Public Safety Risk Assessment CLEARINGHOUSE

Human Decisions and Machine Predictions

The Policy Digest Series introduces emerging and promising scholarly research on risk and needs assessments to a broad audience.

This policy digest summarizes and discusses policy implications of a 2017 study of machine learning applied to pretrial release decisions, “Human Decisions and Machine Predictions.”

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2017). *Human decisions and machine predictions* (No. w23180). National Bureau of Economic Research.

Each day, we make decisions based on the information we have before us. Similarly, judges must decide whether to detain people who are facing criminal changes in jail or allow them to go home until their trial based on the available information. They use the information, such as prior criminal history and the nature of the alleged offense, to assess if a person is likely to appear for his or her future trial, known as “flight risk,” or commit a crime before the trial. It may be possible to improve judges’ pre-trial decision-making with the help of machine learning, according to Kleinberg and colleagues in their 2017 working paper, “Human Decisions and Machine Predictions.” They find that the machine learning model outperforms the current practice of judges by more accurately predicting who will fail to appear for trial. Using the machine learning predictions of flight risk instead of those of judges to select people to detain pre-trial could reduce failures to appear by 24.7% and all other crime types by 11.1% with no change in pre-trial detention rates, or reduce pre-trial detention populations by 42.0% with no change in the failure to appear rate.

Machine learning allows computers to learn from and detect patterns in data to make predictions without being explicitly programmed. Using machine learning to assess risk is increasingly common in criminal justice research, and the results often demonstrate that machine learning is more accurate than the statistical methods commonly in practice. Kleinberg *et al.* take their machine learning analysis a step further by not only explaining that the machine learning model outperforms the decisions of judges, but also detailing the implications of using the model’s predictions to make pre-trial release decisions. They explain that the model outperforms judges by correctly classifying people who should be detained pre-trial more frequently than judges do, using the same case information that is available to the judges. They then consider the potential effects of using the model’s predictions rather than the judges’ decisions on crime and failure to appear.

To develop their machine learning model, Kleinberg *et al.* used over 500,000 cases that included a pre-trial decision between 2008 and 2013 in New York City. For these cases, they knew whether the person failed to appear for trial or was arrested for a new crime before the trial. The information used for model development was the same case information that is available to judges during a pre-trial hearing. Machine learning, although complicated, is transparent in the sense that the researcher developing the model must specifically enumerate which factors go into the model. From the model, the researchers can discern which factors are most important to making the predictions. The prediction model then produces consistent outputs based on relevant case characteristics. In contrast, a host of objective and subjective factors may influence the decisions of judges beyond the basic case information, such as the demeanor of the defendant, making it less clear how their decisions are made or how consistent they are from courtroom to courtroom or even from defendant to defendant before the same judge. Kleinberg *et al.* built the model to predict failure to appear—the only legal basis for pre-trial release determinations in non-felony cases in New York City. From these predictions, the researchers assessed whether a person should be released or detained pre-trial. For simplicity, they did not assess whether bail should be set or what a bail amount should be.

With the finalized model, they considered multiple ways to use these predictions in practice. For example, they re-ranked all the people from lowest risk to highest risk of failure to appear to examine what would happen if they focused detention on people with the highest risk. They found that if only the people predicted to be highest risk were detained pre-trial, the New York City jail population would decrease. Using these predictions would avoid unnecessary incarceration and concentrate resources in a way that helps achieve the most benefit to public safety. They also looked at error patterns in judges' decisions, concluding that judges have the most trouble assessing the flight risk of the people the model predicted to be in the highest 1% of risk, and inappropriately released about half of them. The cases in the top 1% of flight risk had a failure to appear rate 3.6 times higher and a re-arrest rate 2.4 times higher than the rates of the total study population. Detaining people in this top 1% would decrease failures to appear, as well instances of crime committed by people awaiting trial.

Beyond reducing crime rates or jail populations, using machine learning model predictions could also reduce racial disparities in pretrial practices. Race and ethnicity were not included in the development of the machine learning model, and Kleinberg and colleagues tested how the model's release rules could be altered to consider race and ethnicity. They changed the machine learning model's release rule to ensure that the share of Black or Hispanic people jailed pretrial was not higher than the share of those groups in all defendants. Incorporating this racial equity factor increased the failure to appear rate minimally as compared to the original model, leading to a failure to appear rate that was 23.0% lower than the judges' decisions, rather than 24.7%.

The potential gains of using the machine learning model to predict flight risk are clear, but policymakers and practitioners must evaluate their priorities to determine which applications are most useful to them. Kleinberg and colleagues provided examples of how liberty of defendants could be increased or crime could be decreased. If detaining people pre-trial is too costly for the jail system, releasing people predicted to be low risk that are currently detained would lessen the jail population and not change crime rates. Conversely, if reducing crime is a priority, the same number of people could be detained, making sure to detain those defendants with the highest flight risk and releasing those with the lowest risk, and crime rates would decrease.

To achieve decreases in pre-trial jail populations or crime through machine learning, judges would need to believe in the predictions of the machine learning model and be willing to base their decisions upon them. Even a perfect model would not bring about better outcomes if practitioners were unwilling to use the predictions. Kleinberg *et al.* put forward several suggestions to guide using the predictions in practice. First, they suggested creating a warning system that would alert judges when a defendant is above the risk threshold that the court has specified, or "re-ranking" all defendants by their predicted risk and letting the model indicate who to release and detain. Second, to retain judicial discretion, these applications of the predictions could serve as an informative tool to guide judges' decisions rather than replace them. Before implementing the prediction model in either of these ways, policymakers would need to consider the input of judges and the implications of altering their current decision-making processes. Pilots of a prediction tool like this one to assess its feasibility would be helpful before widespread implementation.

Implications for Practice

Pre-trial release decision-making can be subjective and lead to unnecessary detention or exacerbate racial disparities. Using machine learning to predict the flight risk of defendants can assist justice system stakeholders in achieving specific goals regarding pre-trial release decision-making in terms of crime, jail incarceration rates, and even racial disparities. To realize these improvements, the public, policymakers, and practitioners can support the use of this data-driven approach to pre-trial hearings that determine the liberty of defendants, a consequential decision. Furthermore, this practice would have to be implemented with consistency and fidelity to obtain the desired reductions in crime and jail populations.

The **Public Safety Risk Assessment Clearinghouse** is a one-stop resource that provides practitioners and policymakers with up-to-date and objective information about risk assessment as well as training and technical assistance on its use.

For more information, please visit: https://psrac.bja.ojp.gov/

Suggested citation for this publication:

Oglesby-Neal, Ashlin (2017) Human Decisions and Machine Predictions *(Policy Digest Number 2017-02)*. Washington, DC: The Public Safety Risk Assessment Clearinghouse.

This project was supported by Grant No. 2015-ZB-BX-K004 awarded by the Bureau of Justice Assistance. The Bureau of Justice Assistance is a component of the Department of Justice's Office of Justice Programs, which also includes the Bureau of Justice Statistics, the National Institute of Justice, the Office of Juvenile Justice and Delinquency Prevention, the Office for Victims of Crime, and the SMART Office. Points of view or opinions in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.