Identification of markers and artificial intelligence-based classification of radical Twitter data

Mohammad Fraiwan
Computer Engineering, Jordan University of Science and Technology, Irbid, Jordan

Received 14 December 2021
Revised 1 January 2022
2 January 2022
Accepted 26 January 2022

Abstract

Purpose – Social networks (SNs) have recently evolved from a means of connecting people to becoming a tool for social engineering, radicalization, dissemination of propaganda and recruitment of terrorists. It is no secret that the majority of the Islamic State in Iraq and Syria (ISIS) members are Arabic speakers, and even the non-Arabs adopt Arabic nicknames. However, the majority of the literature researching the subject deals with non-Arabic languages. Moreover, the features involved in identifying radical Islamic content are shallow and the search or classification terms are common in daily chatter among people of the region. The authors aim at distinguishing normal conversation, influenced by the role religion plays in daily life, from terror-related content.

Design/methodology/approach – This article presents the authors' experience and the results of collecting, analyzing and classifying Twitter data from affiliated members of ISIS, as well as sympathizers. The authors used artificial intelligence (AI) and machine learning classification algorithms to categorize the tweets, as terror-related, generic religious, and unrelated.

Findings – The authors report the classification accuracy of the K-nearest neighbor (KNN), Bernoulli Naive Bayes (BNN) and support vector machine (SVM) [one-against-all (OAA) and all-against-all (AAA)] algorithms. The authors achieved a high classification F1 score of 83\%. The work in this paper will hopefully aid more accurate classification of radical content.

Originality/value – In this paper, the authors have collected and analyzed thousands of tweets advocating and promoting ISIS. The authors have identified many common markers and keywords characteristic of ISIS rhetoric. Moreover, the authors have applied text processing and AI machine learning techniques to classify the tweets into one of three categories: terror-related, non-terror political chatter and news and unrelated data-polluting tweets.

Keywords ISIS, Classification, Twitter, Radicalization, Arabic

Paper type Research paper

1. Introduction

Recent years have witnessed the rise of the ISIS, also known as the Islamic State in Iraq and Levant (ISIL), or Islamic State (IS) for short. Since 2014, it has grown from a local insurgency into a surprising regional force, such that major Iraqi cities (e.g. Mosul) and large swaths of Syria were taken under control in a very short period. An extreme ideology, gruesome beheadings and a well-run propaganda machine characterize the group. In the past few months, the international alliance has driven ISIS out of its major strongholds and liberated

Availability of data and material: The dataset is available from the following link: https://data.mendeley.com/datasets/8kftmw7rct/draft?a=0fd4d8fc-42e8-478b-bd17-ba61996aad61

Competing interests: The authors declare that they have no competing interests.

Funding: This work was funded by Jordan University of Science and Technology, Deanship of Research (No:20170247).

Applied Computing and Informatics
Emerald Publishing Limited
e-ISSN: 2210-8327
p-ISSN: 2634-1964
DOI 10.1108/ACI-12-2021-0326

the majority of areas under its control. Although major military operations are now focused on counter-intelligence and surgical strikes, the fight is still brewing in cyberspace. The group now aims at an underground strategy and inspiring lone wolf attacks [1] and is still highly active with many attacks in France, Germany, the UK, the USA and the region to name a few.

The number of SN users is projected to reach 3 Billion by 2021 [2]. This huge audience, ease of use and accessibility make these networks an excellent and a dangerous platform for propaganda. SNs provide great means for social engineering, radicalization and the indoctrination of the next generation of attackers/terrorists. The recent events of election meddling in the USA [3, 4] and the attempts to sway Brexit voting [5] demonstrate the power and capabilities of SNs in influencing public opinion. Moreover, terrorists with rare direct physical exposure to ISIS ideology conducted many lone wolf attacks in Europe and the USA. Online propaganda and recruitment drives were successful in driving these individuals to commit their crimes.

Arabs predominantly populate the Middle East and North Africa regions, which are rife with conflicts. Thus, it is natural for a good portion of the SNs and the online discussions to reflect this political and social turmoil. Moreover, the Arabic online chat content is full of religious reference to Allah (i.e. God to Muslims), the name of cities/places, fighting factions and groups and recent events. In this paper, we aim at separating this type of chatter, which the literature regularly marks as terrorist [6, 7], from terrorist propaganda and violence-advocating rhetoric. The contributions of this paper are as follows:

(1) We collect data from suspected terror-leaning Twitter accounts;

(2) We perform expert annotation and arbitration of the tweets to identify terrorist content;

(3) We identify several markers and keywords of ISIS propaganda and its sympathizers;

(4) We develop an AI-based classification of terrorist tweets with high accuracy (i.e. 83.6% F1 score) and

(5) We perform further analysis of some features of the terrorist Twitter accounts (e.g. number of followers/tweets).

2. Background and related work

Traditionally, the studies of users' online behavior have been focusing on identifying shopping or multimedia (e.g. YouTube videos) preferences. In the past few years, the research landscape has shifted dramatically toward analyzing the political implications of online content. There is a general feeling that governments were caught off-guard by the power of SNs in organizing events and forging public opinion. For example, several news outlets reported on a Russian-driven Facebook group that was able to amass 225,000 followers and organize rallies inside the USA from abroad [3, 4]. In this section, we discuss the main efforts in the literature to identify and analyze radical content and communities.

We can roughly classify the research into radicalization and Jihadist terrorism into two categories: radicalization detection and radicalization analysis. Detection refers to the ability to identify hate speech, violence and terrorism by employing text classification methods. Such systems have the potential to aid law enforcement agencies and Internet technology companies in neutralizing radical content and users. On the other hand, analysis goes a step deeper into examining relationships, metadata, behavioral characteristics and organizational structures hidden in SNs interactions. The research in radicalization analysis falls under two umbrellas as follows:

(1) Content-based analysis: Investigate radical posts based on linguistic patterns and develop factors/metrics to quantify the radicalization efforts. Some of the goals include:

- Explore goals, ideologies and online fundraising and propaganda drives of terrorist groups [8].
- Identifying the hidden authors of posts by extracting common stylistic features [9, 10].
- Quantifying the radicalization level toward different events. This is based mainly on typical sentiment analysis research [10, 11].
- Analysis of behavioral patterns of targeted users and their interactions [12, 13].
- Identifying trending topics being discussed by radical users [14, 15].

(2) Network-based analysis: Exploit metadata and online interactions (e.g. likes, re-tweets, comments, mentions, re-blogging and hyperlinks to other pages) to detect communities, social leaders or controllers [16, 17].

3. Methods

3.1 Aim of the study

In this paper, we focus primarily on Arabic Twitter data, which is the primary language used by ISIS members and sympathizers. We screen suspected accounts to include only those with extremist content. Our goals are different from the related literature in that we are trying to distinguish individual violence-advocating tweets. This is important because most of the related work relies on a relatively large material per user to determine their inclination. However, tweets are inherently short (i.e. 280 characters) and terror accounts are short-lived as the service providers block such accounts. In addition, radicals may continuously change their accounts to avoid monitoring. Moreover, the problem we are solving is much harder, given that hashtags, sentiment and keywords are good indicators of the user's inclinations, whereas they may not be so obvious in individual tweets. In addition, we identify several previously unrecognized important keywords commonly used by ISIS propaganda.

3.2 Data collection

Arabic social media datasets are not readily available for the public. Moreover, many terror-related accounts are actively being pursued and deactivated. In addition, most of the common datasets and lists are in English (e.g. Kaggle [6]). We have built our own crawler to download tweets from suspected ISIS accounts. The list of accounts is available from a Twitter hashtag called #OpISIS and provided by the red cult hackers group on their Pastebin website [18]. The majority of these tweets were written in Arabic with a few in other languages. We investigated 14,081 accounts, out of which 2,503 were still active and containing 1,610,448 tweets. However, it was surprising that the list contained many false positives, and many accounts contained normal chatter and not actually radical in an obvious manner. Thus, to select the accounts with the highest number of radical tweets, we applied a manual filtration process by a specialized arbiter based on the content. After exhausting the #OpISIS list, we used the list of accounts from the CtrlSection 1 Twitter group [19]. These efforts resulted in 173 accounts having 24,078 tweets. We used Python 2.7.14 and the Tweepy tool, which enabled us to retrieve the most recent 3,240 tweets per user. This is the maximum set by the Twitter policy [20].

3.3 Metadata

We performed further analysis of the metadata to identify if there are any important features. Most of the tweets (i.e. 97%) were from 2014 and 2015, when ISIS reached the height of its ground control. Figure 1 shows the date and time of publication. The time is typically later in the day, as people may be busy or working during the daytime, whereas the day of publication is almost uniformly distributed over the month. Regarding the users, most of the accounts were new, with many calls in the tweets for support from abuse reports made by vigilantes. Moreover, the number of followers was also small (i.e. less than 50). Figure 2 shows that for most users, the number of tweets is less than 200. It is clear that none of these features is peculiar to the terror tweets and thus were excluded from the classification model.

3.4 Preprocessing for classification

The Arabic language uses many stop words, which appear frequently in writings, yet they are of little utility for our objectives. Removing these words will disclose meaningful words and improve the performance of the classifiers, which is a common practice in the literature [21–23]. To this end, we performed several preprocessing steps on the Arabic tweets before

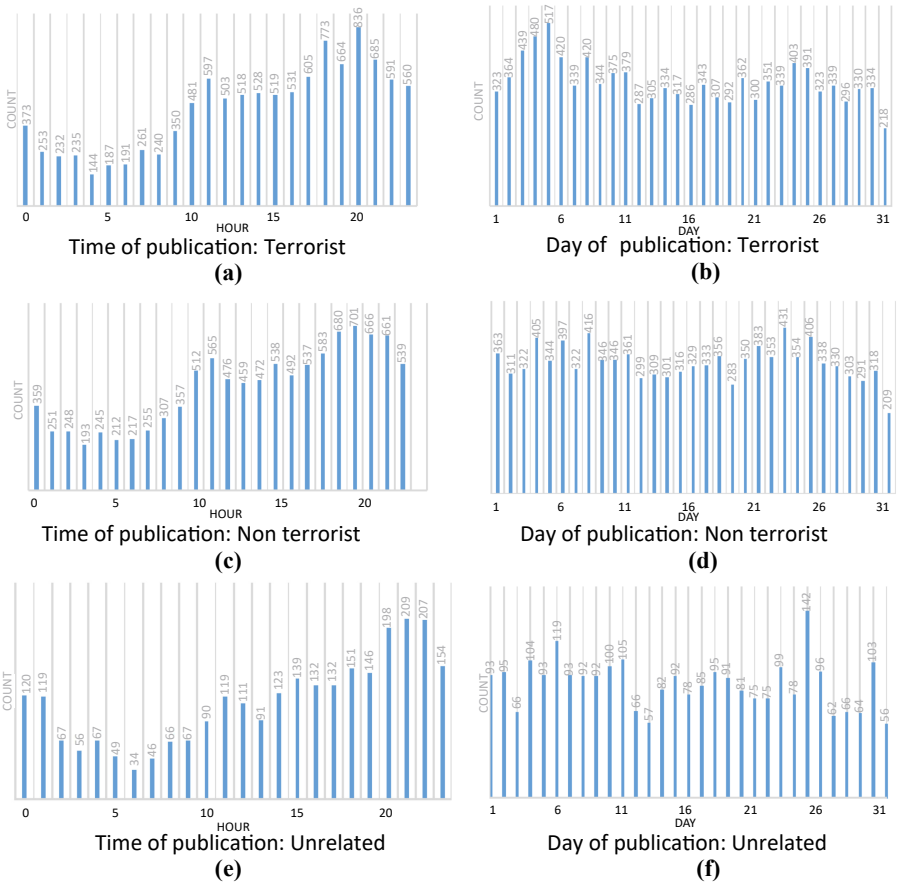
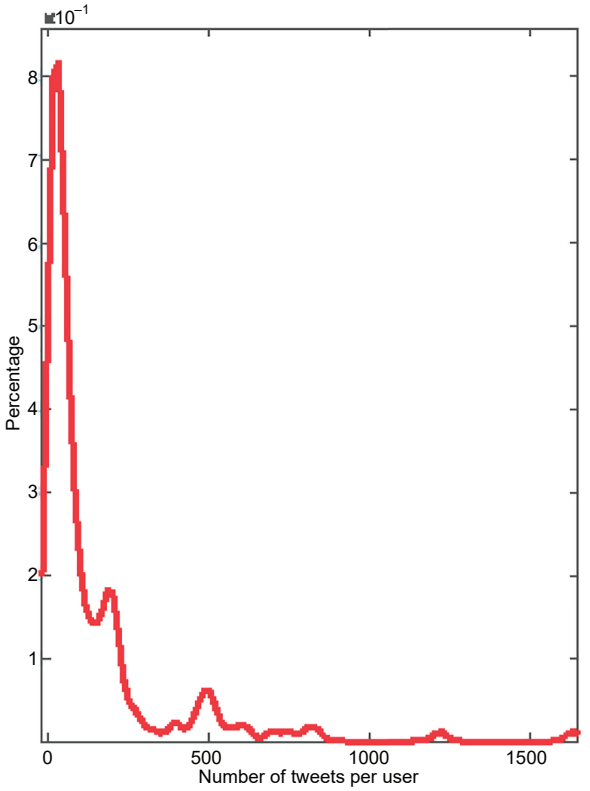


Figure 1.
Comparison of time and day of publication by class



Note(s): The *Y* axis represents the percentage of tweets out of 24,078

Figure 2.
The number of tweets per user

applying the classification algorithms. Such preprocessing is automated and it removes the matches across the dataset and prevents deteriorating the results by stop words (e.g. pronouns and prepositions). The process involves the steps as follows:

(1) Filtering the duplicate tweets;

(2) Tokenizing the tweets into uni/bi/tri-grams;

(3) Removing diacritic marks (i.e. pronunciation signs);

(4) Normalization by deleting repeated letters, along with removing Persian letters similar to the Arabic ones and

(5) Lemmatizing the dataset by removing inflectional endings and unnecessary word beginnings, which leaves only the lemma of the words (i.e. the base or dictionary form) [24].

The experiments were performed on three processed versions of the same data: (1) the raw dataset after removing duplicates (23,936 tweets), (2) normalized data (23,897 tweets) and (3) lemmatized data (23,870 tweets). We report the results from the lemmatized dataset, as it produced the best accuracy.

3.5 Participants

Two religious experts from the armed forces went through the tweets one by one and classified them into three categories as follows:

(1) Religious or political tweets advocating violence and terror (10,793 tweets);
(2) Benign religious or political chatter (10,397 tweets) and
(3) Unrelated topics such as sports, fashion, etc. (2,680 tweets).

The two sets of annotated data had a Cohen's Kappa coefficient of 0.5321 and a disagreement ratio of 29%. Both of these values are consistent with the relevant literature [7]. After interviewing the experts for their opinion, most of the disagreements were shown to result from the interpretation of news reports and their usage in the tweet, individual errors or difference in opinion. A third expert arbitrated and resolved the disagreement to reach a unified dataset with 45% terrorist, 43% non-terrorist and 11% unrelated tweets.

3.6 Expert analysis of tweets

The careful inspection and analysis of the tweets have revealed deep insight into the ISIS chatter with many keywords, hidden meanings, confusing statements and research problems, for reference, see Table 1. In the following, we elaborate these issues in detail along with sample tweets:

Arabic keyword	Description
الولاء والبراء	Loyalty to believers only and disownment of all others
الصحوات	Members of the Iraqi Sunni tribes fighting ISIS. Regarded as apostate
الروافض، المجوس	Shia Muslims. They are considered as disbelievers and need to be killed
البدعه	Anything innovative, in the religious sense, and that did not happen during the days of the prophet Mohammad
الطائفة الممتنعة	Non-practicing Muslims
الردة	Apostates
العملاء الكفرة	Anyone cooperating with the international alliance led by the USA
جنناكم بالذبح	Used when publishing a new video showing killings
صليل الصوارم	The clanking of swords. Used to indicate current fighting
أعوان الظالمين	Those cooperating with the international alliance led by the USA
الطاغوت	All world leaders, except their caliph abu Baker Al-Baghdadi
الصفويون، النصيرية	Syrian Government forces affiliated with Bashar Alassad
الملاحدة	Disbelievers
التحالف الصليبي	The international alliance led by the US
الشرك	Used to describe Yazidis
الخلافة	Caliphate, referring to the ISIS Government
باقية وتتمدد	ISIS staying and expanding
النفير	ISIS call of duty (i.e. fighting)
شفاء الصدور	Payback in terms of beheadings/suicide bombing
اصدار	Press or media release by ISIS
انصار	ISIS supporters
موتوا بغيضكم	Die in your rage. Used to express gloating over ISIS achievements
منهاج النبوة	The path of the prophet
الرايات السود	The ISIS flags
فجر البشائر	The dawn of good news, which is the name of an ISIS mobile app
حكيم الأمة	The wise man of the nation, referring to Al-Qaeda leader Ayman al-Zawahiri
سبيل النجاة	The path to survival (i.e. the ISIS path)
الثبات	Holding fast
ولاية	State, prefecture or province. This term is rarely used in public; however, it is used extensively to describe administrative jurisdictions in ISIS

Table 1.
A list of Arabic keywords in terror-related tweets

(1) The terrorist tweets discuss four conflicts, namely Syria, Iraq, Yemen and Libya, although the majority belongs to the Syrian and Iraqi wars.

(2) There are many sides in the conflicts and certain keywords that may provide indicators to the ideology of the tweeter. For example, Figure 3 shows a tweet about the assassination of a Hezbollah terrorist by an Israeli airstrike. However, the word "'Majoosis" is typically used by ISIS to describe Iranian forces operating in Syria and Iraq. Thus, the tweet is originating from an ISIS sympathizer.

(3) There are political statements that do not indicate terrorism but are related to the conflicts. For example, in Figure 4, the tweet indicates the effect falling oil prices will have on the political future of the Russian and Syrian presidents.

(4) Many statements depend on the context and the original tweet in which they are mentioned. Thus, there is a need to include this in the analysis. For example, in Figure 5, the statement itself is benign but is the description of a terrorist being wise?

(5) The new keyword "'ألنفير" means call for arms/fight. The tweet in Figure 6 appears to be an innocent prayer, but using the "'ألنفير " word completely swayed the statement to indicate terrorism and ISIS recruitment.

(6) The new keyword " منهاج النبوة" means the path of the prophet. The tweet in Figure 7 is a comment about a news article discussing the ISIS ideology, yet the text itself carries few indicators without this keyword.

مقتل المجوسي جهاد مغنية في سوريا بغارةصهيونية

Figure 3.
The assassination of a terrorist, Jehad Maghnyah by an alleged Israeli Strike

RT @BassamJaara: يبدو ان أسعار النفط ستسقط بوتين قبل بشار

Figure 4.
Falling oil prices will lead to the downfall of Putin and Alassad

@tahadulaimi والله والله انك لأعقل العقلاء

Figure 5.
Swearing by Allah three times to describe a tweeter being wise

@mohajera_1415 اللهم يسر لنا الطريق لو تعلمون ان قلوبنا تتمزق للنفير

Figure 6.
Tweet: Praying to Allah to clear the path of any obstacles, if you know our hearts are shredding for the call to fight

أدرك كافر أنها على منهاج النبوه بينما غابت عن كثير من المسلمين!

#دولة_الخلافة_الإسلامية #باقية_وتتمدد https://t.co/LIQ13Ty9M0

Figure 7.
Tweet: The infidel realized that it is on the path of the prophet, but a lot of Muslims did not. Sharing a video with the ISIS and the staying and expanding hashtags

(7) In many cases, the tweets are advocating and celebrating terror events. In Figure 8, the text by itself does not reveal much about the intentions. Thus, there is a need to connect the tweets to other tweets and the linked videos. In addition, the tweet shows repeated letters, emphasizing the importance of preprocessing, for reference, see Section 3.4. Similarly, in regard to Figure 9, which celebrates the burning of the Jordanian pilot captured by ISIS, an important keyword appears in the later tweet, شفاء الصدور, which means healing to our hearts; however, while it is missing the last letter of the first word, the lemma is there.

(8) Some posts contain the Islamic State hashtag, but the subject is totally unrelated to the political issues. We considered such tweets as advocating terrorism. For example, in Figure 10, the tweet is about Saudi club fans and the call to appoint a new coach.

(9) Some users do not refer to the leaders of the religious organizations by name. For example, they refer to Al-Baghdadi, the ISIS leader, as our Shaikh (i.e. a religious leader and not a ruler in the Gulf states), for reference, see Figure 11. Moreover, they refer to the leader of Al-Qaeda, Al-Zawahiri, as the wise man of the nation in another tweet.

(10) In some cases, tweeters manipulate news events in a way to indicate sympathy to terrorism. For example, in Figure 12, the original tweet by *Russia Today* (RT) is modified to include the Arabic word for "'so called" to indicate belittlement of the injured Iraqi officer. Similarly, in Figure 13, the name of the Iraqi prime minister at the time, Al-Maliki, is manipulated to indicate the wish for his demise.

(11) Another interesting aspect that we have noticed is the live monitoring of military airfields by sympathizers in the countries that are part of the International Alliance.

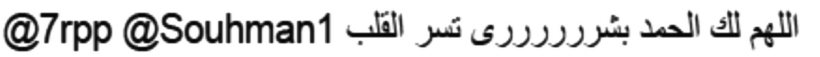


Figure 8.
Tweet: Praise to Allah, good news that makes our hearts happy

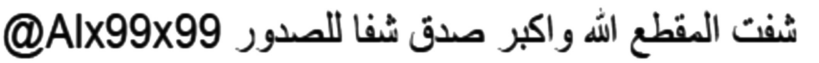


Figure 9.
Tweet: Praise to Allah, good news that makes our hearts happy

Figure 10.
Tweet: Al-Hilal Saudi football club occupies Twitter. A call for Sami Aljaber to be appointed coach

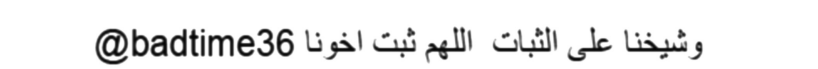


Figure 11.
Tweet: Our Shaikh (i.e. religious leader) is holding fast, praying to Allah to hold our brothers

In Figure 14, the tweet shows a user issuing warnings of a departing fighter jet from Saudi Arabia.

(12) ISIS is actively producing smartphone applications. In Figure 15, the tweet advertises the mobile app, called the dawn of good news.

3.7 Machine learning-based classification

We conducted machine learning classification using KNN, BNB and SVM linear Kernel OAO and OAA classifiers. The supervised learning and classification were performed using four-fold cross-validation. We varied the K value for the KNN from 1 to 20, with 20 achieving the best results. Moreover, we chose the penalty parameter (C), for both SVM-OAO and SVM-OAA, from the set (0.0001, 0.001, 0.1, 1, 10, 100 and 1000). The testing and training time were measured on an HP ProBook 4530s (Windows 7–64 bit, Intel Core i3 2350 M/2.3 GHz, 12 GB DDR3 RAM @ 1333 MHz).

4. Results and discussion

We evaluated the performance using four metrics: precision, recall, F1 score and accuracy. Precision measures the ratio of true positives to all elements identified as positives (including false ones); recall measures the ratio of true positives to all relevant elements (i.e. the actual positives); the F1 score is the harmonic mean of the recall and precision and expresses the accuracy of classification in unbalanced datasets and the accuracy is defined as the ratio of the true positives for all classes to the number of instances (i.e. total tweets in the testing set). The four measures are defined as follows:

RT @IraqiSpringMC: #الرمادي اصابة عقيد الصحوة المدعو (عراك عبيد حمود)
مسؤول صحوة جوبية بمعارك سجارية وانسحاب افراد الصحوة والجيش الى سكة ال...

Figure 12.
Tweet: A news snippet about the injury of an Iraqi officer, and the retreat of an Iraqi officer, and the retreat of the Iraqi Army

RT @mhamdhaif: أي ضربةأمريكيةللموصل هي موجهه لأهل السنة لا لداعش التي
اتخذها نظام الهالكي ومن معه حبل النجاتمن هول صدمةالهزيمةالنكراء لميلي...

Figure 13.
Tweet: American strikes are targeting Sunni Muslims not Daesh. The Iraqi government is using Daesh for religious cleansing

الحذر طائره تخرج من ارض الحرمين لضرب المجاهدين
الحذر الحذر اخوتاه اللهم أحفظ المجاهدين

Figure 14.
Tweet: Live warning to ISIS fighters about the takeoff of fighter jets from a base in Saudi Arabia

أشترك الأن أخبار #الدولة_الإسلامية بجوﷺت الأندرويد#تطبيق_فجر_البشائر
http://t.co/KMGhyeKaq1 فجرالبشائر http://t.co/gfjfdJmMgN

Figure 15.
Tweet: Advertising the ISIS smartphone app

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$F1 = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{3}$$

$$\text{Accuracy} = \frac{\sum_i^3 \text{TP}_i}{\text{No.of Testing Tweets}} \tag{4}$$

where TP is the number of correctly classified tweets (i.e. for each one of the three classes), FP is the number of wrongly classified tweets as another class and FN is the number of tweets missed by the classifier.

Table 2 show that for identifying terrorist tweets, SVM-OAA achieved the highest precision, recall and F1-scores with values of 81.6, 83.6 and 82.6%, respectively. The confusion matrix in Table 3 shows that the overall accuracy was 77.6, 40.8% (317 tweets) of the unrelated tweets were mistakenly classified as non-terrorist (i.e. political in nature but non-terrorist), non-terrorist tweets were misclassified as terrorist in 17.8% of the tested tweets (i.e. 588 tweets) and terrorist tweets were misclassified as non-terrorist in 15% of the tested tweets (i.e. 489 tweets). Thus, the largest source for errors (in terms of percentage) is in classifying unrelated tweets as non-terrorist, which is not a source of concern as we are not trying to distinguish between political and nonpolitical tweets. To evaluate this effect, we retrained the model using two classes with the non-terrorist and unrelated classes merged in one class. The performance measures improved slightly for all of the algorithms, for reference, see Table 4. However, it should be noted that the F1-score of interest relates to the ability of identifying terror tweets as true positives. Thus, retraining the model with the two other classes merged should not have a significant effect on this value. On the other hand, merging the two classes generously improved the accuracy (i.e. correctly identifying each class) of the SVM-OAA algorithm from 77.6 to 82.6%. Consequently, merging the unrelated and non-terrorist classes into one class eliminated the error in distinguishing between them,

Table 2.
The classification precision (P), recall (R) and F1-score for the four classification algorithms using lemmatized data

	KNN			BNB			SVM-OAA			SVM-AOA		
Class/Metric	P	R	F1	P	R	F1	P	R	F1	P	R	F1
N	61.4	54.6	57.8	73.9	67.1	70.3	75.3	77.9	76.6	75.3	77.5	76.3
U	55.7	22	31.6	57.7	74	64.8	69.2	52.3	59.6	71.9	55.2	62.4
T	60.5	75.7	67.2	76.5	78.1	77.3	*81.6*	*83.6*	*82.6*	80.4	82.6	81.5

Note(s): N: Non-Terrorist; U: Unrelated and T: Terrorist. All numbers are percentages

Table 3.
Confusion matrix for the classification of lemmatized data using SVM-OAA

Class	Non-terrorist	Unrelated	Terrorist
Non-terrorist	77.8%	4.4%	17.8%
Unrelated	40.8%	51.7%	7.5%
Terrorist	15%	1.4%	83.6%

Note(s): Overall accuracy was 77.6%

but it does not reflect a greater ability to detect terrorist tweets. Regarding the model overhead, SVM-OAA requires a reasonable amount of training time, which is lower than SVM-OAO and KNN, and the testing time is very close to that of the fastest algorithm, BNB, for reference, see Table 5.

A closer look at the detailed output of the classification model reveals several causes for misclassification in both false positives and false negatives. First, in analyzing reply tweets, it is important to consider the original tweet, as it provides the proper context. For example, if the tweet comments "'good news," then the news being considered determines the classification of the tweet (e.g. a terrorist is captured vs a terrorist act). Similarly, if the tweet offers prayers, then it needs to be considered in context. Second, some news articles contain the same keywords as terrorist tweets (e.g. terrorist nicknames). Moreover, the manipulation of these news articles to include demeaning words may indicate terrorist intent/sympathy (e.g. killed vs perished). Third, local slangs may differ greatly (e.g. Iraqi vs Libyan), which makes the classification model unsuitable if it was trained on formal or other dialects. Thus, regionalized models may provide better detection of terrorist content. Fourth, sarcasm is a style of writing that makes it difficult to perform correct classification. Finally, many tweets contain highly misspelled words due to carelessness, poor education or hastiness. All of these factors complicate the problem and render the detection of radical content an arduous task.

5. Conclusion

SNs have given a global media platform to anyone with a simple device and an Internet connection. Although this can be used for good causes, it can also be used for a greater harm. Terrorist organizations are actively engaging in propaganda and recruitment drives over these media. In the past few years, ISIS has risen to the forefront of the war against terrorism with many directed/inspired attacks all over the world. We have observed a lack of depth and knowledge of the ideology and culture of this organization.

In this paper, we have collected and analyzed thousands of tweets advocating and promoting ISIS. We have identified many common markers and keywords characteristic of ISIS rhetoric. Moreover, we have applied text processing and AI machine learning techniques to classify the tweets into one of three categories: terror-related, non-terror political chatter

Table 4. The classification performance metrics for the four classification algorithms after processing the lemmatized data to merge the Non-Terrorist and Unrelated tweets into one class

	KNN	BNB	SVM-OAA	SVM-OAO
Overall accuracy	76.4%	76.05%	*82.6%*	81.5%
F1-score	69.6%	77.3%	*83.2%*	82.5%
Precision	64.2%	74.7%	*82.2%*	80.7%
Recall	76.6%	80.1%	*84.3%*	83.5%

Note(s): The F1-score, precision and recall metrics used class terrorist as the true positive

Table 5. Testing and training time of the lemmatized data with three classes of tweets

Classifier	Training time (s)	Testing time (s)
KNN	220.7	9.1
BNB	6.8	2.3
SVM-OAA	126.6	2.5
SVM-OAO	2262	27.5

and news and unrelated data-polluting tweets. We have achieved a high accuracy of 78–83% even though the data are mostly political and religious in nature.

The subject is complicated, dynamic and reflective of current events. Nonetheless, the keywords that we have identified are less concerned with specific events and more with the ideology. More research is still required to handle the problems of sarcasm, conversation relationships between tweets, context, local dialects and real-time monitoring.

References

1. Byman DL. What happens when isis goes underground?; 2018. Available at: https://www.brookings.edu/blog/markaz/2018/01/18/what-happens-when-isis-goes-underground/.
2. Statista. Number of social media users worldwide 2010-2021; 2018. [cited 2018 Jun 20]. Available at: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.
3. Bertrand N. Shuttered facebook group that organized anti-clinton, anti-immigrant rallies across Texas was linked to russia; 2017. [cited 2018 Jun 20]. Available at: http://www.businessinsider.com/facebook-group-russia-texas-anti-immigrant-rallies-2017-9.
4. Lucas R. How russia used facebook to organize sets of protesters; 2017. [cited 2018 Jun 20]. Available at: https://www.npr.org/2017/11/01/561427876/how-russia-used-facebook-to-organize-two-sets-of-protesters.
5. Scott M. Cambridge analytica helped 'cheat' brexit vote and us election, claims whistleblower; 2018. Available at: https://www.politico.eu/article/cambridge-analytica-chris-wylie-brexit-trump-britain-data-protection-privacy-facebook/.
6. Kaggle. Datasets – kaggle; 2018. [cited 2018 Jun 20]. Available at: https://www.kaggle.com/datasets.
7. Lara-Cabrera R, Pardo AG, Benouaret K, Faci N, Benslimane D, Camacho D. Measuring the radicalisation risk in social networks. IEEE Access. 2017; 5: 10892-10900.
8. FATF. Financing of recruitment for terrorist purposes; 2018. [cited 2018 Jun 20]. Available at: http://www.fatf-gafi.org/publications/methodsandtrends/documents/financing-recruitment-terrorist-purposes.html.
9. Attestog T, Kukulage SP. Mapping extremist forums using text mining. Master's thesis. Universitetet i Agder/University of Agder; 2013.
10. Scrivens R, Davies G, Frank R, Mei J. Sentiment-based identification of radical authors (sira). Data mining workshop (ICDMW). 2015 IEEE International Conference on. IEEE; 2015. p. 979-986.
11. Figea L, Kaati L, Scrivens R. Measuring online affects in a white supremacy forum. Intelligence and security informatics (ISI). IEEE 2016 Conference on. IEEE; 2016. p. 85-90.
12. Ferrara E. Contagion dynamics of extremist propaganda in social networks. Inf Sci. 2017; 418: 1-12.
13. Kim JJ, Liu Y, Lim WY, Thing VL. An empirical study on collective online behaviors of extremist supporters. International Conference on Advanced Data Mining and Applications. Springer; 2017. p. 445-459.
14. Rios SA, Munoz R. Dark web portal overlapping community detection based on topic models. Proceedings of the ACM SIGKDD workshop on intelligence and security informatics. ACM; 2012, p. 2.
15. Alghamdi HM, Selamat A. Topic detections in arabic dark websites using improved vector space model. Data mining and optimization (DMO). 2012 4th Conference on. IEEE; 2012. p. 6-12.
16. Delavallade T, Bertrand P, Thouvenot V. Extracting future crime indicators from social media. Using open data to detect organized crime threats. Springer; 2017: 167-198.
17. Gialampoukidis I, Kalpakis G, Tsikrika T, Papadopoulos S, Vrochidis S, Kompatsiaris I. Detection of terrorism-related twitter communities using centrality scores. Proceedings of the 2nd International Workshop on Multimedia Forensics and Security. ACM; 2017, p. 21-25.

18. Kelion L. Pastebin: running the site where hackers publicise their at tacks; 2012. [cited 2018 Jun 20]. Available at: https://www.bbc.com/news/technology-17524822.
19. Controlling Section. Ctrlsec - 1 (@ctrlsec1); 2017. [cited 2018 Jun 20]. Available at: https://twitter.com/ctrlsec1?lang=en.
20. Twitter Policies. Rules and policies; 2018. June 365 20, 2018. Available at: https://help.twitter.com/en/rules-and-policies#general-policies.
21. Khasawneh N, Al-Khudair MM, Fraiwan M. On using classification techniques for corpus reduction in arabic text-to-speech systems. Int J Comput Appl. 2011; 33(4): 347-354.
22. Burnap P, Williams ML. Cyber hate speech on twitter: an application of machine classification and statistical modeling for policy and decision making. Pol. Internet. 2015; 7(2): 223-242.
23. Correa D, Sureka A. Solutions to detect and analyze online radicalization: a survey. 2013; 1-30. arXiv:1301.4916v1.
24. El-Beltagy SR, Rafea A. An accuracy-enhanced light stemmer for Arabic text. ACM Trans. Speech Lang. Process. 2010; 7(2): 21-22.

Corresponding author
Mohammad Fraiwan can be contacted at: mafraiwan@just.edu.jo