

# A review of NYPD's patternizr

Mukund Balaji Srinivas  
u7274095

March 17, 2022

# Predictive policing ... what and how

In the past, crime prevention has been more reactive, whereby the focus of policing was on reacting to incidents of crime. Recently however, there is a paradigm shift in policing, law enforcement agencies around the world have been finding out ways to be more proactive than reactive, the argument being preventing crime has more value to the public (Mugari and Obioha 2021) There have been many proactive crime prevention strategies like ILP [Intelligence Led policing], POP [problem-oriented policing] but off late, a data driven strategy that has gained popularity is **predictive policing**.

Predictive policing leverages huge amounts of data and IT systems to predict one of two things: 1) where and when there could be an occurrence of criminal activity, 2) Individuals who are more likely to perpetrate or become victims of crimes (Hälterlein 2021)

## The NYPD *patternizr*

Crime investigators are often interested in groups of related crimes; A person in Bronx NY, armed with a syringe attempted to rob a power drill from a Home Depot store, also having committed a similar crime successfully in a nearby neighborhood, the police could link him to both these instances, however it was not the lead detective who made the connection, rather a machine learning system called *patternizr* (Chohlas-Wood and Levine 2019) Since the December of 2016, NYPD [New York Police Department] has been using this predictive policing tool for crime pattern recognition. The main objective of *patternizr* is to take a “seed” complaint<sup>1</sup> as input and compare it against hundreds of thousands of complaints in the NYPD crime database (Chohlas-Wood and Levine 2019) and output similar complaints based on a “similarity score”.

This system uses supervised machine learning classifiers, in this case each example is a pair of crimes, and the classifier determines how similar they are. Separate models<sup>2</sup> are used for burglaries, robberies, and grand larceny, this is because these crimes have enough prior manually identified patterns<sup>3</sup> that could be used as training data. Furthermore, a portion of this dataset is built from information where the same individual has been arrested for multiple crimes, this is called arrest grouping . Each crime type has about 30,000 complaints which form a pattern, which makes up for a small portion of 200,000-400,000 com-

---

<sup>1</sup>A crime report (Chohlas-Wood & Levine; 2019).

<sup>2</sup>An instance of the machine learning classifier..

<sup>3</sup>A series of crimes committed by the same individual..

plaints, recorded over a span of 10 years (Chohlas-Wood and Levine 2019).

A good portion of the job of building machine learning models is feature engineering, i.e. the process of including / excluding / modifying those relevant feature or attributes that are critical to building a model (Verdonck et al. 2021). In this case, deciding which attributes of a complaint would aid in making the classifier accurate. Features are extracted from the complaints, a complaint typically consists of unstructured text describing the details of the crime, structured information like the time of occurrence, location, crime sub-category and M.O [Modus operandi], which is then processed into the five types of similarities location, date-time, categorical, suspect and unstructured (Chohlas-Wood and Levine 2019).

These five attributes were inculcated into the final model after a lot of formal and informal discussions with crime investigators and enforcement officers. Moreover, metrics were also included to compare the descriptions of suspects. Sensitive information such as race that could potentially bias the model was excluded.

## **The repercussions, foreseen and otherwise**

Wood and Levine hypothesize that when used properly, it could reduce a lot of mundane and uninteresting work for the investigators, instead motivating them to focus their time on the “art of policing” in a fair and unbiased way (Chohlas-Wood and Levine 2019). However, Professor Barry Friedman argues that algorithms don’t need to look at race to be racist, “They rely heavily on criminal records. Much of street policing in recent years — stop and frisk, marijuana enforcement, catching fare-beaters — has been deployed disproportionately against minorities and in poor neighbourhoods” (Friedman 2018).

Furthermore, in recent times, predictive policing systems are emerging enterprise grade software systems with lucrative government contract, hence the incentive to market exaggerated police efficacy figures raises the stakes for governing bodies to seek outside help auditing them.

# Ethical Considerations

The key argument for implementing predictive policing at scale is that predicting crime before it happens is good for the public, however for that argument to hold ground, it must be measured against a code of ethics. We shall be examining it based on two well-known guidelines issued which are the ACS Code (Board 2014) and the US ACM statement (Computing Machinery 2017)

Key stake holders in the case of patternizr would be the developers, enforcement and investigation institutions, crime suspects and the public. Adhering to the recommendations of the USACM, the developers should create awareness by issuing public notices regarding predictive policing and its inherent biases. Moreover, the developer should document the process, inputs, and outputs of these programs in such a way that it is understood by all the stakeholders.

The patternizr is not an open-source software which could potentially be reviewed by anybody, and therefore it should make those decisions and the reasoning behind it more accessible to the ones affected by it for examples defence attorneys and family members of the victim and the suspect, eventually paving the way for a grievance management system. Furthermore, NYPD should be held accountable for all the errors and wrong predictions.

Since the patternizr relies on past criminal records, which is inherently biased towards minorities and by virtue of the data collection method, it does not consider the privacy of the convicted to be of much importance. However, it has opened the algorithms and data for the public to review thereby improving auditability. The machine learning algorithm used by the patternizer are tested against a lot of examples and are also validated by law enforcement and investigating personnel.

## Conclusions

While predictive policing enhances the efficiency of law enforcement personnel, it also introduces some complexities. In the instance of patternizr, the training dataset might not be totally unbiased since any system that relies on historic datasets must contend with racism and segregation.(Griffard 2019) Therefore, it is recommended that the feature engineering process is extensive and includes tools that help identify features and instance that might propagate bias. Furthermore, the law enforcement authorities must audit and update datasets and algorithms very often and publish reports mentioning the details of arrests made based on patternizr recommendations. If patternizr were to inculcate all the recommendations of USACM, then it could truly set a precedent for other law enforcement organisations. On the other hand however, an outright ban on such technologies (Griffard 2019) would induce a policing nightmare. Keeping in mind the pros and cons of patternizr and other similar predictive policing tools, it is wise to include all the stakeholders and update the patternizr to something more holistic that protects the interest of the public in a way that does not induce bias.

## References

- Board, Professional Standards (2014). *Code of Professional Conduct*. Government Document.
- Computing Machinery, Association for (2017). *Statement on Algorithmic Transparency and Accountability*. Government Document.
- Friedman, Barry (2018). In: *The New York Times*. URL: <https://perma.cc/5FG6-53JS>.
- Chohlas-Wood, Alex and E. S. Levine (2019). “A Recommendation Engine to Aid in Identifying Crime Patterns”. In: *INFORMS Journal on Applied Analytics* 49.2, pp. 154–166. DOI: 10.1287/inte.2019.0985. URL: <https://pubsonline.informs.org/doi/abs/10.1287/inte.2019.0985>.
- Griffard, Molly (2019). “A Bias-Free Predictive Policing Tool?: An Evaluation of the NYPD’s pattternizr”. In: *Fordham Urban Law* 47.1.
- Hälterlein, Jens (2021). “Epistemologies of predictive policing: Mathematical social science, social physics and machine learning”. In: *Big Data Society* 8.1. ISSN: 2053-9517 2053-9517. DOI: 10.1177/20539517211003118.
- Mugari, Ishmael and Emeka E. Obioha (2021). “Predictive Policing and Crime Control in The United States of America and Europe: Trends in a Decade of Research and the Future of Predictive Policing”. In: *Social Sciences* 10.6. ISSN: 2076-0760. DOI: 10.3390/socsci10060234.
- Verdonck, Tim et al. (2021). “Special issue on feature engineering editorial”. In: *Machine Learning*. ISSN: 0885-6125 1573-0565. DOI: 10.1007/s10994-021-06042-2.