

Предсказание результатов теста Р. Кеттелла на основе профилей пользователей социальной сети*

Г. Е. Рязанцев¹, В. Д. Олисеенко², М. В. Абрамов², Т. В. Тулупьева^{1,3}

¹ Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

² Санкт-Петербургский Федеральный исследовательский центр РАН, Санкт-Петербург, Россия

³ Северо-Западный институт управления РАНХиГС, Санкт-Петербург, Россия

Аннотация. В статье описывается исследование цифровых следов пользователей в социальной сети и результаты прохождения 16-факторного теста Р. Кеттелла. Метод исследования заключается в применении статистических инструментов и релевантных алгоритмов машинного обучения к анкетным данным полученным со страниц пользователей. Основные результаты эксперимента заключаются в выявлении корреляции между факторами, оцениваемыми тестом Р. Кеттелла, и цифровыми следами, а также построении предсказательных моделей. Лучшие результаты среди методов машинного обучения для предсказания результатов теста Р. Кеттелла показали алгоритмы градиентного бустинга с максимальным значением метрики F1-micro 0,606, которое удалось достичь на факторе «эмоциональная чувствительность» (фактор I). Практическая значимость работы заключается в разработке инструмента для автоматизированного предсказания результатов теста Р. Кеттелла на основе цифровых следов пользователя. Теоретическая значимость заключается в разработке метода для автоматизации оценки выраженности личностных особенностей пользователей социальных сетей по их цифровым следам.

Ключевые слова: машинное обучение, корреляционный анализ, социальные медиа, психологический тест, тест Р. Кеттелла, профилирование.

DOI 10.14357/20718632240106

EDN ESXDTN

Введение

Вопрос оценки выраженности личностных особенностей человека является важной задачей, которая имеет множество приложений в различных сферах информационного общества: медицинской [1–3], банковской [4–6], маркетинговой [7–9], информационной безопасности [10, 11] и др. В контексте информационной безопасности выявленные личностные особенности

человека могут косвенно влиять на успешность атак с элементами социальной инженерии [11]. Особый интерес может лежать в предсказании такой оценки по косвенным признакам — общедоступной информации о человеке, которую он сам размещает в своих социальных медиа. В классическом варианте проведения психологического исследования человека для оценки выраженности личностных особенностей используются тесты, проективные методики, психофизиологические

* Работа выполнена в рамках проекта по государственному заданию СПб ФИЦ РАН № FFZF-2022–0003; при финансовой поддержке гранта Президента Российской Федерации МК-5237.2022.1.6.

подходы и другие методы [12]. С одной стороны, при решении реальных задач может возникнуть необходимость провести анализ большого количества людей в сжатые сроки (например, всех сотрудников компании), что обуславливает требования к наличию профильных специалистов, времени, финансовых и иных ресурсов для такого анализа. С другой стороны, множество людей активно вовлечены в использование социальных медиа, имеют там свои личные профили и оставляют цифровые следы, которые могут содержать пол, возраст, фотографии, посты, подписки, интересы и другую информацию. Одним из наиболее ярких примеров такого социального медиа в русскоязычном сегменте интернета является социальная сеть «ВКонтакте», которую ежемесячно посещают 100 млн. человек, оставляющих ежедневно 15 млрд. сообщений¹.

Согласно исследованиям [13–18] цифровые следы могут содержать в себе информацию о некоторых личностных особенностях пользователей. Так в [15] авторы показали, что анализируя текст и изображения, размещенные пользователями в социальной медиа, можно предсказывать депрессию. В другой публикации [16] рассматривалось поведение, связанное с выкладыванием автопортретов, созданных при помощи смартфона, и результаты теста "Большая пятерка". В [17] авторы показали, что лайки в социальных сетях могут отражать много личной информации: такой, как отношение к употреблению алкоголя, политические взгляды, выраженность личностных особенностей и др. Несмотря на достаточно высокую публикационную активность по данной тематике единого подхода, решающего проблему выявления взаимосвязи (и ее автоматизации) между оценкой выраженности личностных особенностей пользователей и их цифровыми следами, нет.

Таким образом, возникает несоответствие, заключающееся в большом количестве общедоступной информации о пользователях социальных медиа, наличии потребности в оценках выраженности личностных особенностей и отсутствии соответствующих методов, моделей и

алгоритмов исследования. Цель работы заключается в ускорении экспресс-оценки выраженности личностных особенностей пользователей с использованием информации из их социальных медиа, посредством выявления взаимосвязи между оценкой выраженности личностных особенностей по 16-факторному личностному тесту Р. Кеттелла² и цифровыми следами, оставляемыми пользователями на своей странице в социальной медиа. Практическая значимость работы заключается в разработке инструмента для автоматизированного предсказания результатов теста Р. Кеттелла на основе цифровых следов пользователя. Теоретическая значимость заключается в разработке метода для автоматизации оценки выраженности личностных особенностей пользователей социальных сетей по их цифровым следам. Новизна полученного результата заключается в адаптации существующего теоретического и практического инструментария для решения новой прикладной задачи.

1. Релевантные работы

В данной работе в качестве источника информации о выраженности личностных особенностей пользователей будут рассматриваться результаты прохождения ими 16-факторного теста Р. Кеттелла [19]. Этот тест дает более широкую оценку выраженности личностных особенностей человека за счет большего количества рассматриваемых факторов, чем другие известные авторам тесты: "Большая пятерка" (Big Five), тест Майерс-Бриггс (МВТИ), тест Айзенка ПЕН и др. [20]. Вопрос выявления взаимосвязей между контентом профилей пользователей в социальных медиа и результатами данного теста является мало изученным. При поиске по ключевым словам и их производным *correlation Cattell test*, *predicting Cattell test*, «предсказание результатов теста Кеттелла», «взаимосвязь тест Кеттелла и социальных сетей» и другим производным на сайте национальной электронной библиотеки³ релевантных статей не нашлось. Но было обнаружено множество исследований, в

¹ Статистика «ВКонтакте» [электронный ресурс]. URL: <https://vk.com/about>

² 16-факторный личностный опросник Р.Б. Кеттелла [электронный ресурс]. URL: https://psychojournal.ru/tests_online/365-16-faktornyy-lichnostnyy-oprosnik-r-b-kettella.html

³ Национальная электронная библиотека [электронный ресурс]. URL: <https://elibrary.ru/>

которых авторы пытались выявить взаимосвязи между результатами прохождения теста "Большая пятерка" и контентом в социальных медиа.

Рассмотрим эти исследования на предмет отбора статистических методов и методов машинного обучения, которые имеют перспективу быть эффективными в нашей задаче, решаемой в данной работе. Так в [21] авторы исследуют взаимосвязь между данными из социальных сетей Facebook⁵ (количество друзей, частота обновления статуса), Twitter (частота использования, количество подписчиков) и нарциссизмом. Взаимосвязь проверялась с помощью корреляционного анализа. Авторы показали, что все четыре признака статистически значимо коррелируют с нарциссизмом. В [22] авторы используют корреляционный и иерархический регрессионный анализ для выявления взаимосвязи между данными из социальной сети Facebook⁴ (пол, возраст, время в социальной сети, оцениваемое испытуемыми, и количество друзей) и симптомами депрессии. В результате эксперимента авторы выявили, что количество друзей, возраст и время, проведенное в социальной сети, не взаимосвязаны с симптомами депрессии, а пол статистически значимо взаимосвязан (у мужчин выше оценка депрессивных симптомов). В [23] авторы предсказывают результаты теста "Большая пятерка" с помощью алгоритмов машинного обучения с учителем, сравнивая следующие модели: дерево решений, метод опорных векторов, нейронные сети и градиентный бустинг над решающими деревьями, а также метод опорных векторов. Для сравнения моделей использовалась F1-мера. В результате исследования градиентный бустинг над решающими деревьями и метод опорных векторов показали лучшие результаты классификации.

Статья [24] наиболее близка к нашему исследованию. Ее авторы изучают взаимосвязь между данными пользователей из «ВКонтакте» и результатами прохождения ими тестов "Большая пятерка", Шварца и Равена. При этом рассматривались следующие модели: логистическая регрессия (LogisticRegression), логистическая регрессия с L1- и L2-регуляризацией (ElasticNet),

градиентный бустинг (XGBoost). По результатам тестирования моделей лучше всего предсказания выдала модель XGBoost. В ранее представленной работе авторов [25] был предложен подход на основе анализа постов пользователей социальной сети для предсказания результатов 16-факторного личностного теста Р. Кеттелла. В данной статье будет описан другой вид данных для такого предсказания — атрибуты из профилей пользователей.

Таким образом, среди рассмотренных подходов можно выделить следующие: статистический (регрессионный и корреляционный анализ) и машинное обучение, где одним из лучших алгоритмов является алгоритм градиентного бустинга, но не один из них не был использован для предсказания результатов теста Р. Кеттелла.

2. Эмпирические данные и их описание

Для исследования использовалась обезличенная выборка из 300 пользователей, которые прошли 16-факторный тест Р. Кеттелла в приложении⁵ «ВКонтакте», разработанном на базе лаборатории теоретических и междисциплинарных проблем информатики Санкт-Петербургского Федерального исследовательского центра РАН. Из профилей пользователей были использованы следующие цифровые следы (атрибуты):

- *bdate* — дата рождения,
- *education_form* — форма обучения (очное или заочное),
- *faculty* — идентификатор факультета,
- *videos* — количество видео,
- *friends* — количество друзей,
- *education_status* — статус (например, «выпускник (специалист)»),
- *city* — город,
- *subscriptions* — количество подписок,
- *photos* — количество фотографий,
- *graduation* — год окончания обучения,
- *groups* — количество сообществ,
- *articles* — количество статей,
- *followers* — количество подписчиков,

⁴ Организация Meta, которой принадлежит Facebook, признана экстремистской, запрещена на территории РФ

⁵ Психологические тесты [электронный ресурс]. URL: <https://vk.com/ticpsystests>

- *audios* — количество аудиозаписей,
- *clips* — количество клипов,
- *universities* — список университетов,
- *albums* — количество альбомов,
- *sex* — пол,
- *pages* — количество объектов в блоке «Интересные страницы»,
- *counts* — количество постов на странице пользователя.

Кроме этого, в результате анализа указанных цифровых следов в собранной выборке возникла необходимость их предобработки для последующего использования в алгоритмах:

- Дата рождения преобразуется в полное количество лет.
- Выделены 5 самых популярных городов, оставшиеся города отнесены к шестому классу (для снижения размерности признака).
- Из атрибута *education_status* было выделено 4 класса:
 - не указавшие,
 - те, кто ещё учится,
 - те, кто уже окончил обучение,
 - абитуриенты.
- Атрибут *universities* был преобразован в категориальную переменную.
- Атрибут *education_form* был преобразован в категориальную переменную (очно, заочно и отсутствует).

Тест Р. Кеттелла измеряет следующие черты пользователя:

«А: общительность»,
«В: уровень интеллекта»,
«С: эмоциональная устойчивость»,
«Е: доминантность»,
«F: беспечность»,
«G: моральная нормативность»,
«Н: смелость»,
«I: эмоциональная чувствительность»,

1 до 10. Но зачастую специалистам в области психологии важно, насколько черта выражена в одну из сторон. Таким образом, можно свести задачу предсказания результатов теста к задаче предсказания трех классов для каждого из 16 факторов:

- от 0 до 3 — низкая выраженность фактора;
- от 4 до 7 — отсутствие выраженности фактора;
- от 8 до 10 — высокая выраженность фактора.

3. Статистический анализ

Для проверки взаимосвязи между цифровыми следами пользователей социальной медиа «ВКонтакте» и результатами их прохождения 16-факторного теста Р. Кеттелла (в шкале [0;10]) была рассчитана корреляция Пирсона⁶ (Рис. 1).

Ось X (подписи по горизонтали, внизу рисунка) обозначает факторы личности, описанные в разделе 2; ось Y (подписи по левому краю вертикали рисунка) — атрибуты, содержащиеся в цифровых следах из социальной сети «ВКонтакте». В каждой ячейке записана корреляция (в интервале [-1; +1]) фактора личности и данных пользователя из социальной сети. Для удобства оценки взаимосвязи каждая ячейка была раскрашена: чем темнее цвет, тем выше корреляция, т.е. сильнее линейная взаимосвязь между цифровыми следами и факторами личности. Из Рис. 1 видно, что многие черты личности имеют

«L: подозрительность»,
«M: мечтательность»,
«N: дипломатичность»,
«O: тревожность»,
«Q1: восприимчивость к новому»,
«Q2: самостоятельность»,
«Q3: самодисциплина»,
«Q4: напряженность».

В представленном тесте содержится 187 вопросов с тремя вариантами ответа по каждому вопросу. В результате тестирования каждый фактор оценивается в системе стенов оценкой от

низкую корреляцию с данными из «ВКонтакте», но все-таки между некоторыми данными существует взаимосвязь. В качестве подтверждения этого был проведен t-тест⁷ для проверки оценки

⁶Коэффициент корреляции Пирсона [электронный ресурс]. URL: http://www.machinelearning.ru/wiki/index.php?title=Коэффициент_корреляции_Пирсона

⁷ T-тест [электронный ресурс]. URL: <https://www.britannica.com/science/Students-t-test>

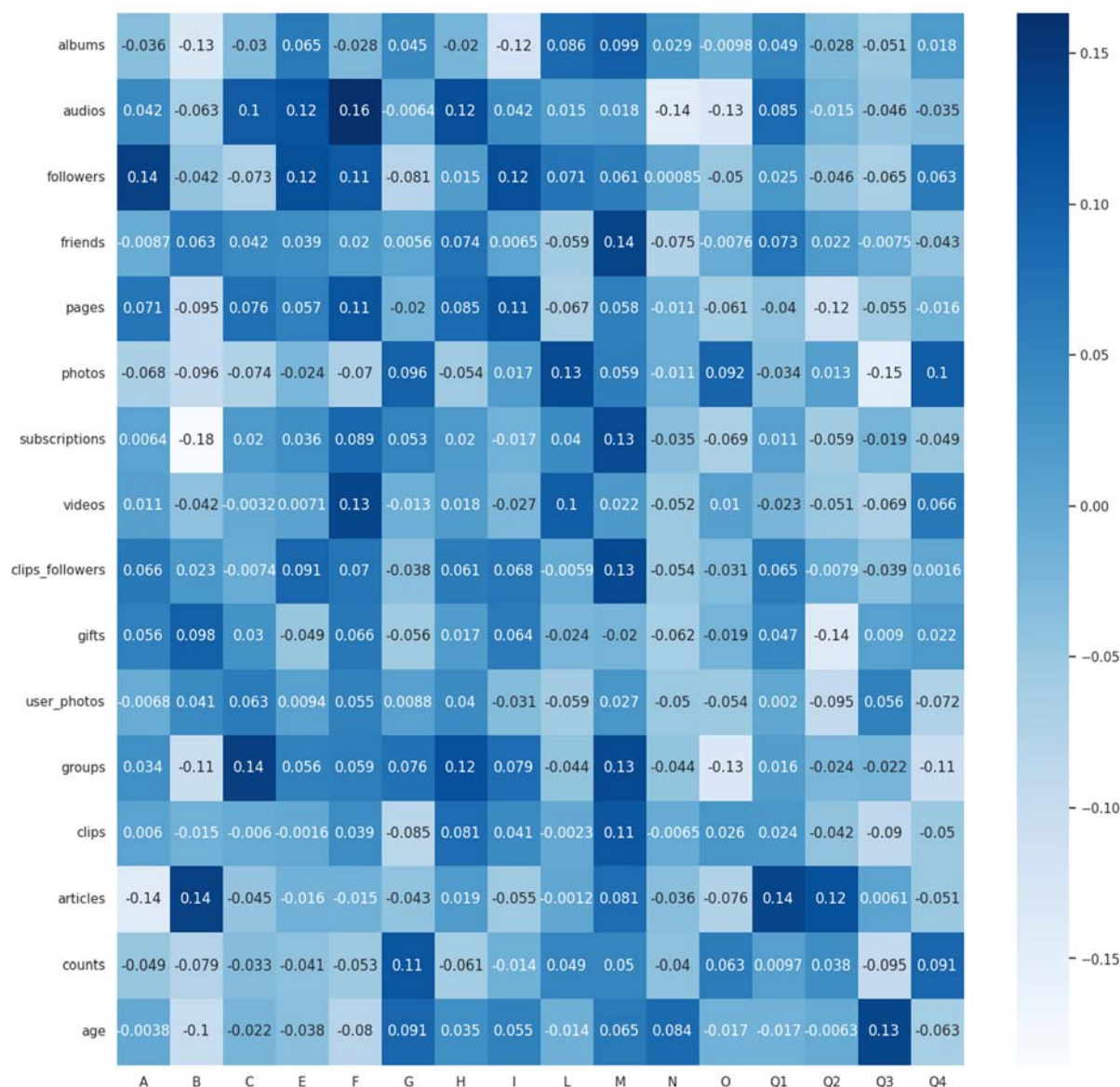


Рис. 1. Корреляция между чертами личности и данными пользователей

значимости коэффициента корреляции, с подсчетом p-value. В качестве гипотезы проверялась линейная зависимость между результатами теста и рассматриваемыми цифровыми следами.

Ниже приведены факторы и цифровые следы, извлеченные из «ВКонтакте», p-value корреляции которых был меньше 0,05, т.е. корреляция была значима:

A — <i>articles</i> (0,013)	L — <i>photos</i> (0,022)
B — <i>subscriptions</i> (0,004)	M — <i>friends</i> (0,021)
B — <i>articles</i> (0,014)	M — <i>groups</i> (0,022)
E — <i>audios</i> (0,026)	N — <i>audios</i> (0,006)
E — <i>followers</i> (0,006)	O — <i>audios</i> (0,020)
E — <i>pages</i> (0,032)	O — <i>groups</i> (0,022)
F — <i>videos</i> (0,022)	Q3 — <i>groups</i> (0,13)
I — <i>albums</i> (0,023)	Q3 — <i>age</i> (0,025)

Из результатов корреляционного анализа можно сделать следующие выводы:

- Черта А (общительность) положительно связана с количеством статей, опубликованных пользователем на своей странице.
- Черта В (уровень интеллекта) отрицательно связана с количеством подписок, но положительно связана с количеством статей, которые опубликовал пользователь.
- Черта Е (доминантность) положительно связана с количеством аудиозаписей, подписчиков и интересных страниц.
- Черта F (беспечность) положительно связана с количеством видеозаписей.
- Черта I (эмоциональная чувствительность) отрицательно связана с количеством фотоальбомов.
- Черта L (подозрительность) положительно связана с количеством фото.
- Черта М (мечтательность) положительно связана с количеством друзей и групп.
- Черта N (дипломатичность) отрицательно связана с количеством аудиозаписей.
- Черта О (тревожность) отрицательно связана с количеством аудиозаписей, но положительно связана с количеством групп.
- Черта Q3 (самодисциплина) положительно связана с возрастом и отрицательно с количеством фото.

4. Применение методов машинного обучения для оценки выраженности факторов Р. Кеттелла

В данном разделе будет произведена попытка автоматизации оценки выраженности личностных особенностей пользователей. Для этого необходимо построить 16 классификаторов, каждый из которых на вход получает вектор ранее собранных и предобработанных значений атрибутов (возраст, форму обучения, количе-

ство видео, количество друзей, статус образования, город, количество подписок, количество фотографий, год окончания обучения, количество сообществ, количество статей, количество подписчиков, количество аудиозаписей, количество клипов, университет обучения, количество альбомов, пол, количество объектов в блоке «Интересные страницы», количество постов на странице пользователя), а на выход выдает класс — низкую, высокую оценку выраженности рассматриваемого фактора или ее отсутствие (пример приведения оценок представлен в конце раздела 2. Для этого будут использованы следующие алгоритмы машинного обучения, которые были выбраны в результате анализа предметной области:

- метод опорных векторов⁸ (SVM) со стандартными гиперпараметрами;
- случайный лес⁹ (Random Forest) со следующими гиперпараметрами:
 - количество деревьев от 10 до 100 с шагом 15,
 - максимальная глубина от 2 до 15 с шагом 2,
 - минимальное количество экземпляра для конечного узла от 1 до 8 с шагом 2,
 - минимальное количество экземпляров для разделения узла от 2 до 10 с шагом 3;
- алгоритмы градиентного бустинга (XGBoost¹⁰, Catboost¹¹) со следующими гиперпараметрами:
 - количество деревьев от 10 до 200 с шагом 50,
 - максимальная глубина от 2 до 11 с шагом 3,
 - темп обучения: 0,05, 0,3, 0,5,
 - минимальное количество экземпляра для конечного узла от 1 до 8 с шагом 2;
- логистическая регрессия¹² (LogReg) с гиперпараметрами по умолчанию.

Для подбора гиперпараметров алгоритмов машинного обучения использовался метод gridsearch, который позволяет найти наилучшую для заданной метрики комбинацию гиперпараметров (в данной работе гиперпараметры подбирались для метрики F1-micro). Для сравнения

⁸ Метод опорных векторов [электронный ресурс]. URL: <https://scikit-learn.org/stable/modules/svm.html>

⁹ Случайный лес [электронный ресурс]. URL: <https://scikit-learn.org/stable/modules/ensemble.html#random-forests>

¹⁰ XGBoost [электронный ресурс]. URL: <https://xgboost.readthedocs.io/en/stable/>

¹¹ CatBoost [электронный ресурс]. URL: <https://catboost.ai/en/docs/>

¹² Логистическая регрессия [электронный ресурс]. URL: https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

алгоритмов выборка была разбита на обучающую и тестовую части (в соотношении 0,7:0,3) и затем на тестовой части были посчитаны метрики F1-macro, F1-micro и F1-weighted, релевантные для данной задачи [26].

Согласно [25], метрика F1-micro отражает долю верно классифицированных экземпляров по каждому из классов на общее количество классифицированных объектов. Однако при дисбалансе классов в выборке данная метрика может показывать высокий результат, даже если классификатор всегда предсказывает один класс. Для избежания данной ситуации полезна метрика F1-macro. F1-weighted же отражает то, насколько хорошо работает модель, когда в каждом классе разное количество экземпляров (несбалансированные данные), позволяя повышать точность предсказания классов с большим количеством экземпляров.

5. Результаты и выводы

После применяя указанных моделей обучения в разделе 4. к собранному набору данных получились следующие результаты, представленные в Табл. 1.

В колонках таблицы представлены факторы (1 колонка); метрики, по которыми сравнива-

лись результаты алгоритмов машинного обучения (2, 3 и 4 колонка), здесь же в скобках указаны метрики случайного угадывания классов; алгоритмы (5 колонка), которые показали лучшие метрики для каждого рассматриваемого фактора.

По результатам анализа полученных результатов, представленных в таблице, можно сделать следующие выводы:

1. Модели градиентного бустинга показывают результаты лучше, чем другие протестированные модели, что соответствует существующим исследованиям [27].

2. На двух факторах (I, Q4) удалось добиться результата метрики F1-macro равной 0.667 и 0.596 соответственно, что превышает случайное угадывание класса.

3. Благодаря тому, что для оценки результатов классификации использовались три метрики, удалось заметить, что модель показывает большие различия на разных метриках, что может быть вызвано небольшой обучающей выборкой.

4. Для повышения качества классификации также могут быть использованы дополнительные цифровые следы, извлеченные из профилей пользователей (фото, посты, подписки и др.).

Табл. 1. Факторы личности, метрики и название модели, показавшей лучший результат

Фактор	F1-micro	F1-macro	F1-weighted	Лучшая модель
Общительность (A)	0,409 (0,333)	0,475 (0,315)	0,451 (0,347)	CatBoost
Уровень интеллекта (B)	0,407 (0,333)	0,626 (0,306)	0,566 (0,359)	SVM
Эмоциональная устойчивость (C)	0,325 (0,333)	0,465 (0,321)	0,411 (0,341)	CatBoost
Доминантность (E)	0,338 (0,333)	0,495 (0,28)	0,493 (0,382)	XGBoost
Беспечность (F)	0,438 (0,333)	0,626 (0,296)	0,556 (0,365)	LogReg
Моральная нормативность (G)	0,324 (0,333)	0,596 (0,275)	0,544 (0,39)	CatBoost
Смелость (H)	0,353 (0,333)	0,394 (0,317)	0,391 (0,343)	XGBoost
Эмоциональная чувствительность (I)	0,606 (0,333)	0,667 (0,279)	0,662 (0,383)	XGBoost
Подозрительность (L)	0,319 (0,333)	0,434 (0,306)	0,374 (0,355)	CatBoost
Мечтательность (M)	0,27 (0,333)	0,394 (0,311)	0,367 (0,353)	LogReg
Дипломатичность (N)	0,361 (0,333)	0,646 (0,279)	0,586 (0,381)	LogReg
Тревожность (O)	0,421 (0,333)	0,545 (0,309)	0,518 (0,354)	CatBoost
Восприимчивость к новому (Q1)	0,371 (0,333)	0,566 (0,3)	0,529 (0,361)	CatBoost
Самостоятельность (Q2)	0,36 (0,333)	0,545 (0,282)	0,543 (0,384)	LogReg
Самодисциплина (Q3)	0,354 (0,333)	0,606 (0,298)	0,528 (0,365)	LogReg
Напряжённость (Q4)	0,466 (0,333)	0,596 (0,3)	0,58 (0,365)	XGBoost

5. В раннее представленной работе авторов [25], где в качестве цифровых следов использовались текстовые посты пользователей, а для предсказания — нейронные сети и языковые модели, для каждого из 16 факторов были получены более высокие метрики. Однако используемый в работе [25] набор данных не позволяет сравнить метрики между собой, т.к. он был собран на несколько лет раньше рассматриваемого в данной работе набора данных. В будущих исследованиях будет проведена попытка сравнения полученных подходов, а также проведено объединение их результатов.

Заключение

В данной работе была рассмотрена возможность автоматизации предсказания оценки выраженности личностных особенностей человека, оцениваемых с помощью 16-факторного теста Р. Кеттелла и цифровыми следами, оставляемыми пользователем в социальной медиа. Для достижения автоматизированной оценки выраженности личностных особенностей человека были обучены и протестированы различные алгоритмы машинного обучения. Полученный классификатор на двух факторах (I, Q4) предсказывает результат теста Р. Кеттелла на достаточном уровне (F1 макро 0.667 и 0.596 соответственно), по остальным факторам точность предсказания можно попытаться повысить, если использовать более сложные модели машинного обучения и больше цифровых следов со страницы пользователя, например, фотографии и тексты постов.

Литература

1. Kern M.L., Friedman H.S. Personality and Pathways of Influence on Physical Health // *Social and Personality Psychology Compass*. 2011. № 5. P. 76–87. Doi: 10.1111/2Fj.1751-9004.2010.00331.x.
2. Luo J., Zhang B., Estabrook R., Graham E. K., Driver C. C., Schalet B. D., Turiano N. A., Spiro A., Mroczek, D. K. Personality and health: Disentangling their between-person and within-person relationship in three longitudinal studies // *Journal of Personality and Social Psychology*. 2020. Vol. 122. № 3. P. 493–522. Doi: 10.1037/pspp0000399.
3. Luo J., Zhang B., Cao M., Roberts, B.W. The Stressful Personality: A Meta-Analytical Review of the Relation Between Personality and Stress // *Personality and Social Psychology Review*. 2022. №27(2). P. 128–194. Doi: 10.1177/10888683221104002.
4. Ho S., Wong A. The role of customer personality in premium banking services // *Journal of Financial Services Marketing*. 2022. № 28. P. 1–21. Doi: 10.1057/s41264-022-00150-3.
5. Agyei J., Sun S., Abrokwa E., Penney E.K., Ofori-Boafo R. Mobile Banking Adoption: Examining the Role of Personality Traits // *SAGE Open*. 2020. Vol.10. №2. Doi: 10.1177/2158244020932918
6. Netzer O., Lemaire A., Herzenstein M. When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications // *Journal of Marketing Research*. 2019. Vol. 56. №6. P. 960–980. Doi: 10.1177/0022243719852959.
7. Wu W., Chen L., He L. Using personality to adjust diversity in recommender systems // *HT 2013 Proceedings of the 24th ACM Conference on Hypertext and Social Media*. 2013. P. 225–229. Doi: 10.1145/2481492.2481521.
8. Smith, T.A. The role of customer personality in satisfaction, attitude-to-brand and loyalty in mobile services // *Spanish Journal of Marketing - ESIC*. 2020. Vol. 24. №. 2. P. 155–175. Doi: 10.1108/SJME-06-2019-0036.
9. Camoiras-Rodriguez Z., Varela C. The influence of consumer personality traits on mobile shopping intention // *Spanish Journal of Marketing - ESIC*. 2020. Vol. 24. № 3. P. 331–353. Doi: 10.1108/SJME-02-2020-0029.
10. Abramov M.V., Tulupyeu A.L., Suleymanov A.A. Analysis of users' protection from socio-engineering attacks: social graph creation based on information from social network websites // *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2018. Vol. 18. № 2. P. 313–321. Doi: 10.17586/2226-1494-2018-18-2-313-321.
11. Тулупьева Т.В., Абрамов М.В., Тулупьев А.Л. Социально-инженерные атаки: социальные сети и оценки защищенности пользователей. Издательство: Санкт-Петербургский государственный университет аэрокосмического приборостроения. 2018. 266 с. ISBN 978-5-8088-1317-5.
12. Кудинов С. И., Кудинов С. С. Психодиагностика личности. Издательство: Российский университет дружбы народов (РУДН). 2018. 232 с. ISBN: 978-5-209-08188-3
13. Bushmelev F., Khlobystova A., Abramov M., Livshits L. Deep Machine Learning Techniques in the Problem of Estimating the Expression of Psychological Characteristics of a Social Media User // *Artificial Intelligence in Models, Methods and Applications. AIES 2022. Studies in Systems, Decision and Control*. 2023. Vol 457. P. 315–324. Doi: 10.1007/978-3-031-22938-1_22
14. Frolova M.S., Korepanova A.A., Abramov M.V. Assessing the Degree of the Social Media User's Openness Using an Expert Model Based on the Bayesian Network // *2021 XXIV International Conference on Soft Computing and Measurements (SCM)*. 2021. P. 52–55. Doi: 10.1109/SCM52931.2021.9507111.
15. Ignatiev N., Smirnov I., Stankevich M. Predicting Depression with Text, Image, and Profile Data from Social Media // *International Conference on Pattern Recognition Applications and Methods*. 2022. P. 753–760. Doi: 10.5220/0010986100003122
16. Choi T., Sung Y., Lee J., Choi S. Get behind my selfies: The Big Five traits and social networking behaviors through

- selfies // *Personality and Individual Differences*. 2017. Vol. 109. P. 98–101. Doi: 10.1016/j.paid.2016.12.057.
17. 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. Vol. 110. № 15. P. 5802–5805. Doi: 10.1073/pnas.1218772110.
 18. Li L., Li A., Hao B., Guan Z., Zhu T. Predicting Active Users' Personality Based on Micro-Blogging Behaviors // *PLOS ONE*. 2014. Vol. 9. № 1. Doi: 10.1371/journal.pone.0084997.
 19. Райгородский Д.Я. Практическая психодиагностика. Методики и тесты. Учебное пособие. Самара: Издательский дом «БАХРАХ-М». 2001. 672 С. ISBN 5-89570-005-5.
 20. Matz S., Chan Y.W.F., Kosinski M. Models of Personality // *Emotions and Personality in Personalized Services*. 2016. P. 35 - 54. Doi: 10.1007/978-3-319-31413-6_3.
 21. Singh S., Farley S., Donahue J. Grandiosity on display: Social media behaviors and dimensions of narcissism // *Personality and Individual Differences*. 2018. Vol. 134. P. 308–313. Doi: 10.1016/j.paid.2018.06.039.
 22. Chow T., Wan H. Is there any 'Facebook Depression'? Exploring the moderating roles of neuroticism, Facebook social comparison and envy // *Personality and Individual Differences*. 2017. Vol. 119. P. 277–282. Doi: 10.1016/j.paid.2017.07.032.
 23. Soury A., Hosseinpour S., Rahmani A.M. Personality classification based on profiles of social networks' users and the five-factor model of personality. // *Human-centric Computing and Information Science*. 2018. Vol. 24, № 8. Doi: 10.1186/s13673-018-0147-4.
 24. Titov S., Mararitsa L. Full-scale Personality Prediction on VKontakte Social Network and its Applications // 25th Conference of Open Innovations Association (FRUCT). 2019. P. 317–323. Doi: 10.23919/FRUCT48121.2019.8981513.
 25. Олисеенко В.Д., Абрамов М.В. Предсказание результатов 16-факторного теста Р. Кеттелла на основе анализа текстовых постов пользователей социальной сети // *Научно-технический вестник информационных технологий, механики и оптики*. 2023. Т. 23, № 2. С. 279–288. doi: 10.17586/2226-1494-2023-23-2-279-288
 26. Grandini M., Bagli E., Visani G. Metrics for Multi-Class Classification: an Overview. 2020. Doi: 10.48550/arXiv.2008.05756.
 27. Tanha J., Abdi Y., Samadi N. Boosting methods for multi-class imbalanced data classification: an experimental review. *Journal of Big Data*. 2020. Vol. 7. № 1. P. 70. Doi: 10.1186/s40537-020-00349-y.

Рязанцев Григорий Евгеньевич. Санкт-Петербургский государственный университет, Санкт-Петербург, Россия. Бакалавр. Область научных интересов: машинное обучение, анализ социальных сетей. E-mail: st088141@student.spbu.ru

Олисеенко Валерий Дмитриевич. Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, Россия. Младший научный сотрудник. Область научных интересов: обработка естественных языков, глубокое обучение, анализ социальных сетей. E-mail: vdo@dscs.pro

Абрамов Максим Викторович. Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, Россия. Старший научный сотрудник, кандидат технических наук. Область научных интересов: информационная безопасность; социоинженерные атаки; анализ защищенности пользователей. E-mail: mva@dscs.pro

Тулупьева Татьяна Валентиновна. Северо-Западный институт управления РАНХиГС, Санкт-Петербург, Россия. Доцент, старший научный сотрудник, кандидат психологических наук. Область научных интересов: психология личности, психология управления, психодиагностика. E-mail: tvt@dscs.pro

Predicting the Results of the R. Cattell Test Based on the Social Network User Profiles

G. E. Ryazantsev¹, V. D. Oliseenko^{II}, M. V. Abramov^{II}, T. V. Tulupyeva^{III,II}

¹ Saint Petersburg State University, Saint Petersburg, Russia

^{II} St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, Russia

^{III} North-West Institute of Management Branch of the Russian Presidential Academy of National Economy and Public Administration, Saint Petersburg, Russia

Abstract. Digital footprints of users in the social network and the results of passing the 16-factor R. Cattell test. The method consists in applying statistical methods and relevant machine learning algorithms to personal data on the user's page. The main results of the experiment are the identification of a significant correlation between the factors evaluated by the R. Cattell test and digital footprints, and the construction of predictive models. The best results among the machine learning methods for predicting the results of the R. Cattell test were shown by gradient boosting algorithms with the maximum value

of the F1-micro metric of 0.606, which was achieved on the factor “emotional sensitivity” (factor I). The practical significance of the work lies in the development of a tool for automatically predicting the results of the R. Cattell test based on the user's digital footprints. The theoretical significance lies in the development of a method for the automated evaluation of the expression of personality traits of social network users on their digital footprints.

Keywords: machine learning, correlation analysis, regression analysis, social media, psychological test, R. Cattell test, social network psychological profile, profiling.

DOI 10.14357/20718632240106

EDN ESXDTN

References

1. Kern M.L., Friedman H.S. Personality and Pathways of Influence on Physical Health. *Social and Personality Psychology Compass*. 2011; 5(1):76-87. doi: 10.1111/2Fj.1751-9004.2010.00331.x.
2. Luo J., Zhang B., Estabrook R., Graham E. K., Driver C. C., Schalet B. D., Turiano N. A., Spiro A., Mroczek, D. K. Personality and health: Disentangling their between-person and within-person relationship in three longitudinal studies. *Journal of Personality and Social Psychology*. 2020; 122(3): 493-522. doi: 10.1037/pspp0000399.
3. Luo J., Zhang B., Cao M., Roberts, B.W. The Stressful Personality: A Meta-Analytical Review of the Relation Between Personality and Stress. *Personality and Social Psychology Review*. 2022; 27(2):128-194. doi: 10.1177/10888683221104002.
4. Ho S., Wong A. The role of customer personality in premium banking service. *Journal of Financial Services Marketing*. 2023; 28(2):285-305. doi: 10.1057/s41264-022-00150-3.
5. Agyei J., Sun S., Abrokwhah E., Penney E.K., Ofori-Boafo R. Mobile Banking Adoption: Examining the Role of Personality Traits. *SAGE Open*. 2020; 10(2). doi: 10.1177/2158244020932918.
6. Netzer O., Lemaire A., Herzenstein M. When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications. *Journal of Marketing Research*. 2019; 56(6):960-980. doi: 10.1177/0022243719852959.
7. Wu W., Chen L., He L. Using personality to adjust diversity in recommender systems, HT 2013 - Proceedings of the 24th ACM Conference on Hypertext and Social Media, 2013, pp. 225-229. doi: 10.1145/2481492.2481521.
8. Smith, T.A. The role of customer personality in satisfaction, attitude-to-brand and loyalty in mobile services. *Spanish Journal of Marketing - ESIC*. 2020; 24(2):155-175. doi: 10.1108/SJME-06-2019-0036.
9. Camoiras-Rodriguez Z., Varela C. The influence of consumer personality traits on mobile shopping intention. *Spanish Journal of Marketing - ESIC*. 2020; 24(3):331-353. doi: 10.1108/SJME-02-2020-0029.
10. Abramov M., Tulupyeva T., Suleymanov A. Analysis of users' protection from socio-engineering attacks: social graph creation based on information from social network websites. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2018; 18(2):313-321. doi: 10.17586/2226-1494-2018-18-2-313-321.
11. Tulupyeva T., Abramov M., Tulupyeva A. Socioengineering attacks: social networks and user security assessments. *Izdatel'stvo: Sankt-Peterburgskij gosudarstvennyj universitet ajerokosmicheskogo priborostroenija, Saint-Petersburg*. 2018; 266 p. ISBN 978-5-8088-1317-5. (in Russ).
12. Kudinov S. I., Kudinov S. S. Psychodiagnostics of personality. *Izdatel'stvo: Rossijskij universitet družby narodov (RUDN)*. 2018; 232 p. ISBN: 978-5-209-08188-3 (in Russ).
13. Bushmelev F., Khlobystova A., Abramov M., Livshits L. Deep Machine Learning Techniques in the Problem of Estimating the Expression of Psychological Characteristics of a Social Media User, Artificial Intelligence in Models, Methods and Applications. *AIES 2022. Studies in Systems, Decision and Control*. 2023; 457:315-324. doi:10.1007/978-3-031-22938-1_22.
14. Frolova M.S., Korepanova A.A., Abramov M.V. Assessing the Degree of the Social Media User's Openness Using an Expert Model Based on the Bayesian Network. *2021 XXIV International Conference on Soft Computing and Measurements (SCM)*. 2021; p. 52-55. doi: 10.1109/SCM52931.2021.9507111.
15. Ignatiev N., Smirnov I., Stankevich M. Predicting Depression with Text, Image, and Profile Data from Social Media. *International Conference on Pattern Recognition Applications and Methods*. 2022; p. 753-760. doi: 10.5220/0010986100003122.
16. Choi T., Sung Y., Lee J., Choi S. Get behind my selfies: The Big Five traits and social networking behaviors through selfies. *Personality and Individual Differences*. 2017; 109:98-101. doi: 10.1016/j.paid.2016.12.057.
17. 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; 110(15):5802-5805. doi: 10.1073/pnas.1218772110.
18. Li L., Li A., Hao B., Guan Z., Zhu T. Predicting Active Users' Personality Based on Micro-Blogging Behaviors. *PLOS ONE*. 2014; 9(1):1-11. doi: 10.1371/journal.pone.0084997.
19. Raygorodsky D.Y. Practical psychodiagnostics. Methods and tests. Tutorial, Samara: Izdatel'skij dom «BAHRAH-M». 2001; p.192-239, ISBN 5-89570-005-5 (in Russ).
20. Matz S., Chan Y.W.F., Kosinski M. Models of Personality. Emotions and Personality in Personalized Services. 2016; p. 35-54. doi: 10.1007/978-3-319-31413-6_3.
21. Singh S., Farley S., Donahue J. Grandiosity on display: Social media behaviors and dimensions of narcissism. *Personality and Individual Differences*. 2018; 134:308-313. doi: 10.1016/j.paid.2018.06.039.

22. Chow T., Wan H. Is there any ‘Facebook Depression’? Exploring the moderating roles of neuroticism, Facebook social comparison and envy. *Personality and Individual Differences*. 2017; 119:277-282. doi: 10.1016/j.paid.2017.07.032.
23. Sourì A., Hosseinpour S., Rahmani A.M. Personality classification based on profiles of social networks’ users and the five-factor model of personality. *Human-centric Computing and Information Science*. 2018; 8(1). doi: 10.1186/s13673-018-0147-4.
24. Titov S., Mararitsa L. Full-scale Personality Prediction on VKontakte Social Network and its Applications, 25th Conference of Open Innovations Association (FRUCT). 2019; p. 317-323. doi: 10.23919/FRUCT48121.2019.8981513.
25. Oliseenko V.D., Abramov M.V. Predicting the results of the 16-factor R. Cattell test based on the analysis of text posts of social network users. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023; 23(2):279–288. doi: 10.17586/2226-1494-2023-23-2-279-288.
26. Grandini M., Bagli E., Visani G. Metrics for Multi-Class Classification: an Overview, 2020. doi: 10.48550/arXiv.2008.05756.
27. Tanha J., Abdi Y., Samadi N. Boosting methods for multi-class imbalanced data classification: an experimental review. *Journal of Big Data*. 2020; 7(1). doi: 10.1186/s40537-020-00349-y.

Ryazantsev Grigori E. Student, St. Petersburg State University, 7-9 Universitetskaya Embankment, St Petersburg, Russia, 199034. E-mail: st088141@student.spbu.ru

Oliseenko Valerii D. Junior researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences. 39, 14th Line V.O., St. Petersburg, 199178, Russia. E-mail: vdo@dscs.pro

Abramov Maxim V. PhD, Senior researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences. 39, 14th Line V.O., St. Petersburg, 199178, Russia. E-mail: mva@dscs.pro

Tulupyeva Tatiana V. PhD, Assistant professor, Northwestern Institute of Management RANEPA, 57/43, Sredny prospect VO, St. Petersburg, 199178, Russia. Senior researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences. 39, 14th Line V.O., St. Petersburg, 199178, Russia. E-mail: tvt@dscs.pro