2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS)

Fostering Global Stability: Machine Learning for Anticipating Terrorism

Archana Naik
Department of Computer Science and Engineering
Nitte Meenakshi Institute of Technology
Bangalore,India
archana.naik@nmit.ac.in

Ushashree P
Department of Computer Science & Engineering
Geethanjali College of Engineering & Technology
Hyderabad, India
ushashree.sgs@gmail.com

Basavaraj Muddapu
Department of Computer Science & Engineering
Nitte Meenakshi Institute of Technology
Bangalore,India
1nt19cs049.basavaraj@nmit.ac.in

Uttkarsh Dutt Guttedar
Department of Computer Science & Engineering
Nitte Meenakshi Institute of Technology
Bangalore,India
1nt19cs205.uthkarsh@nmit.ac.in

Muddam Venkatesh
Department of Computer Science & Engineering
Nitte Meenakshi Institute of Technology
Bangalore,India
vv860563@gmail.com

***Abstract*— The threat of terrorism remains a significant global concern, with governments and security agencies continually seeking advanced methods for mitigating and preventing attacks. Machine learning has emerged as a powerful tool for predictive analysis, offering the potential to anticipate and prepare for potential terrorism incidents. This paper aims to develop a machine learning model that can predict the likelihood of a terrorism attack based on historical data and relevant features. The results of the study will offer valuable insights into the potential application of machine learning to enhance counterterrorism efforts. It's important to emphasize that terrorism is a multifaceted issue and should be seen as a supplementary tool rather than a comprehensive solution.**

Keywords*— Decision Tree, KNN, Naïve Bayes, Random Forest, Regression, Support Vector Classifier*

I. INTRODUCTION

The global spread of terrorism is a growing concern. According to the United Nations, terrorism is characterized as an act driven by political motives and intended to cause significant harm or loss of life. The factors contributing to violence evolve over time, influenced by a range of political and social factors. Identifying the actual responsibility for an attack and understanding its underlying motives can be challenging. The prevalence of violent behavior remains unclear, and the current analysis approaches either take the form of case studies or employ diverse methods like regression analysis. Case studies have limitations in their applicability to specific situations, while interviewing affected individuals restricts the scope of the analysis. Estimates often rely on factors such as the weapons used in the attack and the number of casualties. Other types of analysis involve investigating unusual behavioral patterns or interviewing incarcerated individuals for crime-related insights.

This paper proposes a machine learning (ML) -based approach to analyzing terrorism at both regional and national levels. By leveraging the Global Terrorism Database (GTD) and employing supervised ML models, this study aims to extract valuable insights regarding patterns of terrorist behavior. With countless lives and resources lost worldwide due to terrorism, it is intriguing to observe that some organizations occasionally claim responsibility for such attacks, while in other instances, no group takes credit. The primary objective of this research is to develop a robust tool by analysing past events that can aid in forecasting future terrorist incidents and provide more concrete insights.

II. LITERATURE SURVEY

Global stability is very much essential and necessary in the networked and independent world. It addresses global peace, security and equilibrium. There are various reasons that play a crucial role in the maintenance of stability across the nations. Essential ones are political cooperation, environmental protection and preventing terrorism. This paper makes an attempt to address the problem of preventing terrorism by predicting it in advance. Efficient and accurate outcomes from ML classifiers are crucial to the faster response and subsequent analysis of terrorist attacks. This study focuses on enhancing classifier performance by utilising hybrid ML algorithms to improve accuracy.

The work in [1] provides methods for constructing basic classifiers such as Naïve Bayes, K-Nearest Neighbour, Decision Trees, Support Vector Machines, and Multi-Layer Perceptron. These basic classifiers were then combined to create hybrid models, and their performance was compared against the basic classifiers. The study revealed that optimal performance for all classifiers was achieved with a more balanced class through an improved resampling rate. Based on the findings, it was concluded that hybrid classifiers outperformed basic classifiers. The work in [2] explores how ML techniques and other advanced statistical techniques can be combined with field expertise in the social sciences to discover patterns that are difficult to learn from vision alone. Although the authors investigated methods that could generate only small samples and groups for counterterrorism, they found that the operation of classification studies could be improved by better data collection for practice data in the field.

Terrorist attacks have a devastating impact on humanity worldwide and require immediate attention. With the abundance of specific terrorist data, accurately predicting the responsible terrorist group and their tactics becomes a complex challenge. Therefore, the work in [3] focuses on predicting the terrorist organizations accountable for attacks

979-8-3503-1545-5/23/$31.00 ©2023 IEEE

that took place in Turkey during 2016, utilising accessible and effective ML techniques. A possible attempt has been made by the [4] to curb the damage both in terms of materialistic, economic and human life because of terrorism using machine learning algorithms that learn from localized news data in order to predict the likelihood of attack.

Governments have the potential to enact legislation aimed at reducing the economic burden of violence by tackling the root causes of war and terrorism. The articles featured in this special issue, titled Strategic and Experimental Approaches to the Study of Conflict and Terrorism, employ applied game theory or experimental economics to analyze various aspects related to conflict or terrorism. This introduction provides an overview of the research papers included in the special issue and outlines their interconnections [5]. The work in [6] illustrates the unexpected link between action in support of terrorist activities and conservative religious conviction. Similar work [7] provides the adverse effects on the economy and international trade due to terror attacks. The authors describe the influence of terrorism on the political environment of the state. The study also examines implications for partisanship, governance, and terrorism, providing insights into these areas [8].

The article referenced in [9] examines significant incidents in civil aviation that have influenced the evolution of aviation security policies. It provides an overview of these incidents, which have played a pivotal role in shaping the development of aviation security measures. The article discusses threats and security breaches in the industry, such as hijackings and acts of terrorism, and explores the corresponding countermeasures and policy decisions made in response to these challenges. Predicting the responsible terrorist organisation and identifying patterns of indoctrination based on factual information is challenging due to the vast amount of comprehensive terrorist data available. The ML-based study was performed to identify the organisation behind a particular attack. Five ML algorithms were implemented and compared: Naïve Bayes (NB), K-Nearest Neighbour (KNN), Tree Induction (C4.5), Iterative Dichotomiser (ID3), and Support Vector Machine (SVM). The study utilises real data from the Global Terrorism Database (GTD) provided by the National Consortium for the Study of Terrorism and Responses to Terrorism (START). Crime, characterised by deviating from moral norms and causing harm and losses to others, poses a significant burden on society.

The work in [10] states to get initial insight about the dataset and understand the correlation between various attributes. Based on the insights received from the analysis, they have made use of an efficient ML algorithm to analyse the patterns of crime. An attempt at using unsupervised and semi-supervised algorithms has been made in [11]. They have focused on the quality of the data, specifically missing values, and clustering techniques that help analyse the pattern of crime. The prediction of terrorist groups using factual data has been a less explored area due to the limited availability of comprehensive terrorism data that encompasses both attacks and the training activities of these groups. This scarcity of data can be attributed to the secretive and sensitive nature of such information. To address this challenge and forecast the responsible terrorist organisation for a given incident, this study introduces a model called the Terrorist Group Prediction Model (TGPM). It states that the model was trained to recognize commonalities and connections between different terrorist incidents, which enables it to make predictions about the group or organization responsible for those attacks. The model's overall performance demonstrates a reasonable amount of accuracy [12].

The exponential growth of data on the web has created a demand for effective approaches to identify, gather, analyse, and utilize valuable information for businesses and organizations. In response, the field of data mining, including web mining, has introduced various techniques and strategies for acquiring and managing crucial insights, leaving enterprises in search of clarity [13]. The study presented in [14] introduces an analytical model that focuses on time-dependent predictions of attacks. The possibilities of collaborative terror attacks are studied in [15]. A prototype was developed to predict the likelihood of the attack using the autoregressive integrated moving average (ARIMA) model. A mechanism to predict casualties in natural and human-made disasters is proposed in [16]. The back propagation neural network and the random forest algorithm are implemented. The model was basically used to predict the casualties and injuries that occurred during a calamity to support the emergency.

The literature study by the authors in [17] reveals the various causes for the promotion of terrorist activities. The major causes include poverty and limited political and civil freedom. Some of the evidence shared by the media also promotes fear and misconceptions about such activities. There are misconceptions such as that terrorist groups generally hold territory, the ineffectiveness of international treaties in counterterrorism, and the belief that failed states are conducive to future terrorist attacks. The perpetuation of these myths is attributed, in part, to empirical analyses not adequately addressing identification problems, leading to correlations rather than causal determinants being reported. It emphasizes the need for further research to dispel these myths and underscores the importance of addressing endogeneity concerns in establishing causal relationships.

When anticipating a terrorist attack, the government would formulate an action plan to prevent it. If it fails to prevent the attack, there is a need for a coordinated response to overcome its impact. In addition to addressing the physical and environmental damage, authorities must also monitor the mental well-being of individuals affected by such attacks. The study in [18] reveals that levels of mental health vary based on factors such as gender and occupation. The authors suggest that understanding the mental health status and planning preventive measures accordingly will contribute to a better recovery from the psychological impact of the attack. [19] presents a survey on terrorist activities using Twitter data. The authors highlight a shortage of datasets, which hinders the ability to conduct a precise analysis of the prediction model.

Leaders in counterterrorism operations face difficult and significant challenges. Decision-making processes are needed due to the diversity and complexity of today's attacks and the large amount of data that needs to be processed in real time. Most mathematical models are not suitable for emergencies. [20] has proposed a time-dependent terrorist attack analysis model (ATiPreTA) that is consistent with casualty data from two known major terrorist attacks in Tunisia: the Bardo and Soos attacks. The above said paper predicts the casualty count in different cases and compares it with the results obtained from the multi-agent model to refine the process.

The success of the fight against terrorism reflects the terrorist's capacity and lack of security, and the estimated average of the country's annual success rate is important to the government. [21] obtained data on crime in 146 countries from the Global Crime Database (GTD) from 2002 to 2020 and performed two stages of estimation. First is used to predict the average success rate of next year's terrorist attacks by combining the root causes of bad actions with the previous year's results. Second, the predicted output is described using identification methods including Local Effects (ALE) and SHApley Additive Description (SHAP) to provide a damaging countermeasure to countries around the world.

To predict the likelihood of an attack, only the data visible to the naked eye is sufficient. Understanding the possibilities of the attacks goes beyond a normal person's thought process. Taking advantage of machine learning algorithms and analytical tools will provide better support. In this paper, machine learning algorithms such as Random Forest, K-Nearest Neighbours, Linear Regression, Linear Discriminant Analysis, Decision Tree Classifier, Naive Bayes, Support Vector Classifier, and Logistic Regression are applied to terrorist attack data.

III. Proposed System

The proposed methodology for anticipating terrorism is presented in Fig. 1. The collected data is allowed to flow through the steps depicted to be able to build the model.

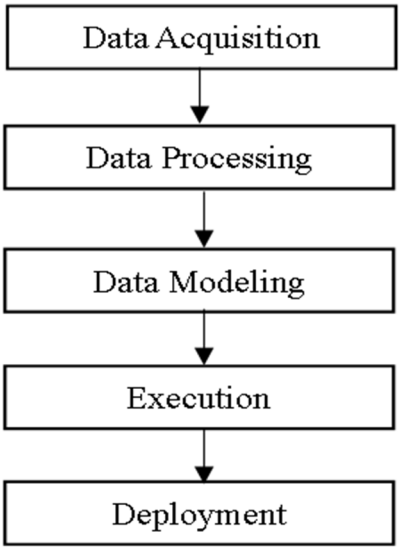

Fig 1: Development Steps

A. Data Acquisition

Data acquisition entails accumulating raw data from diverse sources. This encompasses identifying pertinent data sources, obtaining necessary permissions, and retrieving the data. Sources may include databases, APIs, files, sensors, or other data collection methods.

B. Data Processing

Once data is acquired, it undergoes processing and preparation for analysis. Data processing involves cleansing the data to eliminate inconsistencies, errors, or missing values. Additionally, it may involve transforming the data into a suitable format, aggregating or summarizing it, and handling outliers or anomalies.

C. Data Modeling

Data modeling involves creating a representation that captures the structure, relationships, and rules of the data. This can encompass designing a database schema, defining tables, columns, and constraints, and establishing relationships between data entities. Effective data modeling facilitates efficient organization and storage of data.

D. Execution

In this phase of machine learning, experiments are run, tests are conducted, and adjustments are made. The goal of the phase is to provide a refined result capable of providing the necessary facts for the machine to form opinions. This is done to value the required machine growth and maximize system performance.

E. Deployment

The operationalization and encouragement of machine learning (ML) efforts are crucial, similar to any other software development endeavor. Treating ML as an ad hoc question necessitates its integration into the decision-making system. It is recommended to seamlessly incorporate ML into the product, enabling machines to generate insights directly and reducing the reliance on additional exploratory methods. Various frameworks are available, and this project utilized Flask, a Python-based framework.

IV. Implementation

The prototype is implemented using Python. The software used to run the machine learning code is Jupiter Labs, and the deployment is done on the Flask framework. The project requires a computer with an Ubuntu or Windows Operating System and at least 4 GB of RAM.

A. Dataset

The Global Terrorism Dataset (GTD) provides comprehensive and accurate information on domestic, international, and transnational terrorist acts that occurred during the specified timeframe. Prior to developing the prediction model, we conducted exploratory data analysis to gain a high-level understanding of the dataset's attributes. The GTD draws data from various sources, including newspapers, 25,000 news sources, and a collection of open media sources, making it the world's largest declassified database on terrorist attacks. To predict the success of an attack, we utilized suitable supervised machine learning models and evaluated their performance. The selected attributes for prediction, based on feature importance scores, included the day of the attack, target type, attack type, and other relevant factors.

B. Exploratory Data Analysis

Exploratory Data Analysis (EDA) involves examining and understanding the characteristics, patterns, and relationships present in a dataset. It aims to extract meaningful insights and identify initial trends or patterns to guide further analysis and modelling. During EDA, various statistical and visualization techniques are employed to process the data. This includes calculating summary statistics, visualizing distributions and relationships between variables, identifying outliers or missing values, and assessing data quality. EDA helps in recognizing the structure of the data, detecting anomalies, and forming initial hypotheses for more in-depth analysis.

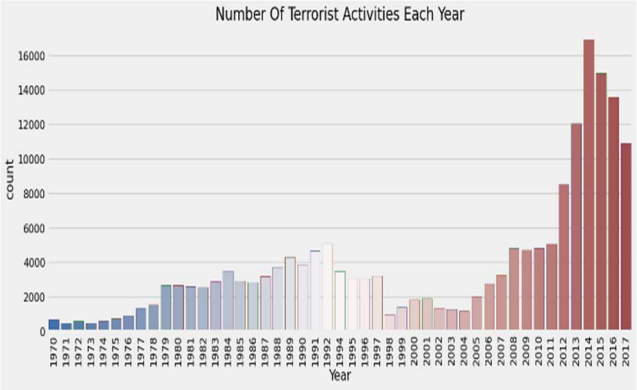

Fig 2. Count of Activities Each Year

Fig. 2 illustrates the occurrence of terrorist activities across different years. The data reveals that the highest number of activities were recorded in 2014, surpassing 16,000 incidents. Figure 3 represents the frequency of attacks that took place in various countries. According to the data, Iraq witnessed the highest number of attacks, with a staggering count of 24,636 incidents.

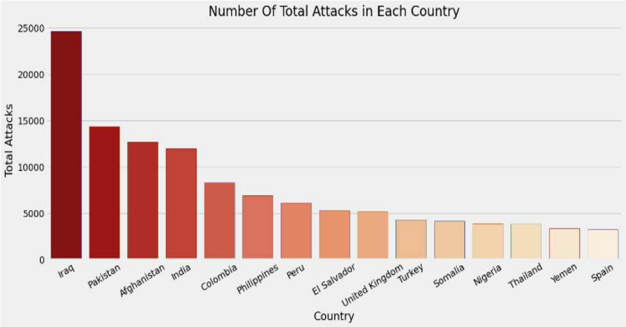

Figure 3. Number of Attacks

C. *Choosing the model*

The selection of machine learning models will prioritize their ability to efficiently handle large datasets, such as the Global Terrorism Dataset. These models should be capable of effectively managing missing and null values, employing techniques such as replacing missing values with mean values and utilizing other machine learning approaches to address these challenges.

D. *Evaluation*

The execution of the models was evaluated through testing and training processes. Notably, Random Forest exhibited a remarkable accuracy of 92%, which led to its consideration as the preferred model.

V. RESULT

To analyze the results and evaluate the performance, the measured accuracy (Acc), recall (R), and precision (P) are considered. The mathematical expressions are provided in (1), (2), and (3).

$$Accuracy = \frac{No.\ of\ True\ Positives + No.of\ True\ Negative}{Total\ No.of\ Cases} \quad -(1)$$

$$Recall = \frac{No.of\ True\ Positives}{No.of\ True\ Positives + No.of\ False\ Negatives} \quad -(2)$$

$$Precision = \frac{No.of\ True\ Positives}{No.of\ True\ Positives + No.of\ False\ Positive} \quad -(3)$$

Fig. 4 illustrates the performance of each of the models used in the study. The Random Forest model got the highest accuracy of 92%, and Naïve Bayes gave the lowest accuracy value of 25.10%.

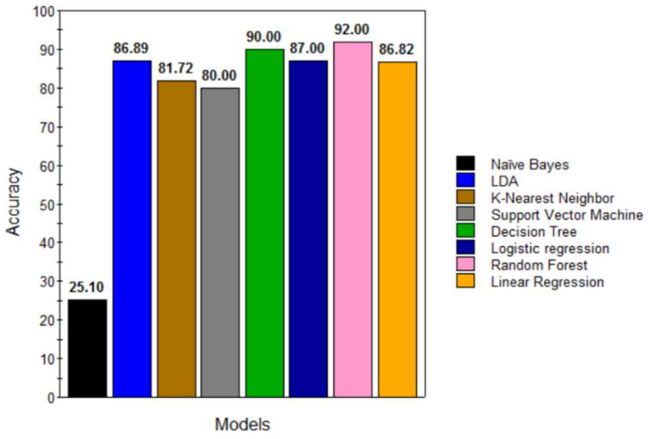

Fig 4. Model Comparison based on Accuracy

In the proposed system, it has been experimented with to check the performance of various machine learning algorithms that would curb or possibly help overcome the terrorist attack. In the phase of experimentation, not all the algorithms are considered, but the most commonly used algorithms are taken into account, such as Naive Bayes, KNN, SVM, and Decision Tree. The process of building and developing a model using the above algorithms was not so effective.

The second most widely used ensemble method, i.e., random forest, is considered, which provides better performance and accuracy in comparison with other techniques. Even the overfitting and underfitting issues are addressed by the ensemble method, and a balance between bias and variance is provided by using different subsets and features of the data. In Figure 4, it is evident that the ensemble method outperforms the other methods.

Table I shows the model comparison of algorithms trained and tested, showing accuracy, recall, and precision. In this paper not only the standard evaluation metrics such as Accuracy is considered but also other metrics such as Recall and Precision is taken into account to evaluate the performance of the model.

Fig. 5 depicts a visual representation called the Receiver Operating Characteristic (ROC) curve, which showcases the relationship between the true positive rate (TPR) and the false positive rate (FPR) in a binary classification model. The curve demonstrates how these rates vary when different threshold settings are applied.

TABLE I. RESULT BASED ON DIFFERENT METRIC

Algorithm	Evaluation Metrics		
	Accuracy	*Recall*	*Precision*
Random Forest	92.00	0.981858	0.934773
K-Nearest Neighbors	81.72	0.849773	0.933941
Linear Regression	86.88	0.8668800	0.868800
Linear Discriminant Analysis	86.89	0.868544	0.868544
Decision Tree	90.00	0.885621	0.885621
Naïve Bayes	25.10	0.251474	0.251474
Support Vector Machine	80.00	0.800000	0.800000
Logistic Regression	87.00	0.868865	0.868865

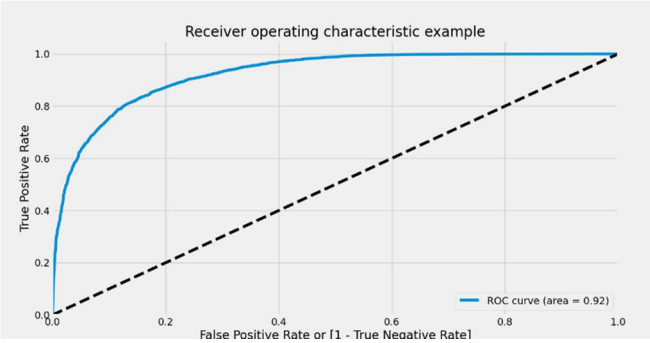

Fig 5. Receiver Operating Characteristic curve for Random Forest

VI. CONCLUSION

The use of random forests for terrorism prediction is a promising approach that leverages the power of ensemble learning and decision trees to analyze historical data and make predictions about future terrorism incidents. It offers valuable insights and a data-driven approach to enhance security and counterterrorism efforts. However, it should be integrated into a broader strategy that combines machine learning with expert knowledge and intelligence analysis for more comprehensive and effective prevention and response to terrorism. While Random Forest is a powerful tool, it is not without limitations. It may struggle with imbalanced datasets and can be computationally expensive, especially when dealing with extensive data.

REFERENCES

[1] Saidi, F., & Trabelsi, Z. (2022). A hybrid deep learning-based framework for future terrorist activities modeling and prediction. Egyptian Informatics Journal, 23(3), 437-446.

[2] Peng, A. (2018). An Integrated Machine Learning Approach To Studying Terrorism.

[3] Mohammed, D. Y., & Karabatak, M. (2018, March). Terrorist attacks in Turkey: An evaluate of terrorist acts that occurred in 2016. In 2018 6th International Symposium on Digital Forensic and Security (ISDFS) (pp. 1-3). IEEE.

[4] Bang, J., Basuchoudhary, A., David, J., & Mitra, A. (2018). Predicting terrorism: a machine learning approach. Predicting terrorism: a machine learning approach.

[5] Mathews, T., & Sanders, S. (2019). Strategic and experimental analyses of conflict and terrorism. Public Choice, 179(3-4), 169-174.

[6] Ravenscroft, I. (2020). Terrorism, religion and self-control: An unexpected connection between conservative religious commitment and terrorist efficacy. Terrorism and political violence, 32(8), 1819-1834.

[7] Khalifa, N. E. M., Taha, M. H. N., Taha, S. H. N., & Hassanien, A. E. (2019, March). Statistical insights and association mining for terrorist attacks in Egypt. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 291-300). Cham: Springer International Publishing.

[8] Wheatley, W., Robbins, J., Hunter, L. Y., & Ginn, M. H. (2020). Terrorism's effect on Europe's centre-and far-right parties. European Political Science, 19, 100-121.

[9] Klenka, M. (2019). Major incidents that shaped aviation security. Journal of transportation security, 12(1-2), 39-56.

[10] 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.

[11] Malathi, A., & Baboo, S. S. (2011). Evolving data mining algorithms on the prevailing crime trend–an intelligent crime prediction model. Int J Sci Eng Res, 2(6).

[12] Sachan, A., & Roy, D. (2012). TGPM: Terrorist group prediction model for counter terrorism. International Journal of Computer Applications, 44(10), 49-52.

[13] Thuraisingham, B. (2003). Web data mining and applications in business intelligence and counter-terrorism. CRC Press.

[14] Kebir, O., Nouaouri, I., Rejeb, L., & Said, L. B. (2022). ATiPreTA: AN Analytical Model for Time–Dependent Prediction of Terrorist Attacks. International Journal of Applied Mathematics and Computer Science, 32(3), 495-510.

[15] I Premkumar, S., Monish, V., Kumar, S., & Saravanan, M. PREDICTION ACCURACY OF TERRORIST ATTACK USING ML. Journal homepage: www. ijrpr. com ISSN, 2582, 7421.

[16] I Hu, X., Hu, J., & Hou, M. (2022). A two-step machine learning method for casualty prediction under emergencies. Journal of Safety Science and Resilience, 3(3), 243-251.

[17] Gaibulloev, K. and Sandler, T., 2023. Common myths of terrorism. Journal of Economic Surveys, 37(2), pp.271-301.

[18] Wesemann, U., Applewhite, B. and Himmerich, H., 2022. Investigating the impact of terrorist attacks on the mental health of emergency responders: systematic review. BJPsych open, 8(4), p.e107.

[19] Leenuse, M.L. and Pankaj, D.S., 2023, May. Detection and Prediction of Terrorist Activities and Threatening Events in Twitter-A Survey. In 2023 International Conference on Control, Communication and Computing (ICCC) (pp. 1-6). IEEE.

[20] Kebir, O., Nouaouri, I., Rejeb, L., & Ben Said, L. (2022). Atipreta: An analytical model for time-dependent prediction of terrorist attacks. International Journal of Applied Mathematics and Computer Science, 32(3).

[21] Luo, L., Li, B., Wang, X., Cui, L., & Liu, G. (2022, December). Interpretable machine learning-based terrorist attack success rate prediction. In 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (pp. 1-6). IEEE.