WEB MINING TO DETECT ONLINE SPREAD OF TERRORISM

Aishwarya Kadam[1], Om Raskar[2], Pranav Sherkar[3], Shivrajsingh Shekhawat[4], Vishrut Shetty[5]

[1]*B. E Professor, Department of Information Technology, Shrimati Kashibai Navale College of Engineering, Pune, India*

[2]*B. E Student, Department of Information Technology, Shrimati Kashibai Navale College of Engineering, Pune, India*
[3]*B. E Student, Department of Information Technology, Shrimati Kashibai Navale College of Engineering, Pune,India*
[4]*B. E Student, Department of Information Technology, Shrimati Kashibai Navale College of Engineering, Pune,India*
[5]*B. E Student, Department of Information Technology, Shrimati Kashibai Navale College of Engineering, Pune,India*

---***---

Abstract - To fight online terrorism, integrating web mining techniques with data mining algorithms is crucial. Various algorithms like Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes (NB), Decision Trees (DT), Random Forests (RF), and Gradient Boosting (GB) are studied for detecting terrorist activities online, extracting data, identifying patterns, and relevant information. Logistic Regression provides a probabilistic framework, KNN uses similarity metrics, SVM constructs hyper-planes, NB assumes feature independence, DT builds decision trees, RF applies ensemble learning, and GB boosts weak learners' performance. These algorithms aid in proactive monitoring and prevention of online terrorism through efficient analysis of structured and unstructured web data. By combining web mining and data mining strengths, this study emphasizes a comprehensive approach to combatting online terrorism dissemination, helping security agencies anticipate evolving threats and prevent terrorist propagation effectively.

*Key Words***:** Terrorism, Naive-bayes, random forest, web mining, Gradient boosting

1.INTRODUCTION

Terrorist associations are using the internet to spread their propaganda and radicalize youth online and encourage them to commit terrorist conditioning. In order to minimize the online presence of similar dangerous websites we need to concoct a system which detects specific keywords in a particular website. The website should be flagged unhappy if the keywords are set up for effective system development.

Data mining consists of textbook mining styles that help us to overlook and prize useful content from unshaped data. Text mining helps us to descry keywords, patterns and important information from unshaped textbooks. Hence, then we plan to apply an effective web data mining system to descry similar web parcels and flag them for farther mortal review. Data mining is a fashion used to prize patterns of applicable data from large data sets and gain maximum perceptivity to the attained results. Web mining as well as data mining are used contemporaneously for effective system development.

The literature check shows the former work that has been carried out on this subject. The being systems have been explained in detail in the paper. The system that we propose to apply significantly improves the current system and eliminates the excrescencies that live in the being system. The methodology and results that we achieved after the perpetration of the proposed system have also been explained in brief further. This system should be helpful in anti-terrorism and cyber security response departments. The system should help the bobbies to track communication held between terrorists and should descry web runners developed in different platforms. The literature check shows the former work that has been carried out on this subject. The being systems have been explained in detail in the paper. The system that we propose to apply significantly improves the current system and eliminates the excrescencies that live in the being system. The methodology and results that we achieved after the perpetration of the proposed system have also been explained in brief further. This system should be helpful in anti-terrorism and cyber security response departments. The system should help the bobbies to track communication held between terrorists and should descry web runners developed in different platforms.

2. LITERATURE REVIEW

1. Detect Online Spread of Terrorism Using Data Mining
AUTHORS: Prem Bhatt, Priya Porwal, Aakash Negandhi, Soham Gawas
ABSTRACT: Terrorism has grown its roots relatively deep in certain corridor of the world. With adding terrorist conditioning, it has come important to check terrorism and stop its spread before a certain time. So as linked internet is a major source of spreading terrorism through speeches and vids. Terrorist associations use internet to brain marshland individualities and promote terrorist conditioning through instigative web runners that inspire helpless people to join terrorist associations. So then we propose an effective web data mining system to descry similar web parcels and flag them automatically for mortal review.

2. A Schematic Approach on Web Data Mining In Online Spread Discovery of Terrorism

AUTHORS: P. Vara Prasad Rao, P.S.V. Srinivasa Rao, B. Sankara Babu, A. Sai Hanuman, G. Charles Babu

ABSTRACT: Dread grounded tyrannizer advancement has extended in unequivocal bits of the world. Dread grounded tyrannizer packs use Facebook, WhatsApp, dispatches to spread their information on the casual network. It's crucial to fete internal fighting and keep its spreading before a particular time. The introductory idea of this undertaking is to reduce or stop spreading of internal mistreatment and to clear all of these records. A dread- grounded tyrannizer is spreading their internal fighting conditioning using the web by talk, content, accounts. Dread grounded tyrannizer clusters are exercising the web as a medium to move irreproachable people to partake in internal tyrannizer practices by goading the each- inclusive community through webpage runners that move worried individualities to take an interest in the dread monger cooperation. This needs a lot of mortal trouble to execute this bid will assemble the information and find the dread monger social affairs. To dwindle the mortal trouble, we execute the structure which perceives dread monger groups in online life. Mental persecution has erected up its retired establishments exceptionally some place down in unequivocal bits of the world. With extending dread- grounded tyrannizer works out, it has ended up being introductory to check internal persecution and stop its spread before a particular time. So as perceived web is an imperative wellspring of spreading dread mongering through addresses and accounts. Dread monger confederations use web to rationally condition individualities and advance internal raider practices through instigative point runners that convert helpless people to join internal tyrannizer confederations. So then we propose a compelling web data mining structure to fete similar web parcels and standard them consequently for mortal review.

3. Saja Aldera, Ahmad Emam, Muhammad Al-Qurishi Majed Alrubaian 3, and Abdulrahman Alothaim Online Extremism Detection in Textual Content: A Systematic Literature Review

The escalating use of social media platforms by extremist groups to disseminate radical ideologies and recruit followers has prompted a pressing need for effective countermeasures. This paper conducts a systematic literature review (SLR) covering 45 studies published between 2015 and 2020, with a primary focus on understanding and defining extremism in the context of online platforms such as Twitter, Facebook, YouTube, blogs, and discussion forums. Identifying gaps in the existing literature, the review highlights challenges and technical pitfalls in previous studies while presenting opportunities for advancements in research methodologies. The critical analysis presented in this study underscores the paramount importance of addressing extremist content and profiles on social media, emphasizing its status as a top priority for counter-terrorist agencies, technology companies, and governments. The findings stress the need for a more nuanced understanding of the landscape and directions of online extremism, providing a foundation for future research to enhance the comprehension and mitigation of this evolving threat. This paper contributes to the emerging field with valuable insights and recommendations for further exploration and development in countering extremism in the digital realm.

4. Unauthorized Terror Attack Tracking Using Web Usage Mining

AUTHORS: Ramesh Yevale, Mayuri Dhage, Tejali Nalawade, Trupti Kaule.

ABSTRACT: Terrorist groups use the Web as their structure for colorful purposes. One illustration is the forming of new original cells that may latterly come active and perform acts of terror. The Terror Tracking using Web operation Mining (TTUM) is aimed at tracking down online access to abnormal content, which may include terrorist- generated spots, by assaying the content of information penetrated by the Web druggies. TTUM operates in two modes the training mode and the discovery mode. In the training mode, TTUM determines the typical interests of a prespecified group of druggies by recycling the Web runners penetrated by these druggies over time. In the discovery mode, TTUM performs real- time monitoring of the Web business generated by the covered group, analyzes the content of the penetrated Web runners, and issues an alarm if the penetrated information isn't within the typical interests of that group and analogous to the terrorist interests. An experimental interpretation of TTUM was enforced and estimated in a original network terrain. An innovative knowledge- grounded methodology for terrorist shadowing by using Web business content as the inspection information is presented. The proposed methodology learns the typical practices ('profile') of terrorists by applying a data mining algorithm to the textual content of terror- related Web spots. The performing profile is used by the system to perform real- time discovery of druggies suspected of being engaged in terrorist conditioning. The Receiver- Operator Characteristic (ROC) analysis shows that this methodology can outperform a command grounded intrusion discovery system.

5. Enhanced Content Discovery system to Detect Online Spread of Terrorism

AUTHORS: K. Monisha, S. Vaishnavi, J. Kiruba, P. Sumitha

ABSTRACT: Terrorist growth reached to certain part of the world. Terrorist groups have spread their information on social network like facebook, whatsapp, dispatches. Before spreading the terrorism conditioning, we need to change and repeal it. The introductory idea is to stop or reduce spreading of terrorism is to remove all this accounts. Web is a noteworthy wellspring of spreading cerebral warfare by converse, content, sound arrangement. Terrorists use internet to convert individualities and it paves the way to involve in terrorist conditioning through instigative webpage that inspires individualities to involve in terrorist groups and association. To execute this study we need

a great deal of mortal exertion to gather the data and discover the fear monger gatherings. To reduce the mortal trouble, we apply the system which detects terrorist groups in social media. To apply this idea we're in need of a lot of mortal trouble to gather the information and find out the terrorist groups who are involved in. therefore we apply this system for major causes which detects terrorist groups in internet and social media. It paves the way to reduce the mortal trouble.

3. PROPOSED SYSTEM

We propose a system with the primary thing of developing a website where druggies can check any webpage or any website for any trace of terrorist exertion. To do so, our website will give the point of entering the URL of the webpage the stoner wants to overlook. After entering the URL, our system will census the words of the whole webpage and census them with the words that are formerly present in our database. Each word that we will store in our database will have a certain score to it. Our system will cost the scores of each word that's present in the stoner's web runner from our database, and in the end it will calculate a total rank of the website. This rank will determine if the stoner's webpage contains any trace of terrorism or not. Our system will descry patterns, keywords and applicable information in unshaped textbooks in a webpage using web mining as well as data mining. Our system will booby-trap webpage using web mining algorithm to mine textual information on web runners and descry those web runners that are applicable to terrorism. Data booby-trapping as well as web mining is used together at times for effective results. Traditionally, there was no similar system to keep an eye on colorful websites or any suspicious words present online. Bobbies were unfit to track the terrorist related website or any person with suspicious information. The rate of terrorism is high in moment's world. There must be a system to track those suspicious word online and bring down the rate of terrorism. In colorful arrangements and have images, videos etc. combined on a single web runner. So we then propose to use dashingly designed web mining algorithms to mine textual information on web runners and descry their applicability to terrorism. In this way we may judge web runners and check if they may be promoting terrorism. This system proves useful in anti-terrorism sectors and indeed search machines to classify web runners into the order. Their applicability to the field helps classify and sort them meetly and flag them for mortal review.
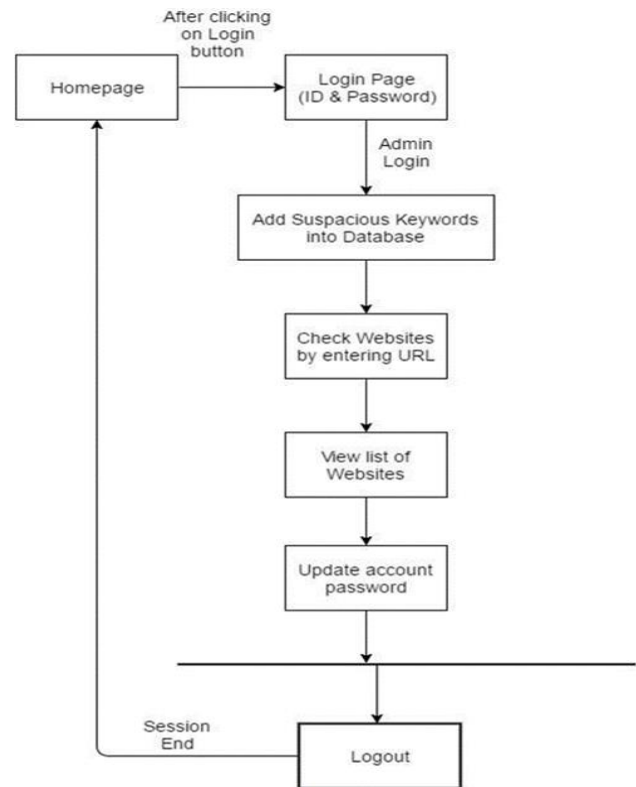

Fig - 1. System block diagram

4. IMPLEMENTATION DETAILS

Upon visiting the website for the first time, individuals are required to create an account by filling in specific fields with details such as their preferred username, full name, a valid email address, mobile number, and a secure password. After the completion of this registration process, users will be directed to the designated sign-in page, where they can input their chosen credentials, consisting of the username and password they previously selected. Upon successfully logging in to their account, users will be greeted with a prominently featured section labeled "url," providing them with an opportunity to enter the web address of any site they desire to examine for potential terrorist-related content. By simply clicking the designated "check" button, the system will swiftly generate a safety percentage related to the specified website.
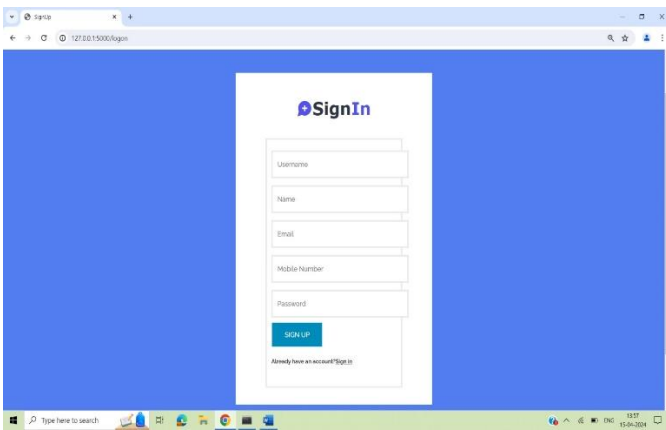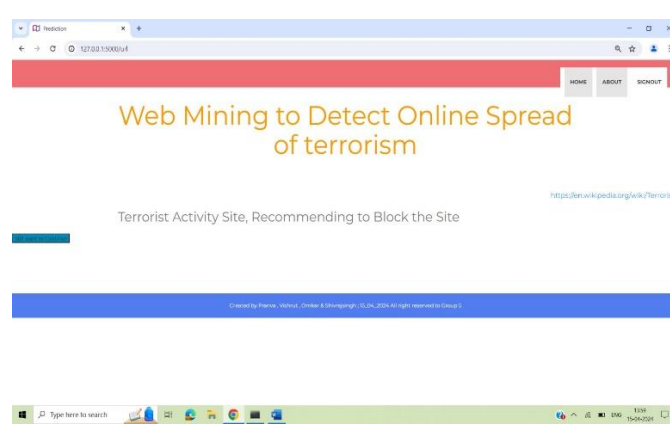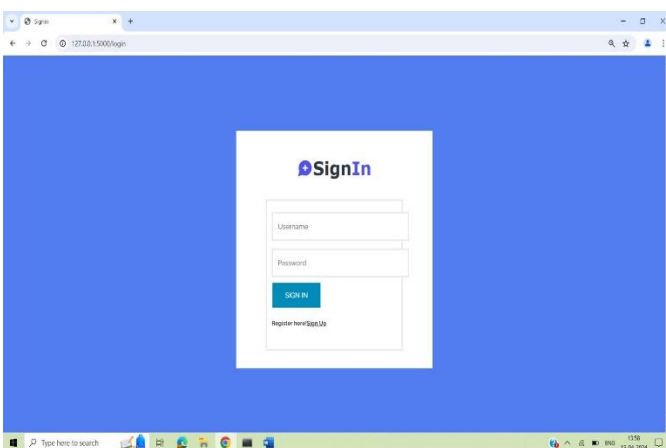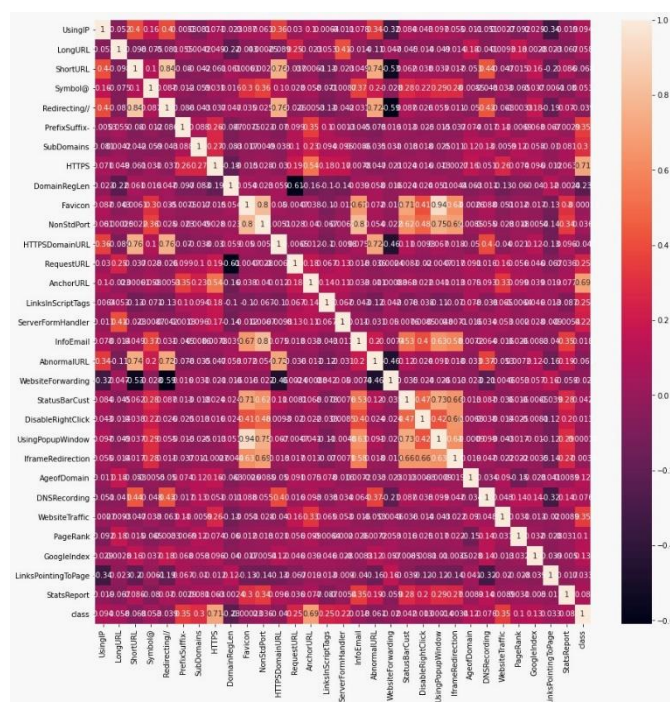
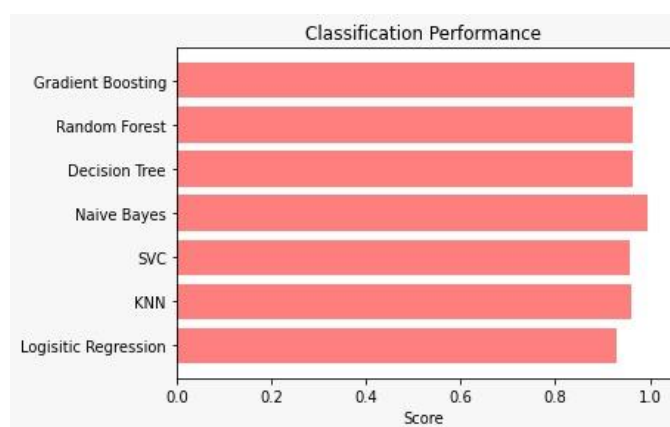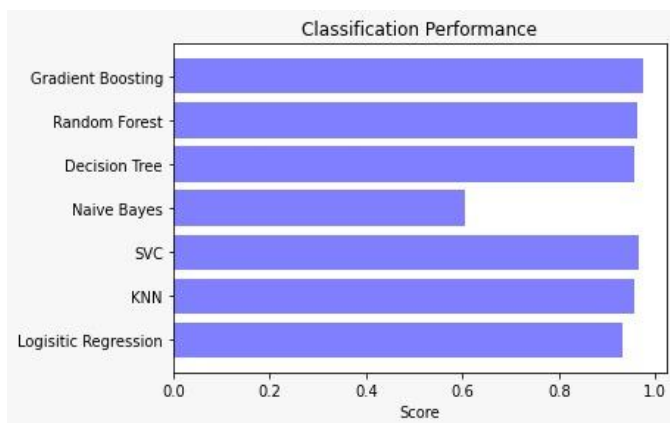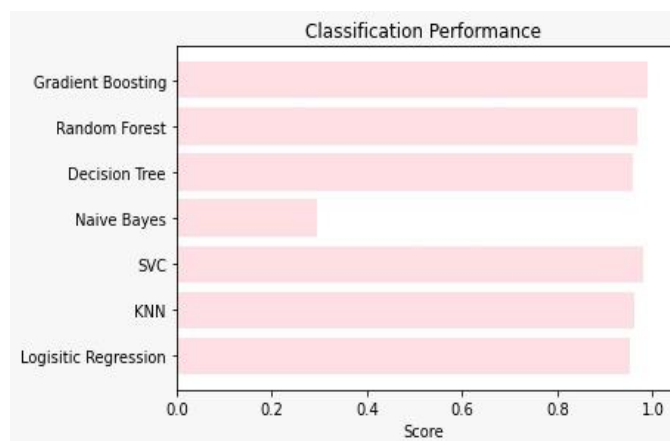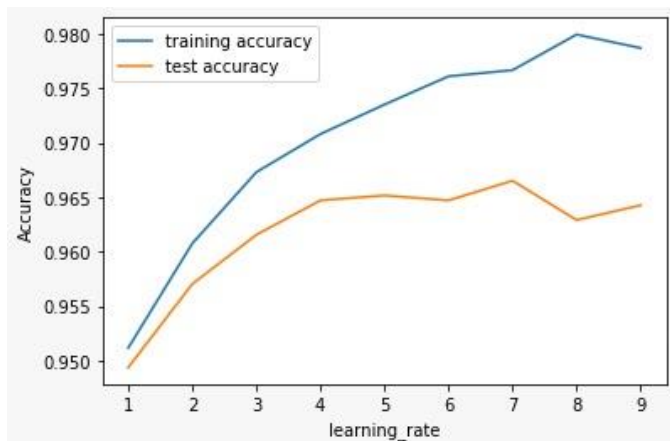1. Home page
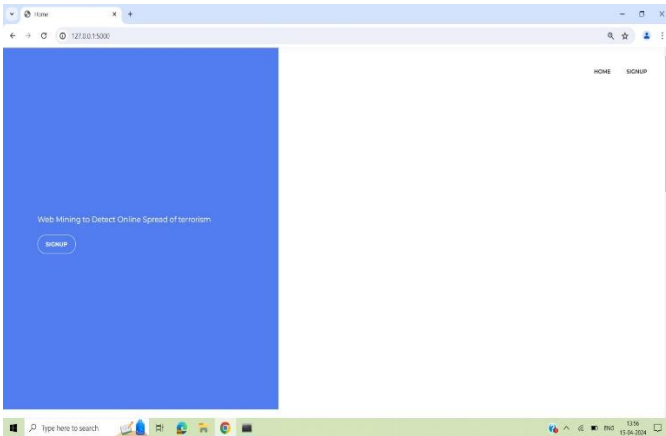
4. URL
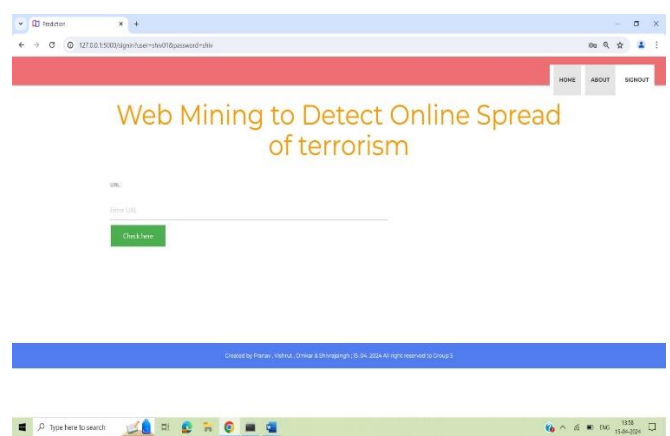
2. SignUp

5. Result

3. SignIn

5. RESULT ANALYSIS
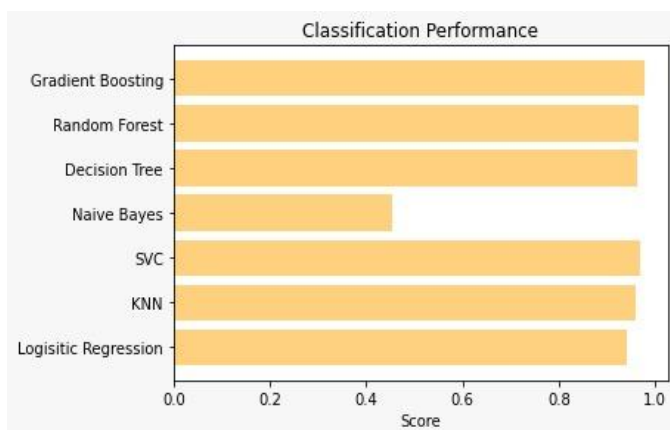
5. CONCLUSIONS

In conclusion, the integration of web mining techniques with a comprehensive suite of data mining algorithms, including Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes (NB), Decision Trees (DT), Random Forests (RF), and Gradient Boosting (GB), represents a robust and versatile solution in the ongoing battle against online terrorism. This amalgamation of methodologies equips security agencies with powerful tools to detect and mitigate terrorist activities on the internet effectively.

By harnessing the capabilities of these diverse algorithms, the system demonstrates remarkable versatility and adaptability. It enables proactive monitoring and rapid response to emerging threats by parsing through diverse web page structures and extracting pertinent information. This not only enhances detection accuracy but also minimizes false positives, optimizing resource allocation and operational efficiency for counter-terrorism efforts.

One of the system's key strengths lies in its ability to identify subtle patterns and indicators of terrorist behavior, thereby enhancing its effectiveness in safeguarding against online radicalization and recruitment. Moreover, its modular design ensures scalability and adaptability, enabling it to evolve alongside the dynamic online landscape and emerging threats.

In essence, the proposed system serves as a critical asset in bolstering global security measures and fostering a safer online environment. By leveraging advanced web mining and data mining techniques, it empowers security agencies to stay ahead of the curve in combating the online spread of terrorism. As the digital realm continues to evolve, this integrated approach remains pivotal in safeguarding against the ever-changing tactics employed by terrorist organizations, ultimately contributing to a more secure and resilient online ecosystem.

REFERENCES

1. Anagha Jagtap1, Amrita Nayak2, Manaswee Adwant3, Reetika Kerketta. Detection of Online Spread of Terrorism International Journal of Engineering Research Technology.

2. Goradia, R., Mohite, S., Jhakhariya A., Pinjarkar, V. (2020, June). Web Mining to Detect Online Spread of Terrorism. International Journal of Engineering Research Technology (IJERT), 9(7), 645–648. http://www.ijert.org.

3. Negandhi A., Gawas, S., Bhatt, P., Porwal, P. (2018, March). Detect Online Spread of Terrorism Using Data Mining. IOSR Journal of Engineering (IOSRJEN), 13, 17–19. https://www.iosrjen.org..

4. Ahsan, S., Shah, A. (2008). Data mining, semantic web and advanced information technologies for fighting terrorism. 2008 International Symposium on Biometrics and Security Technologies,15. https://doi.org/10.1109/isbast.2008.4547644

5. Shaikh, M. A., Jiaxin, W. (2006). Investigative Data Mining: Identifying Key Nodes in Terrorist Networks. 2006 IEEE International Multitopic Conference, 201–206. https://doi.org/10.1109/inmic.2006.358163.

6. Sachan, A. (2012). Countering terrorism through dark web analysis. 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12) https://doi.org/10.1109/icccnt.2012.6396055