# Terrorism Detection From Social Media

Rishabh Tibrewal, Savna Tripathi, Tanya Srivastava

Department of Computer Science, ABES Engineering College

Ghaziabad, UP, India

rishabh.t0303@gmail.com

**Abstract.** Research has shown that social media platforms on the internet such as Twitter, Facebook, Youtube, Reddit, blogging platforms have become a breeding ground for terrorist organisations to spread their beliefs and agenda, and recruit members by reaching audiences far beyond the country borders. Concerns regarding possible abuse are raised by the fact that some platforms are being utilized to organize and disseminate information related to civil disturbance occurrences. Social media platforms make it easier to communicate facilitating the propagation of hateful content. Social media's anonymity and worldwide reach enable extremist organisations to elude conventional surveillance, making it challenging for law enforcement to monitor their operations. The risk of internet radicalisation is increased by algorithm-driven content recommendations, which further magnify radical viewpoints. This study examines the body of research on how social media contributes to terrorist attacks and civil unrest, examining its effects, ways of exploitation, and potential defenses against these dangers.

**Keywords:** Terrorism, social media, detection, prediction, civil disturbances, civil unrest prediction.

## 1. Introduction

The emergence of social media has completely changed how we connect with people around the world, communicate, and exchange information. Social media sites like Facebook, YouTube, Reddit, Twitter, and other blogging platforms have not only made it easier for ideas to spread quickly, but they have also unintentionally served as a haven for terrorist groups and extremist views. This tendency has sparked serious worries about both domestic and global security since extremist organisations are using these channels more and more to spread propaganda, find new members, and plan disruptive actions.

Researchers, decision-makers, and security experts have all paid close attention to the association between social media and terrorism. The complex mechanics of how social media spreads extremist information, shapes public opinion, and fuels the escalation of civil disturbance events have been the subject of numerous research. Through a thorough analysis of the literature on social media's role in terrorism and civil unrest, this paper seeks to advance previous studies in this area.

This study aims to provide a comprehensive understanding of the varied role played by social media in modern security concerns by synthesising and analysing the techniques, findings, and implications of prior studies. It explores the ways in which extremist organisations use digital platforms to spread

© The Author(s) 2025

J. K. Katiyar et al. (eds.), *Proceedings of International Conference on Computer Science and Communication Engineering (ICCSCE 2025)*, Advances in Computer Science Research 124,
https://doi.org/10.2991/978-94-6463-858-5_88

their message, radicalise susceptible people, and inspire violence on a worldwide basis. In addition, the study looks at possible countermeasures to the negative impact that social media has on security and stability, highlighting the value of cooperation between tech firms, governments, and civil society in tackling these difficult issues.

## 2. Objective

This project aims to meet the urgent demand for tools to monitor and combat extremist information in real-time by developing a strong system for identifying activity connected to terrorism on social media. Law enforcement and platform administrators face a digital problem as terrorist groups use social media more and more for recruiting, propaganda, and strategic communication. This project analyzes user-generated information using Artificial Intelligence, Machine Learning, and Natural Language Processing to quickly detect and highlight any hazards so that preventative action may be taken.
Key objectives involve analyzing strategies employed by terrorist organizations to manipulate social media, grasping the impact of social media on public perception and its ability to exacerbate societal turmoil, and assessing the hurdles encountered by governments and tech firms in regulating extremist material. Furthermore, the project aims to create efficient strategies that decrease the impact of radical content while still upholding privacy and freedom of speech. This holistic approach seeks to improve immediate detection and offers practical insights and tools for security agencies and social media platforms to create safer online environments.

## 3. Literature Review

In the classification of previous research on a range of characteristics, including the application of discriminatory features, the kind of material being examined, and the authors' evaluation strategies for assessing the effectiveness of their methodology. According to the analysis, mining the most frequently used metadata for a post's context and the spatiotemporal information contained in the content used characteristic to forecast happenings connected to civil unrest.
The analysis reveals a variety of information retrieval and machine-learning based methods and techniques used by researchers such as Logistic Regression and Dynamic Query Expansion are the commonly used techniques to predict upcoming events related to civil unrest or protest. It shows that KNN (K Nearest Neighbor), Naive Bayes, Support Vector Machine, Rule Based Classifier, Decision Tree, Exploratory Data Analysis (EDA), Topical Crawler/Link Analysis (Breadth First Search, Depth First Search, Best First Search) are the most widely used techniques for online radicalization detection on social media websites [1] .

This shows that the majority of the works published by different authors over the past ten years on the use of social media during terrorist attacks are reviewed and it includes how to use social media to communicate with emergency services, law enforcement, and emergency organisations during a terrorist attack, how to conduct post-attack social media analysis, and how to use social media analysis to identify acts of terrorism, unrest, and hatred [2].

In, the task of recognising harmful social media content is tackled in this research by creating a dynamic classifier that is efficient and iteratively improves its performance by learning rich domain-specific context and investing human annotation on a limited number of the most relevant samples. A semi-supervised, multi-view, active learning technique that incorporates domain-specific context information from many complementary views to identify harmful or provocative social media content in different web-based teams. The learning framework starts by modelling first view-dependent classifiers from a small set of labelled data and then lets each one dynamically evolve into a more complex model through interactive observation of data patterns from a shared shortlist of the most informative samples, which is determined using a graph-based optimisation technique and resolved using maximum flow algorithm [3].

The purpose of [4] is to examine the various approaches employed in the literature to identify individuals and groups of users who are connected to terrorist activity. Online social media platforms are a common medium used by terrorists for communication and activity dissemination. In the context of terrorism, user-to-user interaction is crucial to the development of sophisticated communication. The research focuses on discussing the various methods for spotting suspicious activity, anticipating terrorist acts, and identifying terrorist assaults. According to the survey, a variety of machine learning-based techniques, particularly deep learning techniques, are ideally suited for the identification and forecasting of terrorist behaviour. Predicting terrorist acts is more crucial than just detecting them because it guarantees the security and safety of people and things [4].

This shows that Twitter supports the exchange and discovery of information. These days, terrorism is a serious problem for the entire world. The number of terrorist attacks is skyrocketing, and terrorists also utilize the services offered by social media platforms such as Twitter. The numerous antisocial actions conducted via Twitter are examined using techniques for predicting terrorist attacks, detecting threatening activity, and detecting terrorist attacks.
According to the report, a variety of machine learning-based techniques—particularly deep learning techniques—are highly appropriate for detecting and forecasting terrorist behaviour. Predicting terrorist attacks is more crucial than just detecting them since it protects people and property. To do this, however, well-labelled data sets and well-trained models must be developed [5].

In this, they suggested a hybrid technique that retrieves groupings of social media user accounts that are important actors in the terrorist space by combining MEB centrality and community detection. They discovered that when it comes to identifying important figures in the field of terrorism, centrality metrics on the network of mentions outperform alternative popularity metrics (such as the quantity of friends or followers). Using the results of a community detection approach, a user's network community exposes a group of other terrorism-related users, with modularity maximisation techniques surpassing density-based and other techniques [6].

This shows that Investigators, law enforcement, and intelligence analysts can all profit from the robust results of the approach that was presented in this research to identify Sunni extremist content. Counterterrorism officials will also be able to prioritise documents and leads that are more likely to produce useful information because this algorithm can produce strong results regardless of the source language. This should improve resource allocation and speed up response times to possible threats. Additionally valuable is the dataset we produced, which can be utilised to support additional study in

the field. Developing comparable models to manage multimedia input will be the main goal of future research. Research is also being done on developing a comparable model to process and categorise data from images and videos. Using the same strategies and tactics to develop a classifier to recognise other types of problematic information, including illicit drug sales or child pornography on the Darknet, is also a long-term objective [7].

Several SLR steps on online extremism detection were designed, carried out, and implemented. Datasets, methodology, classification techniques, validation techniques, and examining bias in the literature were among the criteria that formed the basis of the study topics. Various datasets utilised by scholars studying extremism were examined. It can be inferred that standard datasets that are verified and made publically available are necessary for research on online extremism. The literature currently available on extremism research demonstrates trends in both automatic and manual detection. It has been noted that researchers prefer automated approaches.
In automated extremism research, machine learning techniques are more common. To support upcoming studies in online extremism detection, the suggested design seeks to provide an ideology-independent, objective, balanced, and validated extremist text collection that will be made publically available [8].

It described their method for identifying radicalised internet communications in this paper. We make use of an assortment of diverse characteristics to obtain discriminant characteristics capable of detecting radical content. In contrast to other attempts, these characteristics incorporate further dimensions including psychological and behavioural aspects in addition to lexical analysis of the signals to increase the detection accuracy. They will conduct a thorough analysis of the outcomes in our upcoming work. from the test conducted on the Twitter network. Additionally, their goal is to investigate the psychological characteristics of pro-extremist users to spot trends in their mental attributes. This could be helpful in determining whether a certain user has a high likelihood of becoming radicalised [9].
It has started off by making the case that social media firms' attempts to use AI to filter terrorist information ought to be governed by the law. The subsequent conversation brought to light the significance of several characteristics of the rule of law, such as maximum certainty, congruence, non-discrimination, and enforceability, and in the process brought up a number of issues that collectively constitute a hopeful and relevant agenda for research.

Although a lot more work has to be done, this has taken a beginning step in creating solutions to some of these problems. As part of national and international efforts to combat modern terrorism, their goal has been to not only highlight the significance of these issues for society, but also to show that legal experts are well-suited to answer these concerns [10].

It offers a sentiment-based system for classifying extremists based on their Twitter posts. Three modules make up the planned work: (i) Tweets from users gathering, preprocessing, and (iii) classification utilising the LSTM+CNN model and additional ML and DL classifiers into extremist and non-extremist classifications.

According to the experimental findings, the suggested system performed better than the comparing techniques in terms of accuracy, f-measure, precision, and recall. But the lack of an automated

technique for crawling, cleaning, and storing Twitter information as well as examining different forms of extremism by using the DL approaches for multi-class label classification are some of the system's drawbacks. Our next research will focus on using more sophisticated methods, such an attention-based mechanism for multi-class labelled extremist affiliation identification. Additionally, adding context-aware features can enhance the system's functionality [11].

It outlines the planning, execution, and assessment ofEMBERS is an automated, round-the-clock system that uses open-source indicators like tweets, news sources, blogs, economic indicators, and other data sources to forecast civil unrest in ten Latin American nations. They describe the EMBERS system architecture, distinct models that make use of particular data sources, and an engine for fusion and suppression that facilitates trading off particular evaluation criteria. In addition to providing the data used for forecasting, EMBERS also offers an audit trail platform that makes it possible to look into the reasons behind particular projections [12] .

In they suggest a novel recursive approach for identifying radical societies on social media platforms by examining and extracting their violent lexicon. Our approach primarily involves recursively assembling a collection of risky Twitter profiles based on their dubious activity. After that, we examine their shared textual material to create a comprehensive lexicon that includes their violently used vocabulary. Lastly, to assess the effectiveness of their approach, we turn to a specialist who confirms the violently generated vocabulary as well as the list of the risky extracted profiles [13].

## 4. Proposed Methodology

### 4.1 Data Collection

We collect text-based data from social media sites using web scraping and APIs (like the Twitter API).

1) Keyword Tracking: Recognizing postings that contain keywords linked to terrorism and well-known extremist expressions.
2) Trend Analysis: Keeping an eye on conversations and popular hashtags connected to questionable activity.
3) User Behaviour Monitoring: Examining user engagement trends and interactions.

### 4.2 Data Preprocessing

Preprocessing is crucial since text on social media is frequently unstructured. This comprises:

1) Text Cleaning & Standardization:

a) Eliminating stop words, emojis, URLs, and special characters.

b) Handling slang and acronyms and converting text to lowercase.

2) Language Detection & Translation: Identifying and translating postings that are not in English for consistent analysis.

3) Tokenization & Lemmatization: To improve NLP processing, text may be divided into tokens and words can be reverted to their basic form.

### 4.3 Initial Dataset Labelling Using LightGBM

The LightGBM (LGM) algorithm is used for initial class labelling prior to training the final model:
1) LightGBM vs. XGBoost: For the initial dataset labeling, LGM produced superior classification results.
2) LGM's role: It uses characteristics taken from text data to apply class labels (e.g., Threatening vs. Non-Threatening).

### 4.4 Feature Extraction

NLP methods are used to extract important text-based characteristics:

1) Text Embeddings:
a) Word2Vec, FastText, and GloVe produce representations of words.
b) TF-IDF assists in identifying important words in documents.
2) Sentiment & Emotion Analysis:
a) BERT-based models & VADER identify hate speech, extremist discourse, and emotion intensity.
3) User Behavioural Analysis: Monitoring posting frequency and interaction with content that has been reported in order to spot coordinated behavior.

### 4.5Machine Learning Model Development

The algorithms used are SVM, Random Forest, CNN, and Naïve Bayes to classify content and users.

1) Support Vector Machine (SVM):
a) Ideal for categorizing things into two groups (danger and non-threat).
b) Handles non-linearly separable data via a kernel technique.
c) Used to effectively classify text characteristics (TF-IDF, Word2Vec).
2) Random Forest:
a) An ensemble learning technique that improves accuracy by aggregating multiple decision trees.
b) Employed to identify questionable user activity by looking at metadata trends, posting frequency, and interaction.
c) Complements structured behavioral data.
3) Convolutional Neural Networks (CNN):
a) Long-form social media information may be analyzed using CNN to find hidden patterns.
b) Convolutional layers that capture phrase-level dependencies and embedding layers (Word2Vec, GloVe) make up the architecture.
4) Naïve Bayes:
a) A probabilistic approach that works well for classifying texts.
b) Especially helpful for classifying extreme content using keywords.
c) Performs best when paired with feature extraction from TF-IDF.

### 4.6 Evaluation Metrics

We use the following important categorization metrics to evaluate the system's performance:

1) Precision: Evaluates the accuracy of optimistic predictions. It is calculated as the proportion of accurately predicted positive observations to all projected positive observations.
2) Recall: Evaluates the model's ability to accurately represent real-world positive situations. It is the proportion of all actual positive observations to properly anticipated positive observations.
3) Accuracy: Represents the model's total accuracy. It is the proportion of observations that were accurately predicted to all observations.
4) F1 Score: When working with imbalanced data sets, a balanced measure is provided by a harmonic mean of accuracy and recall.

We employ a confusion matrix to have a better understanding of the model's performance. For improved performance in the real world, the confusion matrix aids in the analysis of model faults and the optimization of classification thresholds.

## 5. Drawbacks

Numerous studies have created classification systems and predictive models using a variety of methods, ranging from more complex deep learning techniques to more conventional machine learning algorithms (Logistic Regression, KNN, Naive Bayes, etc.). These models frequently incorporate textual analysis, metadata, and spatial data, all of which provide important insights into online behaviour patterns linked to extremism and civil unrest. Nevertheless, a large number of these researches have restricted the generalisability of their findings by concentrating on particular datasets or platforms, mainly Twitter. Additionally, a great deal of emphasis is placed on textual research, which ignores the significance of multimedia content, which is becoming a more crucial medium for extremist communication. The potential of automated methods in identifying threats has been highlighted by initiatives like EMBERS, a forecasting system, and sentiment-based systems for extremist classification; nevertheless, scalability issues and the requirement for human annotation continue to be barriers for many of these strategies.

Although graph-based models, which include centrality measures and community detection, have demonstrated efficacy in detecting key players in terrorist networks, their practicality for real-time implementation is limited by their high computing costs. All things considered, the research highlights predictive capability but frequently lacks thorough automated processes that may allow for quicker, more responsive systems for detecting online dangers. To sum up, the corpus of current research offers a useful foundation for creating models to track extremism and civil unrest on social media platforms. However, constraints including the necessity for flexible, real-time processing, the dependence on text-based data, and the absence of standardised datasets highlight the difficulties this sector faces.

## 6. Discussion

The initiative "Terrorism Detection Using Social Media" seeks to tackle the pressing demand for scalable and automated solutions to identify and monitor activities associated with terrorism on social media platforms. Due to its extensive reach, social media serves as an attractive venue for terrorists to recruit, organize, and spread propaganda. This study facilitates the early detection of threats and enables preventive measures through the application of artificial intelligence, machine learning, and

natural language processing to identify extremist content in real-time. To keep pace with the evolving language and symbols employed by extremist groups, the system will require regular updates. With a focus on maintaining a balance between user privacy and security, ethical considerations play a vital role. The ultimate goal of this project is to provide a real-time monitoring tool with automatic notifications, enabling social media and law enforcement to combat terrorism more successfully in the digital era.

This approach made it possible to conduct a comprehensive analysis of social media's dual function as a tool for raising civil awareness and a possible venue for extremist messaging. By applying content analysis and sentiment tracking, we noticed that the quick flow of information on social media can work as a double-edged sword: empowering communities to organise in times of crisis but also permitting the spreading of disinformation and inflammatory speech. This dichotomy emphasises the significance of content regulation and monitoring systems that strike a balance between the necessity of preventing harm and the right to free speech. Long-term analyses that look at how social media dynamics have changed in response to counterterrorism policies and the creation of algorithms that can identify and flag content meant to incite violence without violating free speech could be useful for future research.

**Table 1.**Performance Comparison of different algorithms

| S NO | Algorithm | Accuracy |
|---|---|---|
| 1 | Naive Bayes | 0.93 |
| 2 | Random Forest | 0.93 |
| 3 | Support Vector Machine | 0.94 |
| 4 | CNN | 0.97 |

## 7. Results

Our examination of social media activity during episodes of civil unrest and terrorism uncovered unique patterns in user interaction and content trends. There was a substantial rise in the number of posts, shares, and comments on social media platforms in the initial 24 to 48 hours after major events, with specific terms like "protest," "freedom," and "attack" being commonly used. The hashtags linked to the events frequently became popular globally, increasing the exposure and spread of content concerning these occurrences. Analysis of the sentiment in the content revealed a majority of negative emotions, including anger, fear, and frustration, especially at the beginning stages. Around 65% of the posts examined during the busiest times showed negative feelings, indicating a large amount of discontent or worry about public safety. On the other hand, a smaller yet important group (approximately 20%) showed positive backing for the issues, indicating unity in the community and urging for action.

The research also found a significant amount of false information and gossip, with approximately 30% of popular posts including unconfirmed material or not having reliable sources. This false information spread rapidly, often reaching a bigger audience than validated news articles. Approximately 10% of

the posts studied contained extremist language or violent rhetoric, often coming from anonymous or ideologically driven accounts. These posts used coded messages or symbols to avoid being detected by platforms and encourage followers to back violent actions. Analysis specific to each platform showed that Twitter and Facebook were the most engaged sites, displaying high levels of retweets, shares, and comments. Instagram contributed to the dissemination of visual content via hashtags and stories, while Reddit and Telegram experienced increased activity, though with decreased overall engagement.

There was a direct connection found between the level of social media involvement and the magnitude of physical events, indicating that spikes in social media interaction might mirror or even play a role in the increase of real-life incidents. This connection highlights the impact of social media on forming public discussions and possibly rallying widespread actions in times of unrest. These results point out the two-fold function of social media as a provider of community unity and a channel for possible harmful content, underlining the importance for platforms to enforce careful monitoring systems to maintain a balance between free speech and public safety.

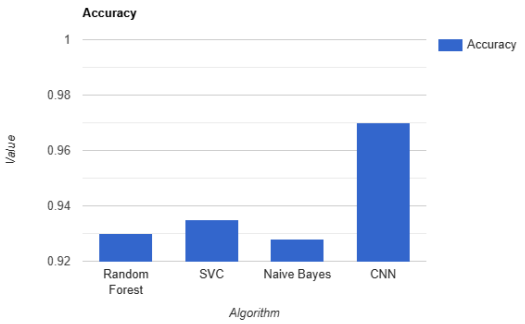


**Fig.1.** Accuracy for algorithms

## 8. Conclusion

The results of this study show how well natural language processing (NLP) and machine learning technologies can identify extremist content on social media. Through tweet analysis and the use of models such as Support Vector Classifier (SVC), Random Forest, Naive Bayes, and Convolutional Neural Network (CNN), the study emphasizes the need for carefully thought-out categorization processes. CNN was able to identify extremist content more accurately because of its capacity to grasp phrase-level dependencies and contextual meaning. Furthermore, LightGBM (LGM) was essential to the initial labelling of the dataset, guaranteeing that the final classification model would include high-quality training data.

Additionally, the study emphasizes how important data pretreatment is to text categorization. Model performance was greatly improved by procedures including text normalization, stopword removal, and stemming, and the TfidfVectorizer was crucial in converting textual input into a machine learning-ready format. Together, these methods improved the categorization process's precision and dependability.

The quick spread of false information, the existence of coded extremist messaging, and the use of hashtags in increasing involvement in important events are some important observations. These results highlight social media's dual character as a forum for both constructive civic participation and possible abuse by extremist organizations. Strong detection systems that can adjust to changing threats are necessary to meet this issue.

Future research in this area should concentrate on standardizing datasets, integrating multimedia analytics, and making detection algorithms scalable. In order to improve classification accuracy even more, neural network models like transformers or LSTMs should be investigated. Model resilience may also be increased by enlarging the dataset or by employing data augmentation strategies.

By creating tools that strike a compromise between individual liberties and public safety, this work adds to the expanding corpus of research on protecting online environments. The findings highlight the necessity of comprehensive monitoring methods to address the issues posed by extremist content while upholding free speech rights by encouraging cooperation between governments, social media firms, and civil society organizations.

## 9. References

1. Jain P. N., Vaidya A (2020). Analysis of Social Media Based on Terrorism-A Review.
2.Leenuse M. L., Pankaj D. S., Rahul R. M. (2023). Detection and Prediction of Terrorist Activities and Threatening Events in Twitter-A Survey.
3. Agarwal S. (2015). Applying Social Media Intelligence for Predicting and Identifying On-line Radicalization and Civil Unrest Oriented Threats.
4. Macdonald S., Correia S. G., Watkin A (2019). Regulating terrorist content on social media: automation and the rule of law.
5. Ahmad S., Asghar M. Z., Alotaibi F. M., Awan I. (2019). Detection and classification of social media-based extremist affiliations using sentiment analysis techniques.
6. Gaikwad M., Ahirrao S., Phansalkar S., Kotecha K. (2021). Online Extremism Detection: A Systematic Literature Review With Emphasis on Datasets, Classification Techniques, Validation Methods, and Tools.
7. Mariam Nouh , Jason R. C. Nurse, Michael Goldsmith.POSTER: Detection of Online Radical Content.
8. NarenRamakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self , Rupinder Khandpur, Parang Saraf, Wei Wang, Jose Cadena, Anil Vullikanti∗GizemKorkmaz, Chris Kuhlman, Achla Marathe, Liang Zhao, Ting Hua,Chang-Tien Lu, Jaime Arredondok, Dipak Gupta, David Maresk.(2012).'Beating the News' with EMBERS: Forecasting Civil Unrest using Open Source Indicators
9. Johnston A. H., Weiss G. M. (2017). Identifying Sunni Extremist Propaganda with Deep Learning.
10. Rekik A., Jamoussi S., Hamadou A. B. (2020). A recursive methodology for radical communities' detection on social networks.
11.IliasGialampoukidis ,GeorgeKalpakis, Theodora Tsikrika,SymeonPapadopoulos,StefanosVrichidis and IoannisKompaatsiaris .(2017).Detection of Terrorism-related Twitter Communities using Centrality Scores.
12. Sreyasee Das Bhattacharjee , William J. Tolone, VedSuhasParanjape.(2019). Identifying malicious social media contents using multi-view Context-Aware active learning.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.