Machine learning as a tool for police to reduce crime

Introduction

Costs as a consequence of crime in England and Wales in 2015/16 was £40bn, whilst costs in response to crime, such as policing and the criminal justice system, was £13.5bn from a total of 16.8m crimes. (Heeks et al., 2018). In that same year, there were 4.5m crimes reported to police (Office for National Statistics, 2016). The latest figures show 6.7 million crimes reported to police (Office for National Statistics, 2023). Whilst for some offences trends in police-reported crime "do not provide reliable trends in crime" (Office for National Statistics, 2023: 6), the figures clearly illustrate an upward trend leading to increasing costs to society. Given the cost constraints on public service, it is therefore asserted that cost-effectively reducing crime would have a positive social and economic impact on society.

Crime is committed by a person or persons, in a place or places. A machine learning model to predict crime could predict if a person is likely to commit a crime, or if a place is likely to be the location of a crime. Crime is often committed in the same location "because criminals tend to operate in their comfort zone" (Perry et al., 2013: 2) so this review will focus on the second scenario only; is it possible to use machine learning to predict where a crime will take place?

The research questions are:

- Which machine learning models have been used to predict crime locations?
- Which machine learning models provide actionable insights for day-to-day police work?

- What would be the ethical challenges, if any, with using the insights in active policing?

Machine learning models for crime prediction

Machine learning includes unsupervised learning, where patterns in data are learnt and clustered into sets with similar characteristics, and supervised learning where input data are labelled and the characteristics of the data are learnt to predict the label of a new unlabelled input (Russell & Norvig, 2021). Both unsupervised learning (clustering) and supervised learning (classification) algorithms will be explored for suitability for assisting law enforcement with crime prediction.

Unsupervised Learning – Clustering

Clustering is a popular form of unsupervised learning that groups unlabelled data into clusters that share similar features (Miroslav, 2021). Crimes can be clustered by type and location using K-Means to show the density of occurrences of a crime type within close proximity to each other (Jain et al., 2017). This could be a useful tool to help direct police resources to reduce specific criminal activities at specific locations, however, it is not a particularly sophisticated predictor of future crime so clustering will not be examined further.

Supervised Learning – Classification

Simple classification models have been built to predict the rate of violent crime as high, medium or low based on features such as average income, poverty and education (Iqbal et al., 2013; Zaidi et al., 2020). When predicting the crime level for a state in the United States of America (USA), Decision Tree (DT) performed better than Naïve Bayes (NB) (Iqbal et al., 2013) and Random Forest (RF) performed better than Support Vector Machine (SVM) (Zaidi et al., 2020). Both studies used the same dataset and similar features, so whilst a direct comparison between them carries some risk, there is a clear indication that both RF and SVM performed significantly better than DT and NB. However, Ahishakiye et al. (2017) successfully used DT to predict the crime level for a county in the USA using the same dataset with better accuracy than the DT in Zaidi et al. (2020), which is surprising because one might expect accuracy to decrease with greater granularity.

These models would be of limited use to policing because the predictions at state level and even at county level lack useful granularity. However, DT is an easily explainable algorithm so for those models it might be useful to work back and discover the most important features to take broader political interventions to reduce crime similar to those proposed earlier for clustering algorithms.

Safat et al. (2021) tested eight classification algorithms on Chicago and Los Angeles crime datasets, using features common to both, to predict the rates of different crimes in districts within each city. XGBoost performed best on Chicago and was only slightly behind K Nearest Neighbour (KNN) on Los Angeles. They went on to use Autoregressive Integrated Moving Average (ARIMA) to produce a time series forecasting model to predict crime levels over time for each city and within district hot spots within each city. This kind of model could be used to assist with policing strategies such as funding and recruitment to deal with emerging hotspots. However, the outputs looked suspiciously uniform, probably because ARIMA is “good for short-term forecasting [but] not built for long-term forecasting” (Hayes, 2023), so whilst it might be useful to inform short-term strategies, the decreasing accuracy over time would make it less useful for medium to long-term planning.

Hybrid – Clustering and Classification

Some researchers have used a combination of clustering and classification algorithms for more sophisticated predictive solutions.

KNN was used to classify crimes then K-Means, Agglomerative Clustering and DBSCAN were compared for clustering crimes of the same category (Sivaranjani et al., 2016). DBSCAN performed best with a precision of 0.95, recall of 0.91 and f-score of 0.93. The output assigned each city to a cluster, with a high, medium or low value for each crime type. This could be useful for macro-level policing policy such as the distribution of funds and police specialists, but it is not considered a useful crime prediction tool to assist with day-to-day policing because even predictions at city level lack sufficient granularity.

Hajela et al. (2020) reverse the approach by clustering first, using K-Means, to identify crime hot spots, and then classify each instance using NB and DT to determine the prevalent crime type at that location. If the location belongs to a hot spot the prediction is high likelihood of the classified crime type happening there. An advantage of this approach over Sivaranjani et al. (2016) is coordinates are used for the initial clusters so arbitrary city boundaries are not used, making clusters potentially more useful. However the classification of each location results in a single crime class, which might help to prioritise certain policing activities, but it would potentially miss other significant predicted criminal activities.

Mapping in grids

Dividing an area into a grid to predict crime within cells in the grid has been successfully tested (Cichosz, 2020; Mohler, 2014; Rummens et al., 2017; Stalidis et al., 2021). The size of the grid is important because bigger cells make the predictions less useful, but smaller cells reduce the features per cell and hence the accuracy drops (Stalidis et al., 2021).

Cichosz (2020) divided a city into a grid of 300m^2 cells, mapped criminal activity and points of interest in each cell, and used classification algorithms to predict crime hotspots in 4 districts with an AUC of circa 0.85 across all districts. Better yet, when a model trained on one district was applied to another the AUC only reduced by 0.05-0.1, suggesting high model transferability (Cichosz, 2020). Four classifier algorithms were tested with RF performing best overall. This model could be used to direct police to crime hotspots based on the nearby POIs. The potential weakness, however, is that the hotpots are 300m^2 big, so identification of a specific criminal location is still vague. The size of the grid cells could be reduced, but that would reduce the POIs in each grid, potentially reducing the accuracy.

Rummens et al. (2017) mapped demographic and environmental variables to 200m^2 cells along with criminal activity divided into day or night over bi-weekly and monthly periods. Logistic Regression (LR) and Artificial Neural Network (ANN) were used to predict crime using a rolling time window of the preceding three years. Precision was used to measure the performance of the models to reflect the “need for efficiency” Rummens et al. (2017: 258), because it penalises false positives which would be a waste of police time. Bi-weekly predictions were circa 0.3 precision and monthly circa 0.55 precision, with similar performance between LR and ANN.

An interesting idea to map crimes using four dimensions; longitude, latitude, time and type of crime was proposed by Stalidis et al. (2021). A city is split into a grid and each grid cell plots crimes in a series of time windows. The cells, output as images, are passed through a Convolutional Neural Network (CNN) to create a feature vector for each cell/timespan instance, which are then fed into a Long-Short Term Memory (LSTM) network to extract feature vectors. “LSTM is more robust to degradation of gradient than the standard [Recurrent Neural Network] RNN” (Noh, 2021: 10), hence using it in preference to an RNN for time series modelling. The model, named Spatial features First Then Temporal (SFTT), was tested on crime datasets for five cities in North America and compared with state-of-the-art classifier algorithms such as DT, Naïve Bayes (NB), SVM and KNN. SFTT outperformed the other algorithms with consistently excellent results in predicting emerging hotspots of crime. The biggest drawback is the high computational effort; between 17 minutes and 1 hour 53 minutes to train the SFTT vs. 2 seconds to train the DT, which still gave reasonable results.

Table 1 summarises the papers in the order discussed in this literature review, including the various machine learning models and their relevance and limitations for day-to-day policing. The increasing usefulness from clustering through increasingly granular classification and finally grid-based classification can be seen.

Reference	Machine learning model(s)	Solution overview	Granularity	Relevence to day-to-day policing	Limitations
Jain et al. (2017)	K-Means	Cluster by crime type and location	Variable; typically a few streets	Low	Not a predictor. More appropriate for strategic resource planning
Iqbal et al., (2013)	Decision Tree Naïve Bayes	Demographic data to predict if a state has high, medium or low violent crime	State	Low	Predictions at state level so limited use in day-to-day policing
Zaidi et al., (2020)	Random Forest Support Vector Machine	Demographic data to predict if a state has high, medium or low violent crime	State	Low	Predictions at state level so limited use in day-to-day policing
Ahishakiye et al. (2017)	Decision Tree	Demographic data to predict if a county has high, medium or low violent crime	County	Low	Predictions at county level so limited use in day-to-day policing
Safat et al. (2021)	Logistic regression Decision Tree Random Forest Multilayer Perceptron Naïve Bayes Suport Vector Machine GXBoost K Nearest Neigbour	Compare classifier algorithms using two datasets then test time series prediction using ARIMA	District	Potential to inform short-term resource planning	Longer-term accuracy is questionable
Sivaranjani et al. (2016)	K Nearest Neighbour K-Means Agglomerative Clustering DBSCAN	Classify crimes with KNN then cluster crimes of the same category to give each city a high/medium/low preopensity per crime type	City	Potentially useful for city-level planning	Lacks granularity to inform day-to-day policing
Hajela et al. (2020)	K-Means Naïve Bayes Decision Tree	Cluster crime hot spots then classify crime type at each hot spot	Variable based on cluster size; typicaly district but centre point is visible	Hot spots are based upon coordinates, so potential for targetted interventions	Each cluster only has one class, so intervetions would be based upon most prevelent crime only
Cichosz (2020)	Logistic Regression Support Vector Machine Decision Tree Random Forest	Points of interest (POI) aggregated into cells in a geographic grid to predict hotpots of different crime types based upon POI atributes. Performed well when trained on one area and tested on another	300m²	Predicting crime types based upon POI could enable targetted policing	The grid doesn't give a precise prediction of where crime will occur
Rummens et al. (2017)	Logistic Regression Artificial Neural Network	Hotspots by crime type plotted in a grid, split by day and night. Predictions using rolling time window	200m²	Predictive patterns by day and night could help with proactive policing	Only three crime types, but more could be added
Stalidis et al. (2021)	CNN LSTM	Time series maps with crime plots passed through CNN to create feature vectors then LTSM to extract feature vectors to predict emerging crime hotspots	500m²	Predicted emerging crime hotspots could be used to target police activity	Reliatively high compute required in pre-processing the maps and running the algorithms

Table 1. Machine learning models for crime prediction

Ethical considerations

The examination of models to predict the locations of crime rather than the perpetrators of crime removed some ethical considerations because personal data (Information Commissioner's Office, N.D.) is not used in any of the models examined. Ethics must still be considered, however. The results of machine learning models could be used to influence policing which could have implications for the public, such as unnecessary and intimidating police presence in "high-risk" areas, or failure to respond to criminal activities in "low-risk" areas because resources were diverted elsewhere. It would therefore be important to be able to explain how the models made their recommendations.

Rather than try to explain black box machine learning models such as SVM and ANN, it is preferable to use interpretable models such as DT instead, especially when the results are similar (Rudin, 2019). This provides a much more transparent explanation of the output of the model, and transparency is a key enabler for ethical machine learning models (Walmsley, 2021). Unfortunately, black box models generally performed better than interpretable models (Cichosz, 2020; Rummens et al., 2017), or only black box models were used (Sivaranjani et al, 2016; Stalidis et al., 2021; Zaidi et al, 2020) and this is backed up by a thorough literature review on a similar topic (Jenga et al., 2023). A useful future study would be to give interpretability a weighting value and assessing models on performance and interpretability combined. This would indirectly include computational overhead too, since black box models are generally more complex and require more compute resources to train (Stalidis et al., 2021). All of the studies used publicly available data so reproducing and comparing them should be relatively simple. Where possible the same dataset should be used to ensure a fair comparison. This will not be possible in all cases because some models used different features, so there will still be some limitations with the comparison that could only be overcome with a single superset of data from which all features for all models could be extracted. This does not exist in the public domain.

Conclusions

There are many ways to predict crime with varying degrees of usefulness. What is clear from this literature review is that the measures of model performance such as accuracy, precision, recall and AUC are not the most important measures for the usefulness of the model because they don’t describe how the results would be used.

Clustering in isolation provides very little benefit to day-to-day policing. Classification at state and county levels is also of limited benefit.

Classification becomes more useful when combined with clustering. Classify then cluster had poor granularity when clustering at city level, but it differentiated between crime types. Cluster then classify was more granular, but only gave one crime class per cluster.

Promising studies were found using grids of varying sizes and features. Demographic and environmental features gave reasonable metrics to predict crime in a rolling window. Here, lower model performance scores are outweighed by the increased usefulness of the output, which would be more actionable to police forces. Another grid-based solution using points of interest as a feature was highly transferable, with the potential to train on one area and deploy to multiple, making adoption simpler.

Interpretability is an important ethical factor barely mentioned in any of the papers. Being able to explain the output would make public acceptance easier, so a further study is recommended to compare the more useful models using the same datasets where possible, and the same performance metrics with the results weighted based on the interpretability of the model to find the best balance between performance and interpretability.

Word count: 2,199

References

Ahishakiye, E., Taremwa, D., Omulo, E.O. & Niyonzima, I. (2017) Crime prediction using decision tree (J48) classification algorithm. *International Journal of Computer and Information Technology*, 6(3): 188-195.

Cichosz, P. (2020) Urban crime risk prediction using point of interest data. *ISPRS International Journal of Geo-Information*, 9(7): 459. DOI: https://doi.org/10.3390/ijgi9070459

Hajela, G., Chawla, M. & Rasool, A. (2020) A clustering based hotspot identification approach for crime prediction. *Procedia Computer Science*, 167: 1462-1470. DOI: https://doi.org/10.1016/j.procs.2020.03.357

Hayes, A. (2023) Autoregressive Integrated Moving Average (ARIMA) Prediction Model. Available from: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp [Accessed 01 December 2023].

Heeks, M., Reed, S., Tafsiri, M., & Prince, S. (2018) *The economic and social costs of crime: Second edition*. London: Home Office

Information Commissioner's Office (N.D.) What is personal information: a guide. Available from: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/personal-information-what-is-it/what-is-personal-information-a-guide [Accessed 22 December 2023].

Iqbal, R., Murad, M.A.A., Mustapha, A., Panahy, P.H.S. & Khanahmadliravi, N. (2013) An experimental study of classification algorithms for crime prediction. *Indian*

Journal of Science and Technology, 6(3): 4219-4225. DOI: https://dx.doi.org/10.17485/ijst/2013/v6i3.6

Jain, V., Sharma, Y., Bhatia, A. & Arora, V. (2017) Crime prediction using K-means algorithm. *GRD Journals-Global Research and Development Journal for Engineering*, 2(5): 206-209.

Jenga, K., Catal, C. & Kar, G. (2023) Machine learning in crime prediction. *Journal of Ambient Intelligence and Humanized Computing*, 14(3): 2887-2913.

Miroslav, K. (2021) *An Introduction to Machine Learning*. 3rd Ed. Cham: Springer.

Mohler, G. (2014) Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting*, 30(3): 491-497. DOI: https://doi.org/10.1016/j.ijforecast.2014.01.004

Noh, S.H. (2021) Analysis of gradient vanishing of RNNs and performance comparison. *Information*, 12(11): 442. DOI: https://doi.org/10.3390/info12110442

Office for National Statistics (2016) *Crime in England and Wales: year ending Mar 2016*. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingmar2016/pdf [Accessed 28 November 2023].

Office for National Statistics (2023) *Crime in England and Wales: year ending June 2023*. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingjune2023/pdf [Accessed 28 November 2023].

Perry, W.L., McInnis, B., Price, C.C., Smith, S. & Hollywood, J.S. (2013) Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations. Santa Monica: RAND Corporation.

Rudin, C. (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5): 206-215.

Rummens, A., Hardyns, W. & Pauwels, L. (2017) The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. *Applied geography*, 86: 255-261. DOI: https://doi.org/10.1016/j.apgeog.2017.06.011

Russell, S. & Norvig, P. (2021) *Artificial Intelligence: A Modern Approach*. 4th ed. Harlow: Pearson Education Limited.

Safat, W., Asghar, S. & Gillani, S.A. (2021) Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. *IEEE access*, 9: 70080-70094. DOI: https://doi.org/10.1109/ACCESS.2021.3078117

Sivaranjani, S., Sivakumari, S. & Aasha, M. (2016) ‘Crime prediction and forecasting in Tamilnadu using clustering approaches’. *2016 International Conference on Emerging Technological Trends.* Kollam, India, 21-22 October. IEEE. 1-6.

Stalidis, P., Semertzidis, T. & Daras, P. (2021) Examining deep learning architectures for crime classification and prediction. *Forecasting*, 3(4): 741-762. DOI: https://doi.org/10.3390/forecast3040046

Walmsley, J. (2021) Artificial intelligence and the value of transparency. *AI & Society*, 36(2): 585-595.

Zaidi, N.A.S., Mustapha, A., Mostafa, S.A. & Razali, M.N. (2020) ‘A classification approach for crime prediction’. *Applied Computing to Support Industry: Innovation and Technology.* Ramadi, Iraq,15–16 September. Switzerland: Springer International Publishing. 68-78.