Crime Analysis

Karl Attard
karl.attard.19@um.edu.mt
University of Malta

Abstract

A concern amongst every citizen in the world is their own safety. One of the major factors that affect citizens' perceived safety is the level of crime around them and how it fluctuates along the years. This research focuses on using data mining techniques within the crime analysis domain. We focus on three main objectives; predicting the type of crime, predicting the future crime rate and the investigation of how much data is required to train our models. Two distinct American public datasets are chosen for this study, one focusing on the city of San Francisco (SF) and the other on New York City (NYC). A comparative analysis is performed on other published studies where we have successfully managed to improve on their results.

Keywords: Data Mining, Machine Learning, Artificial Intelligence, Crime Prediction, NYC Dataset, SF Dataset

1 Introduction

Crimes in urban spaces are a concern for every person's own welfare. Crime has become a socio-economic problem that is continuously affecting the quality of life and economic growth [2]. Law enforcement agencies continuously strive to understand the nature of crimes which are occurring in their responsible city with the aim of identifying who is committing them; where and when they are most likely to occur; what is the likelihood that a certain crime takes place; and how much will the rate vary along the years.

In our research, we focus on tackling crime prediction from an Artificial Intelligence point-of-view, by using both classification and regression techniques to predict the type of crime as well as the future crime rate respectively. Furthermore, we also study to see how much data is required by our models to train on such that their performance and effectiveness remain on acceptable levels.

1.1 Motivation

Nowadays, there are numerous crime datasets which are being publicly available to researches. Criminology is also one of the most important fields when applying data mining techniques, as these can produce effective results [19]. Providing accurate and reliable crime predictions help law enforcement and other entities to effectively prevent crime from happening again as well as to handle them effectively when they occur [9].

1.2 Why the Problem is Non-Trivial

This problem comes up with some challenges which can be foreseen in advance. First of all, finding the right dataset/s which is/are suitable for our use case is an element of concern. Various countries collect data differently since crime information is considered as a sensitive data, and this collection differs amongst the different countries. Additionally, not all countries publish their crime data, and thus, makes it more difficult for us to find the most suitable data. For instance, up to our knowledge, no such dataset exists for the Maltese islands within the public domain. It is also vital to ensure that any available datasets are not only comprehensive but are also ideal for our research topic objectives.

Another notable problem is the fact that datasets can be very large in size. Hence, dealing with these datasets can be challenging at times since creating both classical Machine Learning (ML) models as well as more advanced Deep Learning (DL) models can be computationally expensive to execute on a normal computer machine.

1.3 Aims and Objectives

The primary aim for this research is to employ different data mining techniques on urbanised city's open crime datasets to predict the type of crime and forecast the expected number of future crimes, using classification and regression methods on San Francisco (SF) and New York City (NYC) datasets respectively. We also aim to investigate the amount of data our classification and regression models built require to train on such that they remain effective in their predictions.

This research aim is formalised through the following objectives:

- **Objective 1 (O1)**: Identifying the best classification models, ranging from classical machine learning to deep learning methods, which best predict the type of crime given initial crime observations - namely temporal and spatial information, and whether an arrest was made or not.
- **Objective 2 (O2)**: Finding the best regression models, also ranging from classical machine learning to deep learning methods, in order to forecast the expected crime rate along future years.
- **Objective 3 (O3)**: Investigating the amount of training data all implemented models require such that their performance is not hindered, and thus, retain their effectiveness.

2 Background Research and Literature Review

From the background research conducted on other researchers work, the application of different data mining techniques within the crime analysis domain is vast and can take many different forms. These include analysing criminal networks, calculating the likelihood that an individual becomes a victim, creating effective patrol routes as well as forecasting where a perpetrator is likely to commit a future crime based on their previous activity and their previous crimes [20]. However, as already highlighted in Section 1.3 above, this research focuses on crime prediction and thus, an overview of others' work on crime prediction will be discussed next.

2.1 Crime Prediction

Crime prediction is one of the most popular objectives which researchers aim to attain within the crime analysis domain [1, 5, 7, 8, 11, 13, 14, 16]. This objective is often handled through supervised machine learning algorithms, and are applied into two forms, either as a classification task or as a regression task. A classification task is one which uses a model that classifies a discrete value, for instance, labels or categories. On the other hand, a regression task is one that utilises a model to predict some continuous value, such as a quantity value or a probability value.

2.1.1 Datasets Utilised in Research. To solve this objective, it is imperative to have data, which can either be collected manually [10, 11, 17] or publicly available on the internet [3, 5, 12, 13]. In this section, we will only be describing those datasets that are generally used in literature.

Pradhan et al. in [12] used an open-source dataset consisting of crimes together with the temporal and spatial features, collected by the San Francisco Police Department (SFPD) and available through the Open Data initiative [4]. Similarly, Sardana et al. [13] used a subset of [4], which is found on Kaggle[1]. On the other hand, both Catlett et al. [3] and Elluri et al. [5] decided to use NYC crime data respectively. This dataset includes all misdemeanor, violation and felony crimes which were reported to the New York Police Department (NYPD) for the years of 2003 up until 2019[2]. Essentially, this dataset is an extension of [4] in terms of feature variables available since it also includes data related to the victim and suspect (age group, race, gender respectively). Moreover, Elluri et al. integrated weather data as well to this current NYC dataset.

2.1.2 Classification Approach. Generally, the authors from the reviewed literature utilised ML methods to perform predictions. The most common ML models used are Decision Trees, Logistical Regression Classifier and Naïve Bayes Classifier [5, 12, 13].

Pradhan et al. used Naïve Bayesian classifier and Decision Tree classifier models to predict the type of crime in a particular location at specific times [12], using the SF dataset [4]. Similarly, Yuki et al. in [18] also performed crime type prediction using temporal and spatial features but on the city of Chicago. Multiple ML algorithms and Ensemble methods are implemented, for instance, Random Forest Tree and Bagging respectively. Furthermore, Sardana et al. [13] implemented similar solutions to [12]. The main difference between these two papers is that [13] enhanced their model's scores by grouping crime types into three super classes according to the corresponding USA's law category (Misdemeanour, Infraction and Felony). Sardana et al. implemented traditional ML models, such as Logistical Regression Classifier, Decision Tree Classifier and much more, to predict this new categorical variable. Hence, a comparative study analysis was carried out to find the best model which is predicting the most accurate for their research problem. It is worth noting that in [13], a subset of the dataset in [4] is used, having the same features and date range, which is publicly available on Kaggle.

On the other hand, Elluri et al. [5] implemented similar solutions as well, but decided to integrate additional data to crime. Apart from crime, temporal and spatial data, Elluri et al. also decided to include weather data. They focused their work on NYC crime data and their main aim was to study if weather data can help in predicting the type of crime. Hence, they performed a comparative study between predicting crime types with and without weather data. Additionally, they went deeper in their implementations by developing more ML algorithms (example, SVMs) and DL methods, including NNs, CNNs and RNNs. Similarly, Stalidis et al. aimed at examining DL architectures for crime prediction by comparing ML models with DL models [15]. They experimented on 5 different open datasets and from the results obtained, Stalidis et al. conclude that these DL architectures outperform the existing best performing methods.

2.1.3 Regression Approach. Similar systems from literature indicate that Linear Regression and Auto-Regressive Integrated Moving Average (ARIMA) are amongst the most common models used for this problem [3, 11]. However, some authors also used DL models, particularly the Multi-Layer Perceptron (MLP) model [6, 11].

Marzan et al.'s main aim was to predict the future crime trends [11]. They focused on predicting the number of crimes in each district of Manila City's crime data for the future days and weeks. This is achieved by using Time Series Forecasting methods, a data mining technique which is dependent of time. Subsequently, these authors developed several ML and DL methods that predict future crime rates, which include Linear Regression, Gaussian Processes, MLP and

[1] https://www.kaggle.com/competitions/sf-crime/data; [Last Accessed: 6th April 2022]
[2] https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i?ref=hackernoon.com; [Last Accessed: 2nd May 2022]

SMOreg. Similarly, Catlett et al. also tackled this objective by predicting the crime rate of Manhattan borough in NYC as well as predicting the crime rate in Chicago [3]. To achieve this, these authors initially implemented a clustering algorithm (in their case, DBSCAN) to find regions having crime densely hotspots. Hence, every *k* cluster corresponds to a region, and Catlett et al. then chooses the top 3 most dense regions to predict their future number of crimes individually as well as overall (for whole city). Several ML algorithms are implemented, including Decision Tree Regression and ARIMA, with the latter model performing the best across both datasets.

Huang et al. developed a crime prediction framework called DeepCrime [6]. This framework was built on a Deep Neural Network architecture which uncovers dynamic crime patterns and captures time-evolving dependencies between crimes as well as ubiquitous data in urban space. In their research study, each individual crime occurrence was given an input weight vector, and thus, this weight vector denotes the importance of each incident in their respective crime category (types of crime, for instance, felony assaults and robberies) which leads to better prediction of crimes. Furthermore, a MLP is appended in their proposed framework, and hence, they can assign an importance weight to each category. Finally, these authors compared their framework with other existing baseline models (for example, ARIMA) to evaluate the performance of their framework, where they have concluded that their framework performs better than these baseline models.

3 Methodology

In this section, we outline the approach taken towards the objectives of crime type prediction and crime rate prediction. Subsequently, we explain the process and decisions taken in our research, highlight any pre-processing steps taken as well as document any observations made. Figure 1 illustrated below outlines an overview of our research study.

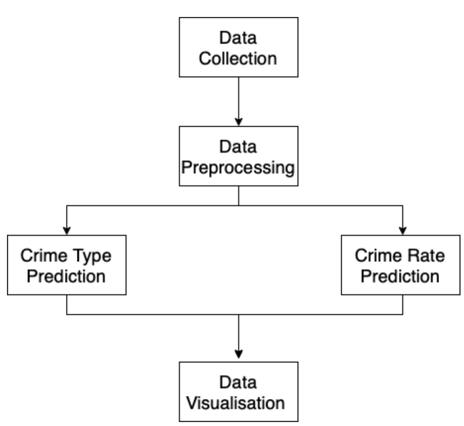

Figure 1. Overview of Solution

3.1 Datasets Used

In our research, we intend to compare our results obtained with those in literature [3, 13]. For this reason, we decided to utilise the same datasets which the respective authors used in their research study to provide this comparative analysis. The dataset used for O1 is SF's dataset found on Kaggle, as used by Sardana et al. [13], and contains 878049 rows of crime incidents between the years of 2003 to 2015. On the other hand, for O2, NYC crime data was used, as utilised by Catlett et al. [3], and contains a total of 1328134 historical crimes between January 2006 to December 2016.

Before implementing our models, it was vital to prepare the respective datasets. Therefore, pre-processing techniques commonly used in literature [11, 14] were applied to the data:

1. **Data Reduction**: reducing the data, for instance, removing unneeded columns.
2. **Data Cleaning**: dealing with missing and inconsistent values. Fixing column values which are misspelled or uneasy to understand.
3. **Data Integration**: integrating new attributes from current features, or from external sources into the dataset.
4. **Data Transformation and Discretisation**: transforming feature values into groups and encoding categorical features before implementing the models to solve our objectives.

Since our research study consists of two different datasets, different pre-processing steps had to be applied for the respective data. Furthermore, to guarantee a fair comparative analysis, any pre-processing steps done by the authors were replicated from our end.

Our main aim for O1 was to replicate and extend Sardana et al.'s research [13] in such a way to implement models that classifies crime with a higher level of accuracy than theirs. Therefore, we utilised the same SF dataset, performed the same pre-processing techniques but implemented more ML models and developed a DL model, ranging from Logistical Regression Classifier to MLP neural networks. Hence, we directly compared our obtained results with the findings of [13]. Additionally, in our research, we retained the *resolution* feature (how the crime was resolved, for example, an arrest or not) as opposed to [13]. Furthermore, we decided to optimise *hour*, *day* and *month* columns by mapping their respective values into their corresponding predefined categories to further balance our dataset. Hence, we ended up with new features, including *DayGroup*, *WeekGroup* and *Season* which contains the corresponding categories. For example, the original *month* feature contained 12 values, each denoting a month whilst the new *Season* feature now contains 4 values, each denoting a season.

For O2, we focused on forecasting the number of crimes likely to happen in the next three years individually (2014, 2015 and 2016 respectively), based on previous crime data. We used regression-based ML models as well as DL models to predict the crime rate for the years of 2014, 2015 and 2016 separately, similar to Catlett et al. [3]. In this way, we directly compared our obtained results with those found in [3] as well as compared each regression model having different training features combinations. The main difference between our research and [3] is that we integrated additional data (*population*, *unemployment rate* and *median income*) to our dataset to help our models when making predictions. Thus, we also analysed which external feature was the most important one when predicting future crime rate.

Finally, for O3, we investigated the amount of training data our classification and regression models required, such that their optimal performance was retained, and hence, remain effective. We began by splitting our respective datasets into a training set and a test set, and proceeded to diminish the training set while keeping the test set unchanged. We performed multiple experiments, each time re-running the models and checking their performance. This process was repeated until the minimum training size such that all models retain their effectiveness.

4 Evaluation

In this section, we explain our evaluation strategy, present our results and evaluate all objectives. We split this section into three sub-sections, each corresponding to the respective objective.

4.1 Evaluation of Crime Type Prediction

For O1, we aimed to classify the type of crime based on temporal, spatial and whether an arrest was made or not, using classical ML and DL models, being Logistical Regression Classifier (LRC), kNN, Decision Tree Classifier (DTC) as well as MLP classifier neural network. As highlighted in Section 3 above, we began by replicating the findings of Sardana et al. [13], and then proceeded to improve on their results. Hence, a total of three experiments were carried out, two of which concerns replicating Sardana et al.'s system using the same original features as used in [13] (first one predicts all crime types whilst the second predicts the three crime super classes categories), and the last one was an extension to Sardana et al.'s system to expand on their work (including *resolution* feature and developing additional models, including a DL model). Since the SF dataset has an imbalanced class distribution, we used the weighted f1 score, as utilised by [13]. Furthermore, 10-fold cross validation was also performed on every model implemented in order to avoid possible overfitting.

Once the *resolution* feature was introduced as part of the training features, the results improved for every evaluation metric. Thus, this implies that such a feature impacts positively the performances of each developed model. When comparing our findings with [13], we observed that our models produced better overall results, where the f1 score has increased from 0.57 up to 0.71, as shown in Table 1 above. Furthermore, the model which performed the worst is kNN whilst LRC, DTC as well as MLP produced better and identical results.

Table 1. Model Evaluation F1 Scores for Classifying Crime Super Classes with and without *Resolution* Feature

Model	F1 Score (without res.)	F1 Score (with res.)
LRC	0.57	**0.71**
kNN	0.53	0.68
DTC	0.57	**0.71**
MLP	0.57	**0.71**

Table 2. Comparing MAPE Scores without external data (orig) and with unemployment rate external data (ext)

Year	LR		RFR		MLP	
	orig	ext	orig	ext	orig	ext
2014	1.11	3.12	4.28	2.4	8.71	0.68
2015	0.19	5.97	4.63	6.32	9.12	4.1
2016	3.88	0.18	2.44	0.56	6.89	1.48
Avg	**1.72**	2.06	3.78	5.46	8.24	**0.74**

4.2 Evaluation of Crime Rate Prediction

As explained in Section 3 above, in O2 we predicted the number of crimes that will take place in future years, by implementing classical ML (non-DL) and DL models on NYC's dataset. Hence, we implemented Linear Regression (LR), Random Forest Regressor (RFR) and a MLP regressor neural network models. For each technique, we performed 5 experiments; the first experiment using only previous crime incidents, the second using crime and population data, the third using crime and median income data, the forth using crime and unemployment rate data and the last using crime, population and unemployment rate data. Then, for each experiment conducted, we provide the MAPE evaluation scores, a common evaluation metric used by [3, 11**?**], which will allow us to observe the quality of the developed models and directly compare our findings with [3]. The set of features which are helping our predictions include previous crime data and unemployment rate data, and hence, we are showing the table for this only.

When comparing our findings with those in [3], we can infer that we have managed to obtain better scorings than Catlett et al.'s study. The main differences between our study and [3] is that we integrated external data to the dataset, such as *unemployment rate* and also decided to implement other

ML models, such as Linear Regression and a DL model, being the MLP deep neural network. In their study, Catlett et al. concluded that the ARIMA model fared best overall, with an average MAPE of 5.67. However, in our study we managed to implement a MLP model having an average MAPE score of 0.74, which is substantially less than the best score in [3]. From the results obtained in Table 4 above, one can observe that the *unemployment rate* feature is the most important feature overall when predicting crime rate due to its positive impact on the predictions.

4.3 Evaluation of Training Data Required

In this section, we evaluate O3 where we analyse the minimum amount of data our built models require to train on such that their overall effectiveness is not hindered. To the best of our knowledge, no one has attempted to solve an objective similar to this, and thus, we cannot directly compare our findings with the read literature. This section is split into two sub-sections; the first sub-section refers to our classification models whilst the second one concerns our regressor models.

4.3.1 Classification-based Models. For this part of O3, we analysed the effectiveness of the classification models when their training data size is reduced. In Section 3 above, we discussed the process of how this objective was achieved. Through the process, we have derived the size of the training set needed to train our models and the respective f1 scores obtained for each model. Hence, in this experiment, the training set consisted of the first 40% of crime records from the whole SF dataset (351219 records), whilst the test set is always the last 20% rows. It is also worth re-mentioning that the *resolution* feature together with the other features, such as *DayGroup* and *Season* were utilised throughout the entire list of experiments carried out.

Table 3. Model Evaluation F1 Scores for Classifying Crime Super Classes using *Resolution* Feature on the full and the diminished Training Set

Model	F1 Score (O1)	F1 Score (O3)
LRC	0.71	0.70
kNN	0.68	0.67
DTC	0.71	0.70
MLP	0.71	0.70

Table 3 above outlines the f1-scores obtained for both O1 (full training set) and O3 (diminished training set). As observed in this table, the results obtained fared almost identical to each other, having the highest score of 0.7 for LRC, DTC and MLP neural network models, as compared to the highest score of 0.71 in O1. It is also interesting to observe that there is a similarity between the obtained scores when using the full dataset versus when using the reduced dataset because all the respective models which are trained on the minimal training size have f1-scores that are 0.01 less than their corresponding model performance in O1. For instance, the kNN model in O3 produced an f1 score of 0.67 whilst for O1, it produced the result of 0.68. Therefore, a training set which is 40% of the whole dataset is enough for the models to train on and yet produce effective results.

Table 4. MAPE Scores for Crime Rate Prediction using Crime, Population and Unemployment Rate Data on the full and the diminished Training Set

Year	LR		RFR		MLP	
	O2	**O3**	**O2**	**O3**	**O2**	**O3**
2014	3.7	5.4	6.77	3.73	2.72	0.21
2015	3.34	160.71	7.12	4.07	3.58	0.65
2016	0.19	397.57	4.88	1.9	1.36	1.52
Avg MAPE	**2.41**	187.89	6.25	3.23	2.55	**0.79**

4.3.2 Regression-based Models. After having conducted the process of finding the minimal training size, we observed that training data between the years of 2010 to 2013 is sufficient for the models to train on and still produce optimal results. Hence, we performed the same experiments as done in O2 but on the diminished training data and tabulated the results obtained, showing only Table 4 where the best overall result was produced.

From Table 4 above, it can be observed that the least overall MAPE score of 0.79 was produced by the MLP model when using crime, population and unemployment rate data as training features. When compared to the best overall score of 0.74 obtained in O2, it can be inferred that the results obtained fared very similar to each other, and thus, minimising the full training data to approximately half is sufficient for the models to train on and remain effective in their predictions. Similarly to O2, the *unemployment rate* feature was also the most important feature since it had the most positive impact on the model's predictions (when only one external data is combined with crime data) and the MLP model is also the best overall model in terms of accurate predictions since it produced the lowest MAPE value.

5 Conclusion

In this section, we first provide an overview of the main conclusions achieved and then conclude this research by discussing possible future work within the crime analysis domain.

5.1 Overview of Findings

In our research, we managed to solve every objective set up in this research. For both O1 and O2, we improved on the findings of [13] and [3] respectively. In the first objective, we inferred that including the *resolution* feature improves the classification models f1-scores from 0.57 up to 0.71. In the second objective, we found out that the *unemployment rate* feature is the most important feature since it has the greatest positive impact on the predictions, resulting in an MAPE value of 0.74 produced by the MLP model. Additionally, we noted that the DL model built in O1 was not effective when compared to the other ML models. On the contrary, in O2 this was not the case because the MLP DL model performed the best overall, deeming it as more effective than the other models. Finally, in O3, we deduced that diminishing the training data by approximately half is sufficient for all models to remain effective in their predictions.

5.2 Future Work

Despite solving all our objectives, we still believe that there is always room for improvement. Hence, for future work, we would perform crime hotspot detection using clustering algorithms to identify crime dense regions and predict crime for these regions detected. This can be extended by studying the application of hierarchical spatial methods which are able to split clusters when their sizes are huge. Furthermore, for crime rate prediction, we can also study the correlation between the crime trends and other city events to identify a relationship amongst them.

For crime type prediction, it could be very interesting to include demographic data, such as victim and suspect race, to observe if this has any effect of the performances of the models. Finally, further adding advanced state-of-the-art DL models, such as LSTM and RNN, would be interesting to see if such complex models can help in predicting crime more effectively.

References

[1] Tahani Almanie, Rsha Mirza, and Elizabeth Lor. 2015. Crime Prediction Based On Crime Types And Using Spatial And Temporal Criminal Hotspots. *International Journal of Data Mining and Knowledge Management Process* 5 (08 2015). https://doi.org/10.5121/ijdkp.2015.5401

[2] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, and Alex Pentland. 2014. Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data. (09 2014). https://doi.org/10.1145/2663204.2663254

[3] Charlie Catlett, Eugenio Cesario, Domenico Talia, and Andrea Vinci. 2019. Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments. *Pervasive and Mobile Computing* 53 (2019), 62–74. https://doi.org/10.1016/j.pmcj.2019.01.003

[4] City and County of San Francisco. 2012. Police Department Incident Reports: Historical 2003 to May 2018. https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmnf-yvry

[5] Lavanya Elluri, Varun Mandalapu, and Nirmalya Roy. 2019. Developing Machine Learning Based Predictive Models for Smart Policing. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*. 198–204. https://doi.org/10.1109/SMARTCOMP.2019.00053

[6] Chao Huang, Junbo Zhang, Yu Zheng, and Nitesh V. Chawla. 2018. DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Torino, Italy) *(CIKM '18)*. Association for Computing Machinery, New York, NY, USA, 1423–1432. https://doi.org/10.1145/3269206.3271793

[7] N Kanimozhi, N V Keerthana, G S Pavithra, G Ranjitha, and S Yuvarani. 2021. CRIME Type and Occurrence Prediction Using Machine Learning Algorithm. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. 266–273. https://doi.org/10.1109/ICAIS50930.2021.9395953

[8] Suhong Kim, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri. 2018. Crime Analysis Through Machine Learning. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. 415–420. https://doi.org/10.1109/IEMCON.2018.8614828

[9] Jay Yoon Lee, U. Kang, Danai Koutra, and Christos Faloutsos. 2013. Fast Anomaly Detection despite the Duplicates. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13 Companion)*. Association for Computing Machinery, New York, NY, USA, 195–196. https://doi.org/10.1145/2487788.2487886

[10] Olta Llaha. 2020. Crime Analysis and Prediction using Machine Learning. 496–501. https://doi.org/10.23919/MIPRO48935.2020.9245120

[11] Charlie S. Marzan, Maria Jeseca C. Baculo, Remedios de Dios Bulos, and Conrado Ruiz. 2017. Time Series Analysis and Crime Pattern Forecasting of City Crime Data. In *Proceedings of the International Conference on Algorithms, Computing and Systems* (Jeju Island, Republic of Korea) *(ICACS '17)*. Association for Computing Machinery, New York, NY, USA, 113–118. https://doi.org/10.1145/3127942.3127959

[12] Isha Pradhan, Katerina Potika, Magdalini Eirinaki, and Petros Potikas. 2019. Exploratory data analysis and crime prediction for smart cities. *IDEAS '19: Proceedings of the 23rd International Database Applications & Engineering Symposium*, 1–9. https://doi.org/10.1145/3331076.3331114

[13] Divya Sardana, Shruti Marwaha, and Raj Bhatnagar. 2021. Supervised and Unsupervised Machine Learning Methodologies for Crime Pattern Analysis. *International Journal of Artificial Intelligence & Applications* 12 (01 2021), 83–99. https://doi.org/10.5121/ijaia.2021.12106

[14] Shiju Sathyadevan, M.S. Devan, and S. Surya Gangadharan. 2014. Crime analysis and prediction using data mining. In *2014 First International Conference on Networks Soft Computing (ICNSC2014)*. 406–412. https://doi.org/10.1109/CNSC.2014.6906719

[15] Panagiotis Stalidis, Theodoros Semertzidis, and Petros Daras. 2021. Examining Deep Learning Architectures for Crime Classification and Prediction. *Forecasting* 3, 4 (2021), 741–762. https://doi.org/10.3390/forecast3040046

[16] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. 2016. Crime Rate Inference with Big Data. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 635–644. https://doi.org/10.1145/2939672.2939736

[17] Jacky Yu, Wei Ding, Ping Chen, and Melissa Morabito. 2014. Crime Forecasting Using Spatio-Temporal Pattern with Ensemble Learning. *Advances in Knowledge Discovery and Data Mining* 8444. https://doi.org/10.1007/978-3-319-06605-9_15

[18] Jesia Quader Yuki, Md. Mahfil Quader Sakib, Zaisha Zamal, Khan Mohammad Habibullah, and Amit Kumar Das. 2019. Predicting Crime Using Time and Location Data. In *Proceedings of the 2019 7th International Conference on Computer and Communications Management* (Bangkok, Thailand) *(ICCCM 2019)*. Association for Computing Machinery, New York, NY, USA, 124–128. https://doi.org/10.1145/3348445.3348483

[19] K. Zakir Hussain, M. Durairaj, and G. Rabia Jahani Farzana. 2012. Criminal behavior analysis by using data mining techniques. In *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012)*. 656–658.

[20] Xiangyu Zhao and Jiliang Tang. 2018. Crime in Urban Areas: A Data Mining Perspective. *SIGKDD Explor. Newsl.* 20, 1 (may 2018), 1–12. https://doi.org/10.1145/3229329.3229331