

# Detecting Online Spread of Terrorism on Twitter Using Machine Learning

Kokane Saurav Bhairu, Kshtij Kanojiya, Mane Shidhaji  
Vijaykumar , Chormale Pooja Balaso

*(student, Shree Ramchandra College of Engineering, Pune)*

*(student, Shree Ramchandra College of Engineering, Pune)*

*(student, Shree Ramchandra College of Engineering, Pune)*

*(student, Shree Ramchandra College of Engineering, Pune)*

Submitted: 10-11-2021

Revised: 24-11-2021

Accepted: 27-11-2021

**ABSTRACT:**(11 Bold)Detecting terrorist related content on social media is a problem for law enforcement system due to large amount of data. This work is aiming at detecting tweets that are supporting terrorism or sent by terrorist orgs. To do this we use machine learning approach where we make use of two set of features: data dependent features and data independent features. The data dependent features are features that are heavily influenced by the specific dataset while the data independent features are independent of the dataset and can be

used on other datasets with similar result. By using this approach we hope that our method can be used as a baseline to classify violent extremist content from different kind of sources since data dependent features from various domains can be added.

In this system, we have developed a front-end system for real-time viewing of the tweets from twitter that are detected using this system. We also have compared the performance of the two different machine learning classifiers, Support Vector Machine (SVM) and Multinomial Logistic Regression and found that the first one works better than the second one.

**KEYWORDS:**Social media, Terrorism, twitter, Extremism, machine learning, real-time tweets.

## I. INTRODUCTION:

Nowadays, we have a lot of means of communication. The era called as era of internet. Internet is highly dynamic means of communication. We can use internet for many reasons like to get any type of information, entertainment, business, microblogging, social media and many all. People also like to share their views on internet. For that they use social media. Nowadays microblogging is the popular ways to share views.

Twitter is most popular microblogging platform. Twitter is most useful way of interaction. Twitter has 380 million regular users and there are 500 million tweets everyday.

But everything has their pros and cons. As we use these platforms for share our views and to tell people what we are doing in our daily lives, some people use them to share their bad intent. Terrorists are one of these people. They use twitter for promote their bad views, to spread fear among people. They also use twitter for recruitment as well as fund raising.

There is our social responsibility to stop this. So we are building this web app using real time tweets for detecting extremist, radical, elementary and terrorist supporting tweets. And to report and share this information to authorized body to take action.

Rupali Patil al. [1] The procedure of sentiment analysis and its visualization is explained in detail concerning the topic Article 370. Sentiment analysis and opinion mining require detailed knowledge of how twitter and its python client Tweepy works to obtain the results. Python libraries like matplotlib and pandas are also used for simpler analysis and visualization of the tweets acquired. As a whole, from this paper it can be concluded that Pakistan is comparatively more concerned about the impact on its trade and has been somewhat more cynical when it comes to the sentiments of its Twitter users whereas India, on the other hand, is more concerned about increase in terrorism with a slightly positive attitude towards the revocation from the country's Twitter users.

M. Ashcroft et al. [3] made an attempt to detect jihadist messages from Twitter. They used sentiment analysis to detect if a message supports ISIS or not. They used some keyword to extract tweets from twitter feed. The advantage of this

work is it uses three different features such as Time-based features, Sentiment based features and Stylometric features to detect jihadist text. They got almost 90% accuracy using these features. There is a limitation and that is, they didn't used real time validation of tweets.

Md. Abrar and team of Chittagong university et al. [4] developed a framework for detect real time tweets to detect terrorist activities. They use logistic regression and SVM to analyze tweets. They use tweepy model to create their dataset from real time tweets. They do classification of tweets as they are terrorist supporting or not. They The limitation of this project was, it could not classify terrorist related tweets and terror non supporting tweets.

Mariam Nouth et al. [4] developed a project named Understanding radical mind on twitter. They mostly used magazines to find data about terrorism. They used term frequency inverse frequency bit trigger and word embedding language model for differentiating tweets.

Use of feature engineering was significance in model building.

They use NLP for tokenizing tweets and understanding sentiment of tweet. They used pro-ISIS Paris2015 attack database from Kaggle databases. There was no real time tweet validation in this project.

Akshay Karale et al. [5] made project on Framework for analysing real time activities on Twitter main features ware countering terrorism and protect human rights as manually reinforcing

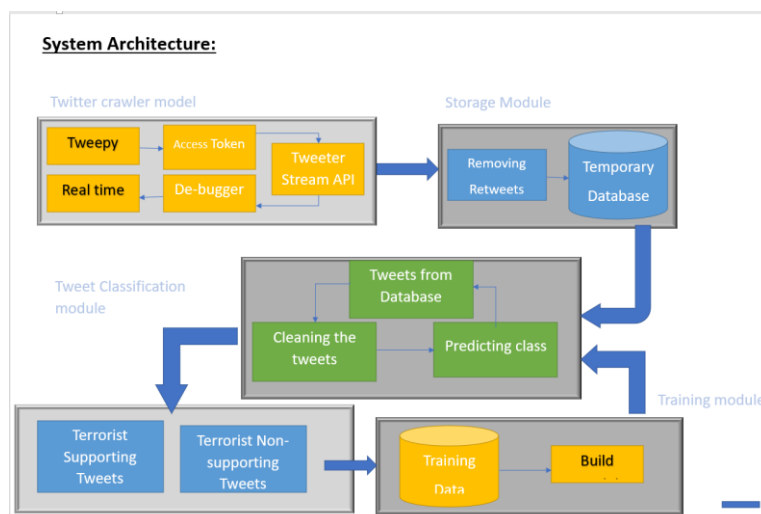
Countering violent extremism and radicalisation that lead to terrorism

They used database produced by bye Twitter API.

Used algorithm SVM and logistic regression.

It was a desktop application. The data set as input it and output text and account name of those tweets which are terrorism supporting.

## II. SYSTEM ARCHITECTURE:



The system architecture of Twitter Terrorism Detection Framework comprises five basic modules:

- 1) Twitter data crawler module,
- 2) Storage module,
- 3) tweet classification module,
- 4) Output module and
- 5) Training module

## III. EXPERIMENTATION:

### A. Crawling Data from Twitter

We have crawled real-time tweets from the twitter by using twitter streaming API. To do this we have provided four keys that we have collected from the Twitter developer website. The 4 keys are:

1. Access Token
2. Access Secret Key
3. Consumer Token
4. ConsumerSecret Key.

These 4 keys are needed to get access to Twitter API. Using these 4 keys we can set up a

twitter real listener. This listener will allow us to collect real-time tweets from Twitter. As we have used python to build our framework, we have used a python library that helped us getting access to the Twitter streaming API. We have used a python library 'Tweepy' API. Algorithm 1 illustrates the Twitter. While collecting real-time tweets we have made sure to handle all kinds of error that would break the connection the connection breaks, the crawler module will stop working. That's why we have checked for several exceptions such database exception, limit exceeded exception etc.

Algorithm 1: Crawl Real-Time Tweets

Input: Developer Access keys

Require: Real-Time tweets streaming from twitter

1. Begin
2. Call Twitter API
3. Call Twitter Streaming API
4. SetaccessTkn = ""
5. Set accessTknSec = ""
6. SetconsumerKey = ""
7. SetconsumerSec = ""
8. tweets = STREAM-LISTENER(accessKeys)
9. Create a table named all\_tweets having the field username, tweet, tweet\_id, type
10. if tweets != null then
11. if tweets['retweet'] = False then
12. Insert tweets['text'] , tweets['username'] and tweets['tweet\_id'] into the database
13. End

1. URL is removed
2. Any user mention is removed
3. Hash (#) from the hashtag is removed
4. Contracted words are converted to their long form
5. Tokenized the tweet

B. Pre-Processing Crawled Data:

After data is crawled from the twitter the tweets are in raw form using algorithm 2. We can't use these tweets to classify or train. So, we have cleaned the tweet before using n classification or training. To handle the special component of a tweet, we have done the following pre-processing tasks.

Algorithm 2: Cleaning raw tweets.

Input: raw tweets

Require: clean the raw tweets

1. Begin
2. remove URL from raw tweets
3. remove hash (#) symbol of hashtags from raw tweets
4. remove user mentions form raw tweets
5. remove retweet symbol RT from raw tweets

6. convert the raw tweets into lowercase form
7. search for contracted form in tweets
8. if contracted form found then
9. replace it with long form
10. search for stop words in tweets
11. if stop words found then
12. remove the stop words
13. tokenize the tweets
14. apply stemming on the tweets
15. End

C. Building Model and Generating Output

To predict the class of the tweet we needed a mathematical model which can classify the tweets based on their features. We have used two classification algorithm. These are SVM (Support Vector Machine) and Logistic Regression. Using our

training dataset we built a model that can classify the tweets accurately. By using the model that we have built in the previous steps, we can classify a tweet. The classification result is 0 or 1 or 2. According to this result, we can show the type of tweets. Algorithm 3 is used to classify real-time tweets.

Algorithm 3: Classification of real-time tweets

Inputs: model file

Require: Classification of the tweets

1. Begin
2. classifier = load(model)
3. for each tweet in the all tweets table in the database do
4. clean the tweets
5. type = classifier. Predict (clean tweets)
6. if type = 0 then
7. result = "Terrorism Supporting"
8. elseif type = 1 then
9. result = "Terrorism Non-Supporting"
10. elseif type = 2 then
11. result = "Random"
12. show the result
13. End

#### IV. EXPECTED OUTCOMES:

This project will differentiate between tweets As they are terrorism supporting or terrorism related or there is an unsupported or others.

After the completion of this project we expect that spread of terrorism on Twitter for different social media sites will be reduced for this project will help it to reduce. This project will expect that in this system or model will work efficiently tu to differentiate b tweets and report to Twitter as they contain any extremist thoughts.

## V. CONCLUSION:

In this presentation tweeter terrorism detection framework to detect tweets that support terrorism from real tweet stream. Our framework will collect real time tweets by using tweeter streaming API and analyses them. It can be classified into three classes and based on category of tweet they shown in different screens of web applications.

## REFERENCES:

- [1]. A Framework for Analyzing Real-Time Tweets to Detect Terrorist Activities Akshay Karale<sup>1</sup>, Pranav Shinde<sup>1</sup>, Pushpak Patil<sup>1</sup>, Sanjay Parmar<sup>1</sup>, Prof. Niyamat Ujloomwale<sup>2</sup> <sup>1,2</sup>Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra, India
- [2]. Understanding the Radical Mind: Identifying Signals to Detect Extremist Content on Twitter Mariam Nouh<sup>\_</sup>, Jason R.C. Nursey, and Michael Goldsmith<sup>\_</sup> <sup>\_</sup>Department of Computer Science, University of Oxford, UK
- [3]. Detection and classification of social media-based extremist affiliations using sentiment analysis techniques
- [4]. Shakeel Ahmad<sup>1\*</sup>, Muhammad Zubair Asghar<sup>2</sup>, Fahad M. Alotaibi<sup>3</sup> and Irfanullah Awan<sup>4</sup>
  - [www.ieeexplore.ieee.org](http://www.ieeexplore.ieee.org)
  - <https://hcis-journal.springeropen.com/articles/10.1186/s13673-019-0185-6#Tab4>
  - Intelligent projects using python by packt.com
  - <https://www.tweepy.org>
  - <https://www.devloper.twitter.org>
  - And many more newspapers, research papers, websites, magazines etc.