Detection of online Terrorism using keywords

Yashwanth A
yashwanth476cs@gmail.com
M. S. RAMAIAH UNIVERSITY OF APPLIED SCIENCES Bengaluru

Sushma Yadav M M
sushmaydv@gmail.com
M. S. RAMAIAH UNIVERSITY OF APPLIED SCIENCES Bengaluru

Nagavi S R
nagavisr@gmail.com
M. S. RAMAIAH UNIVERSITY OF APPLIED SCIENCES Bengaluru

Shree Ganesh A
nisargashree1872@gmail.com
M. S. RAMAIAH UNIVERSITY OF APPLIED SCIENCES Bengaluru

Divya BS
divyabies@gmail.com
M. S. RAMAIAH UNIVERSITY OF APPLIED SCIENCES Bengaluru

***Abstract* -- This paper tackles the problem of spotting online terrorism on Hindi websites. With the rise of digital platforms, extremists are using new ways to spread their ideas. This study suggests a new method that looks at keywords to find and keep an eye on potential terrorism-related content in Hindi online spaces. Hindi keywords had been recognized linked to extremist ideas to make this system better at finding online terrorism. The advanced Python and Tkinter techniques were used to analyze content, creating a proactive plan for spotting and dealing with possible threats. This approach adds to the conversation about stopping online radicalization by paying attention to the unique language used on Hindi online platforms. This research not only helps us understand online terrorism detection better but also gives practical ideas for making targeted tools and policies for Hindi digital spaces.**

Keywords—Terrorism, Terrorist organization, cyber security, security Threat, potential threats.

I. INTRODUCTION

The internet is being used by terrorist groups to propagate their message, radicalize young people, and inspire them to carry out acts of terrorism. A system must be developed that recognizes particular terms on a given website in order to reduce the amount of dangerous websites that are available online. If the terms for effective system development are found on the website, it ought to be marked as inappropriate. Text mining techniques are a subset of data mining that enable us to search through unstructured data and extract relevant material [6].

From unstructured texts, text mining enables us to extract critical information, trends, and keywords. Therefore, in order to identify certain web properties and flag them for additional human evaluation, appropriate technologies like web data mining system is used here. Data mining is a technique used to maximize insights into the results gained by extracting meaningful patterns from massive data sets. Data mining and web mining are applied concurrently for effective system development [7].

The literature review displays the prior research that has been done on this topic. The document provides a detailed explanation of the current systems [14]. The system was intended to put into place greatly enhances the current system and gets rid of its shortcomings. A brief explanation of the technique and outcomes were attained following the installation of the suggested system has also been provided. Departments dealing with counterterrorism and cyber security response ought to find this system useful. Technologies that ought to make it easier for law enforcement to follow held between terrorists and ought to identify websites created on various platforms.

II. LITERATURE SURVEY

Online content information is the most prevalent form of material found on the internet, reflecting the creator's viewpoint. With the recent advancements in remote internet and mobile devices, such as iPhones, the volume of online data has significantly increased without any constraints on time or location [1],[4]. This paper introduces a method for identifying terrorist activities based on the keywords found on websites. By analyzing specific URLs, the system determines the locations where terrorist activities are occurring. This approach has been tested in various countries, encouraging citizens to report ongoing terrorism activities to their respective governments. Unlike other research papers as shown in table 1[4],[7]. that rely solely on English language algorithms, This strategy does not incorporate machine learning algorithms. Content analysis is employed to uncover obscure but legitimate patterns and connections within extensive data sets.

Sl No	Paper	Algorithms	Scope	Research Gaps
1	Aakash Negandhi, SohamGawas, Prem Bhatt , PriyaPorwal "Detect Online Spread of Terrorism Using Data Mining"	•Logistical regression • Decision Tree • Random Forest	•Finds the words that can be pegged as related to terrorism	Data quality issues and ethical considerations, as well as challenges in defining relevancy and avoiding false positives.
2	T.Anand "Terror Tracking Using Advanced Web Mining" AYCCollege of Engg Mayiladuthurai, India	• Decision Tree • Naïve Bayes.	•Uses WEKA •Finds the sentiments of the words	Potential challenges in feature selection and the creation of a reliable scoring system for words in sentiment analysis, as well as scalability concerns in handling multiple languages effectively.
3	Fawad Ali, Farhan Hassan Khan, Saba Bashir, and Uzair Ahmad, Counter Terrorism on Online Social Networks Using Web Mining Technique. Department of Computer Science, Federal Urdu University of Arts, Science and Technology (FUUAST), Islamabad, Pakistan	• Naïve Bayes • KNN • Decision Tree • Logistical regression • Random Forest	•Uses various techniques like facial recognition. • Uses text mining on OSN	The gaps in methods using facial recognition and text mining on Online Social Networks include privacy and ethical concerns as well as challenges related to data quality, context understanding, bias, scalability, and legal compliance.
4	Naseema Begum A. Detection of online spread of terrorism using web data mining. Institute of Engineering and Technology, Coimbator	• Decision Tree • Random Forest	Performs a well defined cleaning of data and also data storage.	The downfalls of data cleaning and storage mechanisms can include data loss, resource intensity, complexity, storage costs, security risks, scalability challenges, data silos, redundancy, governance issues, and ongoing maintenance burdens, requiring careful management and investment.

Table 1. Literature survey

III. PROPOSED SYSTEM

To track websites built on various platforms, with various algorithms, and with various programming languages. The suggested method can determine whether websites and online content are endorsing and disseminating acts of terrorism, and it can also identify and filter propaganda linked to terrorism. The suggested technique is used to find and examine websites, categorize them as legitimate users and those connected to terrorism, and identify them as either terrorists or regular users.

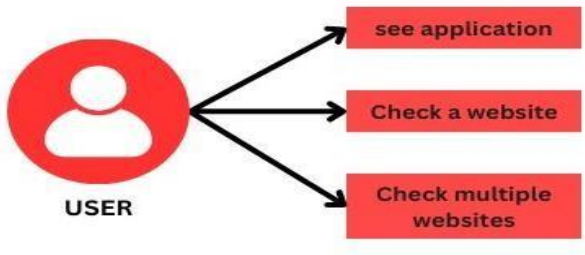

Figure 1. Sequence diagram

Users (or actors) and a system communicate visually through a sequence diagram, as shown in Figure 1. It gives a high-level summary of the capabilities or functionalities the system has to offer. The user can view applications, switch themes, browse websites, and browse numerous websites from this location.

User interaction with the Application, the user initiates the sequence by accessing the application, symbolized by the 'User' lifeline. Upon entering the application, the user is presented with various functionalities, including the ability to check websites for potential securityconcerns.

Checking websites for Security concerns, Captures the user's action of checking websites for potential security threats. As the user engages with the 'Check Websites' functionality, the application initiates a series of operations. This involves the application sending requests to external websites for keyword analysis and pattern recognition, examining the websites for any indications of terrorist activities. The user, represented by the lifeline, remains engaged in the process while the application conducts the necessary checks.

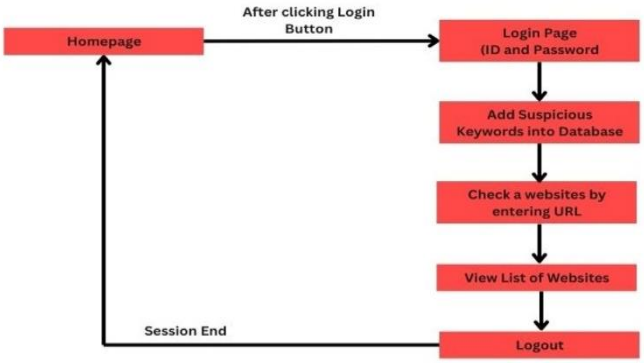

Figure 2. workflow

The workflow diagram as shown in Figure 2. delineates the sequential processes and user interactions within the project, encompassing essential functionalities such as navigating the homepage, accessing the login page, managing suspicious keywords, viewing a list of websites, and logging out.

Homepage access is the workflow commences with the user accessing the homepage of the application. This initial point serves as the entry into the system.

Login page interaction is the user proceeds to the login page, indicating the need for authentication. At this stage, the user provides their credentials (username and password) to gain access to the application. The application verifies the provided credentials, granting access upon successful authentication.

Add suspicious keyword is nothing but Once authenticated, the user navigates to the feature for managing suspicious keywords. This involves inputting specific terms that the system will monitor for potential security concerns. The application processes the entered keywords, incorporating them into its monitoring system.

View list of websites is with suspicious keywords set, the user can now view a list of websites that are being monitored for the specified terms. This step involves the application's retrieval and presentation of relevant data based on the defined criteria.

Logout is user having completed their tasks, initiates the logout process. This involves the termination of the user's session and the return to the login page or homepage. The application ensures the secure logout of the user, safeguarding the integrity of the system

IV. MODULES

Subprocess Module used to install external libraries (requests, bs4, ttkthemes, lxml, tkinter, PIL) using pip. The subprocess. check_ call function is called with a list of arguments specifying the Python interpreter (sys.executable), the -m flag for module, and the library name to install[10].

Tkinter Module used for creating the main application window (MyApp) and various GUI components like frames, labels, buttons, entry boxes, etc. Tkinter is the standard GUI toolkit for Python. You create instances of Tkinter classes to build the graphical user interface. Widgets like Frame, Label, Button, Entry, and others are used to construct the user interface elements[10].

Requests Module used to fetch the content of a webpage for analysis. The requests.get(url) function sends an HTTP GET request to the specified URL (url). The response is then used for further processing, such as extracting content with BeautifulSoup[10].

BeautifulSoup Module used for parsing and extracting text content from the webpage.BeautifulSoup provides functions to parse HTML or XML documents. In this code, BeautifulSoup(result.content, 'lxml') is used to parse the content of the webpage obtained through the requests module[10].

PIL (Pillow) Module used to open and display an image. The Image.open() method opens an image file, and ImageTk.PhotoImage is used to create a Tkinter-compatible image. The image is then displayed using ttk.Label[12].

ThemedStyle Module (from ttkthemes) used to set the theme of the Tkinter application.ThemedStyle is used to set

the theme of the Tkinter application. In this code, the theme is set to "plastik" with style.set_theme("plastik").

Python - Text Translation: Written in Python, translate is a straightforward yet effective translation tool that supports a variety of translation services. It can be integrated with DeepL's free and pro APIs, Microsoft Translation API, Translated MyMemory API, and LibreTranslate.

V. RESULTS AND ANALYSIS

The input illustrates the crucial input configuration aspect of the system, where the user contributes Hindi keywords and URL links associated with terrorism activities.

Simultaneously, the user provides URL links that are associated with terrorism activities. These links may lead to websites or online platforms where extremist content is disseminated. By incorporating these URLs into the system, the application extends its reach to actively monitor and analyze the specified web locations.

By combining Hindi keywords and terrorism-related URLs as inputs, the system establishes a comprehensive framework for detecting and preemptively addressing potential security threats. This input configuration ensures that the application remains adaptive to the evolving landscape of online extremist activities, contributing to the overall efficacy of the counterterrorism strategy. The subsequent stages of the system's workflow, as depicted in the previously discussed diagram, leverage these inputs to black-list suspicious websites, effectively curbing the online footprint of terrorist propaganda and activities.

The user inputs specific Hindi keywords relevant to extremist ideologies, terrorist activities, or radicalization. These keywords serve as linguistic markers that the system will monitor within the content of websites, aiming to identify potential threats.

The homepage as shown in Figure 3. serves as the initial point of interaction, presenting a user-friendly interface to the application. Users can easily navigate through different features and access relevant functionalities from this central hub.

Figure 3. Home page
(picture belongs to one of the member)

Detection page as shown in Figure 4. Upon successful login and input configuration, users are directed to the detection page. This page visually communicates the system actively monitoring for suspicious activities based on the provided Hindi keywords and terrorism-related URLs. It may display real-time status indicators or logs, conveying the ongoing detection processes.

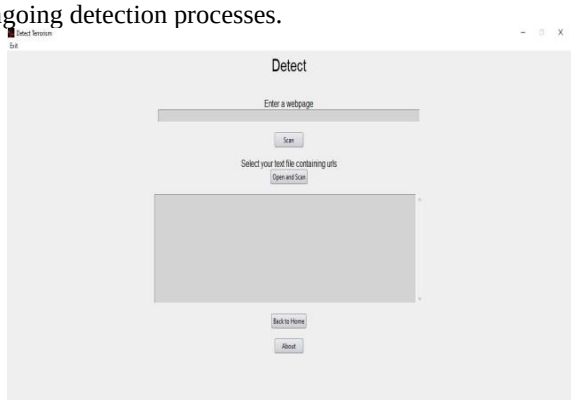

Figure 4. Detection page

The scan button allows users to perform specific queries or investigations. Users can input keywords, URLs, or other parameters to conduct targeted searches within the system. This button serves as a dynamic tool for users to explore and gather information related to the detection of terrorism activities.

Searched Results as shown in figure 5. Here In this process redacted the URL in the output to avoid disclosing the specific website. system's analysis and detection processes, the results provide a detailed output. This could include a list of websites flagged for suspicious content, associated metadata, and any relevant information indicating potential security threats. The results are presented in a clear and comprehensible format for the user's review and further action. The performance matrix percentage, denoted as PMP, can be calculated using the formula:

$$PMP = \frac{\text{Number of Matched Keywords}}{\text{Total Number of Keywords}} \times 100$$

This formula quantifies the percentage of matched keywords related to terrorism and violence in the given text, providing a measure of the extent to which such topics are discussed. It displayed an output of 102 keywords matched for the particular URL.

Figure 5. Searched result

The about page as shown in figure 6. furnishes users with contextual information about the application, such as its purpose, features, and the methodology employed for terrorism detection. This section contributes to user awareness and understanding, fostering transparency regarding the system's objectives and functionalities.

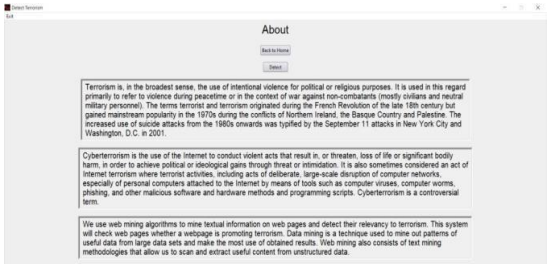

Figure 6. About page

The methods and knowledge for detecting web-based terrorist activities are presented in this paper. This study provides ideas and techniques for identifying terrorists, their actions, and those who encourage them to engage in terrorist-related activities in society by using an Internet connection. Since these acts of terror are a significant drain on society, the people must raise awareness of online terrorism and ensure that no one in the immediate vicinity falls victim to it.

VI. CONCLUSIONS

To assess the possibility of psychological oppression and eliminate the online presence of harmful fear mongering organizations, like ISIS and other websites that promote radicalization. To find and close down websites that spread dangerous content meant to radicalize children and defenseless adults, a sound foundation is necessary. In addition to developing a web-based platform, this project has introduced a number of features that will help with the real-time application of the system. Any admin manager can use this application to distribute it to their administrators so they can keep an eye out for any signs of terrorism spreading. The methods and knowledge for detecting web-based terrorist activities are presented in this paper. The concepts and procedures developed in this study can be used to identify terrorists, their actions, and the people who encourage them to engage in terrorist-related activities in society by using an Internet connection. By preventing people from becoming radicalized by internet websites and other online media, the developed application will stop the spread of terrorism. Because terrorist organizations use the internet as a media platform to spread and promote terrorism, people are prevented from accessing these websites and media. As a result, terrorism is prevented.

VII. REFERENCES

[1] Negandhi, A., Gawas, S., Bhatt, P. and Porwal, P., 2019. Detect Online Spread of Terrorism Using Data Mining. IOSR Journal of Engineering, 13.

[2] Gordon, A., 1998. The spread of terrorism publications: A database analysis

[3] Hanuman, A.S., Babu, G.C., Rao, P.V.P., Rao, P.S. and Babu, B.S., A Schematic Approach on Web Data Mining In Online Spread Detection of Terrorism. International Journal of Recent Technology and Engineering, 8.

[4] Ali, F., Khan, F.H., Bashir, S. and Ahmad, U., 2019. Counter terrorism on online social networks using web mining techniques. In Intelligent Technologies and Applications: First International Conference, INTAP 2018, Bahawalpur, Pakistan, October 23-25, 2018, Revised Selected Papers 1 (pp. 240-250). Springer Singapore.

[5] Abbasi, A., Chen, H. and Salem, A., 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. ACM transactions on information systems (TOIS), 26(3), pp.1-34.

[6] Kiruba, J., Sumitha, P., Monisha, K. and Vaishnavi, S., Enhanced Content Detection Method to Detect Online Spread of Terrorism. International Journal of Engineering and Advanced Technology, 8.

[7] Begum, N., Mohanambal, R. and Aswathy, R.H., 2019. Detection of online spread of terrorism using web data mining. A. Institute of Engineering and Technology, Coimbatore, Tamil Nadu, International Journal of Advance Research, Ideas and Innovations in Technology, 5(1).

[8] Anand, T., Padmapriya, S. and Kirubakaran, E., 2009, July. Terror tracking using advanced web mining perspective. In 2009 International Conference on Intelligent Agent & Multi-Agent Systems (pp. 1-4). IEEE.

[9] Vidic, M., 2009. The use of the internet for terrorist purposes. Journal of Criminal Investigation and Criminology/Ljubljana, 60(3).

[10] Agarwal, S., 2013, December. Data mining: Data mining concepts and techniques. In 2013 international conference on machine intelligence and research advancement (pp. 203-207). IEEE.

[11] Han, J., Pei, J. and Tong, H., 2022. Data mining: concepts and techniques. Morgan kaufmann.

[12] Tan, P.N., Steinbach, M. and Kumar, V., 2016. Introduction to data mining. Pearson Education India.

[13] Goradia, R., Mohite, S., Jhakhariya, A. and Pinjarkar, V., 2020. Web Mining to Detect Online Spread of Terrorism. International Journal of Engineering Research & Technology (IJERT), 9(7), pp.645-648.

[14] Coglitore, A., 2017. Terrorism spreading by way of the internet (Doctoral dissertation, Utica College).

[15] Gialampoukidis, I., Kalpakis, G., Tsikrika, T., Papadopoulos, S., Vrochidis, S. and Kompatsiaris, I., 2017, June. Detection of terrorism-related twitter communities using centrality scores. In Proceedings of the 2nd international workshop on multimedia forensics and security (pp. 21-25).