

УДК 004.8

АВТОМАТИЗАЦИЯ ОЦЕНКИ ТЕМПЕРАМЕНТА ПОЛЬЗОВАТЕЛЕЙ ОНЛАЙН СОЦИАЛЬНОЙ СЕТИ

© 2023 г. В. Д. Олисеенко^{1,*}, А. О. Хлобыстова^{1,**},
А. А. Корепанова^{1,***}, Т. В. Тулупьева^{1,2,****}

Представлено академиком РАН А.Л. Семеновым

Поступило 31.08.2023 г.

После доработки 15.09.2023 г.

Принято к публикации 15.10.2023 г.

В работе рассматривается вопрос автоматизации предсказания результатов теста “PEN” (тест на темперамент) по численным характеристикам, извлеченным из аккаунтов пользователей одной из популярных русскоязычной онлайн социальной сети. Целью работы является автоматизация оценки выраженности личностных особенностей пользователей онлайн социальной сети путем сопоставления результатов прохождения теста и контента, размещаемого самим пользователем в своем аккаунте онлайн социальной сети при помощи методов машинного обучения. Результатом работы является построение классификаторов на основе моделей CatBoost и случайного леса для предсказания характеристик экстраверсии-интроверсии и нейротизма. Теоретическая ценность результата лежит в разработке подхода построения исследования в области автоматизации оценки выраженности личностных особенностей человека. Практическая значимость заключается в разработке программного модуля для создания автоматизированной системы оценки выраженности личностных особенностей человека по онлайн социальным сетям.

Ключевые слова: PEN, темперамент, машинное обучение, предсказание оценки выраженности личностных особенностей человека, онлайн социальные сети

DOI: 10.31857/S2686954323601471, **EDN:** FRPPOF

1. ВВЕДЕНИЕ

Существует множество задач в различных сферах экономической деятельности (нефтепромышленности, столярной промышленности, стеклопромышленности, сферы упаковки и построения логистики и других [1–4]), где ключевым вопросом для оптимизации бизнес-процессов является эффективное управление человеческими ресурсами [5]. Основной сложностью для такого управления является множество недетерминированных характеристик, которые могут быть присущи человеку: характеристики ум-

ственной деятельности (уровень знаний и умений, способность обучаться и усваивать новые знания); физической активности (занятие спортом, хобби, пищевые привычки, режим сна и т.д.); личности (психологические и психические особенности индивида); биологические шкалы (состояние внутренних органов, например сердца, легких, почек и т.д.) и др. [6] и изменчивостью этих характеристик с течением времени [7]. Среди представленных характеристик наиболее интересными в управлении бизнес-процессами могут быть характеристики личности, т.к. именно они определяют, насколько хорошо человеку подходит специфика работы [7, 8], сможет ли он повысить эффективность работы команды [9], как он будет реагировать на стресс и внешние раздражители [10] и т.д. Кроме того, такую информацию возможно оценить из онлайн социальных сетей, которые ведет человек [8, 11]. Именно поэтому в данной работе будет рассматриваться вопрос автоматизации оценки выраженности личностных особенностей человека в онлайн социальных сетях.

Существует множество различных моделей личности человека, например 16-факторный тест Р. Кеттелла [12], темная триада [13], “Большая пятерка” [14], “PEN” [15] и др. Они отражают неко-

¹Лаборатория теоретических и междисциплинарных проблем информатики, Федеральное государственное бюджетное учреждение науки “Санкт-Петербургский Федеральный исследовательский центр Российской академии наук”, Санкт-Петербург, Россия

²Кафедра государственного и муниципального управления, Северо-Западный институт управления РАНХиГС, Санкт-Петербург, Россия

*E-mail: vdo@dscs.pro

**E-mail: aok@dscs.pro

***E-mail: aak@dscs.pro

****E-mail: tvf@dscs.pro

торые аспекты личности человека и могут быть оценены различными методами психодиагностики (опросы, тесты, проективные техники) [16]. В контексте онлайн социальных сетей одним из применимых и надежных психодиагностических методов являются тесты [16, 17], которые легко реализовать в приложении для автоматизированного сбора и представления результатов их прохождения. После чего результаты прохождения при помощи машинного обучения могут быть сопоставлены с контентом, который размещают пользователи в своих личных аккаунтах онлайн социальной сети [17].

В данной работе в качестве способа оценки выраженности личностных особенностей человека используется русифицированная модель [18] “PEN” (или Eysenck Personality Questionnaire) за авторством Ганса и Сибиллы Айзенка [15]. В предложенной модели измеряются три показателя личностных особенностей: экстраверсия (общительность, активность, открытость), нейротизм (тревожность, неполноценность, подавленность) и психотизм (импульсивность, склонность к риску, ответственность). В качестве метода сбора информации (результатов прохождения теста “PEN”) используется разработанное на базе лаборатории теоретических и междисциплинарных проблем информатики СПб ФИЦ РАН приложение с тестами¹ в онлайн социальной сети ВКонтакте. Важным аспектом сбора информации является его законность. Этот вопрос будет рассмотрен далее, также как и тип анализируемой информации аккаунтов пользователей в социальной сети, описание модели личности человека “PEN” и выбора социальной сети ВКонтакте.

Таким образом, целью работы является автоматизация оценки выраженности личностных особенностей пользователей онлайн социальной сети путем сопоставления результатов прохождения теста и контента, размещаемого самим пользователем в своем аккаунте онлайн социальной сети при помощи методов машинного обучения. Теоретическая ценность результата лежит в разработке подхода построения исследования в области автоматизации оценки выраженности личностных особенностей человека. Практическая значимость заключается в разработке программного модуля для создания автоматизированной системы оценки выраженности личностных особенностей человека по онлайн социальным сетям. Новизна предложенной работы заключается в разработке подхода для автоматизации оценки выраженности личностных особенностей пользователей онлайн социальной сети, который отличается от других используемым тестом

(“PEN”), онлайн социальной сетью и уникальным набором данных.

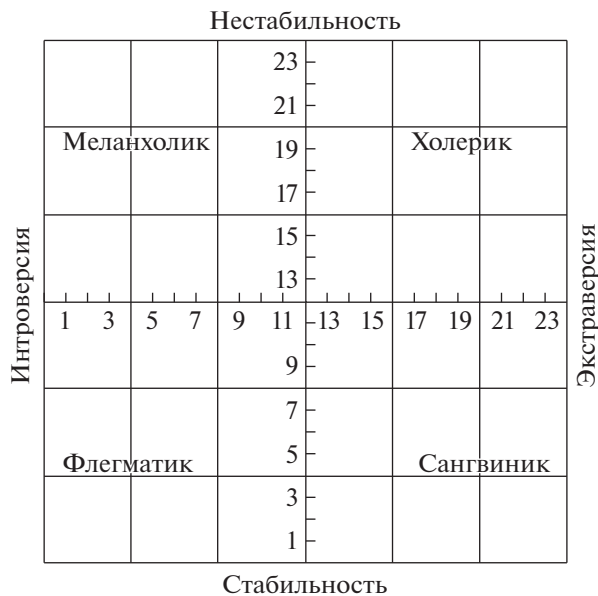
Работа состоит из следующих частей: в первой части приведено описание модели личности человека “PEN”, обоснован выбор социальной сети ВКонтакте, а также описаны правовые рамки проведения эксперимента; во второй части рассмотрены материалы и методы исследования, в т.ч. постановка решаемой проблемы, методология ее решения, описания методики сбора данных и их представления. В части третьей представлены результат эксперимента, сформулировано заключение по проделанной работе.

2. ОПИСАНИЕ МОДЕЛИ ЛИЧНОСТИ “PEN”, ВЫБОРА СОЦИАЛЬНОЙ СЕТИ ВКОНТАКТЕ И ПРАВОВЫЕ ВОПРОСЫ ОБРАБОТКИ ДАННЫХ

Тест “PEN” содержит в себе 101 вопрос с двумя вариантами ответа (“да” или “нет”), результаты которого представляют собой 4 шкалы: экстраверсия-интроверсия; нейротизм; психотизм; искренность отвечающего [15]. Результаты каждого ответа обрабатываются при помощи ключа и соответствуют баллу 1 или 0, который суммируется с другими ответами в выбранной шкале [18] и измеряются в баллах от 0 до 25. Шкала экстраверсия-интроверсия показывает, насколько человек общительный или замкнутый в себе, контролирует свои чувства, склонен к рискованным поступкам и т.д. Согласно расшифровки теста “PEN” [18] результаты в баллах более 12 — экстраверт, от 0 до 12 — интроверт. Шкала нейротизма позволяет оценить эмоциональную устойчивость или неустойчивость (в баллах более 12 — невротик, от 0 до 12 — эмоционально устойчивый). Шкала психотизма определяет склонности к асоциальному поведению, вычурности, неадекватности эмоциональных реакций, высокой конфликтности и другие характеристики человека (в баллах более 10 — высокая выраженность психотизма, от 6 до 10 — повышенный психотизм, от 0 до 5 — выраженность психотизма отсутствует). Первые две шкалы (шкала экстраверсии-интроверсии и нейротизма) определяют типологию личности или темперамент: холерик (сильный, неуравновешенный, подвижный), меланхолик (слабый, неуравновешенный, инертный), флегматик (сильный, уравновешенный, инертный) и сангвиник (сильный, уравновешенный, подвижный) с пограничными состояниями. Формальное разделение данных типологии по баллам шкал показано на рис. 1.

Функциональными и нефункциональными требованиями, предъявляемыми для социальных сетей в данном исследовании, являются доступность в России, наличие API для исследователей, покрытие аудитории пользователей. В соответ-

¹ Психологические тесты. [электронный ресурс] URL: https://vk.com/services?w=app7794698_8126979



шая пятерка” и Майерс-Бриггса [17, 26]. Для таких исследований используются небольшие наборы данных: MyPersonality [27], собранный в социальной сети Facebook и Twitter [28]; Essays [29], состоящий из эссе студентов-респондентов; Pandora [30], состоящий из комментариев пользователей Reddit и др. Основными проблемами представленных наборов данных являются закрытость (например, самый большой набор данных, содержащий результаты “Большой пятерки” MyPersonality недоступен для исследователей вследствие политики конфиденциальности его авторов), небольшой размер и различная природа данных. Например, набор данных [twitter] содержит только твиты и их сопутствующие характеристики (текст, фото, лайки, реплаи и др.), однако не содержит личной информации о пользователях (возраст, пол, родной город, список друзей и т.д.). Таким образом, природа данных и направленность рассматриваемой онлайн социальной сети накладывают ограничения на применимость тех или иных предлагаемых подходов для автоматизации оценки выраженности личностных особенностей человека в других онлайн социальных сетях. Несмотря на это существует некоторое количество подходов, учитывающих цифровые характеристики из аккаунтов пользователей для автоматизации оценки выраженности их личностных особенностей [31]. Например, в работе [32] в качестве данных для предсказания результатов теста “Большая пятерка” используются такие числовые характеристики, как количество лайков на постах, количество обновления статусов, количество мероприятий, в которых участвовал человек, количество подписок на группы, количество хештегов, семейное положение, возраст, пол человека из социальной сети Facebook. После преобразования данные передавались в полносвязную нейронную сеть, решающую задачу многозначной классификации (multi-label), где метками были 5 характеристик личности, оцениваемые тестом “Большая пятерка”. В работе [33] также рассматриваются некоторые числовые характеристики (возраст, пол, город, университет, верифицирован ли аккаунт) из аккаунтов пользователей в китайской онлайн социальной сети Sina Weibo в контексте корреляции с результатами теста “Большая пятерка”. Было сделано заключение, что числовые характеристики имеют высокую корреляцию с результатами теста “Большая пятерка”.

Среди исследований наиболее близкими к данной работе являются [34, 35]. В первом исследовании [34] рассматривается вопрос предсказания результатов теста “Большая пятерка” и теста Шварца и других тестов по извлекаемым из социальной сети ВКонтакте числовым характеристикам (количеству друзей, постов, подписок, подписчиков, количеству записей на стене и т.д.)

при помощи алгоритма XGBoost. Во втором исследовании [35] также была попытка предсказать результаты большой пятерки по постам, комментариям, репостам и числовым характеристикам аккаунта пользователя. В первом исследовании наибольшая выборка была по тесту “Большая пятерка” 2410 результатов, а во втором 1020 результатов. Стоит отметить, что в рассмотренных исследованиях нигде не использовался тест “PEN”, а также только числовые характеристики аккаунта, которые возможно получить при запросе `user.get` к API ВКонтакте.

5. ОПИСАНИЕ НАБОРА ДАННЫХ И ЭКСПЕРИМЕНТА

Набор данных, использованный в данном исследовании, составил 30 116 респондентов. В связи с тем, что некоторые аккаунты респондентов в социальной сети были скрыты настройками приватности после исключения таких аккаунтов, набор данных составил 17 344 респондента. Из аккаунтов для построения предсказательной модели были извлечены 20 указанных выше атрибутов. После анализа атрибутов они были приведены к двум типам — бинарному и количественному. К первому типу атрибутов (`mobile_phone`, `site`, `status`, `home_town`, `city`, `career`, `schools`, `relatives`, `relation`) относились поля в аккаунте, которые являются не обязательными для заполнения пользователями (кроме `sex`), поэтому данные поля были преобразованы в бинарные переменные, обозначающие заполненность поля. Второй тип переменных (`bdate`, `albums`, `followers_count`, `friends`, `pages`, `photos`, `subscriptions`, `user_photos`, `gifts`, `groups`) представлял собой количественные счетчики (кроме `bdate`), поэтому они преобразовывались только заменой выбросов на 95 квантиль. Дата рождения (`bdate`) в аккаунтах указывается в формате ДД.ММ.ГГ, поэтому она была преобразована в возраст, если дата рождения указана с го-

Таблица 1. Статистика бинарных переменных

Атрибут	Поле заполнено	Поле не заполнено
<code>mobile_phone</code>	1058	16286
<code>site</code>	1699	15645
<code>status</code>	9553	7791
<code>sex</code>	11754 (женский)	5590 (мужской)
<code>city</code>	9833	7511
<code>career</code>	8262	9082
<code>home_town</code>	4080	13264
<code>relation</code>	2605	14739
<code>schools</code>	8262	9082
<code>relatives</code>	9082	8262

Таблица 2. Статистики количественных переменных

Атрибут	Медиана	Минимальное	Максимальное	ДИ 95%
albums	1	0	36	(3.13, 3.32)
followers_count	193	0	51 252	(372.55, 402.69)
friends	83	0	508	(123.19, 127.02)
pages	79	0	624	(141.18, 146.11)
photos	13	0	468	(63.32, 66.88)
subscriptions	0	0	26	(3.33, 3.53)
user_photos	0	0	34	(1.3, 1.46)
gifts	0	0	423	(37.14, 39.86)
groups	0	0	420	(31.29, 33.91)

дом, если нет, то возраст заменялся на 0. Статистика для бинарных переменных указана в табл. 1, а в табл. 2 — статистика (медиана, min, max, 95% доверительный интервал для среднего с использованием t-распределения) для количественных характеристик из аккаунтов пользователей.

Указанные числовые характеристики после предобработки были использованы для обучения моделей CatBoost, случайный лес и метод опорных векторов (LinearSVC и RandomForestClassifier из пакета scikit-learn). В качестве метки класса использовались шкалы экстраверсия-интроверсия и нейротизма, переведенные в {0, 1}. Для каждой шкалы строилось 3 модели. В качестве метрик оценки качества построенных моделей были использованы метрики AUC, Accuracy и F1 на тестовых данных с использованием перекрестной проверки на 4 блоках. Также для CatBoost был использован метод GridSearch для подбора оптимальных параметров. Финальные параметры для Catboost следующие: функция потерь — Logloss, параметр остановки обучения early_stopping_rounds = 20, коэффициент L2 регуляризации функции стоимости l2_leaf_reg = 4, глубина деревьев depth = 5. Результаты метрик представлены в табл. 3.

В табл. 3 показаны результаты метрик, усредненные на 4 тестовых блоках при проведении перекрестной проверки. В верхней части таблицы результаты классификации для шкалы экстраверсии-интроверсии, а в нижней нейротизма. Полужирным в таблице отмечены наилучшие метрики в сравнении между различными моделями.

6. ОБСУЖДЕНИЕ И ЗАКЛЮЧЕНИЕ

В данной работе была проведена попытка автоматизации оценки выраженности личностных особенностей пользователей онлайн социальной сети путем сопоставления результатов прохождения теста PEN и контента, размещаемого самим пользователем в своем аккаунте онлайн социальной сети при помощи методов машинного обуче-

ния. По полученным результатам можно сказать, что оценить выраженность личностных характеристик человека экстраверсия-интроверсия и нейротизм возможно предсказать с точностью, значительно отличающейся от случайной, на что указывают метрики, превышающие 0.5. Стоит также отметить, что в качестве признаков для моделей использовались только численные характеристики из аккаунтов пользователей в онлайн социальной сети без извлечения дополнительных признаков при помощи глубокого обучения из текста, фото, видео и других данных. Что делает построение прогноза на большом количестве пользователей достаточно быстрым. Дальнейшими перспективами исследования являются расширение набора данных, улучшение качества модели за счет поиска и добавления новых признаков, а также внедрения полученных результатов в системы управления персоналом.

Цель работы, заключающаяся в автоматизации оценки выраженности личностных особенностей пользователя онлайн социальной сети, достигнута. Теоретическая ценность результата лежит в разработке подхода построения исследования в области автоматизации оценки личности человека. Практическая значимость заключается в раз-

Таблица 3. Результаты метрик

Экстраверсия-интроверсия			
	Accuracy	F1	AUC
LinearSVC	0.43	0.19	0.51
RandomForestClassifier	0.45	0.51	0.45
CatBoost	0.62	0.64	0.63
Нейротизм			
	Accuracy	F1	AUC
LinearSVC	0.52	0.55	0.53
RandomForestClassifier	0.86	0.89	0.85
CatBoost	0.65	0.74	0.60

работке программного модуля для создания автоматизированной системы оценки выраженности личностных особенностей человека по онлайн социальным сетям.

СПИСОК ЛИТЕРАТУРЫ

1. *Shen F., Ren S.S., Zhang X.Y., Luo H.W., Feng C.M.* A Digital Twin-Based Approach for Optimization and Prediction of Oil and Gas Production // *Mathematical Problems in Engineering*. 2021. Vol. 2021. Article ID 3062841. <https://doi.org/10.1155/2021/3062841>
2. *Barni A., Pietraroia D., Züst S., West S., Stoll O.* Digital Twin Based Optimization of a Manufacturing Execution System to Handle High Degrees of Customer Specifications. *J. Manuf. Mater. Process.* 2020. 4. 109. <https://doi.org/10.3390/jmmp4040109>
3. *Zhang H., Liu Q., Chen X., Zhang D., Leng J.* A Digital Twin-Based Approach for Designing and Multi-Objective Optimization of Hollow Glass Production Line. *IEEE Access*. 2017. № 5. P. 26901–26911. <https://doi.org/10.1109/ACCESS.2017.2766453>
4. *Marmolejo-Saucedo J.A.* Digital Twin Framework for Large-Scale Optimization Problems in Supply Chains: A Case of Packing Problem. *Mobile Netw Appl.* 2021. <https://doi.org/10.1007/s11036-021-01856-9>
5. *Tao F., Qi Q., Wang L., Nee A.Y.C.* Digital Twins and Cyber-Physical Systems toward Smart Manufacturing and Industry 4.0 // *Correlation and Comparison, Engineering*. 2019. V. 5. Iss. 4. P. 653–661. <https://doi.org/10.1016/j.eng.2019.01.014>
6. *Sahal R., Alsamhi S.H., Brown K.N.* Personal Digital Twin: A Close Look into the Present and a Step towards the Future of Personalised Healthcare Industry // *Sensors*. 2022. 22. 5918. <https://doi.org/10.3390/s22155918>
7. *Alhendi O.* Personality Traits and Their Validity in Predicting Job Performance at Recruitment: a Review // *International Journal of Engineering and Management Sciences (IJEMS)*. 2019. V. 4. № 3. P. 222–231. <https://doi.org/10.21791/IJEMS.2019.3.21>
8. *Хлобыстова А.О., Абрамов М.В., Столярова В.Ф.* Исследование тенденций взаимосвязи между профориентационными предпочтениями пользователей и их цифровыми следами в социальной сети // *Научно-технический вестник информационных технологий, механики и оптики*. 2023. Т. 23. № 3. С. 564–574. <https://doi.org/10.17586/2226-1494-2023-23-3-564-574>
9. *Curşeu P.L., Ilies R., Virgă D., Maricuţoiu L., Sava F.A.* Personality characteristics that are valued in teams: Not always “more is better”? // *International Journal of Psychology*. 2019. V. 54. №. 5. P. 638–649. <https://doi.org/10.1002/ijop.12511>
10. *De Clercq D., Haq I.U., Azeem M.U.* Time-related work stress and counterproductive work behavior: Invigorating roles of deviant personality traits // *Personnel Review*. 2019. Vol. 48. № 7. P. 1756–1781. <https://doi.org/10.1108/PR-07-2018-0241>
11. *Huang C.* Social network site use and Big Five personality traits: A meta-analysis // *Computers in Human Behavior*. 2019. V. 97. P. 280–290. <https://doi.org/10.1016/j.chb.2019.03.009>
12. *Cattell R.B., Gibbons B.D.* Personality factor structure of the combined Guilford and Cattell personality questionnaires // *Journal of Personality and Social Psychology*. 1968. V. 9. № 1. P. 107–120. <https://doi.org/10.1037/h0025724>
13. *Furnham A., Richards S.C., Paulhus D.L.* The Dark Triad of personality: A 10 year review // *Social and personality psychology compass*. 2013. V. 7. № 3. P. 199–216. <https://doi.org/10.1111/spc3.12018>
14. *Saucier G., Goldberg L.R.* What is beyond the Big Five? // *Journal of personality*. 1998. V. 66. P. 495–524.
15. *Eysenck H.J., Eysenck S.B.G.* Eysenck personality questionnaire-revised. 1984. <https://doi.org/10.1037/t05461-000>
16. *Kline P.* *Handbook of Psychological Testing*. London: Routledge. 2013. 752 p. <https://doi.org/10.4324/9781315812274>
17. *Mehta Y., Majumder N., Gelbukh A., Cambria E.* Recent trends in deep learning based personality detection // *Artificial Intelligence Review*. 2020. P. 2313–2339. <https://doi.org/10.1007/s10462-019-09770-z>
18. *Суходольский Г.В.* Математическая психология. Харьков: Изд-во Гуманитарный центр, 2006. 360 с.
19. *Li L., Li A., Hao B., Guan Z., Zhu T.* Predicting Active Users' Personality Based on Micro-Blogging Behaviors // *PLOS ONE*. 2014. № 9 (1). e84997. <https://doi.org/10.1371/journal.pone.0084997>
20. *Evin M., Hidalgo-Munoz A., Béquet A.J., Moreau F., Tattegrain H., Berthelon C., Fort A., Jallais C.* Personality trait prediction by machine learning using physiological data and driving behavior // *Machine Learning with Applications*. 2022. V. 9. № 100353. <https://doi.org/10.1016/j.mlwa.2022.100353>
21. *Zhao J., Zeng D., Xiao Y., Che L., Wang M.* User personality prediction based on topic preference and sentiment analysis using LSTM model // *Pattern Recognition Letters*. 2020. V. 138. P. 397–402. <https://doi.org/10.1016/j.patrec.2020.07.035>
22. *Kim Y., Kim J.H.* Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study // *Information Processing and Management*. 2018. № 54(6). P. 1101–1114. <https://doi.org/10.1016/j.ipm.2018.07.005>
23. *Omar B., Dequan W.* Watch, share or create: The influence of personality traits and user motivation on TikTok mobile video usage // *International Journal of Interactive Mobile Technologies (IJIM)*. № 14 (4). P. 121–137. <https://doi.org/10.3991/ijim.v14i04.12429>
24. *Kosinski M., Stillwell D., Graepel T.* Private traits and attributes are predictable from digital records of human behavior // *Proceedings of the National Academy of Sciences*. 2013. № 15. P. 5802–5806. <https://doi.org/10.1073/pnas.1218772110>
25. *Stachl C., Pargent F., Hilbert S., Harari G. M., Schoedel R., Vaid S., Gosling S.D., Bühner M.* Personality research and assessment in the era of machine learning // *European Journal of Personality*. 2020. V. 34. № 5. P. 613–

631.
<https://doi.org/10.1002/per.2257>
26. *Mehta Y., Majumder N., Gelbukh A., Cambria E.* Recent trends in deep learning based personality detection // *Artificial Intelligence Review*. 2019. P. 2313–2339.
<https://doi.org/10.1007/s10462-019-09770-z>
 27. *Kosinski M., Matz S.C., Gosling S.D., Popov V., Stillwell D.* Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines // *American Psychologist*. 2015. V. 70. № 6. P. 543.
<https://doi.org/10.1037/a0039210>
 28. *Plank B., Hovy D.* Personality Traits on Twitter or How to Get 1,500 Personality Tests in a Week // *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 2015. P. 92–98.
<https://doi.org/10.18653/v1/W15-2913>
 29. *Pennebaker J.W., King L.A.* Linguistic styles: Language use as an individual difference // *Journal of personality and social psychology*. 1999. V. 77. № 6. P. 1296.
<https://doi.org/10.1037/0022-3514.77.6.1296>
 30. *Gjurković M., Karan M., Vukojević I., Bošnjak M., Šnajder J.* PANDORA talks: Personality and demographics on Reddit // *Proceeding's 9th International ACL Workshop on Natural Language Processing for Social Media (ACL, Pennsylvania)*. 2021. P. 138–152. ISBN 978-1-954085-32-9
 31. *Kosan M.A., Karacan H., Urgan B.A.* Predicting personality traits with semantic structures and LSTM-based neural networks // *Alexandria Engineering Journal*. 2022. V. 61. Iss. 10. P. 8007–8025.
<https://doi.org/10.1016/j.aej.2022.01.050>
 32. *Başaran S., Ejimogu O.H.* A neural network approach for predicting personality from Facebook data // *Sage Open*. 2021. V. 11. № 3.
<https://doi.org/10.1177/21582440211032156>
 33. *Gu H., Wang J., Wang Z., Zhuang B., Su F.* Modeling of User Portrait Through Social Media // *2018 IEEE International Conference on Multimedia and Expo (ICME)*. 2018.
<https://doi.org/10.1109/icme.2018.8486595>
 34. *Titov S., Novikov P., Mararitsa, L.* Full-scale Personality Prediction on VKontakte Social Network and its Applications // *2019 25th Conference of Open Innovations Association (FRUCT)*. 2020. P. 317–323.
<https://doi.org/10.23919/FRUCT48121.2019.8981513>
 35. *Станкевич М.А., Игнатъев Н.А., Смирнов И.В., Кисельникова Н.В.* Выявление личностных черт у пользователей социальной сети ВКонтакте // *Вопросы кибербезопасности*. 2019. № 4 (32). С. 80–87.
<https://doi.org/10.21681/2311-3456-2019-4-80-87>

AUTOMATING THE TEMPERAMENT ASSESSMENT OF ONLINE SOCIAL NETWORK USERS

V. D. Oliseenko^a, A. O. Khlobystova^a, A. A. Korepanova^a, and T. V. Tulupyeva^{a,b}

^a*Laboratory of Theoretical and Interdisciplinary Problems of Computer Science,
“St. Petersburg Federal Research Center of the Russian Academy of Sciences”, St. Petersburg, Russia*

^b*Department of State and Municipal Administration, Northwestern Institute of Management RANEPa, St. Petersburg, Russia*

Presented by Academician of the RAS A.L. Semenov

The paper deals with the problem of automating the prediction of the Eysenck Personality Questionnaire (temperament test) results by numerical characteristics extracted from the accounts of users of a popular Russian-language online social network. The purpose of the work is to automate the evaluation of the expression of personality traits of online social network users by comparing the results of the test and the content posted by the user on his or her account using machine learning methods. The result of the work is the construction of classifiers based on CatBoost and random forest models for predicting the expression of extraversion-introversion and neuroticism. The theoretical value of the result lies in the development of the approach to research design in the field of automating the assessment of human personality traits expression. Practical significance lies in the development of a program module for assessing the expression of human personality traits on online social networks.

Keywords: PEN model, temperament, machine learning, prediction of personality traits, online social networks