COMPAS in the criminal justice system raises serious questions about what kind of society is best. A society that prioritizes efficiency and data or one that upholds the principles of fairness and justice for all.