IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907110 |

Predicting Global Terrorism Activities using Machine learning Techniques

Nishanth R[1], Manish Seena Devadiga[1], Manoj B[1], Nikesh M R[1], Dr.Rajesh K S[2]

UG Student, Dept. of CSE, Rajarajeswari College of Engineering, Ramohalli Cross, Kumbalgodu, Bangalore, India[1]

Associate Professor, Dept. of CSE, Rajarajeswari College of Engineering, Ramohalli Cross, Kumbalgodu, Bangalore, India [2]

ABSTRACT: The goal of this research is to use machine learning to forecast the area and nation of a terrorist strike. The research was conducted using the Global Terrorism Database (GTD), an open database that contains a record of terrorist actions from 1970 to 2017. To improve accuracy, four machine learning methods were applied to a subset of characteristics from the dataset. The findings show that if specific characteristics are available, machine learning algorithms may be trained to anticipate the area and nation of a terrorist strike. The work may be utilized to improve global security against terrorist strikes, according to the theory. This research develops an integrated machine learning method for worldwide terrorist activity categorization and analysis. Machine learning-based data mining may be used to forecast terrorist attack occurrences, allowing specialists to acquire a better understanding of what terrorists are thinking in order to strengthen defences against these coordinated attacks.

KEYWORDS: Machine Learning, Terrorist Attacks.

I. INTRODUCTION

The history of terrorism is a collection of well-known and historically significant individuals, entities, and situations that have been associated with terrorism, whether correctly or incorrectly. Terrorism is a contentious term, according to scholars, and only a small percentage of people labelled terrorists identify as such. In a violent battle, it is typical for one side to refer to the other as terrorists or terrorists in practice.

We depend heavily on data from the Global Terrorism Database (GTD) in our coverage of terrorism, which defines terrorism as "acts of violence by non-state actors conducted against civilian populations with the intent of instilling fear in order to achieve a political aim." Government-initiated violence (state terrorism) and open battle between opposing armed forces, even if they are non-state entities, are not included in its definition. We give the GTD's more detailed definition, as well as others like as the United Nations', in our definitions section.

Terrorist assaults are rapidly expanding over the world. Terrorism is defined by the United Nations as "any action with a political objective designed to cause death or serious bodily harm to civilians." Around 22 thousand occurrences occurred globally in the previous year, resulting in approximately 18 thousand deaths. Terrorism's causes evolve with time since they are influenced by a variety of political and societal factors. Apart from forecasting the motive for the attack, identifying the agencies responsible is also challenging. There has been a scarcity of knowledge about prevalent terrorist behaviour patterns. Terrorist attacks are becoming more common all around the world. Terrorism is defined as "any action with a political objective that is intended to cause death or serious bodily harm to civilians," according to the United Nations. The elements that contribute to terrorism evolve throughout time since they are influenced by a variety of political and social circumstances. Apart from guessing the motive for the attack, identifying the perpetrators is challenging. There has been a dearth of information on the patterns of widespread terrorist activity. Existing analyses make use of case studies or quantitative approaches such as regression analysis.

Investigation of anomalous trends individual behaviours or asking detainees to obtain facts related to the attacks are examples of other types of study. The current study aims to establish a link between terrorism and the elements that cause it. Existing efforts have not been successful in predicting the future. Given the necessary data, machine learning algorithms can help estimate risk of a terrorist attack. The findings of this research can assist security agencies and policymakers in combating terrorism by recommending appropriate and effective solutions.

There has been a scarcity of knowledge about prevalent terrorist behaviour patterns. Case studies or quantitative methods such as regression analysis are used in the existing analyses. The first is limited to specific occurrences, whilst the latter is limited to interviews with citizens who were affected by the attack. The majority of these assessments are based on variables such as the weapons used in the attacks and the number of individuals injured. Investigation of

e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907110 |

anomalous trends in individual behaviours or questioning detainees to obtain facts related to the attacks are examples of other types of study. The current study is attempting to establish a link between terrorism and its consequences.

It's incredibly tough to learn about a terrorist's identity, intentions, plans, and vulnerabilities. However, no other policy initiative is more critical for preventing, anticipating, and responding to attacks. The authors employed data mining techniques in the existing system to find the most useful machine learning algorithm. The analysis was done with the WEKA tool, and the techniques employed were J48 and Bayes Net. The lowest accuracy came out. As a result, it is a less accurate method of prediction.

If specific characteristics are available, it is feasible to train machine learning models in the proposed system to forecast the area and nation of a terrorist strike. We're working on a dataset of worldwide terrorism strikes. Our Naive Bayes method is used to analyses the data. The algorithm's accuracy is expected. As a result, our research includes a global terrorist analysis.

Terrorist assaults are rapidly expanding throughout the world. Terrorism is defined by the United Nations as "any action with a political aim designed to cause death or serious bodily injury to civilians." Around 22 thousand incidents happened globally in the previous year, resulting in approximately 18 thousand deaths.

Terrorism's causes change throughout time since they are influenced by a variety of political and societal variables. As a result, identifying global terrorism is critical to protecting the country's citizens from any terror strikes.

[1]. Gather data from GTD and format it appropriately.

[2] To study all four algorithms and put them into practice in order to determine their accuracy and precision.

[3]. Using GTD to discover the many outcomes that represent data.

[4] Calculating all of the algorithms' accuracy and precision values.

[5] Comparing algorithms to determine which the best is

[6] Using multiple measures to determine accuracy, such as the confusion matrix.

[7] Foreseeing an attack.

After training our models on the month variables, Target type, Attack type and many more independent features to predict the success of Attack. The findings of the presented project can be used in the coming times to improve security against terrorist attacks.

Given the necessary data, machine learning methods can help forecast the risk of a terrorist strike. The findings of this research can assist security agencies and policymakers in implementing appropriate and effective steps to combat terrorism. Machine learning algorithms can assist predict the danger of a terrorist attack if given the proper data. The findings of this study can aid security agencies and policymakers in putting in place suitable and effective counter-terrorism measures.

II. LITERATURE SURVEY

[1] If given the right data, machine learning algorithms can help forecast the likelihood of a terrorist strike. The outcomes of this research can help security agencies and policymakers devise appropriate and successful counter-terrorism strategies. Bayes Net, J48, k-Nearest Neighbours, Support Vector Machines, and Naive Bayes were employed in the study using the WEKA tool. KNN had the lowest accuracy, despite being good on other metrics. The usual study of these datasets, which includes algorithms, is carried out using the Weka tool and is based on real data from the Global Terrorism Database.

[2] Between September 1 and October 11, 2001, the LexisNexis database of media stories was utilized to identify occurrences of hate-related violent actions against Middle Easterners or those believed to be Middle Easterners in the United States. In the 2659 news items examined, a total of 100 incidences of hate-related violence were discovered. Researchers are concerned about the prediction of terrorist actions. Because of the vast number of occurrences, predicting the terrorist group responsible for certain terrorist activity is challenging.

e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907110 |

[3] This study discusses how machine learning may aid in the understanding of terrorism. We emphasize that, despite its reputation for black box prediction, machine learning may give highly complex explanations of terrorism.

Geospatial statistics have been used to analysespatio-temporal evolution assaults on the Indochina Peninsula in this study. With 15 driving elements, Random Forest was used to forecast terrorist strikes on the Indochina Peninsula. According to the findings, Thailand is the most risky country for terrorist strikes, followed by Middle Cambodia and Myanmar. The map in this article depicts the geographic hotspots for terrorist attacks.

[4] The hotspots for terrorist attacks on a more fine-grained geographical level are identified in this study. Meanwhile, it demonstrates that combining machine learning methods (such as RF) with geographic information systems (GIS) has a lot of promise for modelling the danger of terrorist strikes.

They've also produced a new dataset called QFactors Terrorism, which combines event-specific characteristics obtained from the GTD with population-level demographic data from UN and World Bank sources. Approaches such as Naive Bayes, decision trees, Linear Discriminant Analysis, k-nearest neighbours, and random forest have been used. After being trained, the random forest model was effective in determining the cause of an identified occurrence with up to 68 percent accuracy.

[5] WeisiGuo, Kristian Gleditsch, and Alan Wilson suggest that utilizing artificial intelligence to forecast violent outbursts and investigate their origins might save lives. Terrorism's impact on Europe's center- and far-right parties is investigated. These platforms are primarily designed to entice voters to vote in elections.

Future study should focus on understanding nonlinear impacts of factors like assassinations, as well as all other characteristics like democracy and economic institutions, according to the paper. The research looked into voter behavior, such as how they reacted to terrorist events and how they supported for center-right and far-right parties. Far-right parties are seen to profit more than Centre-right parties. From 1975 to 2013, data was collected on more than 30 European nations. The authors looked at supervised machine learning techniques for studying terrorist activities before creating a model to identify past occurrences in the Global Terrorism Database.

III. SYSTEM ANALYSIS

Researchers are particularly concerned about the prediction of terrorist operations. Because of the high number of events, predicting the terrorist group accountable for a certain terrorist act is challenging.

The current study aims to establish a link between terrorism and the elements that cause it. Existing efforts have not been successful in predicting the future. Given the necessary data, machine learning algorithms can help estimate the risk of a terrorist attack. The findings of this research can assist security agencies and policymakers in taking appropriate and effective steps to combat terrorism. As a result, there is a method for evaluating terrorism in a given region or country using machine learning techniques and terrorism-specific knowledge in order to draw inferences about terrorist behavior patterns.

A. *Tools and Technologies*

Python is a high-level programming language that may be used in a variety of situations (human understandable languages are High level programming languages) Python Guido Van Rossam developed it for the National Research Institute in 1989. (NRI) Python was first made available to the general public in 1991 in the Netherlands: FEBRUARY 20TH, 1991 Guido van Rossum of the Centrum Wiskunde en Informatica (CWI) in the Netherlands was the first to adopt Python in December 1989 as a replacement for the ABC dialect (which was preceded by SETL) for dealing with and interacting with the Amoeba working system. Van Rossum, the creator of Python, is still alive and well.

Understanding is simple and clear. JAVA contains about 100 keywords, but Python has only 33. We can't afford to pay for open source or freeware since there isn't a license (we can able to see source code if source is not good I can able to customize our requirements) High-level programming language (human understandable language) Python is platform independent (i.e., I can write a programme once and run it on any machine) (WORA).

Python includes libraries which allow programmers for employing more efficient algorithms. Which makes use of well-known machine learning technique like classification, recommendation and clustering. As a result, before proceeding, a basic introduction to machine learning is required.

e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907110 |

Machine learning is a subfield of computer science concerned with programming systems to learn and improve on their own. Identifying and digesting the input data, as well as making informed judgments based on the information presented, are all examples of learning. When making judgments, it is difficult to consider all available inputs. To overcome this problem, algorithms are being created that use statistical science, probability, logic, mathematical optimization, reinforcement learning, and control theory to construct knowledge from specific facts and prior experience.

Machine learning is a type of learning that is automated and requires little or no human interaction. It includes programming computers such that they may learn from the data they are given. The fundamental objective of machine learning is to research and develop algorithms that can learn from prior data and make predictions based on new data. The input to a learning algorithm is training data, which represents experience, and the algorithm's output is any expertise, which is typically in the form of another algorithm capable of performing a task. The input data to a machine learning system might be numerical, written, audio, visual, or multimodal.

The system's output data can be a floating-point number, like the rocket's velocity, or an integer, like a pigeon or a sunflower from image recognition.

Positive and negative feeling, female and male humans, malignant cancers, unsecure and secure loans, and other classifications are examples of attempts to identify the proper class label. The goal of supervised learning is developing a general rule that maps inputs to respective outputs, and data comes with descriptions, labels, targets, or intended outcomes. Labeled data is a term used to describe this type of learning data. The learned rule is then used to label new data with unknown outputs

Unsupervised learning is used to find anomalies, outliers, such as fraud or malfunctioning equipment, or to group clients who behave similarly for a sales campaign. It is diametrically opposed to guided instruction. There are no data points that have been labelled. When learning data only offers a few hints and no descriptions or labels, it is up to the coder or algorithm to figure out the underlying data's structure, uncover hidden patterns, or decide how to characterize the data. Unlabeled data is a term used to describe this type of learning data.

The learning data provides feedbacks, allowing systems to respond to changing condition in order to attain specific goal. Based on the feedback responses, the system analyses its performance and adjusts accordingly. Chess master algorithm and Self-driving cars are two of the most well-known examples.

IV. SYSTEM DESIGN

The process of establishing the architecture, components, modules, interfaces, and data for a system in order to satisfy specific requirements is known as system design. It may be thought of as a systems theory application in product development. Systems analysis, systems architecture, and systems engineering all have some overlap. Design is the process of obtaining marketing data and translating it into a design for a produced product, if product development as a whole "combined the views of marketing, design, and manufacturing into a unified approach to product creation."

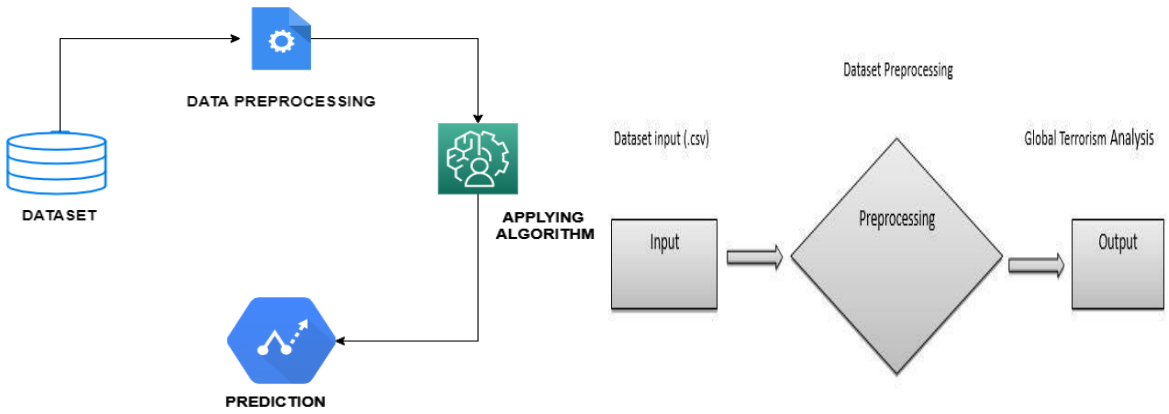

Fig. 1 System ArchitectureFig. 2 IO Design

e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907110 |

1) Naïve Bayes

The Bayes Theorem is used to generate the Naive Bayes classifiers, which are a collection of classification methods. It's a group of algorithms that all work on the same principle: each pair of features to be categorized is independent of the others

2) K Nearest Neighbors

KNN may be used to tackle both classification and regression forecasting problems. However, it is more typically used in the industry when classification problems arise. The K-Nearest Neighbour method is one of the most fundamental Machine Learning algorithms. It is based on the Supervised Learning technique.

3) Linear Discriminant Analysis

In statistics and other fields, Fisher's linear discriminant (LDA) is a method for determining a linear combination of information that separates or differentiates two or more classes of objects or events. A generalization of Fisher's linear discriminant is linear discriminant analysis (LDA), normal discriminant analysis (NDA), or discriminant function analysis. The resultant mixture can be used as a linear classifier or, more commonly, for dimensionality reduction prior to classification.

4) Decision Tree

Each internal node represents a "test" on an attribute (for example, whether a coin flip will land heads or tails), each branch reflects the outcome of the test, and each leaf node offers a class name (decision taken after computing all attributes). The paths from the root to the leaf indicate the classification rules. Using a visual and analytical decision support tool, a decision tree and its closely related impact diagram are used to determine the anticipated values (or expected utility) of competing options in decision analysis.

A. Dataflow Diagram

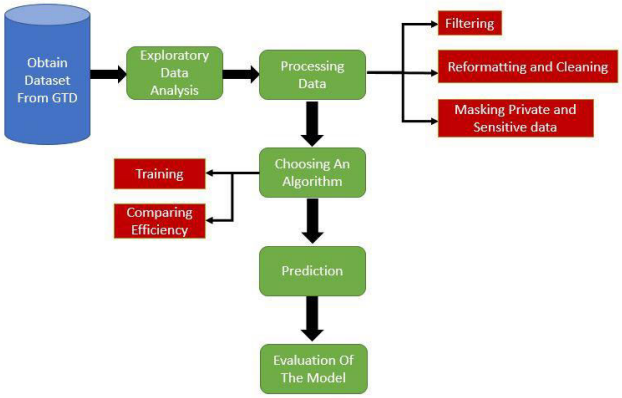

Fig. 3 Flow Design

A data flow diagram is a graphical representation of data "flow" across an information system that represents the process elements. They're typically utilized as the first step in creating a system overview that can then be developed. DFDs can also be used to visualize data processing (structured design). The DFD is also known as a bubble chart. It's a simple graphical representation of a system's input data, various processing operations done on the data, and the system's output data.

B. Algorithm

1) Naïve Bayes

The Bayes Theorem is used to generate the Naive Bayes classifiers, which are a collection of classification methods. It's a group of algorithms that all work on the same principle: each pair of features to be categorized is independent of the others

2) K Nearest Neighbors

KNN may be used to tackle both classification and regression forecasting problems. However, it is more typically used in the industry when classification problems arise. The K-Nearest Neighbour method is one of the most fundamental Machine Learning algorithms. It is based on the Supervised Learning technique.

e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907110 |

3) Linear Discriminant Analysis

In statistics and other fields, Fisher's linear discriminant (LDA) is a method for determining a linear combination of information that separates or differentiates two or more classes of objects or events. A generalization of Fisher's linear discriminant is linear discriminant analysis (LDA), normal discriminant analysis (NDA), or discriminant function analysis. The resultant mixture can be used as a linear classifier or, more commonly, for dimensionality reduction prior to classification.

4) Decision Tree

Each internal node represents a "test" on an attribute (for example, whether a coin flip will land heads or tails), each branch reflects the outcome of the test, and each leaf node offers a class name (decision taken after computing all attributes). The paths from the root to the leaf indicate the classification rules. Using a visual and analytical decision support tool, a decision tree and its closely related impact diagram are used to determine the anticipated values (or expected utility) of competing options in decision analysis.

V. IMPLEMENTATION

A. Flow Diagram

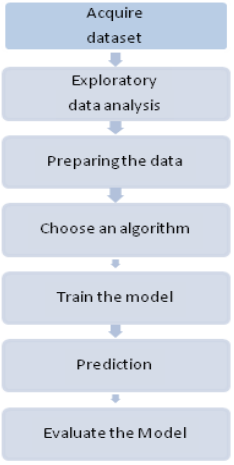

Fig. 3 Data Flow

In the process of analyzing the success we have considered machine learning algorithms to predict the output. Some of the models are taken into consideration to check the efficiency of the machine learning algorithm.

B. Gathering Data

The Global Terrorism Database (GTD) is a free, open-source database that provides statistics on terrorist attacks from 1970 to 2017. It includes detailed information on domestic, international, and global terrorist acts that happened during this time period. There are 180,000 bombings (88,000), deaths (19000), and kidnappings (11000) incidents on the list.

C. Exploratory Data Analysis

Before creating the model, we undertook some exploratory data analysis to gain a high-level understanding of the dataset's features.

e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907110 |

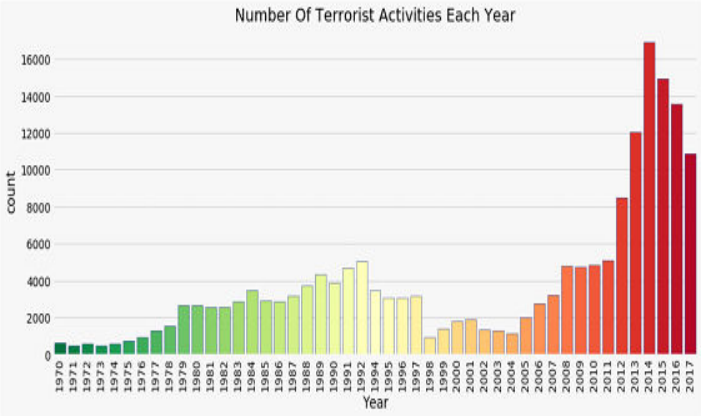

Fig. 4 Terrorist Attack Data

D. Preparing the Data

The process of cleaning and altering raw data prior to processing and analysis is known as data preparation. Reformatting data, making data changes, and merging data sets to enhance data are all part of this crucial stage before processing. For data professionals or business users, data preparation might be time consuming, but it is necessary to put data into context in order to transform it into insights and minimize bias caused by poor data quality. Standardizing data formats, enhancing source data, and/or eliminating NAN values are all common steps in the data preparation process. Which, in any case, cannot be used in a machine learning manner.

VI. SYSTEM TESTING

Testing is a key component of confirming the proposed system's quality and efficacy in (satisfactorily) achieving its goals. In order to build a transparent, adaptable, and secure system, testing is done at various phases of the system design and implementation process. The importance of software testing in the development process cannot be overstated. In a sense, the testing technique verifies that the product being manufactured complies with the specifications for which it was designed. Creating test cases against which the product must be evaluated is part of the testing process.

A. Levels of Testing

1) Unit Testing

Unit testing is the most fundamental type of testing. Unit testing is used to test the smallest unit of software design, the module. A white box is always used to represent a unit test. The requirements generated throughout the module design phase are used to test the various modules. The aim of unit testing is to check the code created during the coding process and to test the modules' underlying logic. The module's programmer is generally in charge of this. Because of its close relationship with coding, the coding phase is sometimes referred to as "code plus unit testing." The unit test may be run in parallel for several modules.

2) Manual and Automated Testing

Manual testing is a type of software testing in which a tester runs test cases manually rather than utilizing automated methods. Manual testing's objective is to identify defects, flaws, and errors in a software product. The most fundamental of all testing methods, manual software testing assists in the detection of major flaws in software programs.

Any new application must be manually tested before it can be automated. Manual software testing takes longer and requires more effort, but it is necessary to see if automation is viable.

Manual testing principles do not need any testing tool knowledge. "100% automation is not possible," one of the Software Testing Fundamentals asserts. Manual testing is required as a result.

3) Functional Test

Functional tests verify that the assessed capabilities are available in a systematic manner, as stated by the business and technical requirements, system documentation, and user manuals. Functional tests are planned and structured around requirements, critical functions, and one-of-a-kind test cases. Furthermore, testing must address systematic

e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907110 |

coverage of business process flows, data fields, preset procedures, and subsequent processes. Before functional testing, more tests are identified and the utility of existing tests is evaluated.

VII. CONCLUSION

We were able to forecast the result of the assault after training our models on month variables, target type, attack type, and a number of other independent factors. Naive Bayes is thought to be accurate in forecasting the success of an attack. The findings of the study provided here can be utilized to increase protection against terrorist attacks in the future.

Terrorism has become a major global danger. Various machine learning systems, artificial intelligence, and data analytics have given us with a mechanism to assist investigators and antiterrorist or counter-terrorist squads in quickly determining the most likely culprit of a terrorist incident.

We've shown how approaches like the k-Nearest algorithm, Naive Bayes, and Decision Tree can help us accurately identify the guilty party eight out of 10 times. This aids anti-terrorist agencies in narrowing down the number of probable suspects and acting quickly to locate and apprehend the genuine criminal.

In the future, we intend to experiment with other techniques and approaches, such as deep learning models and package classifiers, in order to improve the accuracy of the results and, as a consequence, correctly forecast the culprit with more precision and accuracy.

Apart from that, we want to utilize web scraping technologies and emotional analysis in the future to examine various posts and comments on social media sites for hate speech and text, filter them, and develop a classifier to combine the present project with the social media texts.

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. [3] https://cogsci.yale.edu/sites/default/files/files/Thesis2018Peng.pdf DilipKumar S. M. and Vijaya Kumar B. P. ,'Energy-Aware Multicast Routing in MANETs: A Genetic Algorithm Approach', *International Journal of*Computer *Science and Information Security* (IJCSIS), Vol. 2,2009.
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. Klenka, M. (2019). Major incidents that shaped aviation security. Journal of Transportation Security, 1-18.
8. Guo, W., Gleditsch, K., & Wilson, A. (2018). Retool AI to forecast and limit wars. "Sentiment analysis of Twitter data", Apoorv Agarwal, BoyiXie, Ilia Vovsha, Owen Rambow, 2011.
9. "Hao, M., Jiang, D., Ding, F., Fu, J., & Chen, S. (2019). Simulating Spatio-Temporal Patterns of Terrorism Incidents on the Indochina Peninsula with GIS and the Random Forest Method. ISPRS International Journal of Geo-Information, 8(3), 133

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details