Crime Prediction with Geo Hotspots and Heatmaps using Machine Learning

BHAVANA.S

Faculty of Engineering and Technology Ramaiah University of Applied Sciences BENGALURU, INDIA

Dr. PUSHPAVATHI T P

Associate Professor Faculty of Engineering and Technology Ramaiah University of Applied Sciences BENGALURU, INDIA

Abstract—Crime is regarded as the most severe and urgent problem facing our society, and preventing it would be a crucial task. A considerable number of crimes are committed each day. This requires keeping track of all offences and maintaining a database so they may be accessed in the future. However, maintaining a reliable crime record and analyzing this information to help forecast and foundation future crimes is a challenge. This work aims to examine a collection including a selection of crimes and predict the crime that may emerge in the future based on the dispersion of causes. Law enforcement agencies have devised various crime prevention techniques to recognize the gravity of this issue. However, in most situations, the inefficiency and sluggishness of these preventative measures render them incapable of predicting crime patterns for initial prevention. This research provides a regression-based approach incorporating temporal, statistical correlations and other pertinent data to predict the crime patterns of distinct districts concerning states. We employ Auto-Regressive Integrated Moving Average (ARIMA) to evaluate the crime trends of different districts of states for the analysis of occurrences of crimes by employing the crime trends because seasonal information is a valuable addition to applying time-series patterns

***Keywords*—**Crime Analysis, Geo Hotspots, heat maps, ARIMA Model

I. INTRODUCTION

Complicated community marvel leading a crime that has a prominent influence on the public's intellect of safety and can cost humanity. This marvel's intricacy makes it challenging to forecast. Every day, new-fangled crimes of various sorts are committed due to the decisions and monotonous actions of offenders, dupes, and responsible guards, ensuring in several new crime tendencies. Among the primary goals of criminology is the investigation of these crime traits, crime trends, and criminal conduct. Due to rapid technology advancement and newly built technologies, several new criminal investigation possibilities have emerged in recent decades. Criminal investigation increasingly relies on data mining techniques due to the difficulties of dealing with the vast amounts of data created daily. The digitization of society has also resulted in the evolution of various police techniques and ideologies, which are frequently based on massive data. Moreover, the significance of system-related decision-making procedures in police grows [1]. In today's investigation, action, and crime prevention, criminal analysis and the utilization of massive amounts of data show a crucial role.

In the context of police, attempting to predict the forthcoming is not a novel way of intelligent but rather a typical practice. For instance, the primary responsibility of the police is to prevent potential future threats by proactive measures. Two visions are fundamental to the concept of crime prediction: First, hominoid activities can differ owing to differing values and desires, as well as the effect of spatial variations [2]; 2nd, crimes do not happen regularly or arbitrarily in space and time [3]; and third, human actions are not random in space and time [4] (p. 5). The methodical use of digital resolutions, massive data, and the increasing difficulty of data analysis in police are novel developments. By investigating what they may have previously recognized built on their current data, the police exposed the door to big data and concerning data to intellect. In this optimistic mindset, several data analysis techniques have been deployed[19]. These include predicting and implementing a predictive supervising model to measure an all-inclusive crime analysis and forecasting strategy. Due to the vast volume of data relating to criminal activity, predicting crime and identifying offenders is one of the police department's highest priorities.

Therefore, there is a need for technologies that could expedite case resolution [4]. This approach is based on the premise that crimes may be readily anticipated once we can sift through vast amounts of data to identify valuable patterns for determining what needed [5] is. Recent advances in machine learning make this possible. As inputs, it will provide the date, time, location (longitude and latitude), and yield, which provide information on which crimes are probable to occur in that area. It principally reveals the criminal hotspots [6]. According to Figure 1, the top 10 crimes, which statistically account for 83.5% of all records, are larceny/theft, other offences, noncriminal, attack, narcotic, vehicle holdup, vandalism, warrants, stealing, and suspected OCC [17]. Certain crimes are more likely to occur, so it is legitimate to advise investing additional police resources to deal with them.

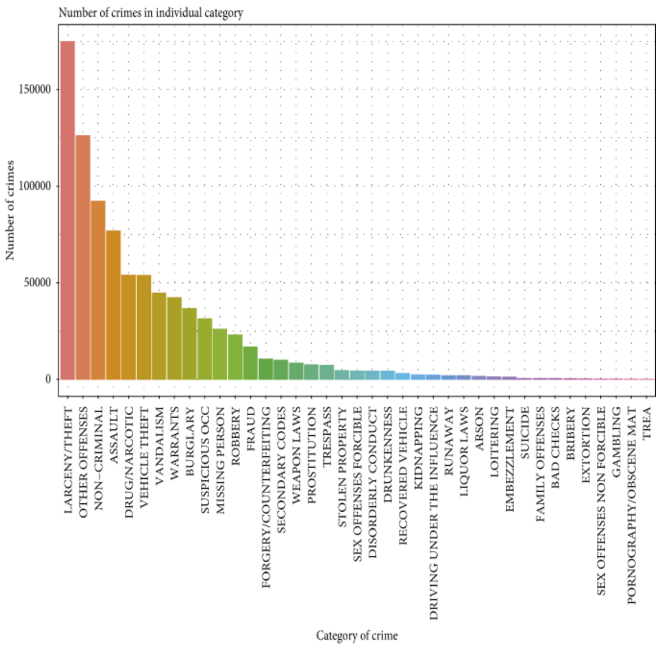


Figure 1: Number of different category crimes

This article provides an overview of the various crime forecasting techniques, including crime predicting, risk terrain modelling, and time series investigation. The temporal component is next discussed, with time series decomposition and ARIMA modelling of crime data taken into account. Finally, inferences are reached on the three techniques, and their potential and limits are examined.

II. RELATED WORK

Algorithmic crime mapping uses current information dispensation technologies to merge GIS data, ordinal maps, and facts to comprehend crime's spread better. Conferring to Mammalian et al. [7], the method of crime plotting allows law supervision organizations to evaluate and connect databases to provide a thorough overview of crime episodes and associated variables within a neighbourhood or other geographic region. It has been functional for various crime kinds, including drug incidents [9], ecological crimes [10], burglary [10], gang violence [11], burglary recurrence victimization [12], inhabited burglaries [13], and serial robberies [14], and is a multipurpose instrument for crime enquiry officers to comprehend the spread of crime [8]. The Crime Mapping Research Center (CMRC) Institute of Impartiality disseminated a national survey to discover whether interventions employed GIS, its purpose and the reasons for refusal [15]. A pilot learning was done to investigate the implementation of crime plotting by selecting 125 police actions from the LEMAS 1999 review of sections with more than 100 police personnel [16]. The presence of a clear relationship between algorithmic crime planning and hot spot techniques in policing and the importance of both fundamental and applied research concerning crime locations and hot spots in algorithmic crime mapping's spread are two additional significant discoveries. Several nations outside the United States were also early adopters of algorithmic crime mapping. The vast mainstream of the hotspot and place-based analytical controlling processes concentrate on crimes primarily described to police by the civic (e.g., theft, theft, attack) rather than arrests [20]. Spatial bundling has been studied to determine where crimes essence in space and time, e.g., to find hotspots or to anticipate the location of future crimes [21]. To further improve hotspot identification, researchers have studied spatiotemporal correlations across extended periods [22]. Spatial ellipses, thematic plotting of terrestrial regions, grid thematic plotting, and Kernel Density Estimation (KDE) are the most used techniques [23]. KDE is one of the most prominent strategies and is accurate in its predictions. In addition, the approach is notable for creating smooth and accurate maps [24]. A novel criminal hotspot mapping device, Hotspot Optimization Tool (HOT), and submission of spatial data analysis to the subject of hotspot mapping is one of many other ways. The advantages and disadvantages of using associated variables in hotspot mapping are examined. [25]. Different exploration demonstrates that calculating co-offending complex observing is an operative method for exploring information regarding criminal societies from bulky real-world crime datasets, particularly police-reported crime facts, which are nearly unbearable to acquire using conventional crime analysis methods [26]. However, such tools and approaches have been connected with several issues. Research has shown that there is racial prejudice in police, including racial profiling of cars [27], pedestrian stops [28], traffic fines, narcotics enforcement and arrests [29], use of force [30], and even the choice to kill white or black illegal distrusts in an exercise simulator [31]. Though the processes underlying these documented outlines of tribal inequality persist hard to separate [32], there is little question that racial inequalities in police results exist. Recent research papers [33] have focused on the racial prejudice of predictive police systems. Regarding place-based analytical regulating approaches that estimate an interval and place where a crime might occur, there is a worry that racially partial enforcement tactics may be aimed toward certain regions as opposed to others. Meaningful they are in a forecast may enhance police officers' consciousness in methods that increase their prejudices. In other words, a minority spotted in a prediction region may be exposed to more biased policy actions than the identical minority viewed outside a forecast area [34]. Additionally, there are serious privacy issues regarding hot spot policing [35]. As police intelligence are public records and many police sections provide online crime mapping tackles, the distribution of spatial crime facts might be tricky when the sites of crimes can be connected to particular residences and persons. Lastly, several research [36] studied the elements that influence the usage of algorithmic crime technology. This study suggests that law enforcement agencies should use algorithmic crime mapping to focus on growing the full-time compensated employees, as long as academy exercises, assigning beat majors to precise areas, and frequently updating knowledge to analyze public difficulties.

III. PROPOSED SYSTEM

The traditional approach to crime forecasting employs predictive policing as a tool for calculating geographical crime risks. Predictive policing entails several phases and procedures that build upon one another, beginning with analyzing the individual incident and collecting and processing the data necessary for crime prediction. The systematic procedure shown in Figure 2 provides insight into the various phases for adopting predictive policing from the perspective of the police department, as was also the case. Variations are possible whenever machine learning methods are used, but similar designs are likely to exist.

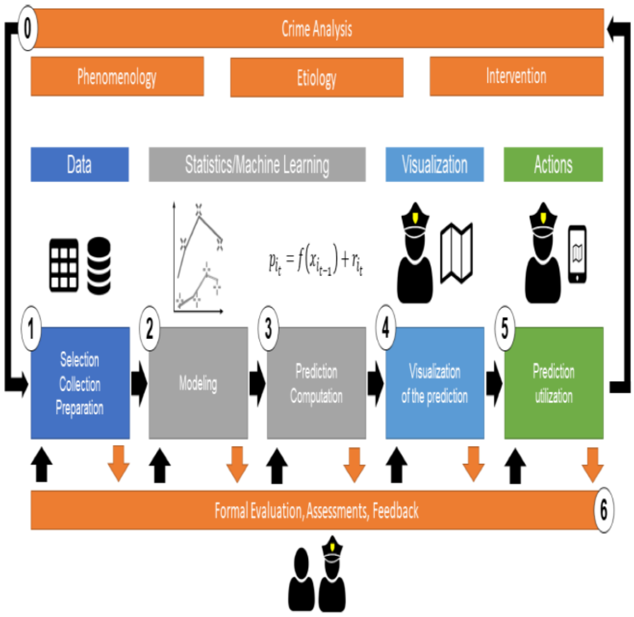


Figure 2: Shows the depiction of a predictive policing process [18]

The underlying premise is that the past is indicative of the future, or at least the current (days for tactical tactics, months to a insufficient years for tactical tactics). Even though this is true for answering the queries about the techniques used will vary subject to the situation and the objective:

- Hot spot analysis, statistical regression, data mining, and near-repeat approaches are often used to determine where a crime will occur over a defined time frame (when, ranging from a day to a year, depending on the method and application) and, therefore, who is likely to be a victim.
- Temporal and spatiotemporal approaches may determine the most probable time for a crime. In addition to identifying victims, these systems also account for the ambient population and nearby inhabitants.
- Risk terrain analysis is suited for identifying the geographical characteristics that contribute to crime risk and identifying potential places for a particular kind of crime (where).

When using analytic approaches to answer the question, "What are the probable reasons of the crime?" practitioners should use extreme caution. In general, they must avoid attributing causal linkages to the outcomes of studies using any of these techniques. For example, a statistical correlation between a factor and an increased likelihood of crime does not always imply that the factor "causes" crime. A well-known illustration shows this issue: It would be incorrect to argue that police are responsible for high crime levels in places where they operate.

Six analytic categories structure the predictive strategies described in this chapter for predicting crime risk. Risk terrain analysis, hot spot investigation, regression approaches, data mining practices, near-repeat procedures, and spatial-temporal investigation. Not every approach under these categories is equally complicated. They are only similar in that they are all used to forecast the timing and location of future crimes and rely on past crime data. Therefore, we classify the practices into four classes:

- Classical statistical techniques: This category consists of conventional statistical procedures, including most regression types, data mining, time-series analysis, and seasonal adjustments.
- Simple procedures: Simple approaches need nothing in complex computation or large quantities of data. For instance, most heuristic approaches are essential, focusing more on checklists and indexes than on analyzing massive data sets.
- Complex applications include new and inventive ways or procedures that demand substantial volumes of data and advanced computing capabilities. This category includes several recent data mining techniques and a few techniques with a high degree of repetition.
- Tailored tactics: In several instances analyzed here, existing techniques were modified to facilitate predictive policing. For instance, traditional statistical techniques may be used to generate basic heat maps with colour-coded grids representing the intensity of criminal activity in a particular location.

A. ARIMA

The ARIMA stands for the autoregressive integrated moving average which is a statistical analysis model.ARIMA is a kind of regression investigation that may be recycled for periodic and non-seasonal data. It may identify the relationship between the reliant variable and varying factors. There are three parameters in an ARIMA model: seasonality, tendency, and clatter denoted as p, q, and d, correspondingly. p is the directive of the AR term, which refers to the autoregressive, that illustrates the influence of an altering adjustable that retreats on its previous ethics. Seasonality allows us to identify the recurring short-term series of a series. It arises when an arrangement exhibits periodic variations (monthly/quarterly/yearly) (i.e. crime surges on a specific month of the year). q is the directive of the MA word relating to the affecting regular, which is the dependence amongst an observed worth and the enduring error from a touching average model functional to prior explanations. The trend is the variable value in the sequence, which may not be lined.

The chance variation in the data that the model cannot account for is called noise (i.e. sudden variations in crime rate for unforeseen incidents) d is the command of the I tenure relating to the different numerical features of the data, including the variance, mean, and autocorrelation, to calculate the differences among the past and present standards for rationalizing the data sequence. In this study, we use an expanded Auto ARIMA ideal in which the model determines the best p, q, and d variables to provide more accurate forecasts. It searches various mixtures of p, q, and d limits using a step-by-step method and finds the optimal model. The choice procedure considers BIC (Bayesian Information Criterion) and AIC (American Information Criterion). Using an evaluation of each copy's quality, these numbers pick the top model from a limited collection of models. The better the model, the lower these numbers. In our study, we implement the following Auto ARIMA architectural steps:

1) Meanwhile, this is a unilabiate regression; we opted to exclude columns additional than the date and amount of offences for each crime category. Therefore, for every crime forecast, we forecast one kind of corruption at a time and generate a fresh data set.

2) Choose Time stages: The amount of time steps is specified as the indices of the time stages in the forecasting process. For example, for a one-step out-of-sample prediction, the value may be set to 1. In our situation, we set it to 12 to get a 12-month prediction.

3) Train Auto ARIMA: Individually fit the model to each univariable series.

4) Predict ethics on evaluation: Once training is complete, we assess our model by predicting values on the confirmation.

5) Compute RMSE: The routine of our model is validated by comparing the foretold values to the actual values using Equation 1.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\text{Predicted}_i - \text{Actual}_i)^2}{N}} \quad (1)$$

IV. RESULT AND DISCUSSION

For this Research work, a crime dataset with several parameters are considered such as INCIDENT_NUMBER,OFFENSE_CODE,OFFENSE_CODE_GROUP,OFFENSE_DESCRIPTION,DISTRICT, REPORTING_AREA,SHOOTING,OCCURRED_ON_DATE,YEAR,MONTH,DAY_OF_WEEK,HOUR,UCR_PART,STREET,LATITUDE,LONGITUDE,location which is of 327820 data are considered for this analysis. Using python, the Auto ARIMA model has been implemented, and it has been iterated. Figure 3 displays the graph of density against the number of samples. Figure 4 shows the autocorrelation of the samples, Figure 5 displays the partial autocorrelation of the samples, Figure 6 displays the sample crime rates in different districts across the states.

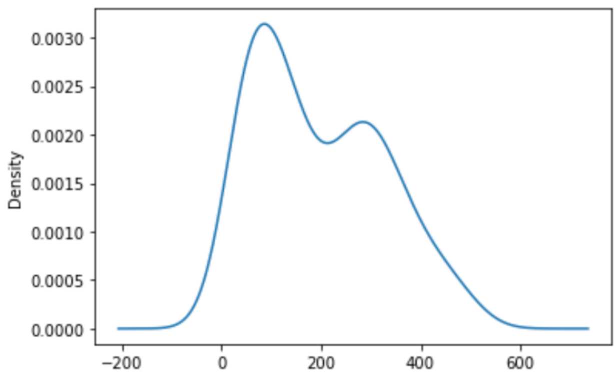


Figure 3: The graph of density against the number of samples

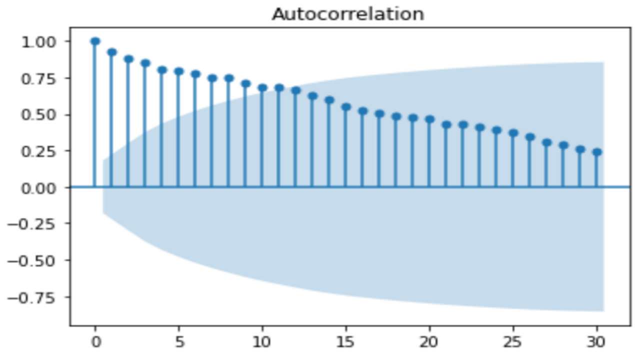


Figure 4: Shows the autocorrelation of the samples

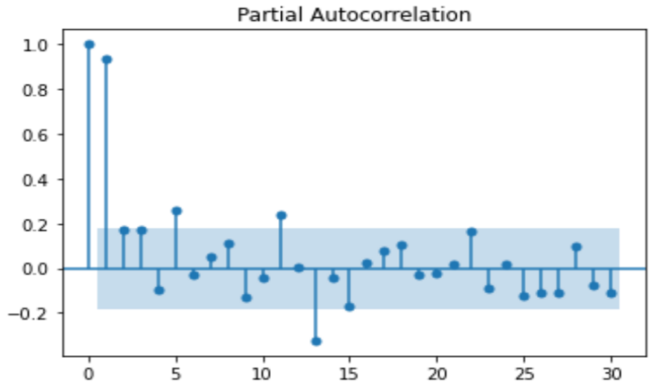
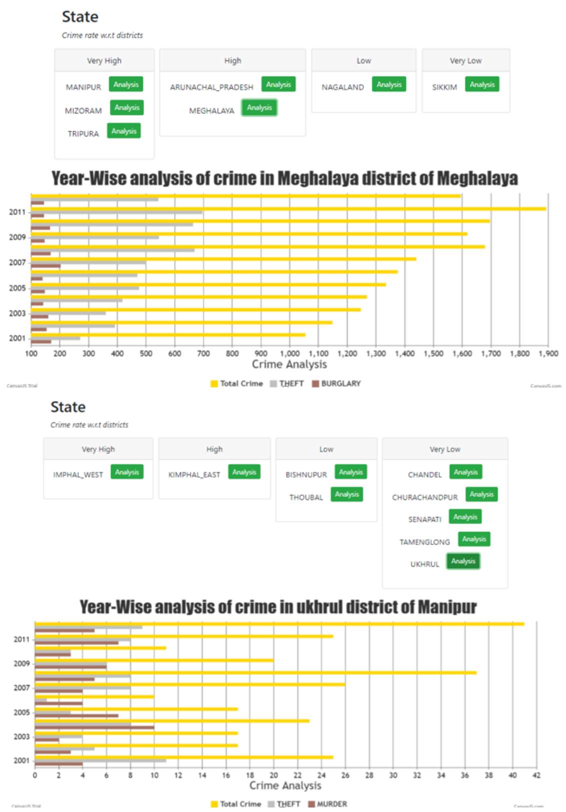


FIGURE 5: DISPLAYS THE PARTIAL AUTOCORRELATION OF THE SAMPLES



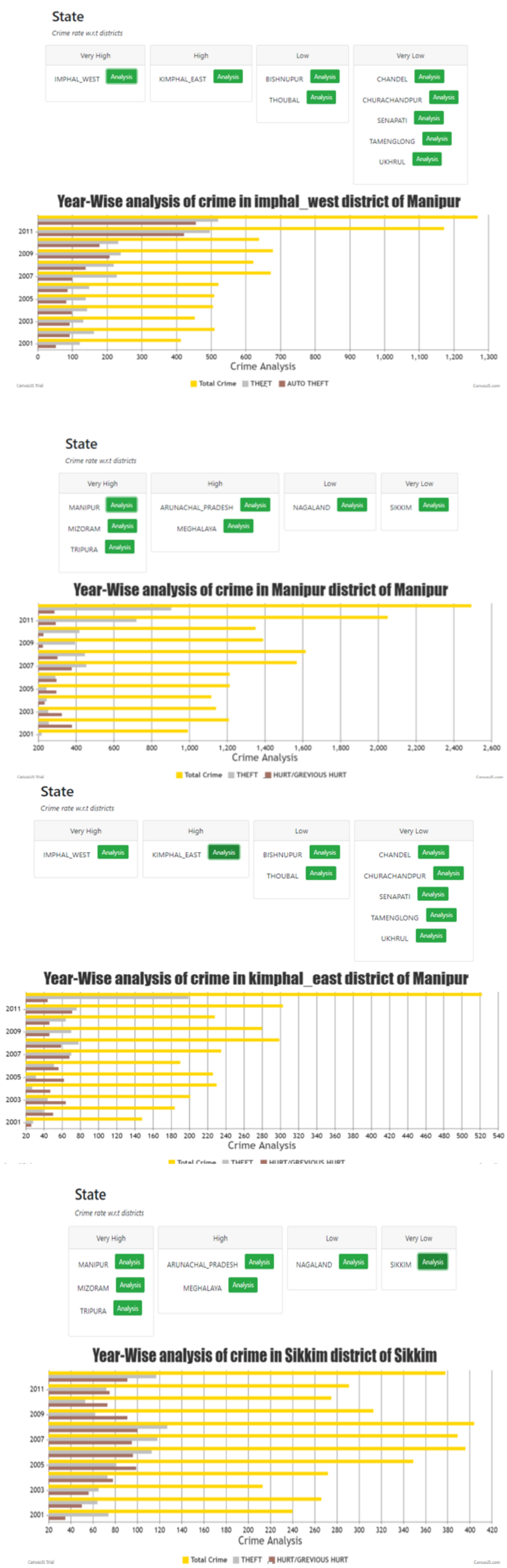


Figure 6: Shows the sample crime rates in different districts across the states

V. CONCLUSION

This study provides a method for anticipating and predicting crimes and fraud inside a district. It emphasizes the utility of a crime prediction tool for law enforcement. This research aims to improve the accuracy of predictions as much as feasible. For this, proposed combined solutions for forecasting crime trends in the districts of states are considered for analysis using the occurrences of crimes using the ARIMA model. This work employs an Auto ARIMA, an extended formulation approach, to produce projections for various periods. During the lockdown, we saw a decline in the majority of offences. Therefore, the experimental data reveal some intriguing crime trends. It is a significant finding that decreased outside activities affected the crime trend. Crime analysts may still use these findings to guide individuals in taking the steps required for crime prevention. The analysis results show a significant drop in the crimes over the period and also derived the crime rates of different districts of the states. It is thought that more accuracy may be attained by using more feature engineering in the address field. This will enable law enforcement organizations to handle the resources in a given region rapidly.

References

[1] Seidensticker, K.; Bode, F. Good policing in times of abstract police. In The Abstract Police; Terpstra, J., Salet, R., Fyfe, N., Eds.; Eleven International Publishing: Den Haag, The Netherlands, 2022; pp. 169–182.

[2] Fotheringham, A.S.; Brunsdon, C.; Charlton, M. Geographically Weighted Regression. The Analysis of Spatially Varying Relationships; John Wiley & Sons Ltd.: West Sussex, UK, 2002.

[3] Ratcliffe, J. Crime mapping: Spatial and temporal challenges. In Handbook of Quantitative Criminology; Piquero, A.R., Weisburd, D., Eds.; Springer: New York, NY, USA, 2010; pp. 5–24.

[4] Lin, Ying-Lung, Tenge-Yang Chen, and Liang-Chih Yu. "Using machine learning to assist crime prevention." In 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 1029-1030.IEEE, 2017.

[5] Kim, Suhong, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri, "Crime Analysis Through Machine Learning." In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 415-420. IEEE, 2018.

[6] Williams, Matthew L., Pete Burnap, and Luke Sloan. "Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns." The British Journal of Criminology 57, no. 2 (2017): 320-340.

[7] Nancy G. LaVigne Cynthia A. Mamalian and the staff of the Crime Mapping Research Center. The Use of Computerized Crime Mapping by Law Enforcement: Survey Results. Technical Report. U.S. Department of Justice.

[8] Alexander Hirschfield and Kate Bowers. 2001. Mapping and Analysing Crime Data: Lessons from Research and Practice.

[9] A M Olligschlaeger. Artificial Neural Networks and Crime Mapping. *Criminal Justice Press/Willow Tree Press, pages* 313 - 347*,* 20531 (1997)*.*

[10] Spencer Chainey and Jerry Ratcliffe. 2013. *GIS and crime mapping*. John Wiley & Sons.

[11] David M. Kennedy, Anthony A. Braga, Anne M. Piehl, and David M. Kennedy Et Al. 1997. The (un)known universe: Mapping gangs and gang violence in Boston. https://citeseerx.ist.psu.edu/viewdoc/similar?doi=10.1.1.483.6248&type=ab.
[12] S. D. Johnson, K. Bowers, and A. Hirschfield. 1997. NEW INSIGHTS INTO THE SPATIAL AND TEMPORAL DISTRIBUTION OF REPEAT VICTIMIZATION. *The British Journal of Criminology* 37, 2 (1997), 224–241. http://www.jstor.org/stable/23638645.
[13] GIS in Law Enforcement: Implementation Issues and Case Studies. https://www.crcpress.com/GIS-in-Law-Enforcement-Implementation-Issues-and-Case-Studies/Leipnik-Albert/p/book/9780415286107
[14] Bryan Hill. 2002. Operationalizing GIS to investigate serial robberies in Phoenix, Arizona.
[15] LaVigne Nancy G. Mamalian, Cynthia A. and Elizabeth Groff. 2008. *Use of Computerized Crime Mapping by Law Enforcement in the United States, 1997-1998*. Technical Report.
[16] David Weisburd and Cynthia Lum. 2005. The Diffusion of Computerized Crime Mapping in Policing: Linking Research and Practice. *Police Practice and Research* 6, 5 (2005), 419–434. DOI:http://dx.doi.org/10.1080/15614260500433004.
[17] T. A. Shen, W. Wang, and S. Chyou, *San francisco Crime Classification*, University of California, California, USA, 2015.
[18] Bode, F.; Stoffel, F.; Keim, D. *Variabilität und Validität von Qualitätsmetriken im Bereich von Predictive Policing*; KOPS: Konstanz, Germany, 2017.
[19] Muzammil Khan, Azmat Ali, Yasser Alharbi, "Predicting and Preventing Crime: A Crime Prediction Model Using San Francisco Crime Data by Classification Techniques", Complexity, vol. 2022, Article ID 4830411, 13 pages, 2022. https://doi.org/10.1155/2022/4830411.
[20] George O Mohler, Martin B Short, Sean Malinowski, Mark Johnson, George E Tita, Andrea L Bertozzi, and P Jeffrey Brantingham. 2015. Randomized controlled field trials of predictive policing. *Journal of the American statistical association* 110, 512 (2015), 1399–1411.
[21] Yifei Xue and D. E. Brown. 2003. A decision model for spatial site selection by criminals: a foundation for law enforcement decision support. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 33, 1 (Feb 2003), 78–85. DOI:http://dx.doi.org/10.1109/TSMCC.2003.809867.
[22] Jameson L. Toole, Nathan Eagle, and Joshua B. Plotkin. 2011. Spatiotemporal Correlations in Criminal Offense Records. *ACM Trans. Intell. Syst. Technol.* 2, 4, Article 38 (July 2011), 18 pages. DOI:http://dx.doi.org/10.1145/1989734.1989742.
[23] Spencer Chainey, Lisa Tompson, and Sebastian Uhlig. 2008. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Security Journal* 21, 1-2 (Feb. 2008), 4–28. DOI:http://dx.doi.org/10.1057/palgrave.sj.8350066.
[24] J. Eck, S. Chainey, J. Cameron, and R. Wilson. 2005. *Mapping crime: Understanding Hotspots*. Report. National Institute of Justice, Washington DC. 1–71 pages. http://discovery.ucl.ac.uk/11291/.
[25] Julia Dressel and Hany Farid. 2018. The accuracy, fairness, and limits of predicting recidivism. *Science Advances* 4, 1 (Jan. 2018). DOI:http://dx.doi.org/10.1126/sciadv.aao5580.

[26] M. A. Tayebi and U. Glasser. 2012. Investigating Organized Crime Groups: A Social Network Analysis Perspective. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 565–572. DOI:http://dx.doi.org/10.1109/ASONAM.2012.96.
[27] William C. Horrace and Shawn M. Rohlin. 2016. How Dark Is Dark? Bright Lights, Big City, Racial Profiling. *The Review of Economics and Statistics* 98, 2 (2016), 226–232. DOI:http://dx.doi.org/10.1162/REST_a_00543.
[28] Sharad Goel, Justin M. Rao, and Ravi Shroff. 2016. Precinct or prejudice? Understanding racial disparities in New York City's stop-and-frisk policy. *The Annals of Applied Statistics* 10, 1 (March 2016), 365–394. DOI:http://dx.doi.org/10.1214/15-AOAS897.
[29] Mona Lynch, Marisa Omori, Aaron Roussell, and Matthew Valasik. 2013. Policing the 'progressive'city: The racialized geography of drug law enforcement. *Theoretical criminology* 17, 3 (2013), 335–357.
[30] Justin Nix, Bradley A Campbell, Edward H Byers, and Geoffrey P Alpert. 2017. A bird's eye view of civilians killed by police in 2015: Further evidence of implicit bias. *Criminology & Public Policy* 16, 1 (2017), 309–340.
[31] E Ashby Plant and B Michelle Peruche. 2005. The consequences of race for police officers' responses to criminal suspects. *Psychological Science* 16, 3 (2005), 180–183.
[32] Ourania Kounadi, Kate Bowers, and Michael Leitner. 2014. Crime Mapping On-line: Public Perception of Privacy Issues. *European Journal on Criminal Policy and Research* 21 (06 2014), 167–190. DOI:http://dx.doi.org/10.1007/s10610-014-9248-4.
[33] Brian Jordan Jefferson. 2018. Predictable policing: Predictive crime mapping and geographies of policing and race. *Annals of the American Association of Geographers* 108, 1 (2018), 1–16.
[34] Andrew Guthrie Ferguson. 2012. Predictive policing and reasonable suspicion. *Emory LJ* 62 (2012), 259.
[35] Ourania Kounadi, Kate Bowers, and Michael Leitner. 2014. Crime Mapping On-line: Public Perception of Privacy Issues. *European Journal on Criminal Policy and Research* 21 (06 2014), 167–190. DOI:http://dx.doi.org/10.1007/s10610-014-9248-4.
[36] Jia Ma. 2013. *How the use of computerized crime mapping was affected by characteristics of law enforcement agencies in 2007*. Ph.D. Dissertation. https://lib.dr.iastate.edu/etd/13616