Analysis of Bias in Crime Prediction Algorithms

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Nikhil Sharma

Spring 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor
Pedro A. P. Francisco, Department of Engineering and Society

Predictive policing is a strategy that law enforcement agencies use to predict and prevent criminal activity before it occurs. This approach relies on data analytics and algorithms to identify areas or individuals that are likely to commit crimes. While this strategy has the incredible potential to reduce crime rates, it also has side effects that can be detrimental to society. The use of predictive policing has been shown to magnify existing biases in the criminal justice system. The reason being, algorithms rely on historical crime data, which is shaped by societal factors such as discrimination and unequal access to resources. As a result, the predictions generated by these algorithms can be biased against certain populations, particularly those who have been historically over-policed or marginalized (like black and hispanic populations in states like Florida and California). Recent algorithms have begun to use datasets from the population that have unchecked information, underreported information and more importantly, biased information (Sun et al., 2020). One example of the negative effects of predictive policing is Operation LASER (Los Angeles Strategic Extraction and Restoration). Operation LASER was criticized for its potential to reinforce racial biases in policing.--

The rationale behind predictive policing is simple: *If I know where a crime will occur, then I can stop it before it happens*. An increase in crime has also led to scrutiny for the police in the United States and questions about their training. (Mayson et al., 2018). The goal is to use information to anticipate and mitigate criminal activity, rather than just reacting to it after it occurs. Advocates of predictive policing also argue that it can help police departments deploy resources more efficiently and effectively, leading to better outcomes and improved public safety. By identifying potential crime “hotspots”, police can focus their efforts in those areas and prevent crime before it occurs. Some algorithms also have the ability to identify individuals that can be targeted by criminals.

However, like a lot of innovations that sound helpful to society, there happens to be side effects that are equally detrimental to society. One noticeable side effect of predictive policing is the over-policing of minority communities, which could result in more arrests but not necessarily lead to a reduction in crime. For example, a predictive policing software used in Chicago was more likely to target Black and Latino communities. The algorithm used in this software relied on data such as prior arrests and convictions, which isn't always the best measurement of actual criminal behavior. By relying on this flawed data, the software reinforced existing biases in the criminal justice system and led to discriminatory policing practices. Another reason for the flaws of predictive policing is the use of historical data in order to make predictions about where crime is likely to occur or the persons likely to be involved. A 2019 study by the AI Now Institute states, this data can be "derived from or influenced by corrupt, biased, and unlawful practices," including racially driven policing practices like stop-and-frisk and even the manipulation of crime statistics (Díaz et al., 2021). This causes the programs that are trained using this data to target minority groups, like black and hispanic americans. This then leads to the marginalization of these groups.

I'll be using case studies to analyze my research question: Does bias in predictive policing algorithms advantage certain groups while marginalizing others? With these case studies I will use literature review and document analysis to dive deeper into the topic. To do this I will find primary sources on certain cases and secondary sources of scholars who have already analyzed the cases as well. I'll be looking at a broad range of predictive policing algorithms from all over the United States as well as bias seen in other fields to support my argument.

STS Framework

I will apply the theory of technological politics to the case of Operation LASER, a LAPD algorithmic risk assessment model to show how the bias involved in these algorithms advantages certain groups while marginalizing others. To do this, let us first describe the theory of technological politics. The theory of technological politics is a framework that seeks to understand the social and political effects of technology (Schraube et al., 2021). The theory is based on the idea that technology is not only used by society, but that it also has a hand in how social and political structures are shaped. Technological politics implies that advancements in technology can have negative side effects, such as the creation of inequality between white and black/latino communities. This framework works perfectly for this case because predictive policing is a technology that changes the structure of society and creates inequality between different communities. I will approach this case using the framework and support my claim using corroborating evidence.

Although the police forces that do use these predictive models don't mean to have biased predictions, they cannot avoid them (Goel et al., 2021). This means that the bias is an unintentional effect of the datasets , making it a prime example of implicit bias. Predictive models use the past to make educated guesses about the future, but when our nation's past has been filled with inequality then it only makes sense that those inequalities have bled into the predictions of the future. As you can see, because crime Operation LASER continued to use these datasets, currently marginalized races will continue to be disenfranchised and targeted by the police in the future. I argue that the implicit bias that crime prediction algorithms like Operation LASER are based on, results in disadvantages in black and latino communities and advantages in other communities. To support my argument I will analyze evidence from a handful of corroborating sources of evidence.

Corroborating Evidence

Let's dive into the prime example of the negative side effects of predictive policing, Operation LASER. LASER was implemented by the LAPD in 2011 and used crime data from various stock sources to identify potential hotspots for crime and deploy police resources accordingly. The program hoped to use data and technology to improve policing, while also building a connection between law enforcement and the community. It used machine learning algorithms to identify patterns and trends in the data, which could be used to predict where crimes were most likely to occur in the future. Then, using these predictions, police officers were deployed to high risk areas and were tasked to crack down on communities that were deemed as "hotspots".

The second source of corroborating evidence stems from the predictive policing practices of Chicago, IL. This source is the most similar to Operation LASER of the LAPD as it focuses on the policing algorithm of a police department in a major United States city. The Chicago Police Department (CPD) mainly uses person based predictive policing to mitigate crime in the city. In such systems, law enforcement may predict individuals or groups most likely to be involved in crimes, either as victims or offenders. Person-based predictive policing could involve social network analysis or regression models using risk factors (Brayne et al., 2015). The CPD developed a predictive policing tool called the Strategic Subject List to form a list of people who have been marked by their algorithm as high risk individuals for being involved in violent crimes. This means they could be the perpetrator or the victim. The Strategic Subject List is designed to analyze multiple data sources like criminal records, arrest histories, gang affiliation and social network analysis. After the multiple sources of data are given to the algorithm it uses machine learning techniques to identify individuals and make predictions about which of them is most likely to commit a crime and which ones are most likely to be victims of a crime. Once this Strategic Subject List is completed by the algorithm, the CPD is able to analyze the list and designate officers to the individuals on the list. Police intervention could range anywhere from higher patrols of the area to more aggressive interventions like directly talking to the individuals on the list or their families too. The CPD can also offer them social services like drug treatment or job training to mitigate the risk of crime in the future for these individuals. The end-goal is the same as Operation LASER: to stop the crime before they are actually committed and to reduce overall violence in their city.

There have been several concerns brought up about potential bias in the strategic subject list used by the CPD. (Brayne et al., 2015). The first concern is that the data fed into the algorithm to produce the list may already be affected by the bias of the criminal justice system. Problems such as over-policing of certain communities and racial disparities in arrests and convictions are forms of bias that can be found in the data we feed into the algorithm. An example of this could be using minor offenses to generate the strategic subject list. Minor offenses include drug possessions or loitering. If this list is formed then communities of color would be more affected than white communities as they have more historical data pertaining to these minor offenses. This would then cause more individuals in these communities of color to be marked by the strategic subject list. When the strategic subject list is generated using these biases it reaffirms the biases by targeting these individuals unfairly and using them as further data points to generate the next list. Machine learning algorithms that are trained on biased data will replicate and amplify the biases.

Another source of evidence to show that implicit bias in data has a major role in marginalizing certain groups is a study in England and Wales that showed racial injustices were a result of bias in datasets that were used to train machine learning algorithms that predicted crime (Babuta et al.,, 2019). Implicit bias in predictive policing in England and Wales has led to the over-policing of certain communities. There is evidence that shows that black, asian and other minority groups get inordinately focused on by the police. This targeting leads to a higher rate of arrests and convictions which then leads to the reinforcement of the bias in the dataset. The algorithm's focus on past arrests leads to a constant cycle of over-policing and criminal "hot-spots" in black and asian communities in England and Wales. Surveys done in this study of the predictive policing algorithms in England and Wales quote a police officer stating, "'young black men are more likely to be stopped and searched than young white men, and that's purely down to human bias. That human bias is then introduced into the datasets, and bias is then generated in the outcomes of the application of those datasets." (Babuta et al., 2019). Predictive technological solutions have been criticized for focusing on low-level 'nuisance' crime, or on areas with high crime levels and thus poor neighborhoods.

A piece of evidence that is in the realm outside of the criminal justice system that we will analyze is in the healthcare system. Implicit bias that leads to inequality can also be seen in healthcare algorithms (Obermeyer et at.,, 2019). Black patients seem to get less healthcare treatment than white people because of the datasets used to tell healthcare professionals where to look when deciding where to put more healthcare efforts. Again, the decision is not intentional, rather the algorithm takes in biased data that makes the decision for the healthcare professionals (Angwin et al., 2016). A way to improve these prediction models would be to use newer algorithms that can detect bias and remove it from the equation when predicting crime (Kim et al., 2018).

A piece of evidence to use to critically analyze the discussion of bias in predictive policing algorithms is the case of Patternizr, which is the New York Police Department's (NYPD) "bias-free" predictive policing algorithm. Let's go over what Patternizr is and the design of it to provide some context. Patternizr was created by the NYPD to help crime analysts with identifying patterns of crimes that were committed by the same suspects. The way analysts can use Patternizr is simple: they feed a crime report into the software and then after a few seconds the seed crime is "patternized" and the software generates a report listing a handful of potentially related crimes from the NYPD database. These crimes are also given a score that shows how similar they are to the seed crime. Once the analyst looks though all the crimes generated by the software they are investigated as a pattern and the data from all these crime reports are pooled together to further the investigation of any known suspects or forma prediction of who the suspect is. While other predictive policing applications that we've looked at have focused on either places (Operation LASER) and people (Strategic Subject List), Patternizr is able to focus on the specific crimes it is fed. It specifically looks at the *modus operandi* (M.O.) of the suspect that is committing the crimes (Griffard, 2019). The reason that the algorithm is "bias free " is because of the attributes the software calls for when asking for crime report data. Some of these attributes include various measures of distance, date and time of occurrence, premise type and name, whether a weapon was used, the number of suspects, suspect height(s), suspect weight(s), property taken, unstructured text, and the complaint narrative (Griffard, 2019)). Notice that one popular attribute that is not asked for in this case is race or skin color. By not including these attributes, the algorithm is able to minimize the marginalization of any specific races/communities by making the algorithm blind to color. Location was also recorded very broadly as an attribute so that this would not cause bias towards minority communities where crimes tend to happen more often.

To test the fairness of their algorithm the creators looked at whether Patternizr generated pairs of crimes with suspects of specific racial groups at a different rate than existing identified patterns or random pairings (Griffard, 2019). After analyzing this, they were unable to find evidence that Patternizr recommends any race at a higher rate than others. This case shows that there is a way to get an unbiased take on predictive policing.

There are potential solutions to the bias involved in predictive policing. One solution is diversifying the data used. Since these programs are usually trained using historical crime data, they can output biased results. Crime data isn't always representative of an entire community and can carry bias towards certain groups that were targeted by the police in the past. Diversifying data sources, such as incorporating data from community surveys or other sources, can help provide a more comprehensive and accurate picture of crime. Another solution is regularly evaluating and updating algorithms. These programs should be regularly updated to ensure they aren't magnifying existing bias. Ways to do this include auditing algorithms for bias, retraining models with diversified data or changing how much weight historical data has in the output of these programs. Using synthetic data is another approach. Synthetic data is data that is artificially generated to look like real data. By creating synthetic data that represents a diverse range of individuals and situations, predictive policing algorithms can be trained on a more representative dataset, reducing the impact of historical biases. Alternative approaches like problem-oriented policing or community policing are also ways to address bias and also reduce crime in communities while building trust. Algorithms like k-nearest neighbor, random forest, support vector machine, and LSTM can all do this job and should be implemented in the future (Zhang et al.,, 2020).

Conclusion

In conclusion, predictive policing has biases that negatively affect society and this can be seen in programs like Operation LASER. Predictive policing has a noble goal of addressing crime before it actually happens so communities don't have to deal with the aftermath of crimes. While it has the potential to reduce crime rates, it also has side effects that must be considered. The use of this strategy can have negative consequences for both individuals and society as a whole. To address these side effects, law enforcement agencies must work to mitigate bias and increase transparency, while also recognizing that predictive policing is not a substitute for addressing the underlying causes of crime.

References

Mayson, S. G. (2018). Bias in, bias out. Bias Out (September 28, 2018), 128.

Díaz, Á. (2021, September 13). *Data-driven policing's threat to our constitutional rights*. Brookings. Retrieved March 15, 2023, from https://www.brookings.edu/techstream/data-driven-policings-threat-to-our-constitutional-rights/

Griffard, M. (n.d.). A bias-free predictive policing tool?: An evaluation of the NYPD's patternizr. FLASH: The Fordham Law Archive of Scholarship and History. Retrieved March 20, 2023, from https://ir.lawnet.fordham.edu/ulj/vol47/iss1/2/

Brayne, S., Rosenbalt, A., & Boyd, D. (n.d.). Data & Civil Rights: A New Era of policing and justice Datacivilrights. Retrieved March 20, 2023, from https://www.datacivilrights.org/pubs/2015-1027/Predictive_Policing.pdf

Goel, S., Shroff, R., Skeem, J., & Slobogin, C. (2021). The accuracy, equity, and jurisprudence of criminal risk assessment. In Research handbook on big data law (pp. 9-28). Edward Elgar Publishing

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias risk assessments in criminal sentencing. ProPublica, May, 23.

Babuta, A., & Oswald, M. (2019). Data analytics and algorithmic bias in policing. The royal united services institute for defence and security studies.

Kim, S., Joshi, P., Kalsi, P. S., & Taheri, P. (2018, November). Crime analysis through machine learning. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 415-420). IEEE.

Sun, W., Nasraoui, O., & Shafto, P. (2020). Evolution and impact of bias in human and machine learning algorithm interaction. Plos one, 15(8), e0235502.

Zhang, X., Liu, L., Xiao, L., & Ji, J. (2020). Comparison of machine learning algorithms for predicting crime hotspots. IEEE Access, 8, 181302-181310.

Schraube, Ernst (2021). View of langdon winner's theory of Technological Politics: Rethinking Science and Technology for Future Society: Engaging science, technology, and Society. Retrieved April 28, 2023, from https://estsjournal.org/index.php/ests/article/view/811/463