



International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

2020 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 2nd Jan 2021. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-12](http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-12)

DOI: 10.48047/IJIEMR/V09/I12/153

Title: **A MACHINE LEARNING APPROACH FOR ENHANCING DEFENCE AGAINST GLOBAL TERRORISM**

Volume 09, Issue 12, Pages: 892-897

Paper Authors

NOOREEN BANU, SAMALA NAVYA, R PRAVALIKA, K.SONY



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

A MACHINE LEARNING APPROACH FOR ENHANCING DEFENCE AGAINST GLOBAL TERRORISM

NOOREEN BANU¹, SAMALA NAVYA², R PRAVALIKA³, K.SONY⁴

^{1,2,3} B TECH Students, Department of CSE, Princeton Institute of Engineering & Technology For Women, Hyderabad, Telangana, India.

⁴ Assistant Professor, Department of CSE, Princeton Institute of Engineering & Technology For Women, Hyderabad, Telangana, India.

ABSTRACT: The objective of this work is to predict the region and country of a terrorist attack using machine learning approaches. The work has been carried out upon the Global Terrorism Database (GTD), which is an open database containing list of terrorist activities from 1970 to 2017. Six machine learning algorithms have been applied on a selected set of features from the dataset to achieve an accuracy of up to 82%. The results suggest that it is possible to train machine learning models in order to predict the region and country of terrorist attack if certain parameters are known. It is postulated that the work can be used for enhancing security against terrorist attacks in the world.

Keywords: Terrorism; prediction; machine learning; accuracy.

I. INTRODUCTION

Terrorist attacks are spreading on a great pace across the world. As per the United Nations definition of Terrorism, "any action with a political goal that is intended to cause death or serious bodily harm to civilians"[1]. In the last year, around 22 thousand events occurred globally, causing over 18 thousand casualties [2]. The factors leading to terrorism change over time since they are dependent upon multiple political and social reasons. Apart from predicting reason behind the attack, identification of the responsible agencies is also difficult. There has been a dearth of the information regarding patterns of widespread terrorist behavior. The existing analyses are either case studies or use of quantitative methods such as regression analysis. The former of these is specific to certain events, while the latter approach is restricted to interviews of civilians impacted by the attack. Most of these

analyses depend on factors such as weapons used for the attacks and number of people harmed. Other type of analysis includes investigation of unusual patterns in individual behaviors or questioning detainees to acquire data pertaining to the attacks. The current research is focused on finding out the correlation between terrorism and its causal factors. Existing efforts have not been good enough for prediction. Machine learning approaches can aid in predicting the likelihood of a terrorist attack, given the required data. The results of this work can help the security agencies and policy makers to eradicate terrorism by taking relevant and effective measures. This paper provides an approach to analyzing terrorism region and country with the machine learning techniques and terrorism specific knowledge to fetch conclusions about terrorist behavior patterns. Through analysis of events using GTD, six supervised machine

learning models (Gaussian Naïve Bayes, Linear Discriminant Analysis, k-Nearest Neighbors, Support Vector Machines, Decision Tree, and Logistic Regression) were built and evaluated on their performances.

II. Related work

Prediction of terrorism activities is an important area of concern for researchers. . The large number of events makes it difficult to predict terrorist group responsible for some terrorist activity. The work in [1] has tested machine learning approaches for classifying and analyzing global terrorist activity. The authors have explored supervised machine learning approaches to study terrorist activity, and then developed a model to classify historical events in Global Terrorism Database. They have released a new dataset as well named QFactors Terrorism, which collaborates event-specific features derived from the GTD with population-level demographic data from sources like United Nations and World Bank. Naive Bayes, decision trees, Linear Discriminant Analysis, k-nearest neighbors and random forest approaches have been implemented. Random forest model has been successful in classifying the reason responsible for an identified incident with up to 68% accuracy after being trained. An evaluation of terrorist acts that occurred in 2016 is presented in [2]. In this paper, the authors have taken into consideration the data of terrorist attacks that occurred in Turkey in 2016. They have used data mining techniques in order to detect the most useful machine learning algorithm. WEKA tool has been used for analysis and approaches used are J48, Bayes Net, Support Vector Machines, k-Nearest Neighbours and Naive Bayes. The lowest

accuracy came out for KNN although it was good in other measures [4]. The work in [5] has used machine learning approaches for the prediction of terrorist attacks. The report has emphasized that future research should focus on explaining non linear effects of variables such as assassinations and all over other parameters such as democracy and economic institutions. Through their approach, the policy makers would be able to predict terrorist attacks and how to deter them and will also help to allocate efficient terrorism fighting resources[5]. The work in [4] lists the root causes of conflict and terrorism in order to help policymakers to specify measures for reducing costs corresponding to violence. They have taken into consideration the tools of applied game theory or experimental economics which in order helps in analyzing issues related to conflict or terrorism. A collection of studies related to this area, explaining the relationships between them is listed in [6]. An unexpected relation between conservative religious commitment and terrorist activism is illustrated in [7]. Jihadist terrorism, Baruch Goldstein's 1994 attack in Hebron, Christian identity groups in the US, and Aim Shinrikyo in Japan has been considered as prominent examples. The hypothesis evolved has been termed as "TERS (Terrorist Efficacy, Religion and Self-control)" which enhances the self control and raises their efficacy as casualties per attack. It also explained that the religious terrorist groups such as AL Qaeda and ISIS have higher efficacy than non religious groups. Terrorist attacks in Egypt have been found to be adversely affecting its economy and also affecting its foreign trade [8]. In this paper, statistical techniques have been applied to tourist attacks information of Egypt in last four

decades. The database used is Global Terrorism Database (GTD). Association mining algorithms have been used to detect frequent hidden patterns in the data in order to understand the nature of terrorist attacks. The effect of terrorism on Europe's centre- and far-right parties is explored in [9]. These platforms are basically created to attract support for elections. The work has studied the behaviour of voters such as how they respond to terrorist attacks and centre and far-right parties. It was found that far-right parties get more benefit than centre-right parties. The data taken was of more than 30 European countries, from 1975 to 2013. The work in [10] goes around the major incidents that happened in civil aviation, and due to which the security policies of aviation have changed over the time period. It basically includes the threats and breaches due to terrorism, and the counter measures and policies made. The 9/11 attack has been studied to explore threats to civil aviation and the international efforts to avoid them. It also elaborates the successful implementation of security measures for security of aviation as aviation in future. The work in [11] tries to explain reasons for violence and making strategies in order to prevent them and which is difficult to predict as they are continuously changing and is multi dimensional. Methods to improve this are defined in this paper that is new machine learning techniques, focusing on causes of conflicts and their resolution and other one is theoretical models which showcases the complexity of social interactions and decision making. The work in [12] has taken into consideration geospatial statistics for analyzing spatio-temporal evolution attacks on the Indochina Peninsula. Random Forest has been applied to predict

terrorist attacks on Indochina Peninsula with 15 driving factors. As per results, Thailand has been marked as the dangerous area for terrorist attacks followed by Middle Cambodia and Myanmar. The study in this paper shows the hotspots for terrorist attacks geographically.

III. Proposed methodology

The objective of our study is to predict the region and country of terrorist attacks. A. Dataset and Features The Global Terrorism Database (GTD) is a database which is open source and includes information on terrorist events for the years 1970-2017. It includes wholesome data regarding domestic incidents, transnational and international terrorist incidents which took place in this duration. The number of cases included is 180,000 {bombings (88,000), assassinations (19000) and kidnappings (11000)}. The parameters include date of incident, month of attack, location of incident, and country of incident, region of incident the weapons used in the incident, nature of the target, type of attack, the number of casualties, the group or individual responsible for the incident. The data source for GTD has been a variety of open media sources, more than 4,000,000 news articles and 25,000 news sources. It has been considered the most comprehensive unclassified database on terrorist attacks in the world. B. Exploratory Data Analysis Before building the model and to gain high level understanding of dataset features we performed some exploratory data analysis.

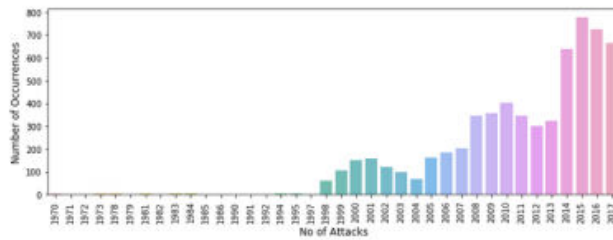


Figure 1: Number of Yearly Terrorist Attacks

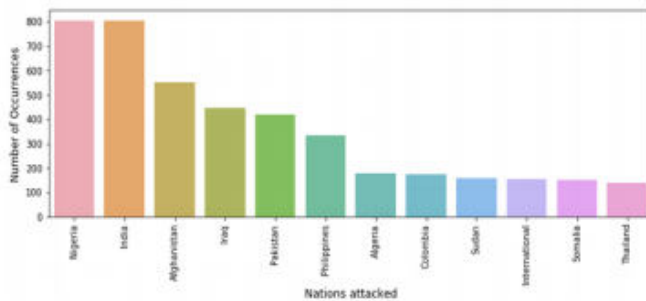


Figure 2: Number of Nation-wise Attacks

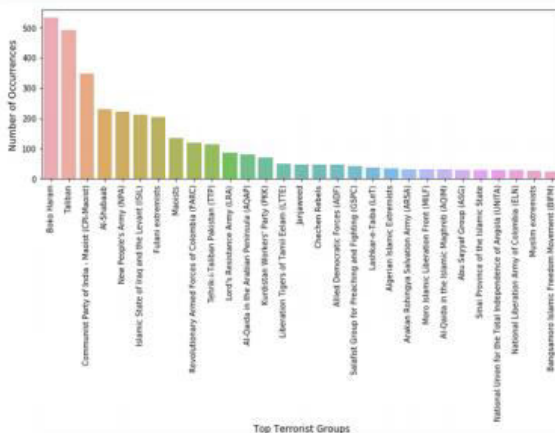


Figure 3: Terrorist Groups

Boko Haram is the terrorist group which has conducted maximum number of terrorist attacks, Taliban being the second and Maoists in India are on the third number of top terrorist groups

C. Machine Learning After exploratory data analysis on GTD (Global Terrorism Database), we further applied six supervised machine

learning models (Gaussian Naïve Bayes, Linear Discriminant Analysis, k-Nearest Neighbors, Support Vector Machines, Decision Tree, and Logistic Regression) for prediction and subsequently evaluated their performances. We have considered month of attack, Target Type1 and attack type1 to predict region and country. Gaussian Naive Bayes: Naive Bayes classifier has been considered as one of the simplest supervised approaches. In this, Bayes theorem provides a way to calculate probability of hypothesis (given prior information), hence the presence of one feature does not affect the presence of other feature. The advantage of NB is it can be easily trained with small and large datasets and the execution time is relatively fast. Linear Discriminant Analysis: LDA is also based on Bayes' Theorem. But instead of directly calculating posterior probability, it estimates multivariate distribution of its distribution. If we see its mathematical aspect, the algorithm does training by first setting the linear combination of predictors (features) that is helpful in separating different classes. The predicted class is classified by detecting the training samples which falls into linear decision boundaries. The advantage of LDA is it always produces an explicit solution and is feasible due to its low-dimensionality, but suffers from the assumption that linear separability is achievable in all classifications. k-Nearest Neighbors: k-NN is another algorithm commonly used for supervised classification problems. First introduced in 1951, the algorithm aims to identify homogeneous subgroups such that observations in the same group (clusters) are more similar to each other than others. Each data points' k-closest neighbors are found by calculating Euclidean or Hamming distance and

grouped into clusters. The k-closest data points are then analyzed to determine which class label is the most common among the set. The most common class is then classified to the data point being tested. For k-NN classification, an input is classified by a majority vote of its neighbors. That is, the algorithm obtains the classification of its k neighbors and outputs the class that represents a majority of the k neighbors. Support Vector Machines: In machine learning, they basically comes under the category of supervised learning which analyze data used for classification and regression analysis. SVM model is a representation of points in space, mapped properly so that the categories get divided by a wide gap. If new examples are mapped, then they fall accordingly into the right side of the gap. Logistic Regression: Logistic regression is the machine learning approach used as regression analysis when the dependent variable is binary. It is also a type of predictive analysis. It basically describes the relationship between one dependent binary variable and one or more independent variables (which can be ordinal, nominal, interval or ratio level). Decision Trees: Decision trees classifiers applies questions and conditions in a tree structure. This approach applies decision rules inferred from the data features to predict the value of target variable and create model accordingly. The condition for categorization is included in the root and internal nodes. Inputs are entered at the top and tree is traversed down, following the branches. Once the input node reaches the terminal node, a class is assigned. The advantage of decision trees is that they can be easily visualized and they can easily handle continuous and discrete data. When the training set is small in comparison with the number of

classes, it also leads to higher classification error rate, hence causing over fitting.

CONCLUSION

After training our models on the variables month, Traget_type, attack_type to predict the region of attack and country of attack it is estimated that Logistic regression, LDA, Naïve Bayes and SVM gives higher accuracy of 82 % in both the cases on predicting Region and country of terrorist attack. The results of the presented work can be used for enhancing defense against terrorist attacks in coming times.

REFERENCES

- [1] S. Sayad. Naïve Bayesian, from Predicting the Future. [Online],
- [2] E.Fix and J.Hodges. Discriminatory analysis: nonparametric discrimination: Consistency properties. PsycEXTRA Dataset, (1951)
- [3]<https://cogsci.yale.edu/sites/default/files/files/Thesis2018Peng.pdf>
- [4] 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).
- [5] Bang, J., Basuchoudhary, A., David, J., & Mitra, A. (2018). Predicting terrorism: a machine learning approach.
- [6] Mathews, T., & Sanders, S. (2019). Strategic and experimental analyses of conflict and terrorism. *Public Choice*, 179(3-4), 169-174.



[7] Ravenscroft, I. (2019). Terrorism, religion and self-control: An unexpected connection between conservative religious commitment and terrorist efficacy. *Terrorism and Political Violence*, 1-16

[8] 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). Springer, Cham

[9] Wheatley, W., Robbins, J., Hunter, L. Y., & Ginn, M. H. (2019). Terrorism's effect on Europe's centre-and far-right parties. *European Political Science*, 1-22.

[10] Klenka, M. (2019). Major incidents that shaped aviation security. *Journal of Transportation Security*, 1-18.