SIAS: Suicidal Intentions Alerting System

Georgios Domalis[1], Christos Makris[1], Pantelis Vikatos[1], Anastasios Papathanasiou[2,3], Efterpi Paraskevoulakou[2] and Manos Sfakianakis[2,3]

[1]*Computer Engineering & Informatics Department, University of Patras, Patras, Greece*

[2]*Department of Informatics, University of Piraeus, Piraeus, Greece*

[3]*Cyber Crime Division, Hellenic Police, Athens, Greece*

Keywords: Alerting System, Classification Model, Suicidal Intention.

Abstract: In this paper, we present an alerting system based on an efficient classification model for detecting suicidal people using natural language processing and data mining techniques. The model uses linguistic features which are derived from an analysis of handwritten and electronic messages/notes. The model was trained and validated with fully anonymised real data provided by the Cyber Crime Division of Greek Police as well as available suicidal notes from social media. The alerting system is intended as a prevention, management tool for automatic detection of suicidal intentions.

1 INTRODUCTION

Suicide is a serious problem of social and public health perspective. According to recent World Health Organization[1] more than 30,000 suicide deaths in the United States and nearly 1 million suicide deaths worldwide occur every year constituting suicide one of the 20 leading causes of death. The Internet as well as social media can play an important role in prevention of suicide cases. In addition, current scientific research (Burnap et al., 2015; Colombo et al., 2016; Sueki, 2015) is focused on discovering patterns and association between users' behavior in social media with suicidal intention. Regarding official statistics from Greek Law Enforcement Agency[2] the rates of interventions Cyber Crime Division, related to suicidal intentions have been extremely increased during the current decade. The efficient investigation and management of cases that involve a report of suicide on the internet requires collaboration between Law Enforcement Agency (Cyber Crime Division) and responsible ISPs. Law Enforcement Agency encounters difficulties in the investigation of such cases and cybercrimes due to the strict legislative framework. First of all the cross-border nature and the involvement of many jurisdictions impede the investigation and detection of such cases. Therefore, it is necessary the international police cooperation in immediate time given the serious and urgent nature of these cases (online suicides). However, there are a lot of constraints and problems encountered by law enforcement authorities in conducting investigations whenever the digital data and suicidal posts are stored on computing cloud environments. More specifically, many challenges that exist in these environments have to be handled involving multiple levels, such as recognition, collection, preservation, examination, interpretation and reporting of digital data and evidence. A major problem arises related to the fact that, even if such data are identified, the characteristics of the cloud environment complicate the collection of evidence through legal procedure, since it is possible by involving many jurisdictions. Another problem is the large amount of data and information stored in cloud computing environments, which adds difficulty to identify those data related to the specific research carried out every time. The policy and guidelines for addressing these challenges is imperative to effectively investigate announced suicide cases on the Internet and law enforcement generally on the World Wide Web.

The goal of this paper is the implementation of an alerting system called SIAS (Suicidal Intentions Alerting System) for predicting suicide intention using an efficient classification model. The classification model introduces a set of characteristics which

[1]*http://www.who.int/mental_health/prevention/en/*

[2]*http://www.hellenicpolice.gr*

Domalis, G., Makris, C., Vikatos, P., Papathanasiou, A., Paraskevoulakou, E. and Sfakianakis, M.
SIAS: Suicidal Intentions Alerting System.
DOI: 10.5220/0006297402910297
In *Proceedings of the 13th International Conference on Web Information Systems and Technologies (WEBIST 2017)*, pages 291-297
ISBN: 978-989-758-246-2

are derived from analysis of texts and are related to cohesion, emotional instability, part of speech, certain emotional lexicons and personality traits. Natural language techniques have been used for the feature extraction directly from the textual sources. We train and evaluate machine learning algorithms in order to conclude with, one which performs efficiently in terms of F-measure for our datasets. We use pre-annotated texts divided into two categories. The first one constitute a set of citizens' texts with the proven suicidal intention and other with normal behavior provided by diverse sources.

The rest of the paper is structured as follows. Section 2 overviews related work, we motivate our research from current challenges regarding and related studies. In Section 3, we provide an overview of the system architecture that we propose with the detailed description of each module. In Section 4 the classification is mentioned with the data preparation and the training/evaluation of the model. Section 5 overviews details of the implementation of the system for modules and sub-modules respectively and presents a reference to our experimental results. Finally, in Section 6, we discuss the strengths and limitations of our approach and we conclude with an outlook to future work.

2 RELATED WORK

The concept of prediction of suicidal people through electronic or hand written messages has gained the interest of researchers in the field of natural language processing, sentiment analysis and machine learning (Burnap et al., 2015; Thompson et al., 2014). In Pestian et al. (Pestian et al., 2010) they introduce methods that distinguish genuinely and elicited suicide notes using a content analysis based on natural language processing.

Also, several studies are focused on predicting military and veteran suicide risk (Soltaninejad et al., 2014; Thompson et al., 2014). (Thompson et al., 2014) describes language models in order to detect suicidal ideation from texts derived from social media and specifically, Facebook posting. The results of this study show that word pairs are more useful than single words for model construction. Study (Burnap et al., 2015) introduces an approach with text mining technique using classification models in order to categorize each text derived from Twitter to multiple classes, related to suicidal ideation and topics such as reporting of a suicide, memorial, campaigning and support. Another study (Sueki, 2015) discovers the association between suicide-related Twitter posts with suicidal behavior. Especially this research focuses on young people using the Internet and social media to identify and predict suicide intentions based on the texts and behavior of Twitter users. The linking of personality traits with the suicidal ideation constitutes the aim of study (Soltaninejad et al., 2014). Regarding to this research personality traits such as neuroticism, extraversion and conscientiousness traits may predict suicidal ideation on the other hand traits such as agreeableness and openness weakly correlate with suicide. An approach using a vocabulary of topics which suicidal persons are used to talking is presented in (Abboute et al., 2014). Twitter messages are parsed and checked using the topics by a process combining natural language processing and learning methods to indicate suicidal risky behavior. In addition, in the study (Lightman et al., 2007) well-known tools for linguistic analysis such as Coh-Metrix (Graesser et al., 2004) and LIWC (Pennebaker et al., 2001) have been used to investigate the correlation of features with suicidal and non-suicidal songwriter's lyrics. In (Mairesse et al., 2007) the authors present firstly a detailed correlation analysis between Big Five personality traits and the features contained in software, e.g. LIWC; then they described an automatic recognition of user personality traits using regression and classification models. In another study (Stirman and Pennebaker, 2001) the word use of suicidal and non suicidal poets are examined and correlations between the suicidal ideation and features from LIWC are extracted.

In this work, we deviate from the above approaches due to the combination of linguistic features with personality traits which are strongly correlated with suicidal intentions improving the performance of prediction.

3 SYSTEM ARCHITECTURE

In this section, we introduce our model a detailed description of the alerting system architecture. The proposed alerting system is focused to immediate identification of suicidal intentions.

The system is composed of the following modules:

- The *targets identification*. The alerting system starts by targeting of certain citizens (subjects) which their psychological disorder could lead to suicidal attempt. The targets are reported by citizens, psychological clinics or Law Enforcement Agency.
- The *user profile creation*. This module deals with the registration of targets' personal info and the creation of users' profile. Coordinating Operational Center of Cyber Crime Division finds or

receives information about the suicidal intention over the Internet or other communication technologies, following a thorough online police investigation, which includes correspondence with concerned websites (Facebook, Twitter, Linkedin etc.) and companies providing telecommunications and Internet Services (ISPs) in order to identify certain citizens.

- The *feature extraction*. The module of feature extraction accepts raw texts for each individual target. In this step, feature vectors are produced by analysis of text, creating characteristics related to cohesion and emotional instability, linguistic and personality traits. The feature vector is introduced to the classification model for suicidal intention declaration.
- The *classification model*. This module constitutes the predictor of the suicidal intention. The classification model has been trained and evaluated with texts derived from, Cyber Crime Division, open data and posts from well-known social media. According to our approach the prediction of suicidal intention constitutes a binary classification problem in which a feature vector is placed in a category (Y/N).
- The *monitoring & alerting*. The monitoring & alerting module is actually the platform which presents in real time the indicators of all targets as well as the classification.

4 CLASSIFICATION MODEL

The efficiency of the proposed system depends on the performance of classification model. The main challenge that we face is the lack of available data sources due to the nature and anonymity of this type of data. Cyber Crime Division of Greek police intervenes in cases related to suicidal intentions in which citizens inform of their attempt in social media, blogs or open sources. Respecting personal information and according to ethical principles and Greek legislation, a real text corpus (fully anonymised) of suicidal cases, in which Cyber Crime Division intervened in the past, was used for the evaluation of our system. In addition the dataset was enriched with text with normal behavior provided by diverse sources. Furthermore, we added suicidal notes from two social media Tumblr[3] and Whisper[4]. Also, we introduced a detailed graph model that makes easy the decision whether a text implies or not a suicidal manifestation as in it is shown in Figure 2; we explain this in the sequel.

[3] *https://www.tumblr.com/*
[4] *https://whisper.sh/*

4.1 Data Preparation

In this section, there is a brief description of the features which are used for the training procedure of different classification algorithms. The choice of cohesion and linguistic metrics were based on previous research concerning the relationship between psychological health and language use as it is described in (Lightman et al., 2007; Stirman and Pennebaker, 2001). In addition, we introduce personality traits as additional characteristics.
There is a fruitful discussion through academics about the relation of personality traits with suicidal intention (Carrillo et al., 2001; Kotrla Topić et al., 2012; Rozanov and Mid'ko, 2011; Soltaninejad et al., 2014). Figure 2 presents a graph with the association of the five basic dimensions of personality, that remain stable in individuals forming the Big Five Model (McCrae and John, 1992), with two factors of Auto-Destructive Behavior and Depression which immediate influence the suicidal ideation. The Depression and Auto-Destructive Behavior are recognized as significant predictors of various suicidal manifestations, in particular, phenomena of hopelessness and suicidal ideation (Kotrla Topić et al., 2012; Rozanov and Mid'ko, 2011). According to Big Five, the human personality is described as a vector of five values of traits. The combination of Big Five personality dimensions explain the dynamics of a personality. For example, a person may be very talkative (high Extraversion), not very tolerant and sensitive (low Agreeableness), systematic and punctual (high Conscientiousness), easily anxious (high Neuroticism) and extremely curious (high Openness). Each personality trait correlates differently with Depression and Auto-Destructive Behavior. More specifically individuals that present open to fantasy, but not to actions are prone to harmful Auto-Destructive intentions and depression respectively (Carrillo et al., 2001). Furthermore, an extrovert person seems to be less risky to deal with depression or to harm himself (Kotrla Topić et al., 2012). Also, people that lack of sufficiency, self-discipline, self-control in decision making increase the probability of belonging to the depression and suicidal class (Rozanov and Mid'ko, 2011; Soltaninejad et al., 2014). In study (Soltaninejad et al., 2014) is mentioned that the high degree of neuroticism which includes anger, hostility, impulsivity and vulnerability constitutes a significant of developing psychiatric disorders such as mood disorders, depression which may lead to suicide (Solta-

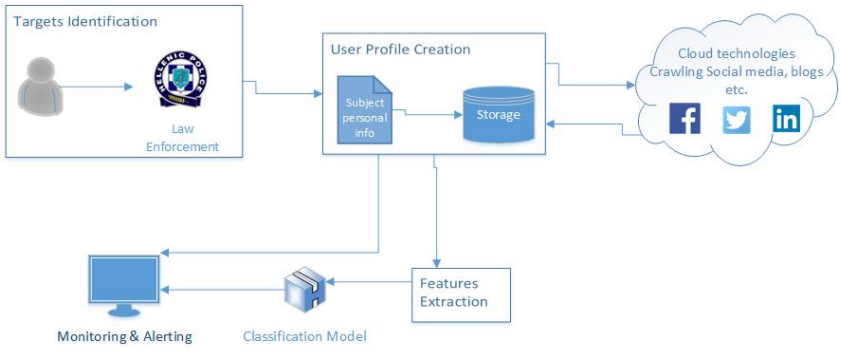

Figure 1: Alerting system architecture

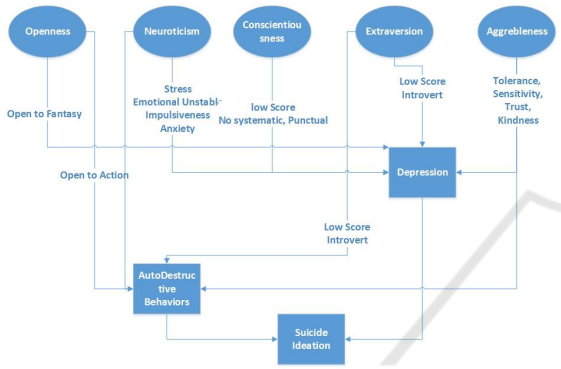

Figure 2: Association of Personality Traits with Suicidal Ideation.

ninejad et al., 2014). Last but not least people who react sensitively, tolerant and kind of situations could lead confront with auto destructiveness and depression (Kotrla Topić et al., 2012).

The personality traits in Big Five model can be extracted based on a questionnaire that determines user personality as described in (John and Srivastava, 1999). However, we follow an approach for unsupervised recognition of personality traits deriving from text analysis (Celli, 2012).

Cohesion & Emotional Instability, Linguistic and Personality Traits constitute the 3 sets of features as it summarily described in Table 1. The prediction features are produced by the linguistic analysis of texts for each individual.

Cohesion and Emotional Instability Metrics:

- Argument Overlap: This index measures aspects of cohesion. The argument overlap index helps to decide readability, complexity and grade level of a particular corpus.
- Latent Semantic Analysis (LSA): LSA is a measure of cohesion like argument overlap. The LSA index compares the contextual similarity between sentences recognizing text's cohesiveness.
- Word Concreteness: This feature measures the terms' concreteness of individuals in the written texts.
- Communication Words: This measure is the ratio of communication words in the sentences by using a certain dictionary of words, such as talk, share and converse. We created these dictionaries adding the synonyms from Wordnet (Miller, 1995).
- Tense/Aspect Ratio: This index declares the degree that sentences are linked by the time relation in order to extract the overall temporal cohesion on the sentences.
- References to Time: The feature exposes the ratio of verbs which are referred to future tense and words related to time in general.
- Death-Themed Words: This measure is the ratio of death-oriented words in the sentences by using a certain dictionary. The dictionary includes words related to death and dying and their synonyms.
- Swear Words: A list of vulgar expressions has been created to extract the ratio of swear words in the sentences.
- Emotion Words: The ratio of words that address affective or emotional processes. We divide these words to positive and negative emotion.

Linguistic Metrics:

- Punctuation metrics: The number of commas, question marks, exclamation marks, parenthesis and all punctuation divided by the total number of tokens in the text.
- Reference metrics: The number of @, first person (singular and plural) pronouns divided by the total number of tokens in the text.
- Part of speech metrics: The number of first person singular pronouns, prepositions, pronouns, first person plural pronouns, second person singular

pronouns, numbers divided by the total number of tokens.

- Emoticons: The number of emoticons that express negative and positive feelings divided by the total number of tokens.
- Word/token metrics: This feature set includes the mean word frequency, the type/token ratio, the number of words longer than 6 letters divided by the total number of tokens. word count: the number of words of the text.
- Special patterns: The number of external links and the number negative particles divided by the total number of tokens in the text.

Personality Traits:

- Agreeableness (A): This personality dimension includes attributes such as affability, tolerance, sensitivity, trust and kindness.
- Conscientiousness (C): Common features of this dimension include organization, punctuality, achievement-orientation and dependency.
- Extraversion (E): This trait includes individuals such as outgoing, talkative, sociable and enjoying social situations.
- Neuroticism (N): Individuals high in this trait tend to be anxious, irritable, temperamental and moody.
- Openness (O): This trait includes curiosity, originality, intellectuality, creativity and openness to new ideas.

4.2 Training & Evaluation

The feature vectors are enriched with the category for each individual Y, N for suicidal and non-suicidal intention respectively in order to form the datasets for training and evaluation of the classifiers. We follow the same approach as previous studies (Burnap et al., 2015; Thompson et al., 2014; Abboute et al., 2014) introducing the dataset to different classifiers to investigate the appropriate one which fits efficiently for this type of data. The performance was evaluated by F-measure. Also, the parameter selection is tested using a greedy approach in each classifier separately.

5 EXPERIMENTAL RESULTS

5.1 Implementation

We gathered 200 samples of suicidal and non-suicidal intention related notes as it is shown in Table 2. The suicidal notes are derived from Cyber Crime Division of Greek police in cases that information about their suicide attempt has been exposed in social media, blogs and open sources. Furthermore, we use suicidal notes from 2 well-known social networks (Tumblr, Whisper). The features were separated into 3 sets which each one produces a different dataset. The first set includes Cohesion & Emotional Instability Metrics (Cohesion), the second set includes the combination of the first set with Personality traits (CohB5) and the third one the aggregation of all features. We selected Naive Bayes, SVM, Rotation Forest, AdaBoost and J48 as representative classifiers in order to examine the performance for this type of data. The classifiers are evaluated by the F-measure metric which is the harmonic mean of precision and recall.

Table 1: The User Profile Feature Vector.

Features	#	Description
Linguistic Metrics	21	Punctuation(5), Reference(2), Part-of-speech(6), Emoticons(2), Word/token(4), Special patterns(2)
Cohesion Metrics	21	Tense/Aspect ratio, References to Time(3), Argument Overlap(9), Communication Words, Death Themed Words, Emotion Words(3), Word Concreteness, LSA, Swears Words
Personality Traits	5	Openness, Neuroticism, Agreeableness, Conscientiousness, Extraversion

We separated each dataset to a train and a test set, using two approaches:

- K-Fold Cross-Validation (K=10 Fold).
- Leave-One-Out Cross-Validation

The concept of using both techniques is that splitting with 10-Fold Cross-Validation, important information can be removed from the train set. However, the Leave-One-Out Cross-Validation technique evaluates the classification performance based on one sample.

The text processing as well as the feature extraction and training of classifiers was implemented in Python

Table 2: Number of Instances per Class.

Class	Instances
Suicidal Intention	114
Non-Suicidal Intention	106

2.7 using NLTK[5] and scikit-learn[6] modules.

5.2 Results & Discussion

In this section, the results of our research are presented. We evaluate our dataset with 10-Cross-Validation and Leave-One-Out methods. Figure 3 and Figure 4 depict performance for all classifiers. According to the result, we observe that the performance of the classifiers in terms of F-measure is better in the dataset with the aggregation of all features than removing the Linguistic and Personality Traits sets. We consider that the add on features improves the performance of all classifiers except SVM and Rotation Forest, which deviates in the Coh - B5 set of features in Leave-One-Out evaluation.

Rotation Forest classifier outperforms in comparison to the other classifiers regarding F-measure. However Rotation Forest presents the highest deviation between the two methods of evaluation. Furthermore, the use of linguistic metrics and personality traits in the training of the classifier enhances the performance 10% approximately in the 10-Cross-Validation. The combination of Cohesion metrics with Big Five traits improves or remains the performance. The aggregation of all features improves the F-measure which is 0.805 in the worst case using J48 classifier and reaches 0.895 in the best case using Rotation Forest.

In our understanding the use of Rotation Forest benefits the efficient prediction of suicidal intention of certain targets and enhance the alerting system. Also the majority of the extracted features from text analysis is language independent and our system can easily be expanded to any language using a specific part of speech and dictionaries for communicative, death and swears words.

Table 3: Cohesion Metrics.

Classifier	10-CV	Leave-One-Out
Naive Bayes	0.727	0.727
SVM	0.795	0.804
Rotation Forest	**0.8**	**0.814**
AdaBoost	0.764	0.736
J48	0.704	0.677

[5] *http://www.nltk.org/*
[6] *http://scikit-learn.org*

Table 4: Cohesion Metrics and Big-Five.

Classifier	10-CV	Leave-One-Out
Naive Bayes	0.737	0.727
SVM	0.777	**0.786**
Rotation Forest	**0.804**	**0.786**
AdaBoost	0.764	0.736
J48	0.741	0.713

Table 5: All Metrics.

Classifier	10-CV	Leave-One-Out
Naive Bayes	0.836	0.764
SVM	0.841	0.85
Rotation Forest	**0.895**	**0.868**
AdaBoost	0.818	0.832
J48	0.805	0.814

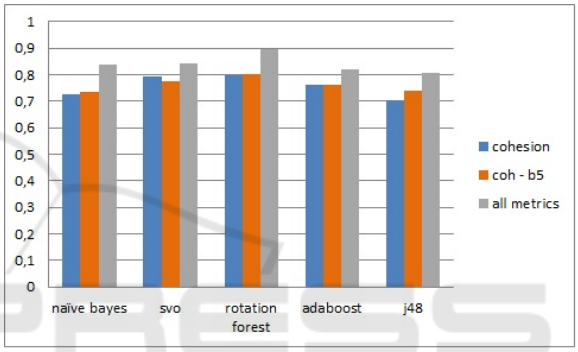

Figure 3: 10-cross validation (F-measure).

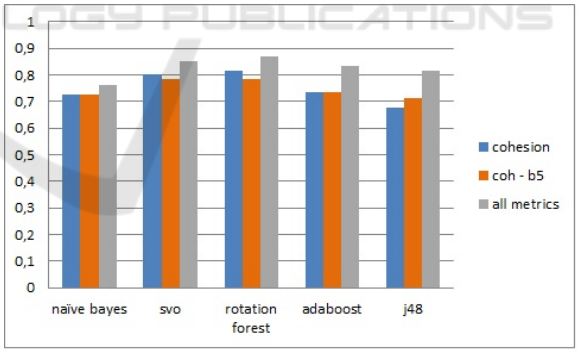

Figure 4: Leave-One-Out (F-measure).

6 CONCLUSIONS

In this work, we looked into the design of the alerting system (SIAS) for the immediate intervention of authorities in suicidal intention. Our methodology includes 5 phases with the Targeting, Using Profile Creation, Feature Extraction, Classification and Alerting. This paper has concentrated on building an efficient classifier with natural language processing and data mining techniques to predict the suicidal intention us-

ing features derived from text analysis. The results depict with high accuracy the suicidal manifestation in the text of users in danger to commit suicide.

As a future work, we are considering to explore alternative factors which are not introduced in our approach and potentially influence the prediction of suicidal intention such as the behavior in social networks.

REFERENCES

Abboute, A., Boudjeriou, Y., Entringer, G., Azé, J., Bringay, S., and Poncelet, P. (2014). *Mining Twitter for Suicide Prevention*, pages 250–253. Springer International Publishing, Cham.

Burnap, P., Colombo, W., and Scourfield, J. (2015). Machine classification and analysis of suicide-related communication on twitter. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, HT '15, pages 75–84, New York, NY, USA. ACM.

Carrillo, J., Rojo, N., Sanchez-Bernardos, M., and Avia, M. (2001). Openness to experience and depression. *European Journal of Psychological Assessment*, 17(2):130.

Celli, F. (2012). Unsupervised personality recognition for social network sites. In *Proc. of Sixth International Conference on Digital Society*. Citeseer.

Colombo, G. B., Burnap, P., Hodorog, A., and Scourfield, J. (2016). Analysing the connectivity and communication of suicidal users on twitter. *Computer Communications*, 73, Part B:291 – 300. Online Social Networks.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., and Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, 36(2):193–202.

John, O. P. and Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138.

Kotrla Topić, M., Perković Kovačević, M., and Mlačić, B. (2012). Relations of the big-five personality dimensions to autodestructive behavior in clinical and non-clinical adolescent populations. *Croatian medical journal*, 53(5):450–460.

Lightman, E. J., McCarthy, P. M., Dufty, D. F., and McNamara, D. S. (2007). Using computational text analysis tools to compare the lyrics of suicidal and nonsuicidal songwriters. In *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, pages 1217–1222. Citeseer.

Mairesse, F., Walker, M. A., Mehl, M. R., and Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *J. Artif. Int. Res.*, 30(1):457–500.

McCrae, R. R. and John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

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

Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., and Leenaars, A. (2010). Suicide Note Classification Using Natural Language Processing: A Content Analysis. *Biomedical informatics insights*, 2010(3):19–28.

Rozanov, V. A. and Mid'ko, A. A. (2011). Personality patterns of suicide attempters: gender differences in ukraine. *The Spanish journal of psychology*, 14(02):693–700.

Soltaninejad, A., Fathi-Ashtiani, A., Ahmadi, K., Mirsharafoddini, H. S., Nikmorad, A., and Pilevarzadeh, M. (2014). Personality factors underlying suicidal behavior among military youth. *Iran Red Crescent Med J*, 16(4):e12686.

Stirman, S. W. and Pennebaker, J. W. (2001). Word Use in the Poetry of Suicidal and Nonsuicidal Poets. *Psychosomatic Medicine*, 63(4):517–522.

Sueki, H. (2015). The association of suicide-related twitter use with suicidal behaviour: A cross-sectional study of young internet users in japan. *Journal of Affective Disorders*, 170:155 – 160.

Thompson, P., Bryan, C., and Poulin, C. (2014). Predicting military and veteran suicide risk: Cultural aspects. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 1–6, Baltimore, Maryland, USA. Association for Computational Linguistics.