

# Racism in, Racism out: Machine Learning is Not the Answer to Crime

## A Research Study of Predictive Policing Algorithms

Wang Yangxuan (700025834,046781)

April 17 2021

### 1 Introduction

People recently pays lots of attention to the risks associated with the use of personal data, and many public discourses on data justice all focus on how to protect individuals' data from discrimination caused by unethical data work (Chapter 9). Noble (2018) in the publication named "Algorithms of Oppression: How Search Engines Reinforce Racism" also pointed out the power of algorithms in the age of neoliberalism and the ways those digital decisions reinforce oppressive social relationships. The application of Machine Learning (ML) in Predictive Policing Algorithms instruments has rapidly developed in recent years. And in the criminal justice system, Machine Learning Algorithms provide substantial convenience to predict crime hotspots or estimate recidivism rates. According to Lum and Issac (2016), more and more police departments in the United States and Europe are using predictive policing algorithms software, which is invented by some technology companies like PredPol, Northpointe, Palantir, HunchLabs, and IBM, to predict some trends of criminal activities, identify some potential criminals, and forecast the locations of some potential crimes.

Machine learning is an application of artificial intelligence, which can observe the trends of big data and calculate these data with high speed, precise accuracy, and small hypothesis deviation. Machine learning is based on complex neural networks, it is designed as a human that has self-learning abilities and independent thinking, it can form relationships through pattern recognition and use these relationships to solve various problems (Marr, B., 2016). According to the definition put forward by the RAND Corporation, the predictive policing algorithms is an application of quantitative techniques, it belongs to analytical techniques application. The predictive policing instruments could identify likely targets for police intervention, and also prevent potential criminal activities or solve existing crime problems, by making statistical predictions (Lum and Issac, 2016).

There are two types of the most common predictive policing instruments that exist today. The first type is location-based algorithms, which use the

relationships between the criminal locations, times, events, and historical crime rates to forecast the probability of crimes and predict when and where crimes would be more likely to occur. These algorithms are usually used to predict those crimes in certain weather conditions or predict some riots in those major sporting events. Therefore, these instruments can determine these hot spots of crimes, and polices could plan their patrol and security routes in advance (Heaven, 2020).

The most common tool, which uses these algorithms is called PredPol. PredPol was designed by UCLA scientists who in collaboration with the Los Angeles Police Department, in order to understand how analytical techniques of crime big data can support uncover patterns of criminal behaviours. It divides a city into hundreds of areas of 500 x 500-foot plots, and printouts jurisdiction maps so that can indicate their prediction of these denoted areas, where and when will occur crimes throughout the day. The company claims its study has indicated that the software they created is twice as accurate as human analysts at forecasting where and when crime will occur. However, no independent research has confirmed these results.

Another type of predictive policing instrument uses personal information data, for example, their age, gender, religion, financial situation, marital status, or criminal record, to forecast who is most likely to participate in potential or future criminal activities. Polices often use these personal information-based tools to forecast crimes or intervene in crimes before they occur, also the justice system or court use these tools to determine whether an arrested person is likely to commit a crime again during a pretrial hearing or sentencing (Heaven, 2020).

For example, the most representative tool is called COMPAS, which was created by Northpointe in 1998, is an algorithm aim at correlational offender management profiling for alternative sanctions. Therefore, COMPAS is used in many jurisdictions or criminal justice systems, to help make decisions about criminal sentencing or pretrial release. It can quantify the possibility of a defendant will re-offend or the likelihood of someone whether being arrested after they are released, which based on statistical score systems between 1 and 10 to make a quantify so that to differentiate individuals between the low-risk and high-risk (Northpointe, 2015).

However, these same technological implications raise concerns, including questions about how to ensure that important decisions made using ML are not discriminatory. Considering the automated nature of these predictive policing tool algorithms, it is the main concern is that those misleading correlations could have a powerful implication, while the inherent bias may generate data that magnify prejudices or biases which already exist in our society. It is crucial that how to ensure that ML can make important decisions that are not discriminatory, and also can maximize the predictive performance in a fairness constraint.

In this article, I will discuss the bias in predictive policing tools and the racial discrimination in these algorithms. I will divide this article into four parts, the first is the background introduction. Second, I will describe the racial biases and discriminations in these predictive police tools in some cases. Then, I will give some reasons to explain why there are such challenges and biases in these algorithms. Finally, I will discuss some training solutions, and summarize the main findings and draw conclusions.

## 2 Racial biases and Discriminations

According to the Coalition for Critical Technology, such work like predictive policing tools is not only scientifically illiterate but also perpetuates prejudice against Blacks and other people of color. Many researches have indicated that the judicial system treats these groups more severely than Whites, moreover the predictive policing tool that trained on those data and algorithms would only amplify and enhance social prejudice and racial discriminations (Vincent, 2020).

Also, a Propublica analysis, which is based on data analyzes of 7,000 cases in Broward County, Florida from 2014 to 2014, it found that there were significant racial differences and discrimination in the results of COMPAS scores (Angwin, J., Larson, J., 2016). The study showed that the correct prediction of the predictive policing assessment tool is 61 percent with a recurrence rate. However, Blacks are labeled a twice higher risk than Whites, who would like to re-offend. In other words, even if the Whites commit crimes, they are more like than blacks to be marked as a low-risk score, and then continue to commit other crimes. They also isolate the influence of race from other factors like criminal history and recidivism, also from age and gender of the defendant, the results of this statistical test were still a bias that the black defendants are still more than 77

The Center of Court Innovation had run a new study, in order to determine whether different predictive tools can be used in different please to reach the same conclusions as Propublica (Schwartzapfel, 2019). The researchers had designed a new simple predictive policing assessment tool, which used 9 questions and focused on the relationships with current allegations and past interactions in the justice system. This tool was covered more than 175,000 defendants in New York City in 2015, including all Blacks, Hispanic, and Whites. To track the results of the investigation, they conducted a follow-up survey two years later to see if the forecastability of these predictive tools were accurate, and the finding was almost identical to those of Propublica (Schwartzapfel, 2019). The results showed that among those people who have not been arrested again, nearly a quarter of black defendants were labeled as high-risk groups, while only 17 percent of Hispanic defendants and only 10 percent of white defendants were classified as high-risk groups. The results of this study are indeed full of prejudice and racial discrimination.

In the same way, Hockenberry and Puzzanchera had indicated that it is indisputable that racial discrimination and ethnic disparities exist in the justice system. In 2014, Eric Holder, then attorney general, cautioned that those predictive policing tools would like to exacerbate the unfound and unfairness gaps, which are already common in our society and justice system. Another bias example was that researchers at Shanghai JiaoTong University had declared that they have created an algorithm, which can forecast crimes based on people's facial features. After this research was published, it suffered a lot of controversy and criticism. Researchers from Google and Princeton University issued a long counter-warning, saying that artificial intelligence researchers are re-examining the pseudoscience of appearance. "When put into practice, the pseudoscience of physiognomy becomes the pseudoscience of scientific racism." wrote the researchers (Vincent, 2020). Therefore, it is obvious that the rapid developments in machine learning, it has brought scientific racism into a new

era, that the machine learning models embed biases with human behavior and racial discriminations, which used for model development.

### 3 Reasons and Challenges

In this sector, I want to discuss what are reasons that account for the racial biases and racial discriminations that exist in Machine Learning, more specifically in predictive policing tools. And I also will discuss some challenges in these questions. I will focus on four reasons dedicate that why there are ethical biases and racial discriminations, the design of such software is full of discrimination bias, the use of false data, the lack of transparency in ML algorithms, and the use of collective statistics data to represent individuals. It is worth mentioning that the role of judges is also a challenge to these predictive policing tools.

The first reason why Machine Learning is discriminatory is quite obvious, most engineers and designers of ML algorithms are usually Whites Men. According to an essay published by Lentz (2021), it indicated that there is more than 90 percent of code engineers are men in the UK. At the same time, women only account for only 11.2 percent of STEM (science, technology, engineering, and mathematics) leadership positions in Europe, and in North America, the figure has risen slightly and is still 18.1 percent. A publication named "Black Box Society" addressed that engineers can easily integrate personal biases and racial discrimination into ML algorithms (Pasquale, 2015). This means due to the lack of diversity in the industry, and the inequality of race and gender constitution of engineers, gender discrimination and racial prejudice have been consciously or unconsciously incorporated into the design of algorithms and codes of Machine Learning, even into these predictive policing tools. The New York Times announced that a well-known example of bias can be found in facial recognition software is that if you are whites, facial recognition is accurate. However, if the skin is dark, it is difficult for the software to distinguish between men and women. As Eubanks (2019) pointed out "automated qualification systems and predictive analytics are best understood as political decision-making machines", while they can't eliminate prejudice. Therefore, in my own view, because of the designer's personal bias, the machine learning algorithm itself is biased. This is an industry development problem, not just the algorithm itself, because the algorithm is just a self-learning tool, but what it embeds is the consciousness of the designers themselves.

The second reason is the use of false data, even use the historical bias data in ML. The Machine Learning Algorithm infers future decisions from the training data. If the training data can represent the population accurately, that the algorithm will operate well, and the behavior of the algorithm would like to be more predictable. However, during the training progress, learning biases would be formed. And when the training data is not sufficiently prepared to use the algorithm, the training data does not represent the target population or is not properly selected, the training of this algorithm will cause harmful learned biases directly (Osaba and Welser 2017). A distinctive feature of predictive policing tools is that it provides a huge number of criminal reports as its training data, which leads to the racial bias displayed by such predictive policing tool is learned from historical bias in crimes reports (Temming 2017). Some studies though that it is a simple way to delete sensitive information such as gender or race

factor from the training data, it can prevent this historical racial discriminatory bias well. However, Osaba and Welser (2017) pointed out such a method may cause inaccuracy that hiding sensitive data from the algorithms. There is a well-known fact that a principle called "garbage in the garbage" (Oxford, 2016) in the field of computer science. In other words, using imperfect data or biased data in an objective algorithm will produce imperfect results inevitably, also it will never be able to overcome such imperfection.

As for the third reason, I want to discuss the transparency in ML algorithms. In fact, most of the crime risk assessments used are very simple, largely dependent on static factors (such as age and criminal records), and basic algorithm tools can be used. To some extent, the predictive policing instruments may be transparent. However, it is not easily available to get some potential verification data that as mentioned above, there may be real problems with how decision-makers use risk assessment tools in practice. In other words, the lack of transparency in some sense is really an issue with these predictive tools (Garrett, 2020). Cathy O'Neill (2018) referred to this oppressive algorithm as a "weapon of mathematical destruction" (WMD). Take the federal "First Step Act" as an example, which requires the development of a risk assessment tool and use it for all qualified federal prison inmates (Garrett, 2020). So an independent research committee to design this tool and the act requires to public its algorithm designs. However the verification data developed by the research committee to design the instrument is not public, so researchers cannot easily evaluate claims about its function (Garrett Stevenson, 2019). Especially in such justice systems, the lack of transparency and easy-to-use complaint procedures severely hinder people's ability to question discriminatory practices.

Lastly, I will point out that the use of collective statistics data to represent individuals is another reason that causes bias and discrimination. There are certainly widespread that discrimination based on personal data, and there are serious problems. However, ethical and racial concerns are not limited to personal data, certain uses of such type of personal data also pose a collective risk. Any social group gathered by a common geographic location, characteristics or purpose or long-term personal ties would like to be labeled the same risk level in some predictive policing tools. Even in the absence of personally identifiable characteristics, such data can be regarded as morally sensitive data (Taylor et al 2017). For example, the predictive policing tool named "PredPoll" is used to predict crime locations and assign policing presence, but it often shows a preference for low-income communities, and locations with a high concentration of ethnic minorities, which has led to an increase in the number of police presence in such areas and increase more recognized crime reports in these areas (Temming, 2017). These biases with using individuals' data to represent all the community make the racial discrimination more severely in such tools.

Last but not least, the role of judges is also a challenge to the discrimination issues among such predictive tools. Due to the existence of the above reasons, police forecasting tools do exist to a certain degree of racial discrimination, the risk assessment information provides by such predictive policing tools is only advisory, the judge should consider or reject the recommendations provided by the instrument as appropriate (State v. Loomis, 2016).

## 4 Solutions and Conclusion

A solution proposed by a sponser of the Marshall Project, the Arnold Ventures, build its own risk assessment application, named Public Safety Assessment (PSA), which is now adopted in more than 200 jurisdictions. According to the organization's statement of principles on pretrial risk evaluation, the plan aims to offer an objective, data-based alternative to replace "seriously flawed" human decisions "reflecting deep-rooted prejudices that are almost impossible to correct". They encourage the use of the predictive algorithm as a part of a decision-making framework that is sensitive to problems of racial justice. The researchers found that if judges make decisions based mainly on the severity of the allegations and then conduct risk assessments, the number of people in jail will be greatly reduced, and the false positive rate of racial differences will almost disappear (Schwartzapfel, 2019). I think that to solve the problem of racial discrimination in these types of prediction tools, adjusting the machine learning algorithm, or optimizing the data set are some solutions. However, because the bias of the algorithm itself is constant, therefore, it is believed that the artificial participation of judges or policemen and the use of algorithm tools to assist in case handling can better and more accurately conduct trials, and at the same time, it can eliminate prejudice to a certain extent.

"The basic goal of machine learning is to generalize beyond the training set of algorithms" (Kreatsoulas and Subramanian, 2018). In particular, when it is applied to justice systems or predictive policing tools, even models that are only used for the prediction can produce causality. The results of these tools will be used to make decisions about the distribution of crime hotspots, the prediction of crime rates and re-offended rates even can determine whether a man should go to prison or not. So in this essay, I list lots of reasons why racial discrimination exists in such predictive tools and consideration of the natural bias of the Machine Learning Algorithm, rooting the roles of judges in an understanding of structural racism will improve the accuracy of the predictive policing tools, and also can avoid discrimination and restrictions on minority people.

## 5 References

### References

- [1] @randyrieland, F., 2021. Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased?. [online] Smithsonian Magazine. Available at: <https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/> [Accessed 17 April 2021].
- [2] Angwin, J., Larson, J. (2016, December 30). Bias in criminal risk scores is mathematically inevitable, researchers say.
- [3] Emspak, J., 2021. How a Machine Learns Prejudice. [online] Scientific American. Available at: <https://www.scientificamerican.com/article/how-a-machine-learns-prejudice/> [Accessed 17 April 2021].

- [4] Eubanks, Virginia (2019) Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor. Picador.
- [5] Floridi, L. (2015), The Ethics of Information. Oxford: Oxford University Press. Garrett, B., 2020. Justice in Forensic Algorithms. 2.1, 2(1).
- [6] Google Books. 2021. The Oxford Handbook of Ethics of AI. [online] Available at: [https://books.google.co.uk/books?hl=en&lr=&id=8vQTEAAQBAJ&oi=fnd&pg=PA253&ots=JkjMteIAu&sig=rqmMLg5\\_vJpXxStsu8I9MCuZP3Q&redir\\_esc=y#v=onepage&q&f=false](https://books.google.co.uk/books?hl=en&lr=&id=8vQTEAAQBAJ&oi=fnd&pg=PA253&ots=JkjMteIAu&sig=rqmMLg5_vJpXxStsu8I9MCuZP3Q&redir_esc=y#v=onepage&q&f=false) [Accessed 17 April 2021].
- [7] Hammarlund, N., 2017. Racial Treatment Disparities after Machine Learning Surgical-Appropriateness Adjustment. SSRN Electronic Journal,.
- [8] Hockenberry, S., Puzanchera, C. (2020, April). Juvenile court statistics 2018. National Center for Juvenile Justice.
- [9] Jackson, E. and Mendoza, C., 2020. Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not. 2.1, 2(1).
- [10] Mack, C. and Mack, C., 2021. Garbage In, Garbage Out: Machines Learn to Discriminate from Discriminatory Data. [online] Extract Systems. Available at: <https://www.extractsystems.com/govnews-blog/2019/3/5/garbage-in-garbage-out-machines-learn-to-discriminate-from-discriminatory-data> [Accessed 17 April 2021].
- [11] Medium. 2021. Anti-racism, algorithmic bias, and policing: a brief introduction. [online] Available at: <https://towardsdatascience.com/anti-racism-algorithmic-bias-and-policing-a-brief-introduction-bafa0dc75ac6> [Accessed 17 April 2021].
- [12] Medium. 2021. Racist Data? Human Bias is Infecting AI Development. [online] Available at: <https://towardsdatascience.com/racist-data-human-bias-is-infecting-ai-development-8110c1ec50c> [Accessed 17 April 2021].
- [13] MIT Technology Review. 2021. Predictive policing algorithms are racist. They need to be dismantled.. [online] Available at: <https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/> [Accessed 17 April 2021].
- [14] Noble, S. U. (2018), Algorithms of Oppression: How Search Engines Reinforce Racism, NYU Press, New York
- [15] Nytimes.com. 2021. Opinion — When an Algorithm Helps Send You to Prison (Published 2017). [online] Available at: <https://www.nytimes.com/2017/10/26/opinion/algorithm-compass-sentencing-bias.html> [Accessed 17 April 2021].
- [16] O’Neil, Cathy. 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown/Archetype.
- [17] ProPublica. 2021. Machine Bias. [online] Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [Accessed 17 April 2021].

- [18] Robinson, W., Renson, A. and Naimi, A., 2019. Teaching yourself about structural racism will improve your machine learning. *Biostatistics*, 21(2), pp.339-344.
- [19] Rudin, C., Wang, C. and Coker, B., 2020. The Age of Secrecy and Unfairness in Recidivism Prediction. 2.1, 2(1).
- [20] The Marshall Project. 2021. Can Racist Algorithms Be Fixed?. [online] Available at: <https://www.themarshallproject.org/2019/07/01/can-racist-algorithms-be-fixed> [Accessed 17 April 2021].
- [21] The Verge. 2021. AI experts say research into algorithms that claim to predict criminality must end. [online] Available at: <https://www.theverge.com/2020/6/24/21301465/ai-machine-learning-racist-crime-prediction-coalition-critical-technology-springer-study> [Accessed 17 April 2021].
- [22] The Verge. 2021. AI experts say research into algorithms that claim to predict criminality must end. [online] Available at: <https://www.theverge.com/2020/6/24/21301465/ai-machine-learning-racist-crime-prediction-coalition-critical-technology-springer-study> [Accessed 17 April 2021].
- [23] Universal Rights Group. 2021. Garbage in, garbage out: is AI discriminatory or simply a mirror of IRL inequalities? — Universal Rights Group. [online] Available at: <https://www.universal-rights.org/blog/garbage-in-garbage-out-is-ai-discriminatory-or-simply-a-mirror-of-irl-inequalities/> [Accessed 17 April 2021].
- [24] Vallor, Shannon (2018) *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press.
- [25] Vox. 2021. Yes, artificial intelligence can be racist. [online] Available at: <https://www.vox.com/science-and-health/2019/1/23/18194717/alexandria-ocasio-cortez-ai-bias> [Accessed 17 April 2021].