RESEARCH ARTICLE OPEN ACCESS

Crime Prediction and Categorization Using Machine Learning Models: Comprehensive Examination of Relevant Literature

Maitanmi Stephen.O, Ariyo Olayemi Abimbola, Izang A.A
Computer Science, Babcock University, Ilisan,Ogun

ABSTRACT
Due to terrible events Every nation's legal system and security measures are vulnerable to criminal activity. Because crime analysis uses the obtained spatial and temporal data to establish the time and location, it consequently has increasing significance. To determine the exact time and location of the incident, however, outdated methods including statistical analysis, analytical "prediction", and documentation is not very successful. Numerous scholars have conducted in-depth study on crime prediction using machine learning. This study investigates several forms of criminal forecasting and evaluation utilizing multiple machine learning procedures that take into consideration the percentage of the previous work's accuracy measure, with the goal of providing a thorough review for integrating these algorithms in crime prediction. In addition to supporting future research focused at developing these techniques for crime analysis with the categories, problems, and prediction systems, it is anticipated that this systematic review will be helpful in introducing these approaches to crime researchers. Thirty chosen machine learning papers that forecast crime, enumerate the relevant knowledge, and emphasize the major limitations found during the study process are the main focus of our attention. This increases the effectiveness of crime prevention while guaranteeing safety and security.
Keywords: —Machine learning; Criminal analysis; Methodologies; Crime prediction; Spatial and temporal data.

I. INTRODUCTION

The task of predicting crime is challenging and involves complex tools for evaluation to bridge the inadequacies in today's detecting systems. Thanks to the advancement of current technology and the increasing amount of crime data available, researchers now have a unique opportunity to investigate and study crime detection using "machine learning" and deep learning methodologies. In order to predict future patterns of 'crime', data related to crime has been analyzed using machine learning procedures. (Kim, S., et al. 2018). For example, algorithms that can accurately 'predict' patterns after being trained on crime data from certain cities of crime include support vector machines, random forests, and decision trees (Raza, D. and Victor, D. 2021). Not only can these algorithms predict patterns of crime, but they can also present insightful information on emerging trends and patterns in crime. By distributing resources and tactics wisely, these qualities enable the successful fight against crime. Additionally, according to Elluri, L., et al. (2019), The location, the weather, and the time of day are examples of demographic and environmental variables that can be utilized to determine the connection between crime rates and machine learning algorithms. With the help of this data, community-specific strategies for crime prevention and prediction can be developed. According to Meijer and Wessels (2019), one significant use of 'predictive policing' uses machine learning to forecast crime. The use of analytics and statistics to combat crime and strengthen police enforcement rates is known as "predictive policing". To pinpoint crime hotspots and forecast future crimes, algorithms that employ machine learning can be used to analyze crime data from a certain geographic area, such as a city or neighborhood. By using this information to concentrate resources where they are most needed, law enforcement operations can be made more successful. This study gives a thorough review of current developments in the field and sheds light on possible uses of 'machine learning' in the anticipation of crime. This study benefits the larger research community by emphasizing the promise of these models and the issues that need to be resolved. Our comprehension of machine learning's function in crime prediction is improved by it. Thus, the following are the main contributions of this work:
- First, it presents a compilation of previous research on neighborhood crime detection that used cutting edge machine learning and deep learning techniques. Furthermore, the study outlined obstacles and suggested avenues for further investigation to close the current knowledge gaps about neighborhood crimes. so logically posing future research goals and/or queries for the scientific community to investigate further.

II. RELATED LITERATURE

Exploring diverse machine learning techniques and algorithms for crime prediction is the goal of this study. In an effort to better understand the contemporary approach and enable upcoming study into the development of more precise and high-performing crime-fighting models, our findings are provided alongside the challenges raised by the researchers. This section discusses and analyzes prior research works that have been related to the topic. These research works vary widely; some focus on applying artificial intelligence (AI) to

crime data, while others use machine learning or data mining—two AI subfields—to forecast and predict violent crimes. Some of these research works also use spatial and temporal data. Methods for spatiotemporal crime hotspot detection and prediction were reviewed by Butt, U. et al. (2020). Because so much data is being collected and made available to the public, the researchers assert that it is now simpler to initiate and continue additional study on the topic of crime and criminal investigation. Historical data makes it possible to forecast future crimes, and its increasing interest in developing meaningful machine learning models to help identify unique traits associated with crime prediction (Butt, M. t., al. 2021). Another systematic analysis was conducted in 2020 that looked at 32 articles related to geographic crime forecasting from 2000 to 2018. The four most effective strategies that have been suggested, as well as baseline techniques utilized among the 32 selected publications are summarized numerous times in this study, along with surveying table specifics regarding research location and time, crime statistics, and forecasting information. The study's discussion of the benefits, drawbacks, dangers, and prospects of the chosen papers resulted in a conclusion that algorithms' capabilities should not be ignored in the future (Kounadi O, et al., 2020). The classification of research papers according to the data mining methodology employed resulted from a thorough analysis of data mining and crime prediction studies made between 2004 and 2018. A gap was discovered in each of the 40 research articles that were analyzed, depending on the questions stated and the number of articles for each approach. This disparity suggests that the overall system performance significantly decreases with increasing dataset sizes (Falade A. et al., 2019). In order to anticipate crime in smart cities, Kawthalkar I. et al. (2020) looked at a number of mapping technologies. The writers performed a comparative analysis, accounting for different representations of criminal behavior. Although a lot of techniques and concepts for crime prediction, according to the authors, have been created, field testing is necessary to make sure that such approaches are workable. Additionally, data mining strategies for crime prediction based on a variety of criteria, such as regional, demographic, spatial-temporal, and socioeconomic characteristics, are examined in a survey done by Saravanan P. et al. (2021)..

III. RESEARCH METHODOLOGY

In the first stage of the methodology, 30 pertinent studies that employ machine learning to predict crimes models are gathered and analyzed; in the second stage, a classification table of each study, the results of several algorithms, the accuracy attained, and a comparison of them are given. Finally, constraints and further research. The publications under review are studies on crime prediction that span the years 2018 through 2022.

A. Machine Learning-Based Crime Prediction

It has been shown that conventional machine learning algorithms are successful in predicting crimes. To find trends that can be used to anticipate criminal activity, crime data has been examined using a range of models, such as random forests, logistic regression, decision trees, and support vector machines, and others. While deep learning necessitates vast quantities of data and intricate neural architecture, typical 'machine learning' models need less data and are simpler to examine. For instance, utilizing characteristics like the place and the time of day, and local demographics, one might utilize a logistic regression model to forecast the chance of a particular kind of crime happening. (Varun, M., et al., 2023). One possible use for a decision tree model is used to determine which critical components that result in the commission of a certain crime. By examining numerous attributes, models such as Random Forests (RF) can be used to predict criminal trends. In addition to these techniques, outlier analysis and anomaly detection in criminal data can also be performed using traditional machine learning models. Law enforcement organizations can identify possible illegal activity and take action to stop it by spotting odd trends or outliers in the information.

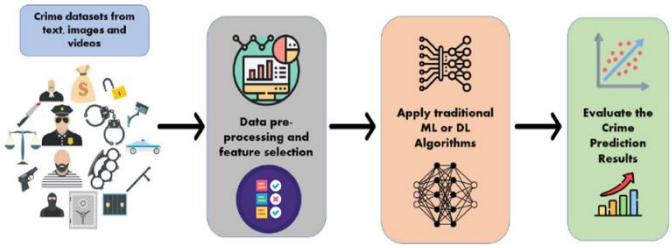

Fig. 1 Architecture flow of Crime Prediction

The process of predicting crimes using machine learning contains multiple crucial steps, as seen in Figure 1. The next step is to gather pertinent data, like crime rates, weather patterns, and demographics. The next stage is known as data preparation, and it entails preparing the data for usage by cleaning and formatting it. In order to create and assess models, training and testing data sets are divided after data preprocessing. The process of choosing significant attributes for the model to be trained on is known as feature engineering, and it comes after. The data can be exposed to various machine learning techniques for training and prediction once the features have been selected. Several performance indicators are used in the evaluation process to assess the trained models' precision and efficiency in forecasting criminal activity. The outcomes have the potential to improve making decisions for programs that prevent crime and enforce the law.

i); ***Stages of Machine Learning***

To efficiently employ 'machine learning models' for crime prediction. The obtained dataset must pass through a few steps before the results will be evaluated. To effectively employ machine learning techniques to make the forecast of crimes, the collected dataset needs to go through a few stages before the results are assessed.

[1] Data Collection

Data collection refers to the procedure of obtaining and estimating data from multiple, discrete sources. We can comprehend the past history of past events by gathering data, which we may then analyze to identify structured patterns. These patterns enable us to build prediction models that identify trends and project future changes using machine learning techniques. Appropriate ways for gathering data are required to build well-functioning technique. The data ought to contain correct information that is pertinent to the current work (Wang, Z. & Wang, J. 2021).

[2] Data Preprocessing

Pre-processing data is the procedure that of transforming unprocessed data into a human-readable format. This phase is critical since raw data is insufficient for machine learning to function. Prior to utilizing machine learning algorithms on the data, the data quality should be preserved. Libraries for Python are preconfigured to perform specific tasks. Importing the required libraries is one of the prerequisites for machine learning data pre-processing. The project made use of the following essential Python libraries: DESlib, SKlearn, Statsmodels, Folium, NumPy, Pandas, Matplotlib, Plotly, and Folium (Cruz R. et al., 2020). Python was utilized to import all of the datasets in the.csv file type utilizing the read_csv() method.

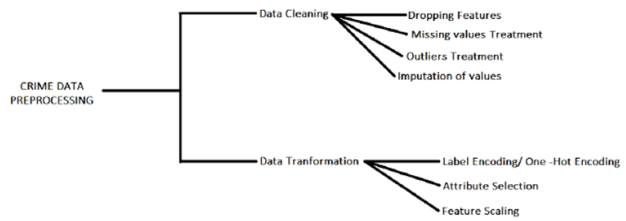

Fig. 2. Various data pre-processing activities

[3] Splitting the dataset

The supplied dataset is split into two sets using the train-test split, which separates it into train and test sets based on how well each machine learning algorithm works (Brownlee, J. 2020). The initial subset that fits the model is called the training dataset. Rather than being used for training, the second subset—known as the test dataset—is the component that the model uses as input. After then, projections are created and compared to the anticipated values (Rácz et al., 2021). 80% of the data in the test-train split are usually classified as train data, and 20% are marked as test data. Two opposing circumstances need to be considered when dividing up a dataset: There will be more variance in the parameter computations with less training data. In addition, less testing data will result in a greater disagreement in the implementation statistic.

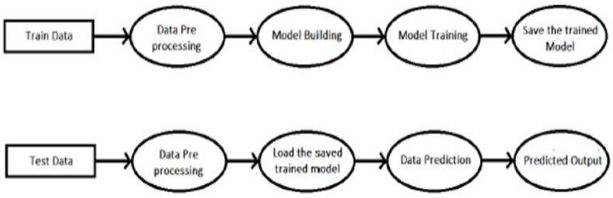

Fig. 3 Train and test steps

[4] Feature Selection

Feature selection is a crucial technique in predictive model development because it reduces the number of input variables. This is important because fewer features improve interpretability and expedite training, both of which reduce the spatial requirements of the model (Khaire & Dhanalakshmi, 2022; Li et al., 2017). By picking more significant features and removing redundant and unnecessary characteristics from our dataset, we were able to improve the test data's projected accuracy (Pilnenskiy & Smetannikov, 2020).

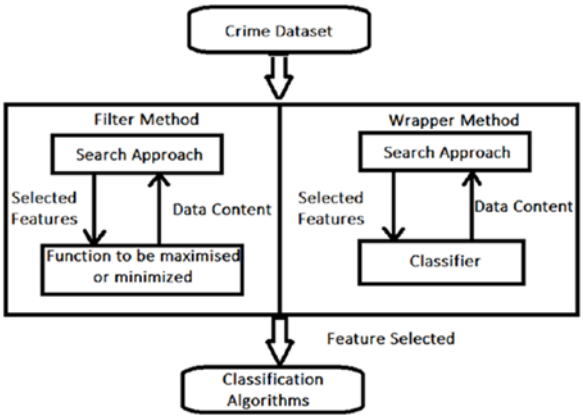

Fig. 4 Various Feature Selection techniques

[5] Evaluating the Performance

Utilizing metrics to evaluate an algorithm's performance is a fundamental part of any machine learning workflow. To show if progress is being made, these indicators employ a numerical representation. Every machine learning model, no matter how complex or basic linear, need a quantifiable value to be determined in order to assess its performance (Yu et al., 2022). The model's performance is monitored and assessed using metrics in both the training and test datasets. In machine learning, every problem is divided into two categories: regression tasks and classification tasks.

B. Classification of Crime Prediction Systems

The literature reviews of machine learning algorithms utilizing various datasets for various cities worldwide are listed in Table 1 of this section. Every study that was chosen was provided with crucial information that will help future researchers identify the most potent categories of crime prediction approaches. Tables include references, machine learning algorithm, source of utilized dataset, and accuracy of each approach based on specific dataset used for a certain city.

Table 1: An Overview on machine learning models used in crime prediction research.

Ref	Year	Methodology	Algorithm	Datasets	Perform (%)
[35]	2018	To forecast the time series data from theft instances that impacted the number of motorbikes, the ARIMAX approach is employed.	The ARIMAX Approach	The City of Yogyakarta	RMSE is 6.68
[2]	2021	Using geographical data to predict crime in New York City neighborhoods	SVM, RF, XGBoost	New York crime data	Accuracy - 43, 50, 52
[9]	2019	Crime statistics are used to identify and map crime-dense regions (CDRs) in Chicago and New York City. These hotspots are then exposed to ARIMA.	ARIMA, (RF), RepTree and ZeroR	Crime in New York City, from 2006 to 2016	2016 Are RMSE: 143.73, CDR 1 - 57.8, CDR 2 - 29.85, and CDR 3 - 16.19
[14]	2018	Using clustering, one may first comprehend the distribution around the city and subsequently forecast the number of crimes happening at each place.	Gradient boosting, LoR, and LR.	Russia's Saint Petersburg Crime (2014-2017)	R-square 0.9
[16]	2021	Using machine learning models and linear regression, crime prediction is used to lower the crime rate.	LR, NN	Baltimore city Data	Accuracy - 95, 94
[31]	2020	The location of a crime, specifically the site of a robbery, is predicted using four machine learning techniques.	Bagging, Extra Tree, DT, RF	Data on the City of Fortaleza	RMSE - 0.00231
[3]	2018	The Random Forest Regressor is used to estimate crime and assess the effects of urban locations. According to the results, unemployment and literacy rates are the two primary causes of crime.	Forest Random Regressor	Brazilian Public Health System's Department of Informatics	97% Accuracy, 80% Average Adjusted R-square
[25]	2021	Predicting Crime rates with Linear Regression	LR	Bangladesh crime data	Accuracy - 73.6%
[38]	2022	OVR-XGBoost: Rest against One OVO-XGBoost: One Vs One. The main distinction stems from the way the dataset is put up for categorization.	XGBoost	H City China: (2019)	Aggregate Accuracy - 85%
[30]	2021	Analytic forecasting and prediction of crime utilizing several machine learning models	MLP, NB, SVM, XGBoost, LR, DT, RF, and KNN	Chicago crime data	90, 66, 77, 87, 73, 66, 94, 88
[11]	2021	Classifying criminal reports using a rules engine and a classifier based on particle swarm optimization.	Multiple classification algorithms	USA, India & UAE Crime articles (2007 - 2017)	Aggregate Accuracy - 79%
[39]	2018	Estimating the chance of crime in the neighborhood using KNN, NB, and DT	KNN, NB, DT	Chicago crime data	Accuracy - 78. 64, 78
[22]	2022	After the XGBoost method was trained and tested using 17 spatiotemporal variables, SMAP was utilized to explain the model predictions.	XGboost model	Southeast China's ZG City 2017 - 2020	Accuracy - 89% & AUC - 0.586
[33]	2018	Grid-based crime forecasting with spatial characteristics	DNN, KNN, SVM	Taoyuan/ Taiwan	Accuracy - 83, 87, 88
[32]	2022	Several algorithms are evaluated using random under- and oversampling strategies.	LoR, RF, XGBoost and LightGBM	Danish national Psychiatric patient register, (474 crime).	F1 Score - 76%
[10]	2018	predicting crime using multiple datasets with CNN and RNN	RNN, CNN	Crime statistics for Chicago, Portland, and other cities	Accuracy - 74.1, 63.8, 72.7, 62.9
[20]	2022	Built on graph theory. Heuristics and decision trees are utilized in the feature selection process, and an	Multiple classification algorithms	Crime Articles: USA, UAE, and India (2008-2016)	F1 aggregate score: 88%

		ensemble classifier is used.			
[29]	2018	Decision tree classification algorithms for the prediction of crime.	KNN, Boosted decision	Vancouver police department	Accuracy 39.9, 43.2
[4]	2019	Determine whether crime is distributed equally across population densities or whether certain socioeconomic characteristics contribute to higher or lower crime rates.	Visual Analysis	London crime dataset	Aggregate Accuracy - 97%
[5]	2018	Utilizes an algorithm-as-a-service (AaaS) architecture and offers insights into public safety platforms and technologies that are currently in use.	K Nearest Neighbor (KNN)	Crime endogenous data sources	Aggregat Accuracy 90%
[23]	2018	Forecasting model using temporal and spatial data	DT, RF, Neural network	Chicago crime data	Accuracy - 38, 59, 81
[6]	2021	Geographical SATA analysis for crime prediction in New York City neighborhoods.	XGBOOST, RF and SVM	New York crime data	Aggregate Accuracy - 52%
[8]	2022	The EADT method is employed to provide Interpretable and Precise Crime Prediction.	Decision Tree (DT)	Pennsylvania state prisons	Aggregate Accuracy 77.6%
[1]	2022	Classification modeling techniques are used to forecast after crimes are clustered using traditional data science procedures.	K-Means	National Crime Records Bureau (India)	Aggregate Accuracy 78%
[27]	2020	Uses seasonal auto-regressive in each densely populated area to forecast the amount of crime incidences in the future using both spatial and temporal data and looks at HDBSCAN to identify hot zones that are more likely to become crime scenes.	HDBSCAN, SARIMA	NYC crimes 2008 - 2017	Aggregate MAE - 11.7
[37]	2021	A vast collection of real forensic casework pictures of drug-related crimes are subjected to two different machine learning classifier models.	Tree-CNN and SVM with Bag of Visual Words	Federated Police of Australia database of illicit drugs.	True posi rate - 89.
[21]	2022	ANN-based method for identifying tweets about crimes is suggested.	NN, SVM	Tweets	Accuracy - 90.33
[34]	2018	Construct scalable prediction models for violent crimes that happen 12–24 months after patients are discharged from safe mental health institutions.	Cox Regression	Swedish psychiatric hospital data 1992 - 2013	Index of aggregate concordance: 0.73
[24]	2022	Outlines a solution for the label scarcity issue in crime prediction using a framework for spatial-temporal self-supervised hypergraph learning (ST-HSL).	SVM, STResNet, DCRNN, STGCN, GWN, and ST-SLARIMA	NYC and Chicago crime data	MAE 0.79 for robbery
[17]	2018	KNN makes use of latitude and longitude. Furthermore, the naive bayes algorithm classifies the type of crime using date.	KNN & Naive Bayes	UK crime data 2015 - 2017	Accuracy 70 – 80%

C. *Categorized Crime Analysis Types*

Estimating the types and classifications of crimes, their time of occurrence, and the kind of crimes that the majority of academics have been discussing. Crime analysis can be done in a number of ways. Table 2 displays the categories for crime analysis employed in the papers under study. We divided the types into four main groups. Social media, criminal activity, and human behavior. Table 3 demonstrates that the most often used method, with 12 research articles employing this strategy, was assessing the neighborhood's crime density. The geographical analysis component was finished in ten (10) research publications. All told, there were three study papers: two were about social media, and one used applied behavioral analysis to forecast crime.

Table 2. Categorized Crime Types

General Category	*Analysis type*	*Freq.*	*Total*
Crime	Crime	6	12
	Neighborhood	4	
	Map crime dense region	2	
Spatial	Clustering	3	10
	Linear regression	3	
	Hot spot analysis	4	
	Temporal-spatial	3	
Human Behavior Analysis	criminal activity	1	3

	facial expression	1	
	point of interest	1	
Social Media	tweets	2	2

IV. DISCUSSION AND FUTURE WORK

The investigation's findings demonstrate that 'machine learning algorithms' in use today are capable of accurately and successfully predicting crimes. Additionally, using hybrid models to forecast crimes yielded positive outcomes. More researchers are being inspired by this to study the creation of hybrid models. K-Nearest Network, XGBoost, and Support Vector Machine are the most widely utilized algorithms for crime prediction among the thirty articles that were reviewed. Six studies employed the support vector machine model, seven papers the k-nearest network, and six publications the XGBoost. Hybrid models that integrated several machine learning approaches were used in the majority of the studies under review. Classification accuracy, which was employed in 20 investigations, is the most regularly utilized performance metric. AUC, F1-score, R squared, Root Mean Squared Error (RMSE), and MAE were among the additional measures that were employed in 2, 2, 2, and 2 studies, respectively. We discovered twenty different datasets that are used in crime prediction algorithms. New York, Chicago, and the United States had the highest frequency of usage of crime datasets in the selected publications. The research yielded positive results when using hybrid models to predict crimes. In the future, more study on hybrid model creation will be encouraged in an effort to increase accuracy. To ascertain the ethical implications of applying deep learning and machine learning to predict criminal activity, more investigation is required. Further study is necessary, especially in light of the focus on the possible privacy impact. Furthermore, future study attempts to enhance the detection of incident location by appending additional location data to tweets regarding crimes.

V. LIMITATION FROM THE REVIEW

The paucity of additional studies on the application of 'machine learning' to predict crimes in the real world contexts represents another important gap in the current assessment of research. Notwithstanding the enormous promise these technologies have demonstrated, more thorough assessments of their precision and effectiveness in practical situations are required. Additional research is also needed to determine how scalable these technologies are and what obstacles arise when integrating them into large-scale systems.

VI. CONCLUSION

The complexity of crimes is increasing in line with technology improvements, posing difficult challenges for law enforcement. Researchers have recently become more interested in utilizing machine learning to forecast crime, with a concentrate on identifying patterns and trends in the incidence of crimes. This study examined research on crime prediction from 2018 to 2023 from a number of angles, such as the kinds and classifications of crimes, the time period of the study, and the methodologies used. The paper reviews fifty studies that look into the various machine learning methods applied to crime prediction. In the end, the following findings arise from comparing strategies for machine learning in crime prediction systems: An algorithm's suitability for a given dataset may vary depending on its kind; for example, an image, text, video, or audio dataset may not yield the same results when applied to another. Particularly when used to identify spatiotemporal crime hot zones, the majority of machine learning models have shown to be extremely accurate in producing valuable information. Hybrid models that were utilized to forecast crimes also produced positive outcomes. Out of 20 research publications, the most frequently used performance parameter is classification accuracy. XGBoost, Support Vector and K-Nearest Neighbor estimation Machines are the most popular machine learning models.

REFERENCES

[1] R. O. Rana, P. Burnap, and N. Alsaedi, "Can we foresee a riot? Using Twitter for disruptive event detection", ACM Transactions on Internet Technology (TOIT), vol. 17, no. 2, pp. 1–26, 2017.

[2] M. Abdalsalam, C. Li, A. Dahou, and S. Noor, "A study of how textual features affect the GTD dataset's ability to predict terrorist attacks," Eng. Lett., vol. 29, no. 2, 2021.

[3] S. Chandra, M. Sharma, and P. Agarwal, "Comparison of machine learning techniques for terrorist attack prediction," in Proc. Twelfth Int. Conf. Contemporary Computing (IC3), 2019, pp. 1–7. [Online]. Available: www.ieee.org

[4] O. Ugbebor and M. Adeosun, "An empirical evaluation of jump diffusion models, symmetric and asymmetric, for the Nigerian stock market indices," African Scientific, vol. 12, no. e00733. [Online]. Available: www.elsevier.com

[5] P. Agarwal, M. Sharma, and S. Chandra, "Assessment of machine learning techniques in terrorist attack forecasting," in Proc. Twelfth International Conference on Contemporary Computing (IC3), 2019, pp. 1-7.

[6] R. Alhamdani, I. Sattar, and M. Abdullah, "Deep learning-based recommender system for worldwide terrorist database," Int. J. Mach. Learn. Comput., vol. 8, pp. 571-576, 2018.

[7] V. Arifin, F. Jallow, A. Lubis, R. Bahaweres, and A. Rofiq, "Predicting terrorists' terms with a deep learning model to plan attacks in real time twitter tweet from rapid miner," in 2022 10th Int. Conf. Cyber IT Serv. Manage. (CITSM), IEEE, 2022, pp. 1-6.

[8] K. Bedjou, F. Azouaou, and A. Aloui, "Detection of terrorist threats on Twitter using SVM," in Proceedings of the 3rd International Conference on Future Networks and Distributed Systems, 2019, pp. 1–5.

[9] R. Bridgelall, "Using natural language processing to categorize the desires stated by terrorists," Social Sciences, vol. 11, no. 1, p. 23, 2022.
[10] M. Benigni, K. Joseph, and K. Carley, "Online extremism and the communities that sustain it: Detecting and supporting community on Twitter," PLoS One, vol. 12, no. 12, p. e0181405, 2017.
[11] J. Brownlee, "Machine Learning Algorithms: Train-Test Split Evaluation," in The Mastery of Machine Learning, 2020.
[12] A. E. Cil, K. Yildiz, and A. Buldu, "Detection of DDoS Attacks with Feed Forward Based Deep Neural Network Model," Expert Systems with Applications, vol. 169, p. 114520, 2021.
[13] R. Cruz, L. Hafemann, R. Sabourin, and G. Cavalcanti, "DESlib is a Dynamic Python Library for Ensemble Selection," Machine Learning Research Journal, vol. 21, 2020.
[14] K. Chatterjee and H. M. Rai, "Hybrid Ensemble Technique and CNN-LSTM Deep Learning Model for Automatic Detection of Myocardial Infarction Using Big Electrocardiogram Data," Applied Intelligence, vol. 2021, no. 2, 2021.
[15] A. Canhoto, "Using Machine Learning from an Affordances Perspective to Combat Money Laundering and Terrorism Financing Globally," Business Research Journal, vol. 131, pp. 441–452, 2021.
[16] A. El Ali, T. Stratmann, S. Park, J. Sch¨oning, W. Heuten, and S. Boll, "Measuring, Understanding, and Classifying News Media Sympathy on Twitter after Crisis Events," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1–13.
[17] Y. Yin, Z. Li, Z. Hu, D. Wang, and Y. Feng, "A Terrorist Attack Casualty Prediction Technique Based on XGBoost," Complex & Intelligent Systems, vol. 6, no. 3, pp. 721–740, 2020.
[18] Ph. Garg, V. Ranga, and H. Garg, "Sentiment Analysis of the Uri Terror Attack Using Twitter," in Proceedings of the 2017 International Conference on Computing, Communication and Automation (ICCCA), 2017, pp. 17–20.
[19] Y. Gao, X. Wang, Q. Chen, Q. Yang, K.-Yang, and T. Fang, "Suspects prediction towards terrorist attacks based on machine learning," in 2019 5th International Conference on Big Data and Information Analytics (BigDIA). IEEE, 2019, pp. 126–131.
[20] A. Géron, "Hands-on Machine Learning with TensorFlow and Scikit-Learn: Ideas, Resources, and Methods to Create Intelligent Systems," 2nd ed. Sebastopol: O'Reilly Media, p. 856.
[21] X. Hu, F. Lai, G. Chen, R. Zou, and Q. Feng, "Quantitative research on global terrorist attacks and terrorist attack classification. Sustainability," 11(5), 1487.
[22] E. Huamaní, M. Alva, and R. Avid, "Using the Global Terrorism Database, Machine Learning Techniques to Visualize and Predict Terrorist Attacks Worldwide," 11: 562–570 in the International Journal of Advanced Computer Science and Applications. [CrossRef].
[23] M. Hao, J. Dong, F. Fangyu, D. Jingying, and C. Shuai, "Using GIS and the Random Forest Method to Simulate Spatial-Temporal Patterns of Terrorism Incidents on the Indochina Peninsula," Geo-Information Journal of ISPRS International 8: 133. [CrossRef].
[24] M. IrfanUddin, F. NazirZada, A. YousafSaeed, A. SyedAtif, M. A. Shah, A. Khasawneh, and M. Marwan, "Prediction of Future Terrorist Activities Using Deep Neural Networks," Hindawi Complexity, 2020, pp. 1-16. (www.hindawi.com)
[25] W. JSPM and K. Tirwa, "Machine learning-based predictive modeling of terrorist attacks," Int. J. Pure Appl. Math., vol. 119, pp. 49–61, 2018.
[26] W. JSPM and K. Tirwa, "Predictive modeling of terrorist attacks using machine learning," Int. J. Pure Appl. Math., vol. 119, pp. 49–61, 2018. (www.ijpam.eu)
[27] G. K¨onig, C. Molnar, B. Bischl, and M. Grosse-Wentrup, "Data processing for prediction relative feature importance," in 2020 25th International Conference on Pattern Recognition (ICPR), pp. 9318–9325.
[28] U. M. Khaire and R. Dhanalakshmi, "Reviewing the stability of the feature selection algorithm," King Saud University's Journal of Computer and Information Sciences, vol. 34, issue 4, 2022. [10.1016/j.jksuci.2019.06.012] is the DOI link.
[29] M. Kokane, S. Saurav, V. Bhairu, C. Kshitij, and B. Kanojiya, "Detecting online spread of terrorism on twitter using machine learning," International Journal of Engineering Research Technology (IJERT), vol. 11, 2022.
[30] S. Shrabanee, S. Debabrata, and N. Mishra, "A Cloud-Based Intelligent Framework for Examination of Terrorism-Related Activities," in New Directions in Management and Decision Science, A. W. H. Ip, M. Tavana, V. Jain, and S. Patnaik, Eds., vol. 1005, Singapore: Springer, 2020, pp. 225–235.
[31] X. Meng, L. Nie, and J. Song, "Predicting terrorist attacks using big data," Computers Electr Eng, vol. 77, pp. 120–127, 2019.
[32] N. Ouassini and A. K. Verma, "Demographic Conditions or Socioeconomic Inequality: A Micro-level Examination of Terrorism in Jharkhand," Victim Justice and Victimology Journal, vol. 1, pp. 63–84, 2018.
[33] O. A. Olabanjo, B. S. Aribisala, M. Mazzara, and A. S. Wusu, "An ensemble machine learning model for the prediction of danger zones: Towards a global counterterrorism," Soft Computing Letters, vol. 3, p. 100020, 2021.
[34] X. Pan, "Quantitative analysis and prediction of global terrorist attacks based on machine learning," Scientific Programming, 2021, pp. 1–15.
[35] K. Héberger, D. Bajusz, and A. Rácz, "Effects of train/test split ratios and dataset size on multiclass QSAR/QSPR classification," Molecular Structures and Dynamics, 2021.

[36] H. Rai, K. Chatterjee, and S. Dashkevich, "Using a novel hybrid unetresnext-50 deep CNN model, automatic and accurate abnormality detection from brain MRI images is achieved," Biomed Signal Processing Control, vol. 66, p. 102477, 2021.
[37] A. Sarker et al., "Improvised technique for analyzing data and detecting terrorist attack using machine learning approach based on twitter data," Journal of Computer and Communications, vol. 8, no. 7, pp. 50–62, 2020.
[38] N. Saiya and A. Scime, "Comparing classification trees to discern patterns of terrorism," Social Science Quarterly, vol. 100, no. 4, pp. 1420–1444, 2019.
[39] K. Srinivasa and P. S. Thilagam, "Crime base: Towards building a knowledge base for crime entities and their relationships from online newspapers," Information Processing & Management, vol. 56, no. 6, p. 102059, 2019.
[40] A. Yang, Y. Song, B. Chen, and N. Hou, "Oil prices and terrorist attacks: An analysis of time-varying causal relationships," [www.elsevier.com], pp. 123340–12350, 2022.
[41] F. Saidi and Z. Trabelsi, "A hybrid deep learning-based framework for modeling and predicting future terrorist activities," Egyptian Informatics Journal, vol. 23, no. 3, pp. 437–446, 2022.
[42] G. Tolan and O. Soliman, "An experimental study of classification algorithms for terrorism prediction," International Journal of Knowledge Engineering-IACSIT, vol. 1, no. 2, pp. 107–112, 2017.
[43] I. Thaseen, C. Kumar, and A. Ahmad, "Using an ensemble of classifiers and chi-square feature selection to create an integrated intrusion detection model," Arab. J. Sci. Eng., vol. 44, pp. 3357–3368, 2019.
[44] A. Zeb et al., "Deep neural network prediction of future terrorist activities," Intricacy, vol. 2020, pp. 1–16.
[45] M. Uddin et al., "Prediction of future terrorist activities using deep neural networks," pp. 1–16, 2020.
[46] A. G. Kissi and A. Ben, "Machine learning-based terrorist act prediction: Tunisia case study," in 17th International Multi-Conference on Systems, Signals Devices (SSD), 2020, pp. 398–403.
[47] G. Kant, C. Weisser, T. Kneib, and B. Säfken, "Topic model—Machine learning classifier integrations on geocoded twitter data," in Biomedical and other applications of soft computing, Springer, pp. 105–120.
[48] F. Llussá and J. Tavares, "Which fear at what price? Regarding the financial fallout from terrorist attacks," Letters on Economics, vol. 110, pp. 52-55. (www.elsevier.com)
[49] L. Luo and Q. Chao, "An analysis of the crucial indicators impacting the risk of terrorist attacks: A predictive perspective," Safety Science, vol. 144, 2021, p. 105442. [CrossRef]
[50] L. Luo and C. Qi, "The tendency of terrorist organizations to explosive attacks: An institutional theory Perspective," Frontiers in Psychology, vol. 13, 2022, p. 747967.
[51] A. Iftene, M. Dudu, and A. Miron, "Scalable system for opinion mining on twitter data. dynamic visualization for data related to refugees' crisis and to terrorist attacks," 2017.
[52] M. Varun, E. Lavanya, V. Piyush, and R. Nirmalya, "A Systematic Review and Future Prospects of Crime Prediction Using ML and DL," ID for Digital Object, 2023. [Online]. Available: 10.1109/ACCESS.2023.3286344.
[53] Z. Wang and J. Wang, "Utilizing Machine Learning for Resource Management and Public Security Information," Science and Technology, 2021. [Online]. Available: 10.1155/2021/4734187.
[54] R. Chen, K. K. Lai, L. Yu, and R. Zhou, "One-Hot Encoding or Imputation for Preprocessing Missing Data for Credit Classification?" Finance and Trade in Emerging Markets, vol. 58, no. 2, 2022. [Online]. Available: 10,1080/1540496X.2020.1825935.
[55] Y. Zhou, "Assault detection and network analysis of pro-ISIS fanboys using Twitter data," in 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), IEEE, 2017, pp. 386–390.