Connecting Suicidal and Help-Seeking Behaviours

Kushal Govindbhai Gevaria
Department of Computer Science
Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, NY 14623
kgg5247@rit.edu

***Abstract*—Although there are lots of researchers and suicide prevention teams out there to help individuals with mentalhealth issues, many cases of suicide happen undetected. Social media provides a platform for users to express themselves online. Can we help such suicide prevention teams on social media to identify individuals who might later have suicidal thoughts due to mental health issues? The main goal of this project is to identify those individuals from social media who have suicidal thoughts and express their thoughts through general issues in their daily life. We have picked Reddit dataset to target users who had suicidal thoughts in the past and those users who just exhibit general issues like depression, anxiety etc. and express the same by posting in certain subreddits. No machine learning and classification algorithms work efficiently on the raw dataset. So, our next step is about preprocessing and data cleaning to extract relevant features out of the user's posts for our classification algorithms. Next we perform topic modeling algorithm and support vector machine classification algorithm to classify the users based on their post in the given subreddits.**

***Keywords*—Suicidal Ideation; Data Preprocessing; Textual Classification; Topic Modeling; Latent Direchlet Allocation; Support Vector Machine**

I. Introduction

Social media contributes an enormous amount of user-generated content on the topics of mental illness and suicidal. Analyzing these data helps to identify the individuals showing mental or health-related concerns that lead to thought of committing suicide. About 80% of the individuals who have mental illness feel like their life is not worth it and express this feeling on online or offline platforms. If there was an automated process to detect the transition mentioned before using a certain score-based system, then such targeted individuals could be recommended for treatment or support from psychologists or natural helpers. Suicide prevention organizations would benefit from such analysis targeting help-seeking individuals. There may be some privacy or ethical issues if this automated process becomes reality. The information pertaining to individuals used to detect the transitions can be misused or even misinterpreted through this automated process.

Suicide is the 18^{th} leading cause of death throughout the world. Nearly 800,000 deaths occurred in the year 2017 [1]. World Health Organization (WHO) provides a yearly statistical report on suicidal attempts and also mentions that that death of person due to suicide leads to 20-30 more deaths. Suicidal ideation is a term used when an individual discusses something that is directly related to death or might eventually lead to having suicidal thoughts [2]. Social media is a common platform for determining suicidal ideation. Thus, a lot of researchers are motivated to find trends or analyze existing data related to suicide to help understand the cause of it and try to prevent it.

In prior work, researchers focused on identifying potential terms related to suicidal ideation using linguistic structure, interpersonal awareness, and interaction as the three major predictors [2]. These predictors supported in identifying multiple psychological factors like depression, mental illness, trauma, stress, unemployment, hardships, social anxiety that are taken into consideration to evaluate the end result of identifying users having suicidal thoughts. They have created a regularized logistic regression model on top of these predictors that ultimately helped them to target around 80% of individuals having suicidal ideation. Even with this accuracy, they don't implicitly mention that the detected individuals are actually having suicidal thoughts. Hence, they mention asking certain questions to the targeted individuals about their well-being with the help of psychologists and natural helpers.

In this paper, we are using an advanced statistical approach called latent dirichlet allocation (LDA) to solve the problem. Latent dirichlet allocation is a topic model which is used to identify trends in a document. The words and phrases in the document that convey the similar meaning are classified into topics like "Individuals having general issues" or "Individuals having suicidal ideation". But still it will provide probabilistic values for each topic that gives document as a mixture of topics rather than being firm on just one particular topic. For better efficiency, we implemented a logistic regression classification algorithm on top of this LDA model. We perform our analysis on the Reddit dataset. It provides user information such as "user_id", "title", "post_description" and other metadata. It also provides subreddits like "SuicideWatch" and "GeneralIssues" to classify users based on their posts.

We extracted 7,168 users with approximately 40,000 raw posts. From those 7,168 users, 4000 users were just subscribers of "GeneralIssues" subsections while remaining 2,168 GI users eventually subscribed to "SuicideWatch" subsection. From those 2,168 users subscribed to "GeneralIssues" subsections, 1,793 users (80% of 2,168 $GI \rightarrow SW$ users) in the year 2015 were classified to demonstrate suicidal ideation. The results were evaluated by cross-validating whether these users actually subscribed to "SuicideWatch" subreddit section. Ultimately, this model turned out to be better than just using the regularized logistic regression algorithm to identify users exhibiting suicidal ideation with an accuracy of 80%. However, this does not imply that the targeted users were actually having suicidal thoughts.

The rest of the paper is as follows: In Section II, we discuss the related work done in this field. Section III, talks about ways to collect data from the Reddit social media and preprocessing steps on the same. In Section IV, we discuss the latent dirichlet allocation (LDA) topic modeling approach. Section V talks about the results and evaluations. Section VI presents the conclusion.

II. Related Work

Given below represents some research papers on the topic related to suicide as the cause of death.

Social media provides a wide range of content that also includes mental or health-related issues and also suicidal ideation related posts and comments. Choudhury *et al.* study suicidal expression on Reddit [2]. In this research, the authors selected two main threads, namely “r/MentalHealth” (MH) and “r/SuicideWatch” (SW). MH focuses on posts related to general mental and health issues. This also includes depression, trauma, eating disorders etc. as the content of the posts. SW focuses more on posts related to helping those having suicidal thoughts or suicidal ideation. It also includes posts with individuals going through certain kinds of psychological therapy. Researchers found out that individuals posting or commenting on the MH threads tend visit the SW thread as well. They created a model to predict such transitions that will eventually help to find out the root cause of having suicidal thoughts, but they are not certain about actually identifying individuals expressing suicidal thoughts. Linguistic, interpersonal, and interaction are the measures adopted from a literature by Chung and Pennebaker, which are used to identify this transition [3]. Besides these three models, they have also used Content as their 4^{th} feature representing unigrams and bigrams. They extracted unigrams and bigrams from the posts and comments text (also referred as tokens). Finally, they performed a regularized logistic regression binary classification to predict whether individuals really made transitions from MH thread to SW thread or not. They performed this analysis over 10 months of data in the year 2014. This Reddit data includes approximately 63000 posts, 209000 comments, and 35000 users. The idea of this research [2] closely relates to the work shown in this paper where a comparison is created between the suicidal ideation data and help-seeking behaviour data available on Reddit.

Signs of suicidal thoughts in an individual are eventually found due to some kind of depression, mental illness or any other disputes. Guntuku *et al.* focus on detecting such signs from the social media content [4]. Social media like Facebook, Twitter, Reddit, and other web forums do provide a wide range of data in this particular domain. They use text posts from this dataset to create n-grams, linguistic inquiry and word count (LIWC), and sentiment features. These features assist them to create a supervised model using logistic regression [5] and also a model using support vector machine [6]. The classified the text posts based on depression or mental illness related topics. They represented results using receiver operating characteristic (ROC) curve with an accuracy of 72%. Data extraction and analysis on this paper [4] closely relates to the comparison carried in this particular project between help-seeking behaviours and suicidal thoughts.

Another evidence of suicidal ideation is in this article represented by Leite *et al.* [7]. Various thoughts about death or the range of depressive thoughts that eventually lead to death as the final thought is the process of suicidal ideation. Authors consider suicidal ideation as an important feature to predict whether the individual having these thoughts would likely commit a suicide or not. They also perform a statistical analysis to find how common is suicide being considered as a measure of death among all possible deaths over the world [8]. Younger people between the ages of 15 and 25 are considered to be the most involved in suicidal. Thus they consider population age as one of the major features [9]. Besides suicidal ideation and population age, the authors also use social isolation as another feature [10]. Social isolation is a term used when individuals don’t get themselves involved in any social activities and also not socially integrated. Often this leads to feeling lonely and causing risk of suicidal ideation. Leite *et al.* mention in this article that social networking with friends and parents benefits individuals with coping up during the times of depression. Features mentioned in this article would definitely help towards finding out better patterns and distinguish population having suicidal thoughts on social media from other kinds of response on the same.

Cheng *et al.* is focused on observing communications related to suicidal topic on the China specific social media named Weibo suicide communication (WSC) [11]. They surveyed the users of Weibo suicide communication and asked about their nature towards factors like anxiety, depression, stress etc that eventually leads to performing suicide ideation on this Chinese Microblog [12]. The Mann-Whitney-Wilcoxon model and the chi-square test model were used to differentiate between WSC individuals and non-WSC individuals thereby giving them a gist of content of suicidal communication from microblog. Eventually, they are trying tweak certain variables and parameters by creating separate models and trying to find out which one performs the best to achieve the final goal of differentiation. The same idea of dealing with social media content helps my project to navigate with proper usage of social media content and the features that it provides.

Perform classification on suicide-related communication online on social media is the main goal of this research paper [13]. Burnap *et al.* have used twitter as a medium of social media to perform this classification analysis. They extracted some important keywords from Tumblr and other web blogs related to suicidal thoughts, and then used term frequency/inverse document frequency (TF.IDF) method to use those keywords and annotate them based on the most frequently used words for suicide as the related topic [14].

Based on these important keywords they extracted those posts from twitter dataset using the twitter API, which are related to suicidal topic. After extracting the feature out of the model, they have used classification models namely support vector machine (SVM), rule based (decision tree), and naive bayes classification (NB). Finally, they performed a comparative analysis on the same. This process of data extraction and preprocessing is similar to the Reddit dataset which will be used for this project.

This research paper represents another example of social media content similar to the previous research paper.Abboute *et al.* have mainly focused on mining meaningful data related to suicide as a support to the previous research paper [15]. Basically they are trying to classify the twitter dataset between risky content and non risky content. In this way, the risky content would include signs of suicidal ideation. This process of mining the data and differentiating between risky and non risky content helps to achieve the same goal towards this project where we eliminate the content which might not be risky. For example, there are certain pro-social individuals or natural helpers who are trying to help individuals having suicidal thoughts. So this research paper will help in eliminating those natural helpers from the Reddit dataset to eventually focus on just individuals having suicidal thoughts or other mental issues.

III. Data

Targeting users with some general concerns and specifically those who express suicidal ideation on massive social media platform is indeed a big challenge. Reddit provides a straightforward way to handle this challenge by providing subreddits, or subforums dedicated to specific topics. First, we will discuss the features and methods of extracting data from Reddit. Following above, we will perform the preprocessing steps on the selected data set.

Data Selection

For our project, we require the posts of the users who reveal some general issues like mental or health-related problems. Also, we would like to extract users who directly express suicidal ideation through their posts. For this purpose, we selected subsections like “r/mentalhealth”, “r/depression”, “r/-trauma”, “r/stopselfharm”, “r/survivorsofabuse”, “r/rapecounseling”, “r/socialanxiety” *etc.* for targeting users with general issues (henceforth GI) [2]. Also, we chose subsection “r/suicidewatch” for targeting users who express suicidal thoughts (henceforth SW).

Data Extraction

Reddit provides an API to help people extract users’ posts and comments for a subreddit, but there are some issues with this API. For our project, we focus on the year 2015 and get the user information that would help to map users from GI subsections to the SW subsection. A single post from Reddit contains attributes like ”user_id”, ”author_name”, ”subreddit”, ”post_creation_time” and finally ”post”. However, the Reddit’s official API does not provide the way to extract data for specific users in a given timestamp of a year. To overcome this challenge, we found a third party Rest API called Pushshift. We used the Pushshift API to collect 30,821 users from GI subsections with 44,543,000 posts and 4,822 users from SW subsection with 19,438 posts for 2015. Out of these 30,821 GI users, we found 2,168 users who later subscribed to the SW subsection with 12,000 posts. These users become the ground truth for our evaluation plan. To balance our dataset, we picked only 5,000 GI users from 30,821 users who did not subscribe to SW subsection. These users had total 28,000 posts. Eventually, our dataset comprises of a total 7,168 users with approximately 40,000 raw posts.

Data Preprocessing

The Raw textual dataset needs some amount of preprocessing before using it as an input for any clustering or classification. Given below are the preprocessing steps:

1) Split each user’s posts into sentences and sentences into words using a technique called tokenization.
2) Remove all the punctuations and transform all the words to lowercase.
3) Remove words less than three characters that are irrelevant.
4) Remove all the stops words.
5) Convert all the words in third person format to the first person format and all the verbs in future and past tense to present tense. This process is called Lemmatization.
6) Reduce all the words to their root form. Convert all the words ending with “ing”. This process is called Stemmatization.

Lets consider an example text that would help us identify why those preprocessing steps are crucial before performing any classification algorithm. One of the posts that we extracted looks something like this:

”I’ve been feeling depressed on and off for about 2 years, recently there has been more triggers, my anxiety ticks have come back and the depression comes more often (last 2 weeks, everyday). The depression gets worse every time, I’ve read so many suicide stories, ways to do it etc. but I doubt I would do it but I haven’t got that deep yet.”

After performing tokenization, removing punctuation and words less than three characters, we get the given below post: *”feelings depressed recently anxiety come back depression comes often last everyday depression worse every time read many suicide ways doubt would got deep yet”*

Next we perform lemmatization and stemming and finally we get the cleaned post show below:

”feel depress recent anxiety come back depress come often last everyday depress worse every time read many suicide way doubt will get deep yet”

As you can see, words in the third person are changed to first person and verbs in past or future tense are converted to present. Hence, depressed word is converted to just depress. Also, words are converted to their root form. For example, words like ”comes” and ”feelings” are converted to ”come” and ”feel” respectively. This preprocessing gets rid of redundant words and also combines words that convey same meaning.

IV. Methods

Even the refined sentences from posts cannot be directly utilized to train our model. We need to extract certain features from these documents. Feature extraction is the following step on the processed posts after performing the above-mentioned preprocessing steps on all the raw textual posts. Given a set of processed textual posts, we built a bag of words model, by extracting each word from each sentence and mapping the frequency count to each of those words. Hence, a collection of words in the post will now additionally have the individual word count associated with the existing posts in the dataset. These frequency counts will help in evaluating how many times a word appears in the whole dataset and how important is that word. For our project, we are going to use the topic modeling algorithm called latent dirichlet allocation. This model accepts the bag of words feature as the input. Latent dirichlet allocation is a generative probabilistic model which is used to classify the collection of composites based on the parts used in it to the number of topics considered. A composite is a textual post, and parts are the words or phrases in each of those textual posts. The model creates two matrices after sampling each sentence. Table I represents the probability of selecting each word given the number of topics to be assigned to in the model. The probability values in Topic 0 across all the words in the preprocessed dataset is 1.0. Since there are 4111 words, Table I shows only subset of high probability words. The hyperparameters include the learning decay which is set to 0.9 and we used two topics.

	Topic 0 General Issues	Topic 1 Suicidal Ideation
improve	0.271	0.004
kill	0.004	0.565
life	0.360	0.004
suicide	0.004	0.425
depress	0.004	0.425

TABLE I
The probability of choosing each word given the number of topics

Table II represents the probability of choosing each sentence given the number of topics. In our project, there are two topics/subreddits namely ”General Issues” and ”Suicide Watch”. Latent dirichlet allocation will provide the probabilistic sampling for each post given these two topics. In the context of LDA, the terms *document* and *post* are used interchangeably that represents a user’s post from Reddit dataset. The words corresponding to each document/post shown in Table II are as follows:

1) Document 0: [stress, depress, quit, kill, suicide, die]
2) Document 1: [severe, pain, feel, headache, improve]
3) Document 2: [kill, suicide, die, pain, quit]
4) Document 3: [disease, never, worse, bed, sad, medical]
5) Document 4: [worry, improve, progress, life, therapist]
6) Document 5: [abuse, social, outcast, fear, unworthy, quit]

	Topic 0 General Issues	Topic 1 Suicidal Ideation
Document 0	0.065	0.935
Document 1	0.924	0.076
Document 2	0.246	0.754
Document 3	0.645	0.355
Document 4	0.924	0.076
Document 5	0.065	0.935

TABLE II
The probability of choosing each sentence given the number of topics

A GI user expressing suicidal ideation must have the high probability value assigned to the Suicide Watch topic by the model for the given sentence. While a GI user not expressing suicidal ideation must have high probability value assigned to the general issues topic by the model. Thus, topic modeling performs soft clustering which does not reveal sufficient learning evidence when evaluating the results. To further improve the performance, we further perform another classification algorithm on top of this model called support vector machine. Eventually, training the model using support vector machine will help in better performance evaluation.

After performing the classification algorithm on the test dataset, the confusion matrix obtained will help in further evaluation. For the evaluation plan, we will take the GI users who were classified to express the suicidal ideation and check whether they had subscribed to the SW subsection or not. Similarly, we will get those GI users who were classified as those who do not express the suicidal ideation and check if they don’t exist in the list of SW subscribers.

You can find this project at Bitbucket repository https://bitbucket.org/kush123gevaria/suicidal_ideation_detection_capstone/src/master/.

V. Results

We split the dataset of 40,000 posts into three sets namely training, testing and validation. The training dataset includes 20,000 posts out of which 6000 posts belong to users who later subscribed to SW subreddit. Remaining 14,000 posts belong to just GI subscribers. We trained the latent direchlet allocation model with this dataset using grid search to find the best number of topics that it can extract from this dataset. The parameters of the best LDA model are learning decay of 0.9 and two topics. The best log likely-hood score obtained is -153175.7175 and the model perplexity is 515.2758. The model was able to classify the testing dataset of 6000 posts with 65% accuracy which is not bad, but it does not give any insight or assurity about the classification all the time. The confusion matrix representing the classification resuls for LDA is shown in table III. It provides different results for each training model. To overcome this issue, we additionally used a hard classification algorithm called support vector machine.

	Actual General Issues 4200	Actual Suicidal Ideation 1800
Predicted General Issues 4118	3122	996
Predicted Suicidal Ideation 1882	1078	804

TABLE III
THE CONFUSION MATRIX FOR LDA WITH AN ACCURACY OF 65.04%

Support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis [6]. The input feature for the support vector machine algorithm is a matrix representation of documents v/s words. Here, documents are those 20,000 posts from the training dataset. Each document in the matrix have a vector representation of the words with binary values. A value of one means that particular word exist in that particular document and zero means that word does not exist in the document. There were total 4211 relevant words as feature for each document that were extracted based on certain frequency count. So finally, the SVM model was trained based on this $20,000 \times 4211$ matrix feature. After training the model, same was tested with the testing dataset of 6000 posts and it classified the posts into two subsections SW and GI with accuracy of 75.88%. The results are shown in the confusion matrix IV.

	Actual General Issues 4200	Actual Suicidal Ideation 1800
Predicted General Issues 3937	3345	592
Predicted Suicidal Ideation 2063	855	1208

TABLE IV
THE CONFUSION MATRIX FOR SUPPORT VECTOR MACHINE ALGORITHM WITH AN ACCURACY OF 75.88%

Further we used the LDA results on top of the SVM model to perform feature reduction. Out of the two topics, we extracted top 200 words as the feature list instead of all the 4211 words. We again trained the SVM model on top of these 200 words for each of those topics and found that the accuracy increased by 5% for the available testing dataset. Thus, some irrelevant words that only appeared once in certain documents we removed and the feature space was reduced. Also, the classification was pretty fast as compared to previous version of the SVM model. The confusion matrix for the modified version of the SVM is shown in the table V.

	Actual General Issues 4200	Actual Suicidal Ideation 1800
Predicted General Issues 4088	3563	525
Predicted Suicidal Ideation 1912	637	1275

TABLE V
THE CONFUSION MATRIX FOR SUPPORT VECTOR MACHINE ALGORITHM WITH AN ACCURACY OF 80.63% AFTER FEATURE REDUCTION

VI. CONCLUSION

In this report, we represented a way to identify individuals exhibiting suicidal ideation by targeting their posts and sentiments. We picked Reddit dataset since it provides a better way to organize user posts based on subreddits where we picked subreddits like suicide watch and other general issue subreddits. Users that belong to suicide watch subreddit are the ones expressing suicidal thoughts through social media. Thus, using the posts of those users, we identified other users who express general issues and later on express to show suicidal ideation through their posts. We came up with data preprocessing steps performed on the raw posts of the Reddit dataset. After performing the preprocessing steps we prepared the feature for the topic modeling algorithm called latent direchlet allocation. Since it is a probabilistic model and also gives different results, we then used support vector machine classification algorithm that gave better accuracy and eventually we used the topic modeling algorithm on top of the support vector machine algorithm to improve the accuracy further. Eventually, we were able to achieve the goal meant for this project of identifying users exhibiting suicidal ideation, but still this does not implicitly means that the users classified to have suicidal thoughts are actually expressing the same. We can use this model to approach those users for some kind of help that would eventually help them to improve their current status and mind set.

VII. ACKNOWLEDGEMENT

I would like to thank my advisor, Professor Christopher Homan for proper guidance and support during the entire process. I was constantly reminded with constructive feedback that helped me embracing this project with quality.

REFERENCES

[1] S. Farrelly, J. Kruchten, P. Williams *et al.*, “The link between mental health-related discrimination and suicidality: service user perspectives,” *Psychological Medicine*, vol. 45, pp. 2013–2022, 2015.

[2] M. Choudhury, E. Kiciman, M. Dredze *et al.*, “Discovering shifts to suicidal ideation from mental health content in social media,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2016, pp. 2098–2110.

[3] C. Chung and J. Pennebaker, “The psychological functions of function words,” *Social Communication*, vol. 1, pp. 343–359, 2007.

[4] S. Guntuku, D. Yaden, M. Kern *et al.*, “Detecting depression and mental illness on social media: an integrative review,” *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, 2017.

[5] J. Neter, M. Kutner, C. Nachtsheim *et al.*, *Applied linear statistical models*, 1996, vol. 4.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[7] B. Leite, V. Amorim, A. Silva *et al.*, "The influence of social networks in suicidal behavior," *International Archives of Medicine*, vol. 8, 2015.

[8] K. Hawton, K. Saunders, and R. O'Connor, "Self-harm and suicide in adolescents," *The Lancet*, vol. 379, pp. 2373–2382, 2012.

[9] S. Dalglish, M. Melchior, N. Younes *et al.*, "Work characteristics and suicidal ideation in young adults in france," *Social Psychiatry and Psychiatric Epidemiology*, vol. 50, pp. 613–620, 2015.

[10] D. Trout, "The role of social isolation in suicide," *Suicide and Life-Threatening Behavior*, vol. 10, pp. 10–23, 1980.

[11] Q. Cheng, C. Kwok, T. Zhu *et al.*, "Suicide communication on social media and its psychological mechanisms: an examination of chinese microblog users," *International Journal of Environmental Research and Public Health*, vol. 12, pp. 11 506–11 527, 2015.

[12] Q. Cheng, S.-S. Chang, and P. Yip, "Opportunities and challenges of online data collection for suicide prevention," *The Lancet*, vol. 379, pp. 53–54, 2012.

[13] P. Burnap, W. Colombo, and J. Scourfield, "Machine classification and analysis of suicide-related communication on twitter," in *Proceedings of the 26th ACM conference on hypertext & social media*. Association for Computing Machinery, 2015, pp. 75–84.

[14] L. Biddle, J. Donovan, K. Hawton *et al.*, "Suicide and the internet," *Bmj*, vol. 336, pp. 800–802, 2008.

[15] A. Abboute, Y. Boudjeriou, G. Entringer *et al.*, "Mining twitter for suicide prevention," in *International Conference on Applications of Natural Language to Data Bases/Information Systems*. Springer, 2014, pp. 250–253.