SSRG International Journal of Computer Science and Engineering
ISSN: 2348 – 8387 /doi:10.14445/23488387/IJCSE-V7I12P103

Volume 7 Issue 12, 9-15, December 2020

Terrorism Detection Model using Naive Bayes Classifier

Francisca Onaolapo Oladipo[1], Ogunsanya Funmilayo Blessing[2], Ezendu Ariwa[3]

[1]*Professor, Computer Science Department, Faculty of Science, Federal University Lokoja, Nigeria*
[2]*Student, Computer Science Department, Faculty of Science, Federal University Lokoja, Nigeria*
[3]*Professor, University of Bedfordshire, Luton, United Kingdom*

Abstract *— The advancement in microblogging has brought an increasing area of interest in sentiment analysis. Terrorist groups have been involved in using social media sites like YouTube, Facebook, and Twitter to propagate their ideology and recruitment of individuals. This work aims to propose a terrorism-related content analysis framework focusing on classifying tweets into terrorist and non-terrorist classes. Based on user-generated social media posts on Twitter, we developed a tweet classification system using supervised learning-based sentiment analysis techniques to classify the tweets as terrorist or non-terrorist. Our results indicate that an automated approach to aid analysts in detecting terrorism content on social media is a promising way forward.*

Keywords — *Classification, Naïve bayes, classification, text Mining, terrorism.*

I. INTRODUCTION

The sentiment is a thought, judgment, or an attitude usually prompted by a feeling. Consequently, sentiment analysis studies the sentiment of people towards certain entities [1]. The process generally determines the attitude of a speaker or a writer concerning some topic. Given the level of subscription to internet opinions, it is one of the resourceful places concerning sentiment information. Users can post their own content and share their thoughts/sentiments online through various social media platforms [2].

According to the survey, about 320 million users per minute on Twitter, about six thousand tweets per second, and over 50 million tweets in one complete day. Due to the high reachability and popularity of social media websites like Twitter worldwide, some users are abusing it by spreading distorted beliefs and negative influence on other users. These include religion, politics, terrorism, fraudsters, and others. Based on research, in 2015, about 12500 users were involved with terrorism on Twitter, leading to the deletion of their accounts [3].

Many hate promoting groups use Twitter to advance their ideology by spreading extremist content among their viewers. Terrorist groups like the Islamic State of Iraq and Syria (ISIS) have exploited Twitter to spread their propaganda and to recruit new members using propaganda, which usually includes virtual messages, presentations, audio and video files that contain explanations, justifications, and/or promotion of terrorist activities [4].

Due to the large volume of content on the internet today, humans manually searching for terrorism-related content is very impractical. This research work describes an application of machine learning to terrorism detection based on tweets extracted from users. The research study centers on building a model that can sense and classify a tweet containing terrorist tweets. The model's accuracy will be evaluated for the automatic classification of tweets into terrorist or non-terrorist through the development and deployment of a web application.

The rest of this paper is organized as follows: The next section presents a review of related research, followed by a description of the methodologies and tools deployed in the research. A discussion of the result is presented in the next section, and the last section concludes the paper.

II. REVIEW OF RELATED WORK

Terrorism is not a new phenomenon and has existed since before the dawn of recorded history [5]. Terrorism is traced back to ancient times, from the Sicarri Horsely (1979) to the Assassins [6]. Later terrorism took on more sophisticated methodologies from gunpowder plot in the 16th century, through nationalist and politically motivated groups of the 20th century, on to the religiously motivated groups of the present 21st century. Terrorism is generally seen as a violent action utilized by the extremist group for criminal or political reasons. According to the FBI [7], terrorism has the following characteristics;

i) Influencing policy of a government by intimidation

ii) To involve violent acts dangerous to human life that violate Federal law.

Since 2011, members of these terrorist groups have issued out media strategies, including promotional hashtags, which help encourage media mujahideen. A guide describing how to use social media platforms and a list of recommended accounts to follow were also released in various forums, including the notorious “The Twitter Guide,” which provided a guide to Twitter usage and outlined its purpose to promote their ideology. ISIS’s social media networks have continued to expand despite

the efforts to disable them [8]. In August 2014, Twitter administrators shut down some ISIS-associated accounts, ISIS recreated and publicized new accounts the following day. ISIS uses dispersed forms of network organization and strategy to send content to users. This interconnected network reconfigures itself from time to time [9]. This fact has made ISIS a challenge for a traditional hierarchical organization to tackle.

Recent studies [10] have proven that with Twitter, it is possible to get people's insight from their profiles in contrast to traditional ways of obtaining information about perceptions. Thus, the need for an approach, such as sentiment analysis to study the behavior of information gotten from social media. It categorizes social media opinions to predict the user interest or behavior, like in our research towards terrorism sentiment.

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that allows the system to learn from observational data, and models are then developed from this data. Machine learning focuses on developing computer programs that can access data and use it to learn from themselves. The technology can also be described as applying computer algorithms (ML Algorithms) to develop models (computer programs) to learn from experience with existing data relating to a task and performance measure, and thereby its performance at a given task increases with experience [11].

Over the years, and with the growth in the use of social media since it reflects people's opinion, experience, and feelings [12]; there has been an increasing interest in the area of sentiment analysis, which involves applying natural language processing methods and determining subjective information in texts as well as extract, identify evaluate and classify online sentiments [13]. Sentiment analysis classifies text into positive or negative or neutral and is majorly concerned with textual information, usually in two forms; opinions and facts. Sentiment analysis is often referred to as subjectivity analysis, opinion mining, and appraisal extraction [12].

One of the most common approaches to solving problems of sentiment analysis is the machine learning technique. [14] performed a sentiment analysis using three machine learning approaches consisting of Naïve Bayes, Maximum Entropy, and Support Vector Machine (SVM) classifiers. They classified movie reviews as positive or negative and performed a comparison between the methods. In their work, Naïve Bayes performs well on large feature spaces. Maximum entropy gave a better result than Naïve Bayes when experimented with large feature space, while the SVM performed very well on large feature space. Similar research by [15] in English and Arabic using the SVM classifier on the movie review dataset produced over 95% accuracy.

Using Lexicon based sentiment analysis on two sets of seed words, [16] created a positive and negative polarity. The sum of polarities of the sentiment words were classified in the document, and the result yielded 71% for the positive categorization of the document while 62% for the negative.

In an approach to addressing inappropriate messages, [17] presents a system to detect inappropriate messages in online social networks through a soft text classifier approach using various ML algorithms. The researchers manually classified messages and labeled each comment as profanity and insult. The model's architecture consists of the training segment and testing segment linked together by a classifier to produce a labeled output.

Research on using terrorism-related tweets to predict support for ISIS groups was conducted by [18]. The authors used Twitter data to study the antecedent of the ISIS support, classified tweets and predicted future support for ISIS. However, this research cannot still interpret the true intentions of the Twitter account holder. Similar research by the same author used ISIS-related tweets to predict future support for the ISIS groups. The bag-of-word features that included individual terms, hashtags and user mention were used to train the SVM classifier. At the end of the experiment, they could predict the future support and opposition of an ISIS user with an accuracy of 70% [19].

Researchers [20] developed a classification system for radical tweets. They built dictionaries by selecting tweets with hashtags like "Al-Qaeda," "Jihad," "terrorism," and "extremism," and a collection of other words relevant for their research. Extracted words were classified under five categories: war, terrorism, operations, extremism, jihads, country, and Al-Qaeda using security dictionaries of enriched themes categorized by semantically related words. The research further vectorized using binary digits, documents based on the presence of security related keywords and adopted a classification rule that uses the presence of one or more words relevant to the category as the final grouping criteria. The classification model obtained an accuracy of 90%, indicating that keywords play an important role in tweet sentiment classification.

Research to detect Jihadist messages on Twitter was conducted by [21]. The work was based on an experiment using three different class features; stylometric, time-based, and sentiment-based. The stylometric features contained the most frequent words in the database; the time-based features contain information about when the tweets were posted. The sentiment-based feature determines the attitude of text in the dataset. Three classification algorithms, SVM, Naïve Bayes, and Adaboost, were applied on the Weka dataset of 4000-7500 tweets giving an accuracy obtained of 98.5%, 96.8%, and 99.5%, respectively.

A machine learning model to detect Twitter accounts supporting jihadist groups and disseminating propaganda content was trained by [23]. To train the model, the researchers used data-dependent features such as most

common hashtags, word bigrams, and most frequent words and data-independent features such as frequency of word length, letters, digits, and emotion words. They showed that their model has significant accuracy for English tweets, while for Arabic tweets, the performance was worse.

The various literature reviewed in this section has essentially described various works that apply machine learning processes and its techniques to solve terrorism analysis content and sentiments. The review has also established that Machine learning algorithms like the Naïve Bayes and Support Vector Machines (SVM) have played a major role in detecting negative emotions in users' written outputs. Our approach is to improve the accuracy of these models through re-engineering.

.

III. METHODOLOGY

The Cross-Industry Standard Process for Data Mining (CRISP-DM) was adopted as the methodology because of its sequential and iterative approach to problem-solving in applying data science and machine learning algorithms (Fig. 1). The dataset consists of tweets extracted from the tweeter in real-time, and to the concerns of fairness, they were not geotagged, and demographics features were excluded. The research relied heavily on techniques of "Natural Language Processing" in extracting significant patterns and features from the large data set of tweets and on "Machine Learning" techniques for accurately classifying individual unlabelled data samples (tweets) into terrorism tweets or not. The research methods are divided into three main parts: data collection, data pre-processing, and classification.

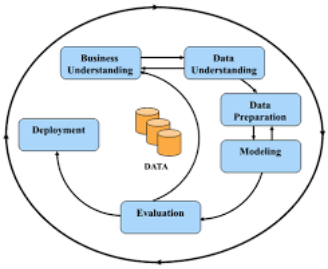

Fig. 1 Cross Industry Process for Data Mining Methodology

The methodology generally involved in sentiment analysis on Twitter follows the essential data collection procedures from Twitter, Data pre-processing, and Sentiment classification. However, this will not still provide an efficient way of differentiating terrorist and non-terrorist tweets. To improve this approach and the accuracy of the classification, our approach consists of four modules; data gathering, pre-processing, labeling and feature extraction, and classification. The terrorism detection model was developed using a Twitter dataset, and the dataset consists of the tweet id, the tweet, and the label. We considered tweets posted by users in the form of hashtags to express their opinions on terrorism, stored the collected tweets in a database, and then pre-processed the dataset, after which we labeled the data. The labeled dataset was spilled into the training, and the test set and the classifiers were applied to the training set to build a classification model (Fig. 2).

In developing the model, the training data fit into the machine for training and the test data is used to test the model accuracy. The Navies Bayes classifiers were used in developing the model. The confusion matrix (Fig. 3) was also computed to describe the classification model's performance on the test data (Fig. 4).

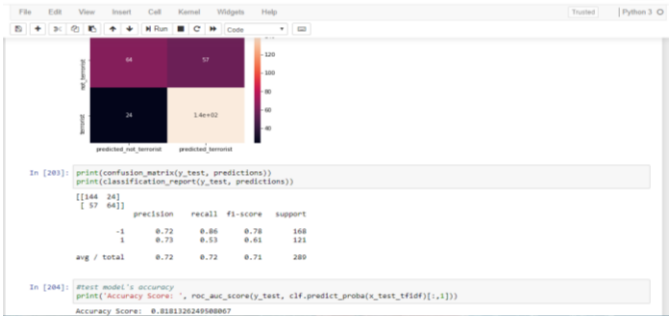

Fig. 2 Confusion Matrix and Accuracy of Naïve Bayes

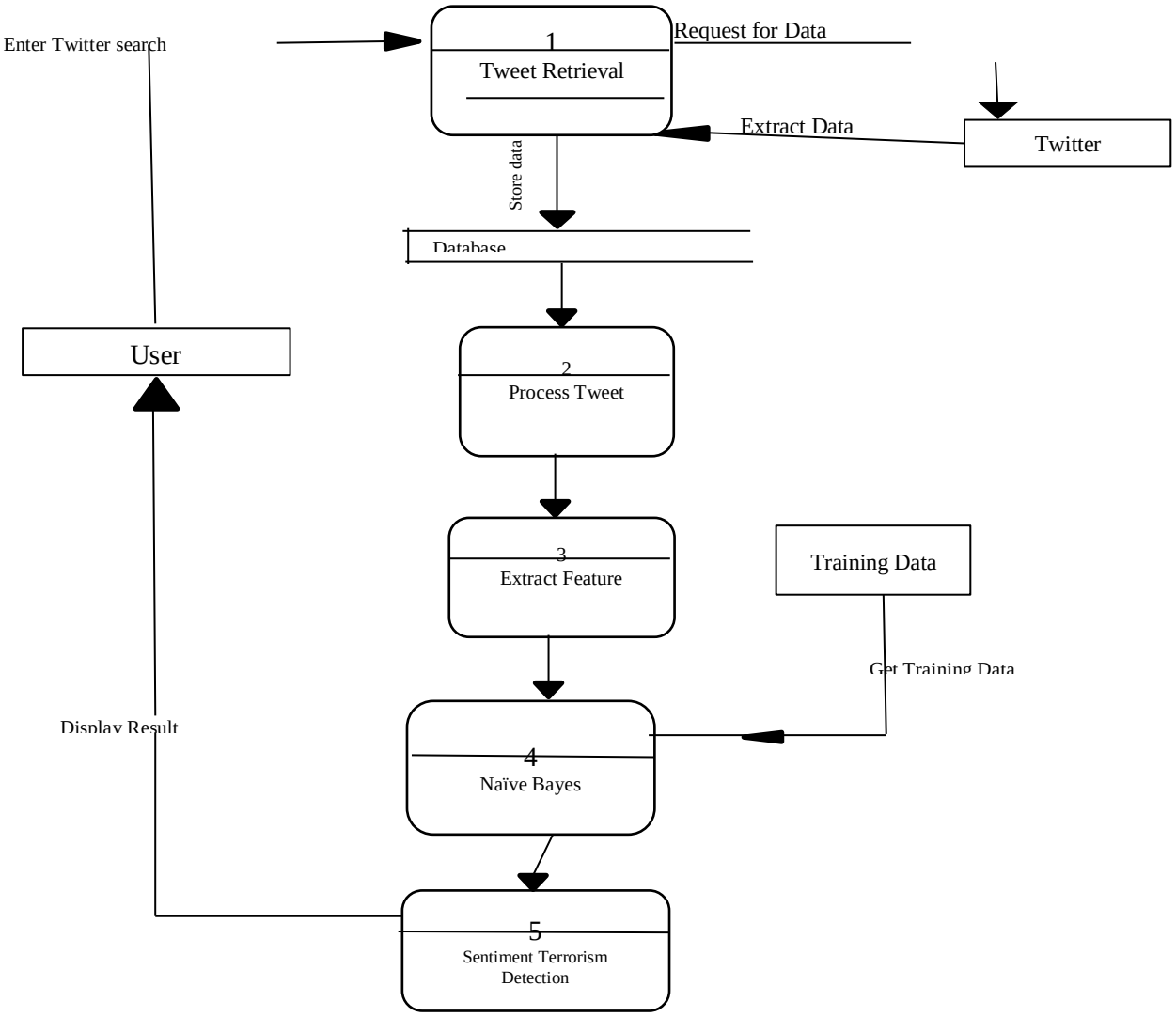

Fig. 3 Detection and classification Methodology

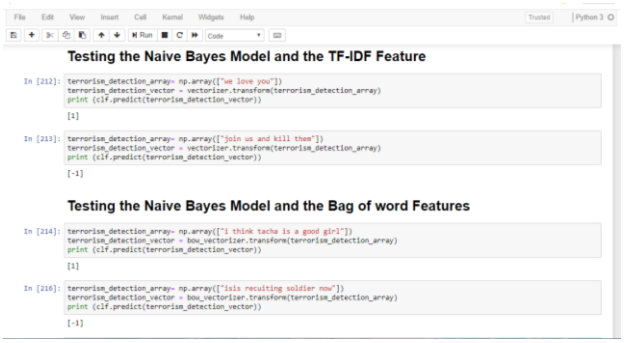

Fig. 4 Testing the Naïve Bayes Classifier

IV. RESULT AND DISCUSSION

The high-level abstraction for the analysis and classification of tweets based on supervised machine learning has been developed in this work. The architecture showing the various components is shown in Fig. 5.

The logical specification of the model components and data, their interaction, relationships, and organization, and the interaction of the processes are showed in a Use Case Description (Fig. 6), the activity diagram (Fig. 7), and a graphical algorithm (Fig.8).

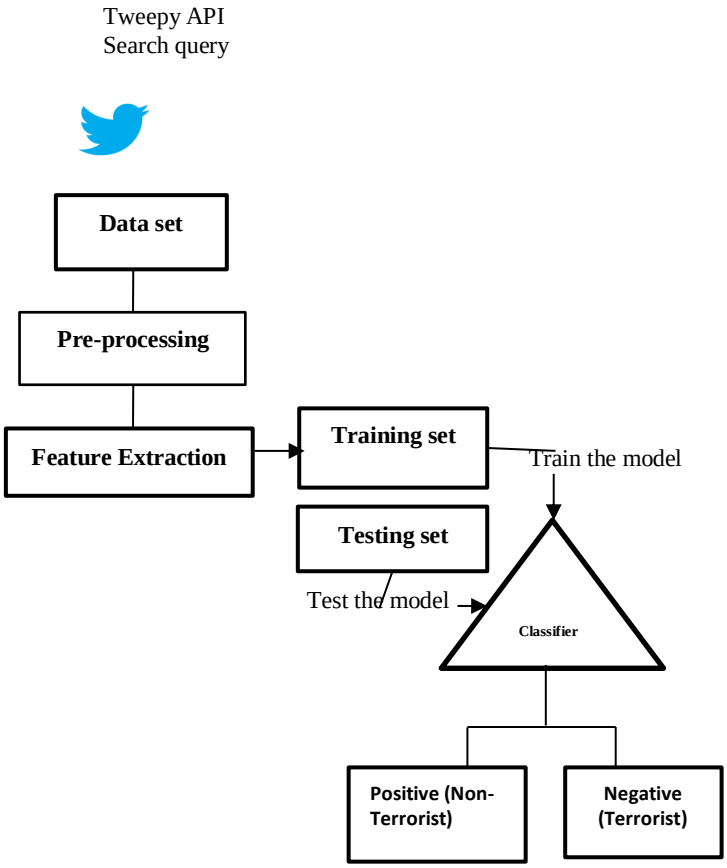

Fig. 5 System Architecture

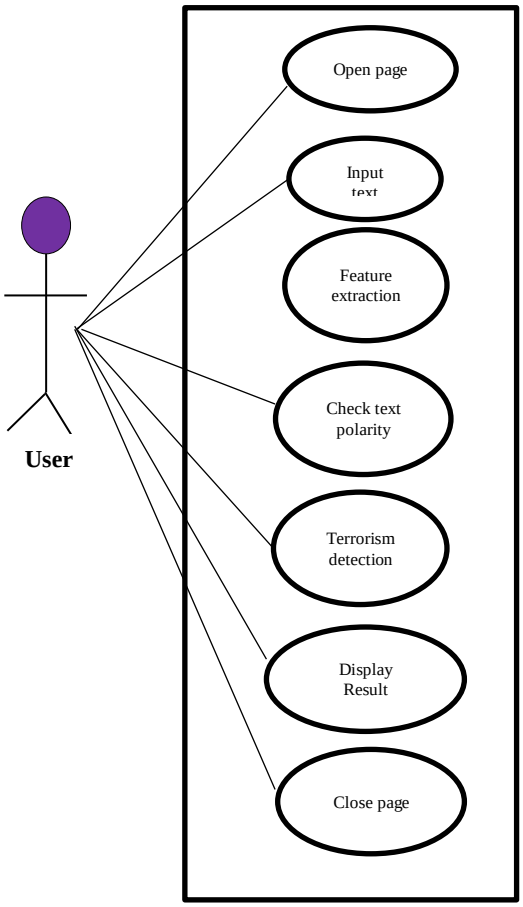

Fig. 6 Use Case Description

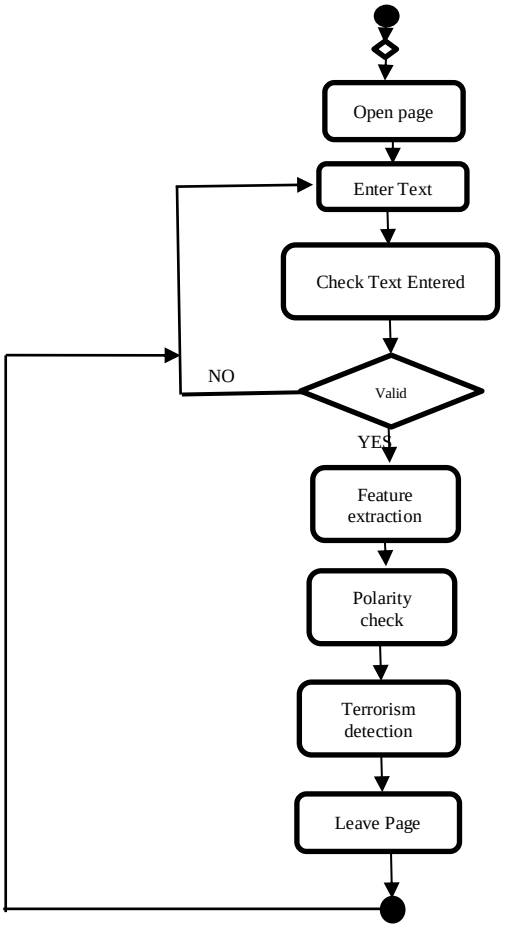

Fig. 7 Activity Diagram

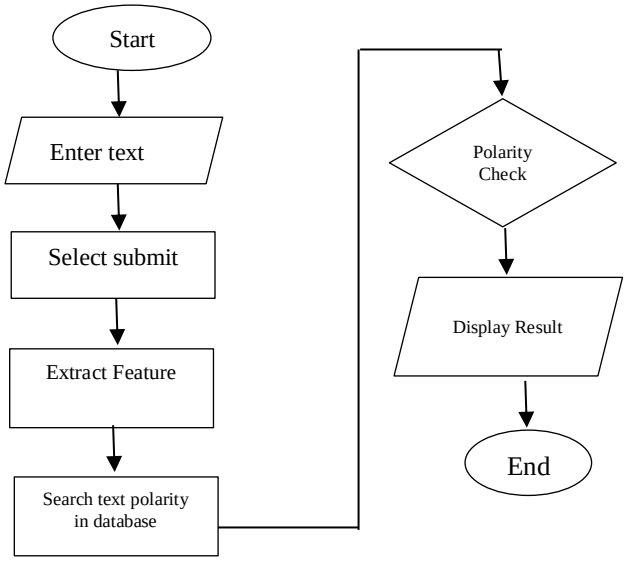

Fig. 8 Graphical representation of the System's Algorithm

VI. CONCLUSIONS

The current sentimental analysis system cannot detect terrorist tweets because some tweets classified as terrorist tweets may be jokes, humor, and sometimes the tweets might be against terrorism. Still, they are detected as terrorist tweets. Our findings revealed that the existing technique of the Support Vector Machine (SVM) could not detect a suspected tweet as a scam, joke, or anti-terrorist tweet. However, by separating such tweets from real terrorist tweets, accuracy can be improved. In this work, we have presented a machine learning approach to classify tweets as containing terrorist content or not. We run our experiments using a Naive Bayes classifier. The study is centered on the detection of terrorism using keywords/hashtags for the sentiment analysis. Due to Twitter platforms' dynamic nature, identifying contents, locating users, and predicting events by keyword-based search is overwhelmingly impractical. The volume of content being posted on social media platforms makes it challenging to discover such content by using hashtags. The sentiment analysis does not consider tweets posted in Non-English Language.

REFERENCES

[1] Lui, B. Sentiment analysis and subjectivity. Handbook of Natural language processing. (2010).
[2] Kim, S.-M., & Hovy, E. Determining the sentiment of opinions. proceedings of the 20th international conference on computational linguistics, 1367, (2004).
[3] Yardon D. (2016). Twitter deletes 125,000 Isis accounts and expands anti-terror teams. The Guardian International Edition. A report published Fri, February 5, (2016). 20:18GMT https://www.theguardian.com/technology/2016/feb/05/twitter-deletes-isis-accounts-terrorism-online
[4] Agarwal, S., & Sureka, A. A focused crawler for mining hate and extremism promoting videos on youtube. Proceedings of the 25th ACM conference on Hypertext and Social media, (2014) 294–296.
[5] Merari, A., & Friedland, N. Social psychological aspects of political terrorism in visualization. IEEE Symposium on Visual Analytics Science and Technology. (2007).
[6] Sloan, S., & Anderson, S. Historical dictionary of terrorism. scarecrow Press. (2009).
[7] FBI. Retrieved March 10, 2019, from http://www.fbi.gov/ (2015).
[8] Fisher, A. N. The call up: The roots of a resilient and persistent jihadist presence on Twitter. (2004).
[9] Fisher, A. Last gang in town. How Jihadist network maintain a persistent presence online, in the perspective of terrorism. (2015).
[10] Miranda Filho, R., Almeida, J., & Pappa, G. Twitter population sample bias and its impact on predictive outcomes: A case study on elections. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Paris, France, (2015) 1254–1261.
[11] Paulo, A., Donald, C., & Ivens, P. The use of Machine Learning Algorithms in Recommender Systems: A systematic review. Expert Systems With Applications, 97. (2015).
[12] Pang, B., & Lee, L. Opinion mining and sentiment analysis. Foundations and Trends in Information extraction, 2, (2008) 1-135
[13] Vishal, A., & Sonawane. Sentiment analysis of Twitter data: A survey of techniques. International journal of computer application, 139. (2016).
[14] Pang, B., Lee, L., & Vaithyanathan, S. Thumbs up?: Sentiment classification using machine learning techniques. Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 10, (2012) 79-86.
[15] Abbasi, A., Chen, H., & Salem, A. Sentiment analysis in multiple

languages: Feature selection for opinion classification in web forums. In ACM Transactions on Information Systems, 26(3), (2008) 1-34.

[16] Harb, A., Planti, M., Dray, G., Roche, M., Trousset, O., & Poncelet, P. Web opinion mining: how to extract opinions from blogs?. Proceedings of the 5th international conference on soft computing as transdisciplinary science and technology. Cergy-Pontoise, France. (2008).

[17] Shetty, R., Nair, K., Singh, S., Nakhare, S., & Upadhye, G. A System to Detect inappropriate Messages in Online Social Networks. International Journal of Advanced Computational Engineering and Networking, 3(3), (2015) 40-43.

[18] Walid, W. I. #failedrevolutions: Using Twitter to study the antecedents of isis support., arXiv preprint arXiv:1503.02401. (2015).

[19] Walid, M., Kareem, D., & Ingmar, W. (2015). #FailedRevolutions: Using Twitter to study the antecedents of ISIS support. doi:10.5210/fm.v21i2.6372

[20] Pooja, W., & Bhatia, M. Classification of Radical Message on Twitter using Security Association. In Case of Studies in secure computing: Achievements and Trends, 273. (2016).

[21] Ashcroft, M., Fisher, A., Lisa, K., Enghin, O., & Nico, P. Detecting Jihadist Messages on Twitter. Intelligence and security informatics Conference(EISIC), European, (2015) 161-164.

[22] Surendiran, R., and Alagarsamy,K., 2010. Skin Detection Based Cryptography in Steganography (SDBCS). International Journal of Computer Science and Information Technologies (IJCSIT), 1(4), 221-225.

[23] Kaati, L., Omer, E., Prucha, N. and Shrestha, A.. Detecting multipliers of jihadism on Twitter. In Data Mining Workshop (ICDMW), 2015 IEEE International Conference, (2015) 954-960.