THE USE OF MACHINE LEARNING IN THE JUSTICE SYSTEM AND IN POLICING

*Kshitij Duraphe**

ABSTRACT

We review the impact of machine learning in the justice system and in policing. The rise of the ease of use of machine learning has led to more and more law enforcement agencies using predictive analytical tools to convict individuals suspected of a crime. However, detractors have accused machine learning systems of being inherently biased; whether it is due to limited data, inexperienced operators, or bias in the data collected. We present a review of how machine learning came to be in used in the justice system and in policing, as well as the pitfalls associated with it. We discuss the algorithms used and the fairness and validity of results given by the algorithm, as well as what questions the algorithm needs to answer. We then look at some historical use cases in the field of recividism. We also analyze some commonly used machine learning tools used primarily in policing known to the public such as PredPol and give a brief slightly mathematical overview of them. At the end, we also present some alternatives to machine learning in the justice system though we acknowledge that these are scanty.

TABLE OF CONTENTS

*Department of Electrical and Computer Engineering, Boston University

INTRODUCTION

This paper seeks to address the use of machine learning in the criminal justice system and policing by considering the roles it plays and its potential biases, in addition to being able to offer alternatives and other means of tackling the issues raised. Machine learning is a field of computer science that aims to give computers the ability to learn without being explicitly programmed. It is a subset of artificial intelligence, which is defined as methods for developing computational models that automate the task of solving problems. Most of the work in machine learning uses supervised learning, where the data or labels are mathematical functions that can be calculated from the input data points. However, there have been many fundamental questions about how to make these models more reliable and effective by using unsupervised learning. Machine learning can be used to predict crime, violent, and non-violent incidents in the criminal justice system. Using machine learning algorithms to classify and predict future criminal behavior has been proposed as a solution to the problem of over-policing by law enforcement. This paper explains the problems that are associated with using these methods and suggests that any algorithm may not be 100% accurate or unbiased when it comes to making judgements. It will discuss the problems with using machine learning and propose another alternative in order to solve these issues. It looks at how machine learning is being used in criminal justice systems worldwide, specifically within policing. It covers the topic of bias in algorithms, as well as the use of other technologies such as open-source programming that can help to minimize, or even eliminate, these biases. Context-specific machine learning approaches will be of great benefit in this context, but can also lead to substantial problems if used inappropriately. Despite these issues, machine learning has become an essential tool in policing and security applications, including facial recognition and threat assessments, with broad implications for governance and democracy. The paper presents a systematic approach for the future generation of criminal justice systems by studying the interaction between machine learning and policing in order to predict, understand, and solve crime problems.

I. AN OVERVIEW OF MACHINE LEARNING IN THE POLICING AND JUSTICE SYSTEM

The intelligence from a machine can be more accurate than that of humans when it comes to detecting crime, and some examples of this are facial recognition in airports, voice recognition for ATMs, or fingerprint analysis for law enforcement agencies. The downside is that there have been reports about these machines leading to false cases (errors) and wrong predictions (false positives). The issue lies within training the machine with enough data so that it can make accurate decisions with what it sees every day. Police try to foresee events and automate labor by using statistics and computer science advancements more and more. Police work is similar to many other professions in this regard; for instance, machines are employed to tally ballots, operate automobiles, forecast the weather, decide loan applications, and more. In the area of security governance, predictive analytics enhance risk management. Examples of cities where police use or have tested predictive policing software include London, Los Angeles, Munich, New Orleans, Philadelphia, and Zurich. This software seeks to either anticipate where crimes are likely to occur or who may be likely to commit a crime in the future. One of the main technologies supporting many of these applications is machine learning. Many people worry that using algorithmic tools to support or automate decision-making has the unintended consequence of reducing accountability. While ML software may rationalize otherwise laborious data-processing tasks, such as sorting through and categorizing a vast cache of documents disclosed in an investigation. The topic of whether using ML models renders individuals unable of accounting for decisions and how they were reached has been raised despite the fact that police accountability was a worry prior to the development of predictive analytics. Making decision-making procedures transparent is vital to keep police accountable for the fairness of their actions and the accuracy of their analysis. For instance, teaching non-statisticians in statistics has also been suggested as a solution to accountability problems. Both approaches assume that when ML models1 are used in a socially significant setting, like policing, increased technical or statistical literacy is required to improve accountability. Even while talks among researchers, practitioners, policymakers, and the general public would undoubtedly benefit from literacy in these domains, it might not be a practical objective. Furthermore, developing ethically righteous and technically competent models requires more than just ML fluency. In-house in police organizations, between police professionals and in-house or commercial developers; stakeholders and affected populations with police and developers; and so on. A variety of actors must consider and discuss the implementation and use of ML software.

A. Background

It is possible to think about predictive policing as a specific technique falling under the larger category of intelligence-led policing. ILP was created as a useful, administrative application for making decisions about police services based on a study of objective facts. The purpose of systematic intelligence gathering and analysis is to increase the cost-effectiveness and effectiveness of measures against crime by offering more precise targeting. Predictive policing prioritizes the objective, scientific selection of strategies and tactics and places a priority on centralized, rationalized, bureaucratic decision-making. It is similar to ILP in that analysis and conclusions are consolidated and rationalized.

B. Accountability

It has long been a concern of police researchers and practitioners to keep police organizations and officials accountable and responsible, which is a crucial aspect of democratic policing. Through accountability systems, whereby police may be held responsible to the public, a bureaucracy, or the law, control over individual and organizational police behaviour has been sought in part. Accountability can refer to political control over the police or to a partnership between the police and the government in which the police are required to justify their actions. This is relevant to the standing of police forces within the democratic system. The use of automated or predictive technologies to enhance decision-making may fundamentally undermine the capacity of officers and organizations to account for decision-making processes and muddle accountability in multi-agent frameworks. The main issue with using algorithms—applied predictive models or automated decision-making systems—remains their lack of transparency. Algorithms are feared to be opaque in the sense that recipients of the algorithm's output infrequently have any understanding of how or why a specific classification was determined from inputs. Either of the aforementioned notions of responsibility is contested when one or more components of the decision-making process are not clear. A statistical model functions as a "black box," processing inputs into output through a calculation that is hidden from the end-user and is generally integrated into commercial, off-the-shelf software. The procedure is essentially opaque to non-experts, albeit it is debatably not fundamentally opaque. If a crucial factor in the formation of the decision-making is fundamentally unknown, how can political control over it be effective? If the police's judgments were partially based on an analysis that they themselves are unable to explain, how can they fully explain their actions? The challenges posed by ML to responsible decision-making have been highlighted, and transparency has been cited as a component of the ideal solution. Information must be both understandable and accessible in order to achieve transparency. However, in regard to partially autonomous learning machines, this is challenging. Thus, some have suggested that accountability may be feasible even in the absence of complete transparency if it is built into the software. The use of algorithmic decision-making or algorithm-supported technology necessitates societal monitoring in addition to technical examination and control.

C. Machine Learning and Policing

According to an often accepted definition of learning in the domain of artificial intelligence, learning has taken place if an agent performs better on subsequent tasks after making observations about the outside environment. In order to have this understanding, all parties must agree on what it means to improve at an activity. How much an agent has learned may be in dispute between two judges of their performance. Accordingly, agreement on how to measure performance must come before agreement on how to judge how well each agent is learning. It's more straightforward to agree on some issues than others. More complicated societal issues, such how to balance performance indicators of law enforcement versus petty offenses, are difficult to come to consensus on. A significant advancement in machine learning, in addition to the ability to learn cognitive tasks, has been the creation of learning algorithms that can approximate complex functions and pick crucial characteristics without overfitting the model to the training sample. The machine can now learn from datasets with hundreds of labelled features, selecting variables and a functional form that are likely to perform well when forecasting future samples thanks to algorithmic advancements. The implication is that the ML algorithm itself chooses the variables included in ML models rather than having human field specialists do so, and that selections are based on a guess by the machine in regards to 'what works' instead of 'what should be used to come to the decision to make it work'. It is essential to contrast ML with human decision-making when discussing whether to utilize ML in police decision-making rather than with ideal decision-making. Machines make decisions in unfavorable circumstances based on ambiguous, confusing, and flawed information. When decisions result in unfair consequences, the procedures may be difficult to follow, and it is rarely easy to determine who should be held accountable for the harm done. However, this is a general issue with decision-making and not specific to decisions made or backed by robots. It is a human strength to learn from cognitive facts. Computers still do not utilize cognitive data as thoroughly as do humans. An important difference between machine and human learning is that ML is based on known algorithms. Since we know the algorithms that machines use (because we create them), and control the data which they are fed (since we record them), it should be in principle that the learning and decision-making processes machines follow are transparent. Rather ironically, one of the main criticisms of decisions made by machine learning is that the decisions made by the machine learning model are opaque. We can always ask humans how they came to their decision, but it is difficult to demonstrate how a machine came to its decision. Although comprehensible, this opacity is concerning since it may allow human players in networks that include both human and machine actors to be disregarded. Although biased policing tactics have also developed from merely human practices, filtering this decision-making process through complex algorithms that few people comprehend gives biased policing tactics unjustified validity. In other words, the output of the machine might seem neutral, leading people to believe it is more objective than it actually is. Many pertinent normative and factual judgments that make up human decisions frequently do not depend on knowing or comprehending the precise interaction of data and algorithm that underlies the decision, nor do they depend on knowing or comprehending the precise neural circuitry of the brain.

D. Fairness and Validity of Machine Learning Results

Decisions can generally be questioned in relation to two main issues: the decision's legitimacy and its fairness. Did the decision result in the desired outcome? is a question we ask to evaluate the model's validity. A reviewer would need to take into account whether the learning model accurately reflects performance based on the established performance metric or whether the performance metric actually measures what we meant to measure in order to determine validity. The breadth of validity issues is likely to intersect with areas outside those of programmers and statisticians since learning goals might be extremely ambiguous and contentious. When evaluating a decision's fairness, one must consider whether the intended outcome and the methods used to attain it were appropriate. Asking questions about the data itself is also relevant. Only crimes that are recorded by the police are turned into crime data, and some crimes are more likely than others to be. Crime statistics go through a vetting process. Legislative action is the initial step in the process (this is where certain actions become crimes). There is still more selection since some crimes go unreported or undiscovered by the general public or the police, and because reporting processes can differ depending on the type of crime and the area. Without mechanisms for inspection or required reporting, some are unlikely to be found, let alone reported. The latter category includes economic crime (tax evasion is an illustration; reporting is reliant on audits and inspections by approved agencies), and the financial sector can employ a variety of practices to thwart financial crime surveillance. Police-generated data bears the imprint of both undesirable and problematic policing methods. An example is provided by a study by the Human Rights Data Analysis Group. The study used police data on drug enforcement in Oakland, California, together with the PredPol algorithm that has been published to create predictive policing forecasts. It compared the predictions to drug usage patterns inferred from data from a national survey on drug use and health. Despite projections suggesting about comparable levels of drug use, the PredPol algorithm would cause black people to be targeted by predictive police at a rate roughly double that of white persons. Other non-Whites beyond Blacks would likewise be disproportionately targeted, as would those with low incomes. This illustration demonstrates how input data used to teach both humans and machines can produce flawed models and unfair methods. In general, an average observer can ask the following questions about data: How are the input data used? What dataset did the model get its training from? Which set is being tested for performance? When and where was the information gathered? Do named variables exist? If so, what are they and which ones influence decisions the most? What are the operations of these variables? How are these mentioned variables measured? Does the input data directly or indirectly capture characteristics that shouldn't have any bearing on the decision? Are any input variables, for instance, connected with sex in such a manner that model choices alter depending on whether you are male or female? Does the data reflect the field that the model's choices will have an impact on? Has the model, for instance, been tested in the environment where it will be used? What are the most noticeable contrasts between the training environment and the current environment? Do any modifications need to be made for specific groups or decisions? How are the statistics gathered? For instance, were they gathered with the purpose that they would be utilized to make choices of this nature? Do we have knowledge of any design- or issue-based selection biases with regard to data collection? Who collects the data?

The observer can also ask the following questions about the ’learning’ done by the ML model: What is the main educational objective? What, for instance, would our society wish to achieve by making these choices? What particular measure(s) or rule(s) are utilized as the benchmark to determine whether a model is learning? What, for instance, are the dependent variable(s)? How is the similarity rule being applied? What behaviors are rewarded or penalized? How is the rule implemented and evaluated? Is the learning objective acknowledged? Does the specific learning goal fully describe what the agent is expected to accomplish? Will maximizing action or decision-making in support of this learning goal detract from or actively undermine other objectives?

The potential problems associated with pattern replication also highlight areas where ML models might supplement human experience and learning. Algorithmic techniques can identify discrimination, but unlike people and organizations, they can be purposefully employed to avoid analyzing shady correlations, such as those between race and crime or ZIP code. However, the work done to identify discriminatory practices and mitigate unfairness in and through algorithmic tools represents opportunities to improve human decision-making. While there is reason to be skeptical of purely technical solutions to protect, for example, a complex social concept such as ’fairness’, algorithmic tools represent an option to improve in human decision-making.

II. Machine Learning and Recividism

Using machine learning to analyze decision-making in the criminal legal system could be a valuable way to identify discrimination and facilitate reconsideration of decisions where justice was inconsistently applied—but reconsideration is still a decision, and stakeholders in criminal law processes should consider whether and how machine learning should play a role in that decision. As the technology progresses, it’s likely that machine learning will be employed in more and more applications. In criminal law, there are many decisions where a machine could be trained to make predictions or recommendations—but should they? Are there situations where human judgment is better-suited than algorithms? In order to answer these questions, we need a better understanding of how machine learning works and the problems it’s designed to solve. When faced with a problem, machine learning algorithms look for patterns in data that can be used to make predictions. The hope is that these predictions will be better than those made by humans or other simple statistical techniques. For example, if we want to predict whether a person will commit another crime after being released from jail on bail, we could use past information about similar defendants who were released and then committed crimes—or we could use information about defendants who were similar but didn't commit crimes. Then, we could feed both sets of data into an algorithm that would try to predict whether a defendant will commit another crime. The hope is that this algorithm would be better at predicting recidivism than humans or simple statistical methods. Machine learning has been used in the past to predict whether a defendant will commit another crime, but not with much success. Machine learning could be used to predict whether a defendant will commit another crime—but only if it's done right.

Most researchers break up the history of predicting recidivism into four generation. The first-generation of risk assessment consists of judgments based on a criminal justice professional's subjective feeling about an offender. Parole board members likely have experience in the criminal justice system and sometimes many years of it. The idea behind first-generation risk assessments at the parole hearing is that the parole board member's experience translates into an accurate gut-level feeling about the offender's likelihood of recidivating. In the second-generation of risk assessment, criminal justice professionals jumped from relying on their judgment to relying on actuarial assessments. In the third-generation, dynamic and criminogenic variables were incorporated into assessments. Four generations are used by researchers to segment the history of recidivism prediction. The first generation of risk assessment consists of conclusions drawn from the individual feelings of a criminal justice practitioner toward an offender. Members of the parole board most often have experience working in the criminal justice system, perhaps for many years. First-generation risk assessments are intended to provide an accurate gut-level evaluation of the offender's chances of recidivism based on the experience of the parole board member. Criminal justice professionals switched from using their own judgment to actuarial assessments in the second generation of risk assessment. Dynamic and criminogenic variables were introduced into assessments in the third generation. The fourth generation saw a shift in emphasis to an integrated implementation of the assessment's findings known as responsively.

A. Accuracy

The accuracy of recidivism risk prediction techniques is its most vital feature. The area under the receiver operator curve captures accuracy more often than total accuracy or correlations. The trade-off between false positives and false negatives is characterized by these categories. When making a binary prediction, the true positives and true negatives are inversely correlated with each other. Higher AUCs are always preferable. The AUCs of the majority of actuarial recidivism assessments range from chance to acceptable. The area under the curves is represented by the AUC. An AUC of.5 would mean that the prediction was no better than random. With an AUC of 1, a perfect prediction would be made.

In meta-analyses, professional judgments significantly underperform actuarial evaluations when evaluated as a whole. Although there are still many unanswered aspects regarding recidivism prediction, it is nearly universally acknowledged that professional judgements fall short of actuarial conclusions. Despite the advancements in machine learning techniques over the past few years, attempts to predict recidivism using machine learning are still somewhat new and infrequent. The methods favored by criminal justice scholars seem to be at odds with the technical literature. Simple regression and choice trees performed the poorest when Duwe and Kim, 2015 compared them to random forests, boosting approaches, decision trees, and decision trees. The most effective techniques were Random Forests and Multiboost, a boosting technique, both of which had an average AUC of.781. Ozkan's April 2017 prediction of widespread reincarceration was based on statistics on inmates freed in 1994 and tracked for three years. He experimented with various machine learning methods and used 150 predictor variables (including XGBoost and random forests). The maximum AUCS were found in Random forests and XGBoost (.824 and .813 respectively). This corresponds to similar research conducted by Curtis, 2018, who calculated that XGBoost would perform better than random forest when developed with a slower learning rate.

III. Real-Life Applications of Machine Learning Systems to criminal justice systems

This section will cover some real-life machine learning applications in justice systems.

A. PredPol

Predictive Policing can roughly be defined as the application of analytical techniques – particularly quantitative techniques – to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions. PredPol is a company that is frequently singled out in discussions about predictive policing, in part because it could be argued that it invented the field, even though it is neither the only player in town nor the cutting edge. What mathematical principles underpin PredPol? According to a Vice report, PredPol had contracts with Utah, California, and Washington (particularly the University of California, Berkeley), making it the most widely used predictive policing firm/software in the US as of 2019. Collaboration between UCLA, the LAPD, and the FBI gave rise to it. PredPol is patented upon the epidemic type aftershock model (ETAS).

1. Epidemic Type Aftershock

The core of ETAS theory is the dynamic occurrence of crime as a continuous time, discrete space epidemic-type aftershock sequence point process. By taking into account a ”parent earthquake” and subsequent background occurrences or aftershocks, point processes are utilized in seismology. The Expectation-Maximization technique used by the ETAS model to predict long-term and short-term hotspots and systematically quantify each's relative contribution to risk. Therefore, the ETAS model can be conceptualized as a branching process: events of the first generation occur according to a Poisson process with constant rate, and afterwards events (from all generations) give birth to N direct offspring events per event. This is a parameterized Poisson random variable. Crime rates rise locally in space as events take place, which causes aftershock crimes. In this model, policing areas are discretized into square boxes. The background rate μ is a (nonparametric histogram) estimate of a stationary Poisson process. The study by Lum and Isaac, which simulates a synthetic population in Oakland, California based on census data, applies a model based on data from the 2011 National Survey on Drug Use and Health (NSDUH) in order to predict an individual's probability of drug use within the past month based on their demographic characteristics, is the first notable quantitative study of the PredPol system. The resulting data collection provides estimates of illicit drug use from a population-based, non-criminal justice data source, and serves as a replacement for the ”ground reality” of drug crime usage data. The authors discover that drug offenses that are reported to police are not an accurate representation of all drug crimes when compared to police records.The authors draw the conclusion that, when they apply their reconstruction of the PredPol algorithm as described above, the model reinforces the apparent biases in the police data rather than correcting for them, suggesting that predictive policing of drug crimes leads to an increase in disproportionate policing of historically overpoliced communities. This is particularly true despite PredPol's assertion that it uses ”just three data points in formulating predictions: prior crime type, crime location, and crime time. It doesn't use any personal information about specific people or groups of people, so any privacy issues and concerns about profiling are gone.”.

B. ClearviewAI

ClearviewAI is an AI tool that in the news recently that has come under public scrutiny because of the vast number of agencies and organizations that employ it. ClearviewAI has been used by American police to create a facial recognition database of protestors, explicitly violating the Fourth Amendment to the American constitution. The Fourth Amendment grew out of the Founding Fathers' desire to not have their privacy invaded without clear legal reasons; in short, there was no power given to the police invade an individual's privacy without a warrant or legal order. It should be noted that while there are, to the best of the author's knowledge, no cases in court where ClearviewAI has explicitly been used to accuse a person of a crime, it is worth mentioning because of the extreme illegal nature of policing it offers law enforcement agencies all over the world. It also is used by organizations like the NBA and Macy's, which makes it all the more reasonable that ClearviewAI has the potential to be abused. Since it does not rely on explicitly matching faces across a variety of datasets and instead pulls millions of images from a variety of websites, ClearviewAI has unprecented power in surveillance and policing because of its ability to match partially covered faces to online profiles with great accuracy. Some argue that this ability could be used by law enforcement to accuse people of crimes they did not commit. It could also be used to surveil and track humans who are marked as politically dangerous. Indeed, ClearviewAI has been rumored to have been sold to government actors.

C. *Jetnopik ALPR*

Jetnopik is a optical technology company that develops a tool for ALPR (Automatic License Plate Reading). It claims that it solves the problem of police not being able to look at license plates and the roads due to the increased speed limits on modern highways. Jetnopik also uses ALPR cameras to identify who is at fault in areas of increased traffic density but low average speed, such as intersections and signals in the middle of cities. It has claimed that its tools aid motorway control, toll control, and civil security. Specifically, the ALPR cameras are capable of scanning multiple lanes of traffic, which was a central problem that previous systems faced. Since Jetnopik ALPR cameras use deep learning over OCR (optical character recognition) technologies that other cameras use, it has prevented other companies from gaining a foothold in the market. This has led to fears that the company could be compromised and the results could be biased, though it is the author's personal opinion that it would be difficult to do so.

IV. Alternative options

There is an argument to be made for using machine learning exclusively in criminal justice systems, even with all the issues discussed above. For example, machine learning systems can be used to develop training programs for new law enforcement officers and justice system trainees. However, there are many challenges that the system faces. Human inspection at each stage of the process is highly likely to be the best possible approach to this problem. However, this overseeing needs to be double-blind in an attempt to remove as much bias as possible. While no algorithm is 100% perfect, it is possible that in the future machine learning algorithms will perform, on average, better than human prosecutors and police at predicting crime and recidivism. This is an emerging field and answers are not very well-known, with many tools being abused rather than studied properly. This is in part due to the unprecedented scale at which machine learning algorithms can be created and deployed. Since machine learning experts are overwhelmingly not using the tools deployed, a situation arises where neither the accused knows why they are accused nor the accuser knows why they are accusing.

CONCLUSION

Machine learning systems in the justice system were examined; a host of issues were brought up. Some pertinent problems were pointed out, particularly in regards to the lack of machine learning expertise in public prosecution. An overview of some patented algorithms were given. What classifies a 'good' model, as well as what a layperson should be able to ask of a machine learning model before attempting to use it, was laid out. In short, human oversight is necessary at every stage of the justice process regardless of the power of machine learning tools available to the prosecution and the police.

V. REFERENCES

Some of the many, many papers and articles read to produce this paper are:

1. Ghasemi, M., Anvari, D., Atapour, M., Stephen wormith, J., Stockdale, K. C., and Spiteri, R. J. (2021). The Application of Machine Learning to a General Risk–Need Assessment Instrument in the Prediction of Criminal Recidivism. Criminal Justice and Behavior, 48(4), 518–538. https://doi.org/10.1177/0093854820969753

2. Annette Vestby, Jonas Vestby, Machine Learning and the Police: Asking the Right Questions, Policing: A Journal of Policy and Practice, Volume 15, Issue 1, March 2021, Pages 44–58, https://doi.org/10.1093/police/paz035

3. Tian An Wong, The Mathematics of Policing

4. Galileo Galilei, Dialogue Concerning the Two Chief World Systems

5. https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/

6. https://predpol.com/

7. https://d3.harvard.edu/platform-rctom/submission/applications-of-machine-learning-in-the-us-criminal-justice-system/

8. Rigano, C. Using Artificial Intelligence to Answer Criminal Justice Needs

9. Idder, A, Coulaux, S., Artificial intelligence in criminal justice: invasion or revolution?

10. https://www.jenoptik.com/

11. https://pubmed.ncbi.nlm.nih.gov/36078307/