Вестник Томского государственного университета Философия. Социология. Политология. 2022. № 65. С. 252–263.
Tomsk State University Journal of Philosophy, Sociology and Political Science. 2022. 65. pp. 252–263.

Original article
УДК 32.019.52, 004.912
doi: 10.17223/1998863X/65/21

PREDICTING POLITICAL VIEWS IN SOCIAL MEDIA: VKONTAKTE AS A CASE STUDY

Anna V. Glazkova[1], Zinaida N. Sokova[2], Valery M. Kruzhinov[3]

[1, 2, 3] *University of Tyumen, Tyumen, Russian Federation*

[1] *a.v.glazkova@utmn.ru*

[2] *z.n.sokova@utmn.ru*

[3] *v.m.kruzhinov@utmn.ru*

Abstract. The authors hypothesize that textual information posted on personal pages on social media reflects the political views of users to some extent. Therefore, this textual information can be used to predict political views on social media. The authors conduct experiments on textual data from user pages and test two machine learning methods to classify pages that declare different political preferences. To undertake a study, the authors collected anonymous open textual data of users of the VKontakte social network (the number of pages is 10 123). Data collection was carried out using the VKontakte Application Programming Interface (VK API). As a result of the analysis of the collected data, the authors discovered two types of textual information. The first is a text filled by the user by selecting one of several possible values (binary or categorical variables). The field "Political Views" is one of these text fields, it provides nine options for selection. The second type of text information includes information entered by the user in an arbitrary form (interests, activities, etc.). The authors trained and tested two machine learning models to predict users' political views based on the remaining text information from their pages: a) linear support vector classifier using text representations from the bag-of-words model; b) neural network using Multilingual BERT text embeddings. The results show that the models sufficiently successfully perform binary classification of users who have polar political views (for example, communists – libertarians, communists – ultra-conservatives). Nevertheless, the results for the groups of users that have close political views are significantly lower. In addition, the authors investigated the assumption that users often indicate "indifferent" political views as "moderate". The authors classified the groups of users who declare indifferent or moderate views (the two largest categories in our dataset) and users who indicated other political preferences. The results demonstrate a sufficiently high performance for the classification of custom pages based on these two political views.
Keywords: social media, opinion mining; VKontakte; political polarization; machine learning

Acknowledgments: The reported study was funded by RFBR and EISR, project number 20-011-32031.

For citation: Glazkova, A.V., Sokova, Z.N. & Kruzhinov, V.M. (2022) Predicting political views in social media: VKontakte as a case study. *Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sotsiologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science*. 65. pp. 252–263. doi: 10.17223/1998863X/65/21

Научная статья

ПРОГНОЗИРОВАНИЕ ПОЛИТИЧЕСКИХ ПРЕДПОЧТЕНИЙ В СОЦИАЛЬНЫХ СЕТЯХ (НА МАТЕРИАЛЕ ВКОНТАКТЕ)

Анна Валерьевна Глазкова[1], Зинаида Николаевна Сокова[2], Валерий Михайлович Кружинов[3]

[1, 2, 3] *Тюменский государственный университет, Тюмень, Россия*

[1] *a.v.glazkova@utmn.ru*

[2] *z.n.sokova@utmn.ru*

[3] *v.m.kruzhinov@utmn.ru*

Аннотация. Исследуется возможность прогнозирования политических предпочтений пользователей социальных сетей с помощью методов машинного обучения. Демонстрируется, что данные методы успешно справляются с бинарной классификацией текстовых данных со страниц пользователей, имеющих полярные политические предпочтения. Однако результаты, полученные для страниц пользователей с близкими политическими предпочтениями, значительно ниже.
Ключевые слова: социальные сети, анализ мнений, ВКонтакте, политическая поляризация, машинное обучение

Благодарности: Исследование выполнено при поддержке РФФИ и ЭИСИ (проект № 20-011-32031).

Для цитирования: Глазкова А.В., Сокова З.Н., Кружинов В.М. Прогнозирование политических предпочтений в социальных сетях (на материале ВКонтакте) // Вестник Томского государственного университета. Философия. Социология. Политология. 2022. № 65. С. 252–263. doi: 10.17223/1998863X/65/21

Social media contains a wealth of useful information for social and political research. Social network users share important aspects of their personal lives, exchange their views, discuss social and political phenomena. As a result, social networks are now a powerful lever of influence on social and political processes, as well as a tool of reflecting the views and the moods of society.

Social network users share opinions on their profiles and express their views on the pages of communities or other users (comments, “likes”, etc.). Information posted by users can be presented in various formats: texts, images, video files, emoji, and so on. Based on the content of the information that users post in the public domain on a social network, important conclusions can be drawn about their belonging to certain social groups, about their views, preferences, and interests. This information can be used for various purposes: as a marketing tool, as a tool for political influence, as a material for sociological research, and many others.

In this work, we study the ability to predict the political views of social media users based on textual information from profiles. We put forward a hypothesis that commitment to certain political views is reflected on user pages: in users’ posts, descriptions of interests, and personal information. We analyze different types of text information on user pages and study how users’ political preferences are expressed in their profiles. This paper focuses on two objectives: i) a binary classification of social network pages belonging to users with different political views; ii) predicting political relation of social media users as politically active and politically inactive based on textual information posted on their pages. We conduct a set of experiments on a self-collected VKontakte dataset of user textual data with two machine learning classifiers.

In the modern world, there is a rapid increase in the role of information factors in the political life of society. Social media materials have become a powerful channel for political agitation, revealing the views and interests of the population, and analyzing public opinion [1–3]. The constant growth in the number of social media users belonging to a wide audience of citizens dramatically increases the possibilities of social media as a tool for studying and tracking political preferences. Various studies analyze the online popularity of political leaders [4], pre-election preferences [5–7], forecasting the results of elections and referendums [8, 9]. The authors of these works emphasize that, although Internet users do not represent the entire population of the country, the analysis of social networks makes it possible to predict the results of elections, and the results also demonstrate a significant correlation between social networks and the results of traditional mass polls. In [10], it is indicated that the predictive power of analysis on social networks increases as the number of citizens expressing their opinions on the Internet increases. A number of authors use data on social media to analyze political preferences of users (on Facebook [11–15], Twitter [16–19], Weibo [20, 21], VKontakte [22; 23]). The authors of [24] attempt to predict the leading political ideology (left, right, or center) or bias of news articles.

In this study, we work with textual data from personal pages of social media users. Textual data include personal information, texts of posts, captions to videos, photos, and so on. We used data from the VKontakte social network. This social network is popular in Eastern Europe, especially in the post-Soviet countries. VKontakte allows users to send messages, create their pages and communities, share images, tags, audio and video recordings, and play browser games [25]. VKontakte users can represent their political views on personal profiles. The user selects political views from among nine proposed types including communist, socialist, moderate, liberal, conservative, monarchist, ultraconservative, indifferent, and libertarian. The political typification of VKontakte is reflected in Figure 1. In outline, there are two main systems for classifying political ideologies. The first is related to the proposed model of the desired society and the second is based on the attitude to social progress and the way of its implementation. The first type divides political ideologies into the right, center, and left (pro-socialist). The second type considers radicals who advocate permanent deep revolutionary transformations and conservatives who strive to preserve the established political order and agree only to the smallest changes in the status quo. There are moderately oriented political forces between them that prefer the way of gradual reforms.

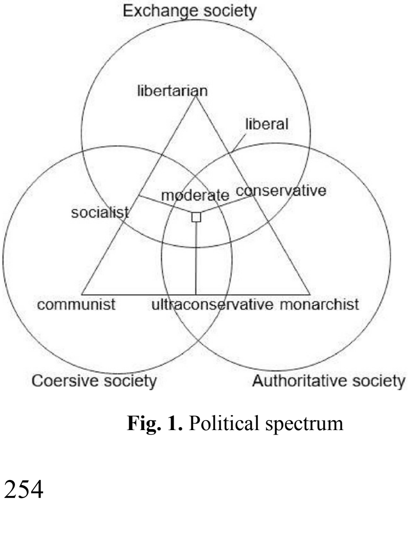


Fig. 1. Political spectrum

For this study, we have collected a dataset of anonymized 10 123 open-access profiles of users who represented their political views on their pages. The data were collected using VKontakte API[1]. The main characteristics of the dataset are shown in Table 1. The average number of words and symbols shows the averaged count of words and symbols respectively in text fields on the user page (for all fields). The standard deviation presents a measure of the amount of variation (dispersion) of a set of values.

Table 1. **Data statistics**

Characteristic	Value	STD
Avg number of words	135.21	3359.25
Avg number of symbols	1790.16	4575.3
% of empty fields	35.49	16.6

The data are presented in .csv format where each row contains the following attributes: political views, gender, age category, education, occupation, relation, military, people main, life main, activities, interests, music, movies, TV, books, about, religion, inspiration, and posts. The detailed description of these fields is shown in Table 2. The last column contains lists of possible values for each attribute and the distribution of these values. As can be seen from the table, our data confirm the fact noted in previous studies (for example, [26]) that men more often show their political views on social networks than women.

Table 2. **Data description**

Column name	Type	Description	Possible values and distributions
Political	Categorical	Political views	"communist", "socialist", "moderate", "liberal", "conservative", "monarchist", "ultraconservative", "indifferent", "libertarian"
Gender	Binary	User's gender	"male", "female", "not specified"
Age category	Categorical	User's age category	"<20", "20–29", "30–39", "40–49", "50–59", "60–69", ">=70", "not specified"
Occupation	Categorical	Current user's occupation	"work", "school", "university", "not specified"
Relation	Categorical	Relationship status	"single", "in a relationship", "engaged", "married", "it's complicated", "actively searching", "in love", "not specified"
Military	Binary	Information about the completion of military service	"specified", "not specified"
People main	Categorical	Important in others	"power and wealth", "intelligence and creativity", "kindness and honesty", "beauty and health", "courage and perseverance", "humor and love of life", "not specified"
Life main	Categorical	Personal priority	"family and children", "career and money", "entertainment and recreation", "science and research", "improving the world", "self development", "beauty and art", "fame and influence", "not specified"
Activities	Free-form text	Personal interests and favorites	–
Interests			
Music			
Movies			
TV			
Books			
Inspired by			
Religion	Free-form text	World view	–
Posts	List of strings	List of user's posts	–

[1] https://vk.com/dev/methods

The distribution of political views in the dataset is shown in Figure 2a. As we can see from the figure, the categories are unbalanced. Most users define their political views as moderate or indifferent. Some categories (for example, libertarians, ultraconservatives) are represented by less than 5% of users. A similar distribution of political views is observed in other sociological studies related to the analysis of VKontakte, such as [27]. Figure 2b illustrates the distribution of categories combined in terms of the proximity of political ideologies. We labeled conservatives, ultraconservatives, and monarchists as “conservative”, communists and socialists as “socialist”, and liberals and libertarians as “liberal”. Thus, Figure 2b shows five aggregated categories the largest of which is 29.22% (moderate) and the smallest is 9.2% (socialist).

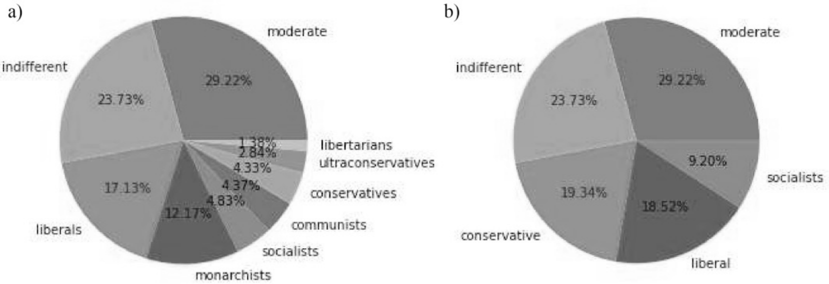


Fig. 2. Distribution of political views: a) original categories; b) combined categories

As can be seen from previous studies [28, 29], most people report moderate political views. This is since many people label themselves as not involved or not feeling strongly about political views. Based on this information, we attempted to combine categories “indifferent” and “moderate” and labeled them as politically inactive users. At the same time, other users were marked as politically active.

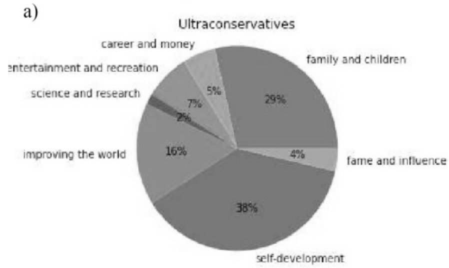
We can see two types of textual information in the dataset. The first type includes fields filled by the user in free-form (“Activities”, “Interests”, “Inspired by”, and so on). The second type represents categorical or binary attributes, such as gender, age, occupation, relation, military, people main, and life main. These attributes are represented as text fields in a profile, but the user selects the text from the given options. For example, Figure 3 shows the life main distribution by category for three radical political ideologies. To calculate the age category, we used the user’s year of birth. Users who did not provide a year of birth were placed in a separate category.

We evaluated two machine learning approaches:

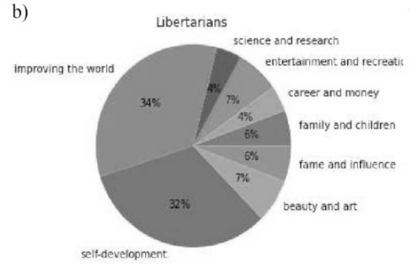
••a classifier based on Linear Support Vector Machines (Linear SVC). The vector of features was a concatenation of categorical features encoded using a one-hot encoding scheme and a bag-of-words (BoW) representation of texts. One-hot encoding is a data representation in a form of new (binary) columns, indicating the presence of each possible value from the original data (e.g. 1 → 100, 2 → 010, 3 → 001 for values from the range [1,3]). The BoW model represents text as a set of words disregarding grammar and even word order but keeping multiplicity (e.g. “it is the best of the best” → [“it” – 1, “is” – 1, “the” – 2, “best” – 2, “of” – 1]). The Linear SVC model was implemented using Scikit-Learn [30] with a tolerance for stopping criteria equal to 1e-2. The number of words in the BoW model was

10000. Table 2 present the most common words for the users' groups of polar political ideologies such as ultraconservatives, libertarians, and communists.

a)



b)



c)

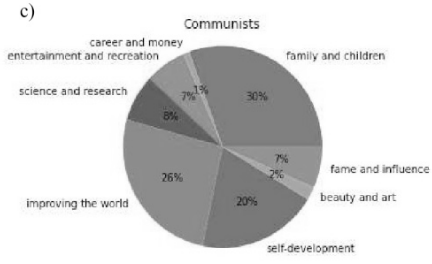


Fig. 3. Life main distribution by category:
a) ultraconservatives, b) libertarians,
c) communists

•a feedforward neural network trained on feature vectors represented concatenation of one-hot encoded categorical features and text embeddings obtained with Multilingual BERT (mBERT) [31, 32]. BERT is a language representation model presented by Google, which stands for Bidirectional Encoder Representations from Transformers. BERT-based models show state-of-the-art results in many natural language processing tasks. The mBERT-based model implementation used Transformers [33] and Pytorch [34] Python libraries. The model architecture is showed in Figure 4.

Table 3. **Top-20 words by category**

No.	Ultraconservatives	Libertarians	Communists
1	music	life	music
2	life	music	good
3	sport	book	book
4	belief	secular humanism	orthodoxy
5	orthodoxy	humanity	movie
6	friend	game	game
7	movie	movie	life
8	girl	travel	atheist
9	book	friend	Soviet
10	history	beloved	sport
11	football	Russian	football
12	miracle	love	friend
13	Russian	psychology	war
14	interest	foreign language	history
15	orthodox	nature	learning
16	Russia	mind	Russian
17	travel	programming	classical
18	heaven	design	civil
19	man	money	family
20	green	pastafarianism	secular

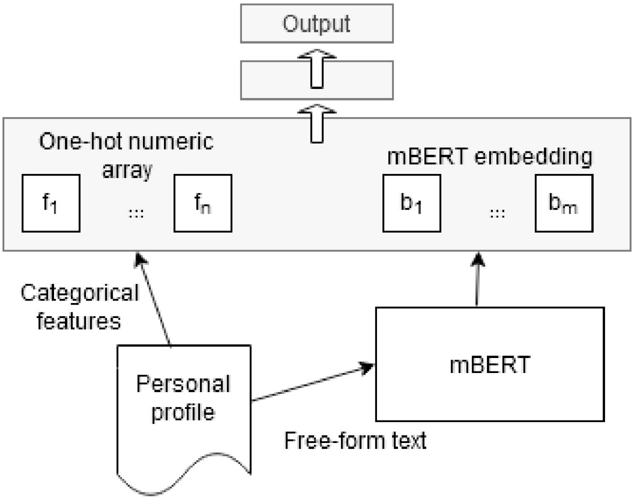


Fig. 4. The architecture of the mBERT-based network

For model evaluation, we used the weighted F1-score that uses averaged weighted precision and recall metrics and can be applied for imbalances classes. The standard F1-score is calculated as follows:

F1-score = 2 * (*Precision* * *Recall*) / (*Precision* + *Recall),*

Precision = *TP* / (*TP* + *FP*),

Recall = *TP* / (*TP* + *FN*),

where *TP* indicates true positive results, when the model *correctly* predicts the *positive* class; *FP* indicates false positive results, when the model *incorrectly* predicts the *positive* class; *FN* indicates false negative results, when the model *incorrectly* predicts the *negative* class.

To pre-process texts, we removed punctuation, merged all free-form text fields from the user's profile into one text, and further lowercased it. For the Linear SVC model, we additionally lemmatized the data using Pymorphy2 [35] and applied a random oversampling technique [36] to reduce the influence of the class imbalance problem.

The results in Table 4 show the performance of our classifiers for both the original and the combined categories. As expected, the highest results were obtained with the state-of-the-art neural architecture BERT.

Table 4. **Results, weighted F1-score, %**

Categories	Linear SVC	mBERT-based classifier
Communists – libertarian	67.02	67.28
Libertarian – ultraconservative	80.02	82.24
Ultraconservative – communist	64.54	65.21
Communist – socialist	33.12	36.46
Liberal – libertarian	23.89	29.5
Socialist (combined) – liberal (combined)	58.67	62.45
Liberal (combined) – conservative (combined)	80.34	81.33
Conservative (combined) – socialist (combined)	74.68	76.01
Politically active – politically inactive (indifferent	64.4	66.57
Politically active – politically inactive (ind. + mod.)	62	65.23

We observe significant differences in performance between the binary tasks. Both models perform much better when they classify polar political views. For example, the mBERT-based model achieved 67.28% (communists vs. libertarians), 82.24% (libertarians vs. ultraconservatives), and 65.21% (ultraconservatives vs. communists). The classification of users with closer political views has significantly worse results, such as 36.46% of F1-score for communists vs. socialists and only 29.5% for liberals vs. libertarians. This observation confirms that textual data from social network pages contain meaningful information about the user's political preferences.

Then, we evaluated the impact of combining small categories into enlarged categories of socialists, liberals, and conservatives. In general, the results are comparable to the classification of the polar original categories. However, both models performed worse at separating the "socialist" and "liberal" classes.

Finally, we attempted to classify users into politically active and politically inactive. In the first case, we labeled "indifferent" users as politically inactive and the others as politically active. In the second case, we combined categories "indifferent" and "moderate" and labeled them as politically inactive users. We can see that the second approach showed the best quality on the test set.

We have explored the task of predicting political views in social media. We created a dataset for this task, which contains the main textual information from VKontakte's pages. We identified two types of user's textual information, such as categorical features and free-form textual fields. Further, we applied two machine learning approaches to the task of binary classification of user pages by their political preferences. The experimental results showed that our classifiers more successfully deal with the classification of profiles with polar political views. We also hypothesized and experimentally confirmed that combining "indifferent" and "moderate" users into one class can improve the quality of politically active users' detection.

Further research should focus on adding new types of features to our models, such as the user's communities, reposts, "likes", and other types of information and activities in social networks. Moreover, we plan to evaluate different text representation models for predicting political views in social networks.

References

1. Maulana, I. (2020) Social Media as Public Political Instrument. In: Adria, M. (ed.) *Using New Media for Citizen Engagement and Participation*. Information Science Reference. pp. 181–197.

2. McGregor, S.C. & Mourão, R.R. (2016) Talking politics on Twitter: Gender, elections, and social networks. *Social media + society*. 2(3). p. 2056305116664218.

3. Ryabchenko, N.A., Malysheva, O.P. & Gnedash, A.A. (2019) Presidential Campaign in Post-Truth Era: Innovative Digital Technologies of Political Content Management in Social Networks Politics. *Polis. Politicheskie issledovaniya – Polis. Political Studies*. 2(2). pp. 92–106. (In Russian). DOI: 10.17976/jpps/2019.02.07

4. Ceron, A. et al. (2014) Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*. 16(2). pp. 340–380.

5. Mengü, S. Ç. et al. (2015) Political preferences of generation Y university student with regards to governance and social media: A study on March 2014 local elections. *Procedia-Social and Behavioral Sciences*. 174. pp. 791–797. DOI: 10.1016/j.sbspro.2015.01.616

6. Oliveira, D., Bermejo, P. & dos Santos, P.A. (2017) Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls. *Journal of Information Technology & Politics*. 14(1). pp. 34–45. DOI: 10.1080/19331681.2016.1214094

7. Volkova, S., Coppersmith, G. & Van Durme, B. (2014) Inferring user political preferences from streaming communications. *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics*. pp. 186–196.

8. Gayo-Avello, D. (2013) A meta-analysis of state-of-the-art electoral prediction from Twitter data. *Social Science Computer Review*. 31(6). pp. 649–679.

9. Huff, C. & Tingley, D. (2015) “Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*. 2(3). p. 2053168015604648.

10. Lacy, S. et al. (2015) Issues and best practices in content analysis. *Journalism & Mass Communication Quarterly*. 4. pp. 791–811. DOI: 10.1177/1077699015607338

11. Chang, C.C., Chiu, S. I. & Hsu, K. W. (2017) Predicting political affiliation of posts on Facebook. *Proc. of the 11th International Conference on Ubiquitous Information Management and Communication*. pp. 1–8.

12. Chiu, S. I. & Hsu, K.W. (2018) Predicting political tendency of posts on facebook. *Proc. of the 2018 7th International Conference on Software and Computer Applications*. pp. 110–114.

13. David, E. et al. (2016) Utilizing Facebook pages of the political parties to automatically predict the political orientation of Facebook users. *Online Information Review*. 40(5). pp. 610–623. DOI: 10.1108/OIR-09-2015-0308

14. Feezell, J.T. & Ortiz, B. (2019) ‘I saw it on Facebook’: an experimental analysis of political learning through social media. *Information, Communication & Society*. pp. 1–20. DOI: 10.1080/1369118X.2019.1697340

15. Vepsäläinen, T., Li, H. & Suomi, R. (2017) Facebook likes and public opinion: Predicting the 2015 Finnish parliamentary elections. *Government Information Quarterly*. 34(3). pp. 524–532. DOI: 10.1016/j.giq.2017.05.004

16. Hernandez-Suarez, A. et al. (2017) Predicting political mood tendencies based on Twitter data. *5th Int. Workshop on Biometrics and Forensics*. pp. 1–6. DOI: 10.1109/IWBF.2017.7935106

17. Makazhanov, A., Rafiei, D. & Waqar, M. (2014) Predicting political preference of Twitter users. *Social Network Analysis and Mining*. 4(1). p. 193. DOI: 10.1007/s13278-014-0193-5

18. Preoţiuc-Pietro, D. et al. (2017) Beyond binary labels: political ideology prediction of twitter users. *Proc. of the 55th Annual Meeting of the Association for Computational Linguistics*. pp. 728–740. DOI: 10.18653/v1/P17-1068

19. Stefanov, P. et al. (2020) Predicting the topical stance and political leaning of media using tweets. *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*. pp. 527–537. DOI: 10.18653/v1/2020.acl-main.50

20. Bolsover, G. & Howard, P. (2019) Chinese computational propaganda: automation, algorithms and the manipulation of information about Chinese politics on Twitter and Weibo. *Information, Communication & Society*. 22(14). pp. 2063–2080.

21. Denemark, D. & Chubb, A. (2016) Citizen attitudes towards China's maritime territorial disputes: traditional media and Internet usage as distinctive conduits of political views in China. *Information, Communication & Society*. 19(1). pp. 59–79.

22. Byzov, L.G. et al. (2020) Ideal'nyy politik dlya sotsial'noy seti: podkhod k analizu ideologicheskikh predpochteniy pol'zovateley [An ideal politician for a social network: an approach to the analysis of users' ideological preferences]. *Problemy upravleniya*. 4. pp. 15–26.

23. Kozitsin, I.V. et al. (2020) Modeling Political Preferences of Russian Users Exemplified by the Social Network Vkontakte. *Mathematical Models and Computer Simulations*. 12. pp. 185–194. DOI: 10.1134/S2070048220020088

24. Baly, R. et al. (2020) We Can Detect Your Bias: Predicting the Political Ideology of News Articles. *Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing*. pp. 4982–4991. DOI: 10.48550/arXiv.2010.05338

25. Kinash, N. et al. (2015) Analysis of large-scale networks using high performance technology (Vkontakte case study). *Creativity in Intelligent, Technologies and Data Science*. 535. pp. 531–541.

26. Ledenev, D.E. (2021) Raskrytie politicheskikh predpochteniy v profile “VKontakte”: gendernyy aspect [Disclosure of political preferences in the VKontakte profile: the gender aspect]. *Psikhologiya cheloveka i obshchestva*. 2(31). pp. 23–41.

27. Zabokritskaya, L.D., Khlebnikov, N.A., Oreshkina, T.A. & Komotskiy, E.I. (2020) Vozmozhnosti izucheniya tsennostey molodezhi cherez profil' sotsial'noy seti “VKontakte” [Possibilities of studying the values of young people through the “VKontakte”profile]. *Monitoring obshchestvennogo mneniya: ekonomicheskie i sotsial'nye peremeny*. 2(156). pp. 148–167.

28. Jackson, J. (2019) Happy partisans and extreme political views: The impact of national versus local representation on well-being. *European Journal of Political Economy*. 58. pp. 192–202. DOI: 10.1016/j.ejpoleco.2018.12.002

29. Oraka, E. et al. (2019) A cross-sectional examination of US gun ownership and support for gun control measures: sociodemographic, geographic, and political associations explored. *Preventive Medicine*. 123. pp. 179–184.

30. Pedregosa, F. et al. (2011) Scikit-learn: Machine learning in Python. *The Journal of machine Learning Research*. 12. pp. 2825–2830.

31. Devlin, J. et al. (2018) *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805.

32. Glazkova, A. (2021) A Comparison of Text Representation Methods for Predicting Political Views of Social Media Users. *Proc. of ITIDMS-2021*.

33. Wolf, T. et al. (2019) *HuggingFace's Transformers: State-of-the-art Natural Language Processing*. arXiv: 1910.03771.

34. Paszke, A. et al. (2019) Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*. pp. 8026–8037. DOI: 10.48550/arXiv.1912.01703

35. Korobov, M. (2015). Morphological Analyzer and Generator for Russian and Ukrainian Languages. In: Khachay, M., Konstantinova, N., Panchenko, A., Ignatov, D. & Labunets, V. (eds) *Analysis of Images, Social Networks and Texts*. AIST 2015. Communications in Computer and Information Science, vol 542. Springer, Cham. DOI: 10.1007/978-3-319-26123-2_31

36. Lemaître, G., Nogueira, F. & Aridas, C.K. (2017) Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *The Journal of Machine Learning Research*. 18(1). pp. 559–563. DOI: 10.48550/arXiv.1609.06570

Список источников

1. *Maulana I.* Social Media as Public Political Instrument // *Using New Media for Citizen Engagement and Participation*. 2020. P. 181–197.

2. *McGregor S.C., Mourão R.R.* Talking politics on Twitter: Gender, elections, and social networks // *Social media + society*. 2016. Vol. 2(3). P. 2056305116664218.

3. *Рябченко Н.А., Малышева О.П., Гнедаш А.А.* Управление политическим контентом в социальных сетях в период предвыборной кампании в эпоху постправды // Полис. Политические исследования. 2019. № 2 (2). С. 92–106.

4. *Ceron A. et al.* Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France // *New media & society*. 2014. Vol. 16, № 2. P. 340–380.

5. *Mengü S.Ç. et al.* Political preferences of generation Y university student with regards to governance and social media: A study on March 2014 local elections // *Procedia-Social and Behavioral Sciences*. 2015. Vol. 174. P. 791–797.

6. *Oliveira D., Bermejo P., dos Santos P. A.* Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls // Journal of Information Technology & Politics. 2017. Vol. 14, № 1. P. 34–45.

7. *Volkova S., Coppersmith G., Van Durme B.* Inferring user political preferences from streaming communications // Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics. 2014. P. 186–196.

8. *Gayo-Avello D.* A meta-analysis of state-of-the-art electoral prediction from Twitter data // *Social Science Computer Review*. 2013. Vol. 31, № 6. P. 649–679.

9. *Huff C., Tingley D.* "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents // Research & Politics. 2015. Vol. 2, № 3. P. 2053168015604648.

10. *Lacy S. et al.* Issues and best practices in content analysis // Journalism & Mass Communication Quarterly. 2015. Vol. 4. P. 791–811.

11. *Chang C.C., Chiu S.I., Hsu K.W.* Predicting political affiliation of posts on Facebook // Proc. of the 11th International Conference on Ubiquitous Information Management and Communication. 2017. P. 1–8.

12. *Chiu S.I., Hsu K.W.* Predicting political tendency of posts on facebook // Proc. of the 2018 7th International Conference on Software and Computer Applications. 2018. P. 110–114.

13. *David E. et al.* Utilizing Facebook pages of the political parties to automatically predict the political orientation of Facebook users // Online Information Review. 2016.

14. *Feezell J. T., Ortiz B.* 'I saw it on Facebook': an experimental analysis of political learning through social media // *Information, Communication & Society*. 2019. P. 1–20.

15. *Vepsäläinen T., Li, H., Suomi R.* Facebook likes and public opinion: Predicting the 2015 Finnish parliamentary elections // Government Information Quarterly. 2017. Vol. 34, № 3. P. 524–532.

16. *Hernandez-Suarez A. et al.* Predicting political mood tendencies based on Twitter data // 5th Int. Workshop on Biometrics and Forensics. 2017. P. 1–6.

17. *Makazhanov A., Rafiei D., Waqar M.* Predicting political preference of Twitter users // Social Network Analysis and Mining. 2014. Vol. 4, № 1. P. 193.

18. *Preoţiuc-Pietro D. et al.* Beyond binary labels: political ideology prediction of twitter users // Proc. of the 55th Annual Meeting of the Association for Computational Linguistics. 2017. P. 728–740.

19. *Stefanov P. et al.* Predicting the topical stance and political leaning of media using tweets // Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 527–537.

20. *Bolsover G., Howard P.* Chinese computational propaganda: automation, algorithms and the manipulation of information about Chinese politics on Twitter and Weibo // Information, communication & society. 2019. Vol. 22, № 14. P. 2063–2080.

21. *Denemark D., Chubb A.* Citizen attitudes towards China's maritime territorial disputes: traditional media and Internet usage as distinctive conduits of political views in China // Information, Communication & Society. 2016. Vol. 19, № 1. P. 59–79.

22. *Бызов Л.Г. и др.* Идеальный политик для социальной сети: подход к анализу идеологических предпочтений пользователей // Проблемы управления. 2020. Vol. 4. C. 15–26.

23. *Kozitsin I.V. et al.* Modeling Political Preferences of Russian Users Exemplified by the Social Network Vkontakte // Mathematical Models and Computer Simulations. 2020. Vol. 12. P. 185–194.

24. *Baly R. et al.* We Can Detect Your Bias: Predicting the Political Ideology of News Articles // Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020. P. 4982–4991.

25. *Kinash N. et al.* Analysis of large-scale networks using high performance technology (Vkontakte case study) // Creativity in Intelligent, Technologies and Data Science. 2015. Vol. 535. P. 531–541.

26. *Леденев Д.Е.* Раскрытие политических предпочтений в профиле «ВКонтакте»: гендерный аспект // Психология человека и общества. 2021. № 2 (31). С. 23–41.

27. *Забокрицкая Л.Д., Хлебников Н.А., Орешкина Т.А., Комоцкий Е.И.* Возможности изучения ценностей молодежи через профиль социальной сети «ВКонтакте» // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 2 (156). С. 148–167.

28. *Jackson J.* Happy partisans and extreme political views: The impact of national versus local representation on well-being // European Journal of Political Economy. 2019. 58. P. 192–202.

29. *Oraka E. et al.* A cross-sectional examination of US gun ownership and support for gun control measures: sociodemographic, geographic, and political associations explored // Preventive medicine. 2019. 123. P. 179–184.

30. *Pedregosa F. et al.* Scikit-learn: Machine learning in Python // The Journal of machine Learning research. 2011. Vol. 12. P. 2825–2830.

31. *Devlin J. et al.* (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

32. *Glazkova A.* A Comparison of Text Representation Methods for Predicting Political Views of Social Media Users // Proc. of ITIDMS-2021. 2021.

33. *Wolf T. et al.* HuggingFace's Transformers: State-of-the-art Natural Language Processing. arXiv: 1910.03771. 2019.

34. *Paszke A. et al.* Pytorch: An imperative style, high-performance deep learning library // Advances in neural information processing systems. 2019. P. 8026–8037.

35. *Korobov M.* Morphological analyzer and generator for Russian and Ukrainian languages // International Conference on Analysis of Images, Social Networks and Texts. 2015. P. 320–332.

36. *Lemaître G., Nogueira F., Aridas C.K.* Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning // The Journal of Machine Learning Research. 2017. Vol. 18(1). P. 559–563.

Information about the authors:

Glazkova A.V. – University of Tyumen (Tyumen, Russian Federation).
E-mail: a.v.glazkova@utmn.ru
Sokova Z.N. – University of Tyumen (Tyumen, Russian Federation).
E-mail: z.n.sokova@utmn.ru
Kruzhinov V.M. – University of Tyumen (Tyumen, Russian Federation).
E-mail: v.m.kruzhinov@utmn.ru

Contribution of the authors: the authors contributed equally to this article.
The authors declare no conflicts of interests.

Сведения об авторах:
Глазкова А.В. – кандидат технических наук, доцент кафедры программного обеспечения Тюменского государственного университета (Тюмень, Россия).
E-mail: a.v.glazkova@utmn.ru
Сокова З.Н. – доктор исторических наук, профессор, профессор кафедры программного обеспечения Тюменского государственного университета (Тюмень, Россия).
E-mail: z.n.sokova@utmn.ru
Кружинов В.М. – доктор исторических наук, профессор, профессор кафедры отечественной истории Тюменского государственного университета (Тюмень, Россия).
E-mail: v.m.kruzhinov@utmn.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации.
Авторы заявляют об отсутствии конфликта интересов.

The article was submitted 26.05.2021;
approved after reviewing 21.01.2022; accepted for publication 03.03.2022

Статья поступила в редакцию 26.05.2021;
одобрена после рецензирования 21.01.2022; принята к публикации 03.03.2022