https://doi.org/10.34074/rere.00108

Article

REDUCING PROFILING BIAS IN CRIME RISK PREDICTION MODELS

Farhad Mehdipour, Wisanu Boonrat, April Love Naviza, Vimita Vidhya, & Marianne Cherrington

ABSTRACT

Crime risk prediction and predictive policing can lead to safer communities, by focusing on crime hotspots. Yet predictive tools should be reliable, and their outputs should be valid, especially across diverse cultures. Machine learning methods in policing systems are topical as they seem to be causing unintended consequences that exacerbate social injustice. Research into machine learning algorithm bias is prevalent, but bias, as it relates to predictive policing, is limited. In this paper, we summarise the findings of nascent scholarship on the topic of bias in predictive policing. The unique contribution of this paper is in the use of a typical police prediction modelling process to unpack how and why such bias can creep into algorithms that have high predictive accuracy. Our research finds that especially when resources are limited, trust in machine learning outputs is elevated; systemic bias of preceding assumptions may replicate. Recommendations include a call for human oversight in machine learning methods with sensitive applications such as automated crime prediction methods. Routine reviews of prediction outputs can ensure unwarranted community targeting is not magnified.

Keywords: predictive policing, ethnic identity, racial profiling, crime risk prediction, algorithm bias, demography

INTRODUCTION

Machine learning (ML) is now a prevalent predictive tool used in diverse applications, as a supportive mechanism in decision-making. Advantages include better accuracy in prediction, but also in realising trends, categorisations or clusters of information that can lead to new or hidden insights. Predictive policing (PP) refers to statistical, analytical or machine learning techniques meant to "identify likely targets for police intervention and prevent crime or solve past crime" (Perry, p. 154, 2013). In an era where resources are limited and organisations are called upon to do more with less, predictive policing applications have transformational merit, providing data-driven prepositioning of assets, better use of resources plus proactive tactical strategy and policy (Meijer, & Wessels, 2019). These predictive tools should support a more interconnected and safer community. Predictive policing and crime risk prediction can revolutionise policing (Egbert & Leese, 2021), but the media is rife with examples of excessive force alongside presumed ethnic or racial profiling. Are predictive tools at fault? Are they misdirecting efforts or escalating the use of force in confrontational situations? Critics and advocacy groups are raising concerns about racial justice; civil liberties concerns are mounting. The algorithms that drive predictive policing require more scrutiny and an evaluative focus on factors that contribute to biased outcomes (Brantingham, 2017); place-based prediction is under scrutiny.

In this paper, we will abridge topics relating to predictive policing and then crime factors, especially as they relate to data utilisation and algorithm design. Next, sources of bias will be investigated, using nascent peer-reviewed scholarly sources. An example of a process used to determine a crime risk prediction model will show that the goal of prediction accuracy can create outputs that have numerous sources of bias. The conclusions and recommendations will suggest methods for reducing bias in predictive ML models.

TOPICS IN PREDICTIVE POLICING

A thorough review of predictive policing literature to April 2017 was undertaken for the benefits and drawbacks of various methods (Meijer & Wessels, 2019). There was little empirical evidence of the benefits or drawbacks of methods; most studies were case driven or anecdotal. A search for peer-reviewed journals in the 41 months from 1 January 2017, was similarly conducted via Google Scholar. The number of articles multiplied almost 250% to 5950. Subsequently, only the first 15 Google Scholar pages (to April, 2021) revealed eight technical journal articles, briefed here, to compare and contrast, Meijer & Wessel findings. Table 1 has abridged results.

Table 1. Benefits and drawbacks of technical remedies to predictive policing bias.

	BENEFIT	DRAWBACK	REFERENCE	BIAS
1.	Empirical links to bias; methodology removes redundant loop.	Runaway feedback loops; call-backs to no incident locales.	(Ensign et al. 2018). Runaway feedback loops in predictive policing.	Bias crime rate (higher) in locales
2.	Random Forest, Neural Network, Kernel Support Vector Machine and Logistic Regression	Tested crime event prediction dynamic features efficacy.	(Rumi et al. 2018). Crime event prediction with dynamic features.	Adds human mobility data - social media.
3.	Random Forest and Logistic Regression Model were used.	Two separate models used to forecast two distinct crime types.	(Martegiani & Berrada, 2019). Crime Prediction Using Data Analytics: The City of Boston.	Predicts victim, place, offences and offenders
4.	Apriori algorithm was used to find patterns. Decision Tree classifier, Naïve Bayesian classifier	Spatio-temporal data in criminal hotspots using two different real-world data sets.	(Almanie, Mirza & Lor, 2015). Crime prediction based on crime types with spatial and temporal criminal hotspots.	Prediction for a particular location and specific time.
5.	K-Nearest Neighbour, Decision Tree, Multi-class Logistic Regression, Naïve Bayes, Random Forest.	Classification with log loss scoring and Naïve Bayes with parallel processing.	(Pradhan, 2018). Exploratory data analysis and crime prediction in San Francisco.	Attributes like seasons affect specific crimes.
6.	Random Forest and Decision Tree ML were used.	Ensemble methods: Extra Trees, Bagging and AdaBoost.	(Yuki et al. 2019). Predicting crime using time and location data.	Crime category predicted for time/ locale.
7.	Deep neural network (DNN) result: DNN model accurate in predicting crime occurrence than other predictions.	Feature-level data fusion method with an environmental context from multi-modal deep learning	(Kang & Kang, 2017). Crime occurrence prevention using crime prevention through environmental design (CPTED)	Broken Windows Theory: CPTED method boosts DNN design

A PROCESS TO DETERMINE A MODEL FOR PREDICTIVE POLICING

To build a predictive model, multiple analyses and experiments must be performed using an array of algorithms most suitable for the data type and context. The norm is to search for a predictive model with a high accuracy rate, but in application, predictive rate may not be the only measure to consider. It is prudent to build on previous research in the field. Using a typical police prediction model process, judgement points can be highlighted; they may introduce model bias (Bekmaganbet, 2021), (Miron et al. 2021).

1) Algorithms are plentiful; many are well known in terms of applicability, benefits or drawbacks. Algorithms are often based on statistical theory and variability and key assumptions are common to all such methods.
2) Data Transformation can be varied in nature. It is quite rare for data sets to be complete, clean, and usable. Most machine learning methods require data pre-preparation methods that vary widely and can be complex.
3) Crime Datasets are often used from open access websites. A New Zealand crime dataset is available from the New Zealand Police website with a good description and explanation (Victimisation time and place, 2018).
4) Descriptive Analysis is always required, to understand data, context, and basic statistical metrics, providing basic information about variables in a dataset and highlighting potential relationships between variables.
5) Baseline Feature and Target Variables must be determined and they will guide the analysis outputs. Domain expertise is helpful, so that the data analyst can support modelling with theoretical and applied performance.
6) Baseline Scores for variables are defined and ready to use for the model training so that scores can be compared and relate to the target label; this is important as it can affect model accuracy.
7) Feature Selection cuts dataset dimension and model complexity for faster training and meagre resource use while Invalid Data Detection is important as some records may be removed for various, valid reasons.
8) Relevant Factors must be chosen, based on any number of metrics, so the relevance to task aim is identified. Coding of variables may be carried out for software requirements; algorithms and equations may be formulated.
9) Predictive Modelling with different algorithms should now be instigated. Model prediction accuracy is typically the aim and cross-validation is undertaken on simulated data. Real-life data can be tested or compared.

At every stage, there are subjective decisions to be made, so partiality, preference or bias may be introduced. It is not enough to argue that the data analyst is data-focused and impartial. It is beneficial that data analysts have domain knowledge, but increasingly with automated systems, they may be the only arbiter of 'excellence' in terms of the context in which the algorithms are intended to be used (Cherrington et al. 2019b). Who knows how machine learning algorithms may be used subsequently, in the workplace?

Predictive policing can lead to life and death consequences (The Guardian, 2015). Seek and you will find. When police are sent to the same locations frequently, suspects will be apprehended with predictive policing systems (The Police Foundation, 2020). If predictive policing has reinforced bias the results are not just bothersome or unfortunate, they may lead to prison sentences, or even a death penalty (Richardson et al. 2019). Although determining bias can be hard to verify (Brantingham, 2017), racial bias has been shown to exist in the U.S.A. for capital cases and across multiple decision-making points which potentially shape the life course of defendants (Petersen, 2017).

OPPORTUNITIES FOR THE INTRODUCTION OF BIAS

A broad definition of bias is an inclination or prejudice for or against, especially in a way considered to be unfair Smith & Noble, 2014); statistical bias can be quantified if the expected value differs from the true estimate of the parameter. The type of data used for ML is the first potential source of partiality; the value of event-based predictive policing, which relies on actual data on crimes that have been committed, should not be ignored (Kirkpatrick, 2017). In terms of the step-wise process above, some specific types of bias were identified.

1) Algorithms are generally evaluated for predictive accuracy, only valid for the test data, 'model shrinkage' is inevitable when the algorithm is applied to new, unfamiliar data (Oswald & Babuta, 2019).
2) Data Transformation is usually limited to numeric data, which is a limitation. Transformation can involve re-labelling or perturbation, which may also introduce bias if one-to-one methods are not used (Bacelar, 2021).
3) Crime Datasets are historical. There is concern that prediction from 'stock data' can reinforce existing bias in policing systems and miss opportunities for new insights into future-based crimes (Sandhu & Fussey, 2020).
4) Descriptive Analysis can bias results if variables are compacted in dimension or if categorisations are utilised for expediency. Visualisations can be biased as they are often only in two or three dimensions (Huff, 1993).
5) Target Variables used will affect results. It is vital that data limitations are known and that the 'question' the algorithm is intending to solve is carefully understood, expressly when used in the real world (Van Brakel, 2016).
6) Baseline Scores for variables are defined and ready to use for the model training so that scores can be compared and relate to the target label; this is important as it can affect model accuracy (Wielenga, 2007).
7) Feature Selection necessarily biases results to improve performance and save resources (Cherrington, 2019c). The issue is that different algorithms will select a different set of features; domain expertise is required.
8) Relevant Factors and software limitations may necessitate the choice of sub-optimal data; insights from scholars who have faced risk assessment throughout the criminal justice system are helpful (Ferguson, 2016).
9) Predictive Models must be tested in practice (Cherrington, 2019c). Independent and methodologically sound trial evaluation is vital for predictive policing models, with event-based evaluation (Oswald & Babuta, 2019).

Predictive policing leads to policing, sentencing and criminal justice (Završnik, 2019). There has been visibility around policing and inequity that causes disparate impacts that exacerbate and prolong social injustice (Selbst, 2017). In critical professions like policing, it is important to understand how predictive models make outputs or decisions so that they are not 'unfair' (Martin, 2019). Are predictive policing models biased? Results show, not always (Brantingham et al. 2018).

Deep learning models with 'black box' outputs have high performance but can be difficult to assess in real-world situations and interpreting complicated models leads to over-reliance on ML systems (Cherrington et al. 2020a, b, c). Efficient procedures with high predictive performance can still escalate unintended behaviours in volatile policing circumstances; surveillance systems are creating reams of evidence (Patil & Bernstein, 2021).

Predictive policing technologies have pros and cons (Martin, 2019); the intention is to target crime hotspots and recidivist offenders (Selbst, 2017). That does not mean that systems can be used 'as is' and unchecked. More evaluation is needed to ensure data is managed as a beneficial asset (Cherrington et al. 2021a, b).

CONCLUSIONS

For predictive policing models, it is important to clarify the purpose of the model and to understand how this is to be used in the police system; this can be very regional in context. Models, therefore, might not transfer well or may have to be refined once again when the context changes. Time may be a contextual factor.

It is important to note what does work and does not work well. It may be that different algorithms can be employed in different contexts. Cross-validation, confusion matrices, and visualisation can test model results.

There can be improved accuracy with deep learning models; the 'black-box' nature of these models may make them unsuitable for some predictive policing objectives. Visualisation techniques can be used in conjunction with models, however, visualisation can both illuminate and obscure information and can be a source of 'bias'.

In this paper predictive policing models were summarised for benefit, drawbacks, and bias; technical journal sources were favoured in the analysis. Several scholarly sources were reviewed to extend a previous scoping review (Meijer & Wessel, 2019) and similar issues were found. Nascent research algorithms were summarised for method and contribution. A process for crime risk detection was provided.

The research finds that police predictive models are meant to support the police by directing them to where they are likely to be needed – wise use of resources. As ML systems become more automated, this elevates trust in the system outputs, often over human decision-making. Given that potential consequences can be life-threatening, it is important that bias in models is understood, and that human checks and balances are always in place. At the very least, routine reviews and audits of prediction policing should be mandatory.

There is considerable scope for future work in predictive policing, especially empirical research and in improvements in ML algorithm design. This is a vital and emotive topic and seasoned data analysts must support scholarship in this field.

Dr Farhad Mehdipour is an academic and Research & Development (R & D) expert with over 20 years' experience both in industry and academia. Farhad has initiated and led several interdisciplinary R & D projects and published 100+ peer-reviewed articles. Farhad is currently a Head of Department, and a Principal Lecturer at Otago Polytechnic Auckland International Campus, and an Adjunct Professor at St Bonaventure University, United States of America (USA). He is a senior member of IEEE. **ORCID: 0000-0002-0357-6182**

Wisanu Boonrat is an Information Technology graduate student from Otago Polytechnic Auckland International Campus. Experienced in web, android, and machine learning development, he has created a wide range of tools for a variety of clients. He is now working in the industry as a Software Engineer at Medi-Map.

April Love Naviza is an Otago Polytechnic Auckland International Campus graduate student in Information Technology with more than ten years of experience in Information Technology (IT) Service Management and currently working in a global Information and Computer Technologies (ICT) company. Most recently, her collaborative research has been applied to algorithms for crime prediction in New Zealand and carbon emissions reduction strategies using technological solutions. **ORCID: 0000-0003- 4912-834X**

Vimitaben Vaidya is a graduate student in Information Technology at Otago Polytechnic Auckland International Campus, with two years of software development experience. Her expertise in programming and collaborative research has been applied to a machine learning-based crime prediction model for New Zealand. Vimitaben is currently working in the IT industry as a software developer.

Marianne Cherrington is a Principal Lecturer at Otago Polytechnic Auckland International Campus with a focus on sustainability and computer science and analytics. A lecturer in disruptive innovation, her research into machine learning feature selection algorithms applies in many fields, producing interesting collaborations with local and international partners in many disciplines and sectors. **ORCID: 0000-0002-1240-2010**

REFERENCES

01 Akpinar, N. J., De-Arteaga, M., & Chouldechova, A. (2021). Effect of differential victim crime reporting on predictive policing systems. In Pro of 2021 ACM Conference on Fairness, *Accountability and Transparency* (p. 838-849).

02 Almanie, T., Mirza, R., & Lor, E. (2015). Crime prediction based on crime types and using spatial and temporal criminal hotspots. *arXiv preprint arXiv:1508.02050*.

03 Almaw, A., & Kadam, K. (2018). Crime Data Analysis and Prediction Using Ensemble Learning. 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS).

04 Bacelar, M. (2021). Monitoring bias and fairness in machine learning models: A review. *ScienceOpen Preprints*.

05 Bekmaganbet, G. (2021). Crime Prediction and Forecasting: Feature Selection and Vulnerable Region Detection Models.

06 Borges, J., Ziehr, D., Beigl, M., Cacho, N., Martins, A., Araujo, A., Bezerra, L. and Geisler, S. (2018). Time series features for predictive policing. In *2018 International smart cities conference* (ISC2) IEEE.

07 Brantingham, P. J. (2017). The logic of data bias and its impact on place-based predictive policing. *Ohio State. Journal of Criminal Law.*, 15, 473.

08 Brantingham, P. J., Valasik, M., & Mohler, G. O. (2018). Does predictive policing lead to biased arrests? Results from a randomized controlled trial. *Statistics and public policy*, 5(1), 1-6.

09 Cherrington, M., Airehrour, D., Lu, J., Thabtah, F., Xu, Q., & Madanian, S. (2019a). Particle swarm optimization for feature selection: A review of filter-based classification to identify challenges and opportunities. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 0523-0529). IEEE.

10 Cherrington, M., Airehrour, D., Lu, J., Xu, Q., Wade, S. and Madanian, S. (2019b) "Feature Selection Methods for Linked Data Limitations, Capabilities and Potentials," in proc. 2019 IEEE/ACM 6th International Conference on Big Data Computing Applications and Technologies (BDCAT) IEEE.

11 Cherrington, M., Lu, J., Airehrour, D., Thabtah, F., Xu, Q., and Madanian, S. (2019c) "Feature Selection: Multisource and Multi-view Data Limitations, Capabilities and Potentials," in proc. *29th International Telecommunication Networks and Applications Conference (ITNAC)*.

12 Cherrington, M., Thabtah, F., Lu, J., & Xu, Q. (2019d). Feature selection: filter methods performance challenges. In *2019 Int. Conference on Computer and Information Sciences (ICCIS)* (pp. 1-4). IEEE.

13 Cherrington, M., Airehrour, D., Lu, J., Xu, Q., Cameron-Brown, D., & Dunn, I. (2020a, November). Features of Human-Centred Algorithm Design. In *2020 30th International Telecommunication Networks and Applications Conference (ITNAC)* (pp. 1-6). IEEE.

14 Cherrington, M., Airehrour, D., Lu, J., Xu, Q., Wade, S., & Dunn, I. (2020b, November). Indigenous Big Data Implications in New Zealand. In *2020 30th International Telecommunication Networks and Applications Conference (ITNAC)* (pp. 1-6). IEEE.

15 Cherrington, M., Lu, Z. J., Xu, Q., Airehrour, D., Madanian, S., & Dyrkacz, A. (2020c). Deep learning decision support for sustainable asset management. In *Advances in Asset Management and Condition Monitoring* (pp. 537-547). Springer, Cham.

16 Cherrington, M., Lu, Z. J., Xu, Q., Thabtah, F., Airehrour, D., & Madanian, S. (2020d). Digital Asset Management:

New Opportunities from High Dimensional Data—A New Zealand Perspective. In *Advances in Asset Management and Condition Monitoring* (pp. 183-193). Springer, Cham.

17 Cherrington, M., Lu, J., Xu, Q., Airehrour, D., & Wade, S. (2021a). Deep learning for sustainable asset management decision-making. *International Journal of COMADEM*, 24(2), 35-41.

18 Cherrington, M., Lu, J., Xu, Q., Airehrour, D., & Wade, S. (2021b). The digital asset management microcosm: a high-dimensional New Zealand view. *International Journal of COMADEM*, 24(2), 21-27.

19 The Police Foundation. (2020, April 20). Data driven policing: Holy Grail or death knell of policing by consent? https://www.police-foundation.org.uk/2018/11/data-driven-policing-holy-grail-or-death-knell-of-policing-by-consent/

20 Degeling, M., & Berendt, B. (2018). What is wrong about Robocops as consultants? A technology-centric critique of predictive policing. Ai & Society, 33(3), 347-356.

21 Egbert, S., & Leese, M. (2021). Criminal Futures: Predictive Policing and Everyday Police Work (p. 242). Taylor & Francis.

22 Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018, Jan.). Runaway feedback loops in predictive policing. In *Conference on Fairness, Accountability and Transparency PMLR*.

23 Ferguson, A. G. (2016). Policing predictive policing. *Wash. UL Rev.*, 94, 1109.

24 Huff, D. (1993). *How to lie with statistics*. WW Norton & Company.

25 Kang, H. W., & Kang, H. B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. *PloS one*, 12(4), e0176244.

26 Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology*, 59(3), 674-692.

27 Kirkpatrick, K. (2017). It's not the algorithm, it's the data. *Communications of the ACM*, 60(2), 21-23.

28 Mann, A. (2017). How Science Is Helping Stop Crime Before It Occurs. URL: https://www. nbcnews. com/mach/science/how-science-helping-stop-crime-itoccurs-ncna805176.

29 Martegiani, G., & Berrada, L. (2019). Crime Prediction Using Data Analytics: The Case of the City of Boston www.academia.edu/25277353/Crime_Prediction_Using_Data_Analytics_the_Case_of_the_City_of_Boston

30 Martin, K. E. (2019). Designing ethical algorithms. *MIS Quarterly Executive*. June.

31 Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. International Journal of Public Administration, 42(12), 1031-1039.

32 Miron, M., Tolan, S., Gómez, E., & Castillo, C. (2021). Evaluating causes of algorithmic bias in juvenile criminal recidivism. Artificial Intelligence and Law, 29(2), 111-147.

33 Mohler, G., Raje, R., Carter, J., Valasik, M., & Brantingham, J. (2018). A penalized likelihood method for balancing accuracy and fairness in predictive policing. In *2018 international conference on systems, man, and cybernetics (SMC)* (pp. 2454-2459). IEEE.

34 O'Neill, R. NZ Police widen use of Auror crime-fighting tools [Internet]. ZDNet. ZDNet; 2016 [cited 2020Jun25]. From: www.zdnet.com/article/nz-police-widen-use-of-auror-crime-fighting-tools

35 Oswald, M., & Babuta, A. (2019). Data analytics and algorithmic bias in policing.

36 Patil, S. V., & Bernstein, E. S. (2021). Uncovering the Mitigating Psychological Response to Monitoring Technologies: Police Body Cameras Not Only Constrain but Also Depolarize. *Organization Science*.

37 Perry, W. L. (2013). Predictive policing: The role of crime forecasting in law enforcement operations. Rand Corporation.

38 Petersen, N. (2017). Examining the sources of racial bias in potentially capital cases: A case study of police and prosecutorial discretion. Race and Justice, 7(1), 7-34.

39 Pradhan, I. (2018). Exploratory data analysis and crime prediction in San Francisco.

40 Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *NYUL Rev. Online*, 94, 15.

41 Rummens, A., & Hardyns, W. (2021). The effect of spatiotemporal resolution on predictive policing model performance. *International Journal of Forecasting*, 37(1), 125-133.

42 Rumi, S.K., Deng, K., & Salim, F.D. (2018). Crime event prediction with dynamic features. EPJ Data Science. 2018; 7(1).

43 Sandhu, A., & Fussey, P. (2020). The 'uberization of policing'? How police negotiate and operationalise predictive policing technology. *Policing and Society*, 1-16.

44 Schneider, S. (2002). Predicting Crime: The Review of Research. Department of Justice Canada.

45 Selbst, A.D. (2017). Disparate impact in big data policing. *Ga. L. Rev.*, 52, 109.

46 Shapiro, A. (2019). Predictive policing for reform? Indeterminacy and intervention in big data policing. *Surveillance & Society*, 17(3/4), 456-472.

47 Smith, J., & Noble, H. (2014). Bias in research. Evidence-based nursing, 17(4), 100-101.

48 The Guardian. (2015). The Counted: People Killed in the us by the Police, available at www.theguardian.com/us-news/ng-interactive/2015/jun/01/the-counted-police-killings-us-database [2 September 2015].

49 The Police Foundation. (2020, April 20). Data driven policing: *Holy Grail or death knell of policing by consent?* https://www.police-foundation.org.uk/2018/11/data-driven-policing-holy-grail-or-death-knell-of-policing-by-consent/

50 Van Brakel, R. (2016). Pre-emptive big data surveillance and its (dis) empowering consequences: The case of predictive policing. *pp. in*, 117-141.

51 *Victimisation time and place*. (2018). New Zealand Police. https://www.police.govt.nz/about-us/publications-statistics/data-and-statistics/policedatanz/victimisation-time-and-place

52 Wielenga, D. (2007). Identifying and overcoming common data mining mistakes. In SAS Global Forum (pp. 1-20).

53 Završnik, A. (2019). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*, 1477370819876762.

54 Yuki, J. Q., Sakib, M. M., Q., Zamal, Z., Habibullah, K. M., & Das., A. K. (2019, July). Predicting crime using time and location data. In Proceedings of the 2019 7th International Conference on Computer and Communications Management (pp. 124-128).