2020 20th International Conference on Control, Automation and Systems (ICCAS 2020)
Oct. 13~16, 2020; BEXCO, Busan, Korea

Machine Learning Approach to Identifying Depression Related Posts on Social Media

Sergazy Narynov[1] , Daniyar Mukhtarkhanuly[1], Batyrkhan Omarov[2-4,*], Kanat Kozhakhmet[5], Bauyrzhan Omarov[3]

[1] Alem Research, Almaty, Kazakhstan (narynov@alem.kz)
[2] International Information Technology University, Almaty, Kazakhstan
[3] Al-Farabi Kazakh National University, Almaty, Kazakhstan
[4] Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan
[5] Open University of Kazakhstan
Almaty, Kazakhstan (batyahan@gmail.com) * Corresponding author

Abstract: According to the latest who data published in 2017, the number of suicides in Kazakhstan was 4855, or 3.55% of the total number of deaths. The age-adjusted death rate is 27.74 per 100,000 population. Kazakhstan is ranked 4th in the world by this indicator. This article compares machine learning algorithms with and without a teacher to identify depressive content in social media posts, with a focus on hopelessness and psychological pain for semantic analysis as key causes of suicide. Suicide is not spontaneous, and preparation for suicide can last about a year, during which time a person will show signs of their condition in our case by posting depressive content on their social network profile. This algorithm helps in detecting depressive content that can cause suicide to help people find confident help from psychologists at the national center for suicide prevention in Kazakhstan. Having obtained the highest score for 95% of the f1 score for a random forest (training with a teacher) with the tf-idf vectorization model, we can conclude by saying that the K-means algorithm(training without a teacher) using tf-idf shows impressive results that are only 4% lower in f1 and accuracy.

Keywords: machine learning, sentiment analysis, natural language processing, depression, social media.

1. INTRODUCTION

Social networks play a significant role in the lives of modern people. Research on the content of pages in social networks of people with a specific community, its impact on others is of great interest. This work is devoted to the search for the main identifiers of participants in the Vkontakte network who committed suicide, highlighting their digital footprint and identifying markers of suicidal motivation in it.

According to data provided by the world health organization, more than eight hundred thousand people die annually due to suicide, i.e. every 40 seconds a suicide is committed, while according to available data, only 30% of those who committed suicide previously reported their intentions [1]. Therefore, there is an objective need to develop methods aimed at identifying individuals who are prone to suicidal behavior and preventing suicide. The most valuable diagnostic tool that allows you to identify the features of the personality's psyche, including its propensity to suicidal behavior, is the analysis of its speech production, including at the formal-grammatical level, which is beyond the control of consciousness.

The widespread use of social media can make it possible to reduce the number of undiagnosed mental illnesses. A growing number of studies have focused on mental health in the context of social media, linking social media use and behavioral patterns to stress, anxiety, depression, suicidality, and other mental illnesses. The largest number of studies of this kind is devoted to depression. Depression is still underestimated, with about half of cases detected by primary care physicians [2] and only 13-49% receiving minimally adequate treatment [3].

Automated social network analysis potentially provides methods for early detection. If an automated process can detect increased rates of depression in a user, that person may be targeted for a more thorough assessment and provided with additional resources, support, and treatment. Research to date has either studied how the use of social media sites correlates with mental illness in users [4-7], or attempted to identify mental illness by analyzing user-generated content.

The necessary information was extracted from the Vkontakte social network [8-10]. Analysis of the data published on this page of the social network allowed us to build a dependency. 30,000 data were received. Then, using machine learning algorithms, we start classifying texts into suicidal and non-suicidal.

2. LITERATURE REVIEW

Predicting suicidal trends in Twitter data using machine learning algorithms (Marouane Birjali 2016). This particular work related to Marouane Birjali talks in this article about the proposal of a suicidal thought detection system that predicts suicidal actions using Twitter data that can automatically analyze the moods of these tweets [5, 11-13]. They are exploring a data extraction tool to extract useful information for classifying tweets collected from Twitter based on machine learning classification algorithms. And provide experimental results showing that their method of detecting suicidal acts using Twitter data and machine learning algorithms test performance in terms of recall, accuracy, and accuracy of mood analysis.

The main difference in our work is related to the volume of analysis, while our colleagues used Twitter data, which consists of short 140-character messages in

the form of headers, we decided to use VK.com, Facebook.com and yvision.kz. social platforms where users are not limited in the amount of text they can publish on their pages, thus text analysis showing us more certain results, so we collected from these sources 35,000 messages with proven depression of the authors of these messages. This data set is used to train and test various AI algorithms with controlled and uncontrolled learning.

Early detection of suicides using real-time big data Analytics (Hardik A. Patel, 2016). This article presents an application that uses predictive analysis to collect comments and posts on social networks to identify individuals on these sites who are prone to suicidal thoughts and tendencies. We also present a model that uses a branch for analyzing big data IDs and sentimental Analytics. User profiles with messages about suicide.

They concluded that the invention is a tool for suicide prevention agencies that can use it for predictive analysis using big data, identifying individuals with suicidal tendencies, which helps them to intervene in a timely manner and save lives.

Unlike Hardik Patel's work, we are not trying to offer a system, but we are constantly improving our algorithms by comparing them with each other, using solutions and world-famous AI algorithms [6].

Exploring different strategies for marking up suicide messages in social networks: a pilot study (Liu, et al. 2017)

The article by Liu and Tong attempts to provide an answer about how to get reliable, compressed data that the authors assume, which, in their opinion, depends largely on what the annotators ask, and what part of the data they mark up [7, 14-16]. They have done several rounds of data markup and collected annotations from crowdsourcing workers and are working with key experts. The resulting labels were combined in various ways to train a series of algorithms based on learning with a teacher.

Their preliminary estimates show that using unanimously agreed labels from multiple annotators is useful for creating reliable machine models. Although this article seems to be the result of a serious study, the same as in the previous article by Hardik Patel, this study is based on Twitter posts also as posts in English, in our work we use VK, FB, Twitter and various popular Russian social networks eliminate the lack of a text base, which, in our opinion, can lead to a decrease in the quality of determining the depressive component in expressions [6].

3. MATERIALS AND METHODS

3.1 Data collection

To build classifiers and compare them, we used data obtained from the Vkontakte microblogging platform. This platform was chosen due to its great popularity among young people in Kazakhstan, convenient tools for obtaining information, as well as its importance for sociological research.

Vkontakte Social Network offers a public API that allows you to programmatically collect posts as they appear, filtering by certain criteria. The VK API was used to monitor Vkontakte for any of the following words or phrases that are consistent with the popular language of suicidal thoughts:

"Suicide; commit suicide; 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 without me; better die; suicide plan; suicide Pact; tired of living; don't want to be here; die alone; fall asleep forever."

When a post corresponding to any of the above terms was identified by this tool, it was saved in this tool along with the Vkontakte profile name.

To collect data in the social network, Telegram Bat was developed, which collects posts published in real time, based on keywords found from the previous topic.

This bot collects all the data it finds in a special file that stores the text of the post, the link to the post, the link to the author of the post.

The bot works around the clock, which allows you to find posts regardless of the time zone of someone who wants to publish a post that contains the keyword that was entered.

The bot uses two technologies simultaneously: the Telegram API and the Vkontakte API. Based on the Streaming API (VK Analytics API), 1% of the published information is sent to the bot (see the documentation for the Streaming API [17]) in the Vkontakte social network.at the time of publication, the bot sends the information to the developer, simultaneously saving it.

Thus, we created dataset of depressive and suicidal intended contents from social media. The dataset was published in open source by [18] address.

Performance was measured using F-score for the positive minority class. Due to the asymmetry of the data, indicators such as accuracy, will contribute to a negative classification. F-score with standard $\beta=1$ was used to provide a harmonic mean between accuracy and recall. For our tasks, both are expected to have the same importance: find what needs to be found, but don't flood the user with false positives. In cases where the recall is of particular significance (for example, for cascading classifiers), we also discuss F-scores with $\beta=2$, so that the recall has twice the accuracy weight in the Fscore calculation.

3.2 Description of collected data

Before classifying information as suicidal or depressive, it is necessary to define the criteria of "danger". One of the solutions is the definition of a set of keywords. It is a method of determining the types of information and is applied in the developed software package. For the definition a set of keywords was compiled, which was used to analyze information on 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.

The implementation of obtaining information may vary depending on the source of information, but maintain the general principle of its construction. The main purpose of the part of the software responsible for obtaining information from open sources is to perform actions quickly and efficiently. To achieve maximum performance, you must use the built-in methods for obtaining information from sources (API), if any. If there are no such methods, then it is necessary to obtain and extract the necessary information from HTTP requests.

There are three separate modules of the software package:

1. Information collection module - is responsible for receiving information from open sources and transmitting it for further processing;
2. Keyword search module - is responsible for finding keywords in a large amount of information;
3. Document ranking module - is responsible for determining whether the information is dangerous.

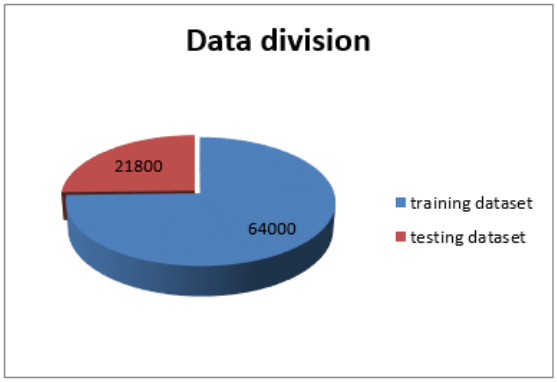

Fig. 1 The caption should be placed after the figure.

4. RESULTS

First, all texts were lemmatized — the process of deleting only endings and returning the base or vocabulary form of a word, which is known as a lemma. For the lemmatization of words in the context of the Russian language, the lemmatizer “MyStem” from Yandex was used, since it demonstrated excellent results. Subsequently, the nltk library for stop words was used to remove the stop word, hence reducing potential noise in the data. Numbers, special characters, not Cyrillic letters have also been deleted.

Secondly, the pre-processed texts were vectorized — the process of representing texts in a vector space for arithmetic operations on the entire data structure. Vector view saves time. For vectorization of texts, the TF-IDF and Word2Vec models were used.

TF-IDF stands for Term Frequency-Inverse Document Frequency, which basically indicates the importance of a word in a package or data set. TF-IDF contains two concepts: term frequency (TF) and reverse document frequency (IDF)

Word2vec is a deep learning technique with a two-layer neural network. Google Word2vec takes data from big data and converts it into vector space. Word2vec basically puts a word into feature space in such a way that their location is determined by their meaning, that is, words that have a similar meaning are grouped together, and the distance between two words also has the same meaning.

To assess the quality of classification, we used such characteristics as: precision, recall and F1 score. Precision shows the percentage of objects in the class that they actually belong to. Completeness (recall) shows how much of the objects belonging to the class were allocated during classification. They can be calculated using the following formulas:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Accuracy and completeness do not depend on the ratio of class sizes. Even if there are orders of magnitude fewer objects of one class than objects of another class, these indicators will correctly reflect the quality of the algorithm.

F1-measure-harmonic mean of accuracy and completeness:

$$F = \frac{2 * precision * recall}{precision + recall}$$

This indicator can be used as a quality criterion based on accuracy and completeness.

For the detection of depressive publications with supervised learning, we used Random Forest, Gradient Boosting [4]. While for teaching without a teacher, we used K-averages with a 2-cluster model.

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. The TF-IDF and Word2Vec models were used for text vectorization.

TF-IDF stands for Term Frequency-Inverse Document Frequency, which basically talks about the importance of a word in a corpus or dataset. TF-IDF contains two concepts: term-frequency (TF) and reverse document frequency (IDF)

Word2vec is a deep learning technique with a two-layer neural network. Google Word2vec takes data from big data and converts it into a vector space. Word2vec basically puts a word in the feature space in such a way that their location is determined by their meaning, meaning words that have a similar meaning are grouped together, and the distance between two words also has the same meaning.

For the experiment, 2 best algorithms were tested according to our previously published work, as well as with TF-idf vectorization [4]:

1. Gradient boosting using word2vec
2. Random forest with word2vec
3. Gradient boosting with tf-idf
4. Random forest with tf-idf

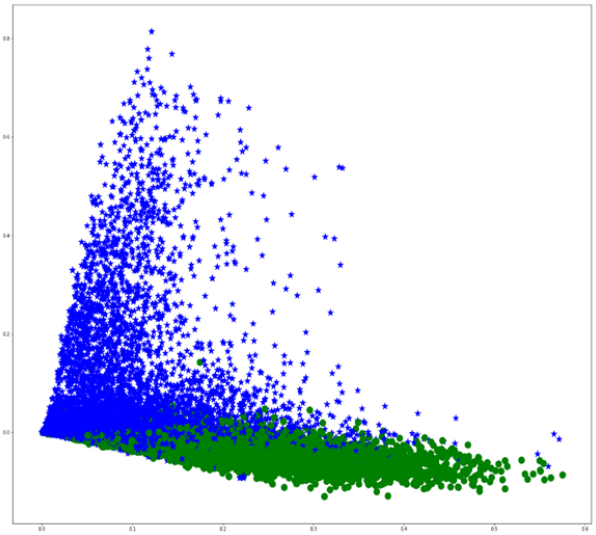

Fig. 2 Graphical representation of word2vec vectors in 2D space, where green placemarks are depressing posts and blue placemarks are normal posts.

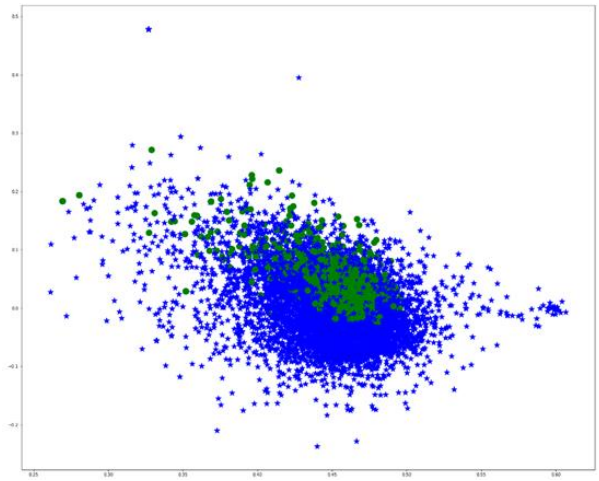

Fig. 3 Graphical representation of tf-idf vectors in 2D space, where green placemarks are depressing posts and blue placemarks are normal posts

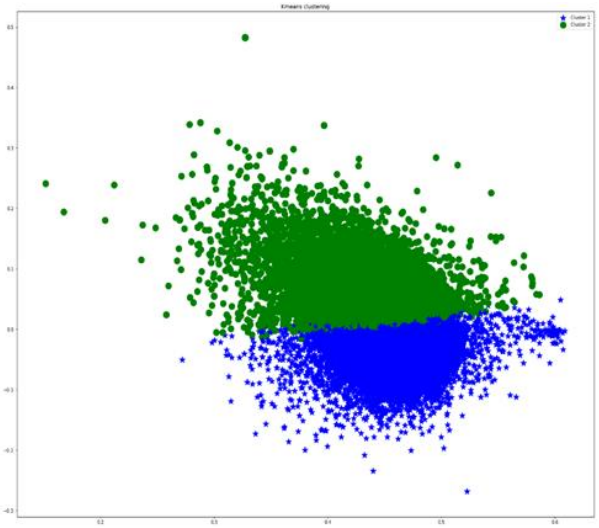

Fig. 4 Graphical representation of word2vec vectors in 2D space, where green placemarks are depressing posts and blue placemarks are normal posts

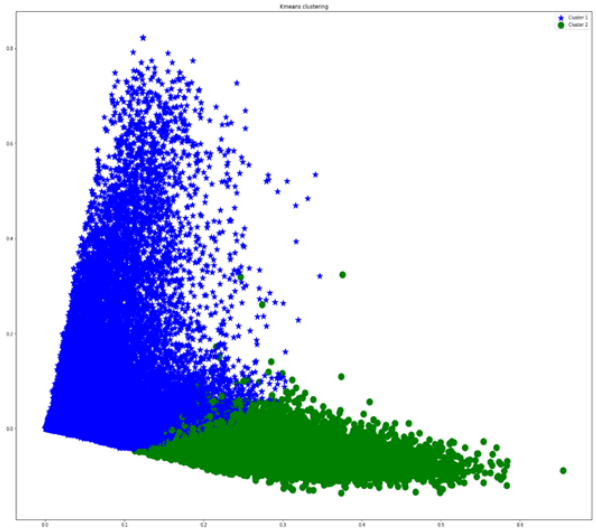

Fig. 5 Graphical representation of tf-idf vectors in 2D space, where green placemarks are depressing posts and blue placemarks are normal posts

Table 1 Depressive post classification results.

Model	Accuracy, %	Precision,%	Recall, %	F1 score, %
Gradient Boosting word2vec	90	91	91	91
Random Forest with word2vec	89	91	90	90
Gradient Boosting with tf-idf	95	96	95	95
Random Forest with tf-idf	96	96	96	96

Table 1 confirm that Gradient Boosting with tf-idf and Random Forest with tf-idf are the best classifiers for the given problem. The best supervised learning algorithm for suicidal ideation detection is Random Forest with tf-idf with 96% accuracy.

Comparison of the results of different algorithms f1-score, we can see that the Random Forest with tf-idf algorithm shows a result of 95%, which is a very good result for the given task.

To make sure that our algorithm is correct, the Receiver Operating Characteristic (ROC) curve with cross-validation was built. ROC curve was applied to understand a performance measurement for classification problem at various thresholds settings.

The "steepness" of ROC curves is also important, since it is ideal to maximize the true positive rate while minimizing the false positive rate.

Figure 9 shows the ROC curve of different train and test datasets, created from K-fold cross-validation. Taking all of these curves, it is possible to calculate the mean area under curve, and see the variance of the curve when the training set is split into different subsets. This roughly shows how the classifier output is affected by changes in the training data, and how different the splits generated by K-fold cross-validation are from one another.

According to the graph, we see a stable result and we can be sure that the algorithm is well-trained to identify depressive posts.

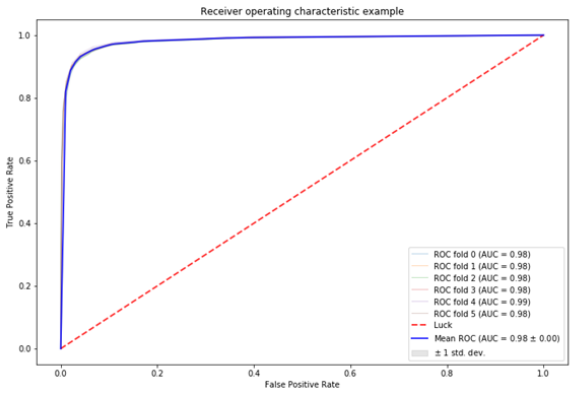

Fig. 6 ROC curves with cross-validation.

Children and teenagers 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.

CONCLUSION

The spread of social networks in the world is increasing, which means that more and more people will be available to participate in research through social networks.

In this paper, we implemented different algorithms of supervised and unsupervised learning methods. We obtained f1-score more than 90% and ROC-area 0.98 with Random Forest with tf-idf vectorization model. By comparing with our previously built algorithm we increased prediction by almost 20%. We also tested how unsupervised model will perform on that dataset and it surprisingly showed great results.

There are several interesting directions of future work. One of them is to implement deep learning models with PyTorch framework. An alerting system will be built for the government to monitor emotional state of a person to prevent possible suicide attempts or any self-inflicting injuries.

We raised a very foundational research question about determining the depressive posts in social media and concerned about anonymity of the data, especially when the topic is sensitive and ambiguous. We controlled parameters of training algorithms, validated it with ROC curve, and visualized results in 2D space. In addition, we made it open-source project, for future commits and changes.

In this regard we have achieved our initial goal. In the next phase of our research, we are going to apply audio, video and text analysis to identify depressed and suicidal people on the social network. We are also going to publish collected data from social media with suicidal, depressive and neutral messages that contain suicidal keywords in a data paper as a data for machine learning purpose to identify suicidal and depressive posts in social networks.

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] Ajdacic-Gross V, Weiss MG, Ring M, Hepp U, Bopp M, Gutzwiller F, et al. Methods of suicide: international suicide patterns derived from the WHO mortality database. Bull World Health Organ 2008;86:726-32.

[2] Sowa, N. A., Jeng, P., Bauer, A. M., Cerimele, J. M., Unützer, J., Bao, Y., & Chwastiak, L. (2018). Psychiatric case review and treatment intensification in collaborative care management for depression in primary care. Psychiatric Services, 69(5), 549-554.

[3] George, D. L. (2018). Assessing the efficacy of a self-administered treatment for social anxiety disorder in the form of gamified mobile application (Doctoral dissertation, Appalachian State University).

[4] Xu, Z., Pérez-Rosas, V., & Mihalcea, R. (2020, May). Inferring Social Media Users' Mental Health Status from Multimodal Information. In Proceedings of The 12th Language Resources and Evaluation Conference (pp. 6292-6299).

[5] https://vk.com/ - Vkontakte Social Network

[6] Souri, A., Hosseinpour, S., & Rahmani, A. M. (2018). Personality classification based on profiles of social networks' users and the five-factor model of personality. Human-centric Computing and Information Sciences, 8(1), 24.

[7] Primack, B. A., & Escobar-Viera, C. G. (2017). Social media as it interfaces with psychosocial development and mental illness in transitional age youth. Child and Adolescent Psychiatric Clinics, 26(2), 217-233.

[8] Kharlamov, A. A., Orekhov, A. V., Bodrunova, S. S., & Lyudkevich, N. S. (2019, December). Social Network Sentiment Analysis and Message Clustering. In International Conference on Internet Science (pp. 18-31). Springer, Cham.

[9] Birjali, M., Beni-Hssane, A., & Erritali, M. (2017). Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. Procedia Computer Science, 113, 65-72.

[10] Baecchi, C., Uricchio, T., Bertini, M., & Del Bimbo, A. (2016). A multimodal feature learning approach for sentiment analysis of social network multimedia. Multimedia Tools and Applications, 75(5), 2507-2525.

[11] Pravalika, A., Oza, V., Meghana, N. P., & Kamath, S. S. (2017, July). Domain-specific sentiment analysis approaches for code-mixed social network data. In 2017 8th international conference on computing, communication and networking technologies (ICCCNT) (pp. 1-6). IEEE.

[12] Romanowski, A., & Skuza, M. (2017). Towards predicting stock price moves with aid of sentiment analysis of Twitter social network data and big data processing environment. In Advances in Business ICT: New Ideas from Ongoing Research (pp. 105-123). Springer, Cham.

[13] Jain, A. P., & Dandannavar, P. (2016, July). Application of machine learning techniques to sentiment analysis. In 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) (pp. 628-632). IEEE.

[14] Fersini, E. (2017). Sentiment analysis in social networks: a machine learning perspective. In Sentiment Analysis in Social Networks (pp. 91-111). Morgan Kaufmann.

[15] Bi, Q., Shen, L., Evans, R., Zhang, Z., Wang, S., Dai, W., & Liu, C. (2020). Determining the Topic Evolution and Sentiment Polarity for Albinism in a Chinese Online Health Community: Machine Learning and Social Network Analysis. JMIR Medical Informatics, 8(5), e17813.

[16] Bharti, S. K., Pradhan, R., Babu, K. S., & Jena, S. K. (2017). Sarcasm analysis on twitter data using machine learning approaches. In Trends in Social Network Analysis (pp. 51-76). Springer, Cham.

[17] https://vk.com/dev/streaming_api - a tool for getting public data from Vkontakte for specified keywords.

[18] Narynov, S., Mukhtarkhanuly, D., & Omarov, B. (2020). Dataset of depressive posts in Russian language collected from social media. Data in brief, 29, 105195.