iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** XII **Month of publication:** December 2020

DOI: https://doi.org/10.22214/ijraset.2020.32551

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XII Dec 2020- Available at www.ijraset.com

Revelation of Terrorism Content

Ishaan Singh[1], Jatin Rawat[2], Shadab Alam[3]
[1, 2, 3]*B. Tech Student, Deptt of Computer Science and Engineering, GGSIPU, New Delhi, India*

Abstract: ***In the ongoing scenario, terrorism has spread in a dramatic way in specific parts of the world. This gigantic development in fear-based oppressor has forced to halt the spread before it makes harm to individual existence or assets. Along with the advancement in innovation, the web has acquired a mechanism to lay out violence with the help of multimedia. Fear based oppressor associations utilize the web to hurt and stigmatize humanity and moreover advance psychological militant affairs by websites that pressurize masses to join militant associations and submit wrongdoings for the sake of those associations. Web mining and information digging are utilized at the same time for the motivation behind the advancement. Web mining indeed, even involves various data mining techniques can be used to examine and extract applicable information from raw texts. Data mining is useful in recognizing different strings, keywords, and critical data in raw texts. Data mining and web scraping are utilized for extracting information from text generally. Data mining techniques are practiced to handle ordered datasets and web scraping can be used for optimizing unstructured websites and information that is accessible over the web. Different websites have different structures and that makes it very hard to peruse for any method.***

Keywords: Psychological Warfare, Web Scraping, Random Forest, URL Flagging, Terrorist

I. INTRODUCTION

Using the internet, different Terrorist organisations are promoting psychological militant exercises among the youngsters and pushing them into the dark world. We need to make a system to stop the spread of such illegal activities which recognizes explicit catchphrases in a specific site.

If the keywords are found in the specific site then it will be flagged for human review. Data mining includes web scraping that assists us with filtering of data, extricating information from raw data. Data mining assists us to recognize keyword, structure and significant data. So, we intend to execute an effective data mining framework to acknowledge these web characteristics and highlight them.

Data extraction is a procedure which separates important information from enormous informational collections and retrieves most extreme experiences to attain outcomes. For system improvement data mining and web scraping are used. Previously work done on this topic is covered under literature review.

In this paper all the frameworks that have been implemented before are discussed. So, to discard the bugs of the current framework and to improvise the efficiency of current frame work we suggest our model. The used procedure and achieved outcomes are discussed briefly. The framework is useful as an enemy of Terrorist and digital security reaction divisions. The framework should help the cops to follow correspondence held among psychological oppressors and ought to recognize site pages created in various stages.

II. LITERATURE REVIEW

1) Reference [1] Paper suggested a framework to investigate terror activities and notify accordingly by making use of the IP address to track and detect eavesdropping. They used clustering techniques to perform operations on the content by grouping them on the basis of the content they have in common.

2) Reference [2] They grouped the website pages into different classifications and arranged them properly. There are two technologies that are utilized in this project which are Data mining and web mining.

3) Reference [3] They used web mining to extract data from URL and determine the correlation of the content with terrorism for which they used various machine learning algorithms (Logistical Recognition, Decision Tree, Random Forest). Their work only finds the words that can be pegged as related to terrorism.

4) Reference [4] They have performed various classification algorithms to scratch the data from the log files and hence proposed the ideology of cleaning the data with the help of classifier techniques or a mixture of classifier techniques.

5) Reference [5] They Examined different strategies by which textual information can be brought to counter psychological warfare on social sites by applying web mining procedures. The work thus implemented uses various techniques like facial recognition and also uses text mining on OSN.

©IJRASET: All Rights are Reserved

S.No.	Paper	Algorithms	Application
1.	J. Kiruba, P. Sumitha, K. Monisha, S. Vaishnavi, "Enhanced Content Detection Method to Detect Online Spread of Terrorism, International Journal of Engineering and Advanced Technology Volume-8, Issue-6S3, September 2019."	Clustering	Uses IP address to detect eavesdrop
2.	Naseema Begum, Aswathy R. H., Hanu Rakavi S., Mohanambal R. "Detection of online spread of terrorism using web data mining"	Random Forest Decision Tree	Performs a well-defined cleaning of data and also data storage.
3.	Aakash Negandhi, Soham Gawas, Prem Bhatt, Priya Porwal" Detect Online Spread of Terrorism Using Data Mining.IOSR Journal of Engineering Volume 13,17 April 2019."	Logistical Regression Decision Tree Random Forest	Finds only when the words that can be pegged as related to terrorism
4.	A. Sai Hanuman, G. Charles Babu, P. Vara Prasad Rao, P.S.V. Srinivasa Rao, B. Sankara Babu "A Schematic Approach on Web Data Mining in Online Spread Detection of Terrorism, International Journal of Recent Technology and Engineering Volume-8, Issue-1, May 2019."	Naïve Bayes Decision Tree K-Star Multi-Class Classifier	Performs Data cleaning to reduce the size of data
5.	Saba Bashir, Fawad Ali, Farhan Hassan Khan, Uzair Ahmed "Counter Terrorism on Online Social Networks Using Web Mining Techniques"	Naïve Bayes KNN Decision Tree Logistical Regression Random Forest	Uses various techniques like facial recognition Uses text mining on OSN

Table I. Existing Techniques

III. PROPOSED METHODOLOGY

The agenda of our proposed system is to flag illegal(terror-promoting) websites from the internet. Our proposed methodology demands the user to login to the system. After login, the user has two choices:

A. Provide the URL which is to be scanned for terror related content.
B. Update the keyword in the database.

Subsequently, the system will scrap the data from the given URL. After scraping of data, the system will tally the keywords and search for terror content. Depending on the keywords present in the database, we will rank the website.
If the rank of the website is more than the rank criteria, then the system will flag the website and after that the site will be sent for human review.

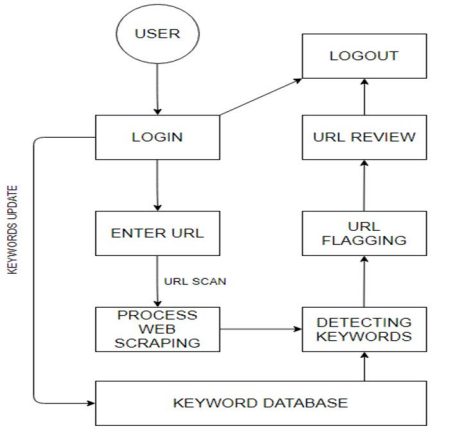


Fig. 1 Proposed Methodology

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XII Dec 2020- Available at www.ijraset.com

IV. ALGORITHMS

A. K-Nearest Neighbors

K-Nearest Neighbors also known as KNN, is used in supervised Machine learning problems. In KNN we give the value of K which will be the closest k instances to the given data. In our framework we used KNN for checking the data nearest to the keywords present in the keyword database.

B. Decision Tree

This supervised machine learning algorithm is used in classification as well as regression problems where we have a predefined target variable which can either be applied in categorical or continuous variables. Here we have a predefined keyword to be matched with the text within the web content that is provided by the user at the time of login.

C. Logistic Regression

This algorithm is used to determine the probability of a certain event existing such as pass/fail, alive/dead where the probability lies between 0 and 1. The logit model has been extended in this project to match the keywords provided by the user from the test data set created. The train dataset created consists of the words that are somewhat related to terror as well as those that do not show any resemblance to any such activity. The probability of the words that do not resemble any terror content is 0, whereas it is 1 for those related to terror content.

D. Random Forest

Random forest is an ensemble learning technique. It is similar to decision tree but not biased as decision tree . Random forest reduces the error rate in the system. We use random forest for classification tasks i.e. matching of keywords of the words which are scraped from the URL.

V. IMPLEMENTATION DETAILS

We use Jupyter notebook software to execute different ML Algorithms such as Decision Tree, Logistic Regression, Random Forest and K-nearest neighbors. It is an open -source software which is free to use for the community. In this we can create, edit and share the ML code . It is also used for visualisation purposes.

S.No.	Algorithm	Accuracy (in percentage)
1.	Decision Tree	78.12
2.	Logistic Regression	77.4
3.	Random Forest	95.5
4.	k-Nearest Neighbors	84.3

Table II. ML Algorithms accuracy

We have performed different algorithms on the data scraped from the URL provided by the user which is to be taken under consideration for review purpose by comparing the keywords from the content of the Website. We have hence concluded to opt for the Random Forest algorithm which has the highest accuracy.

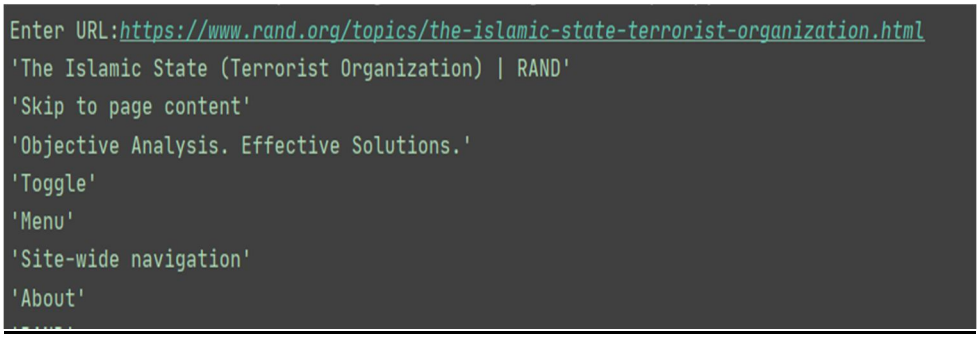


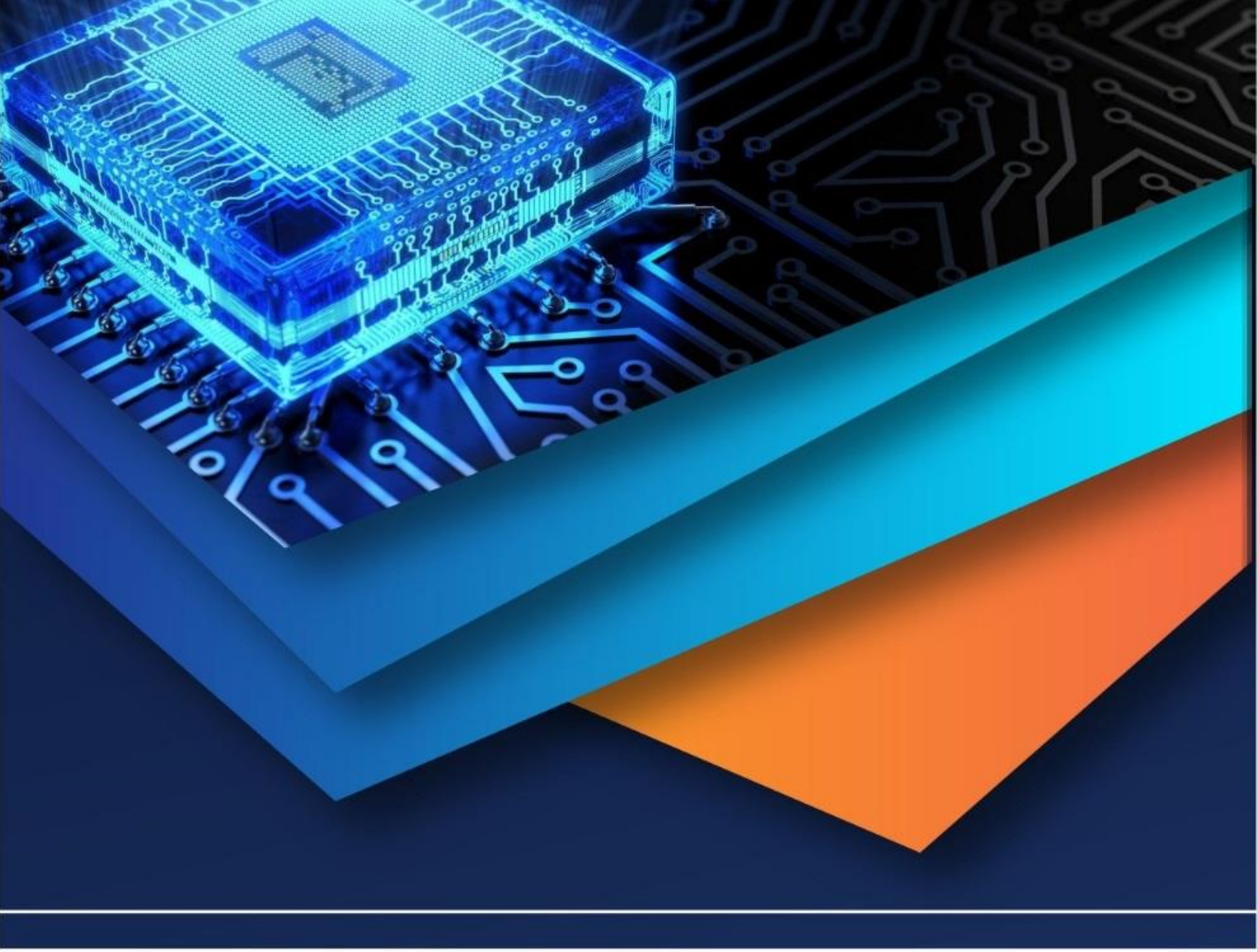
Fig. 2 URL Entry

©IJRASET: All Rights are Reserved

REFERENCES

[1] P. Sumitha, J. Kiruba, K. Monisha, S. Vaishnavi "Enhanced Content Detection Method to Detect Online Spread of Terrorism, International Journal of Engineering and Advanced Technology Volume-8, Issue-6S3, September 2019".
[2] Aswathy R. H, Naseema Begum "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- Volume 5, Issue"
[3] Soham Gawas, Aakash Negandhi, Prem Bhatt, Priya Porwal "Detect Online Spread of Terrorism Using Data Mining. IOSR Journal of Engineering Volume 13, 17 April 2019".
[4] G. Charles Babu, A. Sai Hanuman, P. Vara Prasad Rao, B. Sankara Babu, P.S.V. Srinivasa Rao "A Schematic Approach on Web Data Mining in Online Spread Detection of Terrorism, International Journal of Recent Technology and Engineering Volume-8, Issue-1, May 2019".
[5] Saba Bashir, Fawad Ali, Uzair Ahmad and Farhan Hassan Khan "Counter Terrorism on Online Social Networks Using Web Mining Techniques, Department of Computer Science, Federal Urdu University of Arts, Science and Technology (FUUAST), Islamabad, Pakistan".
[6] Chen H. "sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. ACM Transactions on Information Systems, forthcoming, June 2008".
[7] Anthony Estrada, Michael Grenieri "Down Converter Characterization in a Synthetic Instrument Context 2006 IEEE Autotestcon".
[8] Avishag Gordon "The spread of terrorism publications: A database analysis, Terrorism and Political Violence journal, published in Dec 2007".
[9] S. Padmapriya, T. Anand, E. Kirubakaran "Terror Tracking Using Advanced Web Mining 2009 International Conference on Intelligent Agent & Multi-Agent Systems".
[10] "United Nations Counter-Terrorism Implemantation Task Force The use of the Internet for terrorist purposes, 2009".
[11] Micheline Kamber, Jiawei Han, Jian Pei "Data Mining Concepts and Techniques, Morgan Kaufmann 3nd Edition".
[12] M. Steinbach, P. N. Tan, Vipin Kumar, "Introduction to Data Mining, Pearson Education".

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)