Data Science and Predictive Policing

*Dr. Pronab Mohanty**

Introduction

Crime is a purposive deviant behavior that is an integrated result of different social, economic, and environmental factors. Crime imposes a substantial cost on society at individual, community, and national levels. Criminality worldwide makes trillions of dollars yearly, turning crime into one of the world's top 20 economies according to estimates. Based on the most recent reports, the total cost of crime in Canada alone during 2012 is estimated as $81.5 billion, or approximately 5.7% of its national income. Given such whopping costs, crime reduction and prevention strategies have become a top priority for law enforcement agencies. Policymakers inevitably face enormous challenges deploying notoriously scarce resources even more efficiently to apprehend criminals, disrupt criminal networks, and effectively deter crime by investing in crime reduction and prevention strategies. While data collection from different sources, data preparation and information sharing pose difficult tasks, the big challenge for law enforcement agencies is analyzing and extracting knowledge from their large collection of crime data. Applying data-driven approaches on such data can provide a scientific foundation for developing effective crime reduction and prevention strategies through analysis of offenders' spatial decision making and their social standing. The main idea behind crime prediction techniques is that crime is not random but happens in patterned ways. In the crime data mining process the goal is to understand criminal behaviors and extract criminal patterns in order to predict crime and take steps to prevent it. Although crime analysis has a very long history, it has rapidly grown in the last decades to become

---

* The author is the current Deputy Director General, Unique Identification Authority of India

common practice in law enforcement agencies. Crime analysis aims to assist police in criminal apprehension and crime reduction through systematic study of crime. Crime analysis has two main functions: strategic and tactical. Strategic analysis is about examining long-term crime trends. Tactical analysis concentrates on short-term and immediate problems to investigate the relationship between suspects and crime incidents.

The Need for a Predictive Approach to Policing

Policing resources across the world have become increasingly under pressure, and police governance authorities and governments are struggling to meet the increasing demands of both frontline policing and the complicated financial and social impacts of organized crime on society. Along with these pressures, the world of intelligence gathering has remained relatively stable and consistent in its use of human source information to inform law enforcement authorities on the location and proliferation of organized crime activities in our societies. The research demonstrated in this text shows an alternative evidence-based approach to the standard intelligence gathering process by enhancing law enforcement's preventative capacity in identifying organized crime groups that previously went undetected under standard police intelligence gathering techniques. The utilization of conspiracy networks and geographical analysis provides an unbiased scientific methodology to the intelligence process that in addition to human source techniques increases the productivity and accountability of policing resources in the detection and strength of organized crime groups. Early identification and detection of these groups through predictive policing ensures that both law enforcement and communities can proactively engage and mobilize community efforts to disrupt and remove the threat of organized crime on society.

The Emerging Threat of Organized Crime Networks

Amongst criminal activities, organized crime is seen as a principal threat to public safety. Understanding organized crime as a multifaceted, dynamically changing form of criminality is very challenging. There have been some worthwhile studies, but there is no clear conceptualization of this phenomenon, and lack of clarity, transparency, and uncertainty creates obstacles to combat these organizations. While law enforcement agencies are not aware of any formal modeling of organized crime groups in practice, predictive policing techniques have emerged at the intersection of policing, criminological theories and cutting-edge computer science research to develop viable mathematical models of organized crime groups. One of the newly developed areas of computer science is that of social network analysis. From a social network analysis perspective models have been created to leverage detection of covert and criminal communities embedded in society at large and to identify organized crime groups; Predictive policing techniques are now at hand to develop virtual models of organized criminal networks and to study how they might evolve.

Conventional Crime Analysis and Management

The main purpose of the crime analysis is crime reduction. In the policing approaches few mainstream approaches can be observed which leverage the existing methodologies of conventional crime analysis:

• **Standard model of policing**. The standard model of policing uses law enforcement in a reactive manner. Crime analysis helps in efficient allocation of police resources geographically and temporally.

• **Community policing**. Community policing strategies benefit from partnership and collaboration of the community to understand and solve the problems. The main role of crime analysis in these strategies is providing information to citizens.

• **Law and Order policing**. Disorder policing or broken window policing is applying strict law enforcement procedures to minor offences to prevent happening of more serious crimes. Crime analysis is helpful in evaluating the law and order policing approaches.

• **Problem-Oriented policing**. In problem-oriented policing the goal is diagnosing problems within the community and developing appropriate responses which solve the cause of the problems. Crime analysis is used in all phases of a problem oriented policing strategy including scan, analysis, response, and assess.

• **Hotspots policing**. Hotspots policing is a location-based policing in which the police resources are allocated to different areas proportional to crime rate of each area. Crime analysis is used in identifying the hotspots.

Crime analysis contributed to the operational, tactical, and strategic police decision making for decades. But conventional crime analytics has traditionally focused on crime patterns after the crime has been reported. This is clearly inadequate as the victims of the crimes will find little solace in having a scientific analysis of their suffering prepared and exhibited. The need of the hour in law enforcement is to somehow predict the offence in a meaningful manner and mitigate its impact. In response to this sort of need of the citizens, in the last decade the emergence of data science field and its application to the field of law enforcement has given rise to a a new paradigm in this discipline called *predictive policing* introduced in the next section.

What is Predictive Policing

According to the RAND Corporation, predictive policing is 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". Much like how Amazon and Facebook use consumer data to

serve up relevant ads or products to consumers, police departments across the United States and Europe increasingly utilise software from technology companies, such as PredPol, Palantir, HunchLabs, and IBM to identify future offenders, highlight trends in criminal activity, and even forecast the locations of future crimes.

Predictive policing refers to any policing strategy or tactic that develops and uses information and advanced analysis to inform forward-thinking crime prevention, which involves multiple disciplines to form the rules and develop the models. Given that research strongly supports that crime is not random but rather occurs in patterns, the goal of predictive policing methods is to extract crime patterns from historical data at both macro and micro scales as a basis for prediction and prevention of future crimes. This approach uses data-driven tools that benefit from data mining and machine learning techniques for predicting crime locations and temporal characteristics of criminal behavior.

Predictive analysis for policing can be divided into four classes:

• **Predicting offenders.** The goal is predicting future offenders using the history of individuals such as features of their living environment and behavioral patterns.

• **Predicting victims.** This is about identifying individuals who more likely than others may become victims and predicting risky situations for potential victims.

• **Predicting criminal collaborations.** Predicting likely future collaboration between offenders and the type of associated crime.

• **Predicting crime locations.** This task aims at predicting the location of future crimes at individual and aggregate level.

Transformative Potential of Predictive Policing

The rapid evolution of data science, employing techniques and theories drawn from broad areas such as machine learning and data mining, through availability of massive computational power

increasingly influences our daily lives. Data are collected, modelled, and analyzed to uncover the patterns of human behavior and help with predicting social trends. This is changing the way we think about business, politics, education, health, and data science innovations will undoubtedly continue in the years to come. One particular area that has seen limited growth in accepting and using these powerful tools is public safety. This is somewhat surprising given the important role that predictive analytics can play in public safety. New methodologies emerging in data science can advance crime analysis to the next level and move from tracking patterns of crime to predicting those patterns. This has led to a new paradigm of crime analysis, called *predictive policing* which has been defined in the earlier section. Predictive policing uses data science to identify potential targets for criminal activity with the goal of crime prevention. Successful predictive policing results in more proactive policing and less reactive policing. One of the most important goals of crime analysis is generating information that can enhance decision making for deploying police resources to prevent criminal activity. With predictive policing this process becomes more efficient and effective using the discovered patterns about crime locations, crime incidents, crime victims, criminals, criminal groups, and criminal networks. Nevertheless, predictive policing methods are neither a substitute for integrated solutions to policing nor equivalent to a crystal ball that can foretell the future. Predictive policing can facilitate proactive policing and improve intervention strategies by means of making efficient use of limited resources. These methods give law enforcement agencies a set of tools to do more with less. One of the important tasks in predictive policing is analyzing the relationships between offenders to learn the criminal collaboration patterns. Law enforcement agencies have long realized the importance of analyzing conspiracy networks— networks of offenders who have committed crimes together—for designing prevention and intervention strategies. Despite the importance of conspiracy network analysis for public safety,

computational methods for analyzing large-scale networks are rather premature. Contrary to other social networks, concealment of activities and the identity of actors is a common characteristic of conspiracy networks. Still, the network structure is a primary source of information for predictive tasks. Predictive policing methods can significantly take advantage of discovering collaboration patterns in co-offending networks.

How Predictive Policing Works

Predictive policing is primarily the usage of analytical techniques, specifically the quantitative techniques to identify the most probable areas for urgent police intervention and to prevent crimes or help solve crimes using the statistical prediction and processing capabilities. Predictive policing focuses on harnessing the power of geospatial technologies with a combination of information processing capabilities and evidence based police response models to do crime analysis, detection and prevention. As per Andrew Ferguson, an American Journalist and author, patterns of crime are a "function of environmental factors that create vulnerabilities for victims and spaces at certain times". The basic premise of predictive policing is that crime is not necessarily distributed randomly across places and people but has a set pattern. Thus, instead of simply relying on an investigation officers approximations, this method relies on actionable intelligence derived through the power of analytics. To be effective, predictive policing must be based on analytical predictions so that interventions can be channelized. Successful interventions typically have top-level support, sufficient resources, automated systems to provide needed information and assigned personnel with both the freedom to resolve crime problems and accountability for doing so. Many agencies may find simple methods sufficient for their predictive policing needs, though larger agencies that collect large amounts of data may get more benefited from a little more complex model. Predictive policing doesn't intend to change the

traditional police investigation and analysis methods but puts a focus on the usage of advanced analytics on huge and diverse sets of data from multiple sources in conjunction with the existing response models to help forecast and prevent crime. By applying advanced analytics on various records and data sets, law enforcement agencies can predict what and where an incident is likely to happen and deploy its resources accordingly. By using predictive policing, computer models could be empowered for assisting in law enforcement purposes, like anticipating the likely crime events and informing actions to prevent crime. Prediction can focus on fields such as places, people, groups or incidents. Demographic trends, crowd in that region and economic conditions may all affect crime rates in certain areas. Different models are normally used with the historical crime records to inform police about the interventions that can occur which will eventually reduce the number of incidents.

The basic underlying assumption of predictive policing is that crime is not randomly distributed across crowd or locations. Predictive policing seeks to be dependent on algorithmic insights derived using the power of big data rather than relying on a police officer's intuition. Algorithmic technologies also reduce the biases that exist in human decision making. Predictive policing, as a governance measure, reflects the global trend towards data-driven decision-making and its introduction in India sits well with recent efforts to make India a data-rich jurisdiction. Currently there are four prominent methods that help forecast crimes using predictive policing:

1. **Predicting the time and place with increased risk of crime**
2. **Predicting potential future offenders**
3. **Predicting individual or demographics that are most likely to be the victims**
4. **Creation of a profile for past crimes.**

Benefits of Predictive Policing

Predictive policing method significantly changes the response mechanism from reacting to crime to predicting the likelihood of crime and deploying resources to mitigate crime. The existing traditional approach is leveraged by this new predictive approach by improving their capabilities in intelligence led policing, community led policing and hotspot policing. This capability has enabled practitioners to perform more sophisticated analysis, gain a better understanding of the factors underlying criminal behavior and provide better forecasts of where and when crimes may occur. It further uses analytical models and computing power to anticipate the crime events and provides with actionable intelligence. Predictive analytics can focus on variables such as places, people, groups or incidents. Furthermore it can also assist in analyzing demographic trends, parole populations and economic conditions that may affect crime rates in particular areas. Different models are used with the prior crime records to inform police about the interventions that can occur which will eventually reduce the number of incidents.

Crime anywhere in the world creates huge costs for society at the national, community and individual levels. It is extremely important to estimate the cost to society of different crimes as it also helps in economic evaluation of many social programs. GIS (Geographic Information System) has great potential in reducing the cost of crime. For instance, in the US, property crimes against citizens are fewer and lower in value than automobile crimes that traditionally include stealing a mobile, wallet or bicycle worth few hundred dollars as compared to typical automobile crime involving vandalism or theft costing a few thousand dollars. These crime costs can be significantly reduced through an effective usage of GIS solutions, e.g., predicting the crime hotspots and increasing police patrols in those areas and swift and accurate response to acts of vandalism and thefts. Due to overburdening of work, the police force in India usually faces health and social issues which

mandates the need for better allocation of resources. Police officials usually work seven days a week and often have to work very long shifts. So any technology and system of policing which provides a better allocation of resources is extremely desirable. Through hotspots, risk terrain analysis and near-repeat theories, predictive policing promises analysis of data which is easily available that could empower the law enforcement agencies to identify individuals and locations for targeted policing.

Predictive policing offers the opportunity to the law enforcement agencies to pre-emptively act against predicted crimes by focusing on crime-prone areas and individuals at the risk of offending or being targeted. If crimes can be stopped before they are committed, it has great social and economic value not just for those at the risk of being victims of such crimes, but also for the offenders, as they can be stopped from making life altering mistakes. Predictive policing is a specialized process which aims to bring insights from various fields such as actuarial science, statistics, criminology and constructive understanding of local surroundings. So it is expected that this will lead to a more in-depth and holistic analyses of crime patterns.

Some Illustrations

One of the illustrations of the usage of predictive policing in future can be the reading machines that may get into the main stream by mid-2040's. While this tech will have profound ramifications on how courts operate and how innocence and guilt are proven, it will have an equally insightful impact on how we forecast crime. Consider a scenario where by the late 2040's, laws are passed by governments around the world stipulating the need of handing over the contents of their online profiles, have their psychological profiles assessed by psychologists and their thought profiles documented by the thought reading machines. Once all existing and new prisoners have been assessed, their digital, psychological, thought, demographic and criminal records will be shared

with the national criminal investigations supercomputer. The thought profiles of millions of prisoners will be analyzed and statistically modelled against their public metadata with the goal of isolating a collection of detailed criminal profile types. In other words, the computer will create a series of archetypes that possess a certain set of attributes that predict a certain level of criminal inclination. These criminal profiles will then be compared against the profiles of every citizen in the state.

Another illustrative application of predictive policing can be its usage with the Unmanned Aerial Vehicle or, UAV technology. With the advancements in UAV technology, many sensors that are essential to monitor and help the law enforcement agencies can be built-in various applications. Data from these various sources can then be transferred on a real-time basis to computers capable of performing complex calculations. As the processing of data becomes efficient and with even more refined predictive algorithms, machines can autonomously drive the UAVs remotely to monitor the crime sensitive areas. This step forward in the future would enable the law enforcement agencies to reach places quickly and thus assist them in becoming omnipresent towards protecting the cities. If the cameras on UAVs are enabled with facial recognition and retina scanner, then real-time identification of individuals captured on any camera can simply help map and track the missing persons, fugitive and suspects. If the drone hovering above is able to record any incident, artificial intelligence/ deep learning shall be able to calculate the distance from nearest police station or patrolling vehicle and provide a first-hand information about the incident. Alerting the police department is just a small step, but artificial intelligence/ deep learning is capable of guiding the police patrols to the crime sites with interactive conversations.

Bias in police‑recorded data

But there is also the concern that police‑recorded data sets are rife with systematic bias. Predictive policing software is designed to learn and reproduce patterns in data, but if biased data is used to train these

predictive models, the models will reproduce and in some cases amplify those same biases. At best, this renders the predictive models ineffective. At worst, it results in discriminatory policing. Decades of criminological research, dating to at least the nineteenth century, have shown that police databases are not a complete census of all criminal offences, nor do they constitute a representative random sample. Empirical evidence suggests that police officers – either implicitly or explicitly – consider race and ethnicity in their determination of which persons to detain and search and which neighbourhoods to patrol. If police focus attention on certain ethnic groups and certain neighbourhoods, it is likely that police records will systematically over−represent those groups and neighbourhoods. That is, crimes that occur in locations frequented by police are more likely to appear in the database simply because that is where the police are patrolling. Bias in police records can also be attributed to levels of community trust in police, and the desired amount of local policing – both of which can be expected to vary according to geographic location and the demographic make−up of communities. These effects manifest as unequal crime reporting rates across a police-beat area.

With many of the crimes in police databases being citizen−reported, a major source of bias may actually be community−driven rather than police−driven. How these two factors balance each other is unknown and is likely to vary with the type of crime. Nevertheless, it is clear that police records do not measure crime. They measure some complex interaction between criminality, policing strategy, and community−police relations. Predictive modeling is increasingly being employed to assist human decision-makers. One purported advantage of replacing human judgment with computer models in high stakes settings-- such as sentencing, hiring, policing, college admissions, and parole decisions-- is the perceived "neutrality" of computers. It is argued that because computer models do not hold personal prejudice, the predictions they produce will be equally free from prejudice. There is growing recognition that employing algorithms does not remove the

potential for bias, and can even amplify it, since training data were inevitably generated by a process that is itself biased. The section below illustrates the deleterious effects of data bias on predictive policing.

In late 2013, Robert McDaniel – a 22–year–old African-american who lives on the South Side of Chicago – received an unannounced visit by a Chicago Police Department commander to warn him not to commit any further crimes. The visit took McDaniel by surprise. He had not committed a crime, did not have a violent criminal record, and had had no recent contact with law enforcement. So why did the police come knocking? It turns out that McDaniel was one of approximately 400 people to have been placed on Chicago Police Department's 'heat list'. These individuals had all been forecast to be potentially involved in violent crime, based on an analysis of geographic location and arrest data. The heat list is one of a growing suite of predictive 'Big Data' systems used in police departments across the USA and in Europe to attempt what was previously thought impossible: to stop crime before it occurs. This seems like the sort of thing citizens would want their police to be doing. But predictive policing software – and the policing tactics based on it – has raised serious concerns among community activists, legal scholars, and sceptical police chiefs. These concerns include: the apparent conflict with protections against unlawful search and seizure and the concept of reasonable suspicion; the lack of transparency from both police departments and private firms regarding how predictive policing models are built; how departments utilise their data; and whether the programs unnecessarily target specific groups more than others.

Conclusion and the Way Forward

As has been seen in this discussion, predictive policing promises to be a game changing concept. It is understood that the application of analytical and quantitative approaches will continue to be an important part of police activities. While it is predictive in nature, the effort involves crunching data of past crimes to forecast and thus, in essence it

is largely reactionary policing with a proactive approach. Predictive policing begins with data analysis, so it is important that the law enforcement agencies understand the data and goal of the analysis. It is extremely important to obtain the concurrence of user agencies and to understand the context in which the tools are supposed to be used. In addition, the agencies are required to work closely with analysts to ensure that their findings are tactically useful and able to reduce the rates of crimes prevalent in the society. Going forward, police administrators must deal with the proper scope of data collection, retention and use and be able to explain to the public how data is being used to enhance public safety. As evident, at the center of predictive policing is data: crime data, FIR data, personal data, gang data, social data, associational data, locational data, environmental data, social media data, behavioral data, consumer transactions data, personal communications, surveillance sources data and a growing web of sensors. The use of big data in world of law enforcement is still largely in its early stages but offers more incriminating bits of data to use and study. With the advent of technologies, law enforcement agencies can resort to finding suspicious activity from the digital trail that is left behind by the criminal.

The potential of predictive policing in law enforcement is indisputably real and so is the fear of invading into the privacy of community. The ever increasing law and order rhetoric can also lead to surveillance overreach. Police supervisors, advocates, communities and the corresponding governments are required to address these concerns before venturing into any large scale implementation of such technologies. It is critically important to analyze the applications of predictive policing across the multi-faceted procedures being followed within the criminal justice system. In fact, it has been assessed that in the different stages of a criminal procedure, from starting an investigation to gathering evidence, followed by arrest, trial, conviction and sentencing, as the individual gets subjected to serious sanctions and punitive actions by the law enforcement agencies, the system is able to obtain a higher

standard of certainty about the crimes committed by the individual through predictive policing that also helps in legitimizing the particular action by agencies. Given that the nature of predictive evidence is probability based, it can be inferred that arrest warrants or trials, on their own can have a tangible impact. Predictive policing under the current scenario has been designed to calculate the risk of future crime occurring based on statistical analysis of past crime data. It is also important to deploy need-based and customized applications as per the unique requirements of the agencies. Smaller agencies may not need expensive software and agencies of any size should compare open-source alternatives to commercial products. Larger agencies will want to consider more sophisticated systems. However, the key for agencies of all sizes is to think of the tools as providing situational awareness rather than crystal balls. The systems should help agencies understand the where, when and who of crime and identify the specific problems driving that criminal activity; this information will help support interventions to address these problems and reduce crime.