IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MINING SOCIAL MEDIA FOR TERRORISM ANALYSIS

Pratik Ratanpara
Computer Engineering
Universal College of Engineering
Mumbai, India

Smit Parekh
Computer Engineering
Universal College of Engineering
Mumbai, India

Dravya Gohil
Computer Engineering
Universal College of Engineering
Mumbai, India

Ancy Almeida
Computer Engineering
Universal College of Engineering
Mumbai, India

Abstract: **There has been a tremendous increase in internet usage in the following decade which leads to many great ideas and technologies emerging. But as they say "with good comes bad". So, a lot a malicious activity is also noticed which has been done by unknown groups. Groups can be labelled as fraudster group, terror groups and many other unethical groups. Groups use this social media platform for extracting information, and this piece of information is used by them to expand terrorist activities. This has resulted as a growing activity past decades. The basic idea behind this is to put an end to such kind of activities or curb this activity through technology. The technology that we are using is sentiment analysis. So, in attempt to this, we propose this application system which is efficient enough to detect such web activity and make a count of them to be worked on.**

Keywords - API (Application Programming Interface), IE (Information Extraction), OSN (Online Social Networking)

1.INTRODUCTION

Terror groups use social media as a medium to extract information and recruit's violence in the society. Social media operate through world platforms with open access, which leave digital signatures for law enforcement agencies to tracking them down. Successful operations to curb malicious activities and track those responsible have been performed with the help of open-source information. By acting smoothly on this available information, it also becomes possible to stop incidents before they occur.

Terrorist organisations operate in such a way that that social networks should be able to provide the first point of contact with potential candidates for recruitment. Once initial contact is established further communication is shifted to an encrypted channel where tracking becomes nearly impossible, which is further complicated by privacy and protection laws prevalent in various countries around the world.

Social media can be said as a trend these days which has integrated technology, social interactions & construction of words, pictures, videos and audios. It helps to get an access to friends, tweets and user credentials. The process of Extraction and Analysis is the challenging attribute in Social Media as the data is dynamic and data is unstructured in different sites. This idea creates interest in the students to help them in understanding the steps of extracting content from different sites and come up with new advancement in developing applications by using the active data.

Now a day's new wave of social media technologies such as Facebook, blogs, wikis, microblogging, twitters play a vital role in formal and informal communications. Twitter plays a virtual online world for its users. Virtual world interacts like a real world where location act as an intermediate connection. Commonly used GPS feature in Smartphone and tablet-enabled social media users to attach real-time locations when sending out tweets. Social media can be defined as a web-based application, collection of feeds, news, blogs, and unstructured text which is accessible through the web.

1.1 PROBLEM STATEMENT

In the proposed work considering social media network, we will be searching for keywords and no of post to be searched will be provided to the system. Social Networking Site such as Facebook, Twitter and linkedin, these three are the popular and well-known sites or services. Using Instagram and twitter data, further the algorithm will download the photograph related to the keyword and post along with captions and tags. The posts and tags will be stored in CSV files. The data can be further used for monitoring and indexing that specific keyword. That can be analysed in positive and negative sentiment.

2. LITERATURE REVIEW

Authors in [1] have examine a social media network consisting of terrorism related Twitter accounts. The data was collected through a social media discovery tool executing queries on the Twitter API (https://dev.twitter.com/) using a set of Arabic keywords related to terrorists' propaganda and next they examined which centrality measure is able to detect the most influential. Twitter user and limit the communication effectively within the terrorist network.

[2] Illustrates how big data is used to prevent and combat terrorism in developing countries. First, the author presents a bird-eye view of the analytical procedure for handling Big Data for terrorism detection. In this paper, a use case of Social Media Analysis for Combating Terrorism (SMACT) Model was also presented in order to demonstrate the plausibility of its implementation to solve real security problems that relates to terrorism.

Authors in [3] presented the methods used by terrorists to spread their messages using social media. It is understood that containing terrorism-related material on social media is not advisable. Author analysed the associated problems and tried to propose strategies towards a solution for containment of terrorism-related activities.

[4] illustrates the system that detect patterns and relevant information in texts in web page using data mining. The website will have this characteristics Load balancing, easy accessibility, user friendly, efficient and reliable and easy maintenance. The system used is a basic data mining as well as web mining concept. The system have used Visual Studio and SQL Server 2008 to create and record the project

Authors in [5] illustrated that the system deliver event notification which is used to monitor the activities and delivers notification according to the investigation knowledge. This study results the concepts and methods to detect terrorists and their behavior and their supporters who supports for terror related activities in society using an Internet access.

[6] Illustrates that the framework will be checking website pages, whether a page is advancing fear-based oppression. This framework will be characterizing the website pages into different classes and sort them accordingly. There are two highlights utilized in this framework that is information-mining and web-mining.

3. SYSTEM ARHITECTURE

The system comprises of following steps: -

- Step 1- **Insert Keywords:** After successful login, admin can add keyword which specifies terrorism.
- Step 2- **Check Website** Here, admin can check Instagram and Twitter to scan the Social Media site for any suspicious word.
- Step 3- **Downloading Data**: The system downloads images, tags and caption from the post
- Step 4- **Data Directory**: Tags and captions are stored in a CSV file
- Step 5- **Data classification**: Classification will be done in two types → 1) Positive text

  2) Negative text
- Step 6- **Sending mail:** If the negative data is crossing a set threshold, then it will forward the analysed data to the authorities which handles all this necessary information
- Step 7- **Digital Marketing:** Positively analysed data can be further used for social media marketing.

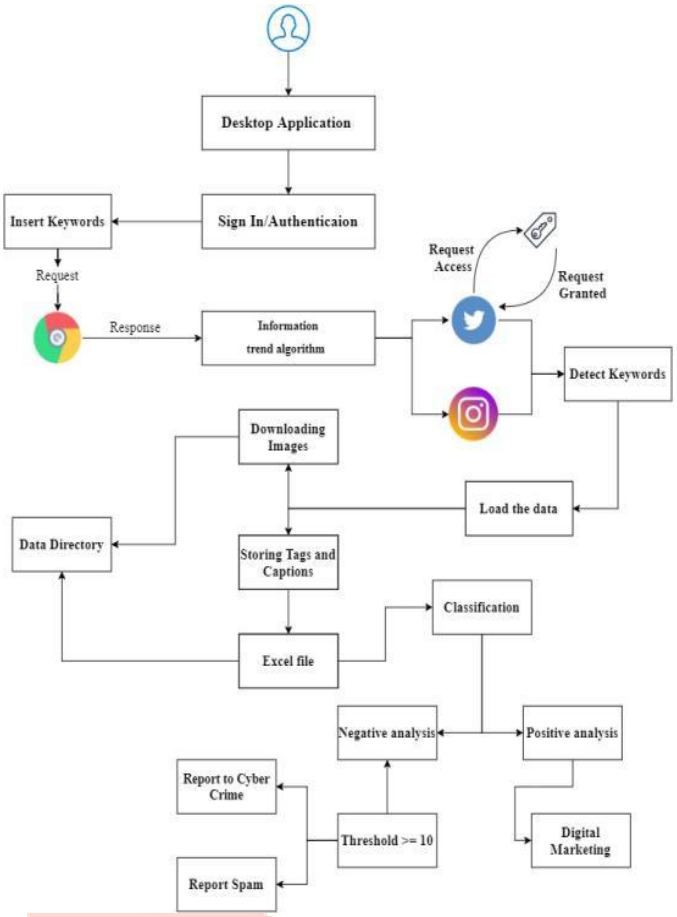

Figure 1 Architecture diagram

4. IMPLEMENTATION

In the proposed work, aim is to extract as much relevant data to the keyword as possible, the keyword is the input provided by user which is used for extraction.
The keyword is then processed through two scrappers built

- Instagram scrapper
- Twitter scrapper

As the names say twitter scrapper will be looking for data in the twitter domain and Instagram will be looking for data in the Instagram feed. The number of searches will also
be asked from user depending on the number, the data will be extracted in the Instagram and twitter that is built from scrap, which is implemented on the algorithm of information trend algorithm and for twitter one needs a consumer key, consumer token as twitter doesn't allow its data to be scrapped unless one has a twitter developer account. Each keyword has its own individual directory named as: data_'keyword'. The data for twitter and Instagram are extracted in a excel file with three sheets (caption, tags, post) and images. After extraction, the data will be going for sentiment analysis. Analysis will be done for the files whether they are appropriate or not, then decision will be made on the basis of results. If the data is terror related that violates the twitter or Instagram policies then action will be taken against by reporting it to authorities taking care of such threats. For e.g., a government body will be measured on whether the number of posts is passing the threshold, if it passed the threshold which is kept as 10 then system will be informing the official bodies. The data is related to terror is decided whether the tweets or posts go through the polarity check, if the value derived is less than 1 then negative and vice versa.

The sentimental module consists of two implementation analysis

1. Pattern analyser
2. Naive bayes analyser

The default technique is pattern analyser but one can overwrite the analyser by passing another implementation.
The visualization plays an important role in the last part as many naive users don't understand technical terms. So, in the end the analysis will be visualised through the pie chart. This pie chart shows 3 portions (positive, negative or neutral) and if the classification falls under the other domain which is positive then the post can be saved, then checked whether the results are repeated or not. If yes, it can used to advertise digitally.

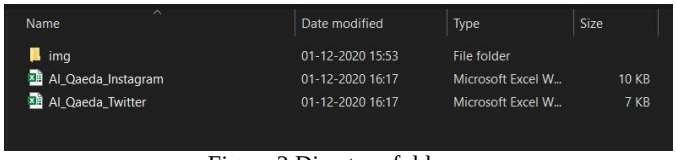

Figure 2 Directory folder

5. RESULTS

While implementing keyword passed is "Al_Queda". According to this keyword, the results are as follows:

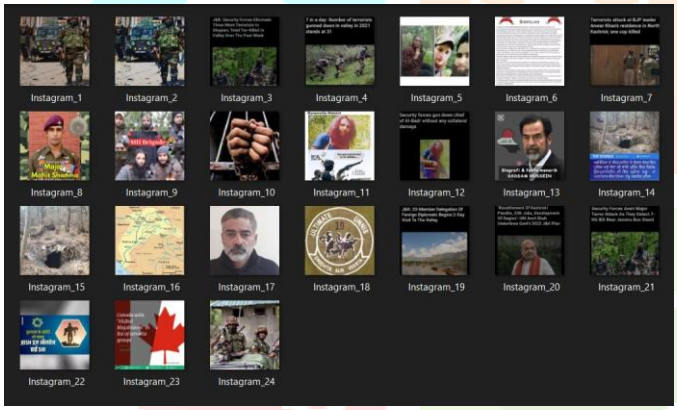

Figure 3 Image Folder

- As per figure 3, the downloaded pictures are saved in the system's directory

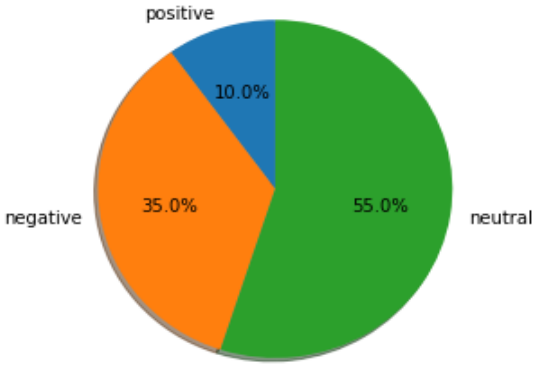

Chart 1

- As per chart 1, the analysis is done on Instagram

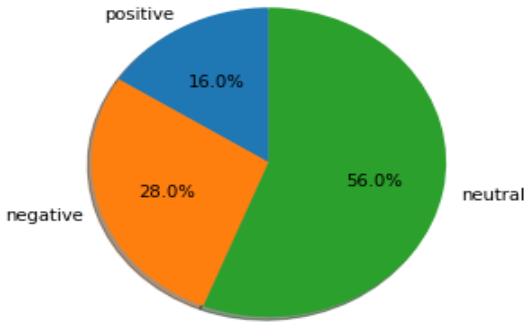

Chart 2

- As per chart 2, the analysis is done on twitter

6. CONCLUSION and FUTURE SCOPE

To curb the possible danger of terrorism and to destroy the online presence of dangerous terrorist organizations like ISIS and other radicalization websites. A need of proper system to detect and terminate website which are spreading harmful content used to influence youth and helpless people Further study will be added with improved algorithm and process of application development. The aim of this research study to predict social media crimes by suing twitter data and Instagram data. The proposed model is currently offline in future work it can be extended for real-time. More crime classes can be added to make the system efficient and robust

REFERENCES

[1]George Kalpakis , Theodora Tsikrika , Stefanos Vrochidis and Ioannis Kompatsiaris (2016), "Key Player Identification in Terrorism-Related Social Media Networks Using Centrality Measures".

[2]Taiwo Kolajo, Olawande Daramola(2017)," Leveraging Big Data to Combat Terrorism in Developing Countries"

[3]Ali Alzahrani, Khalid Bashir Bajwa, Turki Alghamdi and Hanaa Aldahawi(2018)," Countering Terrorism on Social Media using Big Data"

[4]Aakash Negandhi, Soham Gawas, Prem Bhatt, Priya Porwal(2018),"Detect Online Spread of Terrorism Using Data Mining"

[5]J. Kiruba, P. Sumitha, K. Monisha, S. Vaishnavi(2019)," Enhanced Content Detection Method to Detect Online Spread of Terrorism"

[6] Mrs. T. Suvarna Kumari, Mr. Narsaiah Putta (2019)," Web Data Mining to Detect Online Spread of Terrorism"