A Computational Approach to Feature Extraction for Identification of Suicidal Ideation in Tweets

Ramit Sawhney **Prachi Manchanda** **Raj Singh** **Swati Aggarwal**
Netaji Subhas Institute of Technology, New Delhi 110078, India
{ramits, prachim, rajs}.co@nsit.net.in swati1178@gmail.com

Abstract

Technological advancements in the World Wide Web and social networks in particular coupled with an increase in social media usage has led to a positive correlation between the exhibition of Suicidal ideation on websites such as Twitter and cases of suicide. This paper proposes a novel supervised approach for detecting suicidal ideation in content on Twitter. A set of features is proposed for training both linear and ensemble classifiers over a dataset of manually annotated tweets. The performance of the proposed methodology is compared against four baselines that utilize varying approaches to validate its utility. The results are finally summarized by reflecting on the effect of the inclusion of the proposed features one by one for suicidal ideation detection.

1 Introduction

According to World Health Organization, suicide is the second leading cause of death among 15-29-year-olds across the world. In fact, close to 800,000 people die due of suicide each year. The number of people who attempt suicide is much higher. While an individual suicide is often a solitary act, it can often have a devastating impact on families (Cerel et al., 2008). Many suicide deaths are preventable and it is important to understand the ways in which individuals communicate their depression and thoughts for preventing such deaths. (Sher, 2004) Suicide prevention is mainly hinged on surveillance and monitoring of suicide attempts and self-harm tendencies.

The younger generation has started to turn to the Internet (Chan and Fang, 2007) for seeking help, discussing depression and suicide-related information and offering support. The availability of suicide-related material on the Internet plays an important role in the process of suicide ideation. Due to this increasing availability of content on social media websites (such as Twitter, Facebook and Reddit etc.), and blogs (Yates et al., 2017) there is an urgent need to identify affected individuals and offer help. Suicidal ideation refers to thoughts of killing oneself or planning suicide, while suicidal behavior is often defined to include all possible acts of self-harm with the intention of causing death (Costello et al., 2002). Although Twitter provides a unique opportunity to identify at-risk of individuals (Jashinsky et al., 2014) and a possible avenue for intervention at both the individual and social level, there exist no best practices for suicide prevention using social media.

While there is a developing body of literature on the topic of identifying patterns in the language used on social media that expresses suicidal ideation (De Choudhury et al., 2016), very few attempts have been made to employ feature extraction methods for binary classifiers that separate text related to suicide from text that clearly indicates the author exhibiting suicidal intent. A number of successful models (Yates et al., 2017) have been used for sentence level classification, however, ones that are successful for being able to learn to separate suicidal ideation from depression as well as less worrying content such as reporting of a suicide, memorial, campaigning, and support. etc, require a greater analysis to select more specific features and methods to build an accurate and robust model. The drastic impact that suicide has on surrounding community coupled with the lack of specific feature extraction and classification models for the identification of suicidal ideation on social media, so that action can be taken is the driving motivation for the work presented in this paper.

Suicide prevention by suicide detection (Zung, 1979) is one of the most effective ways to drasti-

Proceedings of ACL 2018, Student Research Workshop, pages 91–98
Melbourne, Australia, July 15 - 20, 2018.

cally reduce suicidal rates. The major practical application of this work lies in it's easy adaptability to any social media forum (Robinson et al., 2016), wherein it can be used directly for analyzing text-based content posted by its users and flag it if the content is concerning.

The main contributions of this paper can be summarized as follows:

1. The creation of a labeled dataset for learning the patterns in tweets exhibiting suicidal ideation by manual annotation.

2. Proposed a set of features to be fed into classifiers to improve the performance.

3. Employed four binary classifiers with the proposed set of features and compared them against baselines utilizing varied approaches to validate the proposed methodology.

2 Related Work

Media communication can have both positive and negative influence on suicidal ideation. A systematic review of all articles in PsycINFO, MEDLINE, EMBASE, Scopus, and CINAH from 1991 to 2011 for language constructs relating to self-harm or suicide by Daine et al. (2013) concluded that internet may be used as an intervention tool for vulnerable individuals under the age of 25. However, not all language constructs containing the word suicide indicate suicidal intent, specific semantic constructs may be used for predicting whether a sentence implies self-harm tendencies or not.

A suicide note analysis method for automating the identification of suicidal ideation was built using binary support vector machine classifiers by Desmet and Hoste (2013) using fine-grained emotion detection for classifier optimization with lexico-semantic features for optimization. In 2014, Huang et al. (2014) used rule-based methods with hand-crafted unsupervised classification for developing a real-time suicidal ideation detection system deployed over Weibo[1], a microblogging platform. This approach differs from the proposed approach in terms of both features and the reach of the social media platforms. Topic modeling in Chinese microblogs (Huang et al., 2015) for suicide ideation detection has also proven to be efficient, however for a limited subset with a fairly different set of features.

Studies corresponding to rise in suicidal ideation associated with specific temporal events (Kumar et al., 2015) have also been performed, but do not specifically focus on building a robust system that simply analyzes content coupled with no other factors. Related literature also focuses on building systems that analyze tweets of users who have committed suicide (Coppersmith et al., 2016), that may not specifically hint at suicidal ideation, as opposed to the proposed problem.

3 Data

3.1 Data Collection

Traditionally, it has been difficult extracting data related to suicidal ideation or mental illnesses due to social stigma but now, an increasing number of people are turning to the Internet to vent their frustration, seek help and discuss mental health issues (Milne et al., 2016), (Sueki et al., 2014). To maintain the privacy of the individuals in the dataset, we do not present direct quotes from any data, nor any identifying information.

Anonymised data was collected from microblogging website Twitter - specifically, content containing self-classified suicidal ideation (i.e. text posts tagged with the word 'suicide) over the period of December 3, 2017 to January 31, 2018. The Twitter REST API[2] was used for collection of tweets containing any of the following English words or phrases that are consistent with the vernacular of suicidal ideation (O'Dea et al., 2015):

suicidal; suicide; kill myself; my suicide note; my suicide letter; end my life; never wake up; can't go on; not worth living; ready to jump; sleep forever; want to die; be dead; better off without me; better off dead; suicide plan; suicide pact; tired of living; don't want to be here; die alone; go to sleep forever; wanna die; wanna suicide; commit suicide; die now; slit my wrist; cut my wrist; slash my wrist; do not want to be here; want it to be over; want to be dead; nothing to live for; ready to die; not worth living; why should I continue living; take my own life; thoughts of suicide; to take my own life; suicide ideation; depressed; I wish I were dead; kill me now

The texts were collected without knowing the sentiment. For example, when collecting tweets

[1] http://www.scmp.com/topics/weibo

[2] https://dev.twitter.com/rest/public/search

on hashtag *#suicide*, it is not known initially whether:

- the tweet is posted for suicide awareness and prevention;
- the person is talking about suicidal ideation and/or ways to kill himself;
- the tweet reports a third persons suicide *eg: news report*;
- the tweet uses suicide as a figure of speech *eg: career suicide*

3.2 Data Annotation

Then, text posts equaling 5213 in all were collected which were subsequently human annotated. The Human annotators consisted of both university students fairly active on social media, and aware of aspects of cognitive psychology as well as university faculty in the domain of Psychology and Machine Learning. Human annotators were asked to indicate if the text implied suicidal ideation using binary criteria by answering the question *Does this text imply self-harm tendencies or suicidal intent?*. Each post was scrutinized and analyzed by three independent annotators (H_1, H_2 and H_3), due to the subjectivity of text annotation, wherein ambiguous posts were set to the default level, *Suicidal intent absent*. Posts were examined individually and annotated according to the following classification system:

1. **Suicidal intent present**:
    - Text conveys a serious display of suicidal ideation; e.g., *I want to die* or *I want to kill myself* or *I wish my last suicide attempt was successful*;
    - Care was taken to classify only those posts as suicidal where suicide risk is not conditional unless some event is a clear risk factor eg: depression, bullying, substance use;
    - Posts where suicide plan and/or previous attempts are discussed; e.g., *"The fact that I tried to kill myself and it didn't work makes me more depressed."*
    - Tone of text is sombre and not flippant, eg: *This makes me want to kill myself, lol*, *"This day is horrible, I want to kill myself hahaha"* are not included in this category.

	H_1	H_2	H_3
H_1	–	0.61	0.48
H_2	0.61	–	0.51
H_3	0.48	0.51	–

Table 1: Cohen's Kappa for three annotators H_1, H_2 and H_3

2. **Suicidal intent absent:**
    - The default category for all posts.
    - Posts emphasizing on suicide related news or information; e.g., *Two female suicide bombers hit crowded market in Maiduguri*.
    - Posts such as *Suicide squad sounds like a good option*; no reasonable evidence to suggest that the risk of suicide is present; includes posts containing song lyrics, etc, were marked within this category.
    - Posts pertaining to condolence and suicide awareness; e.g., *"5 suicide prevention helplines in India you need to know about"*, *Politician accused of driving his wife to suicide*.

Annotators were instructed to select only one of the above categories and to select the default level in case of ambiguity. In all, 15.76% (822) of all tweets were annotated to be suicidal, which were then used to train and validate the classifiers presented in the following sections. A satisfactory agreement between the annotators (e.g., 0.61 for H_1 and H_2) can be inferred from Table 1.

4 Proposed Methodology

The overall methodology is divided into three phases. The initial phase consists of preprocessing the text within a tweet, the second phase involves feature extraction from preprocessed tweets for the training and testing of binary classifiers for the suicidal ideation identification, and the final phase actually classifies and identifies tweets exhibiting suicidal ideation. The details of these individual phases are presented below.

4.1 Preprocessing

Preprocessing is achieved by applying a series of filters, based on Xiang et al. (2012), in the order given below to process the raw tweets.

1. Removal of non-English tweets using LingPipe (Baldwin and Carpenter, 2003) with Hadoop.

2. Identification and elimination of user mentions in tweet bodies having the format of @username, URLs as well as retweets in the format of RT.

3. Removal of all hashtags with length > 10 due to a great volume of hashtags being concatenated words, which tends to amplify the vocabulary size inadvertently.

4. Stopword removal.

4.2 Feature Extraction

Tweets exhibiting suicidal ideation lack a semi-rigid pre-defined lexico-syntactic pattern. Hence, they warrant the use of hand engineering and analyzing a set of features (Wang et al., 2016) in contrast to sentence and word embeddings in a supervised setting using Deep Learning Models such as Convolutional Neural Networks (Kim, 2014) (CNN). The proposed methodology utilizes the following set of features for classification.

- *Statistical Features*. These encompass the number of tokens, and their length.

- *LIWC Features*. Features extracted using the Linguistic Inquiry and Word Count program (LIWC) (Pennebaker et al., 2001) capture people's social and psychological states by analyzing the text to generate labels. Owing to the immense similarity in the nature of the problem of Suicidal Ideation detection in text and the background of LIWC in social, clinical, and cognitive psychology, LIWC features are an ideal candidate for inclusion as a subset of features for our overall classification problem.

  As an example, the accompanying tweet is associated with negative emotions and cognitive processes with a high authenticity and emotional tone. *I'm holding a gun and deciding if I want to go through with suicide or not. I want to commit suicide really badly... Help?*

- *Part of Speech counts*. POS counts for each label assigned by the Stanford Part-Of-Speech Tagger (Manning et al., 2014) are used as a feature. POS Tags include nouns, adjectives, adverbs, verbs, etc.

- *TF-IDF*. The Term Frequency-Inverse Document Frequency (TF-IDF) is used as a feature to reflect the importance of a particular word within the corpus and is given by:

$$tfidf(t) = freq(t) \times ln \frac{N}{|\{d \in D : t \in d\}|}$$

  where, t is the word feature, N is the number of tweets, and d is a document in the document set D.

- *Topics Probability.* The probability distribution of each topic over its terms are used as a feature, which is based on the approach that the tweets are represented as random mixtures over latent topics. Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is a generative probabilistic model that is used to describe each such topic as a generative model which generates words of the vocabulary with certain probabilities, and forms the basis of evaluating Topics Probability.

4.3 Classification

Suicidal Ideation detection is formulated as a supervised binary classification problem. For every tweet $t_i \in D$, the document set, a binary valued variable $y_i \in \{0, 1\}$ is introduced, where $y_i = 1$ denotes that the tweet t_i exhibits Suicidal Ideation. To learn this, the classifiers must determine whether any sentence in t_i possesses a certain structure or keywords that mark the existence of any possible Suicidal thoughts. The features presented above are the used to train classification models to identify tweets exhibiting Suicidal Ideation. Linear classifiers such as Logistic Regression as well as Ensemble Classifiers including Random Forest (Liaw et al., 2002), Gradient Boosting Decision Tree (Friedman, 2002) and XGBoost (Chen and Guestrin, 2016) are employed for classification.

Both XGBoost and Gradient Boosting Decision Trees aim to boost the performance of a classifier in a stage-wise fashion by iteratively adding a new classifier to the ensemble to allow the optimization of a differentiable loss function. The Random Forest classifier is one of the most popular ensemble machine learning algorithm based

on Bootstrap Aggregation (Quinlan et al., 1996) or *bagging*. It modifies the bagging procedure so that the learning algorithm is limited to a random sample of features of which to search, which has shown promise in text classification problems.

5 Baselines

Validation of the proposed methodology is done by comparison against Baseline models that act as a useful point for comparison. Comparison in terms of the evaluation metrics presented below are also done with other recent models for Suicidal Ideation classification as follows:

Long Short Term Memory (LSTM) models are more robust to noise in comparison to Recurrent Neural Networks (RNN) (Liu et al., 2016), and better able to capture long-term dependencies in a sequence, due to their ability to learn how to forget past observations. The LSTM model uses $h = 128$ memory units, with a dropout probability of 0.2, and ReLU (Nair and Hinton, 2010) was used for activation. For training, the Adam Optimizer was used to minimize log loss. A batch size of 64 was chosen and trained for a total of 100 epochs. Pre-Trained word2vec word embeddings that were trained on 100 billion words from Google News are employed as features for classification. Support Vector Machines (Desmet and Hoste, 2013) (SVM) have been shown to work well with short informal text (Pak and Paroubek, 2010) and other promising results in the cognitive behavior domain (De Choudhury et al., 2013). The features described in Desmet and Hoste (2013) are used by the SVM for classification. Rule-based approaches focusing on maximizing the information gain aim to reduce the uncertainty of the class a particular tweet belongs to. A J48 decision tree (C4.5) (Quinlan et al., 1996) was used with the features above for classification.
Lastly, a Negation Resolution (Gkotsis et al., 2016) based approach that is relatively recent, that employs parse trees to build a set of basic rules that rely on minimum domain knowledge is used.

6 Results and Analysis

6.1 Analysis: Comparison with Baselines

Table 2 presents the results for both baselines as well as the classifiers with the proposed methodology in terms of four evaluation metrics: *Accuracy, Precision, Recall and F1 Score*. The first four rows represent the results of the proposed features with both Linear and Ensemble classifiers as described in the Classification section above. The final four rows represent the baseline results.

The proposed features used in conjunction with the first four models described in the Classification section supersede the baseline models in terms of performance along most metrics. The LSTM model has the highest recall, owing to its ability to capture long term dependencies, however its overall performance in terms of accuracy and F1 score is relatively less. Both SVM and Rule-based classification don't perform as well as the proposed methodology, owing to the lack of features used in these models that are not suitable for learning how to classify tweets with Suicidal Ideation. Both of these methods are more suitable in a general domain, however, the features in the proposed methodology are more specific to the particular problem domain of Suicidal Ideation detection, particularly the *LIWC* features and Topics probability. Lastly, the Negation Resolution method performs poorly on the dataset, due to its inability to adapt to a vast and highly diverse form of suicidal ideation communication and its implicit rigidity. This in comparison to the proposed methodology, is unable to effectively learn and extract the essential features from input text, and thus does not perform as well.
In conclusion, the proposed methodology consisting of feature extraction coupled with ensemble and linear classifiers supersedes the baselines presented from various domains in terms of performance.

6.2 Classifiers with proposed features

The first four rows of Table 2 represent the results in terms of the evaluation metrics for different classifiers, both Linear and Ensemble, using the proposed set of features. While the performance of the four classifiers is comparable, Random Forest classifiers perform the best. This is attributed to the ability of Random Forest classifiers that tackle error reduction by reducing variance rather than reducing bias. As has been seen with various text classification problems, Logistic Regression performs fairly well despite its simplicity, and has a greater accuracy and F1 score in comparison with both Boosting Algorithms.

Table 3 shows the variation in performance of the Random Forest classifier with the inclusion of

Table 2: Classification Results in terms of Evaluation metrics.

Model	Accuracy	Precision	Recall	F1 score
Logistic Regression	0.830	0.819	0.850	0.832
Random Forest	**0.858**	**0.842**	0.846	**0.844**
Gradient Boosting Decision Tree	0.805	0.802	0.820	0.807
XGBoost	0.817	0.831	0.800	0.812
LSTM	0.789	0.745	**0.874**	0.796
Support Vector Machine	0.792	0.821	0.692	0.754
Rule-based Classification	0.801	0.824	0.743	0.781
Negation Resolution	0.527	0.542	0.752	0.635

Table 3: Variation in performance with the inclusion of features

Features used	Accuracy	Precision	Recall	F1 score
Statistical Features(SF) only	0.596	0.547	0.600	0.569
SF + TF-IDF	0.669	0.663	0.753	0.702
SF + TF-IDF + POS counts	0.789	0.821	0.705	0.721
SF + TF-IDF + POS + Topics Probability	0.807	0.814	0.820	0.817
All Features	0.858	0.842	0.846	0.844

the various features. The precision reduces by a small amount with the inclusion of Topics probability feature implying that a greater subset of tweets is classified as suicidal due to the LDA unigrams included via Topics probability features, but is finally boosted by the inclusion of the LIWC features. The POS counts also lead to a reduction in the recall, which is compensated with the subsequent inclusion of Topics Probability and LIWC features. The drastic improvements are attributed to the TF-IDF, POS counts and LIWC features in terms of most evaluation metrics. It is observed that the proposed set of features perform the best in conjunction with Random Forest classifiers, and the improvement in performance with the inclusion of each feature validates the need for the extraction of that feature.

6.3 Error Analysis

Some categories of errors that occur are:

1. **Seemingly Suicidal tweets:** Human annotators as well as our classifier could not identify whether *"I want to kill myself, lol. :("* was representative of suicidal ideation or a frivolous reference to suicide.

2. **Pragmatic difficulty:** The tweet *"I lost my baby. Signing off.."* was correctly identified by our human annotators as a tweet with suicidal intent present. This tweet contains an element of topic change with no explicit mention of suicidal ideation, but our classifier could not capture it.

3. **Ambiguity:** The tweet *"Is it odd to know I'll commit suicide?"* is a tweet that both human annotators as well as the proposed methodology couldn't classify due to it's ambiguity.

7 Conclusion and Future Work

This paper proposes a model to analyze tweets, by developing a set of features to be fed into classifiers for identification of Suicidal Ideation using Machine Learning. When annotated by humans, 15.76% of the total dataset of 5213 tweets was found to be suicidal. Both linear and ensemble classifiers were employed to validate the selection of features proposed for Suicidal Ideation detection. Comparisons with baseline models employing various strategies such as Negation Resolution, LSTMs, Rule-based methods were also performed. The major contribution of this work is the improved performance of the Random forest classifier as compared to other classifiers as well as the baselines. This indicates the promise of the proposed set of features with a bagging based approach with minimal correlation show as compared to other classifiers. In the future, there is scope for larger amounts of data to be scraped from more social media websites as well as investigate the performance withdeep learning models such as CNNs, LSTM-CNNs, etc.

References

Breck Baldwin and Bob Carpenter. 2003. Lingpipe. *Available from World Wide Web: http://alias-i.com/lingpipe*.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Julie Cerel, John R Jordan, and Paul R Duberstein. 2008. The impact of suicide on the family. *Crisis*, 29(1):38–44.

Kara Chan and Wei Fang. 2007. Use of the internet and traditional media among young people. *Young Consumers*, 8(4):244–256.

Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM.

Glen Coppersmith, Kim Ngo, Ryan Leary, and Anthony Wood. 2016. Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*, pages 106–117.

E Jane Costello, Daniel S Pine, Constance Hammen, John S March, Paul M Plotsky, Myrna M Weissman, Joseph Biederman, H Hill Goldsmith, Joan Kaufman, Peter M Lewinsohn, et al. 2002. Development and natural history of mood disorders. *Biological psychiatry*, 52(6):529–542.

Kate Daine, Keith Hawton, Vinod Singaravelu, Anne Stewart, Sue Simkin, and Paul Montgomery. 2013. The power of the web: a systematic review of studies of the influence of the internet on self-harm and suicide in young people. *PloS one*, 8(10):e77555.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. *ICWSM*, 13:1–10.

Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2098–2110. ACM.

Bart Desmet and VéRonique Hoste. 2013. Emotion detection in suicide notes. *Expert Systems with Applications*, 40(16):6351–6358.

Jerome H Friedman. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367–378.

George Gkotsis, Sumithra Velupillai, Anika Oellrich, Harry Dean, Maria Liakata, and Rina Dutta. 2016. Don't let notes be misunderstood: A negation detection method for assessing risk of suicide in mental health records. In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*, pages 95–105.

Xiaolei Huang, Xin Li, Tianli Liu, David Chiu, Tingshao Zhu, and Lei Zhang. 2015. Topic model for identifying suicidal ideation in chinese microblog. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, pages 553–562.

Xiaolei Huang, Lei Zhang, David Chiu, Tianli Liu, Xin Li, and Tingshao Zhu. 2014. Detecting suicidal ideation in chinese microblogs with psychological lexicons. In *Ubiquitous Intelligence and Computing, 2014 IEEE 11th Intl Conf on and IEEE 11th Intl Conf on and Autonomic and Trusted Computing, and IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UTC-ATC-ScalCom)*, pages 844–849. IEEE.

Jared Jashinsky, Scott H Burton, Carl L Hanson, Josh West, Christophe Giraud-Carrier, Michael D Barnes, and Trenton Argyle. 2014. Tracking suicide risk factors through twitter in the us. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 35(1):51.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Mrinal Kumar, Mark Dredze, Glen Coppersmith, and Munmun De Choudhury. 2015. Detecting changes in suicide content manifested in social media following celebrity suicides. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 85–94. ACM.

Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by randomforest. *R news*, 2(3):18–22.

Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. 2016. Recurrent neural network for text classification with multi-task learning. *arXiv preprint arXiv:1605.05101*.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.

David N Milne, Glen Pink, Ben Hachey, and Rafael A Calvo. 2016. Clpsych 2016 shared task: Triaging content in online peer-support forums. In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*, pages 118–127.

Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

Bridianne O'Dea, Stephen Wan, Philip J Batterham, Alison L Calear, Cecile Paris, and Helen Christensen. 2015. Detecting suicidality on twitter. *Internet Interventions*, 2(2):183–188.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

J Ross Quinlan et al. 1996. Bagging, boosting, and c4. 5. In *AAAI/IAAI, Vol. 1*, pages 725–730.

Jo Robinson, Georgina Cox, Eleanor Bailey, Sarah Hetrick, Maria Rodrigues, Steve Fisher, and Helen Herrman. 2016. Social media and suicide prevention: a systematic review. *Early intervention in psychiatry*, 10(2):103–121.

L Sher. 2004. Preventing suicide. *Qjm*, 97(10):677–680.

Hajime Sueki, Naohiro Yonemoto, Tadashi Takeshima, and Masatoshi Inagaki. 2014. The impact of suicidality-related internet use: A prospective large cohort study with young and middle-aged internet users. *PloS one*, 9(4):e94841.

Yufei Wang, Stephen Wan, and Cécile Paris. 2016. The role of features and context on suicide ideation detection. In *Proceedings of the Australasian Language Technology Association Workshop 2016*, pages 94–102.

Guang Xiang, Bin Fan, Ling Wang, Jason Hong, and Carolyn Rose. 2012. Detecting offensive tweets via topical feature discovery over a large scale twitter corpus. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1980–1984. ACM.

Andrew Yates, Arman Cohan, and Nazli Goharian. 2017. Depression and self-harm risk assessment in online forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2968–2978.

William WK Zung. 1979. Suicide prevention by suicide detection. *Psychosomatics*, 20(3):153–155.