Depression Detection from Social Media Texts

Maxim Stankevich[1], Andrey Latyshev[2], Evgenia Kuminskaya[3], Ivan Smirnov[4,5],
and Oleg Grigoriev[6]

[1] Federal Research Center "Computer Science and Control" of RAS, Moscow, Russia
stankevich@isa.ru
[2] Limited Liability Company "RI Technologies", Moscow, Russia
andrey.latyshev@gmail.com
[3] Psychotherapy and Counselling Psychology FGBNU PI RAE, Moscow, Russia
evgenia.kuminskaya@gmail.com
[4] Federal Research Center "Computer Science and Control" of RAS, Moscow, Russia
[5] Peoples' Friendship University of Russia (RUDN University), Moscow, Russia
ivs@isa.ru
[6] Federal Research Center "Computer Science and Control" of RAS, Moscow, Russia
oleggpolikvart@yandex.ru

Abstract. Nowadays the problem of early depression detection is one of the most important in the field of psychology. Social networks analysis is widely applied to address this problem. In this paper, we consider the task of automatic detection of depression signs from textual messages of Russian social network VKontakte users. We describe the preparation of users' profiles dataset and propose psycholinguistic and stylistic markers of depression in text. We evaluate machine learning methods for detecting signs of depression from social media messages. The results of experiments show that psycholinguistic markers based features achieved 66% of F1-score on the binary classification task which is promising result in comparison with similar works.

Keywords: Depression Detection, Social Networks, Psycholinguistics.

1   Introduction

Nowadays the problem of early depression detection is one of the most important in the field of psychology. Over 350 million people worldwide suffer from depression, which is about 5% of the total population. Close to 800 000 people die due to suicide every year and it is statistically the second leading cause of death among people in 15–29 years old [1, 2]. At the same time, the major number of suicides associated with depression. Recent researches reveal that depression is also the main cause of disability and a variety of somatic diseases.

For example, F. I. Beliakov [3] in his paper summarizes the main results of recent depression, anxiety, and stress investigations and their relation to cardiovascular mortality. His overview shows that increased risk of death from cardiovascular diseases associates with depression and stress. P. G. Surtees et al. conducted a prospective study in the UK that based on the 8.5 years of observation [4]. This study provides that the

presence of major depression is associated with a 3.5-fold increase in mortality from coronary heart disease (CHD). W. Whang et al. demonstrate that women with depression have an increase of fatal CHD by 49% in 9 years of follow-up [5]. These studies demonstrate that depression treatment and stress control, as well as early diagnosis and prevention of symptoms of psychological distress and mental disorders, can increase life expectancy.

Nevertheless, depression is still often falsely associated with a lack of willpower and unwillingness to cope with the "bad mood". There is social stigmatization of this disease, and it is embarrassing to admit it for a person. As a result, people with depression often hide their condition, do not seek help in time, and aggravate the disease.

Online methods and social media provide an opportunity to privately detect the symptoms of depression in time. It would allow people to suggest measures for its prevention and treatment in the early stages. The report of the European branch of WHO (2016) paid special attention to the identification of signs of depression and the personalization of online methods of its prevention.

In this paper, we consider the problem of automatic detection of depression signs from textual messages of Russian social network Vkontakte users. We explored the ability of psycholinguistic and stylistic markers to predict depression from the text of messages. In Section 2, related works are reviewed, in Section 3 we present dataset of Vkontakte profiles, in Section 4 we describe our methods and feature engineering and in last sections, we present and discuss results of experiments.

2   Related Work

Instrumental possibilities of analyzing the behavior of users in social networks are actively developing. In particular, methods of computational linguistics are successfully used in analyzing the texts from social networks.

The computerized analysis method of texts LIWC (Linguistic Inquiry and Word Count) [6] allows assessing the extent to which the author of a text uses the words of psychologically significant categories. The method works on the basis of manually compiled dictionaries of words that fall into different categories: meaningful words (social, cognitive, positive/negative words, etc.), functional words (pronouns, articles, verb forms, etc.). LIWC is used for different languages, including Russian [7], but does not consider the specifics of the language, since it is simply a translation of dictionaries from English to Russian.

A. Yates et al. [8] used neural network model to reveal the risks of self-harm and depression based on posts from Reddit and Twitter and showed the high accuracy of this diagnostic method. The authors indicate that proposed methods can be used for large-scale studies of mental health as well as for clinical treatment.

Seabrook et al. [9] utilized the MoodPrism application to collect data about status updates and mental health of Facebook and Twitter users. It was found that the average proportion of words expressing positive and negative emotions, as well as their variability and instability of manifestation in the status of each user, can be used as a simple but sensitive measure for diagnosing depression in a social network. In addition, it was

found that usefulness of the proposed method may depend on the platform: for Facebook users these features predicted a greater severity of depression, and lower for Twitter.

M. Al-Mosaiwi et al. [10] examined the usage of absolute words (i.e., always, totally, entire) in text writings from various forums devoted to different disorders: depression, anxiety, suicidal ideation, posttraumatic stress disorder, eating disorder, etc. It was found that the number of absolute words in anxiety, depression, and suicidal ideation related forums was significantly greater than in forums from the control group.

Most of the related studies investigate the relationship between mental health and English-speaking social media texts. As an exception, Panicheva et al. [11] and Bogolyubova et al. [12] investigated the relationship between so-called dark triad (Machiavellianism, narcissism, and psychopathy) and Russian texts from Facebook. Using the results of the dark triad questionnaire and profile data of Facebook users the authors conducted a correlation analysis to reveal informative morphological, lexical, and sentiment features.

The study of detecting an early risk of depression based on the experimental task Clef/eRisk 2017 described in the article [13]. The main idea of the task was to classify Reddit users into two groups: the case of depression and non-risk case. The study evaluates the applicability of tf-idf, embeddings, and bigrams models with stylometric and morphological features using Clef/eRisk 2017 dataset and reports 63% of F1-score for depression class.

It should be noted that the use of computational linguistics for analyzing text messages of social networks is mainly limited to lexical approaches. The syntactic-semantic analysis and psycholinguistics markers of the text are still not well evaluated on depression detection task. In this paper, we applied psycholinguistic markers, dictionaries and n-grams models to detect depression in social media texts.

3 Dataset

We asked volunteers from Vkontakte to take part in our psychological research and complete Beck Depression Inventory questionnaire [14]. This questionnaire allows to calculate depression score on 0–63 scale. Before answering questions, users gave access to their public pages under privacy constraints via Vkontakte application. We automatically collected all available information from public personal profile pages using Vkontakte API for the users who completed questionnaire. Posts, comments, information about communities, friends etc. were collected from January 2017 to April 2019 for each user. Overall, information from 1020 profiles were assembled to compile our dataset. All of the personal information that can reveal the identity of persons were removed from data collection.

The scope of our interest were textual messages, namely posts, written in Russian. Therefore, we focused on text messages written by Vkontakte users on their personal profiles and mainly operate with these messages. It is important to note, that social media data contains significant amount of noise and text volume for each user considerably vary from person to person. Before performing on depression detection task, we

accurately cleaned the data. First, we applied constrains on required text volume and number of posts. Secondly, we analyzed scores from Beck Depression Inventory and divided our users into 2 groups: persons with score less then 11 were annotated as control group (users without depression signs); persons with score greater than 29 were annotated as depression group (users with depression signs). In this section, we describe these steps and provide statistics on the data. We refer to the data before any changings as *initial data*, to the data after cleaning as *cleaned data*, and to the data after depression risk grouping as *pre-classification data*.

The *initial data* contained information about 1020 persons who took Beck Depression inventory questionnaire. The distribution of the depression marker across users from *initial data* presented in Fig. 1.
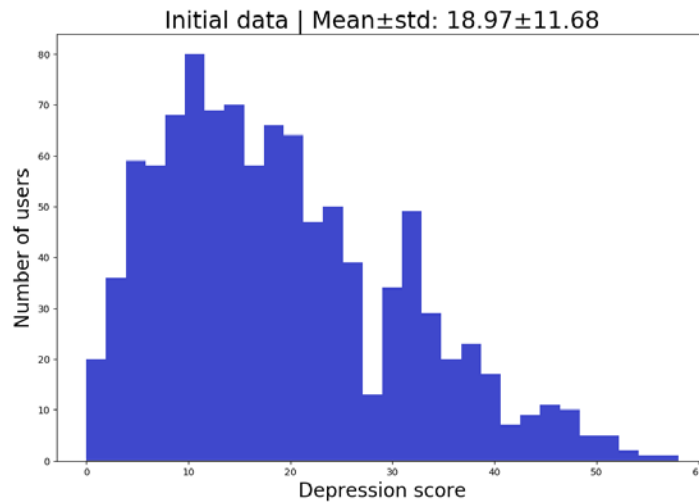

Fig. 1. Depression scores distribution in *initial data*

The mean age in the *initial data* is 25. The gender partition is unbalanced: 699 (68.53%) Females and 321 Males (31.47%). More statistics on the data provided in Table 1. It can be seen from the Table 1, that *initial data* is extremely noisy. Standard deviation values for post, sentence, and word counts are doubled in comparison with their mean values. It was also discovered that 155 of users from dataset did not provide any textual volume. The superficial analysis of the data revealed that data require adjustments and cleaning. As the next step, we performed several actions to adjust the data:

1. Removed all characters which are not alphabet or standard punctuation symbols from texts using regular expressions;
2. Removed all posts with more than 3000 characters or less than 2 words;
3. Removed all users with less than 10 posts or less than 1000 characters provided;
4. Set 100 as the maximum posts count limit for all users.

Applying these steps on the *initial data* yielded 531 user profiles which we annotate as *cleaned data*. We can note, that after data adjustments only 32872 users post left from initial 67.257 posts (see Table 1). We found that limitation of maximum post length is

strongly necessary because the manual observation of the data revealed that the most of long posts (more than 3000 characters) were usually not authored by users themselves.

After adjusting steps, the mean depression score slightly decreased, from what we can make the assumption, that persons with higher level of depression write less text than person without depression signs. The gender distribution become even more unbalanced with 397 females (74.76%) and 134 males (25.23%). The histogram of posts count demonstrated on Fig. 2.
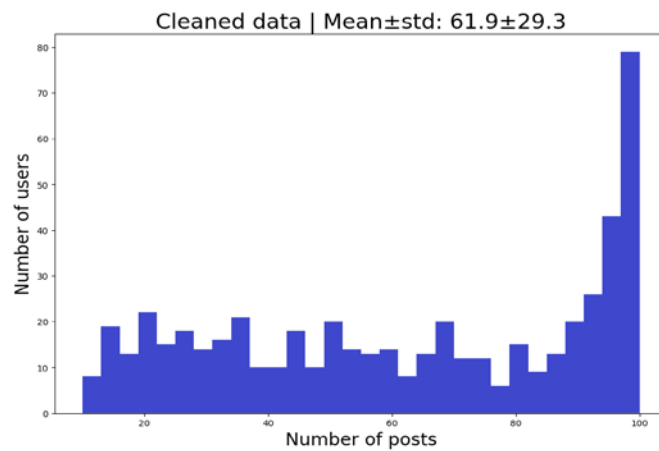

Fig. 12. Posts count distribution in *cleaned data*

After the data cleansing stage, we found this text volume much more suitable for applying natural language processing tools and performing any type of machine learning based evaluation. Anyhow, the depression scores provided by Beck Depression Inventory required some interpretation. We outlined 2 different ways how we can design our research. First one is the regression analysis using raw depression scores, which might be seen as the most appropriate and confident way. But in the other hand, this Russian-speaking social network data is novel, and currently there are no studies related to the depression detection task among Russian-speaking social networks. Most of the English-speaking social networks based depression tasks were designed as a binary classification problem: discover if person depressed or not. To make it possible to compare our results, we decided to perform the similar binary classification task on given data and compare our results with Clef/eRisk 2017 Shared Task [15].

Table 1. Dataset statistics on different data preparation stages. The numbers presented as mean value ± standard deviation

Observed data	Initial data	Cleaned data	Pre-classification data
Number of users	1020	531	248
Males	321 (31.47%)	134 (25.23%)	66 (26.61%)
Females	699 (68.53%)	397 (74.76%)	182 (73.39%)
Age	24.88±6.47	25.99±6.11	25.8±5.69
Depression score	18.97±11.68	17.99±11.04	17.4±15.28
Total number of posts	67257	32872	15238
Avg. posts count	65.93±103.85	61.9±29.3	61.44±29.65
Avg. words count	3114.67±8637.82	1438.01±1244.16	1441.56±1220.59
Avg. sentences count	189.96±492.78	148.98±101.69	148.63±102.81
Words per post	28.75±29.57	22.22±14.42	22.93±15.79
Words per sentence	9.61±4.43	8.98±2.76	9.08±2.72
Sentences per post	2.66±1.96	2.31±0.99	2.34±1.08

Table 2. Statistics between depression and control group

Group	Depression group	Control group
Number of users	92 (35.65%)	156(60.46%)
Males	19 (20.65%)	47 (30.12%)
Females	73 (79.34%)	109 (69.87%)
Age	25.67±6.43	25.87±5.21
Depression score	36.44±6.37	6.17±2.75
Total number of posts	5268	9970
Avg. posts count	57.26±30.13	63.91±29.07
Avg. words count	1328.15±1271.14	1508.44±1184.7
Avg. sentences count	138.61±113.08	154.54±95.75
Avg. Words per post	22.74±18.58	23.04±13.89
Avg. Words per sentence	8.88±2.71	9.19±2.72
Avg. Sentences per post	2.34±1.32	2.35±0.91

As the next step, we observed depression scores and discovered that we cannot simply divide our data by setting boarder value and annotating all users with depression score above this value as a risk group of depression and all user with the depression score bellow boarder value as a non-risk group. In order to form the *pre-classification data,* we annotated all persons with depression score less than 11 as non-risk group (control group). For a risk group we assembled the data of persons with depression scores above 29 (depression group). These values were discussed and proposed by the psychologist experts related to our study. The persons with depression score between these values were removed from observation.

Performing this step reduced the data population to the 248 users, where 156 were labeled as control group (without depression signs) and 92 users were labeled as belonging to the depression group. The general statistics on *pre-classification data* also

presented in Table 1. The statistics between groups on the *pre-classification data* presented in Table 2.

It can be observed from Table 2 that users from depression group tends to write lesser amount of text in the Vkontakte social media. The values of average posts count, average words count, average sentence count are less than in the control groups. The length of posts and sentences are greater in control group. The gender partition is even more biased towards female in depression group.

4   Features and Methods

Before forming the feature sets, all user posts were concatenated into the one text for every user in dataset. We retrieved four groups of features from texts: morphological, syntactic, sentiment and psycholinguistic. We applied MyStem [16] for tokenization, lemmatization, and part-of-speech tagging, and Udpipe [17] for syntax parsing. The sentiment features were calculated using Linis-Crowd sentiment dictionary [18].

Psycholinguistic markers are linguistic features of text that represent psychological characteristics of author and may signal about his psychological disorders. For example, people in stress more frequently use in text pronoun "we" [19]. Psycholinguistic markers are calculated on morphological and syntactic information and in a manner correspond to the writing style of the author. We use more than 30 markers and the most significant of them are the following:

— Mean number of words per sentence;
— Mean number of characters per word;
— (N punctuation characters) / (N words);
— Lexicon: (N unique words) / (N words);
— Average syntax tree depth;
— (N verbs) / (N adjectives);
— (N conjunctions + N prepositions) / (N sentences);
— (N infinitives) / (N verbs);
— (N singular first person past tense verbs) / (N verbs);
— (N first person verbs) / (N verbs);
— (N third person verbs) / (N verbs);
— (N first person pronouns) / (N pronouns);
— (N singular first person pronouns) / (N pronouns);
— (N plural first person pronouns) / (N pronouns).

These psycholinguistic markers were previously utilized for the task of predicting depression from essay in Russian. They are described in more details at [20]. We extend psycholinguistic markers set with postags ratio and following social network specific features: uppercase characters ratio, average number of Vkontakte links per post, number of exclamation marks, number of "sad" and "happy" smiles.

We also formed two n-grams sets: tf-idf matrix computed on the unigrams and tf-idf matrix computed on the both unigrams and bigrams combined. N-grams that appeared less than in 1% of texts were removed from the feature sets. User's lexicon formed while tf-idf set preparation were extremely poor with 5742 unique tokens for unigrams and 10909 unique tokens for both unigrams and bigrams combined. We relate this fact

to the specific of social network language. The writings contain a lot of slang and words with wrong spelling.

Another feature set were retrieved by utilizing dictionaries which was used for the task of detection verbal aggression in social media writings [21]. It is containing following dictionaries: negative emotional words, lexis of suffering, positive emotional words, absolute and intensifying terms, motivation and stressful words, invectives, etc. To calculate features, for every user we calculate the occurrences of words from different dictionaries in user's writings and divide this number on total user's words count.

As it was mentioned before, we designed depression detection task as binary classification. We evaluate 4 different sets of features: psycholinguistic markers (PM), unigrams (UG), bigrams (BG), and dictionaries(D).

5 Results of Experiments

To perform on the task, we utilized scikit-learn machine learning library [22]. Random forest and support-vector machines (SVM) models were used to perform evaluation on the data. All of the feature's sets were normalized and scaled. Hyperparameters of the classification algorithms were tuned by grid-search runs.

Table 3. Classification report

Dummy classifier					
Set	*Precision*	*Recall*	*F1*	*ROC AUC*	*F1-w*
-	45.23±2.38	30.43±8.13	36.0±8.48	48.29±3.76	50.48±6.17
Random Forest					
Set	*Precision*	*Recall*	*F1*	*ROC AUC*	*F1-w*
PM	59.80±6.21	59.80±6.21	54.47±3.66	70.91±6.81	67.98±3.03
UG	51.68±9.89	57.17±3.70	53.84±6.35	64.59±3.79	63.03±8.29
BG	49.64±6.67	58.47±6.06	53.12±3.16	63.18±2.68	61.65±5.95
D	46.21±5.52	56.30±7.20	50.66±5.80	58.07±6.33	59.90±4.88
PM-r	**62.60**±7.77	53.26±7.88	56.59±2.20	74.89±4.05	69.16±2.60
SVM					
Set	*Precision*	*Recall*	*F1*	*ROC AUC*	*F1-w*
PM	55.43±1.99	72.82±1.88	62.92±1.51	71.12±4.46	68.66±1.72
UG	45.63±7.94	83.69±13.53	57.57±3.41	67.72±3.61	49.79±13.77
BG	44.38±6.07	**85.86**±11.24	57.60±2.76	66.88±2.64	47.72±14.90
D	55.68±9.49	55.43±8.34	55.53±8.85	63.57±7.85	66.94±6.89
PM-r	58.40±2.99	77.17±1.88	**66.40**±1.33	**75.11**±3.24	**71.42**±2.21

Since the depression detection task is previously untested on the Russian-speaking social media data, we also demonstrate the accuracy yielded by random based dummy classifier. The metrics for evaluation is weighted mean F1-score of both control and depression group (F1-w) and ROC AUC score. To make it possible to compare our results with Clef/eRisk 2017 Shared Task results, we also demonstrate precision, recall

and F1-score for depression class only. The evaluation metrics were calculated by averaging 5 runs of 4-folds cross-validation. The classification results presented in Table 3.

The evaluation revealed that Psycholinguistic markers performed well on the data. We initially assumed that some of the psycholinguistic markers could work poorly on the data because users usually write very short texts and the volume of concatenated posts cannot be compared to a logically connected text of the same size. This constrains are important for the specific of some psycholinguistic markers. We analyzed feature importance from several Random Forest runs in order to reduce the size of the PM feature vector which can possibly improve classification performance. The reduced version of PM (PM-r) was included in classification report.

The best result on the data yielded by SVM+PM-r model with 75.11% ROC AUC score, 71.42% weighted F1-score and 66.40% F1-score on depression class. The same feature set with the Random Forest algorithm also achieved decent results with 74.89% ROC AUC score and highest precision (62.60%) in our experiments.

The dictionaries based set demonstrated poor results in comparison with other sets. In other hand, considering the general complexity of the depression detection task these dictionaries demonstrated some positive results. These dictionaries should be redesigned and filtered which can make them useful as additional features for PM set.

The surprising result in our experiments is that n-gram and tf-idf based features did not perform well on the data. As it was mentioned before, we relate this fact to the great amount of slang, wrong spelling and another noise in social media language. We should focus this problem by applying term clustering. For example, we can use words embeddings as it was implemented in this work [13].

It should be noted, that we can compare our results with the results of Clef/eRisk 2017 Shared Task evaluation only with some restrictions. First, language of Clef/eRisk 2017 was English, while our data is in Russian. Secondly, the number of data samples and class ratio is different. Finally, depression class in Clef/eRisk 2017 Shared Task was assembled by manual expert examination of profiles from subforum related to the depression disorder. In our study, we operate only with the Beck Depression Inventory scores.

Despites this facts, best F1-score reported in Clef/eRisk 2017 overview [23] was 64% achieved by the model that utilized tf-idf based features on the data with LIWC and dictionary features. In our experiments tf-idf based features demonstrated 57.60% of F1-score with *SVM+BG* model. It is important to mention, that current state-of-art result on Clef/eRisk 2017 data is 73% of F1-score [24]. The best depression detection performance on our Vkontakte data is 66% of F1-score achieved by filtered version of psycholinguistic markers.

6   Conclusion

In the study we performed depression detection task among 1020 users of Russian-speaking social network Vkontakte based on their text messages. By analyzing Beck Depression Inventory scores and processing the initial data we formed the sample of 248 users' posts collections with binary depression/control group labeling. We formed tf-idf and dictionary based feature sets and retrieved novel psycholinguistic features

from users' writings. The experiments were performed using SVM and Random Forest classifiers and results were compared with Clef/eRisk 2017 Shared Task evaluation. The best result in our experiments is 66.40% of F1-score (75.11% of ROC AUC score) achieved by model that based on filtered psycholinguistic markers.

It was discovered that psycholinguistic markers performed well on the data and can be effectively utilized for the depression detection task. We found that Vkontakte textual data is extremely noisy which is resulted in the relatively poor classification results achieved by tf-idf based models. We assume that term clustering methods could improve performance of n-grams models. It is also clear, that dictionaries that we used for feature set should be redesigned and filtered.

Thus, the analysis of depression linguistic markers in social network posts is a promising area that can possibly make the prevention and treatment of depression more accessible to a large number of users. In the future work we planning to examine neural network models for the depression detection task and evaluate regression analysis on the data using Beck Depression Inventory scores.

Acknowledgments

References

1. Turecki, G. and Brent, D.A.: Suicide and suicidal behaviour. The Lancet **387** (10024), 1227–1239 (2016).
2. World Health Organization. https://www.who.int/mental_health/prevention/ suicide/suicideprevent/en/, last accessed 2019/08/19
3. Belialov, F.I.: Depression, anxiety, stress, and mortality. Terapevticheskii arkhiv **88** (12), 116–119 (2016).
4. Surtees, P.G., Wainwright, N.W., Luben, R.N., Wareham, N.J., Bingham, S.A., and Khaw, K.T.: Depression and ischemic heart disease mortality: evidence from the EPIC-Norfolk United Kingdom prospective cohort study. American Journal of Psychiatry **165** (4), 515–523 (2008).
5. Whang, W., Kubzansky, L.D., Kawachi, I., Rexrode, K.M., Kroenke, C.H., Glynn, R.J., and Albert, C.M.: Depression and risk of sudden cardiac death and coronary heart disease in women: results from the Nurses' Health Study. Journal of the American College of Cardiology **53** (11), 950–958 (2009).
6. Tausczik, Y.R. and Pennebaker, J.W.: The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology **29** (1), 24–54 (2010).
7. Kailer, A. and Chung, C.K.: The Russian LIWC2007 dictionary. Austin, TX: LIWC.net (2011).
8. Yates, A., Cohan, A., and Goharian, N.: Depression and self-harm risk assessment in online forums. arXiv preprint arXiv:1709.01848 (2017).
9. Seabrook, E.M., Kern, M.L., Fulcher, B.D., and Rickard, N.S.: Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. Journal of Medical Internet Research **20** (5), e168 (2018).

10. Al-Mosaiwi, M. and Johnstone, T.: In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. Clinical Psychological Science **6** (4), 529–542 (2018).

11. Panicheva, P., Ledovaya, Y., and Bogolyubova, O.: Lexical, morphological and semantic correlates of the dark triad personality traits in russian facebook texts. In 2016 IEEE Artificial Intelligence and Natural Language Conference (AINL) (pp. 1–8). IEEE (2016, November).

12. Bogolyubova, O., Panicheva, P., Tikhonov, R., Ivanov, V., and Ledovaya, Y.: Dark personalities on Facebook: Harmful online behaviors and. Computers in Human Behavior **78**, 151e159 (2018).

13. Stankevich, M., Isakov, V., Devyatkin, D., and Smirnov, I.: Feature Engineering for Depression Detection in Social Media. In ICPRAM, 426–431 (2018).

14. Beck, A.T., Steer, R.A., and Brown, G.K. Beck depression inventory-II. San Antonio **78** (2), 490–498 (1996).

15. Losada, D.E. and Crestani, F.: A test collection for research on depression and language use. In International Conference of the Cross-Language Evaluation Forum for European Languages, 28–39. Springer, Cham (2016, September).

16. MyStem Homepage, https://tech.yandex.ru/mystem, last accessed 2019/08/19

17. Straka, M. and Straková, J. Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 88–99 (2017, August).

18. Koltsova, O.Y., Alexeeva, S., and Kolcov, S.: An opinion word lexicon and a training dataset for russian sentiment analysis of social media. Computational Linguistics and Intellectual Technologies: Materials of DIALOGUE 2016 (Moscow), 277–287 (2016).

19. Pennebaker, J.W. The secret life of pronouns. New Scientist **211** (2828), 42–45 (2011).

20. Stankevich, M., Smirnov, I., Kuznetsova, Y., Kiselnikova, N., and Enikolopov, S.: Predicting Depression from Essays in Russian. Computational Linguistics and Intellectual Technologies, DIALOGUE **18**, 637–647 (2019).

21. Devyatkin, D., Kuznetsova, Y., Chudova, N., and Shvets A.: Intellectual analysis of the manifestations of verbal aggressiveness in the texts of network communities [Intellektuanyj analiz proyavlenij verbalnoj agressivnosti v tekstah setevyh soobshchestv]. Artificial Intelligence and Decision Making, (2), 27–41 (2014).

22. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and Vanderplas, J.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, (12), 2825–2830 (2011).

23. Losada, D.E., Crestani, F., and Parapar, J.: CLEF 2017 eRisk Overview: Early Risk Prediction on the Internet: Experimental Foundations. In CLEF (Working Notes) (2017).

24. Trotzek, M., Koitka, S., and Friedrich, C.M.: Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. IEEE Transactions on Knowledge and Data Engineering (2018).