www.ijiarec.com

ISSN:2348-2079

Volume-8 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

Twitter data stream analysis for self destructive related posts using data mining

A.Inbavalli[1], R.Shenbagavalli[2]

[1]PG Scholar, [2]Assistant Professor

Department of Computer Science, Krishnasamy College of Engineering and Technology, Nellikuppam Main Road, Cuddalore, Tamil Nadu 607109

ABSTRACT

Twitter classification for suicide posts detect and prevent is a machine learning approach in which the systems investigate Social media use continues to grow worldwide and an ever-growing number of people are using various social media platforms to update their social circles on their mental health challenges and suicidal ideations in real time. Number of psychological studies show that expressing suicidal thoughts and attempting suicide happens within a matter of hours and therefore automatic detection and analysis of social media posts by vulnerable users, serves as a critical, real-time window into their health and safety. In this paper, we are interested in classifying data from Twitter and Reddit as 'Suicidal Risky Expression' and 'Non-Risky Expression'. We propose a method which includes automatic collection of tweets (Twitter data) and posts (Reddit data) based on suicidal vocabulary, parsing and tokenizing collected textual data, passing this data through a trained neural network and segregating data into two classes, 'Suicidal data' and 'Non-suicidal data'. Since this process runs in real time, the classified data can be used to support at-risk users either by reporting suicidal content to behavioral crisis response teams or by connecting people to mental-health support resources in real-time window.

Keywords: Social Media, Suicide Ideation, Suicide Prevention, Machine Learning approach, Textual information.

INTRODUCTION

It is recognized that media reporting about suicide cases has been associated with suicidal behavior [1] and concerns have been raised about how media communication may have an influence on suicidal ideation and cause a contagion effect between vulnerable subjects [2]. With the advent of open and massively popular social networking and micro blogging Web sites, such as Facebook, Tumblr and Twitter (frequently referred to as social media), attention has focused on how these new modes of communication may become a new, highly interconnected forum for collective communication and, like news media reporting, lead to contagion of suicidal ideation or at least have the effect of normalizing the desire to self-harm [3].

The concerns about suicide-related communication in social media assume that statements of suicidality within social media platforms are indicators of actual suicidal distress in vulnerable individuals who are posting this material, therefore the affective quality of suicide talk in social media needs to be identified and perhaps responded to. There is some limited evidence of an association between online exposure to suicide-related material and offline suicidal ideation [4] although research on this issue is underdeveloped and online prevention is in its infancy.

Social science and medical research have investigated the impact that communication on the topic of suicide via the World Wide Web may have on vulnerable subjects, with particular attention to

Author for correspondence:
Department of Computer Science, Krishnasamy College of Engineering and Technology, Nellikuppam Main Road, Cuddalore.

the younger generation. [5] Conducted a qualitative study by interviewing young adults who engage in suicidal behaviours and use websites dedicated to these themes. [6], [7] also conducted online searches for Web resources containing suicide-related terms and describing suicide methods. They presented a qualitative analysis of the resources they discovered and concluded that, although neutral and anti-suicide Web sites occurred most frequently, pro-suicide forums and Web sites encouraging suicidal behavior were also present and available, suggesting that more prevention plans specifically focused on Web resources are required. Building on this, [8] have reviewed online suicide intervention and prevention literature, concluding that there is a lack of published evidence about online prevention strategies and more attention is required to develop and evaluate online preventative approaches. [9] Also studied the impact of Facebook suicide notes on suicidal behaviour, reporting that it was not yet clear to what extent suicide notes on online social media actually induce copycat suicides. They note that suicide and social media effects deserve further evaluation and research.

Other studies have focused on the written communication of suicide on the Web via bulletin boards [10], newsgroups [11], chat rooms [12], and web forums [13]. These are mostly qualitative analyses and where quantitative data are used in web-related suicide studies, they tend to rely solely on human classification, which is difficult to implement at scale. Computational methods have only been used in a small number of suicide communication studies.

Some studies report a positive correlation between suicide rates and the volume of social media posts that may be related to suicidal ideation and intent [14, 15]. There is also a developing body of literature on the topic of identifying suicidal language on Twitter [16, 17], but very few attempts to use machine classification to automatically identify suicidal language and differentiate between this and other forms of suicide- related communication, such as awareness raising and reporting of suicides. The differentiation is a requirement for the purposes of analyzing the characteristics of suicidal ideation on social media. [18, 19] study depression and other emotional states expressed via social media. Suicidal language is likely to include emotive content and possible signs of depression but we do not suggest depression and suicidal ideation are synonymous in this paper. Two recent papers presented the results of Twitter studies aiming to classify 'risky' language [20] and levels of 'distress' [21] – both reporting classification performance that has potential for improvement (around 60%–64%). An important step in providing support to suicidal social media users is to understand how suicidal ideation is communicated. Recent studies have shown that people are more likely to seek support from non-professional resources such as social media, rather than risk social stigmatization by seeking formal treatment [21].

Thus, our study aims to contribute to the literature on understanding communications on the topic of suicide in social media by (i) creating a new human-annotated dataset to help identify features of suicidal ideation, (ii) creating a set of benchmark experimental results for machine learning approaches to the classification of suicidal ideation, and (iii) developing a machine classifier capable of distinguishing between worrying language such as suicidal ideation, and flippant references to suicide, awareness raising about suicide and reports of suicide. This last contribution is especially relevant to quantify actual volumes of worrying language on social media for the purposes of understanding risk to human safety, as opposed to all references to suicide. The research presented in this paper comprises an analysis of data collected from the micro blogging website Twitter, the text of which has been classified into one of seven suicide-related categories by a crowd sourced team of human annotators. We then use a range of machine learning classification methods to identify suicidal ideation in tweets and analyse the predictive features of suicidal ideation to help explain the language used by perceived suicidal social media users. We apply this to a data set collected from Twitter over 12 months, to further test the most effective classifier, observe trends over time and estimate demographics.

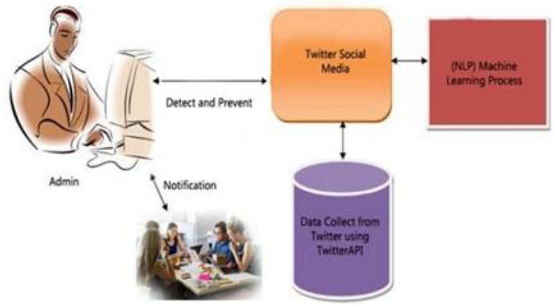

Fig 1. System Architecture

DATA SOURCE

The major criterion for analyzing the service quality and product improvement is about the user's opinion. Blogs, review sites, data and micro-blogs acts as good broadcasting level of the products and services.

Blogs

With an expanding utilization of the web, blogging and blog pages are progressing quickly. Web journal pages have turned into the most prevalent intends to express one's individual feelings. Bloggers record the day by day occasions in their lives and express their sentiments, emotions, and feelings in a web journal (Chau & Xu, 2017)[3]. A large portion of these web journals contain audits on numerous items, issues, and so on. Sites are utilized as a wellspring of suppositions in large portions of the studies identified with sentiment analysis *(Martin, 2015; Murphy, 2016; Tang et al., 2019).*

REVIEW SITES

For any client in doing buying decision, the sentiments of others can be a vital element. The audits for items or services are generally in view of assessments communicated in much of unstructured way. The reviewer's information utilized as a part of a large portion of the feeling characterization studies are gathered from the e-business sites like www.amazon.com (item reviews), www.yelp.com (restaurant surveys), www.CNET download.com (product surveys) and www.reviewcentre.com, which has a great many reviews by shoppers. Other than these the accessible are experts review sites, for example, www.dpreview.com, www.zdnet.com and shopper feeling destinations on wide themes and items, for example, www .consumerreview.com, www.epinions.com, www.bizrate.com (Popescu & Etzioni, 2015; Hu B.Liu, 2016; Qinliang Mia, 2019; Gamgaran Somprasertsi, 2019).

DATASET

A large portion of the work in the field utilizes movie surveys information for grouping. Movie reviews is a piece of information that are available as dataset (http://www.cs.cornell.edu/People/pabo/movie- review-data). Other dataset which is accessible online is Multi- Domain Sentiments (MDS) dataset. (http://www.cs.jhu.edu/mdredze/datasets/feeling). The MDS dataset contains four unique sorts of products surveys extricated from Amazon.com including Books, DVDs, and Electronics and Kitchen apparatuses, with 1000 positive also, 1000 negative audits for every area. Another survey dataset accessible is http://www.cs.uic.edu/liub/FBS/CustomerReviewData.zip. This dataset comprises of surveys of five hardware items downloaded from Amazon and Cnet (Hu and Liu, 2016; Konig & Brill, 2016; Long Sheng, 2018; Zhu Jian, 2018; Pang and Lee, 2014; Bai et al., 2015; Kennedy and Inkpen ,2016; Zhou and Chaovalit, 2018; Yulan He 2019; Rudy Prabowo ,2019; Rui Xia, 2017)

MICRO-BLOGS

Twitter is a prominent micro blogging service where clients make status messages called "tweets". These tweets once in a while express suppositions about distinctive subjects. Twitter messages are likewise utilized as information hotspot for ordering sentiments.

DIFFERENT LEVELS OF ANALYSIS

Parts of speech

We used to the Stanford Part-Of-Speech (POS) Tagger7 to assign each word in a Tweet a POS label. Examples are nouns (broken down into singular, plural, proper), verbs (specifying tenses such as present, past and present participle), 1st vs 3rd person references, adjective and adverbs (comparative, superlative), pronouns (personal, possessive), as well as other tags representing conjunctions, determiners, cardinal numbers, symbols, and interjections. For each of POS we considered the frequency of each in a Tweet as a feature.

Other structural features

For this we considered the inclusion of negations in the sentence (total number), the specific use of a first person pronoun (either singular or plural), and external communication features such as the inclusion of a URL in a tweet or a mention symbol (indicating a retweet or reply).

General lexical domains

- These features represent general lexical categories such as home, religion, psychology, sociology, etc.
- These were extracted using Word Net Domains labels.

Affective lexical domains

These are a set of categories specifically related to domains representing 'affective' concepts. These include concepts representing moods, situations eliciting emotions, or emotional responses such as joy, anger, grief, sadness, enthusiasm, surprise, love, hate, and happiness; but even more specific sub- categories such as amicability, belligerence, bad-temper, unrest, and trepidation; and opposites such as positive-negative concern, negative fear, positive-negative suspense, self-esteem, self consciousness, self-pity, and self-deprecation. These are very appropriate for the specific language we are investigating in this study.

Sentiment score

Using SentiWordNet9 each words is assigned a score between zero and one for both positivity and negativity. The sum all words in a Tweet were used as features.

Words

The most frequently used words and n-grams in terms of (first 100) unigrams, bigrams and trigrams contained in the training set.

Keyword list

We also included each of the 62 keywords derived from the Web form text that were used for the pre-filtering search (e.g. 'asleep and never wake', 'don't want to try anymore', 'end it all', 'isn't worth living', 'my life is pointless', 'kill myself', 'to live any more", 'want to end it', 'want to disappear', 'want to die', etc.). Each of the search terms were included as individual features together with one global binary feature representing the inclusion of any of them in a Tweet.

Feature set 1

Given the psychological and emotional expressiveness of suicidal ideation, we then explored a second set of features by using the Linguistic Inquiry and Word Count LIWC text analysis software [39] to extract more specific labels representing affective emotions and feelings within the text. We refer to these features as Set2. These include a more extensive breakdown of categories that may be more suitable for the particular language of emotional distress that we would expect to be present in suicidal ideation. Examples are related to death, health, money, religion, occupation, and achievement, senses (e.g. feeling, hearing, seeing), and three other groups of terms related to 'cognitive mechanisms', 'affect', and 'social words'. These can be further broken down into labels representing family, friends, humans; anxiety, anger, sadness and positive and negative emotions; and terms related to certainty, inhibition, insight, causal, inclusivity and exclusivity. A subset of these features (sadness) were used in

[21], but we have incorporated a wider range of the feature set to enable us to distinguish between distress and other forms of suicide-related communication (e.g. grief, support and reporting).

Feature set 2

Next, due to the noisy nature of social media, where short, informal spelling and grammar are often used, we developed a set of regular expression (RegEx) and pattern matching rules from our collection of suicide-related posts collected from social networking website Tumblr. We refer to these features as Set3. These were annotated as part of the human annotation process conducted earlier and introduce language from short informal text related to the six suicide related categories to assist the classifier. Examples of these expressions for each class (numbered 1–6 here) include:

1. :‘.+((\cutting\depres|\sui)|\these|\bad|\sad). + (\thoughts| \feel).+’ to represent phrases such as ‘suicidal / cutting / bad / these . . . thoughts / feelings’; ‘.+\wan\w.+d[ie].+’ for expressions as ‘want/wanted/wanting to die’; ‘.+\end.+ (\all|\it|\life).+’ for sentences with ‘end/ending it all’ and ‘end my life’; and‘.+ (can.+|don.+|\take).+(\go|\live|\anymo|cop|\ alive).+’ covering a wide range of phases including ‘can’t take anymore’, ‘can’t/don’t want to live/cope anymore’, ‘don’t want to be alive’, ‘can’t take it anymore’, and ‘can’t go on’. In addition, we added a list of individual words and n-grams including ‘trigger warning’, ‘tw’, ‘eating disorder’, ‘death’, ‘selfharm’ and ‘self harm’, ‘anxiety’, and ‘pain’.
2. :‘.+(\need|\ask|\call|\offer).+\help.+’ related to phrases as ‘call/offer for/of help’ and individual terms as ‘shut’ (e.g. website shut down) and ‘stop’ (e.g. bullying).
3. ‘.+(\kill\hat\throw)’ for phrases including ‘kill/killing /hate myself’, ‘.+(\f**k.+’ for swearwords such as ‘f**k/ f**king’, ‘.+ (\boy\girl).+(\friend)’ for expressions with ‘boy-friend’ and ‘girlfriend’, and ‘.+(\ just)\.+(\like).+’ covering expression including ‘just’ . . . like’. In addition, some words related to general topics such as ‘work’ and ‘school’ have also been included since they are representing contexts more favourable to flippant language rather than genuine expression of distress and suicidal intent.
4. +(\talk|\speak).+\to.+(\one|\some|\any).+’ related to phr-ases as ‘talk / speak to someone/somebody’ and words such as ‘web’, ‘blog’, ‘health’ , and ‘advice’.
5. ‘.+miss.+(\you|\her|\him).+’ related to phrases such as ‘miss/missing you/her/him’ and‘.+(\kill|\die|\comm).+(day| month|year).+’ to represent specific time references.
6. ‘.+(\took|\take).+\own.+\life.+’ covering expressions including ‘took/taken his/her own life’ and words related to suicide methods such as ‘hanged’, ‘hanging’ and ‘overdose’.
7. Note that the regular expressions included in the third class representing flippancy were also identified within those related to the first suicidal class (and vice versa). However, we decided to associate Regex to only one of the two classes according to the nature of the annotated tweets, for example phrases as ‘hate myself’ or ‘kill myself’ were frequently associated with flippant posts whereas terms such as ‘wanted to die’ and ‘want to end it’ were more likely to be included in tweets containing evidence of suicidal thinking.

SUBJECTIVITY AND OBJECTIVITY CLASSIFICATION

Subjectivity/Objectivity ordering is a test that should be tended to work with sentiment analysis issue. The textual information may be useful or not. The subjective sentences are the relevant texts and objective sentences are the irrelevant texts. So, the sorting of sentences should be done for the sentiment analysis. This classification is known as subjectivity classification.

B. Pang and L. Lee displayed a technique for subjectivity identification for the sentiment analysis. This is essential, since the unimportant information from the audits could be dispensed off. This eliminates the handling overheads of large amount of data. The strategy is utilizing minimum cut to create subjective extracts from the content.

The work has been engaged in the sentence level subjectivity extraction.

J. Wiebe presented the Naive Bayesian classifier. They showed the outcomes of creating subjectivity classifiers utilizing un-clarified texts for training. In this work of learning Subjective and Objective sentences, the technique consequently creates training information. This is done by a Rule- based methodology. The rule based subjective classifier orders a sentence as subjective in the event that it contains two or more subjective guessing. Conversely, the principle based target classifier searches for the nonappearance of intimations: it groups a sentence as target if there are no solid subjective enlightens the present sentence, there is atleast one solid subjective educate the past and next sentence consolidated, and at most 2 frail subjective enlightens the present, past, and next sentence consolidated classifiers. They utilize Subjective Precision, Subjective Recall, Subjective F measure, Objective Precision, Objective Recall and Target F measure for the assessment.

CHALLENGES TWITTER SUICIDE ANALYSIS

Several challenges that are considered to the major in the field of twitter analysis were studied. Some challenges were listed below:

Named Entity Extraction

Named entities referred as the definitive noun phrases that specifies about the types of individuals such as organization, persons, dates and so on. The aim is to extract the textual identification of the named entity in a text. It is well suited for the classifier based approach.

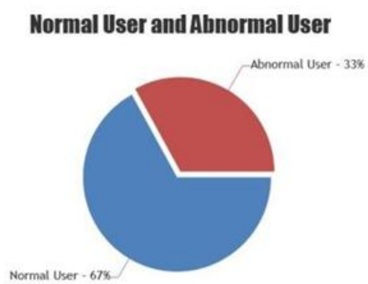

Fig 2. Classification of Normal and Abnormal User

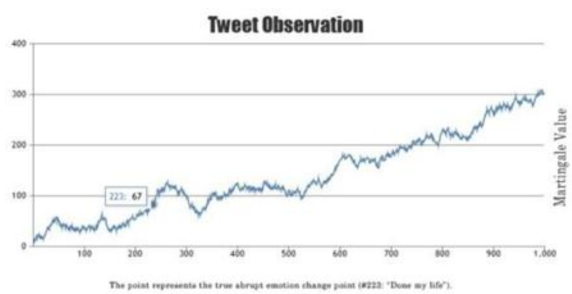

Fig 3. Tweets Observation

Information Extraction

Data may be available in numerous shapes and sizes. The tools in NLP are still not completely skilled to fabricate general purpose representations of context from hidden text. Based on the available information, the structured data may contain regular entities and relationships. This can be applied in the field of business intelligence, media analysis, sentiment detection, patent search, and email scanning.

Sentiment Determination

The role of sentiment determination is to analyze the polarity of a word, sentence or document. The lexicons are considered to be the vital source for sentiment analysis. The adjective part of a sentence possesses more probability to handle the information.

Co- Reference Resolution

The process done in aspect and entity level is termed as 'Co-Reference Resolution'**.** In view of opinionated text, many comparative texts are available. The comparative texts hold the references which are resolved to produce the results.

Relation Extraction

Relation extraction is the assignment of discovering the syntactic connection between words in a sentence. The semantics of a sentence can be discovered by extricating relations between words and this should be possible by knowing the word conditions. This is additionally a noteworthy exploration zone in NLP.

Domain Dependency

Sentiment classifiers that analyze the polarity of a sentence in a domain generate the results under same classifier. Opinion is communicated diversely in distinctive spaces. For example, consider two areas, computerized camera and car. The path in which clients express their thoughts and views about computerized camera will be unique in relation to those of car. In any case, a few similitudes might likewise be available. So, domain dependency is an issue which has high space reliance.

OPINION MINING AND SENTIMENT ANALYSIS

Several researchers studied the sentiment analysis of user opinion data. These data predicts the information based on the polarities of the user reviews. The literature survey is done on two types that include machine learning and semantic orientation.

Machine Learning

The machine learning methodology material to sentiment analysis fits in with supervised classification. In this manner, it is called as 'supervised learning'. In a machine learning based characterization, two sorts of documents are required: Training and testing data. The training data is utilized by a programmed classifier to take in the document quality, and a test set is utilized to approve the programmed's execution classifier. Various machine learning methods have been received to order the reviews. Machine learning procedures like Naive Bayes (NB), Maximum Entropy (ME) and Support Vector Machines (SVM) have accomplished greater results in content classification. The other learning methods are K-nearest neighborhood, ID3, C5, Centroid classifier, Winnow classifier and N- gram model.

Naïve Bayes is an effective classifier. It is mostly applicable to document classification. *(Melville et al., 2019; Rui Xia, 2018; Ziqiong, 2018; Songho tan, 2018 and Qiang Ye, 2019).* For a given document, the joint probabilities of words and categories are estimated. It works on the assumption of word independency. The computation process is very complex in Naïve Bayes classifier. Support Vector Machines (SVM) is a distinctive classifier, widely used in text classification schemes (*Rui Xia, 2018; Ziqiong, 2016; Songho tan, 2018 and Rudy Prabowo, 2019)*. It is invented by Vapnik. It performs on structural risk minimization principle. The decisions are generated based on the support vectors that are selected as the training data points. Multi- class SVM is widely used in sentiment ordering *(Kaiquan Xu, 2016)*. The Centroid algorithm is simple to use. The Centroid vector for training class is generated. And then the similarity between the documents to its Centroid is estimated. The document is assigned to the class based on the most significant Centroid values *(Songho tan, 2018)*.

The k- nearest neighbor (KNN) works on the basis of categories label that are obtained from the training document in relative to the test document. Given a test document d, the system finds the k nearest neighbors among training documents. The similarity score of each nearest neighbor document to the test document is used as the weight of the classes of the neighbor document *(Songho tan, 2018)*. Winnow is a type of classifier, eminently known in online mistake driven method. The text weights are updated at iteration level. In first iteration, it calculates the weights and transmits to the document and receives its feedback. If the feedback is wrong, it again calculates and modifies the weights before continuing the process. In training phase, the process is repeated until the weights are updated and predicted accurately *(Rudy Prabowo, 2019).*

An ensemble technique is one which joins the yields of classification models to frame an incorporated yield. *Rui Xia in 2011* utilized this methodology and made a relative investigation of the adequacy of ensemble method for opinion mining by effectively incorporating diverse capabilities and algorithms to combine a more exact classification procedure. In several studies, the SVM works better than the machine learning system. A good performance system in SVM was proposed by *Ziqiong Zhang in 2011*. An earlier prediction of the knowledge was discovered. The lexical elements were utilized to announce a good conclusion using the Cantonese review.

CONCLUSION

In this paper we developed a number of machine classification models built with the aim of classifying text relating to communications around suicide on Twitter. The classifier distinguishes between the more worrying content, such as suicidal ideation, and other suicide-related topics such as reporting of a suicide, memorial, campaigning and support. We built a set of baseline classifiers using lexical, structural, emotive and psychological features extracted from Twitter posts. We then improved on the baseline classifiers by building an ensemble classifier using the Rotation Forest algorithm, achieving an F-measure of 0.728 overall (for 7 classes, including suicidal ideation) and 0.69 for the suicidal ideation class.

We summarized and attempted to explain the results by reflecting on the most significant predictive principle components of each class to provide insight into the language used on Twitter around suicide-related communication. From this analysis we observed that word-lists and regular expressions (regex) extracted from online suicide-related discussion for a and other micro blogging Web sites appear capable of capturing relevant language 'clues', both in terms of single words, n-grams (word-lists) and more complex patterns. These appear particularly effective for the suicidal ideation class, expressing emotional distress. Lexical and grammar features such as POSs appear mostly ineffective and scarcely present in the principal components (only some mentions as predetermines, existential clauses and superlatives that, however, also relate to more specific 'affective' language features than only pure lexical ones). Affective lexical domains, appear instead very relevant (such as those represented by the WordNet library of 'cognitive synonyms') and able to well represent the affective and emotional states associated to this particular type of language.

Concepts and labels representing broader semantic domains (also derived form the WordNet library) are, on the contrary, not effective. In fact, although they appear rather numerous as attributes within the principle components they reveal to be, on close inspection, for the majority of cases irrelevant and mostly generated by a 'confusion' and 'mis-representation' of words (such as sentences like 'my reason crashed' associated to the 'motor-racing' domain, and 'suicide watch' associated to 'numismatic').

Sentiment Scores generated by software tools for sentiment analysis appear also ineffective and either scarcely or not at all included within the principal features of each class. Note that this is true for both basic tools that only provide a binary representation of positive and negative score values (SentiWordNet) as well as more sophisticated text analysis software that generate sentiment scores over a larger range of labels representing emotional states (LIWC).

A classifier for suicide-related language could potentially make an important contribution to suicide prevention. Monitoring individual social media accounts via keywords that suggest possible suicidal ideation is controversial territory, as shown by the recent withdrawal of the Samaritans Radar app in the UK11 but there is nonetheless potential for such a lexicon to contribute to prevention in some way, as long as acceptability to social media users is thoroughly investigated. The 'real-time' identification of aggregate levels of suicide- related communication at scale in online social networks, which could be facilitated by the ensemble classifier produced in this research, is one possible approach. There is positive potential, for example, for using the classifier to monitor trends at an aggregate level, to inform service provision. Although we found a lack of correlation between the timing of apparently suicidal tweets and actual suicides, nonetheless, a marked increase in the volume of suicidal tweets, such as around the time of high profile celebrity suicides, may well suggest an increased need for helpline and other support for people who are in distress and perhaps at longer-term risk of suicide. Using the classifier to monitor social media communication could help with planning for increased provision.

Our classifier goes beyond the recognition of suicidal language insofar as it also aids identification of other kinds of communication, in recognition that social media platforms can be used for multiple purposes, including the reporting of news and marshalling of campaigns. Monitoring of suicide news reporting in social media is another potential avenue where text mining and machine classification techniques could be applied. The identification of flippant use of suicidal language could be especially useful. The methods needs further development, ideally with a larger sample of social media postings, and application to platforms other than Twitter. Finally, we note that it is important to retain collaboration with domain experts in suicidology throughout the experimental and interpretation phases of future research to improve classification accuracy by incorporating prior knowledge of the characteristics of suicidal language - especially given the significance of the affective features in this paper.

REFERENCES

[1]. Andrea Esuli and Fabrizio Sebastiani, “Determining the semantic orientation of terms through gloss classification”, Proceedings of 14th ACM International Conference on Information and Knowledge Management, Bremen, Germany, 2015, 617-624,.

[2]. Bai, and R. Padman, “Markov blankets and meta-heuristic search: Sentiment extraction from unstructured text,” Lecture Notes in Computer Science, 3932, 2016, 167–187.

[3]. Bing xu, tie-jun Zhao, de-quan zheng, shan-yu wang, “Product features mining based on conditional random fields model, “Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 2018, 11-14.

[4]. Chaovalit, Lina Zhou, Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches, Proceedings of the 38th Hawaii International Conference on System Sciences 2015.

[5]. Chau, M., & Xu, J. Mining communities and their relationships in blogs: A study of online hate groups. International Journal of Human – Computer Studies, 65(1), 2017, 57–70.

[6]. Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng and Chun Jin, “Red Opal: product feature scoring from reviews”, Proceedings of 8th ACM Conference on Electronic Commerce, New York, 2017, 182-191.

[7]. Chunxu Wu, Lingfeng Shen, “A New Method of Using Contextual Information to Infer the Semantic Orientations of Context Dependent Opinions”, 2019 International Conference on Artificial Intelligence and Computational Intelligence.

[8]. Gamgarn Somprasertsri, Pattarachai Lalitrojwong , Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization, Journal of Universal Computer Science, 16(6), 2018, 938- 955.

[9]. Gang Li, Fei Liu. “A Clustering-based Approach on Sentiment Analysis”, 2018, 978-1-4244-6793-8/10 ©2018 IEEE.

[10]. Go Lei Huang and Richa Bhayani. “Twitter Sentiment Analysis”, Project Report, stand ford, 2019.

[11]. Go, Lei Huang and Richa Bhayani, “Twitter Sentiment Classification using Distant Supervision”, Project Report, Standford, 2019.

[12]. Guang Qiu, Xiaofei He , Feng Zhang , Yuan Shi , Jiajun Bu , Chun Chen , “DASA: Dissatisfaction-oriented Advertising based on Sentiment analysis” , Expert Systems with Applications, 37, 2018, 6182–6191.

[13]. Hu, and Liu, "Opinion extraction and summarization on the web", AAAI., 2016, 1621-1624.
[14]. Hu, and Liu, "Mining and summarizing customer reviews", Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 2015, 168–177.
[15]. Hu, Liu and Junsheng Cheng, "Opinion observer: analyzing and comparing opinions on the Web", Proceedings of 14th international Conference on World Wide Web, Chiba, Japan, 2015, 342-351,.
[16]. B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, ACL, 2014.
[17]. J. Wiebe and E. Riloff, "Creating subjective and objective sentence classifiers from unannotated texts," in Computational Linguistics and Intelligent Text Processing. Springer, 2015, 486–497.
[18]. F. L. Gang Li, "A clustering-based approach on sentiment analysis," IEEE, 2019.
[19]. M. X. Y. S. Haiping Zhang, Zhengang Yu, "Feature level sentiment analysis for Chinese product reviews," IEEE, 2017.
[20]. T. K. M. J.K. Sing, Souvik Sarkar, "Development of a novel algorithm for sentiment analysis based on adverb adjective- noun combinations," IEEE, 2018.
[21]. D. A. A. S.M.Shamimul Hasan, "Proximity-based sentiment analysis," IEEE, 2017. [22] Aggarwal Charu C, Zhai Cheng xiang," Mining Text Data", Springer New York Dordrecht Heidelberg London: Springer Science + Business Media, LLC' 12, 2016.
[22]. Kang Hanhoon, YooSeongJoon, Han Donglil,"Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews" Expert SystAppl, 39, 6000-10, 2017.
[23]. Ortigosa-Hernandez Jonathan, Rodriguez Juan Diego, Alzate Leandro, Lucania Manuel, InzaInaki, Lozano, Jose,"
[24]. A. Approaching sentiment analysis by using semi-supervised learning of multi- dimensional classifiers" Neurocomputing; 92, 2015, 98- 115.
[25]. Li Yung-Ming, Li Tsung- Ying, "Deriving market intelligence from microblogs", Decis Support Syst, 2013.
[26]. Jiancheng Fang and Zhanchao Liu, "In- Flight Alignment of POS Based on State- Transition Matrix", IEEE sensors journal, 15(6), 2015.
[27]. Chenliang Li et al, "Tweet Segmentation and Its Application to Named Entity Recognition", IEEE transactions on knowledge and data engineering, 27(2), 2015.
[28]. Zhenghua Li et al, "Joint Optimization for Chinese POS Tagging and Dependency Parsing", IEEE/ACM transactions on audio, speech, and language processing, 22(1), 2014.
[29]. Lorenzo Gatti et al, "Senti Words: Deriving a High Precision and High Coverage Lexicon for Sentiment Analysis", IEEE transactions on knowledge and data engineering, 6(1), 2017.