Detecting Violent Radical Accounts on Twitter

Ahmed I. A. Abd-Elaal[1]
Computer and Systems Engineering
Department Faculty of Engineering
Ain Shams University

Ahmed Z. Badr[2], Hani M. K. Mahdi[3]
Professor of Computer Systems
Faculty of Engineering
Ain Shams University

***Abstract*—In the past few years and as a result of the enormous spreading of social media platforms worldwide, many radical groups tried to invade social media cyber space in order to disseminate their ideologies and destructive plans. This brutal invasion to society daily life style must be resisted as social media networks are interacted with on daily basis. As some violent radical groups such as ISIS has developed well designed propaganda strategies that enables them to recruit more members and supporters all over the world using social media facilities. So it is crucial to find an efficient way to detect the violent-radical accounts in social media networks. In this paper, an intelligent system that autonomously detects ISIS online community in Twitter social media platform is proposed. The proposed system analyzes both linguistic features and behavioral features such as hashtags, mentions and who they follow. The system consists of two main sub-systems, namely the crawling and the inquiring subsystems. The crawling subsystem uses the initially known ISIS-related accounts to establish an ISIS-account detector. The inquiring subsystem aims to detect Pro ISIS-accounts.**

Keywords—Machine learning; ISIS; Daesh; extremism; data mining; social media; Twitter

I. Introduction

Social media become an essential part of everyone’s life style nowadays. Everyone now can easily express his thoughts, feelings and even his emotions through internet and these ideas will spread within seconds among the world. These posts may be viewed by millions who will interact with them. Online Social Network (OSN) has grown significantly over the past decade. Within a population of 7.676 billion human being which is the entire population on earth, there are 4.388 billion Internet users, 3.256 billion smart phones and 3.484 billion active social media users in 2019 [1]. Its existence gave humanity the gift of spreading civilization, literature, science, arts, and others which are means of fulfilling prosperity and welfare all over the world. On the other hand, if this power was misused, many unpleasant sequences may occur. Hatred may be spread instead of constructive ideas, violence instead of literature and science, war instead of welfare.

Many extremist radical hate groups and violent associations are consistently trying to spread their ideologies and hate speech through various social media platforms. In other words, these groups are using the social media facilities to recruit new members and to distribute destructive ideas and plans. It is very important to identify the members of these groups to prevent them from spreading their harmful ideologies, disseminating violence and hatred on social media platforms that may result war and conflicts in peaceful societies. A living example on these radical groups is Islamic State of Iraq and al-Sham “ISIS”, also is known as “Daesh” in Middle East. It is recognized by its adherence to the fundamentalist Salafi faith of Sunni Islam [2]. It gained international fame in early 2014 when it expelled Iraqi government forces from the main cities of Western Iraq then seizing Mosul and committing the massacre of Sinjar. Since its appearance, “ISIS” is continuously trying to leverage its ideology through social media platforms.

Twitter social media is one of the most popular online social media networks in the world with 330 million monthly active users and 500 million tweets per day. A tweet is a message composed of 140 characters that any user can easily share among millions of accounts. Tweets may contain hashtags that highlight the tweet’s main topic. Although there is no accurate statistics shows the existence of ISIS members but in 2014, it was an expected that from 46,000 to 90,000 Twitter accounts that upheld for ISIS or they were controlled by its supporters. Over 1.2 million accounts suspended for terrorist content since August 2015 [3]. Suspension criteria are mainly based on other accounts reporting an account that generates violent radical tweets. However, inspecting the posts of a single user in a social media application such as Twitter is a tedious work. The need for an artificial intelligent approach that mimics the human inspection to solve this problem is a must.

In this paper, a new architecture that can autonomously detect ISIS members’ community in the Twitter social media network is proposed. The proposed system analyzes both linguistic features and behavioral features such as hashtags, mentions and the following lists. These will enable us to continuously have a live image of how ISIS members interact with online social media. The rest of the paper is organized as follows: Section 2 reviews the previous studies and related work. Section 3 clarifies dataset gathering and analysis. Section 4 introduces the proposed architecture. Section 5 discusses the experimental results. Finally, Section 6 includes the conclusions and future work.

II. Related Work

ISIS existence in twitter has been expanded enormously that catch the concerns of the international society. That resulted the appearing of some online groups volunteering their time and resources to report those violent accounts. One of these groups was Ctrl-sec [4] that was responsible of deactivation of about 25,000 twitter accounts in three years by identifying radical accounts manually [5]. Others could suspend about 25,000 Pro-ISIS accounts through crowdsourced reporting [3]. Many researchers were attracted to this topic. They are trying to find new means to discover and limit the vast spread of these accounts. Ashcroft et al. [6] adopted machine learning techniques based on a list of English predefined hashtags related to ISIS to detect ISIS related tweets using three feature classes, namely sentiment, temporal and stylometric features. He found that his approach is highly dependent on data and the approach of detecting radical content should always be a helpful tool to support humans to asset accounts not to replace them. Choudhary et al. [7] tried to detect behavior patterns and key player features to identify terrorist community.

In a trial to understand what happens through the phase of an account being converted from an ordinary account to a Pro-ISIS one sharing Pro-ISIS content, Rowe and Saif [8] defining the Pro-ISIS account by the account that shares more radical Pro-ISIS content than the Anti-ISIS one. Although this approach seems effective but couldn't deal with the lexical diversity.

Klausen et al. [9] studied the communication flow among ISIS members on twitter using 59 manually assessed Pro-ISIS accounts and found that female members has a key role in ISIS propaganda. Carter et al. [10] found that newly bounded ISIS members seek guidance through online spiritual ISIS figures on twitter. Chatfield et al. [11] investigated how ISIS recruits new members through disseminating their ideology on social media platforms. Also Vergani et al. [12] studied how ISIS uses emotional speech and religious quotes to recruit online supporters. Berger et al. [13] found that twitter users who follow ISIS members are highly affected by their ideology. Saif et al. [14] found that semantic features based models out performs other lexical, topic and sentiment based models in detecting Pro-ISIS accounts. Berger et al. [15] found that Pro-ISIS accounts can be identified through their profile description. Agarwal et al. [16] expressed the presence of offensive, war and hate speech terms in ISIS propaganda.

III. Data Collection and Analysis

A. Data Gathering

As the research objective is to automatically detect Arabic speaking violent radical accounts on twitter, dataset should be found and properly cleaned to be analyzed, studied and to apply various machine learning algorithms on it. Although native ISIS members are Arabic speakers as they mainly located in Syria and Iraq, no Arabic ISIS related dataset was found due to the lack of proper Arabic resources in science society especially in machine learning field. In order to prepare a suitable Arabic dataset that can be used in this research, two approaches for collecting data were adopted. The first approach is to collect proper dataset from Arabic speaking twitter accounts that represent Pro-ISIS, Anti-ISIS and non-ISIS. This dataset is referred to hereafter as the collected dataset. The second approach is to collect the published available non-Arabic ISIS related datasets that can be found in online data-science communities and translate them. This dataset is referred to hereafter as the translated dataset.

1) The collected dataset

By studying and examining extremist accounts in Twitter, the most frequent hashtags in ISIS propaganda were manually identified such as وأعدوا, الدولة_الإسلامية, باقية, تتمدد. Using these hashtags, 42 accounts were collected and annotated as the most violent and ISIS influential accounts. Using twitter API [17] the tweets feed of the annotated accounts were collected which resulted to downloading of 21,000 tweets. These accounts will be referenced as "collected Pro-ISIS accounts". Similarly and in order to collect balanced dataset 21,000 Anti-ISIS tweets were gathered in the same way using the following hashtags such as جرائم داعش, داعش تجار الدم, مسلمون ضد داعش, فضائح داعش, بلغ عن داعشas they were the most used hashtags in the most active Anti-ISIS accounts namely "collected Anti-ISIS accounts". 21,000 random non-ISIS related tweets were collected as-well from different domains: "News – Religion – Sports – Art" to represent "collected non-ISIS related accounts". Data pre-processing such as URL and mentions removal, discarding non alpha characters such as (@, #, $, %, _), characters normalization such as (أ, إ, آ, ؤ, ئ) and stop words removal such as (فى, على, أى, من, إلى) , tashkeel removal such as (◌َ, ◌ً, ◌ّ, ◌ٌ, ◌ٍ) and prefix/suffix removal such as (الـ, نى, نا, و) was applied to prepare the collected dataset for training stage.

2) The translated dataset

By searching online data-science communities such as "Kaggle", three non-Arabic ISIS related datasets were found:

a) Fifth-tribe "How ISIS Uses Twitter" dataset [18]. It is consisted of 17,000 tweets was collected from 112 Pro-ISIS twitter accounts from all over the world that supported 2015 terrorist Paris attacks [19]. These tweets are mostly written in English.

b) "Religious Texts Used By ISIS" [20]. 2,685 religious texts dataset which was collected by scrapping 24 issues of Dabiq and Rumiyah English-based ISIS magazines that ISIS uses to spread their ideology in Europe and western world.

c) "Tweets Targeting ISIS" dataset which contained 122,000 ISIS related tweets was collected all over the world in many languages, mostly in English [18]. These tweets were collected by following ISIS related terms such as (ISIS, Daesh, Islamic State, Raqqa, Mosul) 13,000 tweets that were against ISIS ideology and terrorism were translated into Arabic.

Translating into Arabic language was carried out by custom python scripts using Google translating service [21]. Translated dataset was manually reviewed to correct mistranslated words/expressions. Finally, data pre-processing steps were carried out including data cleaning, normalization, stop words removal and stemming as mentioned in the collected dataset subsection.

B. Text Features Vectoring

Dataset was collected from 2-main different sources collected dataset and translated one. These two datasets may suffer from the domain divergence problem [22]. To make sure of suitability of applying any recognition and detection technique for both of these datasets, visual testing was adopted. In the first step to carry out this test, the datasets are first represented as vectors. Then visualization techniques are used to represent the converted words. "Mazajak" word embedding was used [23] to convert dataset corpus into vector domain. Mazajak is considered to be the largest Arabic word embedding models based on a corpus of 250 million tweets converted using skip-gram architecture [24]. In Mazajak, each word token is converted to a 300-D vector. Tweets can be represented by the mean of its contained word embedding's. Similarly the whole user's tweets thread can be represented by the mean of its vector tweets embedding.

The second step in the proposed visual testing is dimensionality reduction of the resulting vectors because it is not proper to visualize 300-D vectors. So embedding vectors dimensions have to be reduced in order to visualize and study the dataset. TSNE [25] is a machine learning technique for dimensionality reduction. This technique is applied here to reduce the vector dimensions from 300-D to just 2-D.

C. Text Features Analytics

After converting the dataset's corpus into a vector form using "Mazajak" word embedding model [23] where the distance between any two vectors is proportional to the difference in the meaning of the words they represent. The vectors should be plotted graphically so they can be studied and analyzed to get better understanding for dataset. That is how the consistency of the dataset can be checked as it was collected from multiple sources and to make sure that vectors that represent the same class can be clustered in spite of the lexical diversity between the collected and translated data.

Fig. 1 illustrates both the collected and the translated tweets. The "collected Pro-ISIS accounts" tweets and both the translated "How ISIS Uses Twitter" tweets and the translated "Religious Texts Used By ISIS." These tweets were labeled as Pro-ISIS in "Red", Anti-ISIS in "Green" and Random in "Blue" colors. From Fig. 1, it can be noticed that some tweets from both Anti-ISIS and Pro-ISIS classes are overlapped with random tweets. This can be easily interpreted because ISIS related tweeters may have other interests in different topics such as sports and news. Also it is noticed that non-ISIS related tweets are distributed over a wide range of fields that belong to different domains such as "news – religion – sports – art".

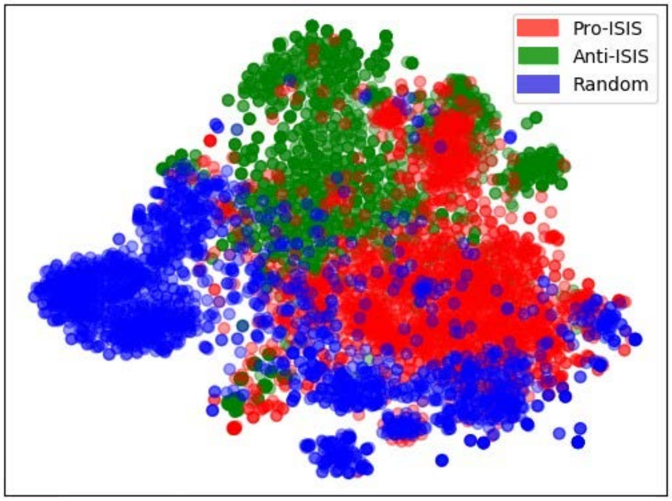

Fig 1. Datasets in 2-D Representation.

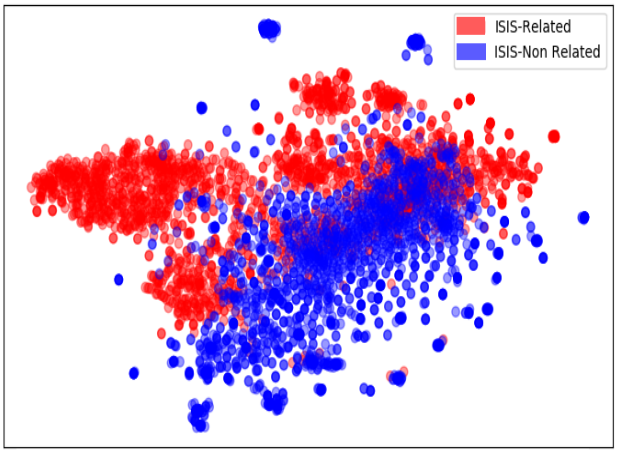

Fig 2. ISIS and Non-ISIS in 2-D Representation.

Non-ISIS related dataset should be collected in order to avoid detection accuracy degradation if the system is not trained to detect non-ISIS related tweets. Fig. 2 represents in "Red" ISIS related tweets from both classes Anti-ISIS and Pro-ISIS tweets, where non-ISIS related tweets are represented in "Blue". It is obvious that ISIS related class is clearly separated from non-ISIS related.

Finally the relation between Pro-ISIS and Anti-ISIS tweets has to be expressed. They may share the same vocabularies, as twitters from both sides (Pro and Anti ISIS) regularly discuss similar subjects in their tweets what will make it harder for us to separate between them. To gather appropriate tweets that represent these two classes, the translated 13,000 tweets from "Tweets Targeting ISIS" dataset and 18,000 manually labeled tweets from collected Anti-ISIS tweets were used to represent the Anti-ISIS data class. On the other side, the translated 12,000 tweets from "How ISIS Uses Twitter" plus 16,200 tweets from collected Pro-ISIS and 2,600 translated texts from "Religious Texts Used By ISIS" were used to represent Pro-ISIS data class. Fig. 3 shows that the two classes can't be separated linearly.

D. Behavioral Features Organization

In addition to the usage of the lexical features of the dataset, it is important to make use of other behavioral features got from the collected data such as mentions, hashtags, likes, retweets and follow lists. These features express the behavior of the user, as the user's used hashtags, retweets and likes defines the topics he is interested in. On the other hand, his mentions and follow list express his spiritual leaders. With the aid of unstructured Mongo database management system [26], the collected mentions, likes, retweets, hashtags and follow lists can be aggregated. Each of these features will have a score that will be identified with its commitment in ISIS community. If a certain hashtag or mention is found more often in ISIS propaganda, it will have higher score which will reflect how much it is related to ISIS. This facilitates capturing the topics they are keen on, how they influence their beliefs and who they follow. Finally a list of the most followed accounts, the most liked and retweeted tweets in addition to the most tweeted hashtags by ISIS supporters will be obtained.

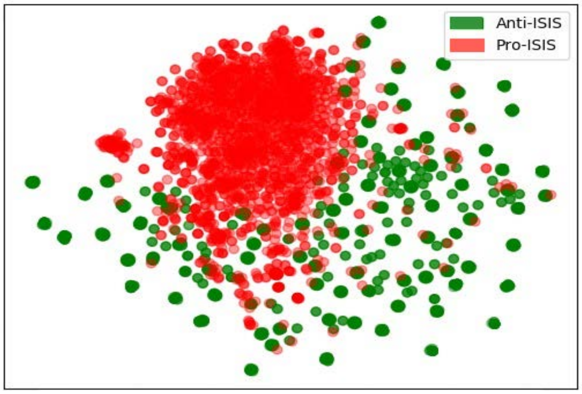

Fig 3. Anti-ISIS, Pro-ISIS in 2-D Representation.

IV. Proposed Architecture

Although Pro-ISIS propaganda is distinguished by the presence of offensive, war and hate speech terms [16], the lexicon itself varies according to the undergoing events [8]. They always use the world trending topics to ensure the wide spread of their propaganda. Other problem is that ISIS propaganda evolves during the process of recruitment itself within many stages that includes religious, emotional and hate speech [7]. These problems cause degradation in the detection accuracy on the long term for non-maintainable detection systems, as they rely on old detection models or outdated lexical dictionaries in the detection process.

The main challenge of the proposed system is to maintain it up to date autonomously. This includes tracking the changes in the online-behavior of ISIS members as well as tracking the evolution of their ideologies and propaganda plans. Moreover, it is essential to identify the fundamental key-players or profound pioneers that originate hate speech content. In a trial to overcome these challenges, the proposed architecture is designed to continuously crawl on ISIS online community updating its corpora and other semantic features such as used hashtags, mentions, who they follow and what they share. That will give us a live image of how ISIS behaves on Twitter social media platform. In order to invade ISIS online community on Twitter, an initial seed of ISIS related accounts will be needed which can be found in the collected dataset. The proposed architecture consists of two main subsystems, namely the crawling subsystem and the inquiring subsystem. The details of these subsystems are explained in the next two subsections. They are depicted separately in Fig. 4 and Fig. 5.

A. The Crawling Subsystem

The Crawling subsystem enables us to invade ISIS online Twitter community. Fig. 4 depicts the subsystem components. The crawling subsystem is started by targeting the most followed accounts from the follow lists in the collected dataset. Using Twitter's REST API [17], the account info, followers list and the last posted 3,000 tweets for each targeted account can be downloaded. The downloaded tweets will undergo data pre-processing steps. The steps include data cleaning, normalization, stop words removal and stemming prior inputting to ISIS-Content detector. ISIS-Content detector will detect ISIS related tweets in order to define for how far the user is involved into ISIS community.

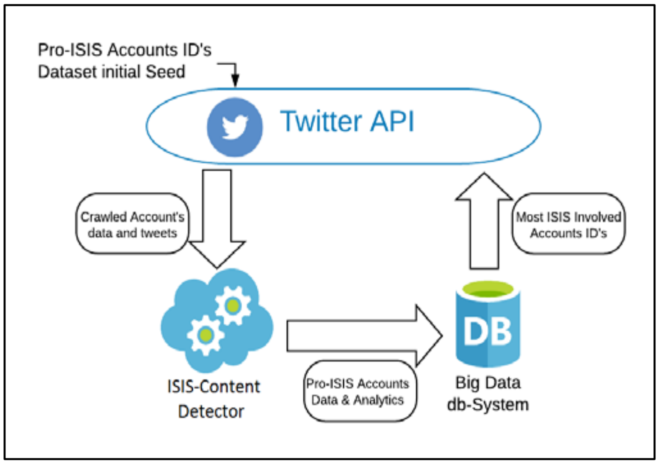
Pro-ISIS Accounts ID's
Dataset initial Seed
Twitter API
Crawled Account's data and tweets
Most ISIS Involved Accounts ID's
ISIS-Content Detector
Pro-ISIS Accounts Data & Analytics
DB
Big Data db-System

Fig 4. The Crawling Subsystem.

As the ISIS-Content detector will assess all of the downloaded tweets for each user, it will distinguish whether each tweet is Pro, Anti or non-related to ISIS in addition to updating hashtags and mention lists in the system's database if the tweet is predicted as Pro-ISIS. If the anticipated Pro-ISIS tweets ratio for the tested account surpassed certain ratio (Pro-ISIS threshold) the account is considered to be Pro-ISIS. Finally all the collected and analyzed data for each account will be stored in the unstructured MongoDB [26]. The collected database can be effortlessly used to investigate and aggregate most followed/active accounts, hashtags and mentions. Updated most followed ISIS accounts will be sent back to be cushioned as a contribution to the following cycle to keep updating the database. As a result of this subsystem, ISIS members, supporters and leaders can be easily tracked in addition to addressing hot topics undergoing in ISIS. Thus the gain of this subsystem is the expansion of the user's knowledge. Likewise an up to date dataset of ISIS-related Arabic labeled tweets will be continuously available that can be used in further studies.

B. The Inquiring Subsystem

Inquiring subsystem enables the user to inspect specific twitter account by twitter ID as an input. Using Twitter API [17], the system will download account data along with the latest posted 3,000 tweets.

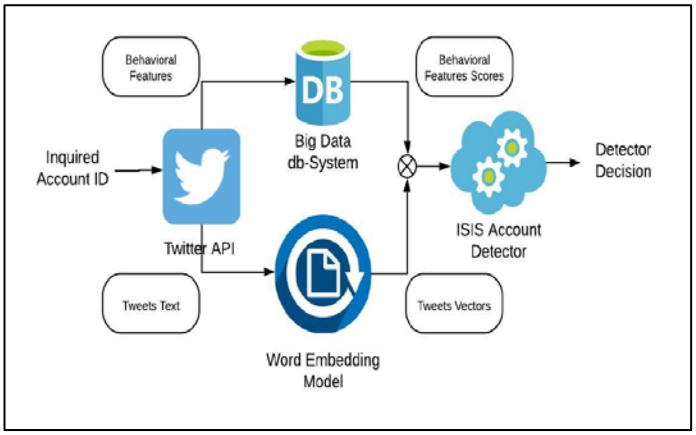

Fig 5. The Inquiring Subsystem.

The downloaded tweets will undergo the same cleaning and data pre-processing steps before usage as in the crawling subsystem. Downloaded hashtags, mentions and following list will be calculated and correlated with the data stored in the system DB which was collected from crawling subsystem to calculate its behavioral features. Behavioral scores will be determined by contrasting the calculated behavioral features by the blacklisted data from crawling subsystem. Behavioral scores along with the pre-processed tweets will be the input to ISIS Accounts Detector which will detect whether the account is Pro-ISIS or not.

C. System Maintenance

In order to maintain the system up to date with high-performance and acceptable detection rates, it must be periodically updated and retrained. It should be introduced to undergoing ISIS related events and its members reactions on them, track their spiritual leaders and key players, learn about their ideologies and propaganda methods. As a result of continues pursue of ISIS community, a live image of how ISIS members spread their news, ideology and even guidance on social media will be obtained. Helpful reports of tracking ISIS members and leaders can be easily developed. Also a proper up to date dataset will be developed which can be used to periodically retrain our detectors on up to date data to avoid deprecation that causes degradation in detection accuracy.

V. Experimental Results

The proposed system includes two pre-trained supervised classification detectors which are key nodes in the system. They detect whether the tweet is Pro-ISIS or not and whether the account is Pro-ISIS or not. Their accuracies define the overall performance of the system as they categorize and define the quality of the crawled data which is important in system evolution. So their accuracy should be boosted and lower down the possible False Positive or False Negative rates.

Experiments were carried out using six different machine learning algorithms namely, Bernoulli Naive Bayes, Decision Tree Classifier, K Neighbors Classifier, Linear Support Vector Classifier, Logistic Regression and Random Forest Classifier [27]. Pre-processed tweets were converted to vectors by multiple word embedding algorithms: “Mazajak” word embedding [23], skip-gram scheme [24] and TF-IDF embedding [28] in [unigram “UG” – bigram “BG”- teragram “TG”] forms. The dataset was then shuffled and spitted into ratios 80% as training dataset and 20% as testing dataset.

A. ISIS-Content Detector

ISIS-Content Detector should process on all of the downloaded tweets downloaded from crawling subsystem that targets ISIS community on Twitter. Although tracking ISIS members in the proposed system, also their tweets should be checked before labeling them where some of ISIS members may have other interests such as news, sports or religion so Pro-ISIS tweets obtained from stalked ISIS members should be inspected in order to increase the quality of the collected dataset.

Other task is to determine how far the stalked account owner is involved into ISIS community, as the ratio of his Pro-ISIS tweets to all of its tweets is calculated. If it exceeds a certain value, the account will be labeled as Pro-ISIS account in the collected database. So a supervised pre-trained detector on labeled data classes that represent Pro-ISIS, Anti-ISIS and non-ISIS related tweets should be prepared. In order to collect these three classes, the translated 13,000 tweets from “Tweets Targeting ISIS” dataset and 18,000 manually labeled tweets from collected Anti-ISIS tweets were used for representing the Anti-ISIS class. For Pro-ISIS data class 12,000 translated tweets from “How ISIS Uses Twitter” plus 16,200 tweets from collected Pro-ISIS and 2,600 translated texts from “Religious Texts Used By ISIS” were used. Finally, the collected 21,000 random tweets were used to represent the non-ISIS related class. Table I shows the results of the testing process.

TABLE I. ISIS-Content Detector's Testing Results

Algorithm	**TF-IDF**			**Skip-gram “Mazajak”**
	UG	BG	TG	
Bernoulli NB	0.88	0.87	0.86	**0.86**
Decision Tree Classifier	0.80	0.80	0.80	0.79
K Neighbors Classifier	0.48	0.51	0.53	0.77
Linear SVC	0.87	0.87	0.89	**0.84**
Logistic Regression	0.85	0.87	0.88	**0.84**
Random Forest Classifier	0.18	0.62	0.64	0.80

Table I shows that TF-IDF embedding outperforms skip-gram embedding in all experiments. Linear SVC achieved accuracy 89% followed by Bernulli NB in “UG” TF-IDF [27] and Logistic Regression [27] in “TG” TF-IDF achieved accuracy 88% regarding F1-score [27]. Detector’s accuracy can be boosted by using all the three detectors as a voting detector as an ensemble learning system [29]. So each tweet will be labeled as the majority of the three detectors decide which will avoid any weaknesses found in any individual one of them in addition to avoiding over fitting.

B. ISIS-Account Detector

ISIS-Account Detector is a key node in the proposed system. It should check if the account inquired by the system user is Pro-ISIS or not. Unlike ISIS-Content detector that operates on discrete tweets, Pro-ISIS accounts detector run on all account data including tweets as textual features, in addition to hashtags, mentions, likes, retweets and following list as behavioral features. Thus beside vectoring dataset corpus by word embedding techniques, other behavioral features such as used hashtags, mentions and following list were vectored by comparing them by the collected data from the crawling subsystem and assigning scores to them.

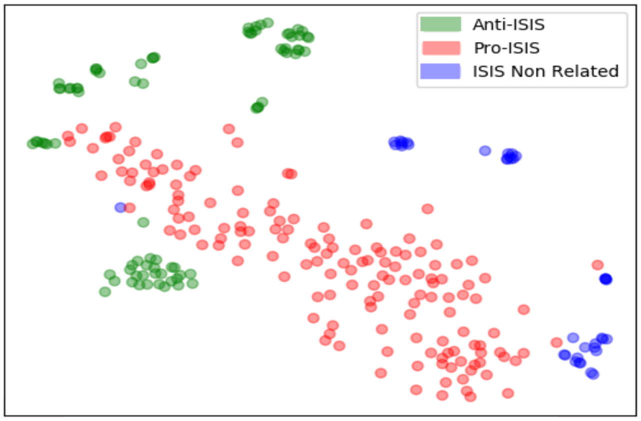

Fig 6. ISIS, Non-ISIS in 2-D Representation.

TABLE II. PRO-ISIS ACCOUNT DETECTOR'S TESTING RESULTS

Algorithm	**TF-IDF**			**Skip-gram “Mazajak”**
	UG	BG	TG	
Bernoulli NB	0.71	0.58	0.55	0.86
Decision Tree Classifier	0.81	0.81	0.83	**0.89**
K Neighbors Classifier	0.68	0.67	0.70	0.87
Linear SVC	0.62	0.62	0.62	**0.94**
Logistic Regression	0.56	0.59	0.62	**0.94**
Random Forest Classifier	0.57	0.45	0.48	0.80

In order to train this detector, two data classes that properly represent Pro-ISIS accounts and non-Pro-ISIS were created. The translated “How ISIS Uses Twitter” dataset in addition to the collected Pro-ISIS accounts were used to represent Pro-ISIS data class. On the other side the collected Anti-ISIS accounts plus collected non-ISIS related accounts were used to represent non-ISIS data class. Non-ISIS related accounts were introduced to the detector to avoid detection accuracy degradation if a non-ISIS related account was inquired in the system. Fig. 6 expresses the scattering of the accounts in the dataset. However Pro-ISIS and Anti-ISIS accounts may share the same lexicons as they are both involved into the same topics, Fig. 6 shows that Pro-ISIS accounts can easily be separated from both Anti-ISIS and non-ISIS-related accounts.

Table II shows that this time skip-gram embedding [24] outperforms TF-IDF embedding. Both Linear SVC and Logistic Regression [27] achieved F1-score 94% followed by Decision Tree [27] by score 89%. Such as “ISIS-Content Detector”, accuracy can be boosted and avoid over fitting by using all three in a voting ensemble learning system [29].

VI. CONCLUSIONS AND FUTURE WORK

A new intelligent system architecture that autonomously detects Pro ISIS-accounts in twitter social media platform is introduced. The system consists of two main sub-systems, namely the crawling and the inquiring subsystems. The kernels of the two subsystems are intelligent detectors. The attributes of these detectors are both linguistic features and behavioral features.

For proper testing the proposed system, a new collected dataset of Pro-ISIS, Anti-ISIS and non-ISIS-related accounts were gathered. Three English datasets about ISIS were translated to Arabic. All datasets were represented as vectors using “Mazajak” word embedding “skip-gram” scheme to 300-D vectors. Vectors dimensions were reduced into 2-D and plotted to check their consistency.

The intelligent detectors kernels for the crawling and the inquiring subsystems were developed using supervised machine learning techniques. The results show that ISIS-Content Detector gave best accuracy 89% according to f1-score by linear SVM algorithm using TF-IDF embedding. They show also that ISIS-Account Detector gave pest accuracy 94% according to f1-score by linear SVM algorithm Skip-gram embedding.

As next steps, extension of the proposed architecture to other social media applications such as Facebook and Instagram is planned. Enlarge the used dataset to include different Arabic delegates.

REFERENCES

[1] S. Akar, Ezgi and Mardikyan, “Analyzing factors affecting users’ behavior intention to use social media: Twitter case,” Int. J. Bus. Soc. Sci., vol. 5, no. 11, 2014.

[2] A. Chaliand, Gerard and Blin, The history of terrorism: From antiquity to ISIS. Univ of California Press, 2016.

[3] K. M. Benigni, Matthew C and Joseph, Kenneth and Carley, “Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter,” PLoS One, vol. 12, no. 12, p. e0181405, 2017.

[4] H. Fernandez, Miriam and Alani, Contextual semantics for radicalisation detection on Twitter. CEUR, 2018.

[5] E. Ferrara, “Contagion dynamics of extremist propaganda in social networks,” Inf. Sci. (Ny)., vol. 418, pp. 1–12, 2017.

[6] N. Ashcroft, Michael and Fisher, Ali and Kaati, Lisa and Omer, Enghin and Prucha, “Detecting jihadist messages on twitter,” in 2015 European Intelligence and Security Informatics Conference, 2015, pp. 161–164.

[7] U. Choudhary, Pankaj and Singh, “A survey on social network analysis for counter-terrorism,” Int. J. Comput. Appl., vol. 112, no. 9, pp. 24–29, 2015.

[8] H. Rowe, Matthew and Saif, “Mining pro-ISIS radicalisation signals from social media users,” in tenth international AAAI conference on web and social media, 2016.

[9] J. Klausen, “Tweeting the Jihad: Social media networks of Western foreign fighters in Syria and Iraq,” Stud. Confl. Terror., vol. 38, no. 1, pp. 1–22, 2015.

[10] P. R. Carter, Joseph A and Maher, Shiraz and Neumann, Greenbirds: Measuring importance and influence in Syrian foreign fighter networks. Citeseer, 2014.

[11] U. Chatfield, Akemi Takeoka and Reddick, Christopher G and Brajawidagda, “Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided twitter networks,” in Proceedings of the 16th Annual International Conference on Digital Government Research, 2015, pp. 239–249.

[12] A.-M. Vergani, Matteo and Bliuc, “The evolution of the ISIS’language: a quantitative analysis of the language of the first year of Dabiq magazine,” Sicurezza, Terror. e Soc., vol. 2, pp. 7–20, 2015.

[13] B. Berger, John M and Strathearn, Who Matters Online: Measuring Influence, Evaluating Content and Countering Violent Exremism in Online Social Networks. International Centre for the Study of Radicalisation and Political Violence, 2013.

[14] H. Saif, Hassan and Dickinson, Thomas and Kastler, Leon and Fernandez, Miriam and Alani, “A semantic graph-based approach for radicalisation detection on social media,” in European semantic web conference, 2017, pp. 571–587.

[15] J. Berger, Jonathon M and Morgan, “The ISIS Twitter Census: Defining and describing the population of ISIS supporters on Twitter,” Brookings Proj. US relations with Islam. world, vol. 3, no. 20, pp. 1–4, 2015.

[16] A. Agarwal, Swati and Sureka, “Using knn and svm based one-class classifier for detecting online radicalization on twitter,” in International Conference on Distributed Computing and Internet Technology, 2015, pp. 431–442.

[17] H. Kumar, Shamanth and Morstatter, Fred and Liu, Twitter data analytics. Springer, 2014.

[18] Fifthtribe, “How ISIS Uses Twitter,” kaggle, 2019. [Online]. Available: https://www.kaggle.com/ activegalaxy/isis-related-tweets. [Accessed: 20-June-2020].

[19] R. K. Cragin, “The November 2015 Paris attacks: the impact of foreign fighter returnees,” Orbis, vol. 61, no. 2, pp. 217–226, 2017.

[20] S. Fuhriman, Christopher and Medina, Richard M and Brewer, “Introducing a Dataset of Multi-Scale Geographies of ISIS Ideology from ISIS Sources,” Terror. Polit. Violence, pp. 1–18, 2020.

[21] H. Somers, “Example-based machine translation,” Mach. Transl., vol. 14, no. 2, pp. 113–157, 1999.

[22] W. Wang, Mei and Deng, “Deep visual domain adaptation: A survey,” Neurocomputing, vol. 312, no. 135–153, 2018.

[23] W. Farha, Ibrahim Abu and Magdy, “Mazajak: An online Arabic sentiment analyser,” in Proceedings of the Fourth Arabic Natural Language Processing Workshop, 2019, pp. 192–198.

[24] A. Neelakantan, Arvind and Shankar, Jeevan and Passos, Alexandre and McCallum, “Efficient non-parametric estimation of multiple embeddings per word in vector space,” arXiv Prepr. arXiv1504.06654, 2015.

[25] G. Maaten, Laurens van der and Hinton, “Visualizing data using t-SNE,” J. Mach. Learn. Res., vol. 9, no. Nov, pp. 2579–2605, 2008.

[26] K. Chodorow, MongoDB: the definitive guide: powerful and scalable data storage. “O’Reilly Media, Inc.,” 2013.

[27] K. Khan, Aurangzeb and Baharudin, Baharum and Lee, Lam Hong and Khan, “A review of machine learning algorithms for text-documents classification,” J. Adv. Inf. Technol., vol. 1, no. 1, pp. 4–20, 2010.

[28] R. Qaiser, Shahzad and Ali, “Text mining: use of TF-IDF to examine the relevance of words to documents,” Int. J. Comput. Appl., vol. 181, no. 1, pp. 25–29, 2018.

[29] T. G. Dietterich, “Ensemble methods in machine learning,” in International workshop on multiple classifier systems, 2000, pp. 1–15.