Security Journal (2023) 36:221–248
https://doi.org/10.1057/s41284-022-00335-4

ORIGINAL ARTICLE

A survey of extremism online content analysis and prediction techniques in twitter based on sentiment analysis

Zouheir Trabelsi[1] · Firas Saidi[2] · Eswari Thangaraj[1] · T. Veni[3]

Accepted: 14 March 2022 / Published online: 18 April 2022

Abstract

Nowadays, extremist organizations use social networks, such as Twitter, to flourish their dark activities. Usually, to polarize new members, these organizations attempt to share their radical propaganda by posting tweets. Practically, Sentiment Analysis (SA) techniques are widely used to classify the polarity of these extremist tweets, to derive appropriate conclusions for decision-making purposes, and to make valuable predictions about future violent and terrorist events. To study the influence of social networks-based malicious activities on human security and safety, this paper surveyed different Machine and Deep Learning based techniques used for Tweets SA, and discussed their strengths and weaknesses and recent trends in the field. Furthermore, the conducted survey work highlighted promising key areas and potential challenges that require further consideration to implement more effective methods to combat extremism in online social networks.

Keywords Social networks extremism · Twitter sentiment analysis (TSA) · Opinion mining (OM) · Machine learning (ML)

✉ Firas Saidi
firas.saidi@ensi.rnu.tn

Zouheir Trabelsi
trabelsi@uaeu.ac.ae

Eswari Thangaraj
eswari@uaeu.ac.ae

T. Veni
veni@nitc.ac.in

[1] College of Information Technology, United Arab Emirates University, Post Box 17551, Abu Dhabi, United Arab Emirates

[2] RIADI Lab, National School of Computer Science (ENSI), University of Manouba, Manouba, Tunisia

[3] Department of Computer Science Engineering, National Institute of Technology, Calicut, India

Introduction

Since the last decade, it is unquestionable that social media has become an inevitable part of daily life for people around the world. In addition, due to the series of lockdowns in 2020, social media and the internet usage has grown at the fastest rate in the recent years with an increase of 4.50 billion users per day. People are using social media such as Facebook, Twitter, or Instagram to not only to share their opinions or but also express their beliefs, emotions to the whole world with the convenience of a click-away (Pai et al. 2020). In addition, whenever there is a catastrophic event occurs, there is a huge surge of text traffic on Twitter, Facebook, Instagram of informative messages, tweets, emotional outbursts, and rumours (Kostakos et al. 2018) as people tend to react faster to negative news than positive news (Esraa Najjar and Salam Al Augby 2021; Berger and Perez 2016; Conway et al. 2019). And it obvious that, due to such wide outreach of these social medial platforms, terrorist organizations such as ISIS, Hezbollah, Al-Qaeda have started using the Online Social Networks (OSNs) as a tool to spread their propaganda or hate speech, raise fund, radicalize, and recruit new members around the world (Berger and Perez 2016; Zerzri 2017).

The weaponization of social media platforms by extremists has led many governments and researchers to focus on developing new methods to counter cyber extremism. From the period August 2015 to December 2017, micro-blogging platform, Twitter has suspended 1,210,357 accounts for violations which are related to terrorism (Berger and Perez 2016; Conway et al. 2019; Aleroud et al. 2020). Though such tech giants become obliged to form regulations and make tools to counter such online extremism (Torregrosa et al. 2021), yet we witnessed live-streamed Christchurch attack in New Zealand in 2019 (Aleroud et al. 2020; Gaikwad et al. 2021). Also, it is often claimed that nowadays extremists have been deploying countermeasures to come back on Twitter and increased their usage to spread their propaganda (Conway et al. 2019). This activity ushered many governments to fund more money on counter extremism research. Having said that, as most of extremism related data published online are based on text and images. Hence it is tedious to analyse such huge chunk of data manually and draw a conclusion. Thanks to the Artificial Intelligence (AI) technologies, researchers made numerous contributions for extremism research. Sentiment Analysis (SA) and Opinion Mining (OM) are two emerging areas used to classify the sentiment of vulnerable tweets to reach appropriate conclusions and make predictions about future mass violent events. There has been a colossal number of research publications on analysing the sentiment of a particular tweet (Esraa Najjar and Salam Al Augby 2021). Most of these papers use one of two widely popular approaches—lexicon based or machine-learning based approach. Lexicon-based approach uses manually pre-classified sentiments for certain words and is further divided into dictionary-based approach and corpus-based approach. Whereas in machine-learning approach numerous algorithms for sentiment analysis—namely Maximum Entropy (ME), Naive Bayes (NB), Support Vector Machines (SVM) and Neural Network (NN) models are used for classifying tweets.

Motivations and contributions

In light of the rise of social media platforms, extremists make use of the chance to portray themselves as saviours and foster and recruit vulnerable youth to commit violent or lone-wolf attacks (Torregrosa et al. 2021; Gaikwad et al. 2021; Dadkhah et al. 2021; Rowe and Saif 2016). To understand such extremism act and behaviour, several research contributions were made using both manual and automated techniques. After enormous research from various perspectives including availability of dataset, proposed detection techniques, performance validation methods and tools, we found that a very few surveys approached the problem conceptually and whereas others focussed on identification and classification of extremists. However, such literatures have some limitations. First, the lack of standard discussion on data sources or dataset selection criteria and custom-made datasets should be studied and standardized to fulfil the research gaps. Secondly, some studies focussed only on a specific process of detection instead of providing much attention to validation techniques. This study focusses on covering the gap between existing research work and its limitations by shedding lights on various data sources, identification and classification, tools, and validation of performance metrics. This article is a systematic review of collection of voluminous literatures and analysed the details systematically based on comparative approach. In addition, it will present a state of the art of the counter terrorism research opportunities based on following research objectives:

RO1: Outline the availability of various data sources or datasets and tools pertaining for combatting online extremism.
RO2: Summarize how sentiment analysis techniques used in the field of extremism research.
RO3: Present current topic and contributions from machine learning techniques to extremism research
RO4: Discuss data validation techniques steps required for sentiment analysis in extremism research.
RO5: Throw light on future directions and challenges of the domain based on this study.

This study broadly surveys the existing Twitter sentiment analysis methodologies and techniques for combating against cyber terrorist activities. Various performance metrics are discussed to compare the performance of the existing approaches for determining which might be the best fit approach to use in future to predict terrorist attacks. In "Research Methodology" section, presents research methodology and in "Sentiment Analysis in Twitter" section, outlines Sentiment Analysis, level, challenges, and feature selection types. In "Twitter Sentiment Analysis Approaches" section, discusses various approaches of Sentiment Analysis and in "Data Sets and Collection Strategies" section, Data Sets and Collection Strategies will be discussed. Then "Discussion and Future Research Directions" section presents the discussion and future research directions whereas in " Conclusion" section we conclude this study.

Research methodology

Based on a systematic approach (Misra 2021), a survey of articles that contributed to the detection of extremism using sentiment analysis techniques was conducted. The articles were extracted from various databases such as ScienceDirect, Scopus, IEEE transactions and Web of Science. While searching for the articles we used various keywords not limited to the following: ((“Extremism Detection” OR “Cyber Terrorism” OR “ISIS” OR “Jihadist” OR “Propaganda” OR “Radicalization” OR “Classification of Extremist”) AND (“Sentiment Analysis” OR “Polarity Analysis” OR “Topic Detection” OR “Opinion Analysis” OR “Emotion analysis”)). In the extremism domain, following are a few important key terms widely used. They are *Extremism* can be defined as “an ideology or supporting belief, not based on civil or ethical values of a society and uses various methods like verbal or physical violence to achieve its goals" (Harb 2019). *Radicalization* is believing in extremism or violence due to the changes in belief and *Propaganda* can be defined as biased information to justify point of view specific group of people or political cause (Berger and Perez 2016; Garg et al. 2017; Misra 2021). A general screening process was conducted to check closeness and quality of the articles. This is done by checking the title, abstract, clearly description of proposed methodology and algorithm, validation techniques, and most importantly datasets used. 283 articles have been found as initial result. After the general scrutiny of textual analysis, 68 articles were included for this survey as shown in the Fig. 1.

Sentiment analysis in Twitter

Sentiment analysis is a process that automates attitude mining of opinions, and emotions from text, audio, video or from any database sources through Natural Language Processing (NLP). Figure 2. shows the basic flow of Sentiment Analysis in Twitter (SAT). Analyzing the sentiment in Twitter is crucial in decision-making process where it is involved in classifying opinions in text into categories like "positive"

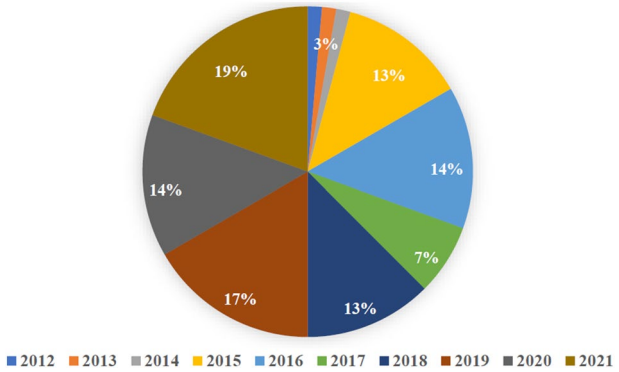

Fig. 1 Article surveyed in the review

Fig. 2 Sentiment analysis process

or "negative" or "neutral". This is an automated detection and quantification of thoughts and emotions in a tweet. It is also referred as opinion mining, and subjectivity analysis. The words such as sentiment, opinion, and belief are used synonymously but there are differences among them (Pai et al. 2020; Kharde and Sonawane 2016; Sharma et al. 2018; Giachanou and Crestani 2016).

Sentiment analysis levels

Generally, sentiment analysis is based on what is the object, object features and opinion about the object. There are three levels of analysis that can be done to analyse the polarity of the object or tweets (Ali 2015; Kolkur et al. 2015).

Document level

Document level analysis analyses a piece of text or document to determine if the text has positive or negative or neutral sentiment. The entire document of opinionated text is assumed as a single unit of information. This works better in case of a movie or a product review.

Sentence level

Sentence level analysis considers each sentence as individual unit and have a different opinion. This has two sub tasks namely, subjectivity classification and sentiment classification. In subjectivity classification, every sentence is classified into objective or subjective sentence, where subjective sentence has opinions and the later has only facts. Sentence can be classified as positive, or negative, or neutral depending upon the opinion words present in a sentence (Torregrosa et al. 2021; Kolkur et al. 2015).

Feature level

Feature level analysis works on labelling each word with their opinion and classifying the data towards where the sentiment is directed. Feature engineering concerns with identifying and extracting aspects or features from given data.

Sentiment analysis challenges in Twitter

Social media monitoring and listening through sentiment analysis on Twitter (SAT) is a special kind of social media monitoring. It is a non-trivial and challenging task of mining and preprocessing unstructured tweet like text and reviews, for feelings and evaluations of a specific event or service.

Unlike traditional sentiment analysis in websites, blogs or forums, Twitter possesses some unique challenges while analyzing the sentiments such as Text Length, Topic Relevance, Incorrect English, Negation, Stop words, Tokenization (Giachanou and Crestani 2016). Figure 3. depicts the challenging scenarios of pre-processing of the tweets. In addition to that, we discuss some of the vital challenges of Sentiment Analysis in Twitter.

Text Length: One of the major differences between conventional sentiment analysis and SAT is tweet length and it can be up to 280 characters. However, considering the topic relevance while analyzing the sentiment orientation of tweet, a lot of existing works considered the existence of a word. In addition, a few other studies considered the hashtags as a reliable indicator of the tweet's relevance about a certain topic. Having said that, due to shorter length of the tweets it is easy to classify tweets than categorizing longer documents such as review pages and blogs.

Data Sparsity and Negation: Due to contemporary casual way of communication with length restriction, sometimes tweets may contain a lot of noise such as incorrect English and misspellings. In addition, the occurrence of negation words has a vital part in finding tweet polarity. Identifying the negations is a crucial and challenging task when analysing the sentiment because it may change the sentiment polarity (Giachanou and Crestani 2016).

Multilingual and Multimodal Content: Tweets are written in 34 different of languages. The challenge is sometime tweets are written in mixed languages such as English + Arabic or English + Spanish. Identifying the correct polarity among mixed language short length tweet is yet to be explored by researchers. Moreover,

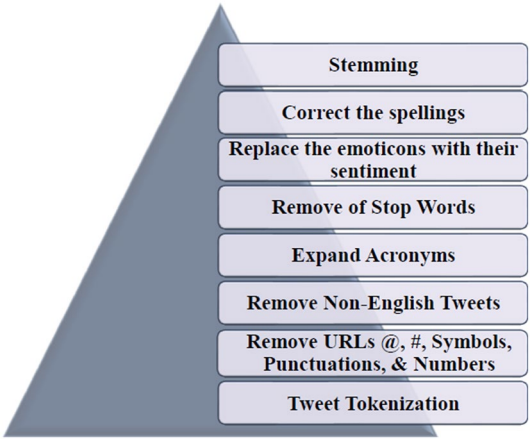

Fig. 3 Scenarios of pre-processing tweets

tweets may contain images or videos. analysing the multimodal content is also under-explored area in sentiment analysis.

Feature extraction

In most of the SAT methods, the accuracy of sentiment analysis depends on feature selection. The selected features and their combination play an important role in analyzing the sentiment of tweet. The selected key features are known as feature vectors which are required for the subsequent classification tasks. Here, we present a few key features used in the existing works (Kharde and Sonawane 2016).

Syntactic features include n-grams, dependency trees and part of speech tags, these are used to understand subjectivity patterns (Kaati et al. 2015). Most widely used N-gram features are unigrams, bigrams and n-gram models with their frequency counts. Parts of speech includes adjectives, adverbs, verb clusters and nouns are good indicators of subjectivity and sentiment. We can generate syntactic dependency patterns by parsing or dependency trees (Ngoge 2016).

Opinion Words and Phrases: Apart from words, sometimes idioms and phrases convey sentiments can be used as features (Omer 2015) Opinion word is considered as a binary-valued feature vector, which indicates that the word availability in the sentence or not where 1 denotes the occurrence of word and 0 denotes absence of word (Sharma et al. 2018) Sometimes such words frequency also considered and compared to analyse the sentiment polarity in a sentence.

Stylistic and Twitter-specific features include individual writing style using emoticons, abbreviations, and intensifiers, hashtags, URLs, followers, and retweets. Feature Extraction techniques has many benefits including improvement of accuracy, overfitting risk reduction, acceleration of training, better data visualization, and increase in annotation of the classifying model. However, before deciding on feature vectors, one should analyze on the features should be used. Because using less features improves the information retrieval of the model.

Performance metrics for sentiment analysis on Twitter

The performance of Twitter sentiment analysis is evaluated by few metrics such as accuracy, precision, recall, and F-score. Sentiment analysis is a classification problem, which involves classifying the tweet opinions in text into categories like "positive" or "negative" or "neutral" (Giachanou and Crestani 2016; Kolkur et al. 2015). Once the classifier model is developed, the next phase is to calculate the performance of the developed model. Confusion Matrix is a tool which contains information about actual and predicted classifications to determine the performance of the classifier.

Figure 4, presents the performance metrics used in the existing works of Sentiment Analysis in Twitter (Kostakos et al. 2018).

Sensitivity or Recall is a degree of positive samples labelled as positive by classifier. It is calculated as in:

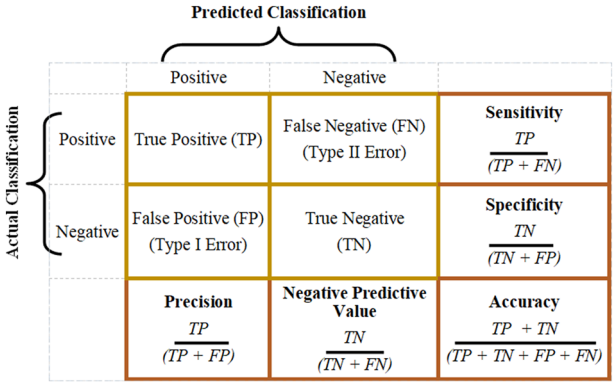

Fig. 4 Confusion matrix

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Precision is ratio of total number of correctly categorised positive samples and the total number of projected positive samples. It shows correctness attained in positive prediction. That is:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Accuracy is widely used metric which is proportion of the total number of predictions that are correct. That is:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

F-Score is a weighted average of the recall and precision. This is also known as F1-score, or F-measure accuracy and is calculated as

$$\text{F - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Twitter sentiment analysis approaches

Sentiment analysis is an emerging field used for classifying sentiment or polarity of vulnerable tweets to reach appropriate conclusions. There are two widely used approaches namely lexicon based and machine learning based approach. Figure 5. show the classification of sentiment analysis approaches.

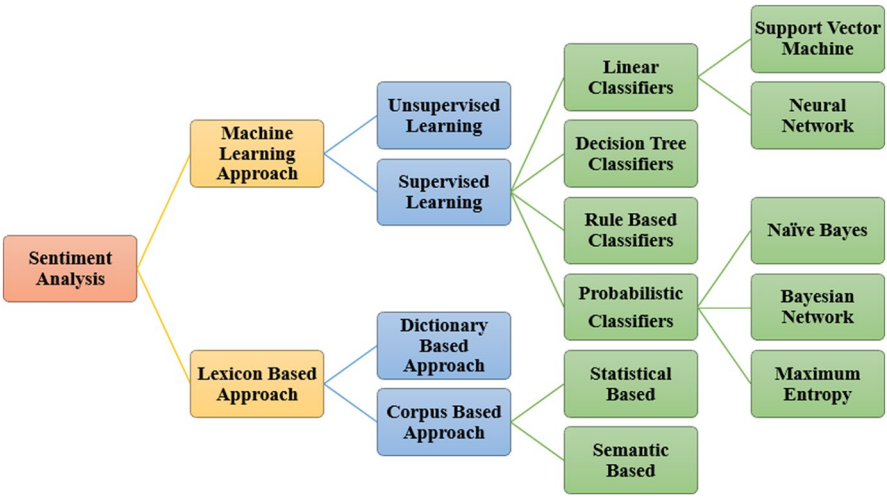

Fig. 5 Sentiment analysis approaches

Machine learning based approaches

Omer (2015) et al. presented a new machine learning based approach using AdaBoost classifier. Initially, author collected three datasets which are supporters of ISIS (TW-PRO), anti-supporters of ISIS (TW-CON), and random tweet dataset that have no connection with ISIS (TW-RAND). The total number of collected tweets was 135,608, and selected 619 features based on stylometric, time-based, and sentiment-based feature selection processes. He used three classifiers namely support vector machine (SVM), AdaBoost Naive Bayes (NB), and obtained impressive outcomes using AdaBoost with 100% on correctly classifying the instances, while NB was 99.9% and SVM was 99.1%.

Smedt et al. (2018) et al. developed a technique based on NLP and machine learning (ML) to automatically detect jihadist hatred speech. They gathered around 45,000 tweets over the period from October 2014 to December 2016. LIBSVM machine learning algorithm was used for balanced training of data. They analysed the sentiment based on accuracy that varied according to the language being spoken. For instance, 80% for French, 79% for English, Farsi was 80%, 84% for Arabic, and Portuguese was 81% with overall accuracy was 82%.

Mirani and Sasi (2016) et al. proposed a unique approach of combining geolocation with data mining algorithms. They contributed an innovative system for categorizing ISIS-related Tweets based on polarity-based classification. Using "Jeffrey Breen" algorithm, they compared five hashtags #ISLAMICSTATE, #DAESH, #ISIS, #ISIL, #IS. The algorithm performance was measured by accuracy, F value, recall and precision. Among all the five hashtags #ISIL provided the highest accuracy while testing the dataset on SVM, maximum entropy, bagging, random forest, and decision tree. Average accuracy 90% was achieved after tenfold cross-validation test. Maximum accuracy 99% was obtained from maximum entropy classifier.

Nouh et al. (2019) et al. presented a new approach to automatically analyze extremism propaganda materials and radical content in the tweets. The authors collected data from 3 datasets including two Kaggle dataset (How ISIS uses Twitter—17,000 tweets and Tweets targeting ISIS—1,22,000 tweets) and one TwitterAPI crawled dataset consists of 8000 tweets from 1000 users. They applied TF-IDF and LIWC dictionary as feature extraction methods and achieved 80% accuracy by training them in various machine-learning models such as SVM, K Nearest Neighbor, and Random Forest. Davidson et al. also applied same feature engineering techniques for 24,802 tweets and achieved 0.91 Recall value with SVM. Table 1 depicts comparison analysis of machine leaning based approaches.

In Kaati et al. (2015), a method based on machine-learning technique to recognize twitter accounts, support jihadist groups and distribute propaganda content online was demonstrated. Feature engineering was performed by analysing data dependency and classified the features into data-dependent and data independent features and the combination of both. The authors used two datasets (English tweeps and Arabic tweeps) with tweets including hashtags related to jihadists and especially ISIS. They used binary text classification method to detect tweeps involved in media mujahideen and applied linguistic features to train the AdaBoost classifier. While performing validation test, accuracy for data dependent, data independent and the combination were 99.07%, 98.82%, and 99.51%, respectively, for English tweets. Whereas in Arabic, accuracy of data independent 82.4%, 84.66% from dependent data and 86.38% combined features. The results of English tweets had high accuracy, precision, and recall ratios than Arabic tweets.

In Ferrara et al. (2016), a sentiment analysis technique to predict the polarity in their interaction with extremists was discussed. For this they used public dataset available in the name Lucky Troll Club which has 3,395,901 tweets from 25,000 user accounts. They used metadata as feature with greedy selection algorithm and classified the sentiment using Logistic Regression and Random Forest classifier models. Araque and Iglesias (2020) found a machine learning based approach to identify racial text on Twitter and online press or magazines. They have used twitter datasets such as Pro-Anti and Pro-Neu and online magazines to analyse the sentiment and generated the distributed representations of the text that are fed into Linear SVM and Logistic regression classifiers to compute the similarity between the analysed text and a particular lexicon. In addition, they proposed a novel approach that uses the emotion dictionary to calculate statistical summary of emotions in analysed text. They evaluated the performance of Pro-Neu, Pro-Anti, Magazines using F1-score metric and achieved 92.41%, 77.21% and 72.22%, respectively.

In Rehman et al. (2021), authors contributed in a work to identify the radical text in social media and believed that religious languages play a major role in radicalization on Twitter. They employed religious features and radical features to train the algorithm. They have taken 7000 tweets from 15 October 2019 to 20 October 2019 and performed feature engineering with radical features and religious features using TF-IDF technique. Then, they applied Naïve Bayes, SVM and Random Forest to predict the polarity. tenfold cross-validation test applied to validate the results and achieved F-Score value 0.87, which higher than existing works.

Table 1 Comparison of machine learning based approaches

Article	Feature extracted	Technique	Dataset	Performance Metrics	Drawbacks
Smedt et al. (2018)	Part of speech tagging	LIBSVM	45,000 tweets over the period of October 2014 to December 2016	Accuracy more than 80%	Analysis is based on only certain keywords in the Hate Corpus
Mirani and Sasi (2016)	TF-IDF LDA	SVM, Random Forest, Bagging, Decision Trees, Maximum Entropy	ISIS related Tweets	83% accuracy for SVM 83% accuracy for RF 82% accuracy for BG 80% accuracy for DT 86% accuracy for ME	Requires more data on geo location and time to improve the results
Fadel and Cemil (2020)	Part of speech tagging	Majority Voting between SVM, Naïve Bayes, and Logistic Regression	96,679 Tweets from 22 May 2017 to 31 October 2017	Accuracy = 94.8% F1 score = 95.9%	Accuracy for negative polarity is very less. Also, the work is based on simple feature selection method V + Adj and could be achieved better results if more feature extraction techniques used
Araque and Iglesias (2020)	Emotion based Features Simon Method	Linear SVM Logistic regression	Magazines and Twitter data sets such as Pro-Neu, Pro-Anti	F1-Score Pro-Neu—92.41% Pro-Anti—77.21% Magazines—72.22%	Dataset selection should be broader and usually text in news paper or magazines are less ideologic than online
Smith et al. (2020)	LIWC function	Logistic regression	40,053 tweets from 110 users related to Daesh supporters (ISIS) and 215,008 tweets from 109 users	Accuracy—89% F-score—0.89 Recall—0.88 Precision—0.90	Other feature selection methodologies could have given better results

Table 1 (continued)

Article	Feature extracted	Technique	Dataset	Performance Metrics	Drawbacks
Aleroud et al. (2020)	TF-IDV LDA Neighborhood over Lapping Term Correlation	SVM KNN Decision Tree Random Forest	Kaggle Data set 17,000 Tweets from 112 Pro- ISIS accounts 77, 813 Tweets from 95, 725 Anti—ISIS accounts	F1-Score 88% on original data 94% after data reduction	Has discrepancies between obtained and predicted model results due to topic modelling technique
Nouh et al. (2019)	LIWC Dictionary Bi-Grams	SVM KNN Neural Network Random Forest	Same as above two Kaggle datasets + 8000 Tweets from 1000 users	80% accuracy on SVM	Bi, tri grams decrease performance on binary classification
Omer (2015)	Stylometry based, Time based, Sentiment	AdaBoost SVM Naïve Bayes	135, 608 tweets from three dataset TW-PRO, TW-CON, and TW-RAND	AdaBoost—100% accuracy for correctly classifying the instances, NB—99.9% accuracy SVM—99.1% accuracy	Labelled dataset would have performed better with this technique
Hartung et al. (2017)	BoW,Bi-grams, Emoticons, Identity	SVM with Linear Kernal	45,747 Tweets	Accuracy 95%, Precision 25%	Shallow features extraction technique
Garg et al. (2017)	–	Combination of Naïve Bayes and SVM (NB-SVM)	59,988 tweets over a period from 16 September 2016 to 15 October, 2016	Not addressed	Lacks overall clarity of the objective of the work

Table 1 (continued)

Article	Feature extracted	Technique	Dataset	Performance Metrics	Drawbacks
Kaati et al. (2015)	Data-dependent Data-independent	AdaBoost Classifier	Tweets related to ISIS from 25th of June 2014 and 29th of August 2014	English Tweeps—Accuracy Data dependent- 9.07% Data independent–98.82% Combination of both – 99.51% Arabic Tweeps—Accuracy Data dependent– 84.66% Data independent – 82.4% Combination of both – 86.38%	Both selected feature extraction techniques are weak and did not solve the objective and lack of result validation
Ferrara et al. (2016)	Greedy Feature Selection	Logistic Regression, Random Forest	3,395,901 tweets from 25 K Twitter accounts from January 2015 to June 2015	AUC is between 72 and 83%	Dataset has suspended accounts; hence prediction of negative sample is very less
Jain and Vaidya (2021)	–	K-Means Clustering algorithm Naïve Bayes	Twitter data	Not addressed	Conceptual study
Omar et al. (2021)	BoW N-gram TF-IDF	Linear SVC, Logistic Regression, Random Forest	14,000 Tweets 33,000 Facebook Posts	Linear SVC with N-gram (1,2) achieved highest accuracy 97.92%	BoW and N-Grams increase false positives which greatly affects overall performance
Masood and Abbasi (2021)	TF-IDF Hashtags Bi-grams	SVM Random Forest Logistic Regression Gaussian Naïve Bayes	284,000 Tweets	Not addressed	Less performance over single graph embedding features

Table 1 (continued)

Article	Feature extracted	Technique	Dataset	Performance Metrics	Drawbacks
Rehman et al. (2021)	TF-IDF Text Corpus	Naïve Bayes, SVM and Random Forest	7000 tweets from 15 October 2019 to 20 October 2019	F-Score value 0.87	As data sample is very less leads false positives which affects system accuracy
Sharma and Jain (2020)	Word2Vec + K Means	Logistic Regression	Kaggle Data set 17,000 Tweets from 112 Pro- ISIS accounts	85.8% Accuracy	Lack of validations of other performance metrics
Sharif et al. (2019)	N-gram TF-IDF	Naïve Bayes SVM Ensemble Classifier Random Forest	3380 Tweets	84% accuracy on SVM	Biased results on detection as feature selection is weak

In An et al. (2021), authors presented a technique based on supervised machine learning to foresee terrorist events or potential risks using microblog entries (tweets). The authors used a combined approach of Word2Vec and K-means clustering to identify topics which will be further used for emotion analysis. Logistic regression classifier was used and achieved 85.8% accuracy. Garg et al. (2017) studied survival and sentiment from post-terror attack tweets. They considered the features like last retweet, number of retweets, number of favourites to study the information flow on Twitter. They adopted the combination of Naïve Bayes and SVM (NB-SVM) classifier to find the polarity of the information flow from 59,988 tweets taken over the period from 16 September 2016 to 15 October 2016. Moreover, the results shown the negative tweets lasted long than the positive tweets, though the number of negative tweets is significantly lesser than positive tweets.

Authors in Smith et al. (2020), created a paradigm to detect and predict the changes in users mind when they are in psychological group memberships through Twitter posts. They analyzed the longitudinal changes in individual user's twitter post over time. For this, they collected 40,053 tweets from 110 users which related to support of Daesh (ISIS) and compared them with baseline Twitter timelines of 215,008 tweets from 109 users. They used logistic regression classifier to classify the accounts into baseline users or Daesh supports, and they validated the results using 10 – fold cross validation testing and achieved 89% accuracy, F-score 89%, recall 88%, and precision 90%.

Authors in Omar et al. (2021), developed a technique to find correlation between hate speech and topics available in online social media. They collected 14,000 tweets and 33,000 Facebook posts and developed a multi-label Arabic dataset and performed manual annotation by dividing them into 11 classes. To perform multi-label classifications, they applied machine learning classifiers such as Linear SVC, Logistic Regression, Random Forest with feature representations N-gram, TF-IDF, and Bow to classify the sentiment polarity into positive, negative, and neutral and achieved highest accuracy of 97.92% in Linear SVC with N -gram (1,2) classifier.

In Dadkhah et al. (2021), authors proposed a method to detect online hostile activities automatically by analysing the polarity of online news content. They investigated many datasets in various dimensions such as role, influential level, vulnerabilities, and distribution pattern. Authors implemented the detection system using machine-learning techniques, deep-learning models, NLP, and Social Network Analysis techniques and analysed the data based on bot score, credibility score, classification score, topic modelling, name entity recognition, truthful score, sentiment score, risky score, and community detection. They contributed a visual data analytics framework to provide a complete understanding of cyber activities at several levels and results were evaluated with tenfold effectiveness test and achieved 95% approximately. Hartung et al. (2017) et al. demonstrated an idea to detect whether a Twitter user is Right-Wing extremist or non-extremist using 45,747 Tweets as dataset. They used Bag of Words (BoW), Bi-grams, Emoticons, Identity as features and achieved 95% accuracy in SVM Classifier.

Authors in Jain and Vaidya (2021), proposed an idea to analyze the sentiment of people on Uri, Pulwama and Surgical Strike attack. They collected tweets related to these attacks hashtags to find sentiment based on user's geolocation. The authors

used K-Means Clustering algorithm to find geolocation of the users and Naïve Bayes classifier to classify the orientation of user's sentiment with emotions such as anger, anxiety, and sadness towards these attacks from their tweets into positive, negative, and neutral polarity.

In Aleroud et al. (2020), authors suggested a methodology based on feature augmentation which is used to categorize the twitter accounts into Pro-ISIS and Anti—ISIS accounts. Terms from the tweets are considered as nodes in a graph, clustered them based on similar terms. They have collected 2 Kaggle Datasets where first has 17,000 Tweets from 112 Pro-ISIS accounts and the second has 77, 813 Tweets from 95, 725 Anti-ISIS accounts. They tested the data on SVM, KNN, Decision Tree, and Random Forest and achieved F1-Score of 88% on original data 94% after data reduction.

In Masood and Abbasi (2021), authors proposed a framework called Supervised Rebel Identification to identify the rebel users on Twitter. They developed a unique methodology to structure the tweets into directed user graph. The user graph then converted into graph embedding to use the semantics within the machine-learning classifiers such as SVM, Random Forest, Gaussian Naïve Bayes, and Logistic regression. They used to 284,000 tweets to classify them into rebel user, counter rebel and normal user. Similarly, Abrar et al. (2019) proposed a machine-learning technique for real-time analysis of terrorist-related tweets. They extracted feature from N-grams methods on 55,123 tweets and classified using AVM, Multinomial Logistic Regression models.

Deep-learning based approaches

Deep Learning is based on artificial neural networks in which multiple layers of processing are used to progressively extract high-level features from data (Nizzoli et al. 2019). Table 2 summarizes the techniques based deep learning techniques. Harb and Becker (2018) et al. found an approach based on the study of emotional reactions of Twitter users on a few terrorist events that occurred in United Kingdom. They have used two deep-learning architectures to create an emotion classifier and developed an analysis on tweets related to terrorist events to understand whether there is an emotional shift due to the terrorist attack and whether the emotional reactions are dependent on the incident, or on the demographics of the users (Harb et al. 2019). Both models, based on convolutional and recurrent neural network architectures, offered almost similar performances. The analysis shown an emotion shift due to the events and a difference in the reactions to each specific event, where gender is the most critical factor the results were obtained with the precision above 70%, recall is above 70% and F-Measure is below 60%.

Authors in Alhalabi et al. (2021), developed artificial intelligence-based terrorist behavior detection system. In this work, the authors proposed a distinct value proposition is based on unified methodology to characterize the Arabic tweets available on Twitter. The system uses advanced social mining techniques to detect terrorist behavioral patterns, provides enhanced visualization and decision-making. They

Table 2 Comparison of deep learning-based approaches

Article	Feature extracted	Technique	Dataset	Performance metrics	Drawbacks
Alhalabi et al. (2021)	Fuzzy ratings	Deep Learning	10,000 Tweets from July 2018 and October 2018	Not addressed	Expert reviews may be wrong in judging the true sentiment of the tweeters and lack of validation of performance of the system
Harb et al. (2020)	GloVe embedding	Deep Learning -CNN biLSTM BERT	Tweets from pre/post mass shooting events	Average F1 – score 76%	Yield better outcome if include multilingual emotional features
Harb and Becker (2018) and Harb et al. (2019)	Keyword Dictionary	Convolutional Neural Network and LSTM-CNN model	Tweets based on US and UK attacks	Precision is above 70% Recall is above 70% F-Measure is below 60%	Focused only on negative emotions hence overall performance will be affected by biased hypothesis
Zinovyeva et al. (2020)	TF-IDF	Deep Learning - SVM Random Forest Logistic Regression LightGBM	Twitter data and online contents	Average precision .996 approximately on all the datasets and models	GloVE and Word2Vec are most preferrable feature engineering techniques and can be used to best outcome
Ahmad et al. (2019)	Word embedding	LSTM+CNN	25,000 Tweets	Accuracy 92.66%	Lacks automated data crawling and did not consider visual and social context features for modeling

collected 10,000 tweets over the period from July 2018 to October 2018 and analyzed the polarity using deep-learning models.

Authors in Ahmad et al. (2019), presented a deep learning-based technique to analyse the sentiment and classifies the tweets into extremism or non-extremism categories. Their proposed work operated in three segments such as users' tweet collection, pre-processing, and classification of tweets with respect to extremist and non-extremist classes using LSTM + CNN model and obtained accuracy 92.66%.

In Zinovyeva et al. (2020), authors elaborated the detection of anti-social online behaviour using NLP deep learning. The authors compared their work with existing deep-leaning-based detection methodologies. For this they have considered four data sets of online social media including Twitter. They used SVM, Random Forest, Logistic Regression and LightGBM (Light Gradient Boosting Model) to classify the data and obtained average precision 99.6% approximately on all models. Harb et al. (2020) et al. developed a framework to analyse the emotional responses over various mass shooting events and its influential features. They have collected tweets from two days before and five days after eight different mass shooting events. They created emotion classifiers using tree different deep learning strategies such as Convolutional Neural Network, biLSTM, and BERT and classified the emotions into anger, fear, sadness, surprise, disgust, and no-feel and achieved the average F-measure nearly 75%.

Lexicon-based approaches

Lexicon-based methods employs word list or annotated dictionary by polarity score to determine opinion score of given data. This method does not require training data. Analysing tweets for find the polarity using lexicon is challenging because of the ever-changing colloquial expressions and hashtags (Giachanou and Crestani 2016). However, there has been quite a few existing works have been proposed using lexicon-based approaches as shown in Table 3. Simon et al. (2014) et al. developed a document level sentiment analysis to analyze the sentiment from original tweets communicated from the field by emergency organizations and their managers during the Kenya Westgate Mall attack. The authors used corpus-based approach to analyze the positive and negative classifications of the tweets and recommended that emergency organizations dispatchers from the field and the communication center should minimize the use of negative emotion during their communication with the public at that time. They have used 67,849 tweets, collected from 21 to 25, September 2013 and got 59.6% accuracy for positive classifications of manager tweets 46.5% accuracy for negative classifications of emergency organizations tweets.

In Mansour (2018), authors proposed an approach to examine the sentiment of people from western countries and eastern how they look at or their sympathy ISIS entity. The author used text sentiment analysis to analyze the word frequency and sentiment of the words using Term Frequency -Inverse Document Frequency (TF -IDF) tool from 6853 tweets over the period of Sep 2017- Dec 2017. Moreover, the author obtained positive accuracy 29% to 33% and negative accuracy 67% to 71%. Fadel and Cemil (2020) created a model for automatically classifying users' reviews

Table 3 Comparison of lexicon-based approaches

Article	Feature extracted	Technique	Dataset	Performance metrics	Drawbacks
Mansour (2018)	TF-IDF	Corpus-based approach	6853 tweets From Sep 2017—Dec 2017	Positive Accuracy 29% to 33% Negative Accuracy 67% to 71%	Lack of validation on performance metrics and quality of text feature could be improved to increase the positive accuracy
Kharde and Sonawane (2016)	Part of speech tagging	Lexicon based methods	Tweets after Manchester attacks and the Las Vegas shooting	Not addressed	Lack of representation of results and it is a conceptual work
Simon et al. (2014)	Dictionary	Corpus-based approach	67,849 tweets Dated, 21 to 25, Sep, 2013	59.6% accuracy for positive classifications of managers tweets 46.5% accuracy for negative classifications of organizations tweets	In depth analysis required as ambiguous result visualization and the objective of study
Al-Khalisy and Jehlol (2018)	Bag of words	Dictionary-based approach and Naïve Bayes	10,322 Tweets	Not addressed	Selected feature is not suitable and related to the objective of the paper
Ferrara et al. (2016)	Dictionary	Lexicon based methods	154 K Twitter Users & 3200 Tweets per user collected during 2014	Not addressed	Based on two hypotheses with lack of performance analysis
Kumar et al. (2017)	Dictionary	Lexicon based methods	How ISIS uses Twitter—17,000 tweets	Precision – 0.05	Shallow method with more false positives
Deven et al. (2018)	N-grams	Unsupervised machine learning	Tweets from 3325 accounts	F1-Score Iteration 1—96% Iteration 2—87% Iteration 3—82%	As N-grams reduces the classification accuracy, better to use corpus based approaches

on Twitter after a terrorist attack. The model was developed using lexicon and machine-learning approaches. Lexicon approach was used to create labelled training dataset while machine-learning approach was used to build the model. Scores of some domain related words were neutralized to avoid their negative effect. Features were selected based on Part of Speech tags such as VER, ADJ and the combination of both VER + ADJ. The author used majority voting between NB, SVM and LR machine-learning classification algorithms was applied. The performance of classification algorithms was measured using accuracy and F1 scores. Negative polarity tweets were categorized as terrorist supporters while positive polarity categorized as non-supporters. The results were compared to identify the best classification algorithm for features selection. This model achieved 94.8% accuracy with 95.9% F1 score.

Authors in Kostakos et al. (2018), carried out a comprehensive study based on the events—Manchester attacks and Las Vegas shooting to analyze the reactions shown and the way those reactions spread over the incident timeline in Twitter. They found "echo chambers" that is group of people sharing similar interest about the same event. They assigned positive and negative scores for each tweet using two lexicon-based methods. First, SentiWordNet 3.0. scores were determined for every single word by examining negative and positive values from a lexicon by the word and its PoS tag. The second SentiStrength was used to find final sentiment score, which was calculated by adding the positive and negative scores and then divided by number of words in tweet. Though the sentiment analysis technique used by the authors classifies the real news and fake news, they should provide the results based the certain performance metrics such as accuracy, precision, and F-score.

In Rekik et al. (2020), authors developed a recursive method to detect radical groups on social media mainly Twitter. Their analysis is based on violent vocabulary and suspicious interactions by anti-social communities, and they computed the danger degrees of the recognized users to find radical communities. They have performed an unsupervised learning analysis on tweets from 3325 accounts and iterated the analysis for 3 times and measures the performance using F-measure.

Authors in Ngoge (2016), proposed a machine-learning-based technique to determine the level of twitter terrorism and to identify the terrorist activities. To achieve this, they have implemented Maximum Entropy, SVM, and Naïve Bayes classifier with Lexicon-based approaches to classify the trends in 346 tweets pertaining to terrorist attack for seven days in Kenya. They achieved 73% accuracy, 15% recall and precision rate 60% while predicting the real-time sentiment over the attack. Simon et al. (2014) et al. developed a methodology to determine the time of radicalization among twitter users using divergent behavior analysis. They considered 154 K users and created a lexicon-based corpus to analyze and found that only 727 users shown interests towards Pro-ISIS behavior.

In Al-Khalisy and Jehlol (2018), authors used data mining techniques to extract useful information such as supporter data such as location, account name and terrorism propaganda. They gathered around 10,322 tweets related to the keyword terrorism and then they performed preprocessing and converted it into a text corpus. Their proposed work consists of two modules such as analyzing twitter data and next was about mapping the sentiment with GeoJSON to find the location of terrorists. They

used manually created word list which has synonyms and antonyms from dictionary and used it to analyze the polarity. They also employed word bag feature by calculating the total number of the word points in tweets indicating the training data. Depending on the training data, Naive Bayes classifier classified 7122 tweets as negative.

Data sets and collection strategies

Dataset collection is a critical step in any research process. Nowadays, the collection of data on online extremism groups and activities has become extremely a hard process, since online extremism is considered a highly sensitive domain for its risk and security reasons. However, many researchers tried to collect their own data or used publicly available datasets for analysis purpose. By default, social media platforms are gatherers of user data, it acts as source to researcher as well. Among all other social network platforms, Twitter is notorious platform as it was widely used by extremists and become popular data source for researchers (Kostakos et al. 2018; Berger and Perez 2016; Conway et al. 2019; Aleroud et al. 2020; Torregrosa et al. 2021; Gaikwad et al. 2021). A lot of existing works who used their own dataset obtained textual data through web crawling tools (Sheth et al. 2021) like TwitterEcho or TwitterCrawl using keywords related to extremism (Kaati et al. 2015; Ngoge 2016; Smedt et al. 2018; Harb et al. 2019; Harb et al. 2020; Kumar et al. 2017). Such datasets contain information such as, basic account details along with the metadata such as followers and following details, tweet text, retweets, and mentions. However, such datasets have some serious limitations on their data collection process like lack of characterizing account inclusion errors and errors caused by lack of filtering and standard validation process (Rowe and Saif 2016; Ferrara et al. 2016; Deven et al. 2018; Berger and Morgan 2015).

On the other hand, publicly available dataset from Kaggle.com (Tribe et al. 2015; Dataset 2016) was widely used in online extremism research to avoid hustles in custom dataset. Such datasets are mainly based on supporters of ISIS and anti-supporters of ISIS (Aleroud et al. 2020; Omer 2015; Sharma and Jain 2020). We found that standard datasets have some problems. First, even after a mass Twitter account suspension during 2016, still these datasets have accounts of suspended users and Table 4. Depicts sources of standard datasets. In addition, to get new insights on counter terrorism research authors may try to use new dataset instead of using same old dataset over and over.

Discussion and future research directions

Notwithstanding ongoing headways in Twitter Sentiment Analysis, it is as yet an open area for research and many issues are underexplored. The most important challenges are lack of verification of datasets, lack of benchmarks in research fields since most of the existing systems were based on existing theories, solution to the multilingual and multimodal content (Fernandez and Alani 2021; Softness 2016). In this

Table 4 Shows the sources of publicly available datasets

Article	Dataset	Type	Size
Tribe et al. (2015)	How ISIS Uses Twitter	Public	17,410 Tweets from 112 users
Dataset (2016)	Tweets targeting ISIS	Public	122,000 tweets from 95,725 users
Davidson et al. (2017)	Automated Hate Speech Detection and the Problem of Offensive Language	Public	24,802 tweets
Olteanu et al. (2015)	Crisis Lex Dataset	Public	Unspecified
Li et al. (2012)	UDI-TwitterCrawlAug2012	Public	50,000,000 tweets from 147,909 users
Li et al. (2013)	ATMTwitterCrawl-Aug2013	Public	50,00,000 tweets

section, we will discuss the attainment of research objectives to emphasize on some of the potential prospects to target the problem of online extremism (Torregrosa et al. 2021; Gaikwad et al. 2021; Narula and Jindal 2015) The attainment of research objectives of this work will be the learning outcome of this literature review. The attainment level of the objectives can be checked through the insights obtained from the review process of this article.

RO1: Outline the availability of various data sources or datasets and tools pertaining for combatting online extremism

Based on this comprehensive review, Twitter is the most identified OSN when it comes counter extremism research. Hence data can be obtained directly from Twitter using Twitter API or data crawling techniques. Public datasets can also be downloaded from various websites such as Kaggle.com. From the period of 2016 to 2017, Twitter has suspended 1,210,357 ISIS accounts for strict safety policies and most of the researchers created and used the dataset which closely around that period. Due to this, there is a low availability of standard datasets on the internet. On the other hand, most of the available datasets lacks inter-rater agreement which often reflects on less accuracy during classification. A summary of the custom-made or public datasets with dataset size, articles used, are shown in "Data Sets and Collection Strategies" section.

RO2: Summarize how sentiment analysis techniques used in the field of extremism research

In the literature, feature extraction methods such as TF-IDF, Part-of-speech Tagging, Topic Modelling (LDA), and N-grams were combined with various machine-learning techniques to identify the sentiment of the tweets. Researchers used sentiment analysis techniques not only to identify the most common terms related extremism but also to find the polarity of emotions of tweets after some real-life events, reactions to comments, message content analysis and to detect abnormal or hostile activities of extremists.

RO3: Present current topic and contributions from machine learning techniques to extremism research

Machine-learning-based extremism research has increased over the years, especially after 2015 Paris attack as extremists used Twitter to communicate their agenda. Since then, counter terrorism seems to be popular among researchers by

proposing more ideas to prevent radicalization or community identification for avoiding future terrorist attacks. Most of the existing machine-learning algorithm are based on basic feature extraction techniques such as TF-IDF, Bag of Words, N-grams and PoS tagging and classifies the polarity using SVMs, Naïve Bayes, Logistic Regression classifiers. Since a few years back, Deep learning methods have been gaining popularity in fast pace as researchers are using various deep learning models such as BERT, LSTM and Convolutional Neural Network along with effective feature extraction methods such as Word2Vec, and Word Embeddings to classify extremism content.

RO4: Discuss data validation techniques steps required for sentiment analysis in extremism research

Most of the existing surveys related to extremism research lacks to discuss about the importance of data validation techniques while working on text mining problems (Ferrara et al. 2016; Tang et al. 2015). During data collection stage, the researcher should check for data quality to avoid inclusion of irrelevant accounts and exclusion extremist account, and this can be solved by collecting the data using appropriate keywords. In the next stage, context identification is critical as it is the most challenging part of the research. Researchers should choose the right extremism context such behavior, religious or psychology as their core work. Finally, obtained results or performance metrics should be evaluated to identify the data imbalance or micro/macro differences in accuracy, precision, recall and F-value.

RO5: Throw light on future directions and challenges of the domain based on this study

Preparing a dataset is a critical step in sentiment analysis (Fadel and Cemil 2020). Especially, obtaining a dataset which is related to extremism content is not an easy task. The availability of data sources will remain as one of the more challenging tasks to confront online extremism (Adek and Ula 2021). Moreover, it has been observed that most of the researchers (Mansour 2018) collected the datasets on their own (Smedt et al. 2018) and presented results based on the dataset. Some of the studied works based on event based (Kharde and Sonawane 2016; Ngoge 2016; Harb et al. 2019; Simon et al. 2014) data set collection (Rehman et al. 2021; Alhalabi et al. 2021; Zinovyeva et al. 2020; Berger and Morgan 2015; Jaki and Smedt 2019; Berger 2016). Hence, if there is an opportunity in future for preparing and sharing full datasets with other researchers by following proper protocols and ethics, it will be a new dimension for researchers to delve into this area and produce quality outcomes.

Another issue is that the interest in analyzing online extremism is mostly reliant on pre-existing feature selection methods, but not on the insight extraction. For effective insight extraction, researchers should be aware of terrorism context such as psychology, ideology, and belief before developing models (Gaikwad et al. 2021; Lara Cabrera et al. 2019). At the same time, feature selection is also critical as one of the main drawbacks of machine learning algorithms is that the efficacy of the approach depends on the extracted features. Hence, there should be a balance between appropriate feature selection and insight extraction instead choosing same old TF-IDF or N-grams or BoW techniques. Having said that, 95% of the surveyed research work did not address the problem of negation detection. Most of the

existing analysis is based on SVM, Naïve Bayes, Bayesian Networks and Logistic Regression. In all these different analysis the area of negation detection remains underexplored (Omer 2015; Berger 2018; Zucco et al. 2019). Having said that, an interesting dimension would be studying the efficiency of the neural network algorithms on negation handling in Twitter Sentiment Analysis (Esraa Najjar and Salam Al Augby 2021; Zinovyeva et al. 2020).

Twitter supports 34 different languages and allows the user to use more than one language in the same tweet. This presence of multiple languages in tweets is incredibly challenging during sentiment analysis. Especially, in twitter terrorism, only a few researchers addressed multilingual sentiment analysis of tweets. The use of new approaches, such as word embedding (and, specially, those that recognize word variations) could be the right direction to follow here, together with the creation of specific lexicons for different types of extremism. One more challenge lies in analyzing the polarity based on correlations among various factors such as geographical locations, gender, and age.

The extension of tweet length from 140 to 280 opens new possibilities in sentiment analysis by providing us with more data to analyse. However, this length extension may also lead the users to use more informal language such as the usage of emoticons, and slang. In some cases, longer tweets also mean discussion of different topics in the same tweet – which poses a new challenge of isolating these different topics.

Limitations of the study

This comprehensive review on extremist content on Twitter might have been investigated and discussed much less papers than an average survey paper does in different domains. However, this is due primarily to the fact that after the Paris Attack in 2015, Twitter has suspended a huge number of extremist accounts on its platform. Moreover, another factor that contributed to limit the data sets contents, is that the scope of this study is limited to the analysis of Jihadism related content. In summary, only 32 studies were available for this review with the data ranging from 2014 to 2019.

Conclusion

In this digital age, social networks have an inevitable presence in our daily lives and it is difficult for people to survive without them. Twitter is an extremely popular platform among all organizations, including terrorist organization, to reach the masses with their message. This study was motivated by the continuing increase in online activities by terrorist organizations in Twitter, where there is lack of automated techniques to predict such terrorism-related activities. The survey looked at different sentiment classification techniques and algorithms that have been tested by various researchers working with different datasets. This research work contributes to provide a better understanding of Twitter Sentiment Analysis using Lexicon based

methods, such as Dictionary-based approach, Corpus-based and Machine Learning-based approach such as SVM, Bayesian Networks, Maximum Entropy, Naive Bayes, and Neural Networks. Furthermore, based on the above analysis, Machine learning-based approaches were the most common methods used by researchers. Though Support Vector Machine and Naïve Bayes were the two most frequently adopted methods; yet the highest accuracy was achieved by AdaBoost classifier. Thus, this survey provides a comprehensive overview of the existing SAT methods and highlights promising future research directions in confronting cyber terrorism.

References

Abrar, M.F., M.S. Arefin, and M.S. Hossain. 2019. A framework for analysing real-time tweets to detect terrorist activities. *Proceedings of International Conference on Electrical, Computer and Communication Engineering*. https://doi.org/10.1109/ECACE.2019.8679430.

Adek, and Bustami Ula. 2021. Systematics review on the application of social media analytics for detecting radical and extremist group. *Materials Science and Engineering* 1071: 012029. https://doi.org/10.1088/1757-899X/1071/1/012029.

Ahmad, Shakeel, Muhammad Zubair Asghar, Fahad M. Alotaibi, and Irfanullah Awan. 2019. Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. *Human Centric Computing and Information Sciences*. https://doi.org/10.1186/s13673-019-0185-6.

Aleroud, Ahmed, Nisreen Abu-Alsheeh, and Emad Al-Shawakfa. 2020. A graph proximity feature augmentation approach for identifying accounts of terrorists on twitter. *Computers and Security* 99: 102056. https://doi.org/10.1016/j.cose.2020.102056.

Al-Khalisy, Muhanad A. E., and Hashem B. Jehlol. 2018. Terrorist affiliations identifying through twitter social media analysis using data mining and web mapping techniques. *Journal of Engineering and Applied Sciences* 13: 7459–7464. https://doi.org/10.36478/jeasci.2018.7459.7464.

Alhalabi, Wadee, Jari Jussila, Kamal Jambi, Anna Visvizi, Hafsa Qureshi, Miltiadis Lytras, Areej Malibari, and Raniah Samir Adham. 2021. Social mining for terroristic behavior detection through arabic tweets characterization. *Future Generation Computer Systems* 116: 132–144. https://doi.org/10.1016/j.future.2020.10.027.

Ali, Mah-Rukh. 2015. Isis and propaganda: How isis exploits women. *University of Oxford archive*. https://reutersinstitute.politics.ox.ac.uk/our-research/isis-and-propaganda.

An, Lu., Yuxin Han, Xingyue Yi, Gang Li, and Yu. Chuanming. 2021. Prediction and evolution of the influence of microblog entries in the context of terrorist events. *Social Science Computer Review*. https://doi.org/10.1177/08944393211029193.

Araque, Oscar, and Carlos A. Iglesias. 2020. An approach for radicalization detection based on emotion signals and semantic similarity. *IEEE Access* 8: 17877–17891. https://doi.org/10.1109/ACCESS.2020.2967219.

Berger, J.M. 2016. Nazis vs. isis on twitter: A comparative study of white nationalist and isis online social media networks. *Archive*. https://extremism.gwu.edu/sites/g/_les/zaxdzs2191/f/downloads/Nazisv.ISIS.pdf.

Berger, J.M. 2018. The alt-right twitter census: Defining and describing the audience for alt-right content on twitter. *Archive*. https://www.voxpol.eu/download/vox-pol_publication/AltRightTwitterCensus.pdf.

Berger, J.M., and J. Morgan. 2015. The isis twitter census. https://www.brookings.edu/wp-ontent/uploads/2016/06/isis_twitter_census_berger_morgan.pdf.

Berger, J.M., and H. Perez. 2016. The islamic state's diminishing returns on twitter: How suspensions are limiting the social networks of english-speaking isis supporters. George Washington University archive. https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/downloads/JMB%20Diminishing%20Returns.pdf. Accessed Feb 2016.

Conway, Maura, Moign Khawaja, Suraj Lakhani, Jeremy Reffin, Andrew Robertson, and David Weir. 2019. Disrupting daesh: Measuring takedown of online terrorist material and its impacts. *Studies in Conflict & Terrorism* 42 (1–2): 141–160. https://doi.org/10.1080/1057610X.2018.1513984.

Dadkhah, Sajjad, Farzaneh Shoeleh, Mohammad Mehdi Yadollahi, Xichen Zhang, and Ali A. Ghorbani. 2021. A real-time hostile activities analyses and detection system. *Applied Soft Computing Journal* 104: 107175.

Davidson, Thomas, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. Data retrieved from *Github*. https://github.com/t-davidson/hate-speech-and-offensive-language

De Smedt, Tom, Guy De Pauw, and Pieter Van Ostaeyen. 2018. Automatic detection of online jihadist hate speech. *Computational Linguistics & Psycholinguistics Technical Report Series*-007.

Deven, Parekh, Amarnath Amarasingam, Lorne L. Dawson, and Derek Ruths. 2018. Studying jihadists on social media: A critique of data collection methodologies. *Perspectives on Terrorism* 12 (3): 3–21.

Fadel, Ibrahim A., and Ö.Z. Cemil. 2020. A sentiment analysis model for terrorist attacks reviews on twitter. *Sakarya University Journal of Science* 24 (6): 1294–1302. https://doi.org/10.16984/saufenbilder.711612.

Fernandez, Miriam, and Harith Alani. 2021. Artificial intelligence and online extremism—challenges and opportunities. Predictive Policing and Artificial Intelligence 1st Edition. Taylor & Francis. London: Routledge.

Ferrara, E., W.Q. Wang, O. Varol, A. Flammini, and A. Galstyan. 2016. Predicting online extremism content adopters and interaction reciprocity. *Proceedings of International Conference on Social Informatics*. https://doi.org/10.1007/978-3-319-47874-6_3.

Fifth Tribe - Kaggle dataset. 2015. How isis uses twitter. https://www.kaggle.com/ fifthtribe/how-isis-uses-twitter.

Gaikwad, M., S. Ahirrao, S. Phansalkar, and K. Kotecha. 2021. Online extremism detection: A systematic literature review with emphasis on datasets, classification techniques, validation methods, and tools. *IEEE Access* 9: 48364–48404. https://doi.org/10.1109/ACCESS.2021.3068313.

Garg, Pulkit, Himanshu Garg, and Virender Ranga. 2017. Sentiment analysis of the uri terror attack using twitter. *Proceedings of the International Conference on Computing, Communication and Automation*, vol. 17.

Giachanou, Anastasia, and Fabio Crestani. 2016. Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys*. https://doi.org/10.1145/2938640.

Harb, Jonathas G.D. 2019. Using a convolutional neural network to compare emotional reactions on Twitter to mass violent events. https://lume.ufrgs.br/handle/10183/198208.

Harb, Jonathas G.D., and Karin Becker. 2018. Emotion analysis of reaction to terrorism on twitter. *SBC 33rd Brazilian Symposium on Databases.* Brazil: Rio de Janeiro. https://sbbd.org.br/2018/wp-content/uploads/sites/5/2018/08/097-sbbd_2018-fp.pdf.

Harb, Jonathas G.D., R.'egis Ebeling, and Karin Becker. 2019. Exploring deep learning for the analysis of emotional reactions to terrorist events on twitter. *Journal of Information and Data Management* 10 (2): 97–115.

Harb, Jonathas G.D, Régis. Ebeling, and Karin Becker. 2020. A framework to analyze the emotional reactions to mass violent events on twitter and influential factors. *Information Processing & Management* 57 (6): 102372. https://doi.org/10.1016/j.ipm.2020.102372.

Hartung, M., R. Klinger, F. Schmidtke and L. Vogel. 2017. Identifying right-wing extremism in german twitter profiles: A classification approach. Lecture Notes in Computer Science. Cham: Springer 10260: 320–325. https://doi.org/10.1007/978-3-319-59569-6_40

Jain, Pooja N., and Archana S. Vaidya. 2021. Analysis of social media based on terrorism—A review. *Vietnam Journal of Computer Science* 8 (1): 1–21.

Jaki, S., and T. De Smedt. 2019. Right-wing german hate speech on twitter: Analysis and automatic detection. *Archive*. http://arxiv.org/abs/1910.07518.

Kaati, Lisa, Enghin Omer, Nico Prucha, and Amendra Shrestha. 2015. Detecting multipliers of jihadism on twitter. *Proceedings of 15th IEEE international conference on data mining work*: 954–960.

Kaggle Dataset. 2016. Tweets targeting isis. https://www.kaggle.com/isis-related-tweets/metadata.

Kharde, Vishal A., and S.S. Sonawane. 2016. Sentiment analysis of twitter data: A survey of techniques. *International Journal of Computer Applications* 139 (11): 0975–8887.

Kolkur, Seema, Gayatri Dantal, and Reena Mahe. 2015. Study of different levels for sentiment analysis. *International Journal of Current Engineering and Technology*.

Kostakos, P., M. Nykanen, M. Martinviita, A. Pandya, and M. Oussalah. 2018. Meta-terrorism: Identifying linguistic patterns in public discourse after an attack. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. https://doi.org/10.1109/ASONAM.2018.8508647.

Kumar, Manish, Rajesh Bhatia, and Dhavleesh Rattan. 2017. A survey of web crawlers for information retrieval. *Wiley Interdisciplinary Reviews: Data Mining Knowledge Discovery* 7 (6): e1218. https://doi.org/10.1002/widm.1218.

Lara Cabrera, R., A. Gonzalez-Pardo, and D. Camacho. 2019. Statistical analysis of risk assessment factors and metrics to evaluate radicalisation in twitter. *Future Generation Computer Systems* 93: 971–978.

Li, Rui, Shengjie Wang, Hongbo Deng, Rui Wang, and Kevin Chen-Chuan Chang. 2012. Dataset-udi-twittercrawl aug2012. Data retrieved from *Wiki.illinois*. https://wiki.illinois.edu/wiki/display/forward/Dataset-UDI-TwitterCrawl-Aug2012.

Li, Rui, Shengjie Wang, and Kevin Chen-Chuan Chang. 2013. Dataset-atm-twittercrawl-aug2013. Data retrieved from *Wiki.illinois*. https://wiki.illinois.edu/wiki/display/forward/Dataset-ATM-TwitterCrawl-Aug2013

Mansour, Samah. 2018. Social media analysis of user's responses to terrorism using sentiment analysis and text mining. *Procedia Computer Science* 140: 95–103. https://doi.org/10.1016/j.procs.2018.10.297.

Masood, Muhammad Ali, and Rabeeh Ayaz Abbasi. 2021. Using graph embedding and machine learning to identify rebels on twitter. *Journal of Informetrics* 15: 101121.

Mirani, T.B., and S. Sasi. 2016. Sentiment analysis of isis related tweets using absolute location. *International Conference on Computational Science and Computational Intelligence*. https://doi.org/10.1109/CSCI.2016.0216.

Misra, Sanjay. 2021. A step-by-step guide for choosing project topics and writing research papers in ict related disciplines. *Communications in Computer and Information Science* 1350: 727–744.

Najjar, Esraa, and Salam Al Augby. 2021. Sentiment analysis combination in terrorist detection on twitter: A brief survey of approaches and techniques. *Advances in Intelligent Systems and Computing* 1254: 231–240. https://doi.org/10.1007/978-981-15-7527-3_23.

Narula, S., and N. Jindal. 2015. Social media, indian youth and cyber terrorism awareness: A comparative analysis. *Journal of Mass Communication & Journalism.* 5: 2.

Ngoge L.A. 2016. Real-time sentiment analysis for detection of terrorist activities in Kenya. *Strathmore University archive*. http://hdl.handle.net/11071/4826

Nizzoli, L., M. Avvenuti, S. Cresci and M. Tesconi. 2019. Extremist propaganda tweet classification with deep learning in realistic scenarios. *Proceedings of the 10th ACM Conference on Web Science*, 203–204. https://doi.org/10.1145/3292522.3326050.

Nouh, M., J.R.C. Nurse, and M. Goldsmith. 2019. Understanding the radical mind: Identifying signals to detect extremist content on Twitter. *Proceedings of IEEE International Conference on Intelligence and Security Informatics*. https://doi.org/10.1109/ISI.2019.8823548.

Olteanu, Alexandra, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2015. Crisis datasets collection. Data retrieved from *CrisisLex*. https://crisislex.org/data-collections.html.

Omar, Ahmed, Tarek M. Mahmoud, Tarek Abd-El-Hafeez, and Ahmed Mahfouz. 2021. Multi-label arabic text classification in online social networks. *Information Systems* 100: 101785.

Omer, Enghin. 2015. Using machine learning to identify jihadist messages on twitter. *Examensarbete 30 hp*. https://www.diva-portal.org/smash/get/diva2:846343/FULLTEXT01.pdf

Pai, Siddhesh, Vaibhav Bagri, Shivani Butala, and Pramod Bide. 2020. Survey of sentiment analysis of political content on twitter. *Lecture Notes in Electrical Engineering* 630: 169–180. https://doi.org/10.1007/978-981-15-2305-2_14.

Rehman, Zia Ul, Sagheer Abbas Muhammad Adnan. Khan, Ghulam Mustafa, Hira Fayyaz, Muhammad Hanif, and Muhammad Anwar Saeed. 2021. Understanding the language of isis: An empirical approach to detect radical content on twitter using machine learning. *Computers, Materials & Continua* 66 (2): 1075–1090. https://doi.org/10.32604/cmc.2020.012770.

Rekik, Amal, Salma Jamoussi, and Abdelmajid Ben Hamadou. 2020. A recursive methodology for radical communities' detection on social networks. *Procedia Computer Science* 176: 2010–2019. https://doi.org/10.1016/j.procs.2020.09.237.

Rowe, M., and H. Saif. 2016. Mining pro-isis radicalisation signals from social media users. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13023/12752.

Sharif, Waqas, Shahzad Mumtaz, Zubair Shafiq, Omer Riaz, Tenvir Ali, Mujtaba Husnain, and Gyu S. Choi. 2019. An empirical approach for extreme behavior identification through tweets using machine learning. *Applied Sciences* 9 (18): 3723. https://doi.org/10.3390/app9183723.

Sharma, Dipti, Munish Sabharwal, Vinay Goyal, and Mohit Vij. 2018. Sentiment analysis techniques for social media data: A review. *Advances in Intelligent Systems and Computing*, vol. 1045.

Sharma, Sanur, and Anurag Jain. 2020. Role of sentiment analysis in social media security and analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10 (5): e1366.

Sheth, A., V.L. Shalin, and U. Kursuncu. 2021. Defining and detecting toxicity on social media: Context and knowledge are key. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2021.11.095.

Simon, Tomer, Avishay Goldberg, Limor Aharonson Daniel, Dmitry Leykin, and Bruria Adini. 2014. Twitter in the crossfire - the use of social media in the westgate mall terror attack in Kenya. *PLoS ONE* 9 (8): e104136.

Smith, Laura G.E., Laura Wakeford, Timothy F. Cribbin, Julie Barnett, and Wai Kai Hou. 2020. Detecting psychological change through mobilizing interactions and changes in extremist linguistic style. *Computers in Human Behavior* 108: 106298.

Softness, Nicole. 2016. Terrorist communications: Are facebook, twitter, and google responsible for the islamic state's actions? *Journal of International Affairs* 70 (1): 201–215.

Tang, Duyu, Bing Qin, and Ting Liu. 2015. Deep learning for sentiment analysis: Successful approaches and future challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 5: 292–303.

Torregrosa, Javier, Gema Bello-Orgaz, Eugenio Martinez-Camara, Javier Del Ser, and David Camacho. 2021. A survey on extremism analysis using natural language processing. *Computers and Society*. https://doi.org/10.48550/arXiv.2104.04069.

Zerzri, Mayssa. 2017. The threat of cyber terrorism and recommendations for countermeasures. C. A. Perspectives on Tunisia No. 04. https://euagenda.eu/upload/publications/untitled-145478-ea.pdf.

Zinovyeva, Elizaveta, Wolfgang Karl Härdle, and Stefan Lessmann. 2020. Antisocial online behavior detection using deep learning. *Decision Support Systems* 138: 113362. https://doi.org/10.1016/j.dss.2020.113362.

Zucco, Chiara, Barbara Calabrese, Giuseppe Agapito, Pietro H. Guzzi, and Mario Cannataro. 2019. Sentiment analysis for mining texts and social networks data: Methods and tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10 (1): e1333. https://doi.org/10.1002/widm.1333.