DOI https://doi.org/10.5281/zenodo.10049652#55
CC BY NC SA
TWIST, 2023, Vol. 18, Issue 4, pp. 308-316

TWIST

Journal homepage: www.twistjournal.net

A Data Mining Based Online Terrorism Detection and Prediction System

Ele B. I.*
Department of Computer Science, University of Calabar, Calabar, Nigeria
[*Corresponding author]

Egete D. O.
Department of Computer Science, University of Calabar, Calabar, Nigeria

Obono I. O.
Department of Computer Science, University of Calabar, Calabar, Nigeria

Abstract
Terrorism has grown its roots quite deep in certain parts of the world. With increasing terrorist activities, the realization of this by the researcher and the thought that it is important to drastically reduce terrorism and its spread before a certain time motivated the researcher towards this work. However, Anti-terrorist organizations have identified internet as a major source of spreading terrorism through speeches and videos; as terrorist organizations use internet to brain wash individuals and also promote terrorist activities through provocative web pages that inspire helpless people to join terrorist organizations. Thus, an efficient web data mining system to detect such web properties and flag them automatically for human review is a great benefit from the proposed system. Data mining is a technique used to mine out patterns of useful data from large data sets and make the most use of obtained results. The dataset in this consists of 18000 records and the clustering algorithm was used to mine patterns of useful data from the dataset. Web pages are made up of HTML (Hyper Text Markup Language) in various arrangements and have images, videos, etc. intermixed on a single web page. In this way we can judge web pages and check if they can be promoting terrorism. In this study, a data mining based online terrorism detection system was developed and implemented using the following programming and technological tools: PHP, Python, Jupyter Notebook, Anaconda, Django, HTML/CSS, Tableau, Python packages (Mathplotlib, Pandas and Numpy), MYSQL, and WAMP server. This system proves useful in anti-terrorism sectors and even search engines to classify web pages into various categories.

Keywords
Data mining, Terrorism, Clustering algorithm, Prediction, Dataset, Online terrorism detection system

INTRODUCTION

Terrorism can be defined as the calculated use of unlawful violence or threat of unlawful violence to inculcate fear; intended to coerce or to intimidate governments or societies in the pursuit of goals that are generally political, religious, or ideological. It is an intentional violence perpetrated by unlawful individuals or groups to achieve religious or political aims. Terrorism is also "An illegal preconceived use of physical or psychic violence (or threat of it) for further political goals aimed at civilians or non-combatants etc., to change current policies, their methods and structure". The roots of terrorism grow deeply in some parts of the world. Terrorist groups use the Internet as their infrastructure for different purposes that might lead to the formation of new local cells in the terrorist network. These local cells later become active and perform acts of terror. Terrorist associations are utilizing the website to spread their propaganda and radicalize youth on the online and urge them to commit terrorist activities.

The Internet is an efficient communication infrastructure that is increasingly used by terrorist organizations to safely communicate with their affiliates, coordinate action plans, spread propaganda messages, raise funds, and introduce new supporters into their networks (Birnhack and Elkin-Koren, 2012). Governments and intelligence agencies are calling

to invest major efforts in development of new methods and technologies for identifying terrorist activities on the web in order to prevent future acts of terror.

According to AU's article 1 (3), Terrorism is defined as any act which is a violation of the criminal laws of a State Party and which can endanger the life, physical integrity or freedom of, or cause serious injury or death to any person, or group of persons or causes damage to public or private property, natural resources, environmental or cultural heritage and is calculated or intended to:

i. Intimidate, put in fear, coerce or induce any government, body, institution, the general public or any segment thereof, to do or abstain from doing any act, or to adopt or abandon a particular standpoint or to act according to certain principles; or
ii. Disrupt any public service, the delivery of any essential service to the public or to create a public emergency; or
iii. Create general insurrection in a State.

The AU, in article 3(1), also notes that:

i. The struggle waged by peoples in accordance with the principles of international law for their liberation or self-determination, including armed struggle against colonialism, occupation, aggression and domination by foreign forces.
ii. Political, philosophical, ideological, racial, ethnic, religious or other motives shall not be a justifiable defense against a terrorist act.

More so, in recent decades, it become clear that the internet can be used to connect people both for good and for ill, Facebook and other social media companies have made it a priority to minimize the way criminals can use new technology. One of the greatest challenges in this arena has been terror groups, which have embraced the internet as a way to spread propaganda and recruit others to their extremist ideologies. These groups do not confine their efforts to one social media service or online distribution mechanism. Instead, they try a variety of avenues to get their message out. But even as governments, companies and nonprofit organizations have battled terrorist propaganda online, social media companies have faced a complex question over what is the best way to tackle a global challenge that can proliferate in different ways, across different parts of the web.

Often analysts and observers will ask that, with the vast databases and advanced technology, why not social media companies block nefarious activity using technology alone. The truth remains that both technology and people are needed to stop the spread of terrorist content across the entire internet. It is based on this that the researcher painstakingly undertakes the voyage through the **Online Terrorism Detection and Prediction System** which uses data mining (a technique used to mine out patterns of useful data from large datasets) to analyze texts, audio and visual content for terrorism propaganda in our communities, and also to predicting terrorist attempts to recruit people to their cause, and then spread fear in our society.

PROBLEM DEFINITION

The manual means of fighting terrorism has failed drastically both in the past and in recent times. These methods includes:

i. Relying on community policing or vigilante groups;
ii. Declaring state of Emergency in some communities through selective engagement;
iii. Securitizing the physical relationships or contacts of individuals, terrorists, intending terrorists and sponsors of terrorism
iv. Unintentionally giving the appearance that the police support particular individuals or groups, which could either undermine the legitimacy of those in a position to exercise a positive influence within the community or alienate other community members or communities.

Thus, these and more are the problems facing the traditional method of detecting and handling terrorism or terrorist movements and efforts.

PURPOSE OF THE RESEARCH

The aim of this research is to develop an online terrorism detection and prediction system using data mining techniques that will assist in detecting terrorism, terrorist sites and planned attacks. In consensus with this aim, the specific objectives for achieving the aim are:

i. To apply clustering algorithm to mine data in order to control the spread of terrorism and terrorist activities;
ii. To develop a terror tracking system to analyse the content of information accessed by Internet users, predict likely terrorist sites and ensures that any person who participates in terrorist acts is tracked down;
iii. To design a database using MySQL to keep records of terrorists, terrorist sponsors and terrorism in general; and
iv. To implement the data mining based online terrorism detection and prediction system using HTML, PHP and Python programming languages.

SIGNIFICANCE OF THE STUDY

The implementation of the proposed system will help to curb or drastically reduce the spread of terrorists' activities. The proposed Anti-Terrorist Detection System (ATDS) will enable individuals, organizations, and government to effectively discover terrorist sites, terrorist' activities, terrorist cells and any terrorist related activities and alert the security operatives in Nigeria for immediate response and rescue operations.

Specifically, the proposed system contributes immensely in the following ways:

i. The system will avoid or keep in check the activities of terrorists thereby reducing it to the barest minimum, if not completely eliminated;
ii. It will reduce the risk in loss of lives and properties as the face to face combats will be reduced;
iii. Ensure that there is a clear distinction between counterterrorism operations and community police work and also ensure that there is effective co-ordination between these operations;
iv. It will create an avenue for members of the armed forces to be trained and properly managed in counterterrorism operations; and
v. The online terrorist detection system is aimed at detecting online access to abnormal content that could embrace terrorist-generated websites by analysing the content of knowledge accessed by the net users.

ANALYSIS OF THE EXISTING SYSTEM

The law enforcement agents and other organizations responsible for terrorist activities and terrorism matters used the manual methods to handle issues of terrorism detection, prediction and handling before now. When persons around begin to suspect some movement around his/her environment, individuals sluggishly go to the appropriate authorities to write long reports which is used in investigating the case and process; as almost eyewitness account is needed from the first day till the process is ended and the culprits arrested and brought to book.

Limitations of the Existing System

The problems of the Existing system include

i. The citizens out numbers the police in most countries
ii. Innocent individuals are harassed and affected during investigations
iii. The force men may aid and abate the crime on a low key
iv. A lot of money is spent on movements from one location to the other
v. Evidences may not be really clear

Analysis of the Proposed System

The proposed system is designed with the aim to detect and predict terrorists in a stress reduced manner, help in hiding the faces of persons in charge for security reasons. The system creates more opportunities for terrorists to be caught as they have no idea they are being tailed. Also, it takes and keeps records which is a strong evidence to be presented against the terrorists.

Justification of the Proposed System

The new system which is computerized has so many benefits that will be fetched if introduced into full operation:

i. Preventing men and women from becoming terrorists;
ii. Providing opportunities and support to individuals on a path to, or involved in terrorism to disengage;
iii. Denying terrorism suspects the support, resources and means to organize themselves or to plan and carry out attacks;
iv. Preparing for, and protecting against, terrorist attacks, in order to decrease the vulnerability of potential targets, in particular critical infrastructure;
v. Pursuing terrorist suspects to apprehend them and bring them to justice; and
vi. Responding to terrorist attacks through proportionate measures to mitigate the impact of such attacks and to assist victims.

Architecture of the Proposed System

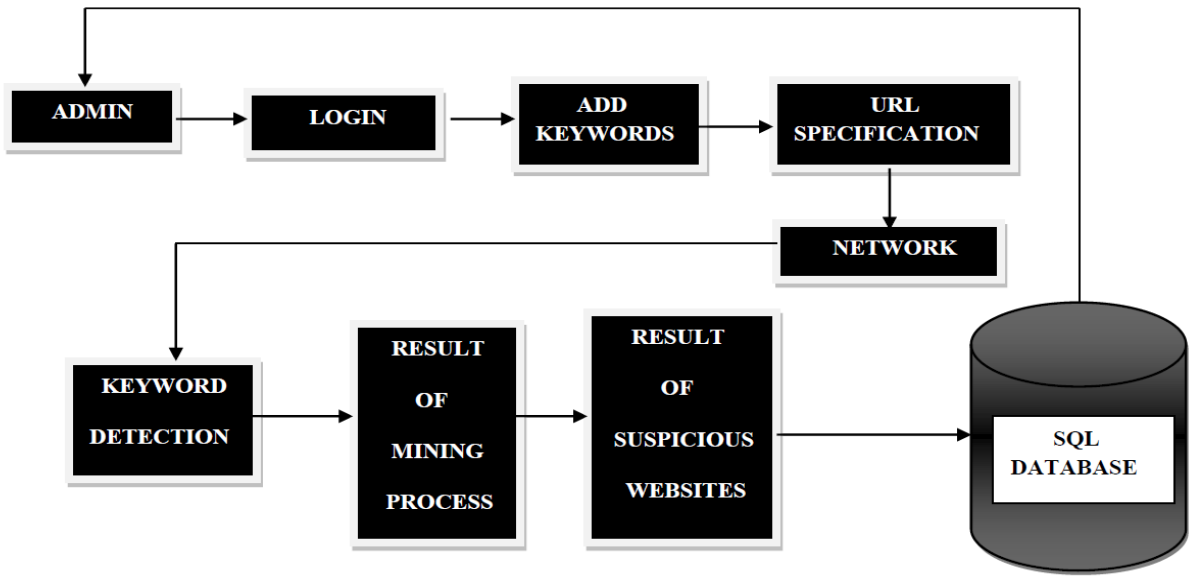

Fig. 1 Proposed System Architecture

In the first step, the admin login using the user id and password. If the admin is new, then the new admin should register the user id and password to access further. After the login process, the admin can add suspicious keywords and the websites to the database. In the next process, it includes the detection of suspicious keywords from the user's browser history, if a user is found to be using any kind of suspicious keywords or websites continuously, then the particular user's IP address will be mined and get stored in a database, which will be helpful to detect the particular user and stop the spread of terrorism.

Algorithm Design of the Proposed System

Algorithm used in this study is the word extraction algorithm. The word extraction algorithm is given below:

Input: Web URL's i.e. log file

Output: Extracted meaningful word Begin

Step 1: Take the browser history
Step 2: Using Vector extract the word from that document
Step 3: Match that word with standard database
Step 4: Use clustering algorithm to do the cluster of similar words
Step 5: If match found and shows terror behaviour then go to step 6 Else go to step 7
Step 6: Report the problem using alarm or beep sound
Step 7: Close the connection
Step 8: Exit

Flowchart of the Proposed System

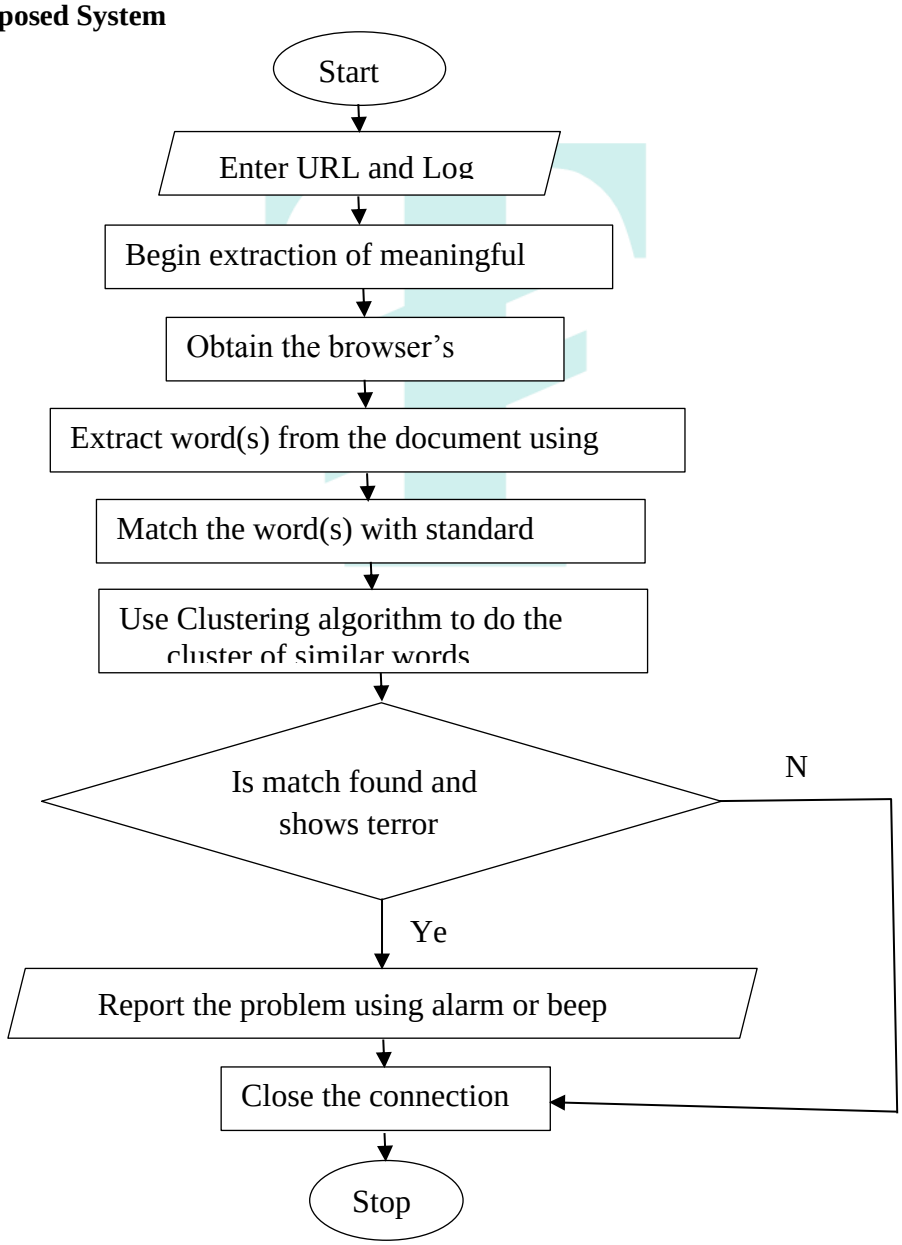

Fig. 2 Flowchart of the Proposed System

RESULTS AND DISCUSSION

The result of this research is the development of a data mining based online terrorism detection and prediction system that is capable of efficiently detecting, predicting and alerting security operatives of terrorism, terrorist sites and planned terrorist attacks.

Initially, we run the program to get the login page where the cybersecurity officials or security operatives enter their details to get access to the software. Using this, they can add new keywords or websites, update already existing data and get connected with the database. They can also detect and predict such websites in order to block them.

Login Interface

The login interface depicts the login page where the cybersecurity officials can select it to enter their details. See Fig. 3.

Fig. 3 Login Page

User Details Interface

This interface depicts the user details page. Here a new user can register and create an account provided the user have authentication to do so. The user can directly login if he/she has already registered. This page consists of two components, which are the user id and password. These details are kept highly confidential for information security. See Fig. 4.

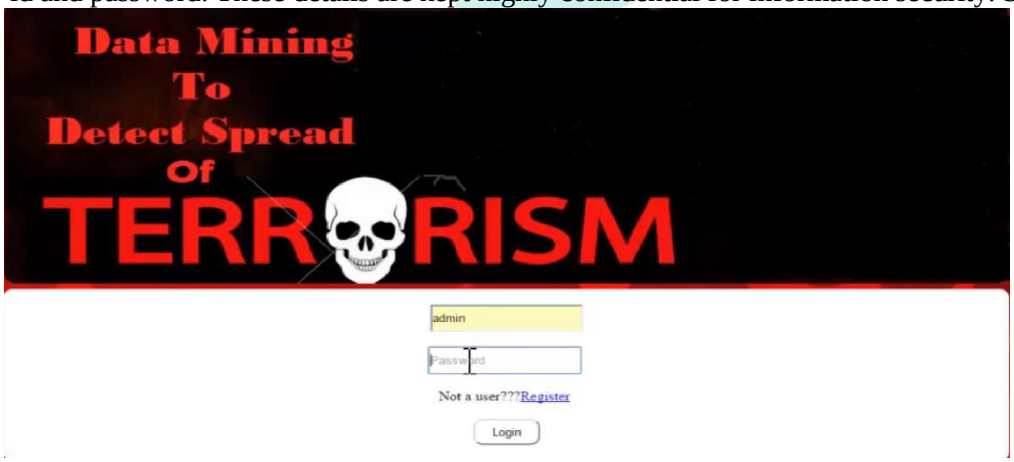

Fig. 4 User Details Page

Sections Interface

In this interface, there are five sections, which are add keywords, check the website, and check all websites, update password, view feedback and log out. Add keywords is used to add keywords to the already existing database. Check all website is used to scan the browser history for any suspicious websites. Check website is to add suspicious site to the database. View feedback is to get feedback from the authorities. Logout is to log off one's account. See Fig. 5.

Fig. 5 Sections Page

Killings in Global Terrorism

Fig. 6 shows the total number of attacks based on years. The interesting part to notice here is to compare different years based on the total number of attacks. Each block represents a year. The larger the block the more the number of attacks faced in that year. Surprisingly six major blocks belong to the current decade (2010s). This observation emphasizes that some of the most violent years in the history of terrorism happened in this decade especially in an era when terrorism is been given priority more than ever before. Different shades are used to make neighbouring blocks more distinctive among each other.

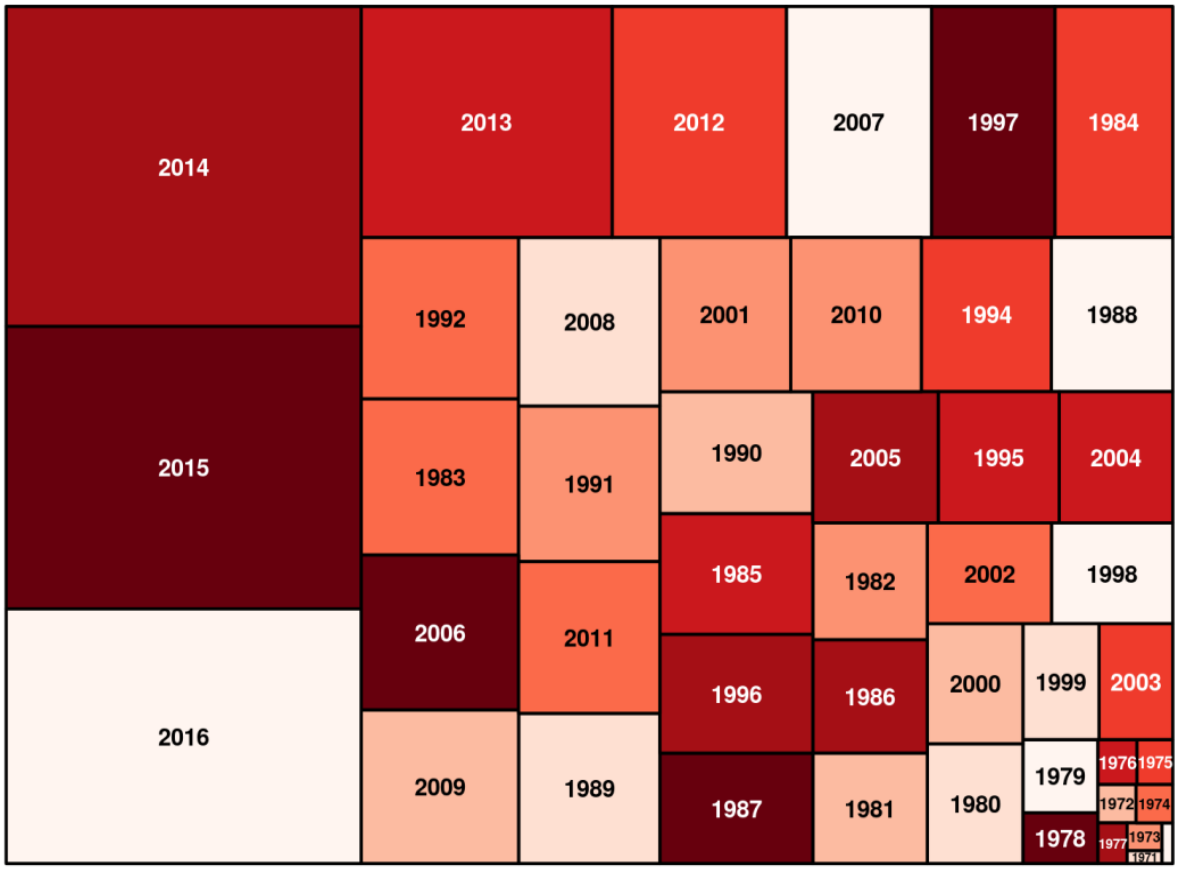

Fig. 6 Killings in Global Terrorism

Attacks to Kill Comparison

Fig. 7 lists countries mostly affected by terrorism based on the total number of attacks. Another bar along with each countries' attack count is the number of total victims killed in those attacks combined for that country. It analyze kills to attack ratio for the most affected countries.

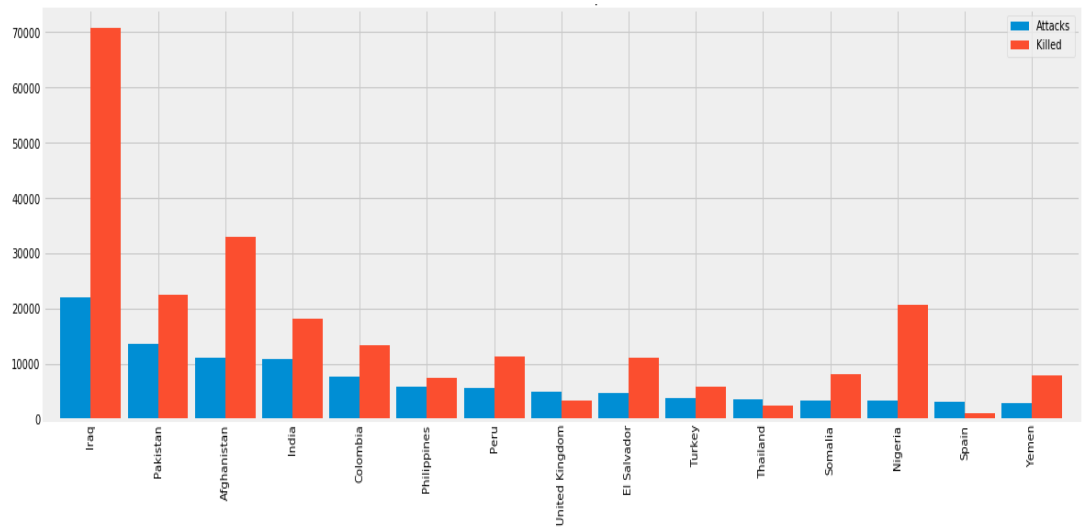

Fig. 7 Attacks to Kill Comparison

Terrorist Activities in Each Region

Based on the geographic location of countries, they have been subcategorized into twelve regions to compare the rate of terrorism in each one of them as shown in Fig. 8. Middle East and North Africa have the highest number of attacks followed by South Asia and South America. Terrorism here does not show an equal distribution among all regions. As a result, based on the number of attacks, different level of attention is required for each individual region.

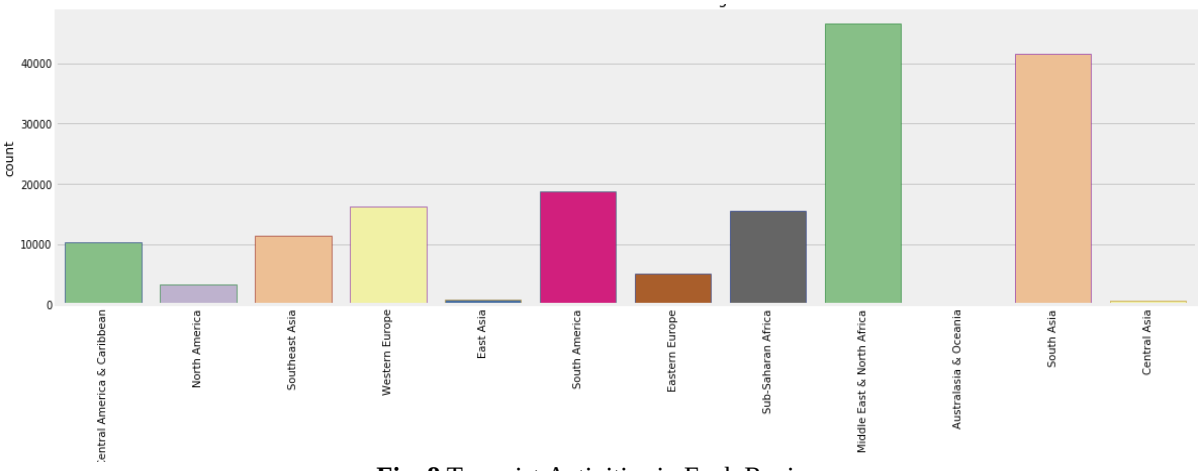

Fig. 8 Terrorist Activities in Each Region

Attacking Methods by Terrorists

Different types of weapons and methods have been used by attackers. There are 8 categorical values for the defined attack type which are unarmed assault, Infrastructure attack, kidnapping, barricade incident, hijacking, bombing/explosion, armed assault and assassination. This figure does reveal the potential target or focus of the attacker. For instance, unarmed assault attacks are usually focusing against specific individuals or a group of small people. Explosives and bombing are targeted towards a larger audience. Hijacking aims to achieve some sort of ransom in return. Here the graph pictures the total number of kill counts with respect to specific attacking methods used. Fig. 9 uses the nine most used means of attacking based on the causalities caused. Explosions are the most common followed by armed assaults, assassinations, hostages and so on. Here the total number of casualties by explosive weapons is almost double than the next most attack which is armed assault. This observation indicates that most of the attacks were intended to civilians for the purpose of spreading terrorism among widespread targets.

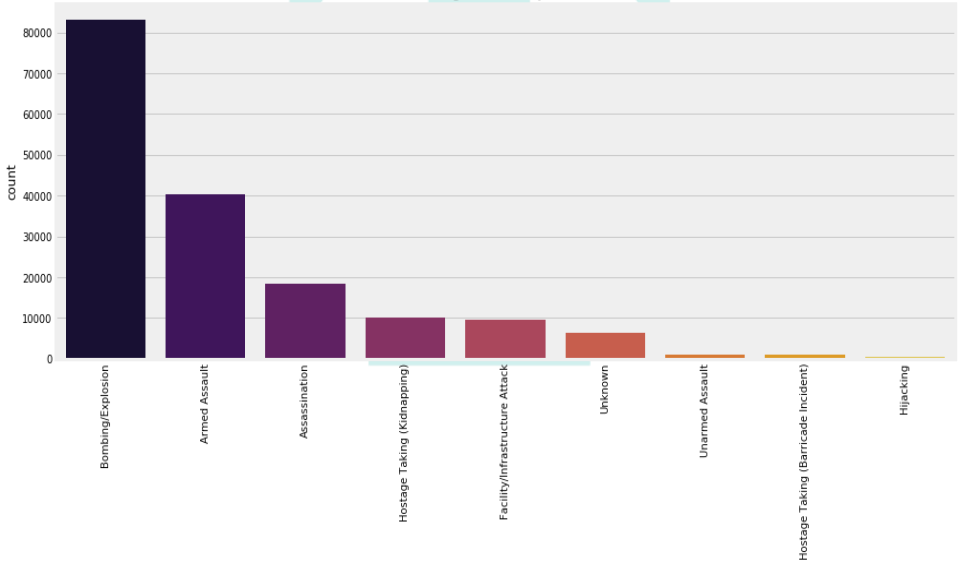

Fig. 9 Attacking Methods by Terrorists

Success Rate of Terrorism in the Top Ten Countries

Fig. 10 shows the output of the code for Success Rate of terrorism in the ten top countries. For example, from this observation, we can deduce that 90.70% of the total attacks Iraq has faced were successful. Sorting is done based on the total number of attacks for each country.

```
Success Rate in the Top Ten Countries
Iraq  :  90.70 %
Pakistan  :  88.24 %
Afghanistan  :  88.67 %
India  :  86.21 %
Colombia  :  92.53 %
Philippines  :  86.44 %
Peru  :  94.28 %
United Kingdom  :  80.63 %
El Salvador  :  98.08 %
Turkey  :  91.37 %
Thailand  :  94.34 %
Somalia  :  92.46 %
Nigeria  :  92.24 %
Spain  :  86.74 %
Yemen  :  84.98 %
Sri Lanka  :  94.18 %
United States  :  82.48 %
```

Fig. 10 Success Rate of Terrorism in the Top Ten Countries

CONCLUSION

Terrorist's use of the internet and other telecommunications devices is growing both in terms of reliance for supporting organizational activities and for gaining expertise to achieve operational goals. Tighter physical and border security has also encouraged terrorists and extremists to try to use other types of weapons to attack the target.

However, Persistent Internet and computer security vulnerabilities, which have been widely publicized, can gradually encourage terrorists to continue to enhance their computer skills, or develop alliances with criminal organizations and consider attempting a cyber-attack against targets critical infrastructure. Cybercrime has increased dramatically in past years, and several recent terrorist events appear to have been funded partially through online credit card fraud.

Thus, the researcher understands all these, as the proposed system uses one of the most critical methods (Data Mining) to detect contents online. This makes it quite difficult for anybody to post anything in form of video, photo or text without it being thoroughly scrutinized.

Combating terrorism, terror attacks and other criminal activities requires the adequate attention of the Government, law enforcement agents, individuals and even private security operatives. Security Agencies and the Government should better be equipped in this information age on how to use computer and computer analysis to track the nefarious activities of the terrorists.

Furthermore, this system can only help the law enforcement agencies and are not intended to replace them. Also, the system will not come up and say that the group is this or that rather it will help the detectives and law enforcement agencies in fighting terrorism through detecting and extracting contents.

RECOMMENDATIONS

Based on the findings of the study, the researcher generally recommends that:

i. Government agencies should setup data mining agencies within the law enforcement agencies where various data should be consolidated and mined.
ii. Data on individuals such as voter's registration, national identity, population census information, etc. should be linked together to profile the identity of an individual. Phone numbers, bank accounts and other related activities could easily be traced to any individual.
iii. Companies and organization to cooperate with the law enforcement agencies more by reporting cases of irregular transaction pattern to the government agencies on real-time basis. They do not have to wait until the law information agencies request them to provide information on some certain transaction.
iv. To complement the above, a centralized data warehouse should be established by the police where data on each individual could be established. This will enable the mapping of real data to the data mining attributes. Data mining techniques will not replace detectives or tell at once whom a terrorist is but a careful study and analysis can help identify ways by which crime could be reduce.
v. Law enforcement to be better equipped in the area of computer technology and computer analysis in order to crack some of the grueling data mining techniques and be able to link cases to suspected individuals. Adequate training and retraining should be provided for the staff of the law enforcement agencies.
vi. Government should also be encourage to invest more into communication technology as well as information technology by creating adequate enabling and necessary field for investment and development. Though much has been done recently on this regard, however other supporting factors such as adequate electricity supply, good road network to mention but a few.

REFERENCES

1. Agrawal, A., Johri, S., Agrawal, A., Tyagi, V. & Kumar, A. (2012). "Multi Agent Based Approach For Network Intrusion Detection Using Data Mining Concept". Journal of Global Research in Computer Science, 3(3), 29-32.
2. Agrawal, R. (2014). "K-Nearest Neighborn for Uncertain Data", *International Journal of Computer Applications (0975–8887)*, vol. 105, no. 11, pp. 13-16, 2014.
3. Balasubramaniyan, J.S., Garcia-Fernandez, J. O., Isacoff, D., Spafford, E., Zamboni, D. (2008). An architecture for intrusion detection using autonomous agents, *Proceedings 14th Annual*
4. Birnhack M. D. and Elkin-Koren, N. (2012). Fighting Terror On-Line: The Legal Ramifications of September 11. Internal Report, The Law and Technology Center, Haifa University. *Computer Security Applications Conference*, IEEE Comput. Soc, Los Alamitos, CA, USA.
5. Corbin, J. (2002) *Al-Qaeda: In Search of the Terror Network that Threatens the World, Thunder's* Mouth Press / Nation Books, New York.
6. Debar, H., Dacier, H., Dacier, M., Wespi, A. (2009). Towards a taxonomy of intrusion-detection systems, *Computer Networks*, **31**, 805–822.
7. Dongjin, C., Byeongkyu, K., Heesun, K., and Pankoo, K., (2014). Text analysis for detecting terrorism-related articles on the web. Journal of Network and Computer Applications, 38, 1621.
8. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (2016). From data mining to knowledge discovery in databases. AI magazine. 17(3), 37-45.
9. Han, J., Kamber, M. (2011) *Data Mining: Concepts and Techniques*, Morgan Kaufmann. http://www.apfn.org/apfn/WTC_why.htm
10. Ingram, M. (2011) Internet privacy threatened following terrorist attacks on US.

11. Jain, A. K., Murty, M. N., Flynn, P.J. (1999) Data Clustering: A Review, *ACM Computing Surveys*, **31**, 3:264-323.
12. Kelley, J. (2012) Terror Groups behind Web encryption, USA Today.
13. Lemos, R. (2002) What are the real risks of cyberterrorism?, *ZDNet*.
14. Mohammad, J. H. and Mohammad, N. O. (2009). Detecting Terror-Related Activities on the Web Using Data Mining Techniques", 2009 Second International Conference on Computer and Electrical Engineering. *Networks*, **34**:547-570.
15. Nisha, C., Mradul, D., Akhilesh, T. and Gupta, R. K. (2012). A Survey on Terrorist Network Mining: Current Trends and Opportunities. *International Journal of Computer Science Engineering Survey (IJCSES)* 3(4), 234-246.
16. Parsons, L., Haque, E., and Liu, H. (2014). Evaluating subspace clustering algorithms. In: Workshop on Clustering High Dimensional Data and its Applications, SIAM Int. Conf. on Data Mining. Citeseer; 2014. 48–56.
17. Ramesh, Y., Mayuri, D., Tejali, N., and Trupti, K. (2014). "Unauthorized Terror Attack Tracking Using Web Usage Mining", *(IJCSIT) International Journal of Computer Science and Information Technologies*, ISSN: 0975-9646, 5 (2) , 1210-1212.
18. Richards, K. (2009). Network Based Intrusion Detection: A Review of Technologies, *Computers & Security*, **18**:671-682.
19. Sequeira, K., Zaki, M. (2012) ADMIT: Anomaly-based Data Mining for Intrusions, *Proceedings of SIGKDD 02*, pp. 386-395, ACM.
20. Spafford, E.H., Zamboni, D. (2010) Intrusion detection using autonomous agents, *Computer*.
21. Thuraisingham, B. (2008). Data mining, national security, assurance and basic opportunities. SIGKDD Explorations.
22. Weiss, A., Charbonnier, E., Ellertsdóttir, E., Tsirigos, A., Wolf, C., Schuh, R., Pyrowolakis, G., Affolter, M. (2010). A conserved activation element in BMP signaling during Drosophila development. Nat. Struct. Mol. Biol. 17(1): 69-76.