Full-scale Personality Prediction on VKontakte Social Network and its Applications

Sergey Titov[1,2], Pavel Novikov[1], Larisa Mararitsa[1,3]

[1]Humanteq, Moscow, Russian Federation
[2]State Academic University for Humanities, Moscow, Russian Federation
[3]National Research University Higher School of Economics, Saint-Petersburg, Russian Federation
{sergey.titov, pavel.novikov, larisa.mararitsa}@humanteq.io

***Abstract*—In this work, we present our prediction results for 17 psychological scales based on online data from the Russian social network VKontakte. As part of this project, we predicted a variety of psychological traits, starting with personality traits from the OCEAN inventory and ending with Raven's Matrices. We then introduced 2 models based on the source of digital data: the first consisting of semi-structured social profiles and the second of public page subscriptions. These models cover most public non-textual data on VKontakte. Lastly, we applied this model to real data with a view to construct psychographics for the audiences of different brands.**

I. Introduction

The emergence of big data for individual users in the form of digital traces has engendered the development of a new direction in the social sciences: computational social psychology [1]. One of its subfields is the inference of classical psychological characteristics from digital footprints.

Most work in this field focuses on the prediction of specific psychological characteristics - in our work, we aim to perform full-scale personality prediction. To this end, we chose to evaluate stable, culturally independent and widely-acknowledged characteristics in classical psychometrics. The result was a series of personality measurement inventories – BIG-5 traits (5 scales) [2], Schwarts' values inventory (10 scales) [3], Raven's intelligence inventory (1 scale) [4] and Golovin's vocabulary inventory [5] (as a Russian substitute of the Mill Hill vocabulary scales [4]). This set of characteristics provides a wide enough view on a person that we might call our models personality prediction models. In this article, we provide an example of a digital study of an audience based on open data from a virtual social network. Such a full-scale personality prediction may prove useful for describing a "buyer persona" or target audience, for marketing purposes.

II. Prior work

M. Kosinski, one of the most notable researchers in this field, has shown on multiple occasions that it is possible to predict personality traits from various digital sources [6]–[8]. His works were mostly centered around myPersonalty dataset collected from Facebook users who agreed to complete BIG5 questionnaire [2] and share some of their Facebook data. On this data set, Kosinski and his team reached the top prediction accuracy - 0.43 Pearson correlation coefficient - in the Openness trait. The least accurate predictions were achieved for Conscientiousness - 0.27 Pearson correlation coefficient [6]. In later works his team achieved an even higher correlation for openness - 0.51 [9], however, this was reached across a relatively small subsample and the metrics were calculated on cross-validation.

For the BIG-5 prediction benchmark, we will rely on more recent results that were published in the 2018 meta-analysis [10]. After reviewing 14 works eligible for analysis, the team estimated the following meta-analytic correlations, as outlined in Table I.

TABLE I. Meta-correlations for big-5 precitions estimated in [10]

Trait	Estimated correlation
Openness	0.39
Conscientiousness	0.35
Extraversion	0.40
Agreeableness	0.29
Neuroticism	0.33

It is important to understand the two limitations of these results: (1) They were mostly determined from Facebook data and English-speaking users - only one of the works was dedicated to the analysis of Chinese-speaking users. (2) The scores were determined through analyzing different works with incomparable preprocessing techniques - thus making it impossible to estimate the significance of any single feature in prediction accuracy.

These two limitations could entail a significant deviation in our results since we are analyzing a Russian-speaking sample from the VKontakte network, which has its own set of unique features.

As for values prediction - latest result was published in 2019 Kalimeri's work [11]: based on apps that users had on their smartphone and internet traffic data, authors achieved average AUROC over binary classification task around 60%.

Additionally, we would like to touch on a series of works that sought to infer further knowledge of a user's personality from less obvious data sources. There were attempts to glean personality traits from images using image aesthetic properties

and deep learning generated features, with good results [12], [13]. For instance, in the 2018 work [13], the author achieved a 0.24 correlation coefficient between the predicted and the true score in the Openness trait, using images from the social network Flickr.

One of the main drivers of progress in this field is the appearance of real-world applications for this technology. We can differentiate at least two fields that could benefit from personality prediction: recommender systems and marketing.

A deeper knowledge of user personality traits could improve the precision of recommender systems and solve the cold start problem. Personality scores could also be used as additional features in existing algorithms or become data for new meta-properties of algorithms, like in Nguyen's work [14]. Nguyen's work hypothesized that people with different psychological trait profiles would have different preferences that could not be inferred directly from ratings. The authors introduced three recommender system characteristics: popularity, serendipity, and diversity, assessing these with information about user personality and feedback. As a result, they observed a series of significant interaction effects between recommender system properties and psychological traits. Such improvements do not require previous user history within a recommendation service and thus could also be used as a possible solution for the cold start problem.

The second application of personality prediction is using predicted profiles in marketing activities which involve personalization, engagement, and audience attraction. The main work in this field is outlined in the Kosinski paper on personalizing Facebook ads [7]. In this work, it is claimed that personality-driven changes (based on user scores in Extraversion and Openness) in advertising could "yield up to 40% more clicks and up to 50% more purchases than their mismatching or unpersonalized counterparts". Although such results provoked a dispute around the validity of this experiment [15], [16], it is important that from the very beginning computational psychology establishes itself not only as an academic field but an applied science.

In our work, we wanted to recreate the full pipeline for the prediction of psychological traits. With this in mind, we used a science-friendly questionnaire app for data collection, developed a prediction pipeline using Python, and applied it to real-world data from the VKontakte network.

III. Data collection

We used the DigitalFreud app for digital and psychological data collection. Users were asked to complete a few psychological tests and provide access to data from their digital footprint. For all users who completed at least one of the tests, the app generated personal psychological insights based on the provided data and test results. All DigitalFreud users agreed to share their data for subsequent research.

For every pair of modalities (a type of data access given by the user) and psychological test, we received slightly different volumes of data because not all users had completed all of the tests (and/or shared all possible data). We collected the following scales:

- BIG-5, BFI-44 version (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) [17]
- Schwartz's value inventory (Stimulation, Benevolence, Achievements, Conformity, Hedonism, Power, Security, Independence, Universalism, Tradition)) [18].
- Golovin vocabulary amount inventory (Verbal Intelligence) [5].
- Raven's progressive matrices (Intelligence)[4].

All of these scales were transformed into a 0-1 scale. Further to this, we rejected data with replies that were too fast and/or identical, as low-quality answers. Lastly, we calculated the quality for the BIG-5 and Schwartz scales using Cronbach alpha, the results of which are displayed in Table II.

For scales like the value of tradition, stimulation, and conformity, we – somewhat predictably - witnessed low scores across the board, due to unreliable measurements. We presume that some of these results were due to specifics of the sample. Cronbach-alpha is usually low in the case of low variance of the signal. In our case, we had a few questions with very low variance; a question about the importance of religion only received a few unique answers - perhaps our sample had a very common view on religion. Additionally, we provided sample statistics for every psychological scale used in the analysis (Table III). We provide data for the entire dataset, but in the analysis, we used only parts due to the scarcity of digital footprint data.

IV. Analisis

The analysis problem was set as a regression task: we wanted to predict every available psychological scale on a score from 0 to 1. In order to assess the performance of the models, we used several standard metrics like predicted and test correlation (Spearman). For testing and training, we used a 0.7 / 0.2 / 0.1 partition of datasets (train, test, holdout). All result metrics were calculated on the holdout datasets.

Additionally, the number of serious misclassifications were also reported. We split the dataset into three equal percentile partitions and checked the first and last ones for mutual misclassification. As had been demonstrated in field overview, we did not expect a very high accuracy in the scores of psychological prediction tasks. Even humans struggle to pinpoint the discrepancies between two extraverts, while distinguishing an introvert from an extravert is quite realistic. Thus, we expect our work to produce plausible prediction metrics and a low percentage of significant errors.

A. Profile model

The first model is based on VKontakte profile data. For this model, we used data that could be collected from a user's personal page on the social network, i.e. their date of birth, home town, relationship status, et cetera. In addition to this, we also used some information from their personal feed.

1) Features and sample: In this model, we mostly used raw data that we collected from the VKontakte API. We also performed some additional feature engineering. The features and their brief description and descriptive statistics are outlined in Table III:

TABLE II. PSYCHOLOGICAL SCALES DESCRIPTIVE STATISTICS

Inventory	Scale	N	Mean	Sigma	Min	Max	Cronbach-alpha
Golovin	Intelligence	6399	19.85	5.32	0	30	-
Raven	Verbal intelligence	6082	60760.4	10564	16000	80030	-
Big-5	Openness	13803	36.07	6.33	12	50	0.77
	Conscientiousness	13803	28.34	6.08	9	45	0.81
	Extraversion	13803	24.09	5.96	8	40	0.79
	Agreeableness	13803	29.98	5.54	9	45	0.72
	Neuroticism	13803	27.04	6.15	8	40	0.83
Schwartz's	Stimulation	6866	11.83	3.21	3	21	0.62
	Benevolence	6866	16.33	3.99	4	28	0.72
	Achievements	6867	16.65	4.32	4	28	0.79
	Conformity	6866	14.33	4.12	4	28	0.66
	Hedonism	6867	13.33	3.26	3	21	0.75
	Power	6867	10.44	3.53	3	21	0.69
	Security	6866	18.68	5.01	5	35	0.68
	Independence	6867	18.43	3.57	4	28	0.68
	Universalism	6866	26.09	5.49	6	42	0.74
	Tradition	6867	12.37	3.97	4	28	0.6

TABLE III. VKONTAKTE PROFILE FEATURES DESCRIPTION STATISTICS

Feature name	Feature description	N	Mean	Sigma	Min	Max
Friends	Number of friends	3262	142.78	392.24	1	10000
Posts	Number of posts	1249	390.01	2289.32	1	45462
Subscriptions	Number of subscriptions	3606	180.5	299.43	1	2873
Followers	Number of followers	3072	119.39	203.33	1	3244
Tv	Number of different tv shows in corresponding section of personal page	249	2.18	2.68	1	17
About	Length of text in about field in symbols	498	113.02	293.86	1	3258
Participants	Number of participants in user's posts feed (amount of commentators on posts)	386	19.81	40.63	1	304
Friend_participant_portion	Ratio of friends in participants	278	0.57	0.21	0.1	1
Repost_portion	Ratio of reposts out of all posts in personal feed	5430	0.35	0.14	0	0.5
Original_posts	Number of original posts	6166	513.51	1980.08	1	49487
Original_len	Average length of original posts	5251	323.36	502.6	0.33	10122
Repost_len	Average length of reposts	5310	46.45	183.35	0.01	8192
Subscriptions_ids	Number of user subscriptions	5723	200.34	246.2	1	2873
Friends	Number of friends	3262	142.78	392.24	1	10000
Posts	Number of posts	1249	390.01	2289.32	1	45462

In the analysis, we also included some properties which were very specific to VKontakte.

- We checked whether the user had uploaded an original photo or was using a stock VK photo as a binary feature.
- It is known that VKontakte attributes the user ID in chronological order of registration - the older the account, the smaller the ID number used. This was leveraged as a sort of innovative metric.
- We used data from the "life views" profile section, using fields with pre-selected VK classifications. We used the following fields: the user's confession, political views, attitudes towards smoking and alcohol, relationship status, the most important traits in others, and life values.

We also used information from the 'hometown', 'education' and 'inspiration' fields. In order to represent information from the 'hometown' field, we took 5 of the most popular cities as variables - in other cases we used the 'other city' feature. For the 'education' field we only used information about their educational status (student/dropped out/finished) and degree level (bachelor/ master/specialist). For the 'inspiration' field we took the top 3 inspirations as binary variables of our sample: books, people and music. All metric features were scaled to [-1,1] interval.

2) Algorithms overview: XGBoost was our algorithm of choice for this task due to its ability to use missed values as features for prediction [19]. For comparison, we also tested elastic net and linear regressions. For regressions, we used a simple imputation of average for missed values..

Regression analysis was performed with the Sklearn[20] library using gridsearch versions of algorithms. For XGBoost, we used XGBoost library with hyperparameter search via Hyperopt, with 300 iterations per model.

3) Results: We achieved considerable results in predicting some scales, as demonstrated in Table IV.

From the results, it is evident that XGBoost outperforms linear regression methods. We suppose this is mainly due to the superior handling of missing values. It is also worth mentioning that some of the scales had near-random results when it came to values of power, hedonism, traditions and conformity. These scales had one of the lowest test quality results - we suppose that for successful prediction we will require a more diverse sample (in terms of values).

B. Subscription model

1) Features and sample description: For the subscription model, we collected data from all user subscriptions. In summary, this amounted to 2,410 users with completed BIG-5 and 199,535 unique subscriptions, and 1,872 users with completed Schwartz's inventory and 162,913 unique subscriptions. In order to avoid high data sparsity, we decided to create embedding over VK subscriptions. To build this, we used the word2vec[21] algorithm. Besides the readily available data, we scraped additional data from users with publicly available profiles, increasing available data to 399,668 unique subscriptions. As a result, we gained a vectorized representation for subscriptions in a 100-dimensional space.

For every user, we transformed their subscription to vector in a newly constructed space. Next, we calculated the mean for all user subscriptions per element in the subscription vector. As a result, we got a 100-element vector of means over user subscriptions for every user and then used it in the prediction.

2) Algorithms Overview: For the prediction, we compared 3 algorithms: XGBoost, LinearSVR, and ElasticNet regression. The first was chosen for its overall performance in prediction tasks [22], the second is a good alternative to XGBoost, and the third was used as a baseline model with basic regularizations. Hyperparameter search was performed for all three models.

3) Results: Results shown in table V. Here we don't have a clear winner between algorithms in terms of prediction accuracy. In this case, XGBoost demonstrates low performance, probably because of data embedding.

The subscription model significantly outperforms the model based on profile data. It is especially visible in the case of BIG-5 prediction, as all traits have much better scores except the agreeableness correlation, which is close to random guess in subscription model. Same drop of prediction performance happened with some of values scales – benevolence and achivments.

It is challenging to compare our results with the works of others due to the absence of conventional metrics around prediction quality. We reported the resulting scores based on holdout part of the dataset which we didn't use in any part of model training. It was important for us to gain more reliable results since we designed this model to be applicable where model robustness is crucial. We also achieved close results with works with the same strictness in score reporting [23].

It was imperative to base all our models on easily-available data. Firstly, the data had to be public. Secondly, the data could not be copyrighted or copyrightable. Lastly, all features had to be easy to acquire and calculate. While it is possible that some other social networking private data (friends list, posts, profile pictures, music) could provide better scores in personality prediction, using them is far more difficult or even unacceptable owing to technical or legal constrictions.

V. APPLICATIONS

In order to demonstrate applications of such models, we decided to carry out simple marketing research with two brand audience comparisons, using our models for psychological profiling. For this task, we parsed publicly available profile and subscription data of users participating in BMW and Mercedes-Benz official VKontakte groups.

We applied our models, took an average of the results and calculated a percentile for these based on our sample. We then averaged per user results and got the mean psychological profile for each group.

TABLE IV. PROFILE MODEL PREDICTION RESULTS

Inventory	Scale	XGBoost correlation	Elastic net correlation	Linear regression correlation	XGBoost p of inverse class	Elastic net p of inverse class	Linear regression p of inverse class
Golovin	Intelligence	0.384	0.39	0.395	0.177	0.184	0.18
Raven	Verbal intelligence	0.215	0.253	0.263	0.215	0.234	0.229
Big-5	Openness	0.189	0.121	0.081	0.25	0.283	0.3
	Conscientiousness	0.179	0.158	0.165	0.275	0.281	0.265
	Extraversion	0.278	0.173	0.16	0.225	0.26	0.264
	Agreeableness	0.123	0.061	0.063	0.283	0.303	0.289
	Neuroticism	0.153	0.135	0.138	0.294	0.282	0.301
Schwartz's	Stimulation	0.339	0.239	0.183	0.194	0.229	0.254
	Benevolence	0.209	0.046	0.139	0.237	-	0.279
	Achievements	0.194	0.044	0.023	0.235	0.291	0.321
	Conformity	0.027	0.093	0.079	0.335	0.287	0.299
	Hedonism	0.077	0.026	-0.023	0.316	0.289	0.316
	Power	0.099	0.024	0.008	0.296	0.317	0.339
	Security	0.115	0.123	0.051	0.281	0.265	0.286
	Independence	0.103	0.01	0.021	0.3	-	0.311
	Universalism	0.102	0.065	0.018	0.308	-	0.338
	Tradition	0.048	0.087	0.114	0.303	0.312	0.293

TABLE V. SUBSCRIPTION MODEL PREDICTION RESULTS

Inventory	Scale	XGBoost correlation	Elastic net correlation	Linear regression correlation	XGBoost p of inverse class	Elastic net p of inverse class	Linear regression p of inverse class
Golovin	Intelligence	0.507	0.507	0.513	0.129	0.129	0.123
Raven	Verbal intelligence	0.371	0.37	0.332	0.17	0.17	0.196
Big-5	Openness	0.304	0.295	0.29	0.209	0.209	0.231
	Conscientiousness	0.396	0.384	0.385	0.133	0.128	0.139
	Extraversion	0.324	0.321	0.32	0.172	0.192	0.192
	Agreeableness	0.02	0.036	-0.033	0.321	0.321	0.348
	Neuroticism	0.211	0.211	0.227	0.261	0.256	0.25
Schwartz's	Stimulation	0.176	0.166	0.107	0.276	0.27	0.303
	Benevolence	0.045	0.021	0.082	0.312	0.355	0.333
	Achievements	0.091	0.116	0.094	0.313	0.306	0.32
	Conformity	0.124	0.129	0.152	0.259	0.296	0.244
	Hedonism	0.171	0.136	0.102	0.224	0.28	0.294
	Power	0.282	0.289	0.257	0.197	0.19	0.219
	Security	0.275	0.303	0.204	0.209	0.202	0.227
	Independence	0.202	0.2	0.155	0.245	0.266	0.266
	Universalism	0.164	0.179	0.069	0.244	0.244	0.289
	Tradition	0.199	0.201	0.186	0.279	0.26	0.286

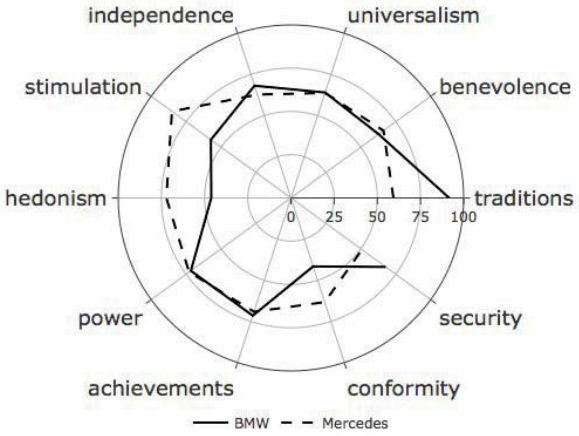

Fig. 1. Values medians comparsion (percentiles) between BMW and Mercedes groups followers

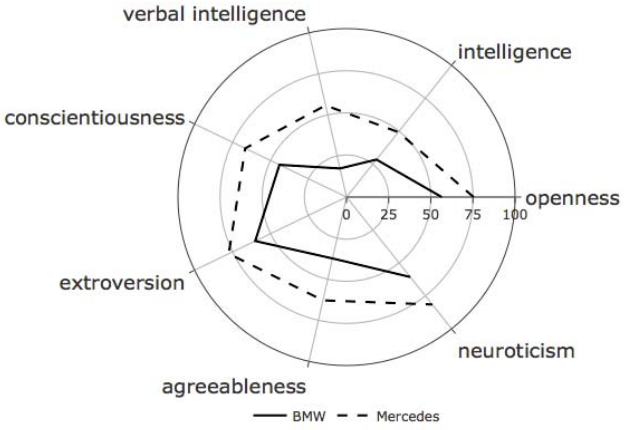

Fig. 2. BIG-5 traits and intelligence inventories medians comparison (percentiles) between BMW and Mercedes groups followers

We performed a Wilcoxon test for every pair of predicted values. All scales except power and universalism demonstrated significant ($p < 0.001$) differences. The most significant properties of the BMW audience were determined to be traditions and security, while Mercedes followers were significantly more extraverted and valued stimulation.

The result of this was that we gained an audience description in psychological terms. This opens some possibilities. First of all, we can now compare how brand values and positioning correspond to their real audience. Secondly, knowledge of the audience's psychological traits makes it possible for businesses to build better communicative strategies.

Lastly, this analysis was performed using publicly available data and could be applied to a competitors' audience, for better market understanding and positioning.

VI. Conclusion

In this paper, we have demonstrated the full psychographic inference pipeline, from data collection to model training and application. Such models not only demonstrated considerable performance but also delivered interpretable and actionable results when applied to a real-world task.

We achieved plausible prediction scores for verbal and general intelligence and BIG-5 inventory, while some of the Schwartz values achieved very low scores. We suppose the quality of prediction relies heavily on the quality of the psychological construct - both inventories with non-self-report scales (Golovin and Raven) demonstrated much better results. Another reason for the lower prediction scores regarding the Schwartz values could be the complexity of representing values in data. It is much easier to demonstrate intelligence than values of universalism or conformity.

We believe that personality prediction using digital footprints is the first step in developing computational psychology. Preferences (page likes, subscriptions) are one of the best sources of information about personality, as shown in many studies, including ours. Research which seeks to reconstruct a user's personality from objective big data helps us understand personality itself. Thus, the use of personality prediction models

will certainly lead to improvements in recommender systems, personal assistants, emotion recognition, marketing, and the HR field [14], [24].

REFERENCES

[1] D. Lazer *et al.*, "Computational social science," *Science (80-.).*, vol. 323, no. 5915, pp. 721–723, 2009.

[2] L. R. Goldberg, "The structure of phenotypic personality traits.," *Am. Psychol.*, vol. 48, no. 1, p. 26, 1993.

[3] A. Bardi and S. H. Schwartz, "Values and behavior: strength and structure of relations.," *Personal. Soc. Psychol. Bull.*, vol. 29, pp. 1207–1220, 2003

[4] J. C. Raven and others, *Raven's progressive matrices and vocabulary scales*. Oxford pyschologists Press, 1998.

[5] G. V. Golovin, "Receptive vocabulary size measurement for russian language", *Social and psycholinguistic reserch*, no. 3, pp. 148–159, 2015.

[6] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 15, pp. 5802–5805, 2013.

[7] S. C. Matz, M. Kosinski, G. Nave, and D. J. Stillwell, "Psychological targeting as an effective approach to digital mass persuasion," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 48, pp. 12714–12719, 2017.

[8] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our twitter profiles, our selves: Predicting personality with twitter," *Proc. - 2011 IEEE Int. Conf. Privacy, Secur. Risk Trust IEEE Int. Conf. Soc. Comput. PASSAT/SocialCom 2011*, no. October, pp. 180–185, 2011.

[9] W. Youyou, M. Kosinski, and D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 4, pp. 1036–1040, 2015.

[10] D. Azucar, D. Marengo, and M. Settanni, "Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis," *Pers. Individ. Dif.*, vol. 124, no. December 2017, pp. 150–159, 2018.

[11] K. Kalimeri, M. G. Beiró, M. Delfino, R. Raleigh, and C. Cattuto, "Predicting demographics, moral foundations, and human values from digital behaviours," *Comput. Human Behav.*, vol. 92, pp. 428–445, 2019.

[12] S. C. Matz, C. Segalin, D. Stillwell, S. R. Müller, and M. W. Bos, *Predicting the Personal Appeal of Marketing Images Using Computational Methods*, vol. 29, no. 3. 2019.

[13] Z. R. Samani, S. C. Guntuku, M. E. Moghaddam, D. Preoţiuc-Pietro, and L. H. Ungar, "Cross-platform and cross-interaction study of user personality based on images on Twitter and Flickr," *PLoS One*, vol. 13, no. 7, pp. 1–19, 2018.

[14] T. T. Nguyen, F. Maxwell Harper, L. Terveen, and J. A. Konstan, "User Personality and User Satisfaction with Recommender Systems," *Inf. Syst. Front.*, vol. 20, no. 6, pp. 1173–1189, 2018.

[15] D. Eckles, B. R. Gordon, and G. A. Johnson, "Field studies of psychologically targeted ads face threats to internal validity," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 23, pp. E5254–E5255, 2018.

[16] S. C. Matz, M. Kosinski, G. Nave, and D. J. Stillwell, "Facebook's optimization algorithms are highly unlikely to explain the effects of psychological targeting," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 23, pp. E5256–E5257, 2018.

[17] G.G. Knyazev, L.G. Mitrofanova, A.V. Bocharov, "Validization of russian version of Goldberg's "Big-five factor markers" inventory", *Journal of psychology*, vol. 31, no. 5, pp. 100–110, 2010.

[18] V. N. Karandashev, "The conception of cultural values by S. Schwartz" *Questions of psychology*, no. 1, pp. 81–96, 2009.

[19] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, vol. 13-17-Augu, pp. 785–794.

[20] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[21] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," pp. 1–12, 2013.

[22] D. Nielsen, "Tree Boosting With XGBoost-Why Does XGBoost Win" Every" Machine Learning Competition?," NTNU, 2016.

[23] S. Kleanthous, C. Herodotou, G. Samaras, and P. Germanakos, "Detecting personality traces in users' social activity," in *International conference on social computing and social media*, 2016, pp. 287–297.

[24] R. Buettner, "Predicting user behavior in electronic markets based on personality-mining in large online social networks: A personality-based product recommender framework," *Electron. Mark.*, vol. 27, no. 3, pp. 247–265, 2017.