Journal of Theoretical and Applied Information Technology
15th November 2020. Vol.98. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

APPLYING NATURAL LANGUAGE PROCESSING TECHNIQUES FOR SUICIDAL CONTENT DETECTION IN SOCIAL MEDIA

[1]SERGAZY NARYNOV, [1,2] KANAT KOZHAKHMET, [1]DANIYAR MUKHTARKHANULY, [3]NURBEK SAPARKHOJAYEV

[1]Alem Research, Almaty, Kazakhstan

[2]Open University of Kazakhstan

[3]Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan

E-mail: narynov@alem.kz

ABSTRACT

This research paper discusses computer solutions in the field of detecting and analyzing Internet texts containing suicidal and depressive content in social networks. Experience in the analysis and detection of suicidal messages, self-harming nature, accumulated by research teams in different countries, provides an opportunity to assess progress in this area and the difficulties that must be taken into account when developing solutions for the local Internet content.

Keywords: *Natural Language Processing, Suicidal Content, Suicidal Ideation Detection, Machine Learning, 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], or attempted to identify mental illness by analyzing user-generated content.

The necessary information was extracted from the Vkontakte social network [5]. 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.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. LITERATURE REVIEW

The study of social networks is becoming more and more relevant every year due to the growing need to ensure the safety of the population and monitor public attitudes. Basically, the purpose of most research is to identify implicit relationships between user groups, as well as their behavior. At the same time, it is important to analyze the individual data set of each user, including information about their interactions, the content generated by this user, as well as information about interaction with the content of other users. Based on the results of the analysis of this data set, it is possible to draw conclusions about the user's personality, character and mores [6], which contributes to the assessment of the situation in the self-organization of society and its individual groups.

A few years ago, such analysis was very difficult to implement, because computer computing power was much lower. SEI hour, using modern technologies and tools, you can independently analyze the open data of users of social networks.

According to social research data [7] more than 2 billion people worldwide are active users of blogs and social networks. All these people leave a huge number of messages every day, reflecting the positions of citizens from different countries and different strata of society. Analysis of such content can help assess changes in the attitudes of many users and can be used in political and social research, including research on consumer preferences. Most often, to solve such problems, they resort to analyzing the tone of text messages.

Processing texts manually in order to extract useful information requires too much time and human resources. In most cases, so many are given that this makes the task impossible. To solve this problem, there are various automatic methods for analyzing text in natural languages, including automatic analysis of the tonality of texts.

Tonality is the emotional attitude of the author of a statement to some object expressed in the text. There are several methods for determining the tone of a text. let's consider a method based on rules and dictionaries, as well as machine learning with a teacher. Text classification as an approach used to obtain knowledge from unstructured data, including for determining the tone of text messages, is described in detail in [8].

Sentiment analysis is a class of content analysis methods in computational linguistics designed for automated detection of emotionally colored vocabulary in texts and emotional evaluation of the text in relation to the objects in question in the text. The main task of analyzing the tonality of texts is to extract emotional color from the texts. Emotional color can be defined as "positive"," negative", or it can take values from a certain range, for example, from 0 to N. More and more research is being conducted in this area, and there are currently many different approaches that are already being used to solve a large number of practical problems.

Modern methods used for automatic sentiment analysis can be divided into several large groups: methods based on rules and dictionaries [9-12], methods using machine learning [13-16], and hybrid methods where lexical resources (tonal dictionaries) are combined with machine learning techniques [17].

2.1 Rule and dictionary based method

Using this method, tonal dictionaries and rules are compiled using linguistic analysis, which are used to search for emotive vocabulary. Further, the set of negative emotive vocabulary is evaluated on a scale containing the number of negative and positive vocabulary.

The following algorithm is usually used:

- assign each word in the text a key value from the dictionary (if it is present in it);

- calculate the overall tonality of an entire text by summing the tonality of individual words.

The disadvantages of this method is a significant amount of labor, because it requires the creation of a large number of rules.

2.2 Social Network Analysis

Social networks involve an increasing number of participants in their infrastructure. The virtualization of society is in an active stage and the use of the Internet has become widespread, which makes it possible to talk about the total intervention of virtual reality in the life of humanity. Society is moving to a new socio-economic formation, as indicated by the development of the information space, one of the components of which is the virtual space of the global Internet. In the context of globalization of all spheres of activity, it is possible

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to identify changes occurring in the psychological state of Internet users.

Each person has distinctive features of behavior in social networks, and the behavior can be deviant, triggering the mechanism of anti-social behavior of the user

The expression of negative emotions is not a serious problem in most cases, but regular or unexpected publication of posts, including emojies or hashtags that contain content of aggression, cruelty or irritation, may indicate the presence of psychological problems.

The time of publication may also indicate some deviations. Publications late at night or early in the morning, in some cases, are signs of insomnia, which is often a consequence of depression or other mental problems. Unusually late or early activity in social networks is not always a worrying sign in itself, but in combination with other indicators, it can be used for a detailed analysis [18].

2.3 Supervised Learning

Not only social networks, but also virtual reality and machine learning are used to identify signs that indicate the manifestation of deformity. Scientific approaches offer unexpected opportunities to identify interrelated social, psychological, biological and cultural factors that affect the change in the psychological state of an individual. Experiments using virtual reality and artificial intelligence identify those who are most at risk, and make it possible to predict who is most susceptible to deviant behavior. Together, they represent the best prevention in the future [19].

Supervised Machine Learning is the most common method. Its essence is to build a model on a collection of pre-marked texts, and then use it to analyze new documents.

This method is based on the creation of a machine classifier, which at the first stages is trained on pre-marked texts, and then, using this knowledge, builds a model for analyzing new documents. The short algorithm of the method is as follows:

– first, a collection of documents is collected, on the basis of which the machine classifier is trained;

– each document is laid out as a vector of features (aspects) for which it will be studied;

- specify the correct tone type for each document;

– selects the classification algorithm and method for training the classifier;

– the resulting model is used to determine the tone of documents in the new collection.

I would like to note that modern society is increasingly using virtual space to expand opportunities and prospects in all spheres of life. It should also be noted that there is a transfer of reality to the virtual space. When conducting a detailed analysis of changes in personal qualities using the proposed methods of social network analysis, as well as the use of machine learning methods, conclusions can be drawn about the following individual psychological characteristics: determining the level of stress, awareness and management of their own emotional States.

As a disadvantage of the method of teaching with a teacher, the need for a sufficiently large amount of data for training can be identified, but in the future, the model obtained from this data will allow automatic analysis of new texts.

In the article [20], the main approach is to use such classifiers as the naive Bayesian classifier, the support vector machine (SVM) method, and MaxEntropy. This allows you to achieve accuracy of up to 82.9% in determining the emotional color of documents belonging to a specific subject area. In [21], we showed a strong dependence of the accuracy of tonality classification on the topics of the training and test collection. Data sets from various subject areas were used to evaluate the model: news articles on various topics and movie reviews. The results of testing a model obtained on a training set from a different subject area are much worse than on a test set of the same subject.

2.4 Unsupervised Learning

Using machine learning methods without a teacher does not require pre-marked texts, but this approach is significantly less accurate than machine learning with a teacher.

The author [22] carried out classification using the average value of semantic orientation of phrases that contain adjectives and adverbs. Finding the meaning of semantic orientation was based on calculating the mutual information [23] of phrases with the words ”excellent” and ”poor”. Statistics obtained by the search engine were used to calculate the mutual information. Reviews of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

restaurants, laptops, and hotels were used as data sets. The average classification accuracy in this paper is about 74%.

Thus, we can conclude that this problem is very relevant and there are many approaches to solve it. However, each method has its own characteristics that must be taken into account when used in solving real practical problems.

2.5 Related Works

[24] was also based on Instagram material, but unlike the previous One, it looked at messages that involved direct self-harm and were provided with corresponding German hashtags. The choice of material was due, on the one hand, to the fact that Germany occupies one of the first places in terms of non-suicidal self-harming among young people, on the other hand, to the authors ' desire to focus on one national segment of the Internet (English hashtags are used by residents of many countries) The starting point of the analysis was photos published by users for 4 weeks in April 2016, which openly depicted such damage. The number of such photos was 2,826 in 1,154 accounts. The content of photos, comments to them, and hashtags were studied. Most of these photos showed cuts on the hands and feet, which were classified by experts as minor or moderate injuries.

The step-by-step method of material selection used by the authors is noteworthy. At first, they uploaded messages from open Instagram profiles within 48 hours with a single hashtag, which is reportedly the most popular for non-suicidal self-harming content in Germany [25]. 1,135 images were obtained. In the second stage, 30 hashtags were identified that were used for these images at least twice, and photos tagged with these hashtags were uploaded within 72 hours. The number of uploaded photos was 5,588. Then two independent experts determined whether these photos were relevant for the purpose of the study, that is, whether they contained a direct image of a wound or scar (other people's photos copied from known sources were not included in the review). In the resulting sample, there were 293 photos, based on which the next stage of refining the list of hashtags was carried out. Hashtags that marked at least 5% of relevant photos were used for the final upload of the material. Then, for 4 weeks, photos with these hashtags were uploaded from Instagram every hour, which eventually amounted to 32,182 photos from 6,721 accounts. This number is particularly impressive given that 10 out of 16 hashtags used receive an automatic warning, in accordance with Instagram's moderation policy. However, an expert assessment of these photos revealed that only 2,826 (8.8%) of them were relevant, published in 1,154 accounts. This result is interesting in itself, as it demonstrates that self-harming hashtags in a message are rarely accompanied by a direct image of damage.

In recent years, studies using these software tools have analyzed the texts of diaries, letters (including those published on the Internet) and lifetime interviews of people who have committed suicide [26]. a review of this kind of work is presented in D. Lester [27].

The task of developing a mathematical model that would predict the probability of suicidal behavior based on the analysis of speech production is set in the Durkheim Project funded by the us Department of defense [28], the material of which is the texts, including Internet communications, of war veterans, some of whom subsequently committed suicide [29]. Based on the developed models, the researchers plan to further identify individuals with an increased risk of suicidal behavior online based on the analysis of their language features of their texts. Currently, the mathematical model developed by the authors of the project, based on machine learning methods, takes into account only the frequency of individual full-meaning words and phrases (the accuracy of classifying the author of the text as suicidal/non-suicidal was 65 %) [30]. Formal grammatical parameters of the text that are not controlled by the author's consciousness were not analyzed, while the informativeness of such parameters for diagnosing various personality parameters (gender, age, psychological characteristics, etc.) was proved in numerous studies ([31]).

In the work of [32], a model was proposed that automatically predicts the suicidal tendency of a message, created and trained on the basis of data collected in the Twitter network. From an array of 14,701 tweets found for suicide-related keywords, 2,000 tweets were randomly selected for experts who classified them as "causing serious concern", i.e. containing a serious and open expression of intent to commit suicid3 (14%); "possibly worthy of concern" was the default category that included tweets that did not contain signs of the other two (56%); "safe enough to ignore" (29%). The consensus of experts ' opinions was 76%.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The model was assigned the task of automatically detecting messages that "cause great concern". A total of 1,820 tweets were used for machine learning and model testing, divided into two sets: set A – 829 tweets (746 for training, 83 for testing) and set B=991 (891 and 100, respectively). The model correctly identified 80% of messages that "cause severe anxiety", which is significantly better than the results of the model [33] created on the material My Space, and achieved only 14% accuracy in determining suicide risk from blog texts. It is also important that the machine has reached the same level of accuracy as the results of manual analysis.

1. Word processing included the following stages:

2. Lemmatization.

3. Converting words to vectors.

4. Calculation of tf-idf 4 coefficients that take into account the information content of a word based on its frequency.

The authors conducted tests and calculated the accuracy (precision), the completeness (recall) and F1-measure 8 for each combination. All these measures are important for evaluating the quality of the model. High accuracy means that the model is rarely wrong when it claims that fasting is suicidal. In turn, the high completeness indicates that the model does not miss enough "disturbing" posts. Thus, the model may have high accuracy but low completeness (an uncertain model), or high completeness and low accuracy (an overly confident model). The first case is good when we need to block posts automatically, and we don't want to block "good" posts. The second case is good if a secondary review (moderation) is planned later. The authors were not able to achieve high accuracy and completeness at the same time, but the results obtained can be used as a basis for a better algorithm.

3. TOOLS

To solve this problem, we used the following technologies: Python, SciPy, Pandas, VK API [34] and machine learning algorithms. The extraction of the necessary information from the social network was carried out using the VK API.

The API — application programming interface) is an intermediary between the application developer and any environment that the application should interact with. The API simplifies code creation by providing a set of ready-made classes, functions, or structures for working with existing data. The VK API is an interface that allows you to get information from a database vk.com using http requests to a special server. You don't need to know in detail how the database is organized, what tables and fields of what types it consists of — just that the API request "knows"about it. The syntax of queries and the type of data they return are strictly defined on the service side. For more information about API features, see the official documentation [35].

3.1 Machine Learning Techniques

Automated social network analysis is performed by building predictive models that use "features" or variables extracted from social network data. For example, commonly used functions include user language encoded as the frequency of each word, message time, and other variables. Then the objects are considered as independent variables in the algorithm (for example, linear regression [36] with built-in variable selection [37] or support vector machines (SVM)) [38-40] to predict the dependent variable of the result of interest (for example, the mental health of users). Predictive models are trained using an algorithm on a part of the data (the training set) and then evaluated on another part (the test set) to avoid over — matching-a process called cross-validation. The forecast results are then presented as one of several possible indicators.

4. MATERIALS AND METHODS

4.1 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; 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 [41]) in the Vkontakte social network.at the time of publication, the bot sends the information to the developer, simultaneously saving it.

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 β=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 β=2, so that the recall has twice the accuracy weight in the Fscore calculation.

4.1 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.

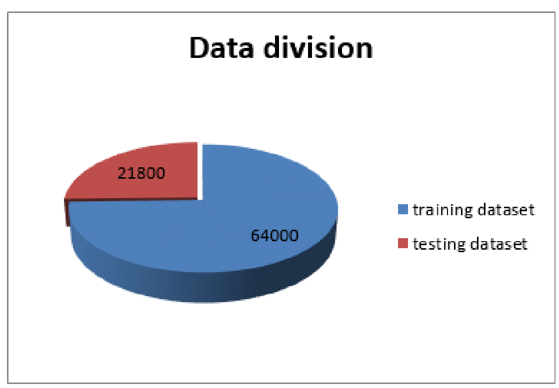

Figure 1: Data division for machine learning process

We collected 35,000 messages in Russian for people diagnosed with depression at various levels: severe, chronic, manic, persistent depressive disorder, and so on. We also used about 50,000 personal posts from social networks with negative attitudes on various topics to create a homogeneous

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

text base, divided into a training base for learning algorithms and test bases for evaluating the performance of algorithms. Training and testing dataset divided by 74.5% to 24.5% as illustrated in Figure 1 [42].

The data differs significantly from the usual texts used in classification tasks. To write tweets, users use informal speech, which may include mentions of other users, links, hashtags, and emoticons.

To improve the quality of classification and reduce the dimension of the feature vector, the following actions were performed:

- Delete stop words.

- Replacing all hyperlinks with ‘LINK’, and all user references with 'USERNAME'.

- Removing punctuation marks.

- Removing special characters, emoticons, and emojis.

- Deleting hashtags.

– All uppercase letters are reduced to lowercase.

The corps consisting of automatically placed posts was divided into two parts. 80% was used for training, 20% for testing. Then the quality of classification was evaluated on the same data set.

Figure 2-7 illustrates characteristics of collected data. Figure 2 demonstrates distribution of text length in the collected dataset.

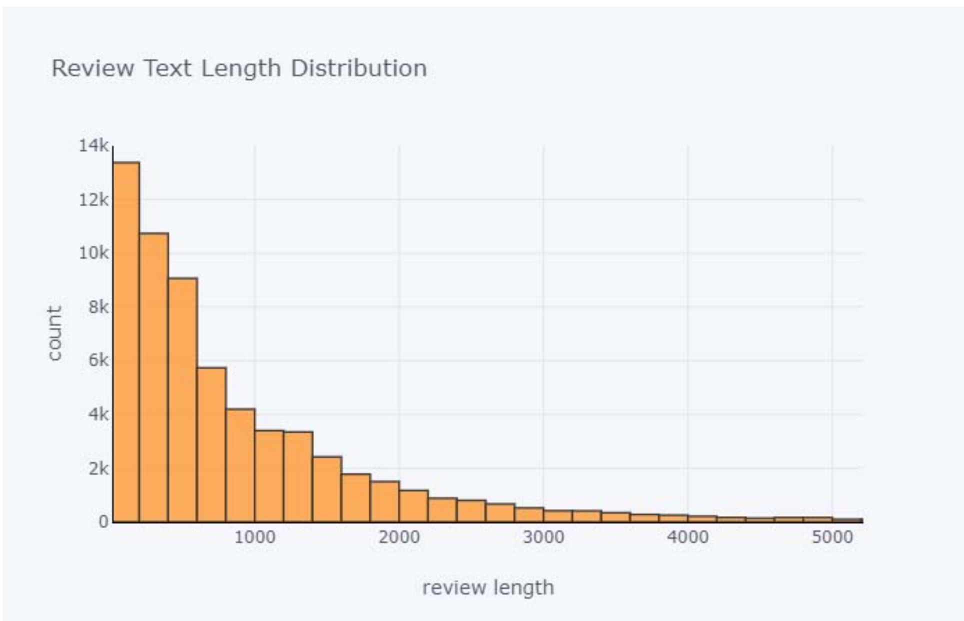

Figure 2: Distribution of text length

Figure 3 illustrates age distribution of suicidal post authors.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

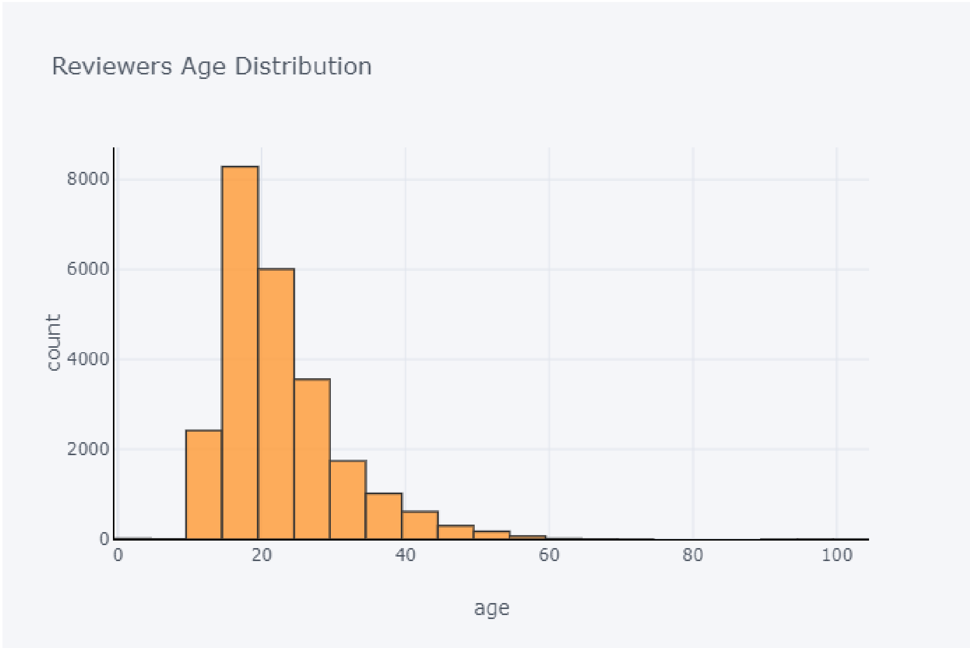

Figure 3: Age distribution

Figure 4 illustrates top unigrams before removing stop words.

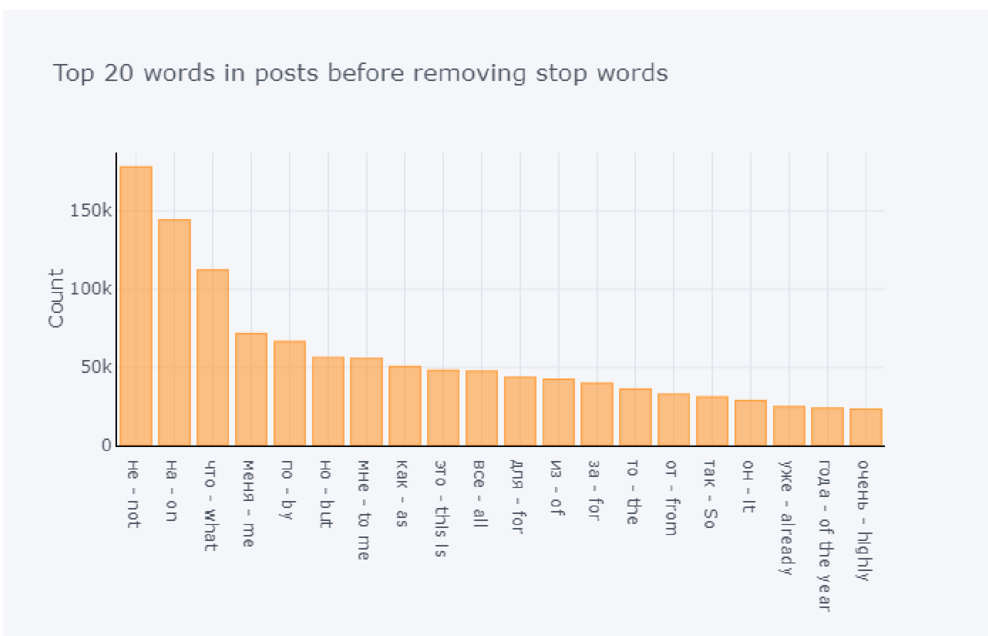

Figure 4: Top unigrams before removing stop words

Figure 5 illustrates top unigrams after removing stop words.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

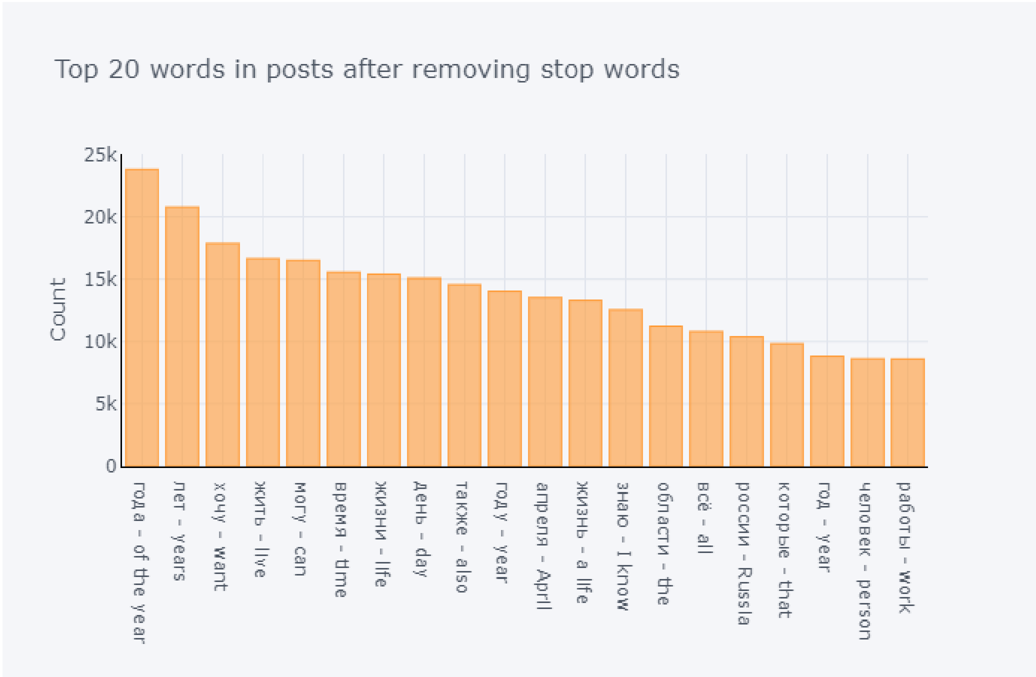

Figure 5: Top unigrams after removing stop words

Figure 6 shows top bigrams before removing stop words.

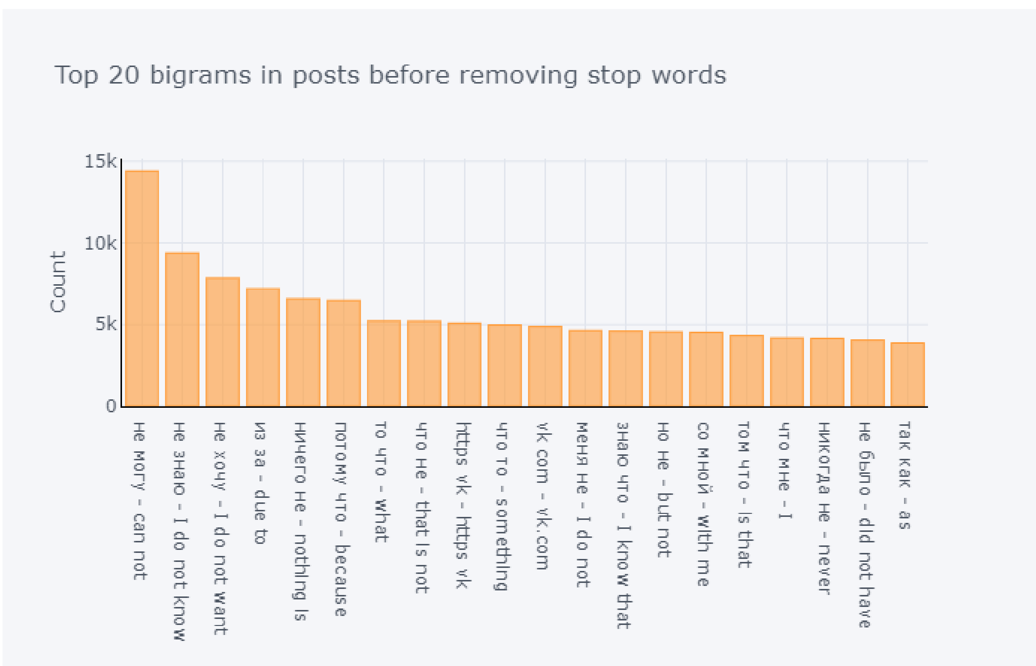

Figure 6: Top bigrams before removing stop words

Figure 7 shows distribution of top bigrams after removing stop words.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

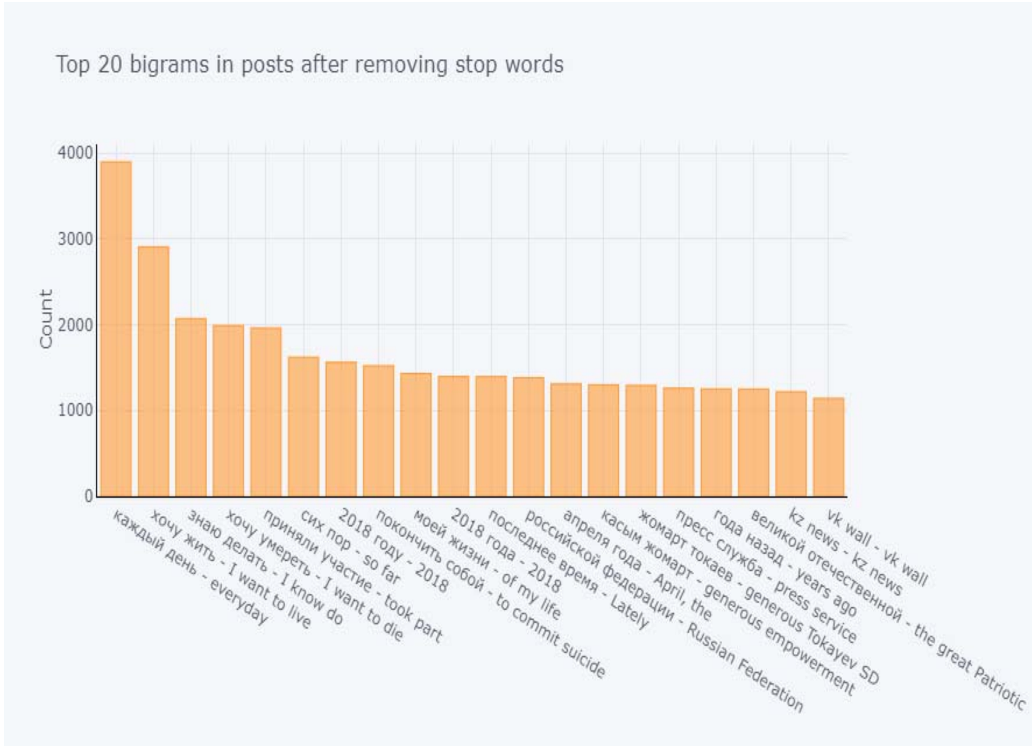

Figure 7: Top bigrams after removing stop words

Figure 8 illustrates the distribution of post length by each label.

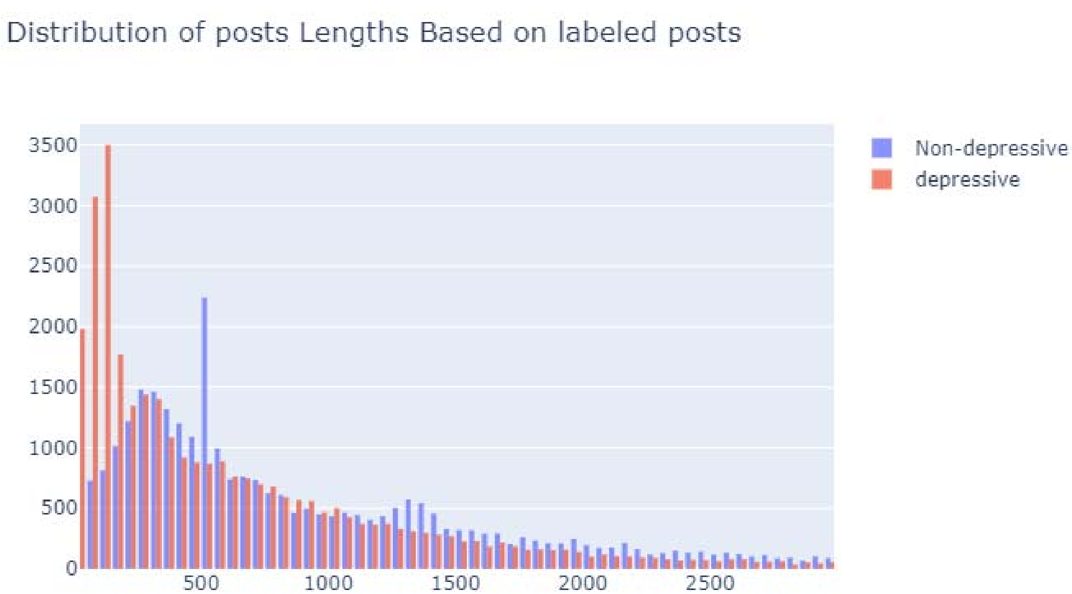

Figure 8: Distribution of post length

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. RESULTS AND DISCUSSION

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 [43] and F1 score [44]. 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.

Table 1. Supervised learning algorithm for suicidal ideation detection

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

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

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.

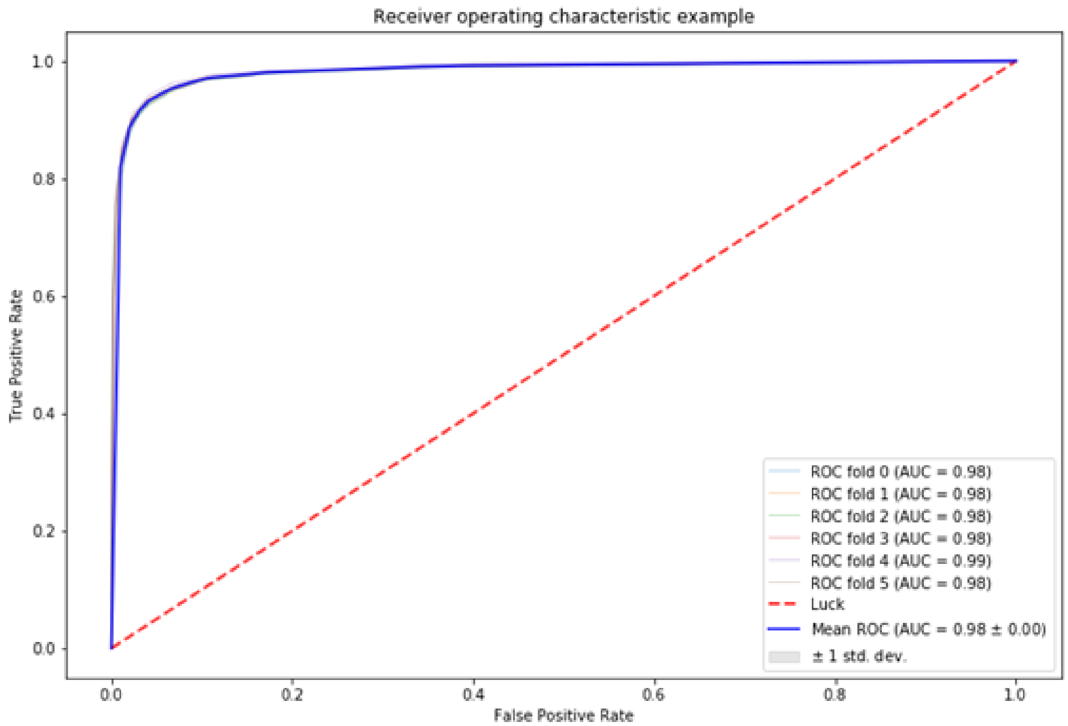

Figure 9: ROC[i] curves with cross-validation

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.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

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.

REFRENCES:

[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

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

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] Sapountzi, A., & Psannis, K. E. (2018). Social networking data analysis tools & challenges. Future Generation Computer Systems, 86, 893-913.

[18] Ding, Y., Yan, S., Zhang, Y., Dai, W., & Dong, L. (2016). Predicting the attributes of social network users using a graph-based machine learning method. Computer Communications, 73, 3-11.

[19] Saqr, M., Nouri, J., & Santolini, M. (2019). Towards group-aware learning analytics: using social network analysis and machine learning to monitor and predict performance in collaborative learning.

[20] Aldarwish, M. M., & Ahmad, H. F. (2017, March). Predicting depression levels using social media posts. In 2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS) (pp. 277-280). IEEE.

[21] Chiroma, F., Cocea, M., & Liu, H. (2019, September). Detection of suicidal Twitter posts. In UK Workshop on Computational Intelligence (pp. 307-318). Springer, Cham.

[22] Turney Peter D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews // Proceedings of the 40th annual meeting on association for computational linguistics / Association for Computational Linguistics. — 2002. — P. 417–424.

[23] Khan, F. H., Qamar, U., & Bashir, S. (2016). SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection. Applied Soft Computing, 39, 140-153.

[24] Scherr, S., Arendt, F., Frissen, T., & Oramas, J. (2019). Detecting intentional self-harm on Instagram: development, testing, and validation of an automatic image-recognition algorithm to discover cutting-related posts. Social Science Computer Review, 0894439319836389.

[25] Omarov, B. (2017, October). Exploring uncertainty of delays of the cloud-based web services. In 2017 17th International Conference on Control, Automation and Systems (ICCAS) (pp. 336-340). IEEE.

[26] Gesi, C., Carmassi, C., Miniati, M., Benvenuti, A., Massimetti, G., & Dell'Osso, L. (2016). Psychotic spectrum symptoms across the lifespan are related to lifetime suicidality among 147 patients with bipolar I or major depressive disorder. Annals of general psychiatry, 15(1), 1-7.

[27] Omarov, B., Suliman, A., & Tsoy, A. (2016). Parallel backpropagation neural network training for face recognition. Far East Journal of Electronics and Communications, v.16, no.3, pp.471 – 491. http://dx.doi.org/10.17654/EC016030471

[28] Malgaroli, M., Hull, T. D., Bantilan, N., Ray, B., & Simon, N. (2020). Suicide Risk Automated Detection Using Computational Linguistic Markers From Patients' Communication With Therapists. Biological Psychiatry, 87(9), S444.

[29] Altayeva, A., Omarov, B., Suleimenov, Z., & Im Cho, Y. (2017, June). Application of multi-agent control systems in energy-efficient intelligent building. In 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS) (pp. 1-5). IEEE.

[30] 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.
[31] Jain, S., Narayan, S. P., Dewang, R. K., Bhartiya, U., Meena, N., & Kumar, V. (2019, May). A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter. In 2019 IEEE Students Conference on Engineering and Systems (SCES) (pp. 1-6). IEEE.
[32] Huang, Y., Liu, X., & Zhu, T. (2019, August). Suicidal Ideation Detection via Social Media Analytics. In International Conference on Human Centered Computing (pp. 166-174). Springer, Cham.
[33] Just, M. A., Pan, L., Cherkassky, V. L., McMakin, D. L., Cha, C., Nock, M. K., & Brent, D. (2017). Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. Nature human behaviour, 1(12), 911-919.
[34] https://vk.com/dev/methods
[35] https://vk.com/dev/products
[36] Rohini, K. R., Rajendran, G., Arjun, V., & Poornachandran, P. (2019, December). Early Detection of Suicidal Predilection on Instagram Social Networking Service. In Symposium on Machine Learning and Metaheuristics Algorithms, and Applications (pp. 255-263). Springer, Singapore.
[37] Rajesh Kumar, E., Rama Rao, K. V. S. N., Nayak, S. R., & Chandra, R. (2020). Suicidal ideation prediction in twitter data using machine learning techniques. Journal of Interdisciplinary Mathematics, 23(1), 117-125.
[38] Sawhney, R., Manchanda, P., Mathur, P., Shah, R., & Singh, R. (2018, October). Exploring and learning suicidal ideation connotations on social media with deep learning. In Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 167-175).
[39] Onalbek, Z. K., Omarov, B. S., Berkimbayev, K. M., Mukhamedzhanov, B. K., Usenbek, R. R., Kendzhaeva, B. B., & Mukhamedzhanova, M. Z. (2013). Forming of professional competence of future tyeacher-trainers as a factor of increasing the quality. Middle East Journal of Scientific Research, 15(9), 1272-1276.
[40] Mann, P., Paes, A., & Matsushima, E. H. (2020, May). See and Read: Detecting Depression Symptoms in Higher Education Students Using Multimodal Social Media Data. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 14, pp. 440-451). Trupthi, M., Pabboju, S., & Narasimha, G. (2017, January). Sentiment analysis on twitter using streaming API. In 2017 IEEE 7th International Advance Computing Conference (IACC) (pp. 915-919). IEEE.
[41] Avola, D., Cinque, L., Foresti, G. L., Lamacchia, F., Marini, M. R., Perini, L., ... & Telesca, G. (2019, October). A Shape Comparison Reinforcement Method Based on Feature Extractors and F1-Score. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 2155-2159). IEEE.
[43] Narynov, S., Mukhtarkhanuly, D., & Omarov, B. (2020). Dataset of depressive posts in Russian language collected from social media. Data in brief, 29, 105195.
[42] AlBeladi, A. A., & Muqaibel, A. H. (2018). Evaluating compressive sensing algorithms in through-the-wall radar via F1-score. International Journal of Signal and Imaging Systems Engineering, 11(3), 164-171.
[43] Mikal, J., Hurst, S., & Conway, M. (2017, August). Investigating patient attitudes towards the use of social media data to augment depression diagnosis and treatment: a qualitative study. In Proceedings of the fourth workshop on computational linguistics and clinical psychology—from linguistic signal to clinical reality (pp. 41-47).