

Artificial Intelligence in Detecting Suicidal Content on Russian-Language Social Networks

Sergazy Narynov¹, Kanat Kozhakhmet^{1,2}, Daniyar Mukhtarkhanuly¹,
Aizhan Sambetbayeva³, and Batyrkhan Omarov^{3,4,5}✉

¹ Alem Research, Almaty, Kazakhstan
narynov@alem.kz

² Open University of Kazakhstan, Almaty, Kazakhstan
³ Al-Farabi Kazakh National University, Almaty, Kazakhstan
batyahan@gmail.com

⁴ International Information Technology University, Almaty, Kazakhstan

⁵ Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan

Abstract. Due to the anonymity of online media and social networks, people tend to express their feelings and suffering in online communities. To prevent suicides, it is necessary to detect messages about suicides and user perceptions of suicides in cyberspace using natural language processing methods. We focus on the social network Vkontakte and classify users' messages with potential suicide and without suicidal risk using text processing and machine learning methods.

In this paper, we tell about suicidal and depressive ideation detection in Russian Language. For this purpose, we create a dataset that consists of 64,000 posts that collected from Russian language social network Vkontakte. The dataset was applied to eight machine learning algorithms.

Keywords: Suicidal ideation detection · Suicide · Machine learning · Social networks · Suicidal content

1 Introduction

Currently, one of the most pressing problems facing mental health professionals is the problem of suicidal behavior, especially in the adolescent population. At the same time, suicide is a serious public health problem and mental health in particular. According to WHO data, the overall level of suicidal activity has increased by 60% over the past 45 years [1], and the number of suicide cases in the world in recent decades has reached about 1 million people a year. Suicide is the 14th leading cause of death and accounts for 1.5% of all deaths in the world, which corresponds to a global death rate of 16.7 people per 100,000 population and is equivalent to the death of one person every 40 s [2, 3].

According to world statistics, the number of suicides exceeds the number of victims of murder, terrorist acts and wars combined [4]. Every 20 s, one person commits suicide, and every 2 s, someone unsuccessfully tries to take their own life. At the same time, men are three times more likely than women to resort to suicide, but women are 2–3

times more likely to attempt suicide. This is based on the increased demonstrativeness of women [5].

Content of this paper as following: Next section make a review about suicide problem. Section 3 describe the dataset development process that contains suicidal post collection from Vkontakte [6] social network, description of the developed dataset. Section 4 shows the experiment results and evaluation process. In addition, we give online link to the dataset in open source and make a conclusion of our research.

2 Background

Suicide is an extremely complex and multifaceted human phenomenon, caused by a variety of causes and circumstances that sometimes contradict each other. If we try to generalize, we can say that suicide is associated with both individual psychological and social, cultural and existential factors [7].

Traditionally, several forms of suicidal behavior are distinguished: suicidal thoughts, suicidal attempts, and completed suicides [8]. Suicidal thoughts are relatively widespread, including among teenagers. About a third of teenagers have experienced them [9]. Suicidal attempts are one of the most accurate predictors of future suicide. The tragedy of completed suicides is not only the death of a teenager, but also the severe consequences for his micro - and macro-social environment.

To prevent suicide, Facebook announced in early March 2017 that it would use artificial intelligence to identify users on the verge of suicide and offer them support. The social network contains detailed information about more than two billion of its users. It will be processed by algorithms, and then pattern recognition technology will highlight posts “with a high probability of containing suicidal thoughts” and transmit them to a team of analysts for human intervention [10].

A couple of weeks later, the Russian network Vkontakte announced similar plans. However, there is an emphasis on image analysis: “dangerous” content, according to the developers, will simply stop getting into users’ feeds. First of all, we are talking about the symbols of “death groups”. Page authors of “dangerous” images are blocked: the next time you sign in, the social network user asked what prompted him to post a particular picture, and depending on the response, send for help, or psychologists, or support [11].

Attempts to teach technology to fight suicide have been going on for a long time: back in 2007, a group of researchers from the Queen Victoria University of new Zealand tried (quite successfully) to analyze user records MySpace.com to identify those who are on the verge of suicide. Among other things, it is known that machine learning algorithms are much more effective than professional doctors can distinguish a real suicide note from a fake one (78% accuracy against 63%), and suicidal tendencies can be indicated, for example, by the duration of vowel sounds during conversation. These searches do not arise from scratch: a recent study found that in fifty years of studying suicide, scientists have not made any visible progress in predicting it. Traditional risk factors identified over the past half century—depression, stress, and substance use—provide no more accuracy than random guessing. According to the study’s author, Joseph Franklin of the University of Florida, “information about previous suicide attempts helps improve the accuracy of the forecast in much the same way that buying a second lottery ticket helps win the jackpot” [12].

The main methodological features of the study were described in detail earlier [13]. In this work, we used the answers of teenagers to questions concerning 28 types of NHS that occurred in the last 6 months. The proposed event options included (in the specified order): Illness of a family member; Appearance of a new family member; Mother/father starts or stops working; problems with parents; Bullying or harassment; Death of a pet; unemployment of parents; Increased workload at school; Change in the financial situation of parents; Theft of personal property; Grades worse than expected; Serious conflict with a teacher; Transfer to another school; Sexual problems; marriage of a sister/brother with whom you are emotionally close; a family Member abusing alcohol or drugs; being taken To the police; Serious quarrel with a close friend; Minor legal violation; Important interview, exam; Failed exam; Serious injury or illness; Separation from a boyfriend/girlfriend; divorce of parents; Death of a close friend; Death of a close family member; Pregnancy; Other. We also analyzed the answers to the question "How often do you feel stress? Stress means a situation in which you feel tense, anxious, irritated, anxious or confused" with the suggested answer options "Never or several times a year, monthly, weekly, several times a week, most days a week". This question is the key one in the P88 questionnaire. The data analysis was performed using the methods of variational and nonparametric statistics. The total number of respondents included in this analysis, taking into account the quality of filling out questionnaires, was 589 [14].

3 Materials and Methods

Suicide can be considered as one of the most serious social health problems in modern society. Many factors can lead to suicide, such as personal problems such as hopelessness, severe anxiety, schizophrenia, alcoholism, or impulsivity; social factors such as social isolation and excessive exposure to death; or negative life events, including traumatic events, physical illnesses, affective disorders, and previous suicide attempts. Thousands of people around the world are suicidal every year, making suicide prevention a critical global public health mission.

According to world statistics, Kazakhstan ranks 3rd–4th in the world in the number of suicides [15], as well as 1st in the number of suicides among girls aged 15 to 19 years worldwide [16–18].

Suicidal thoughts or suicidal thoughts are people's thoughts about suicide. It can be considered as an indicator of the risk of suicide. Suicidal thoughts include fleeting thoughts, extensive thoughts, detailed planning, role-playing, and incomplete attempts. According to a who report [19], an estimated 788,000 people worldwide committed suicide in 2015. Moreover, a large number of people, especially teenagers, have been reported with suicidal ideas. Thus, one possible approach to effective suicide prevention is early detection of suicidal ideas.

3.1 Keywords

What do you mean "keywords, confirming the possibility of determining the post like a suicidal"? There is a certain set of words that are often used by people who have decided to commit suicide. Basically, these words are directly related to the idea of life

and death, but sometimes it happens that in posts that are written by people who are in a drooping, suicidal mood, try to avoid using words that directly mean their attempt at suicide. Nevertheless, they try to use synonyms of these same words, thereby giving us the opportunity to find their posts, using more and more new sets of keywords.

From the previous topic, we identified key words that are associated with suicide. For example, coffin, life, end, etc. These keywords will help you search for suicidal posts in social networks.

As you find suicidal posts, the database of keywords will be updated, thereby providing a more accurate definition of suicidal posts.

In order to more accurately find suicidal posts based on keywords, a meeting with experts in this field will be held to expand the database of keywords and understand the characteristic signature of suicides.

Set of available keywords: kill, die, goodbye, life is shit, coffin, don't love yourself, hate yourself, want to die, hang yourself, suicide, suicide, unrequited love, hate school, die, die, die, don't want to live, to the next world, die, heaven, hell, I'm guilty, depression, die, etc.

3.2 Data Collection

Before classifying information as suicidal or depressive, it is necessary to determine the criteria for "danger". One solution is to define a set of keywords. This method of determining the types of information was used in the developed software package. For this purpose, a set of keywords was compiled, which was used for analyzing information in the social network Vkontakte. The software package, based on the presence or absence of the specified keywords in the text, concludes that the text is suitable for further research. Figure 1 shows the entire data collection, analysis, and job classification scheme.

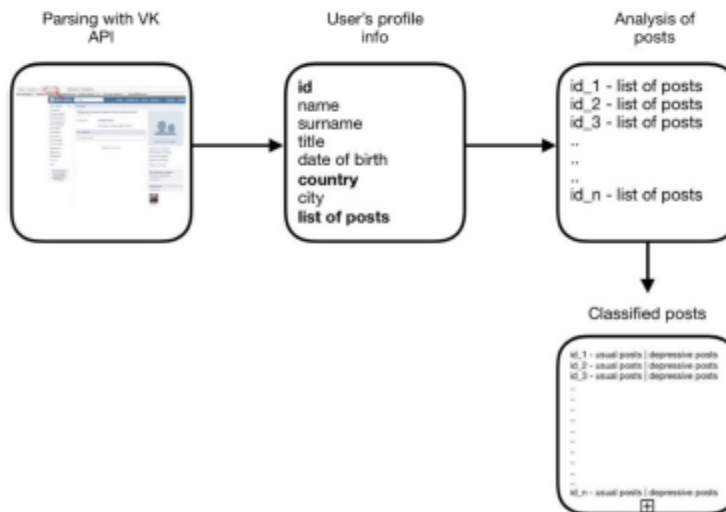


Fig. 1. Scheme of data acquisition, analysis and classification of posts

3.3 Dataset

We collected about 64,000 posts in Russian language that labeled to depressed and not depressed. 32,021 posts are non-depressed posts, when 32,018 are depressed posts. During the applying machine learning models, we divided the dataset into 2 parts as 80% for training and the remaining 20% for testing. Next four figures, from Fig. 2 to Fig. 5 describe characteristics of the developed dataset for depressive post detection.

Figure 2 illustrates characteristics of collected data.



Fig. 2. Illustrates characteristics of collected data.

Figure 3 illustrates distribution of top unigrams and bigrams of the collected dataset.

Figure 4 illustrates distribution of unigrams and bigrams taking into account only depressive posts.

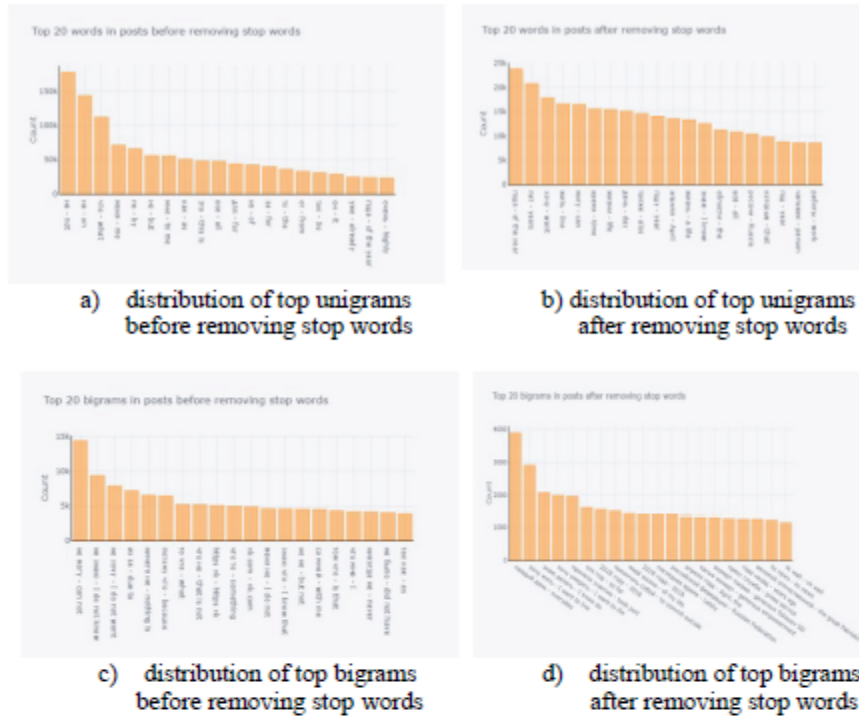


Fig. 3. Distribution of top unigrams and bigrams

Figure 5 illustrates the distribution of post length by each label.

4 Experiment Results

In this section, we tell about applying the developed dataset and results of machine learning based classification.

First, all texts were lemmatized—the process of removing only the endings and returning the basic or dictionary form of a word, which is known as a Lemma. The Yandex “Mistem” lemmatizer was used to lemmatize words in the context of the Russian language, as it demonstrated excellent results. Subsequently, the nltk library for stop words was used to delete the stop word, which reduced potential noise in the data. Numbers, special characters, and not Cyrillic letters were also removed.

Second, preprocessed texts were vectorized—the process of representing texts in a vector space for arithmetic operations on the entire data structure. The vector view saves time. For text vectorization, the IDF TF and Word2Vec models were used.

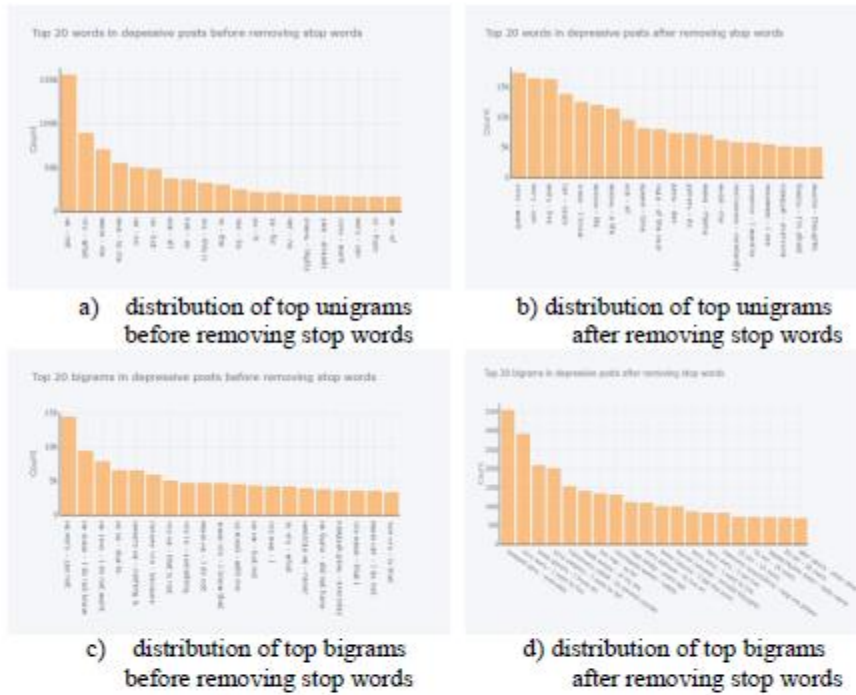


Fig. 4. Distribution of top unigrams and bigrams taking into account only depressive posts

Distribution of posts Lengths Based on labeled posts

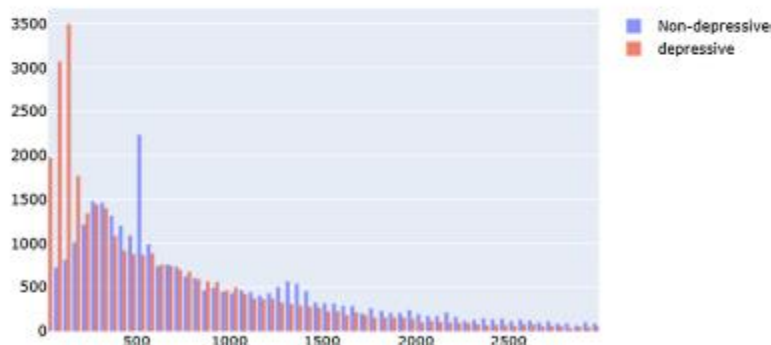


Fig. 5. Distribution of posts lengths by labels

The developed dataset applied to test eight machine learning algorithms. Table 1 demonstrates the results of each method. As evaluation parameters of methods we take accuracy, precision, recall, f1 score, and Area Under The Curve (AUC).

Table 1. Machine learning for suicidal ideation detection

Model	Accuracy	Precision	Recall	F1 score	AUC
Gradient Boosting word2vec	0.9178	0.9245	0.9427	0.8945	0.9748
Random Forest with word2vec	0.9372	0.9765	0.9215	0.9154	0.9578
Gradient Boosting with tf-idf	0.9435	0.9346	0.9215	0.9345	0.9576
Random Forest with tf-idf	0.9537	0.9462	0.9876	0.9642	0.9623
Naïve Bayes	0.8132	0.8254	0.8352	0.8124	0.9348
SVM	0.8635	0.8524	0.8864	0.8754	0.9467
XGBoost with tf-idf	0.8456	0.8325	0.8571	0.8254	0.9425
LSTM with word2vec	0.8654	0.8452	0.8576	0.8457	0.9472

5 Data Availability

The developed dataset is in open source. It can be found in [20] and [21].

6 Discussion

People now reflect their psychological state in social networks in the form of images, posts, and groups that they subscribe to. This is sufficient and even excessive to determine the psychological state of the child. With the help of one post, the psychologist can determine about 40 parameters of the child's psychological state. It is very important to note that only information that is publicly available is used. We see only what man has allowed everyone to see. We do not violate the Constitution or the boundaries of personal territory. Technologically, it is not possible to take information from closed accounts. Ethical standards are not violated.

Since the Internet is dynamic, accessible and, in fact, controlled by its users, and can also be an effective tool for intervention in the psychological state of a person, researchers agree that it is necessary to actively develop the possibilities of this intervention in a positive way. For example, interactive forums created by medical professionals can be a way to inform and support young people in order to minimize the risk of suicide and self-harm among them [22].

The main directions of research. The Internet is generally a multimedia environment, and so is content related to the SU and CX. It includes not only text, but also photos, videos, and music recordings, and all components of this multimedia content are the subject of attention of specialists [23, 24]. The subject of our special interest are the posts in the social networks and methods of its automated (computerized) research and monitoring.

7 Conclusion

The volume of text continues to grow along with the popularization of social networks. And suicide prevention remains an important task in our modern society. Therefore, it

is crucial to develop new methods for detecting online texts containing suicidal ideas, in the hope that suicide can be prevented.

In this paper, we investigated the problem of detecting suicidality in online user content. We claim that most of the work in this area has been done by psychological experts with statistical analysis that is limited by costs and the issue of confidentiality in obtaining data. By collecting and analyzing anonymous online data from the active social network Vkontakte, We provide a huge amount of knowledge that can complement the understanding of suicidal thoughts and behavior. Although we applied feature processing and classification methods to our carefully constructed data sets, Vkontakte. We evaluated, analyzed, and demonstrated that our structure can achieve high performance (accuracy) in contrast to suicidal thoughts in normal messages in online user content.

Acknowledgements. This research was supported by grant of the program of Ministry of Education of the Republic of Kazakhstan BR05236699 “Development of a digital adaptive educational environment using Big Data analytics”. We thank our colleagues from Suleyman Demirel University (Kazakhstan) who provided insight and expertise that greatly assisted the research. We express our hopes that they will agree with the conclusions and findings of this paper.

References

1. World Health Organization. (2014). Preventing suicide: a global imperative. Geneva, Switzerland (2014). http://apps.who.int/iris/bitstream/10665/131056/1/9789241564779_eng.pdf?ua=1&ua=1
2. https://vk.com/topic-78863260_30603285 - About VK Social Network
3. Brown, J.: Is social media bad for you? The evidence and the unknowns. BBC Future, January 2018
4. Naslund, J.A., Aschbrenner, K.A., Bartels, S.J.: How people with serious mental illness use smartphones, mobile apps, and social media. *Psychiatr. Rehabil. J.* **39**(4), 364 (2016)
5. Escobar-Viera, C.G., et al.: Passive and active social media use and depressive symptoms among United States adults. *Cyberpsychol. Behav. Soc. Netw.* **21**(7), 437–443 (2018)
6. VK.com – Vkontakte Social Network
7. Waytz, A., Gray, K.: Does online technology make us more or less sociable? A preliminary review and call for research. *Perspect. Psychol. Sci.* **13**(4), 473–491 (2018)
8. Villalonga-Olives, E., Kawachi, I.: The dark side of social capital: a systematic review of the negative health effects of social capital. *Soc. Sci. Med.* **194**, 105–127 (2017)
9. Wang, J.L., Gaskin, J., Rost, D.H., Gentile, D.A.: The reciprocal relationship between passive social networking site (SNS) usage and users' subjective well-being. *Soc. Sci. Comput. Rev.* **36**(5), 511–522 (2018)
10. Elhai, J.D., Tiarniyu, M., Weeks, J.: Depression and social anxiety in relation to problematic smartphone use: the prominent role of rumination. *Internet Res.* **28**(2), 315–332 (2018)
11. Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., Kross, E.: Do social network sites enhance or undermine subjective well-being? A critical review. *Soc. Issues Policy Rev.* **11**(1), 274–302 (2017)
12. Woods, H.C., Scott, H.: # Sleepyteens: social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *J. Adolesc.* **51**, 41–49 (2016)
13. Primack, B.A., et al.: Use of multiple social media platforms and symptoms of depression and anxiety: a nationally-representative study among US young adults. *Comput. Hum. Behav.* **69**, 1–9 (2017)

14. Frison, E., Eggermont, S.: Exploring the relationships between different types of Facebook use, perceived online social support, and adolescents' depressed mood. *Soc. Sci. Comput. Rev.* **34**(2), 153–171 (2016)
15. Twenge, J.M., Joiner, T.E., Rogers, M.L., Martin, G.N.: Increases in depressive symptoms, suicide-related outcomes, and suicide rates among US adolescents after 2010 and links to increased new media screen time. *Clin. Psychol. Sci.* **6**(1), 3–17 (2018)
16. Andreassen, C.S., et al.: The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: a large-scale cross-sectional study. *Psychol. Addict. Behav.* **30**(2), 252 (2016)
17. Tan, Y., Chen, Y., Lu, Y., Li, L.: Exploring associations between problematic internet use, depressive symptoms and sleep disturbance among southern Chinese adolescents. *Int. J. Environ. Res. Public Health* **13**(3), 313 (2016)
18. Primack, B.A., et al.: Use of multiple social media platforms and symptoms of depression and anxiety: a nationally-representative study among US young adults. *Comput. Hum. Behav.* **69**, 1–9 (2017)
19. Lin, L.Y., et al.: Association between social media use and depression among US young adults. *Depression Anxiety* **33**(4), 323–331 (2016)
20. Narynov, S., Mukhtarkhanuly, D., Omarov, B.: Dataset of depressive posts in Russian language collected from social media. *Data Brief* **29**, 105195 (2020)
21. <https://data.mendeley.com/datasets/838dbcjpxb/1>
22. Ersoy, M.: Social media and children. In: *Handbook of Research on Children's Consumption of Digital Media*, pp. 11–23. IGI Global (2019)
23. Olenik-Shemesh, D., Heiman, T.: Cyberbullying victimization in adolescents as related to body esteem, social support, and social self-efficacy. *J. Genet. Psychol.* **178**(1), 28–43 (2017)
24. Guntuku, S.C., Yaden, D.B., Kern, M.L., Ungar, L.H., Eichstaedt, J.C.: Detecting depression and mental illness on social media: an integrative review. *Curr. Opin. Behav. Sci.* **18**, 43–49 (2017)