JCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol-6, Special Issue-6, July 2018 E-ISSN: 2347-2693

Prediction of Violent Extremism from Online Textual Contents

Varuna. T.V

Department of Computer Science and Engineering, N.S.S College of Engineering, Palakkad

[*]*Corresponding Author: varunatv29@gmail.com, Tel.: +91-9645185723*

Available online at: www.ijcseonline.org

Abstract — Social media plays a central role in society nowadays. Extremism is also a hot topic of discussion. Violent extremist activities causes major issues. It is difficult to identify people engaging in extremist activities thorough public cyber spaces like internet cafes, institutions etc... With the help of machine learning concepts the proposed system tries to analyze the online text content and classify it in to Violent or Nonviolent content. SVM Classifier is used here. Then analyzing the user activities till now it will predict the user status, which helps to detect the extremists. The proposed system helps to detect the extremism contents and extremists as well as the extremism supporting users

***Keywords*—** SVM,Extremism,Violentextremism,Violent,Nonviolent

I. INTRODUCTION

Social media plays a very important role today. Extremism also is a hot topic of discussion. The word extremism refers to the state of being extreme. To be more specific this paper deals with the violent extremism, which refers to the beliefs and actions of people who supports and use ideologically motivated violence to achieve radical ideological, religious or political views. Nowadays, issues related to online extremism, terrorist propaganda and radicalization campaigns holds the main attention of researchers. [1][2]. Social media plays a central role in these issues. For example, the relationship between the social media usage and the rise of extremist groups [3][4]. ISIS's success in increasing its influence among the youngsters has been related in part to use of social media for propaganda and recruitment purposes. One reason is that, until recently, social media platform like Twitter provided a public venue where single individuals, interest groups, or organizations, were given the ability to carry out extremist discussions and terrorist recruitment, without any form of restrictions, and with the possibility of gathering audiences of potentially millions[5][6]. Only recently, based on manual reporting some mechanisms have been implemented to limit these abusive forms of communications. By analysing the scenario we get to know that the online content will be mostly spread through cyber cafes and public institutions.

The entire problem is dealt through the analysis of the online content and by classification of the content among the two classes, violent and Nonviolent. The classification technique, SVM is used for the classification of the online text content. It classifies the new samples based on the labelled training set they have given with. Through continuous analysis the preferences and intention of the users also can be predicted. So the idea of monitoring the user activities without affecting the privacy concerns is also proposed for preventing the extremist users from spreading their ideologies. This can be done in servers in a cyber cafe or public institutions

The proposed method can be done using python and scikit-learn libraries. So we are using Anaconda for implementing the model [8].

II. RELATED WORK

Social media is of great importance in today's world. As Twitter continues to grow in popularity, the interactions having the extremism content is being recognized as a serious problem. Islamic State of Iraq and Syria (ISIS) is the extremist group using twitter as a platform for sharing its ideas and recruiting members/jihadis to its groups. Its activities have also led to large scale displacement in Iraq, Syria and elsewhere. As of the end of 2015, almost 8 million persons were internally displaced in Syria and another 4 million were refugees in neighbouring countries. ISIS related attacks in Iraq had displaced about 2.6 million persons in that country. As a response to the rise of extremist conversation by ISIS on twitter, twitter has been continuously suspending accounts which are believed to be associated with ISIS. In February 2016, twitter announced that it had shut down 125,000 accounts primarily related to ISIS from the middle of 2015 to the beginning of 2016. There moved ISIS-related accounts include ISIS-related media outlets, information hubs, and supporters' accounts. For obvious reasons, twitter has never released its algorithm or strategy for determining whether an account is primarily related to ISIS or not. Here we consider different features that may be useful for a surveillance system attempting to identify accounts exhibiting extremist behaviour on twitter. There are very few existing approaches for detecting extremism-linked conversations in social media [9].

The extremist groups are aiming at spreading online hate, which helps them to convince their ideologies. As per the cases analysed here, the Internet and social media sites helps to promote violence as a strategy through the social learning theory and their ideologies, well documented. This includes terrorist groups such as ISIS who are using the Internet and social media sites, as a tool for propaganda via websites, sharing information, data mining, fund raising, communication, and recruitment. And also the study has found that whilst there was a strong online reaction against ISIS, that there was also an online wave towards the glorification of the role of ISIS and their tactics of propaganda, recruitment and radicalization all emerge within the online virtual space [10].

Countering violent extremism remains a pivotal national security priority. Violent extremism has become more destructive and needed to be countered with great concern. Most of the propaganda messages contain verses from the Qur'an or Hadith (prophetic narrations), carefully selected and painted in an extremist ideology that demand action against the enemies of Islam (e.g., the West). Scholars and policymakers concur that big data analytics offers an effective method of detecting and countering violent extremism [11] [17]. Although much has been written on how terrorist networks utilize social media for recruitment, little has been done to systematically capture the evolving dynamics of online Islamist extremist narratives in a big data analytic context. Thus, a pressing need exists to develop robust mechanisms to capture, evaluate, and counter Islamist extremism online. Capturing of Online Religious Extremism (CORE) is a mechanism to prevent this extremism contents. CORE uses an extremism ontology to guide the system by capturing frequently used terms in extremist content and their relationships. Terms like jihad, infidel, Al-Baghdadi, ISIS (also known as DAESH/ISIL), caliphate, sharia etc. and the relations in between are identified, and they are stored as triples in RDF format. Text and metadata of tweets and interlinked content (if relevant) are processed through NLP technique(s) to identify unique semantic signatures that overlap with parts of the extremism ontology that identify extremist narratives. This approach can differentiate semantic and contextual nuances of potential extremist narratives from others. For example, a mention of jihad in the context of defensive war or self/spiritual struggle is not automatically identified as extremist. Domain experts can monitor the accuracy of identification scheme and intervene when content is misidentified. The experts' intervention is instrumental in assessing the reliability and validity of CORE's ontology. To measure variations in extremist content online, a quantitative score is applied. The extremism ontology is assigned score weight dimension to capture the intensity of extremism associated with various concepts and their present relationships in the content [11].

Nowadays, social networking websites (Twitter, YouTube, Tumblr etc.) are some of the largest repositories of user generated content (contextual data) on web. Therefore, Text Classification KNN, Naive Bayes, SVM, Rule Based Classifier, Decision Tree, Clustering (Blog Spider), Exploratory Data Analysis(EDA) and Keyword Based Flagging (KBF) are the most commonly used techniques to identify hate promoting content on Internet.[12].

The text can be classified based on its content. Support vector machine (SVM) is a powerful supervised learning paradigm based on the structured risk minimization principle from computational learning theory. Which is a state of the art classification algorithm that is known to be successful in a wide variety of applications [16]. High generalization ability of the method makes it particularly suited for high dimensional data such as text. Indeed, it has been shown that SVM outperformed most of the other classification algorithms in text categorization tasks [14].

III. Methodology

This section briefly describes the proposed method for Extremism detection. The proposed system is a new idea to save the naive users from extremists. The analysis has shown that the extremists always use public spaces. So the system proposes to detect the extremist as well as extremism supporting users.

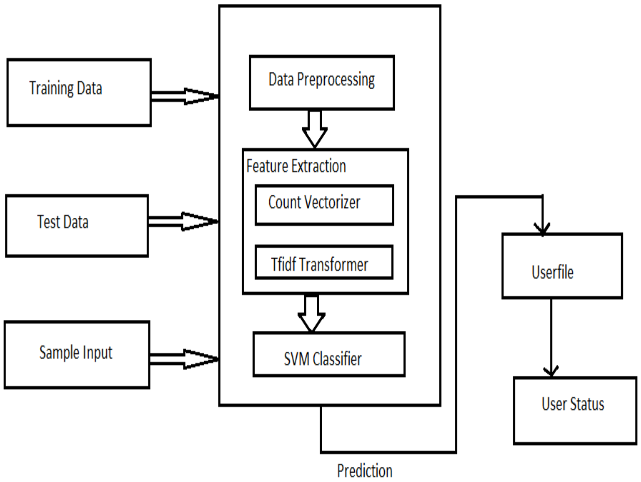

Fig. 1: Illustration of the proposed method

As in figure 1 there will be train data and test data. The text content browsed by the users will be filtered out and is the sample input. Using the training data the proposed system will be trained. Then the sample input will be pre-processed and given for feature extraction process. The feature extraction have two steps count vectorizer and tfidf transformer. The count vectorizer will be creating a vocabulary with the input and counts the occurrence of terms in the vocabulary and this frequency will be calculated. The output from the count vectorizer will be given to the Tfidf transformer. Tfidf transformer will be taking the tfidf (term frequency inverse document frequency) value of each term. By taking this value the weight of each term is calculated in order to get the keywords with its weightage from the inputs. Tfidf value will be calculated using the term frequency and inverse document frequency. Tfidf value is the product of the normalized values of the term frequency and the inverse document frequency as per the equations given below.

$$tfidf_i = tf_{i,d} \times idf_i$$

$$tf_i = 1 + \log tf_{i,d}$$

$$idf_i = \log \frac{N}{df_i}$$

$$tfidf_i = (1 + \log tf_{i,d})(\log \frac{N}{df_i})$$

Then this will be given to the SVM classifier [16]. The content is classified in to two classes, Violent and Nonviolent. The output class will be determined based on the distance between the hyperplane and the sample input, called as decision function. Here the value of decision function will be negative for Nonviolent content which indicates the input is of class Nonviolent and positive when the content is of Violent class. A probability model is needed to compute probabilities of possible outcomes for sample inputs. To obtain results slightly different than the results obtained by predict, using cross validation a probability model is created. The userid with probabilistic value will be passed to the userfile. A userfile is stored for each user in the server. This probabilistic values will be stored there. The values in the userfile will be used to calculate the uservalue and this value is calculated using the equation below

$$uservalue = \sum_{i=0}^{N} \frac{x_i}{N}$$

Which lies in the range 0 to 1. Then according to the analysis done we are dividing the users into the given five status base on the values assumed

- No Risk - when uservalue lies between 0 and 0.2. The users with the No Risk status is the naive users and users who do not supports extreme contents.
- Slightly Risk - uservalue lies between 0.2 and 0.4. The users in the Slightly Risk status also do not support extremism content.
- Moderately Risk – uservalue lies between 0.4 and 0.6. The users is supporting violent and extremism content in a moderate level. These users needed to be monitored.
- Highly Risk - uservalue is between 0.6 and 0.8. The users with this status is assumed to be supporting the extremism and violent content.
- Dangerous – uservalue is greater than 0.8. The users of dangerous status will be assumed to be mostly extremists as well as extremism supporters.

IV. Results and Discussion

The proposed system will be analysing the sample text input and classify it in to one of the appropriate class, Violent or Nonviolent. An example input of the format

- Kill kafirs. We need to build our nation without them. According to our hidayath fighting against kafirs is our duty.

After processing it with count vectorizer and Tfidf transformer and classifier the output class is Violent with a probabilistic value 0.79034.

- Love is the basis of all beliefs. All religions teaches us to love everyone

The output class of the text after count vectorizing, then given to Tfidf transformer and finally SVM classifier is Nonviolent with a probabilistic value of 0.8093.

The classifier after the training, has the accuracy of 83.33%. Then we are passing this probabilistic value along with the class to the userfile stored in the server and by analysing it the user status is determined. The probabilistic value is passed when the text content is violent. Otherwise the 0 value will be passed. By calculating a uservalue using these values in the userfile the status of the user is determined. The user status will be one among the mentioned five status levels

- No Risk
- Slightly Risk
- Moderately Risk
- High Risk
- Dangerous

The users with the dangerous and high risk status will be needed monitoring and further actions to prevent them from spreading the extreme content. According to the proposed system they have more chances of being extremists or extremism supporters. The accuracy as well as the effectiveness of the system can be increased by the implementation of sentimental analysis during the text classification. It will be more helpful to analyse the intensity of the content. It can be done as a future work to enhance the proposed system.

V. CONCLUSION AND FUTURE SCOPE

The main aim of this project is to save the naive users from extremists. The analysis has shown that the extremists always use public spaces. So the system proposes to detect the extremist as well as extremism supporting users through the analysis of the online textual content. Through this project we are implementing a framework helps to recognize and to take proper action against extremists. The text content is taken for the classification process. SVM classifiers are used here for the text classification. The extremism supporting users will be monitored continuously to produce an output status. By this we are analysing the user activities without invading their privacy and helps to recognize the extremists and their supporters. As a future work, by implementing sentimental analysis the accuracy and effectiveness of the system can be incremented.

ACKNOWLEDGMENT

Thanking the Department of Computer Science and Engineering, N.S.S College of Engineering, Palakkad for providing required facilities

REFERENCES

[1] Ferrara E., Wang WQ, Varol O., Flammini A., Galstyan A.,"Predicting online extremism, content adopters, and interaction reciprocity", Social Informatics, vol.10047), 2016

[2] A. Fisher, "How jihadist networks maintain a persistent online presence," Perspectives on Terrorism, vol. 9, no. 3, 2015

[3] Q. Schiermeier, "Terrorism: Terror prediction hits limits." Nature, vol. 517, no. 7535, p. 419, 2015.

[4] S. Reardon, "Terrorism: science seeks roots of terror," Nature, vol. 517, no. 7535, pp. 420–421, 2015.

[5] J. Stern and J. M. Berger, ISIS: The state of terror. Harper, 2015

[6] P. Cockburn, "The rise of Islamic State: ISIS and the new Sunni revolution". Verso Books, 2015.

[7] M. Weiss and H. Hassan, ISIS: Inside the army of terror. Simon and Schuster, 2015.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in python," The Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[9] Yifang Wei,Lisa Singh and Susan Martin, "Identication of Extremism on Twitter", International Conference on Advances in Social Networks Analysis and Mining, 2016.

[10] Imran Awan, "Cyber-Extremism: Isis and the Power of Social Media", SOCIAL SCIENCE AND PUBLIC POLICY, Vol. 54, pp. 138–149, 2017.

[11] Budak Arpinar, Ugur Kursuncu and Dilshod achilov, "Social Media Analytics to Identify and Counter Islamist Extremism: Systematic Detection, Evaluation, and Challenging of Extremist Narratives Online", International Conference on Collaboration Technologies and Systems, vol. 18, no. 7,pp. 1527–1554, 2016.

[12] Swati Agarwal, Ashish Sureka, and Vikram Goyal ,Open Source Social Media Analytics for Intelligence and Security Informatics Applications, LNCS, vol. 9498,pp. 21–37, 2015.

[13] J.Berger and B.Strathearn, "Who matters online: measuring infuence, evaluating content and countering violent extremism in online social networks", Int. Centre for the Study of Radicalisation,2013.

[14] Zi-qiang Wang, Xia Sun, and De-xian Zhang, "An Optimal SVM-Based Text Classification Algorithm", International Conference on Machine Learning and Cybernetics, 2006.

[15] L.Breiman, "Random forests, Machine learning", vol. 45, no. 1, pp. 532, 2001.

[16] https://machinelearningmastery.com/support-vector-machines-for-machine-learning

[17] Conway, "Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research", .StudiesinConflictandTerrorism, 2016.

[18] Hussain,G., Saltman, "Jihad Trending: A Comprehensive Analysis of Online Extremism and How to Counter it, Quilliam Foundation" , The United Kingdom,2014.

[19] Klausen, Tweeting the Jihad: social media networks of western foreign fighters in Syria and Iraq,StudiesinConflictandTerrorism,volume.38,pp.1-22,2015.

[20] H. Chen, W. Chung, J. Qin, E. Reid, M. Sageman, and G. Weimann, "Uncovering the dark web: A case study of jihad on the web," Journal of the American Society for Information Science and Technology, vol. 59, no. 8, pp. 1347–1359, 2008.