RB Journal of Lib & Information Science ISSN: 0972-2750
(UGC Care Group I Listed Journal) Vol-07 Issue-01 No.01: 2017

SUICIDE PREDICTION FOR MILITARY PERSONNEL USING ML

N.V. Ganapathi raju

Associate Profesor, Department of I.T., GRIET, Hyderabad, Email: nvgraju@griet.ac.in

ABSTRACT

About 800 000 people commit suicide every year and detecting suicidal people remains a challenging issue as mentioned in a number of suicide studies. With the increased use of social media, we witnessed that people talk about their suicide plans or attempts in public on these networks. This paper addresses the problem of suicide prevention by detecting suicidal profiles in social networks and specifically twitter. First, we analyze profiles from twitter and extract various features including account features that are related to the profile and features that are related to the tweets. Second, we introduce our method based on machine learning algorithms to detect suicidal profiles using Twitter data. Then, we use a profile data set consisting of people who have already committed suicide. Experimental results verify the effectiveness of our approachin terms of recall and precision to detect suicidal profiles. Finally, we present a Java based prototype of our work that shows the detection of suicidal profiles.

INTRODUCTION

Social media has changed the world. It has become an everyday part of our lives. Many people are nowadays active on several popular social networks such as Facebook, twitter, Instagram, etc. They share photos and posts on their daily life and experiences such as their food, their clothes, and their trips. Some people are more active on social networks, while others are less so. On the other hand, social networks can reflect different social phenomena such as diseases, depression, suicide, etc. In particular, suicide is a complex and dangerous phenomenon that should be considered and studied in order to reduce mortality rates. A recent study1 revealed that close to 800 000 people commit suicide every year, which means one person every 40 seconds. Thus, this growing phenomenon presents one of the biggest challenges the world is facing today. Understanding the symptoms related to suicidal tendencies is important to prevent such deaths In this respect, many studies on suicide prevention have become more prevalent in recent years. Indeed, one of the greatest things that characterize social networks is their use in extracting emotional thoughts and feelings of depression. For that reason, many researchers rely on social networks to study suicide. As an example, Twitter has become a very popular social network where millions of users share their opinions and feelings using short texts called tweets, which contains semantic expressions such as emoticons, hashtags, special characters, etc. Consequently, twitter provides a rich source of data for text mining. Most suicidal people who are active in social networks give signals of their intentions. For example, they make statements such as "I want to kill myself," "I hate my life", "I have lived long enough "or “I’m so tired”. The best way to prevent their suicide is to catch these signals and predict other hidden signals behind their posting content in order to react to them and take appropriate actions. Generally, a user in twitter is characterized by a profile and a set of tweets. The profile features describe his/her persona such as name, age, location, date of birth. On the other hand, tweets refer to the content shared by the user such as text, photos or videos. Some existing works (Jain et al., 2013) in this context utilize publicly shared attributes including name, gender, location, and other information to identify user profiles in social networks. However, due to the privacy settings, user’s attributes are not available in many cases and this makes these existing works fragile. In addition, some researchers address the problem of suicide only through tweets (Kavuluru et al., 2016), (Colombo et al., 2015). However, even though tweets contain rich information that can identify users, they can miss some significant details that maybe available on the user profile public attributes and may contribute to a higher accuracy of suicide detection. Apart from these works, we utilize in our approach both user shared information that we call account features and tweets as an attempt to solve the problem of suicidal profiles detection. First, posted tweets pose important challenges to infer more information about users. The most relevant challenge is semantic features that are difficult to extract directly from user's posted tweets such as stylometry, writing style, sentiments, emojis, hashtags, n-grams, etc. Instead of many existing studies that ignore these features to identify users, we analyse tweets and extract as much as possible of semantic features. Second, adding account features to the user's posted tweets can help to improve the suicide detection task since they may reflect the habits and characteristics of users. Although there are many studies (Sueki, 2014), (O'dea, 2018) that focused on the particular problem of suicidality detection in social networks, they do not take into account the profile itself. They only considered suicide related-communication with the aim of classifying text relating to suicide. However, the biggest challenge for the suicide task is how to detect users who want to commit suicide from their public profiles in social networks. In this paper, we consider the challenge of suicidal profiles detection in Twitter. We analyse posted tweets to extract semantic features including linguistic, emotional, stylometric, etc. These features allow us to distinguish between the writing styles of different users and thus to facilitate the final classification of users into suicidal or not suicidal. In particular, posted tweets contain temporal information that can indicate the real time of user's posting. Such information is very relevant to enrich the user identification and improve the suicide detection. We also use account features related to publicly shared information such as profile photo, location, biography, followees, etc. We exploit these features to infer other implicit ones and build a rich profile that can help us to predict suicidal users. We adopt different data mining tools and techniques for the extraction process. We also introduce a supervised machine learning model to learn the features identifying each user. Moreover, we adopt several classification techniques to classify profiles into suicidal and not suicidal. We apply our method to a data set collected from Twitter and including profiles whose owners committed suicide

RELATED WORK Social media have become increasingly popular and the number of active users continues to increase. Several phenomena such as suicide are now visible on social media. To address suicide and reduce the related mortality rates, many studies were conducted on suicidality in social networks. Kavuluru et al., 2016 conducted a suicide study by classifying text relating to suicide on Twitter. They built a set of account classifiers using lexical, structural, emotive and psychological features extracted from Twitter posts. Their aim was to distinguish between the more worrying content, such as suicidal ideation, and other suicide-related topics. Other studies (Kavuluru et al., 2014) have focused on the writing styleusing the LIWC tool as a sampling technique to identify 'sad' Twitter posts that were subsequently classified using a machine learning classifier into levels of distress on an ordinal scale, with around 64% accuracy in the best-case. Additionally, (Birjali et al., 2017) based their work on WordNet to analyse semantically Twitter data. They address the lack of terminological resources related to suicide by constructing a vocabulary associated with suicide. A case study (O'dea et al., 2015) used both human coders and a machine classifier to confirm that Twitter is used by individuals to express suicidality and that it is possible to distinguish the level of concern among suicide-related tweets. In another work, (De Choudhury et al., 2016) considered online platforms such as Reddit and applied topic analysis and linguistic features to identify behavioural shifts and mental health issues such as suicidal ideation, thus highlighting the risks of supposedly helpful messages in such online forums. Furthermore, (Colombo et al., 2015) investigated the characteristics of the authors of Tweets containing suicidal thinking, through the analysis of their online social network relationships rather than focusing on the text of their posts. More recently, (O'dea et al., 2018) used a dataset of suicide related posts to study how Twitter users respond to suicide-related content compared to nonsuicide related content. They found that the rate of reply to the suicide-related posts was significantly faster than that one for non-suicide related posts, with the average reply occurring within 1 hour. Finally, (Braithwaite et al., 2016) classified text from Twitter users as suicidal or non-suicidal using affective markers and machine classification algorithms – stopping short of examining texts for other forms of suicidal communication. Existing works on suicide prevention are mainly focused on identifying suicidal thinking or ideation and detecting suicidal posts. However, there are no significant research works that focused in particular on suicidal profile detection. Thus, our study aims to contribute to the literature on understanding communication on the topic of suicide in social networks by detecting suicidal profiles on Twitter.

EXISTING SYSTEM

- Reilly N. Grant, AnaM. León, David Kucher, Daniela S. Raicu, Jonathan F. Gemmell, and Samah J. Fodeh(2018), This paper is evaluated suicide-related posts from a user on social media.To analyze the post, they use the clustering algorithm. A Reddit data is feed to clean words and phrases after that the data is a pass to word2vec which is used to create word vector. A k-mean clustering method is utilized to aggregate comparative expressions of Reddit information to find suicide related exercises. This paper endeavors to talk about the specialized and social points of view of the content mining investigation of Reddit information [1].
- Bridianne 'Deaa, Melinda R., Philip J. B, Mark E. Larsena, Alison. Calearc, H Christensena (2018), this paper is a proposed online networking stage Twitter has been utilized by people to impart self-destructive musings and aims. To perceive how customers twitter, respond to a suicide-related substance when diverged from non-suicide-related substances.
- By the use of dataset for suicide and non–suicide-related posts, answers, re-tweets, and likes were looked and dissected. The pace of answer to the suicide-related posts was additionally specifically quicker than of the non–suicide related posts. Mean and the standard deviation is utilized to discover the pace of the answer [2]. Fatima Chiroma, Mihaela Cocea, Han Liu (2018), Right now, evaluated the show of four well-known machine classifiers, for example, RF, DT, NB, and SVM, in grouping suicide-related Tweets. The consequences of the trials indicated the best execution was F-proportion of 0.78 for the suicide-related correspondence (flippant and suicide classes).
- To improve the presentation of machine learning procedures for classification suicide-related correspondence, it's required to additionally analyze and look at the exhibition another machine learning strategies with the consequence of test. In this manner, we expect to examine the presentation of group learning draws near, which might be increasingly exact and strong, as appeared by another examination [3]
- Jingcheng Du, Jianhong Luo, Yaoyun Zhang, Yuxi Jia, Cui Tao, Qiang Wei And Hua Xu (2018), This paper is proposed a Suicide was one of the fundamental wellsprings of passing in the USA. Mental stressors are a noteworthy explanation for suicide. An acknowledgment of mental fatigue in an in peril people will support early balancing activity of reckless habits or suicide. The essential effort to remove mental fatigue from Twitter data using significant learning based tactics. Connection with standard AI estimations indicates the prevalence of significant learning-based systems.

- CNN has been driving the introduction by recognizing suicide-correlated tweets with an exactness of 78.00% and an F1 extent of 83.00%, beating SVM, Decision Trees (DT), etc. RNN based mental fatigue affirmation secures the best F1 extent of 53.25% precise equal and 67.94% ambiguous facilitate, beating CRF. Furthermore, move picking up commencing clinical notes intended for the Twitter corpus beats the planning with Twitter corpus just with an F1 extent of 54.90% via equal caution. The results exhibit the upsides of significant learning-based methodologies for the robotized fatigue affirmation of online life [4].

DISADVANTAGES

- In the existing work, the system does not analyze offline data.
- This system is less performance due to lack of Pre-Processing on data sets.

PROPOSED SYSTEM

A. Input Data

1) Online:- Numerous individuals wherever all through the world may get automatic access to Twitter information using twitter Application Programming Interfaces (APIs).

2) Offline:- We have collected offline tweets from Website https://github.com/npanwar/SuicideTweets_Analysis

B. Pre Processing Data pre-processing is a data mining system that includes converting unprocessed data into a legitimate course of action. Authentic data is normally lacking, clashing, and also disgusting in categorical practices or inclinations, and most likely to contain various bungles. Information pre-preparing is an appropriate method for resolving such concerns [6, 26]. Twitter information is unstructured information. It should be handled before it very well may be utilized. Consequently, the tweets got are cleaned to expel undesirable inconsistencies and hold just data that will help in deciding the hidden feeling [28, 23, and 24]. This makes information simpler to process in the later stages.

The system for pre-preparing comprises of the accompanying advances:

1. Removing non-English tweets.

2. Converting every one of the tweets gathered into lower case.

3. Remove URLs – eradicated all string that portrays connections or hyperlinks introduce in the tweets.

4. Replacing any username introduce in the tweets to @username – expelled the username and considering the way that these are not considered for estimations

5. Converting the hash labels to typical words since hash marks can give a few supportive information, so it is important to displace them with the same word without the hash. For example #Happy replaced with Happy.

6. Removing any superfluous characters, additional spaces, and so forth.

Pre-Processing is implemented on the Twitter dataset to expel pointless words from the tweets , Exclude punctuations, stop words and hash labels to improve exactness in the aftereffect of examination.

ADVANTAGES

- In the proposed work, the model works on online and offline tweets which is collected from the tweeter API and which related to suicide ideation.
- To The system is more effective due to presence of SVM, NB, and RF.

MODULES

SERVICE PROVIDER

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as View All Tweet Data Set Details,Search Tweet Data Set Details,View Suicide Related Activity By Hybrid Features,View All Remote Users,View Positive Retweets By Hybrid Features,View Negative Retweets By Hybrid Features,View Sentiment Accuracy Details,View Tweet Score Results,View Sentiment Results By Hybrid Features.

VIEW AND AUTHORIZE USERS

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

REMOTE USER

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like ADD TWEET DATA SETS,SEARCH ON TWEET DATA SET DETAILS,VIEW YOUR PROFILE.

CONCLUSIONS

In this paper, we worked on detecting user profiles that are at risk of suicide. We worked on twitter and defined a detection model using a set of rich features including linguistic, emotional, facial, timeline as well as public features to identify twitter profiles. We used several machine learning methods (mainly classifiers) for the suicidal detection. Moreover, we implemented a Java based tool to detect suicidal profiles. To evaluate our work, we conducted a series of experiments using a data set of profiles that committed suicide. Results were promising with an average recall of 86%.

REFERENCES

[1] Nock, M. K., Deming, C. A., Fullerton, C. S., Gilman, S. E., Goldenberg, M., Kessler, R. C., ... & Ursano, R. J. (2013). Suicide among soldiers: a review of psychosocial risk and protective factors. Psychiatry, 76(2), 97-125.

[2] Ribeiro, J. D., Pease, J. L., Gutierrez, P. M., Silva, C., Bernert, R. A., Rudd, M. D., & Joiner Jr, T. E. (2012). Sleep problems outperform depression and hopelessness as cross-sectional and longitudinal predictors of suicidal ideation and behavior in young adults in the military. Journal of affective disorders, 136(3), 743-750.

[3] Kochanski-Ruscio, K. M., Carreno-Ponce, J. T., DeYoung, K., Grammer, G., & Ghahramanlou-Holloway, M. (2014). Diagnostic and psychosocial differences in psychiatrically hospitalized military service members with single versus multiple suicide attempts. Comprehensive psychiatry, 55(3), 450-456.

[4] Umhau, J. C., George, D. T., Heaney, R. P., Lewis, M. D., Ursano, R. J., Heilig, M., ... & Schwandt, M. L. (2013). Low vitamin D status and suicide: a case-control study of active duty military service members. PloS one, 8(1), e51543.

[5] Bryan, C. J., Rudd, M. D., Wertenberger, E., Young-McCaughon, S., & Peterson, A. (2015). Nonsuicidal self-injury as a prospective predictor of suicide attempts in a clinical sample of military personnel. Comprehensive psychiatry, 59, 1-7.

[6] Chapman, S. L. C., & Wu, L. T. (2014). Suicide and substance use among female veterans: A need for research. Drug and alcohol dependence, 136, 1-10.

[7] Bryan, C. J., Clemans, T. A., & Hernandez, A. M. (2012). Perceived burdensomeness, fearlessness of death, and suicidality among deployed military personnel. Personality and Individual Differences, 52(3), 374-379.

[8] Khazem, L. R., Houtsma, C., Gratz, K. L., Tull, M. T., Green, B. A., & Anestis, M. D. (2015). Firearms matter: The moderating role of firearm storage in the association between current suicidal ideation and likelihood of future suicide attempts among United States military personnel. Military Psychology, 28(1), 25-33.

[9] Selby, E. A., Anestis, M. D., Bender, T. W., Ribeiro, J. D., Nock, M. K., Rudd, M. D., ... & Joiner Jr, T. E. (2010). Overcoming the fear of lethal injury: Evaluating suicidal behavior in the military through the lens of the interpersonal–psychological theory of suicide. Clinical psychology review, 30(3), 298-307.

[10] O'Keefe, V. M., Wingate, L. R., Tucker, R. P., Rhoades-Kerswill, S., Slish, M. L., & Davidson, C. L. (2014). Interpersonal suicide risk for American Indians: investigating thwarted belongingness and perceived burdensomeness. Cultural Diversity and Ethnic Minority Psychology, 20(1), 61.

[11] Bryan, C. J., Elder, W. B., McNaughton-Cassill, M., Osman, A., Hernandez, A. M., & Allison, S. (2013). Meaning in life, emotional distress, suicidal ideation, and life functioning in an active duty military sample. The Journal of Positive Psychology, 8(5), 444-452.

[12] Bachynski, K. E., Canham-Chervak, M., Black, S. A., Dada, E. O., Millikan, A. M., & Jones, B. H. (2012). Mental health risk factors for suicides in the US Army, 2007–8. Injury Prevention, 18(6), 405-412.

[13] Gilman, S. E., Bromet, E. J., Cox, K. L., Colpe, L. J., Fullerton, C. S., Gruber, M. J., ... & Kessler, R. C. (2014). Sociodemographic and career history predictors of suicide mortality in the United States Army 2004–2009. Psychological Medicine, 44(12), 2579-2592.

