Using Machine Learning Algorithms to Detect Suicide Risk Factors on Twitter

Samah J. Fodeh
Department of Emergency Medicine &
Yale Center for Medical Informatics
Yale University
New Haven, USA
samah.fodeh@yale.edu

Taihua Li
College of Computing and Digital Media
DePaul University
Chicago, USA
tli35@mail.depaul.edu

Kevin S. Menczynski
College of Computing and Digital Media
DePaul University
Chicago, USA
kmenczy1@mail.depaul.edu

Todd M. Burgette
College of Computing and Digital Media
DePaul University
Chicago, USA
tburget1@mail.depaul.edu

Andrew K. Harris
College of Computing and Digital Media
DePaul University
Chicago, IL
gharr107@mail.depaul.edu

Georgeta I. Ilie
College of Computing and Digital Media
DePaul University
Chicago, IL

Satyan K. Rao
College of Computing and Digital Media
DePaul University
Chicago, IL
srao8@mail.depaul.edu

Jonathan F. Gemmell
College of Computing and Digital Media
DePaul University
Chicago, IL
jgemmell@cdm.depaul.edu

Daniela S. Raicu
College of Computing and Digital Media
DePaul University
Chicago, IL
draicu@cdm.depaul.edu

***Abstract*: The goal from this study is to identify suicide risk factors on Twitter. We propose a machine learning framework that could be potentially useful for suicide prevention interventions. We applied search terms from the suicidal ideation tracking framework proposed by Jashinsky et al. and downloaded 12,066 public tweets from 3,873 users via Twitter's application programming interface (API). We created "HighRisk" or "AtRisk" labels for users based on their suicidal ideation terms' usage and applied three topic discovery algorithms to find underlying suicide risk factors among users, which were subsequently used to classify users into "HighRisk" or "AtRisk". Algorithms applied included Latent Semantic Analysis, Latent Dirichlet Allocation, Non-negative Matrix Factorization, Decision Tree and K-means Clustering. Our topic discovery approach detected 7 out of 12 suicide risk factors proposed by Jashinsky et al. Using a decision tree classification model that utilized these factors, we achieved 0.844 in precision, 0.912 in sensitivity, and 0.829 in specificity in classifying users into "HighRisk" and "AtRisk" groups. The development of this framework supplements suicide researchers and suicide prevention efforts, with a potential to be employed at run-time.**

Keywords- Suicide ideation, Topic modeling, Text analysis, Latent Semantic Analysis

1. Introduction

Suicide is the 10th leading cause of death in the United States with an estimate cost of $51-billion annually [1], making suicide prevention not only a public health issue, but also an economic one. Center for Disease Control and Prevention (CDC) reported that for youth between the ages of 10 and 24, suicide is the third leading cause of death. [1] Even more concerning, the results of a recent study on 32 U.S. children's hospitals which show that rates of suicide and serious self-harm in children and adolescents have increased steadily from 2008 to 2015. [2] Social media has been identified as playing a possible role in contributing to suicide through copycat actions, mainly in vulnerable and impressionable youth [3]. Twitter is a social network platform where users share messages limited to 140 characters. It has been shown that 21% of Americans use Twitter, with 36% of them age between 18 and 29 [4]. Twitter has been known to be used as a platform for suicidal messages and suicide notes [5]. Social media platforms have even been used as venues to live-stream suicide attempts showing that these warning signs need to be taken seriously [6]. Twitter has recognized this serious risk and has put a service in place for people who are, or know of somebody who is, suicidal to reach out and get help [7]. However, Twitter is not proactive in identifying users at risk, and reporting is at the discretion of the user and not in real time. To make suicide prevention timely and effective, the suicide-related data has to be collected, analyzed and reported in a timely manner, so that interventions can be made before the person commits suicide.

Prior research has identified specific terms or phrases in tweets indicative of suicide risk factors [5]. While most research has focused on using machine learning in conjunction with human annotators and/or suicide research experts, our study dropped the human factor to reduce cost and utilized machine learning approaches to increase efficiency in identifying users at high risk of suicide. For example, in [5], human annotators determined the suicide risk factors and linked them to certain terms and phrases in the tweets. In our previous work [8, 9], we utilized the suicide risk factor framework by [5] to detect relational features and language patterns indicative of suicidal ideation. In this study, we applied reproducible machine learning algorithms including Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) to identify suicide related topics and themes of discussion in tweets and studied the user cluster formations and user profile classifications.

2. Related Work

The study of Twitter data in suicide analysis has become more prevalent as there are known issues with recall and context bias in the psychological assessments that have been studied in the past [10]. There is a lack of textual data for review in traditional analysis of suicide and the review of Twitter data offers an insight into the day-to-day feelings of individual users [11].

2375-9259/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDMW.2019.00137

This analysis seeks to find quantifiable signals of suicide, ways to visualize linguistic data, automatic emotion classification of tweets to help create a possible map of emotions leading up to a suicide attempt, and looks to incorporate results into current work in the field of psychology [11]. The overarching goal of the research is to create a tangible screening process by which users could be analyzed in a way that is non-invasive and cost effective to different research groups [11]. As social media has become a focus of suicide detection and prevention, various effective techniques have emerged to perform this analysis. Linguistic Inquiry and Word Count (LIWC) has been used to analyze single users for both classification and time-series analysis [12, 13]. These analyses were centered around the linguistic characteristics of users' tweets, leveraging previously categorized types of words (pronouns, past tense, positive emotions, etc.).

Logistic regression and linguistic analysis have been done to identify suicidal users using LIWC analysis to pre-process data [14]. O'Dea's research has led to expanded feature sets from purely word counts into other high-level semantic features, such as the number of total words and the number of pronouns in each post. The goal of this work was to identify different populations of suicidal users with different levels of suicide risk and distinguish linguistic profiles between strongly concerning users. Strongly concerning suicide-related tweets have a higher word count, higher usage of the words defined in the library, increased use of first-person singular pronouns, and increased references to time and death. It was also noted that this group of tweets generally have fewer user-tags, indicating an isolation in the Twitter social network [14]. Similar analysis has been conducted looking to define a level of concern for different Twitter users through the use of a Support Vector Machine (SVM) classifier on word frequency with Term Frequency-Inverse Document Frequency (TF-IDF) transformations performed on the data [15]. SVM has been used to define tweets that warrant further investigation or review. During testing, SVM and the human coders achieved the same level of accuracy. When performing this type of classification, an identified challenge is that it can be difficult to determine if a tweet is sarcasm or requires actual intervention [15]. Despite this issue, this type of study gained validation when researchers were able to identify 53 users who ended up attempting to attempting suicide based on their analysis on a Chinese social media site similar to Twitter [16]. A previous research has tried to obtain labels for suicide-related tweets through crowdsourcing [17]. This is one of the most difficult tasks that researchers are faced with when studying suicide, as there are not definitive labels to be used for classification. Crowdsourcing is the most common solution for generating labels, but they are still subjected to the reviewers' bias. In a different study, authors acquired confirmed suicide cases by scanning traditional media outlets and then traced back the deceased user's' tweets [18], but this is still not a perfect science as Twitter account information cannot guarantee the identity of the user. Another approach that has been undertaken is performing topic identification, which is the concept of identifying broader ideas or sentiments across multiple tweets and users. Rather than identifying users who have specific risk levels, this type of analysis seeks to describe groups of users in ways that can aid in feature extraction. Latent Dirichlet Allocation (LDA) is one technique that has proven successful for topic mining in Twitter data [19]. LDA along with Latent Semantic Analysis (LSA) can also be used to compare similarities between words and sentences, which can help identify word vectors and allow comparison across different suicidal users at different risk levels [20]. These techniques also provide probabilities (LDA) and loadings (LSA) for each word as it contributes to each topic. These are unsupervised methods, but can also be used for classifying users into one of the discovered topics [20].

3. Data

3.1 Twitter Data Collection

Previous suicide-related research has shown that people who are at risk of committing suicide can be detected and tracked using the twelve-suicide risk factor related terms [5]. Using these search terms, we collected tweets posted between January 1th, 2015 until June 8th, 2016, from 3,873 unique users.

3.2 Data Preprocessing

Processed tweets were filtered based on the suicide risk factors and associated language, which were then used to create a user-term frequency matrix. Since most of the tweets contained a keyword indicative of the feeling or behavior in combination with a descriptor to complete the meaning, we filtered the keywords into 48 representative terms, as shown in *Table 1*. After filtration, there were 12,066 tweets left and used to build a final user-term frequency matrix.

Table 1: Terms Used in The User-Term Matrix

Abused	Prozac	Cut	Panic Disorder
Depressed	Pills	Bully	Social Anxiety
Hopeless	Suicide Abused	Bullied	Fight Dad
Worthless	Suicide Pain	Suicide Gun	Fight Mom
Empty	Suicide Tried	Suicide Shoot	Fight Parents
Anxious	Suicide Mom	Schizophrenia	Fight Sister
Sleeping	Suicide Sister	Anorexia	Fight Brother
Irritable	Suicide Brother	Bulimia	Argue Dad
Restless	Suicide Friend	OCD	Argue Mom
Alcohol	Suicide Thought	Bipolar	Argue Parents
Sertraline	Suicide Kill	PTSD	Impulsive
Zoloft	Suicide Think	Borderline Personality	Suicide Before

3.3 User Labeling: "HighRisk" and "AtRisk"

Users were labeled and divided into two groups, "HighRisk" and "AtRisk" following [8]. "HighRisk" users were those whose tweets contained language pertaining to suicide related behavior, and other users were annotated as "AtRisk". If a user did not have any risk factor term from *Table 1*, he or she was removed. In the filtered dataset, there were 280 "HighRisk" users and 1,614 "AtRisk" users.

3.4 Addressing Class Imbalance

To deal with the class imbalance issue (minimal number of "HighRisk" users), two additional balanced data-sets were created using random down-sampling and *K*-means clustering. The first balanced dataset contained all the 280 "HighRisk" users and 280 randomly-selected "AtRisk" users. The second balanced dataset contained 280 "HighRisk" users and 285 "AtRisk" users, which were selected from 15 clusters found using *K*-means clustering; for each cluster, we selected 19 most representative users.

4. Methodology

We first applied three topic clustering algorithms in this study to discover topics discussed by users on Twitter, including Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), and Non-Negative Matrix Factorization (NMF). Second, we explored cluster formations and user classifications into "HighRisk" or "AtRisk" using the identified topics. LDA is a probabilistic approach that maximizes the log likelihood of each term appearing in each topic, while LSA is a matrix factorization approach that rotates topic vectors in the term space to best capture the variation of terms. However, LSA produces topics which contain terms that are negatively correlated with those topics, and such results are hard to be interpreted. To account for this disadvantage, NMF, a non-negative rank factorization approach, is performed to ensure the topics are better in being interpretable and well separated.

In order to classify users into either "HighRisk" or "AtRisk", two machine learning approaches were employed: Decision Tree classification and *K*-means clustering. While Decision Tree is a supervised approach that learns to classify instances based on relations between the data points and corresponding ground truth labels, K-means is an unsupervised approach that partitions users based on their similarities Since the ground truth labels are unknown when such applications are deployed in real-world, we applied *K*-means clustering to examine the effectiveness of using topics discovered in separating "HighRisk" and "AtRisk" users.

5. Experiments and Results

5.1 Topic Identification

Based on the framework proposed in [5], three topic identification algorithms were applied with the assumption that there are 12 underlying suicide related topics, corresponding to the 12 suicide risk factors. As shown in *Table 2*, each topic identified from results of LDA, LSA, and NMF is associated with one of the 12 risk factors. To determine the topic definition using one of the twelve suicide risk factors, a threshold of 0.25 is used on terms' weights associated with the topics. For example, in the first topic discovered by LDA, *sleeping* and *cut* had loadings that are greater than the threshold, and since the term *sleeping* has the greatest loading value, the topic is assigned with *Depressive feelings*, according to the framework in [5]. Furthermore, topics that are assigned *Self-harm* and *Bullying* must have dominating terms such as *Cut* and *Bullied*, respectively. By comparing the suicide risk factor assignments across three algorithms, 5 topics are found consistent in the results, including *Depressive Feelings*, *Drug Abuse*, *Psychological Disorders*, *Self-harm*, and *Bullying*. *Depressive Symptoms* appears to be another topic found by LSA and NMF, and the top contributing terms for that topic include *sleeping*, *alcohol*, and *empty*. In addition, *Family Violence/Discord* is discovered by LSA. In total, using the three topic identification algorithms, 7 out of 12 suicide risk factors were identified.

5.2 User Classification

After confirming the definitions of topics, the user-term matrix *A* was transformed into the user-topic matrix, where Decision Tree and *K*-means Clustering were applied to classify users.

5.2.1 Decision Tree

Decision Tree models were built using the Chi-square Automatic Interaction Detector (CHAID) algorithm with gain ratio as the splitting criterion, which produces a more balanced tree that is less likely to be overfitting than using other splitting criteria. With these parameter settings, number of minimum memberships in child and parent nodes were varied in experiments, where the restriction in the number of members in child nodes is half of that in the parent nodes, and the parent nodes membership restriction is varied between the range of 10 and 180.

5.2.2 *K*-means Clustering

In total, 54 *K*-means Clustering experiments were performed. For each of the dataset prepared, two *K*-means Clustering analyses are conducted for each of the similarity measures. Two assumptions were made: the clusters are well separated, and the clusters can be partitioned into two groups and be labeled "AtRisk" and "HighRisk". Therefore, *K*=2 is chosen for the clustering analysis. Second, the clusters are disjointed and in order to separate all groups and label each of them into "AtRisk" or "HighRisk", *K*=N is chosen for the clustering analysis, where N is determined by the number of terminal nodes in the corresponding Decision Tree. For example, for the LDA unbalanced dataset, there are 27 terminal nodes in the best decision tree trained, therefore, the *K*-means clustering is conducted with *K*=27.

5.3 Classification Results

After clusters are formed, the cluster label is based on the majority class label within the cluster. For example, if a cluster is dominated by "AtRisk" users, that cluster will be labeled as "AtRisk" and all users will be assigned the label "AtRisk" as predicted labels. Four metrics are used to evaluate the performances of decision tree and *K*-means clustering, and they are precision, sensitivity (recall), specificity, and AUC; to calculate each of these metrics, "HighRisk" class is treated as the positive class.

As shown in *Table 3*, in general, Decision Tree models perform better than *K*-means Clustering in identifying "HighRisk" users, measured by precision. Using the unbalanced dataset,

Table 2: Topic Clustering Results

	LDA Topics	Top Contributing Terms with Probabilities
1	Depressive Feelings	Sleeping (0.605), Cut (0.347)
2	Drug Abuse	Zoloft (0.328), Prozac (0.298)
3	Psychological Disorders	Panic Disorder (0.411)
4	Self-harm	Cut (0.31)
5	Bullying	Bullying (0.796)
6	Self-harm	Cut (0.395)
7	Self-harm	Cut (0.354), Depressed (0.324)
8	Drug Abuse	Alcohol (0.768)
9	Depressive Feelings	Empty (0.964), Worthless (0.375)
10	Drug Abuse	Pills (0.747)
11	Self-harm	Cut (0.995)
12	Family Violence/Discord	Suicide (0.295)
	LSA Topics with Variance Explained	**Top Contributing Terms with Loadings**
1	Self-harm (0.463)	Cut (0.995)
2	Bullying (0.353)	Bully (0.988)
3	Drug Abuse (0.031)	Zoloft (0.56), Alcohol (0.475), Prozac (0.372)
4	Psychological Disorders (0.023)	Panic Disorder (0.825)
5	Bullying (0.021)	Bullied (0.978)
6	Depressive Symptoms (0.017)	Sleeping (0.54), Alcohol (0.438), Empty (0.417)
7	Depressive Symptoms (0.015)	Sleeping (0.698)
8	Drug Abuse (0.014)	Pills (0.75)
9	Drug Abuse (0.012)	Pills (0.533), Empty (0.33)
10	Depressive Feelings (0.012)	Empty (0.687), Depressed (0.415)
11	Bullying (0.012)	Abused (0.997)
12	Psychological Disorders (0.008)	Bipolar (0.65), Schizophrenia (0.442)
	NMF Topics	**Top Contributing Terms with Loadings**
1	Self-harm	Cut (2.22)
2	Psychological Disorders	Panic Disorder (0.89)
3	Drug Abuse	Alcohol (0.651)
4	Bullying	Bully (2.275)
5	Depressive Symptoms	Sleeping (0.961)
6	Drug Abuse	Pills (2.154)
7	Depressive Feelings	Empty (2.002)
8	Depressive Feelings	Worthless (2.106), Bullied (0.77)
9	Depressive Feelings	Depressed (1.143)
10	Psychological Disorders	Bipolar (0.572)
11	Depressive Feelings	Anxious (1.513), Hopeless (0.421)
12	Depressive Feelings	Abused (1.246)

neither Decision Tree nor *K*-means Clustering is able to clearly distinguish between “AtRisk” and “HighRisk” users.

Decision Tree tends to classify most of users into “High-Risk”, and therefore, all Decision Tree models built on the unbalanced dataset have sensitivity close to one while the specificity is low. On the other hand, *K*-means Clustering tends to assign all users into the “AtRisk” class, and therefore, the *K*-means Clustering results have zero sensitivity and a specificity of one. Using the balanced dataset whose “AtRisk” users were randomly sampled, Decision Tree models tend to perform better than *K*-means Clustering as well. *K*-means Clustering with *K*=N, where N is determined as described in Section 5.2.2, can capture more "HighRisk" users than that of the Decision Tree model; the *K*-means Clustering result has a sensitivity of 0.771 and the Decision Tree model has a sensitivity of 0.736. Overall, both *K*-means clustering, and decision tree perform better using the balanced dataset generated by the *K*-means approach compared to those using the other datasets. Decision Tree outperforms *K*-means Clustering in precision, sensitivity, specificity and AUC. However, using LDA’s balanced dataset with *K*-means approach, *K*-means Clustering is able to separate the “AtRisk” user better than the Decision Tree; *K*-means Clustering result has a specificity of 0.86 compared to that of the Decision Tree which is 0.829. In addition, *K*-means Clustering is achieving a precision (0.828) that is close to that of the Decision Tree model (0.844).

K-means Clustering results show that the “AtRisk” and “HighRisk” user clusters are disjointed in the risk factor space and can be properly separated using the clustering approach, and specifically, the clustering approach can almost perfectly separate the “AtRisk” users with 0.993 specificity. Furthermore, such disjoint clusters can be better separated using the supervised approach, which in this study is the Decision Tree model. Among all the Decision Tree models, using NMF’s balanced dataset generated using the *K*-means Clustering approach, it has the best performance of all, where the precision is 0.853, the sensitivity is 0.933, the specificity is 0.836, and the AUC is 0.885.

6. Discussion

The results have indicated that by utilizing the supervised and unsupervised machine learning algorithms combined with topic identification techniques, users who are “AtRisk” and “High-Risk” of suicidal ideation can be identified using their Twitter data. Using the topic clustering techniques, the result also shows that with minimum human interpretation, 7 out of the 12 suicide risk factors confirmed by suicide researchers were discovered. Without looking at any additional linguistic features of tweets, decision tree models are able to distinguish the “AtRisk” and “HighRisk” users. Additionally, the classification results in combination with the topic clustering provide qualitative interpretation of the users’ psychological state which could be potentially useful in future prevention efforts. Among the three topic discovery approaches, Latent Semantic Analysis is able to provide information on the dominating topics or the popularly used suicide-related terms. In *Table 2*, the variance captured by, or the popularity of each topic is reported next to the topic definition. For example, topic 1, which is "Self-harm" (0.463), indicates it’s a self-harm related topic and it captures 0.463 of the variance of the original dataset, and its dominating term is “Cut” which has a weight of 0.994. By observing these variances and term weights, the LSA results show that “Self-harm” is the dominating topic in the dataset, followed by “Bullying” as the second mostly discussed topic. Between these two topics, “Cut” and “Bully” are the most commonly used terms. While LDA and NMF do not produce ordered topics with a parametric approach, the popularity of topics discussed can be measured by the frequency of the topic discovered. Since LDA is a generative probabilistic approach, and Self-harm appear in 4 out of 12 topics discovered, which has the highest frequency compared to other topics, it can be argued that Self-harm is the most probable topic being discussed in the dataset. Among NMF’s topics discovered, Depressive Feelings is the dominating topic. Such results, however, do not provide insights on what topics are important in distinguishing between “AtRisk” and “HighRisk” users. Such information can be observed from decision tree plots.

Among the classification experiments, decision tree models built using balanced datasets generated using the *K*-means approach performed the best. Therefore, conclusions concerning the importance of topics shall be drawn from those trees. In a Decision Tree model, the topics that are used to first partition the users are considered to be the most important topics. Observing three Decision Tree models’ top nodes, LDA’s most important topics are *Self-harm*, *Depressive feelings*, *Drug Abuse*, and *Bullying*, LSA’s top important topics are *Drug Abuse*, *Self-harm*, *Psychological Disorders*, *Bullying*, and *Depressive Symptoms*, and NMF’s top important topics are *Drug Abuse*, *Psychological Disorders*, *Self-harm*, and *Drug Abuse*. Overall, it can be concluded that *Self-harm*, *Drug Abuse*, *Bullying*, *Depressive Feelings*, *Depressive Symptoms*, and *Psychological Disorders* play important roles in determining “AtRisk” and “HighRisk” users. Furthermore, since the ground truth labels are usually unknown in real-world applications, and with a specificity of 0.993 performance by applying *K*-means clustering on topics identified, we believe such suicidal ideation detection technique can be studied further and eventually deployed as a real-time application to accomplish in-time suicide prevention work.

7. Limitations

The suicide risk factor framework designed by [5] was based on previous research ranging from 1994 to 2012. With the evolving language usage on social media, an up-to-date suicide-related lexicon should be considered for this type of framework development and incorporated into the suicide detection and prevention research. This is indicative that in the dataset used in this study, there might be missing information that conveys the ideation of the other 5 suicide risk topics. Furthermore, the use of emoticons, hashtags, and other twitter meta-data that could potentially indicate suicide ideation are not included in the dataset. In addition, there is an unsolved issue in the selection of tweets and users who are displaying sarcasm or making statements in jest compared to users who have actual suicidal intent. As advancements are made to natural language processing techniques, this piece of the study can be improved. This study is evaluated based on the previously established framework in suicide-related research using Twitter data instead of the ground truth: whether the Twitter user committed suicide. For this work to be applied in the suicide prevention domain, ground truth data should be collected and used to evaluate the true performance of this study. Other possibilities could include working to determine the number of different classes within suicidal users. This study operates off of the "AtRisk" and "HighRisk" structure and could miss other insights.

8. Conclusion

Suicide is a serious social and economic problem in the United States. Many efforts have been made in studying the language formation of suicide-related tweets and a few are made to detect suicidal ideation using open social media data. In the most recent study in achieving such task, [5, 18] tweets are manually tagged by human annotators before machine learning algorithms are applied to classify if users are at risk of committing suicide, which is not efficient enough to detect suicidal ideation to support suicide prevention. In this study, we propose a suicidal ideation detection framework that requires minimum human efforts in annotating data by incorporating unsupervised topic discovery algorithms. Three techniques are tested in this study, including Latent Semantic Analysis, Latent Dirichlet Allocation, and Non-Negative Matrix Factorization. Using these algorithms, we were able to discover 7 out of 12 suicide risk factors proposed by [5], and using those topics, we were able to represent of user profiles in a more compact format using topics. Furthermore, by conducting K-means clustering analysis on the transformed datasets, we concluded that "AtRisk" and "HighRisk" user groups are disjointed and cannot be well distinguished by partitioning them into clusters. However, as shown in the experimental results using Decision Tree, we were able to achieve 0.844 of precision, 0.912 of sensitivity, and 0.829 of specificity, where "HighRisk" users are the positive class. This framework shows that with minimal human interpretation of social media data, it is possible to detect suicidal ideation using the combination of supervised and unsupervised machine learning algorithms.

Table 3: User Classification Result

Abbreviations for Balanced Sets: R: Random Sampled K: *K*-means Sampled		***K*-Means**						**Decision Trees**		
		$K = 2$			$K = \mathrm{N}$			-		
		Unbalanced	Balanced (R)	Balanced (K)	Unbalanced	Balanced (R)	Balanced (K)	Unbalanced	Balanced (R)	Balanced (K)
LSA	Precision	0.000	0.511	0.546	0.000	0.631	0.647	0.865	0.747	0.845
	Sensitivity	0.000	0.886	0.739	0.000	0.689	0.746	0.993	0.750	0.839
	Specificity	1.000	0.154	0.396	1.000	0.596	0.600	0.111	0.746	0.843
	AUC	0.500	0.520	0.568	0.500	0.643	0.673	0.552	0.748	0.841
LDA	Precision	0.000	0.538	0.648	0.686	0.554	0.828	0.881	0.640	0.844
	Sensitivity	0.000	0.432	0.571	0.086	0.771	0.686	0.979	0.736	0.912
	Specificity	1.000	0.629	0.695	0.993	0.379	0.860	0.239	0.586	0.829
	AUC	0.500	0.530	0.633	0.539	0.575	0.773	0.609	0.661	0.870
NMF	Precision	0.000	0.520	0.606	0.556	0.575	0.734	0.873	0.711	0.853
	Sensitivity	0.000	0.882	0.604	0.018	0.661	0.700	0.984	0.739	0.933
	Specificity	1.000	0.186	0.614	0.998	0.511	0.751	0.171	0.700	0.836
	AUC	0.500	0.534	0.609	0.508	0.586	0.725	0.578	0.720	0.885

References

1. Center for Disease Control and Prevention, "Suicide Data Sheet," Center for Disease Control and Prevention, 1 January 2015. Available: https://www.cdc.gov/violenceprevention/pdf/suicide-datasheet-a.pdf. (accessed 26 December 2017).
2. G. Plemmons, M. Hall and W. Browning, et al. "Trends in Suicidality and Serious Self-Harm for Children 5-17 Years at 32 U.S. Children's Hospitals, 2008-2015," in Pediatric Academic Societies 2017, Toronto, 2017.
3. D. Luxton, J. D. June and J. M. Fairall, "Social media and suicide: a public health perspective," American journal of public health, vol. 102, no. S2, PS195--S200, 2012.
4. Pew Research Center, "24% of online adults (21% of all Americans) use Twitter," 10 November 2016. Available: http://www.pewinternet.org/2016/11/11/social-media-update-2016/pi_2016-11-11_social-media-update_0-04/. (accessed 26 December 2017).
5. J. Jashinsky, S. H. Burton, and C. L. Hanson, et al. "Tracking suicide risk factors through Twitter in the US," Crisis, vol. 35, no. 1, P51-9, 2014.
6. B. Stelter, "Web Suicide Viewed Live and Reaction Spur a Debate," 24 November 2008. Available: http://www.nytimes.com/2008/11/25/us/25suicides.html. (accessed 26 December 2017).
7. Twitter, "About self-harm and suicide," 1 January 2017. Available: https://support.twitter.com/articles/20170313. (accessed 26 December 2017).
8. S. Fodeh, J. Goulet, and C. Brandt, et al. "Leveraging Twitter to better identify suicide risk," in Machine Learning Rearch, Halifax, 2017.
9. R. Grant, D. Kucher and A. León, et al. "Automatic Extraction of Informal Topics from Online Suicidal Ideation," in 11th International Workshop on Data and Text Mining in Biomedical Informatics, Singapore, 2017.
10. S. Shiffman, A. A. Stone and M. R. Hufford, "Ecological momentary assessment," Annual Review of Clinical Psychology, vol. 4, no. 1, P1-32, 2008.
11. G. Coppersmith, K. Ngo and R. Leary, et al. "Exploratory analysis of social media prior to a suicide attempt," in Third Workshop on Computational Linguistics and Clinical Psychology, San Diego, 2016.
12. J. F. Gunn and D. Lester, "Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death," Suicidologi, vol. 17, no. 3, P28-30, 2012.
13. S. R. Braithwaite, C. Giraud-Carrier and J. West, et al. "Validating machine learning algorithms for Twitter data against established measures of suicidality," JMIR mental health, vol. 3, no. 2, P e21, 2016.
14. B. O'Dea, M. E. Larsen and P. J. Batterham, et al. "A linguistic analysis of suicide-related Twitter posts," Crisis, vol. 38, no. 5, P319-329, 2017.
15. B. O'Dea, S. Wan and P. J. Batterham, et al. "Detecting suicidality on Twitter," Internet Interventions, vol. 2, no. 2, P183-188, 2015.
16. X. Huang, L. Zhang and D. Chiu, et al. "Detecting suicidal ideation in Chinese microblogs with psychological lexicons," in Ubiquitous Intelligence and Computing, 2014 IEEE 11th International Conference on and IEEE 11th International Conference on and Autonomic and Trusted Computing, and IEEE 14th International Conference on Scalable Computing and Communications and Its Associated Workshops (UTC-ATC-ScalCom), Bali, 2014.
17. T. Liu, Q. Cheng and C. M. Homan, et al. "Learning from various labeling strategies for suicide-related messages on social media: An experimental study," CoRR, vol. abs/1701.08796, no. abs/1701.08796, P1, 2017.
18. P. Burnap, W. Colombo and J. Scourfield, "Machine classification and analysis of suicide-related communication on twitter," in 26th ACM Conference on Hypertext & Social Media, New York, 2015.
19. K. D. Rosa, R. Shah and B. Lin, et al. "Topical clustering of tweets," in 34th international ACM SIGIR conference on Research and development in Information Retrieval, Beijing, 2011.
20. V. Rus, N. Niraula and R. Banjade, "Similarity measures based on latent dirichlet allocation," in International Conference on Intelligent Text Processing and Computational Linguistics, Samos, 2013.