The New York Times

Even Imperfect Algorithms Can Improve the Criminal Justice System

A way to combat the capricious and biased nature of human decisions.

By Sam Corbett-Davies, Sharad Goel and Sandra González-Bailón

Dec. 20, 2017

In courtrooms across the country, judges turn to computer algorithms when deciding whether defendants awaiting trial must pay bail or can be released without payment. The increasing use of such algorithms has prompted warnings about the dangers of artificial intelligence. But research shows that algorithms are powerful tools for combating the capricious and biased nature of human decisions.

Bail decisions have traditionally been made by judges relying on intuition and personal preference, in a hasty process that often lasts just a few minutes. In New York City, the strictest judges are more than twice as likely to demand bail as the most lenient ones.

To combat such arbitrariness, judges in some cities now receive algorithmically generated scores that rate a defendant's risk of skipping trial or committing a violent crime if released. Judges are free to exercise discretion, but algorithms bring a measure of consistency and evenhandedness to the process.

The use of these algorithms often yields immediate and tangible benefits: Jail populations, for example, can decline without adversely affecting public safety.

In one recent experiment, agencies in Virginia were randomly selected to use an algorithm that rated both defendants' likelihood of skipping trial and their likelihood of being arrested if released. Nearly twice as many defendants were released, and there was no increase in pretrial crime.

New Jersey similarly reformed its bail system this year, adopting algorithmic tools that contributed to a 16 percent drop in its pretrial jail population, again with no increase in crime.

Algorithms have also proved useful in informing sentencing decisions. In an experiment in Philadelphia in 2008, an algorithm was used to identify probationers and parolees at low risk of future violence. The study found that officers could decrease their supervision of these low-risk individuals — and reduce the burdens imposed on them — without increasing rates of re-offense.

Studies like these illustrate how data and statistics can help overcome the limits of intuitive human judgments, which can suffer from inconsistency, implicit bias and even outright prejudice.

Algorithms, of course, are designed by humans, and some people fear that algorithms simply amplify the biases of those who develop them and the biases buried deep in the data on which they are built. The reality is more complicated. Poorly designed algorithms can indeed exacerbate historical inequalities, but well-designed algorithms can mitigate pernicious problems with unaided human decisions. Often the worries about algorithms are unfounded.

For example, when ProPublica examined computer-generated risk scores in Broward County, Fla., in 2016, it found that black defendants were substantially more likely than whites to be rated a high risk of committing a violent crime if released. Even among defendants who ultimately were not re-arrested, blacks were more likely than whites to be deemed risky. These results elicited a visceral sense of injustice and prompted a chorus of warnings about the dangers of artificial intelligence.

Yet those results don't prove the algorithm itself is biased against black defendants — a point we've made previously, including in peer-reviewed research. The Broward County classifications are based on recognized risk factors, like a documented history of violence. The classifications do not explicitly consider a defendant's race.

Because of complex social and economic causes, black defendants in Broward County are in reality more likely than whites to be arrested in connection with a violent crime after release, and so classifications designed to predict such outcomes necessarily identify more black defendants as risky. This would be true regardless of whether the judgments were made by a computer or by a human decision maker.

It is not biased algorithms but broader societal inequalities that drive the troubling racial differences we see in Broward County and throughout the country. It is misleading and counterproductive to blame the algorithm for uncovering real statistical patterns. Ignoring these patterns would not resolve the underlying disparities.

Still, like humans, algorithms can be imperfect arbiters of risk, and policymakers should be aware of two important ways in which biased data can corrupt statistical judgments. First, measurement matters. Being arrested for an offense is not the same as committing that offense. Black Americans are much more likely than whites to be arrested on marijuana possession charges despite using the drug at similar rates.

As a result, any algorithm designed to estimate risk of drug arrest (rather than drug use) would yield biased assessments. Recognizing this problem, many jurisdictions — though not all — have decided to focus on a defendant's likelihood of being arrested in connection with a *violent* crime, in part because arrests for violence appear less likely to suffer from racial bias.

Many jurisdictions additionally consider flight risk, and in this case the act of skipping trial can be perfectly observed, which circumvents the potential for biased measurement of behavior.

The second way in which bias can enter the data is through risk factors that are not equally predictive across groups. For example, relative to men with similar criminal histories, women are significantly less likely to commit future violent acts. Consequently, algorithms that inappropriately combine data for all defendants overstate the recidivism risk for women, which can lead to unjustly harsh detention decisions.

Experts have developed gender-specific risk models in response, though not all jurisdictions use them. That choice to ignore best statistical practices creates a fairness problem, but one rooted in poor policy rather than the use of algorithms more generally.

Despite these challenges, research shows that algorithms are important tools for reforming our criminal justice system. Yes, algorithms must be carefully applied and regularly tested to confirm that they perform as intended. Some popular algorithms are proprietary and opaque, stymieing independent evaluation and sowing mistrust. Likewise, not all algorithms are equally well constructed, leaving plenty of room for improvement.

Algorithms are not a panacea for past and present discrimination. Nor are they a substitute for sound policy, which demands inherently human, not algorithmic, choices.

But well-designed algorithms can counter the biases and inconsistencies of unaided human judgments and help ensure equitable outcomes for all.

Sam Corbett-Davies is a Ph.D. student at Stanford; Sharad Goel is an assistant professor at Stanford and executive director of the Stanford Computational Policy Lab; and Sandra González-Bailón is an assistant professor at the University of Pennsylvania.