Survey: Crime Prediction using Machine Learning Approach

Esraa Faisal Khalaf[a]*, Dr. Ali Hasan Taresh*[b]*

[a]*Informatics Institute for Postgraduate Studies (IIPS), Iraqi Commission for Computers and Informatics (ICCI), Iraq, Baghdad,*
E*mail: ms202020598@iips.icci.edu.iq*

[b]*University of Information Technology and Communications (UoITC), Baghdad, Iraq,*
 Email: alihtaresh@uoitc.edu.iq

A R T I C L E   I N F O

A B S T R A C T

The presented studies have proven that machine learning algorithms The field of machine learning (ML) is expanding as more people realize how important it can be in a variety of crucial applications, including data mining, natural language processing, picture recognition, and expert systems. As it is well known that ML offers possible answers in all these domains and more, it is destined to be one of the pillars of our future civilization. This article presents an outline of the function of machine learning in prediction. algorithms have excelled in solving prediction and classification problems. Below We highlight the machine learning algorithms and techniques used in predicting crimes in particular, and the accuracy of the results obtained by each study or research. We see the challenges faced by a study or researcher who used machine learning algorithms and we hope with this paper, providing the researcher, in particular, with information covering the most important studies or research presented during the past five years to abbreviate the time of the researcher, and in the interest of his effort.

MSC..

1. Introduction

Crime forecasting has become an urgent matter that requires research in the field of data science and with the availability of reliable and documented crime data being adopted by police departments and stations around the world. This paper investigates machine learning algorithms used to predict the occurrence of crime, some used ML algorithms with geographic information systems (GIS) data, others used machine learning algorithms to perform spatial analysis of geographic data, and some used machine learning algorithms with demographic data, while others were content with adopting based on study Statistical analysis. Crime is a prevalent social concern that plays a significant role in society's vitality and economic prosperity. Law enforcement agencies continue to demand advanced technologies and novel approaches to improve crime analysis and better safeguard their communities as crime rates rise. In this paper, we try to review and cover the most important studies that dealt with this topic over the past five years from 2015 to 2021.

NOMENCLATURE

∗Corresponding author

Email addresses:

Communicated by 'sub etitor'

Aradius of

Bposition of

Cfurther nomenclature continues down the page inside the text box

---

2. Survey Classification

 Many studies pay great attention to expanding and intensifying the study on the topic of crime prediction due to the availability of data that has become open, circulated and documented electronically, which made it somewhat easier for the researcher to study and analyze, so there is a shortening of the collecting step for these data and saving a lot of effort and time than in the past. The specified articles are classified based on the following factors.

- Using socioeconomic, spatial-temporal, and demographic elements to predict crime.

- Identifying and predicting crimes based on their characteristics.

- Predicting crimes based on victim characteristics (i.e., women, children, elderly people)

- Predicting crimes by using different algorithms (i.e., machine learning)

3. Literature Review

    We review what these studies and research have provided. In one of these papers, the authors used various data collection techniques to analyze the crime trend in Bangladesh, where they used the K-nearest. through the analysis of the data, they were able to discover and determine the crime rate in many places. Thus, researchers were able to know the safe route from the unsafe route, depending on the prediction ratio resulting from the analysis of crime data and the extent of the relationship of crime rate in different places with age and gender data. Thus, they came to know the relationship between Crime rate based on region, monthly levels and data of research was about of last three years of crime and showing crime forecasting trend. The table below shows the accuracy of using different approach (see Table I):

KNN provides the crime trend prediction system with the highest percentage [1].

Table 1 - precision.

The Year	Approach	precision
2019 – 2018-2017	LR	73.61403
2019 – 2018-2017	NB	69.5087
2019 – 2018-2017	KNN	76.9298

The criminal records from the past were studied in this study in order to forecast crime in amelioration and develop countermeasures. Various sorts of crimes and tools used in the crimes were analyzed and forecast using crime data from Baltimore between 2011 and 2016. The data world's crime data was used, and the crime kinds and tools were predicted using a variety of machine learning methods, including Multiclass Logistic Regression and Multiclass NN. The dataset used is Chicago crime data, which is used to make predictions and show the patterns and trends of various crimes. To anticipate crime categories and tools used in the incident, the author used a victim-based crime data set. The Multiclass LR and Multiclass NN algorithms were used to predict the crime type

and the crime tool. The Multiclass LR approach had superior accuracy, precision, and recall for weapon forecasting. The following are the outcomes of the weapon evaluation model utilizing the Multiclass LR technique. averaged recall: 0.64856, overall accuracy: 0.789046. And the results of the weapon estimate model utilizing the Multiclass NN technique. averaged recall: 0.612495, overall accuracy: 0.754892. [2].

Another work uses crime data to introduce six distinct kind of machine learning techniques, covering KNN and DT, NB, LR, CART Classification and Regression Tree, and SVM. These algorithms are put to the test to see how accurate they are. The primary goal of this study is to assess the effectiveness and applicability of ML algorithms in data analytics. The author explained about the highest crime rate in the data he dealt with in his study from 2001 to 2012, for records collected for crime against women in 29 Indian states, it was found that India in 2012, recorded the highest statistic of crimes committed against women. The results concluded that the KNN algorithm is more efficient than other ML techniques. This paper presented a statistical study on the highest crime area and the type of crime that occurs regularly in India. The author hopes that this study and the result obtained will be useful for controlling violence against women in the future [3].

Another study aimed is to find a machine learning process for shortlisting accused persons for a new offense, with a focus on Modus Operandi. The goal is to create an optimal criminal suspect shortlist from the criminal database that contains the prospective offender in a prioritized manner, rather than to single out a specific offender. The machine learning algorithm scours the database for similarities between crimes in an ever-increasing pattern, attempting to find parallels in method of operation (MO). We might locate an offender within the pattern they can discover offenders proportionately inside the pattern, which will aid in criminal shortlisting. Out of the 21 grave crime categories in Sri Lanka, housebreaking and stealing is considered the category with the most unique MOs, hence the focus of the research is on that specific category [4].

There is work introduce suggestion the predication model with varied of layers depend on two types of information, namely historic crime data and location information by using model in a mathematical form. using Machine learning and DL algorithm applied on crime dataset from the city of Chicago in the US of America from 2001 to 2018 year, the result was following (see Table II) [5]:

Table 2 - Comparison of tests.

	NB	DT	RF	NN	SVM
Classifier	Multinomial NB Bernoulli NB Gaussian NB From sklearn	DT Classifier from sklearn	RF Classifier from sklearn	TensorFlow	SVC from sklearn
Accuracy	----	38%	59%	81%	---

Another project objective to develop a ML model for crime predicting based on geographical characteristics for a variety of crime categories. (OSM) spatial data is returned using the reverse geocoding approach. This research is also aimed at identifying hotspots. Then, based on the location of distinct hot spots for various forms of crime, the spatial distance feature was generated, and this value was employed as a feature for classifiers [6].

Another research, we examine how dynamic variables derived from Foursquare check-in data can be utilized to forecast short-term crime events with fine spatiotemporal precision. Despite the fact that crime event prediction has gotten a lot of attention because of its social importance, its success rate isn't very high. There are few studies

that focus on human movement via social media, as most studies focus on relatively static variables like regional characteristics, demographic data, and subjects retrieved from tweets. Based on criminology research findings, we define a number of dynamic traits and report their associations with various types of criminal events and report their correlations with various sorts of criminal events based on criminology research findings in this study. We find that some types of criminal activity, such as theft, drug offenses, fraud, and unauthorized entrance, are more strongly linked to dynamic factors than others. A total of two data sets have been gathered. One is from Brisbane, Queensland's metropolis city, while the other is from New York City, New York. Between January 1st and September 9th, 2013, the authors used four prediction models: (RF), (NN) kernel, (SVM), and (LR) Model [7].

Other paper focused on classifying cluster crimes based on their frequency over different years. For the real crime data set reported from 1990 to 2011, a theoretical model based on data mining techniques such as clustering and classification was deployed [8].

Machine learning applications have made it a significant study area. One of the most important domains in which data mining can be used is criminology. The decision tree (J48) approach was used to create a crime forecasting prototype model was examined in this work since it has been characterized in the literature as the most effective ML technique for forecasting crime data. The J48 algorithm correctly identified the unknown category of crime data, according to the testing findings. 94.25287 percent of the time, which is good enough for the system to be trusted for future crime prediction [9].

Other studies described crime prediction using ML algorithms such as (SVM), RF, GBM, and NN, here this study of crime forecasting focus on location and time and relating these features with demographic data.

Prediction results with first dataset (SVM) got the overall 57.3%, RF got the results 66.07% ,gradient boosting machines got the results 63.70% , Prediction Results with second dataset after the model were trained using SVM with got the overall 79.39% , trained a model using RF deep when they set number of tree to 250 the result of accuracy get to (65.79% accuracy) and when trained a model using Gradient Boosting the perfect accuracy results occur when they set number of tree to 300 (61.67% accuracy) [10].

The NB classifier is utilized to derive criminal decisions in another study. The crime dataset is synthesized using from another the work [11] , which employs a parametric model based on the K-means technique and a Gaussian mixture model. The proposed system is tested for the criminal forecasting trouble using cross-validation, and the results reveal that it may be used in criminology with an 83 percent success rate to aid security professionals in resolving incidents [12].

Other study, to anticipate the future of crime trends in Bangladesh, a linear regression model was used the real-time statistics of crime. This dataset is then subjected to the LR model. Following the training of the model, crime forecasting is carried out for dacoit, robbery, murder, women and child repression, kidnapping, burglary, theft, and other crimes in Bangladesh's various regions. the abstract of this paper is forecast the status of crime [13].

The goal of another study was to identify crime hotspots in both space and time. It analyzes two distinct real-world crime datasets for Denver, Colorado, and Los Angeles, California, utilizing a statistical analysis and several visuals to compare the two datasets. The article goes on to explain how the Apriori algorithm was utilized to generate surprising common patterns for criminal hotspots. In addition, the study shows how we used Decision Tree and Nave Bayesian classifiers to identify possible criminal categories. They conducted an analysis research by integrating our findings from the Denver crime dataset with theirs, which resulted in a prediction accuracy of 51 percent in Denver and 54 percent in Los Angeles information on demographics [14].

To build new prediction models, researchers from informatics, computer science, mathematics, and statistics have collaborated with criminologists, sociologists, and others. In addition to typical crime prediction models, Models for predicting crime, as well as crime forecast models integrated with data from social media data were able to have a significantly higher success rate in certain types of crime (Corso, 2015; Gerber, 2014; Wang & Gerber, 2015; Wang et al., 2012). In contrast to the majority of crime forecasting approaches are used for retrospective crime forecasting, which takes historical crime data into account [15].

Another study introduced a web mapping and visualization-based crime prediction tool that was built in R Language using its various supporting mapping and GIS libraries. The suggested framework used a variety of visualization tools to demonstrate crime trends and the various ways that machine learning algorithms can be used to forecast crimes. The study used data from the UK Police Department official website [16] and followed the processes outlined in the source [17], in which the essential stages are data collecting, data pre-processing, data visualization, and model construction, which are covered in greater detail in the source [16]. In terms of analysis, this is correct. A construction phase is used to apply different machine learning algorithms when developing a prediction model [18] .

There are papers initiated a discussion on the impact of economic issues on crime, discussed the Indian economic crisis, and attempted to demonstrate the impact of the economy on crime rates. From several official portals, crime data and economic indicators such as Gross State Domestic Product (GSDP), Net State Domestic Product (NSDP), Per capita Net State Domestic Product, unemployment rate, and Consumer Price Index (CPI) were gathered for the years 2004-2013. The decision trees, random forest, linear regression, and neural networks algorithms were used to evaluate independent variables (Gross District Domestic Product (GDDP) and the unemployment rate) as well as dependent variables (theft, burglary, and robbery). Out of the four, linear regression produced the most accurate findings [19].

A document was published that collected data on the population from a variety of sources, including police department crime data, census data, location-based social networks, subway journeys, and taxi rides. There are two types of crime prediction models in use. Long-term crime prediction models look at crime rates over a period of 1–5 years. Short-term crime prediction models, on the other hand, focus on a short time frame, which might range from one day to one month. Foursquare and other location-based social networks provide information on people on the move and their mode of transportation. The public data was compared to the criminal data obtained from the police department. Three alternative tree-based machine learning models were used to evaluate census, geographical, and temporal characteristics. The characteristics of the ambient population are delineated by the residential population and spatial-temporal features, resulting in effective crime prediction by machine learning models. This human movement data-based model produces better results than census-based models [20].

4. Conclusion

All the studies presented above agreed that crime constitutes a major social problem that the whole world faces challenges to solve as it affects the growth of the economy, threatens the stability of countries, stresses life and impedes development in all its fields. Governments, security organizations and the competent authorities strive with effort with availability of technological development and documented data, which are harnessed to serve the public interest. In order to solve crimes, crime prevention and detection has become an important and difficult trend. Various approaches for solving crimes have been discovered in numerous studies and have been applied in a variety of applications. These studies may aid in the speeding up of the criminal justice system and the automatic detection of criminals by computerized systems. Furthermore, quickly growing technologies can aid in the resolution of such problems.

References

[1]   S. Mahmud, M. Nuha, and A. Sattar, Crime Rate Prediction Using Machine Learning and Data Mining, vol. 1248, no. June. Springer Singapore, 2021.

[2]   K. Kyoung-Sook and J. Yeong-Hoon, "A Study on Crime Prediction to Reduce Crime Rate Based on Artificial Intelligence," Korean J. Artif. Intell., vol. 9, no. 1, pp. 15–20, 2021.

[3]   P. Tamilarasi and R. U. Rani, "Diagnosis of Crime Rate against Women using k-fold Cross Validation through Machine Learning," Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020, no. Iccmc, pp. 1034–1038, 2020, doi: 10.1109/ICCMC48092.2020.ICCMC-000193.

[4]   M. Munasinghe, H. Perera, S. Udeshini, and R. Weerasinghe, "Machine Learning based criminal short listing using Modus Operandi features," 15th Int. Conf. Adv. ICT Emerg. Reg. ICTer 2015 - Conf. Proc., no. November, pp. 69–76, 2016, doi: 10.1109/ICTER.2015.7377669.

[5]   [H. Aitelbour, S. Ounacer, Y. Elghomari, H. Jihal, and M. Azzouazi, "A crime prediction model based on spatial and temporal data," no. November 2018, 2019, doi: 10.21533/pen. v6i2.524.

[6]   A. Soares, "Predicting Crime Using Spatial Features," no. March, 2018.

[7]   S. K. Rumi, K. Deng, and F. D. Salim, "Crime event prediction with dynamic features," EPJ Data Sci., vol. 7, no. 1, 2018, doi: 10.1140/epjds/s13688-018-0171-7.

[8]   K. A. Bokde, T. P. Kakade, D. S. Tumsare, and C. G. Wadhai, "Crime Detection Technique Using Data Mining and K - Means," vol. 7, no. 02, pp. 223–226, 2018.

[9]   E. Ahishakiye, E. O. Omulo, D. Taremwa, and I. Niyonzima, "Crime prediction using Decision Tree (J48) classification algorithm," Int. J. Comput. Inf. Technol., no. May, 2017.

[10]   T. T. Nguyen, A. Hatua, and A. H. Sung, "Building a Learning Machine Classifier with Inadequate Data for Crime Prediction," J. Adv. Inf. Technol., vol. 8, no. 2, pp. 141–147, 2017, doi: 10.12720/jait.8.2.141-147.

[11] M. S. Vural, M. Gok, and Z. Yetgin, "Generating incident-level artificial data using GIS based crime simulation," 2013 Int. Conf. Electron. Comput. Comput. ICECCO 2013, no. May 2017, pp. 239–242, 2013, doi: 10.1109/ICECCO.2013.6718273.

[12] M. S. Vural and M. Gök, "Criminal prediction using Naive Bayes theory," Neural Comput. Appl., no. September, 2017, doi: 10.1007/s00521-016-2205-z.

[13] A. Awal, J. Rabbi, S. I. Hossain, and M. M. A. Hashem, "Using Linear Regression to Forecast Future Trends in Crime of Bangladesh," no. May, 2016, doi: 10.1109/ICIEV.2016.7760021.

[14] T. Almanie, R. Mirza, and E. Lor, "Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots," Int. J. Data Min. Knowl. Manag. Process, vol. 5, no. 4, pp. 01–19, 2015, doi: 10.5121/ijdkp.2015.5401.

[15] A. Ristea and M. Leitner, "Integration of social media in spatial crime analysis and prediction models for events," CEUR Workshop Proc., vol. 2088, 2017.

[16]   U.K. Crime data, https://data.police.uk/data/.

[17]   https://en.wikipedia.org/wiki/Data_analysis

[18]   H. K. R. Toppireddy, B. Saini, and G. Mahajan, "Crime Prediction & Monitoring Framework Based on Spatial Analysis," in Procedia Computer Science, 2018, vol. 132, pp. 696–705, doi: 10.1016/j.procs.2018.05.075.

[19]   Mittal M, Goyal LM, Sethi JK, Hemanth DJ (2019) Monitoring the impact of economic crisis on crime in India using machine learning. Comput Econ 53(4):1467–1485.

[20]   Kadar C, Pletikosa I (2018) Mining large-scale human mobility data for long-term crime prediction. EPJ Data Sci 7(1):26.