International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03478
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Web Mining to Detect Online Spread of Terrorism

S.Dinesh Kumar Reddy
UG Student Dept of CSE(CTIS) Jain University
Kanakapura,Karnataka,India.

R.Purushotham
UG Student Dept of Cse(CTIS)
Jain University Kanakapura,Karnataka,India.

B.Dheeraj
UG Student Dept of CSE(CTIS)
Jain University Kanakapura,Karnataka,India

Dr.N.Raja Praveen Professor
Dept of CSE-CTIS Jain University
Kanakapura,Karnataka,India.

Abstract: **In the recent times, terrorism has grown in an exponential manner in certain parts of the world. This enormous growth in terrorist activities has made it important to stop terrorism and prevent its spread before it causes damage to human life or property. With development in technology, internet has become a medium of spreading terrorism through speeches and videos. Terrorist organizations use the medium of the internet to harm and defame individuals and also promote terrorist activities through web pages that force people to join terrorist organizations and commit crimes on the behalf of those organizations. Web mining and data mining are used simultaneously for the purpose of efficient system development. Web mining even consists of many different text mining methods that can be helpful to scan and extract relevant data from unstructured data. Text mining is very helpful in detecting various patterns, keywords, and significant information in unstructured texts. Data mining and web mining systems are used for mining from text widely. Data mining algorithms are used to manage organized data sets and web mining algorithms can be helpful in mining and extracting from unstructured web pages and text data that is available across the web. Websites built in different platforms have varying data structures and that makes it quite difficult to read for a single algorithm.**

Keywords: Terrorism, naïve-bayes, random forest, online spread

I. INTRODUCTION

Terrorist organizations are using the internet to spread their propaganda and radicalize youth online and encourage them to commit terrorist activities.In order to minimise the online presence of such harmful websites we need to devise a system which detects specific keywords in a particular website. The website should be flagged inappropriate if the keywords are found for efficient system development. Data mining consists of text mining methods that help us to scan and extract useful content from unstructured data. Text mining helps us to detect keywords, patterns and important information from unstructured texts. Hence, here we plan to implement an efficient web data mining system to detect such web properties and flag them for further human review. Data mining is a technique used to extract patterns of relevant data from large data sets and gain maximum insights to the obtained results. Web mining as well as data mining are used simultaneously for efficient system development. The literature survey shows the previous work that has been carried out on this subject. The existing systems have been explained in detail in the paper.The system that we propose to implement significantly improves the current system and eliminates the flaws that exist in the existing system.The methodology and results that we achieved after the implementation 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 cops to track communication held between terrorists and should detect web pages developed in different platforms.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03478
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

III. LITERATURE REVIEW

[1.]S,DineshKuma reddy et al. applied various machine learning algorithms in “Detect Online Spread of Terrorism Using Data Mining” to mine textual information on web pages and detect their relevancy to terrorism.

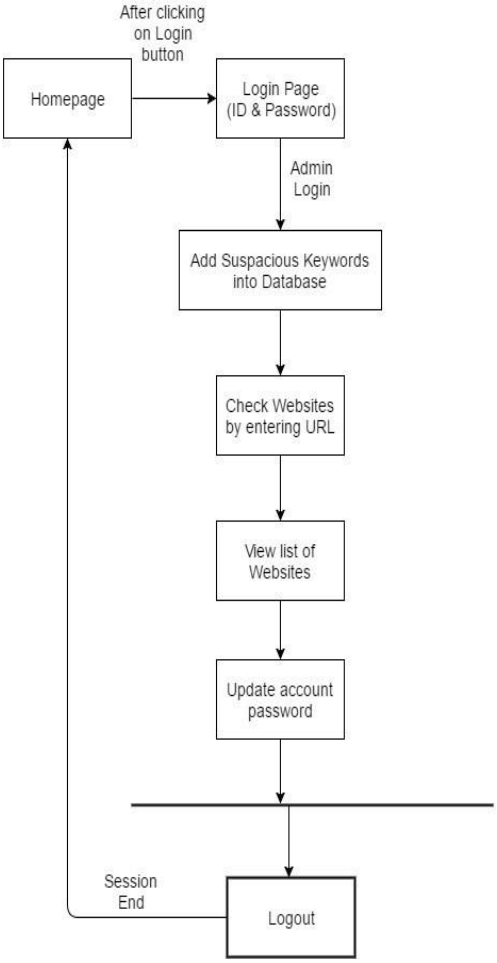
[2.]B.Dheeraj, H. et al. used the features of sentiment analysis to segregate the words of a web page, classify them and assert a score to each word in "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums."

[3.] R.Purushotham et al. studied various methods by which textual data can be fetched and scanned and executed them to counter Terrorism on Online Social Networks using web mining techniques.

TABLE 1. COMPARISON OF EXISTING SYSTEM

Sr No.	Paper	Algorithms	Scope
1.	S.Dinesh Reddy “Detect Online Spread of Terrorism Using Data Mining”	• Logistical regression • Decision Tree • Random Forest	• Finds only the words that can be pegged as related to terrorism
2.	R.Purushotham“Terror Tracking Using Advanced Web Mining” AYCCollege of Engg Mayiladuthurai, India.	Decision Tree Naïve Bayes.	Uses WEKA • Finds the sentiments of the words
3.	Counter Terrorism on Online Social Networks Using Web Mining Technique Fawad Ali, Farhan Hassan Khan, Saba Bashir, and Uzair Ahmad, 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

4.	Detection of online spread of terrorism using web data mining Naseema Begum A. Institute of Engineering and Technology, Coimbator	Decision Tree Random Forest	• Performs a well defined cleaning of data and also data storage.



IV. PROPOSED SYSTEM

We propose a system with the primary goal of developing a website where users can check any webpage or any website for any trace of terrorist activity. To do so, our website will provide the feature of entering the URL of the webpage the user wants to scan. After entering the URL,our system will tally the words of the whole webpage and tally them with the words that are already present in our database. Each word that we will store in our database will have a certain score to it. Our system will fetch the scores of each word that is present in the user's web page from our database, and in the end it will calculate a total rank of the website.

This rank will determine if the user's webpage contains any trace of terrorism or not.

Our system will detect patterns, keywords and relevant information in unstructured texts in a webpage using web mining as well as data mining. Our system will mine webpage using web mining algorithm to mine textual information on web pages and detect those web pages that are relevant to terrorism. Data mining as well as web mining is used together at times for efficient results.

Machine Learning algorithms:

- Random Forest:
  Random forest algorithm, like its name implies, consists of a large number of individual decision trees that operate together. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. [Deziel, M. et al.]

- Decision Tree:
  A decision tree is a tree-like graph with nodes representing the place where we pick an attribute and ask a question; edges represent the answers to the question; and the leaves represent the actual output or class label. They are used in non-linear decision making with a simple linear decision surface.

- Naïve Bayes:
  Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

- Logistic Regression:
  The logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

- K-nearest Neighbours:
  K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non- parametric technique.

Figure 1. System block diagram

Traditionally, there was no such system to keep an eye on various websites or any suspicious words present online. Cops were unable to track the terrorist related website or any person with suspicious information. The ratio of terrorism is high in today's world. There must be a system to track those suspicious word online and bring down the ratio of terrorism. In various arrangements and have images, videos etc. intermixed on a single web page. So we here propose to use smartly designed web mining algorithms to mine textual information on web pages and detect their relevancy to terrorism. In this way we may judge web pages and check if they may be promoting terrorism. This system proves useful in anti-terrorism sectors and even search engines to classify web pages into the category. Their relevance to the field helps classify and sort them appropriately and flag them for human review.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03478
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

IV. IMPLEMENTATION DETAILS

We implemented various machine learning algorithms using the tool WEKA (Waikato Environment for Knowledge Analysis) which is a free software licensed under the GNU General Public License, and the companion software to the book "Data Mining: Practical Machine Learning Tools and Techniques".

Sr No.	Algorithm	Accuracy(in percentage)
1.	Logistic Regression	77.47
2.	Naive Bayes	88.23
3.	Decision Tree	71.44
4.	k-Nearest Neighbors	84.96
5.	Random Forest	98.66

TABLE 2: COMPARISON OF MACHINE LEARNING ALGORITHMS

We compared all of the algorithms on the basis of their accuracy and correctness (tallying the words and score stores in the database and the words on the webpage that the user wants to check) by applying these algorithms on our dataset and chose the one which has the highest accuracy: Random Forest. Above table shows each of the implemented algorithms and their accuracy. Once you login, it will redirect you to the page where you can enter the URL of the webpages that you want to check for any trace of terrorism. On entering the URL and clicking on 'Search', it will show you the complete webpage that its checking along with the words that have the maximum occurrences and that are tagged in the database as related to terrorism.

The below images show the complete result.

Figure 2. URL page

Figure 3. Final score

Figure 4. History of visited websites

V. CONCLUSION AND FUTURE SCOPE

To curb the menace of terrorism and to destroy the online presence of dangerous terrorist organizations like ISIS and other radicalization websites. We need a proper system to detect and terminate websites which are spreading harmful content used to radicalizing youth and helpless people. We analysed the usage of Online Social Networks (OSNs) in the event of a terrorist attack.

We used different metrics like number of tweets, whether users in developing countries tended to tweet, re-tweet or reply, demographics, geo-location and we defined new metrics (reach and impression of the tweet) and presented their models. While the developing countries are faced by many limitations in using OSNs such as unreliable power and poor Internet connection, still the study finding challenges the traditional media of reporting during disasters like terrorist's attacks. We recommend centres globally to make full use of the OSNs for crisis communication in order to save more lives during such.

REFERENCES

[1] W. Khan, A. Ahmad, A. Qamar, M. Kamran, and M. Altaf, ‘‘SpoofCatch: A client-side protection tool against phishing attacks,’’ IT Prof., vol. 23, no. 2, pp. 65–74, Mar. 2021.

[2] B. Schneier, ‘‘Two-factor authentication: Too little, too late,’’ Commun. ACM, vol. 48, no. 4, p. 136, Apr. 2005.

[3] S. Garera, N. Provos, M. Chew, and A. D. Rubin, ‘‘A framework for detection and measurement of phishing attacks,’’ in Proc. ACM Workshop Recurring malcode, Nov. 2007, pp. 1–8.

[4] R. Oppliger and S. Gajek, ‘‘Effective protection against phishing and web spoofing,’’ in Proc. IFIP Int. Conf. Commun. Multimedia Secur. Cham, Switzerland: Springer, 2005,
pp. 32–41.

[5] T. Pietraszek and C. V. Berghe, ‘‘Defending against injection attacks through context- sensitive string evaluation,’’ in Proc. Int. Workshop Recent Adv. Intrusion Detection. Cham, Switzerland: Springer, 2005, pp. 124–145.

[6] M. Johns, B. Braun, M. Schrank, and J. Posegga, ‘‘Reliable protection against session fixation attacks,’’ in Proc. ACM Symp. Appl. Comput., 2011, pp. 1531–1537.

[7] M. Bugliesi, S. Calzavara, R. Focardi, andW. Khan, ‘‘Automatic and robust client-side protection for cookie-based sessions,’’ in Proc. Int. Symp. Eng. Secure Softw. Syst. Cham, Switzerland: Springer, 2014, pp. 161–178.

[8] A. Herzberg and A. Gbara, ‘‘Protecting (even naıve) web users from spoofing and phishing attacks,’’ Cryptol. ePrint Arch., Dept. Comput. Sci. Eng., Univ. Connecticut, Storrs, CT, USA, Tech. Rep. 2004/155, 2004.

[9] N. Chou, R. Ledesma, Y. Teraguchi, and J. Mitchell, ‘‘Client-side defense against web- based identity theft,’’ in Proc. NDSS, 2004, 1–16.

[10] B. Hämmerli and R. Sommer, Detection of Intrusions and Malware, and Vulnerability Assessment: 4th International Conference, DIMVA 2007 Lucerne, Switzerland, July 12- 13, 2007 Proceedings, vol. 4579. Cham, Switzerland: Springer, 2007.

[11] C. Yue and H. Wang, ‘‘BogusBiter: A transparent protection against phishing attacks,’’ ACM Trans. Internet Technol., vol. 10, no. 2, pp. 1–31, May 2010.

[12] W. Chu, B. B. Zhu, F. Xue, X. Guan, and Z. Cai, ‘‘Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing URLs,’’ in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2013, pp. 1990–1994.

[13] Y. Zhang, J. I. Hong, and L. F. Cranor, ‘‘Cantina: A content-based approach to detecting phishing web sites,’’ in Proc. 16th

Int. Conf. World Wide Web, May 2007, pp. 639–648.

[14] D. Miyamoto, H. Hazeyama, and Y. Kadobayashi, ‘‘An evaluation of machine learning-based methods for detection of phishing sites,’’ in Proc. Int. Conf. Neural Inf. Process. Cham, Switzerland: Springer, 2008, pp. 539–546.

[15] E. Medvet, E. Kirda, and C. Kruegel, ‘‘Visual-similarity-based phishing detection,’’ in Proc. 4th Int. Conf. Secur. privacy Commun. Netowrks, Sep. 2008, pp. 1–6.

[16] W. Zhang, H. Lu, B. Xu, and H. Yang, ‘‘Web phishing detection based on page spatial layout similarity,’’ Informatica, vol. 37, no. 3, pp. 1–14, 2013.

[17] J. Ni, Y. Cai, G. Tang, and Y. Xie, ‘‘Collaborative filtering recommendation algorithm based on TF-IDF and user characteristics,’’ Appl. Sci., vol. 11, no. 20, p. 9554, Oct. 2021.