

УДК 004.056

doi:10.31799/1684-8853-2022-1-8-18

## Определение психического состояния пользователей социальной сети Reddit на основе методов машинного обучения

А. А. Браницкий<sup>а,б</sup>, канд. техн. наук, старший научный сотрудник, [orcid.org/0000-0003-3104-0622](https://orcid.org/0000-0003-3104-0622), [brantskiy@comsec.spb.ru](mailto:brantskiy@comsec.spb.ru)

Я. Д. Шарма<sup>в</sup>, студент, [orcid.org/0000-0003-2491-0167](https://orcid.org/0000-0003-2491-0167)

И. В. Котенко<sup>а,б</sup>, доктор техн. наук, профессор, [orcid.org/0000-0001-6859-7120](https://orcid.org/0000-0001-6859-7120)

Е. В. Федорченко<sup>а,б</sup>, канд. техн. наук, старший научный сотрудник, [orcid.org/0000-0001-6707-9153](https://orcid.org/0000-0001-6707-9153)

А. В. Красов<sup>б</sup>, канд. техн. наук, доцент, [orcid.org/0000-0002-9076-6055](https://orcid.org/0000-0002-9076-6055)

И. А. Ушаков<sup>б</sup>, канд. техн. наук, доцент, [orcid.org/0000-0002-6988-9261](https://orcid.org/0000-0002-6988-9261)

<sup>а</sup>Санкт-Петербургский Федеральный исследовательский центр РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ

<sup>б</sup>Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича, Большевиков пр., 22-1, Санкт-Петербург, 193232, РФ

<sup>в</sup>Санкт-Петербургский государственный электротехнический университет «ЛЭТИ», Профессора Попова ул., 5, Санкт-Петербург, 197376, РФ

**Введение:** диагностирование психических заболеваний представляет собой сложный процесс, который включает проведение диалоговых бесед, анализ поведения обследуемого и прохождение им специализированных тестов. На успешное решение данной задачи может влиять как отсутствие знаний и опыта психолога, так и наличие противоречивых или неполных исходных данных со стороны пациента. Для устранения последнего недостатка разрабатываются экспертные или интеллектуальные системы. **Цель:** разработать методику определения психического состояния пользователей социальной сети. **Результаты:** с помощью методов машинного обучения разработана методика, предназначенная для определения типа психического состояния пользователей социальной сети. Новизна предлагаемой методики заключается в наличии двухшаговой процедуры предварительной обработки текста и построении нескольких наборов признаков, описывающих эмоциональное настроение пользователей социальной сети на уровне публикуемых ими сообщений. В качестве исходных данных привлекались текстовые сообщения пользователей социальной сети Reddit. В методике выделяются три этапа: 1) сбор данных, 2) предварительная обработка данных, 3) разметка постов и построение признаков. Оценка функционирования программного средства, построенного на основе данной методики, проводилась по четырем показателям: достоверность, точность, полнота и F-мера. Наилучшие результаты демонстрирует ансамбль, построенный на основе подхода One-vs-Rest, где в качестве базовых решателей выступают линейные машины опорных векторов. **Практическая значимость:** результаты исследования могут применяться при построении вспомогательных систем, которые направлены на поддержку принятия решений специалистами-психологами при определении психических нарушений.

**Ключевые слова** – машинное обучение, социальная сеть, психические нарушения, эмоциональное настроение, машина опорных векторов, сверточная нейронная сеть.

Для цитирования: Браницкий А. А., Шарма Я. Д., Котенко И. В., Федорченко Е. В., Красов А. В., Ушаков И. А. Определение психического состояния пользователей социальной сети Reddit на основе методов машинного обучения. *Информационно-управляющие системы*, 2022, № 1, с. 8–18. doi:10.31799/1684-8853-2022-1-8-18

For citation: Branitskiy A. A., Sharma Y. D., Kotenko I. V., Fedorchenko E. V., Krasov A. V., Ushakov I. A. Determination of the mental state of users of the social network Reddit based on machine learning methods. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 1, pp. 8–18 (In Russian). doi:10.31799/1684-8853-2022-1-8-18

### Введение

Современные интернет-платформы, такие как социальные сети, предоставляют своим пользователям множество функций, которые позволяют им обсуждать интересующие их темы, делиться друг с другом графической и текстовой информацией, выражать собственные мнения и эмоции. Благодаря наличию таких открытых социальных сетей, как Facebook, Twitter, Instagram, Snapchat, Reddit, люди стали часто использовать

их для создания сообществ и обсуждения общих вопросов. Рост популярности таких сетевых сервисов приводит к генерации огромного объема данных, которые включают анкетные данные пользователей, содержимое постов и сообщений, комментарии, отметки о количестве просмотров, загруженные аудио- и графические файлы. С другой стороны, в социальных сетях могут распространяться ложные новости, присутствовать пропаганда нездорового образа жизни и вредных привычек, содержаться призывы к выполнению

противозаконных действий. Наличие такой информации негативно влияет на поведение пользователей и приводит к нарушениям в их психическом здоровье.

Для оценки психического состояния исследуемого человека специалисты используют текстовую информацию, полученную в результате диалога с ним. Результаты обследования являются конфиденциальными, в связи с чем ими нельзя воспользоваться при проведении открытых исследований. В то же время такая социальная сеть, как Reddit, предоставляет возможность построения и организации сообществ (subreddits), внутри которых ведется открытое обсуждение вопросов, к примеру, связанных с настроением или психическим здоровьем участников определенной группы. Кроме того, извлекаемая из постов социальной сети информация может отражать эмоциональное состояние автора (на основе наличия определенных ключевых слов и последовательностей знаков препинания). Поэтому социальную сеть можно рассматривать в качестве платформы, подходящей для сбора и анализа информации о психическом состоянии ее пользователей.

Правильность и своевременность определения типа и степени психического расстройства позволяет сформировать корректный план лечения для пациента и вовремя предотвратить у него развитие возможных болезней. Успех решения этой задачи во многом зависит как от степени квалификации психолога, так и от уровня открытости обследуемого пациента. В то же время в связи с многообразием и популярностью интернет-ресурсов (форумов, каналов в мессенджерах, сообществ в социальных сетях) пользователям становится проще излагать свои мысли, выражать эмоции и консультироваться со специалистами в анонимном формате, не прибегая к реальному диалогу. В связи с этим актуальной является задача, заключающаяся в разработке методик и информационных систем, предназначенных для анализа открыто публикуемых в социальных сетях постов и направленных на обнаружение отклонений в психическом состоянии их авторов.

### Анализ релевантных работ

Выявление нарушений (депрессии, тревоги, стресса) в психическом состоянии человека является важной задачей, решению которой посвящено множество работ. С учетом массовой доступности Интернета социальные сети могут рассматриваться как платформа для сбора информации об определении психического состояния их пользователей. К примеру, в работе [1] выполнялось обнаружение стрессового состояния у пользователей социальной сети Twitter. Предложенный

подход был направлен на определение текущего статуса и возможных изменений в поведении подростков, находящихся под психологическим давлением. Аналогичные задача и социальная сеть были рассмотрены в [2], где авторы добавили статистические атрибуты, которые были получены в результате обработки сообщений, опубликованных в течение недели. Отмечается, что использование расширенного таким образом набора признаков позволяет анализировать совокупность событий, возможно, послуживших причиной нарушения психического состояния у адресанта сообщения.

Состояние депрессии является распространенной формой нарушения психического здоровья, обнаружению которого посвящено исследование [3]. Авторы этой статьи отмечают возможную корреляцию между депрессивным состоянием пользователя и наличием эмоциональных слов в его сообщениях. В [4] обнаружение депрессивного состояния пользователей Twitter выполнялось на основе анализа временных рядов, представляющих собой последовательность параметров активности каждого пользователя в течение года. Отмечается, что у людей, страдающих депрессивными расстройствами, наблюдается низкая социальная активность, преобладание негативных эмоций и завышенное внимание к своей личности. В [5] при решении аналогичной задачи был определен минимальный период (длительностью два месяца), в течение которого построенный набор статистических признаков, описывающих активность пользователей Twitter, доставляет наибольшую точность (69 %) достоверного определения депрессивного состояния.

В отличие от Twitter, в социальной сети Facebook допустимый размер публикуемых постов практически не ограничен, что позволяет извлечь больший объем анализируемой информации из каждого поста. В [6] данная социальная сеть послужила источником для проверки гипотезы о сезонности проявления депрессии. Степень тяжести этого состояния может колебаться в зависимости как от перемен в личной жизни, так и от факторов окружающей среды. Для количественного прогнозирования этой величины авторы разработали регрессионную модель, при обучении которой использовались результаты опроса пользователей и данные, касающиеся обновления статуса в их профилях. Согласно построенной модели уровень прогнозируемого депрессивного состояния у пользователей зимой оказывался выше, чем летом.

Кроме текстовой информации, фотографии, размещенные в профиле социальной сети Facebook, также могут быть информативными при определении особенностей характера человека [7]. Было отмечено, что существует прямая зависимость между

наличием депрессии в поведении человека и преобладанием темных тонов и замкнутых пространств на его фотографии. Похожая задача, связанная с выявлением психических заболеваний на основе анализа изображений, решается в [8], однако в ней рассматривается другая социальная сеть, а именно Instagram. В результате проведения исследования было выявлено наличие корреляции между цветовыми схемами в изображении и депрессивным состоянием ее автора. В данном исследовании извлекаемые признаки были разбиты по двум типам. Для построения признаков первого типа использовались показатели, связанные с активностью пользователя и его подписчиков (количество комментариев и отметок «нравится» для каждого поста). Признаки второго типа подразумевали детальный анализ изображений (количество человеческих лиц на изображении; средние значения показателей пикселей: оттенка, насыщенности и яркости; наличие Instagram-фильтра на изображении). Явное задание набора вычисляемых признаков, как это показано в [8], подразумевает извлечение из изображений таких индикаторов, которые кажутся с экспертной точки зрения репрезентативными для выявления депрессии. С другой стороны, существуют исследования, в которых процесс формирования этих признаков перекладывается непосредственно на модель машинного обучения. Так, в [9] исследовалась применимость двух типов глубоких нейронных сетей для анализа сообщений из социальной сети Reddit и их классификации в соответствии с 11 типами психических заболеваний. Для увеличения точности прогнозирования таких заболеваний в [10] предлагается использовать несколько бинарных классификаторов, каждый из которых предназначен для выявления наличия только одного психического расстройства и построен на основе XGBoost или сверточной нейронной сети. Другой вариант повышения качества анализа постов в социальных сетях может включать использование современных моделей обработки естественного языка, например BERT [11]. Применение этой модели и ее модификации в виде RoBERTa [12] рассматривается в [13], где отмечается их превосходство над сетью с долгой краткосрочной памятью. При этом проведенные для социальной сети Reddit эксперименты выполнялись таким образом, что в качестве входных данных анализировались как посты и заголовки по отдельности, так и их объединение.

Дополнительный прирост производительности классификаторов возможен за счет построения разнородного списка признаков, покрывающих анализ комментариев, изображений и профиля на странице пользователя социальной сети. В [14] с использованием такого списка признаков исследовалась применимость нескольких типов нейронных сетей. В рамках задачи прогнозиро-

вания подверженности пользователей социальных сетей деструктивным воздействиям наилучшие результаты были достигнуты при помощи нейронной сети с тремя скрытыми слоями.

В отличие от представленных работ, разработанная методика отличается наличием двухшаговой процедуры предварительной обработки текста, а также возможностью построения нескольких наборов признаков, описывающих эмоциональное настроение пользователей социальной сети на уровне публикуемых ими сообщений.

### Методика определения психического состояния пользователей социальной сети Reddit

Разработанная методика определения психического состояния пользователей социальной сети Reddit включает три этапа. Первый этап — сбор данных. С этой целью использовался общедоступный интерфейс прикладного программирования API Pushshift Reddit Dataset [15]. В Reddit пользователи объединяются в сообщества со схожими интересами и убеждениями. В рамках одного сообщества его участники могут отправлять друг другу сообщения, комментировать сообщения других людей и голосовать за или против. После сбора данных из социальной сети необходимо провести их предварительную обработку, выполняющуюся на втором этапе методики. Для этого используется двухшаговая процедура, включающая фильтрацию сообщений и обработку их содержимого. Наконец, на третьем этапе выполняется разметка постов и извлечение признаков. При присвоении постам меток психических расстройств их авторов использовались следующие шесть классов: депрессия, тревога, членовредительство, стресс, гнев и норма. В качестве классификаторов использовались следующие два типа моделей машинного обучения:

1) линейные классификаторы: линейная машина опорных векторов (МОВ) и ансамбли, использующие в качестве базовых классификаторов МОВ и построенные на основе одной из двух стратегий комбинирования One-vs-One (OvO) или One-vs-Rest (OvR);

2) текстовые классификаторы: fastText и сверточная нейронная сеть (СНС).

### Сбор данных

Каждая запись в загруженном наборе данных содержит семь полей, включая автора и заголовок сообщения, а также время его создания (табл. 1). Наиболее информативным среди входных данных является поле, обозначенное Selftext.

■ **Таблица 1.** Список полей с их описанием и типом данных в наборе данных

■ **Table 1.** List of fields with their description and data type within the dataset

| Поле        | Описание   | Тип данных  |
|-------------|--|-------------|
| Author      | Имя пользователя, который опубликовал запись   | Строка      |
| Created_utc | Метка времени публикации записи (в формате POSIX-времени)  | Целое число |
| Title       | Заголовок публикации   | Строка      |
| Selftext    | Содержимое публикации  | Строка      |
| Score       | Рейтинг публикации, вычисляемый как разность количества голосов «за» и количества голосов «против» | Целое число |
| URL         | URL-адрес публикации   | Строка      |
| Subreddit   | Название сообщества, в котором размещена публикация  | Строка      |

В Reddit нет ограничений на размер публикуемых сообщений, поэтому пользователи могут свободно выражать свои мысли и идеи, используя неограниченное количество слов. В зависимости от состояния психического здоровья участников можно выделить несколько сообществ. В настоящем исследовании рассмотрены такие классы психических нарушений, как депрессия, тревога, членовредительство, гнев и стресс. В табл. 2 для каждого из этих классов и класса «норма» приводится краткое описание соответствующих им сообществ, а также перечисляются сведения о количестве участников и сообщений. В результате анализа собранных данных было выявлено, что наибольшая доля постов относится к классу «депрессия». Также для этого класса характерно наибольшее количество участников. Рассматриваемые данные были собраны в период с января 2018 года по апрель 2021 года.

### Предварительная обработка данных

Этап предобработки данных заключается в приведении исходных данных к определенному формату и удалении избыточных данных. Данный этап выполняется в два шага.

■ **Таблица 2.** Reddit-сообщества, связанные с вопросами психического здоровья

■ **Table 2.** Reddit-communities related with questions of mental health

| Сообщество и класс                 | Описание сообщества   | Количество участников | Количество постов |
|------------------------------------|---|-----------------------|-------------------|
| r/depression, депрессия            | Поддерживающее сообщество, которое помогает любому, кто борется с депрессией, и предоставляет открытое пространство для разговоров и обсуждений | 757 тыс.              | 1,2 млн           |
| r/depressed, депрессия             | Сообщество людей, страдающих депрессией или находящихся в депрессивном состоянии  | 74 тыс.               | 30 тыс.           |
| r/Anxiety, тревога                 | Сообщество, в котором проводятся обсуждения, связанные с тревожными расстройствами  | 455 тыс.              | 410 тыс.          |
| r/Anxietyhelp, тревога             | Сообщество, в котором участники делятся статьями, видео- и текстовыми сообщениями в блогах, чтобы справляться с тревогой                        | 91 тыс.               | 24 тыс.           |
| r/selfharm, членовредительство     | Сообщество, в котором участники обсуждают членовредительство и его аспекты  | 71 тыс.               | 141 тыс.          |
| r/SuicideWatch, членовредительство | Сообщество, которое поддерживает людей, размышляющих о самоубийстве, или лиц, подверженных риску самоубийства                                   | 274 тыс.              | 580 тыс.          |
| r/Anger, гнев                      | Сообщество, в котором участники обсуждают проблемы гнева и способы борьбы с ним   | 27 тыс.               | 14 тыс.           |
| r/Stress, стресс                   | Сообщество, обсуждающее причины стресса и методы управления стрессом  | 11 тыс.               | 7,5 тыс.          |
| r/philosophy, норма                | Сообщество, в котором участники обсуждают философские вопросы   | 15,7 млн              | 174 тыс.          |
| r/AskReddit, норма                 | Сообщество, в котором участники задают наводящие на размышления вопросы с целью получить на них ответы  | 32,3 млн              | 27 млн            |

**Шаг 1.**

1. Удаляются сообщения, авторы которых исключились из социальной сети.
2. Удаляются сообщения, у которых часть содержимого (Selftext) недоступна.
3. Удаляются сообщения, имеющие схожее содержимое, но опубликованные в разных сообществах.
4. Удаляются сообщения наименее активных пользователей, имеющих менее 50 опубликованных постов.
5. Удаляются сообщения пользователей, являющихся возможными ботами и имеющих более 5000 постов.

**Шаг 2.**

1. Заголовок и содержимое поста объединяются в единое сообщение.
  2. Удаляются все ссылки и URL-адреса, символы новой строки и табуляции заменяются символом пробела.
  3. Сленговые выражения заменяются их полными эквивалентными формами, например, idk — I do not know, ur — your, fam — family.
  4. Сокращения заменяются их полными аналогами, например, I'll — I will, let's — let us, couldn't've — could not have.
  5. Удаляются все специальные символы из текста.
  6. Удаляется запятая из чисел.
  7. Все числа в тексте нормализуются путем их замены на текстовую форму этого числа.
  8. Эмодзи-символы заменяются на их текстовые обозначения.
  9. Исправляются опечатки при помощи python-пакета wordninja.
- Статистические сведения об экспериментальном наборе данных после этапа предобработки приведены в табл. 3.

■ **Таблица 3.** Статистические сведения об экспериментальном наборе данных

■ **Table 3.** Statistical data about experimental dataset

| Класс              | Количество постов | Среднее число слов в посте | Среднее число символов в посте |
|--------------------|-------------------|----------------------------|--------------------------------|
| Депрессия          | 40 003            | 127.075                    | 618.99                         |
| Тревога            | 10 164            | 138.257                    | 687.271                        |
| Членовредительство | 25 734            | 114.211                    | 550.039                        |
| Гнев               | 9260              | 210.887                    | 1040.01                        |
| Стресс             | 3954              | 171.799                    | 867.976                        |
| Норма              | 40 633            | 47.9625                    | 254.905                        |

**Разметка постов и построение признаков**

Разметка данных по принадлежности к классам депрессии и тревоги выполнялась при помощи поиска самовыражений [16], указывающих на соответствующее психическое нарушение. Например, выражение «I (was|am) diagnosed with (depression|anxiety)», в котором присутствует местоимение первого лица, может свидетельствовать о наличии такого нарушения. Для классов «членовредительство», «стресс» и «гнев» разметка постов выполнялась аналогичным образом, но поиск осуществлялся только с использованием ключевых слов. Записям, не связанным с психическими нарушениями, была присвоена метка класса «норма».

Процесс построения вектора признаков подразумевает вычисление таких показателей, которые по отдельности или в совокупности будут находиться в корреляционной зависимости от прогнозируемой метки класса. Поэтому нахождение таких показателей требует от исследователя определенных знаний предметной области. В данном исследовании были выделены представленные ниже признаки.

1. Количество эмодзи-символов в публикации (emojiCount). Вычисление этого признака выполнялось перед предварительной обработкой текста. На рис. 1 показано распределение величины emojiCount в зависимости от метки класса.

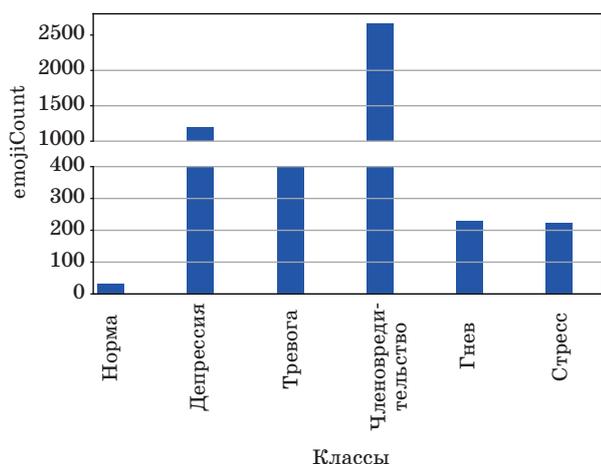
2. Количество местоимений первого и второго лица (firstPropnCount, secondPropnCount). Согласно [17] посты пользователей с психическими нарушениями характеризуются большим количеством личных местоимений. Для подтверждения этой гипотезы была выполнена оценка среднего количества местоимений в сообщении с разбивкой по классам (рис. 2).

3. Количество специальных наречий частотности, таких как абсолютно, постоянно, всегда, никогда, целиком, полностью (absWordCount). В соответствии с [18] пользователи с психическими нарушениями при выражении своих мыслей используют больше абсолютных слов по сравнению со здоровыми людьми. На рис. 3 показана зависимость среднего количества абсолютных слов в сообщении от метки класса.

4. Количество положительных (www.enchantedlearning.com/wordlist/positivewords.shtml) и отрицательных (www.enchantedlearning.com/wordlist/negativewords.shtml) слов (posWordCount, negWordCount). На рис. 4 показана зависимость среднего количества таких слов в сообщении от метки класса.

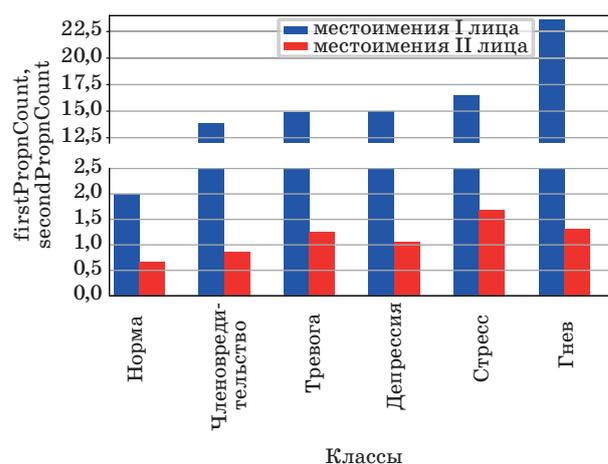
5. Общее количество слов в сообщении (wordCount).

6. Сентимент-оценка сообщения (afinnScore) [19].



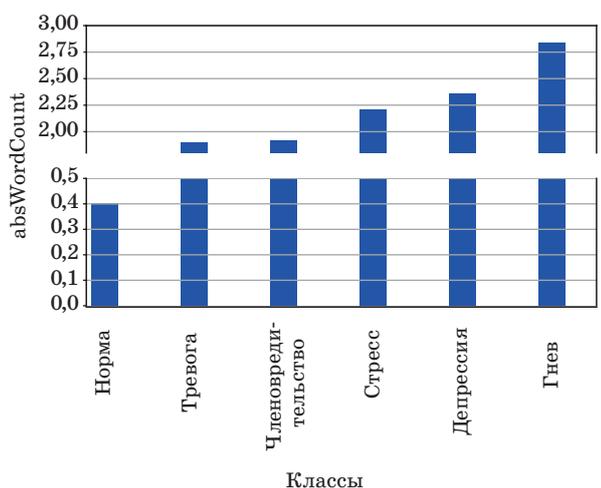
■ **Рис. 1.** Распределение величины emojiCount в зависимости от метки класса

■ **Fig. 1.** Distribution of emojiCount value in dependence of class label



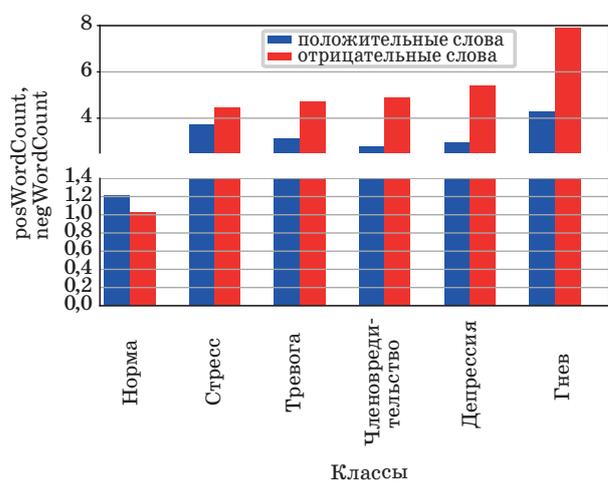
■ **Рис. 2.** Зависимость среднего количества местоимений в сообщении от метки класса

■ **Fig. 2.** Dependence of average quantity of pronouns within message from class label



■ **Рис. 3.** Зависимость среднего количества абсолютных слов в сообщении от метки класса

■ **Fig. 3.** Dependence of average quantity of absolute words within message from class label



■ **Рис. 4.** Зависимость среднего количества положительных и отрицательных слов в сообщении от метки класса

■ **Fig. 4.** Dependence of average quantity of positive and negative words within message from class label

7. Признаки, построенные на основе мешка слов (bowVec).

8. Признаки, построенные на основе TF-IDF (tfidfVec).

### Эксперименты

При проведении экспериментов использовались несколько наборов признаков (НП):

1) I НП — семь признаков (emojiCount, firstPropnCount, secondPropnCount, absWordCount, negWordCount, posWordCount, afinnScore);

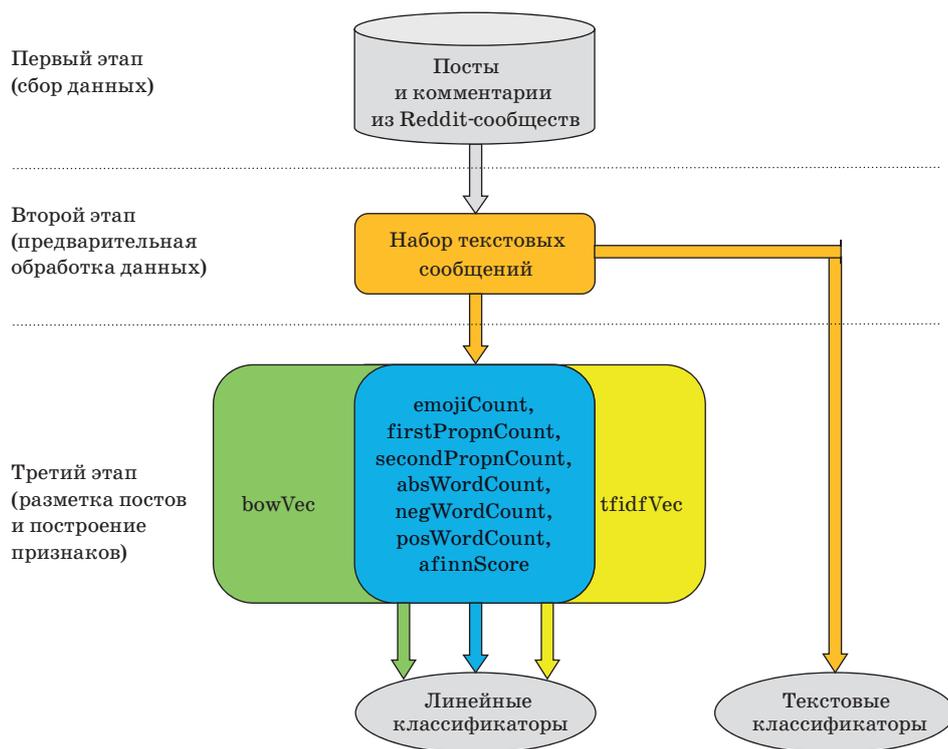
2) II НП — восемь признаков (emojiCount, firstPropnCount, secondPropnCount, absWord-

Count, negWordCount, posWordCount, afinnScore, bowVec);

3) III НП — восемь признаков (emojiCount, firstPropnCount, secondPropnCount, absWordCount, negWordCount, posWordCount, afinnScore, tfidfVec);

4) IV НП — текстовые сообщения.

Схема проведения экспериментов согласно разработанной методике представлена на рис. 5. Обучение линейных классификаторов выполнялось с использованием I, II и III НП, обучение текстовых классификаторов — с использованием только IV НП. Для нахождения оптимального набора гиперпараметров у указанных классификаторов применялся python-модуль GridSearchCV.



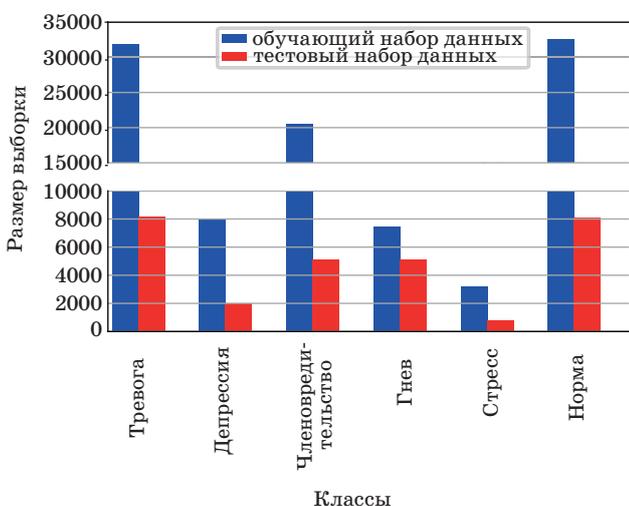
■ *Рис. 5.* Схема проведения эксперимента  
 ■ *Fig. 5.* Experiment scheme

Для линейных классификаторов наилучшие результаты были достигнуты при использовании III НП.

Разбиение набора данных на обучающую и тестовую выборки выполнялось в отношении 4:1 (рис. 6). При проведении экспериментов использовалась пятиблочная перекрестная проверка [20], что позволило оценить показатели эффек-

тивности классификаторов на нескольких дизъюнктных выборках из исходного набора данных. Полученные для каждого классификатора результаты в виде матриц неточностей представлены на рис. 7, а-д.

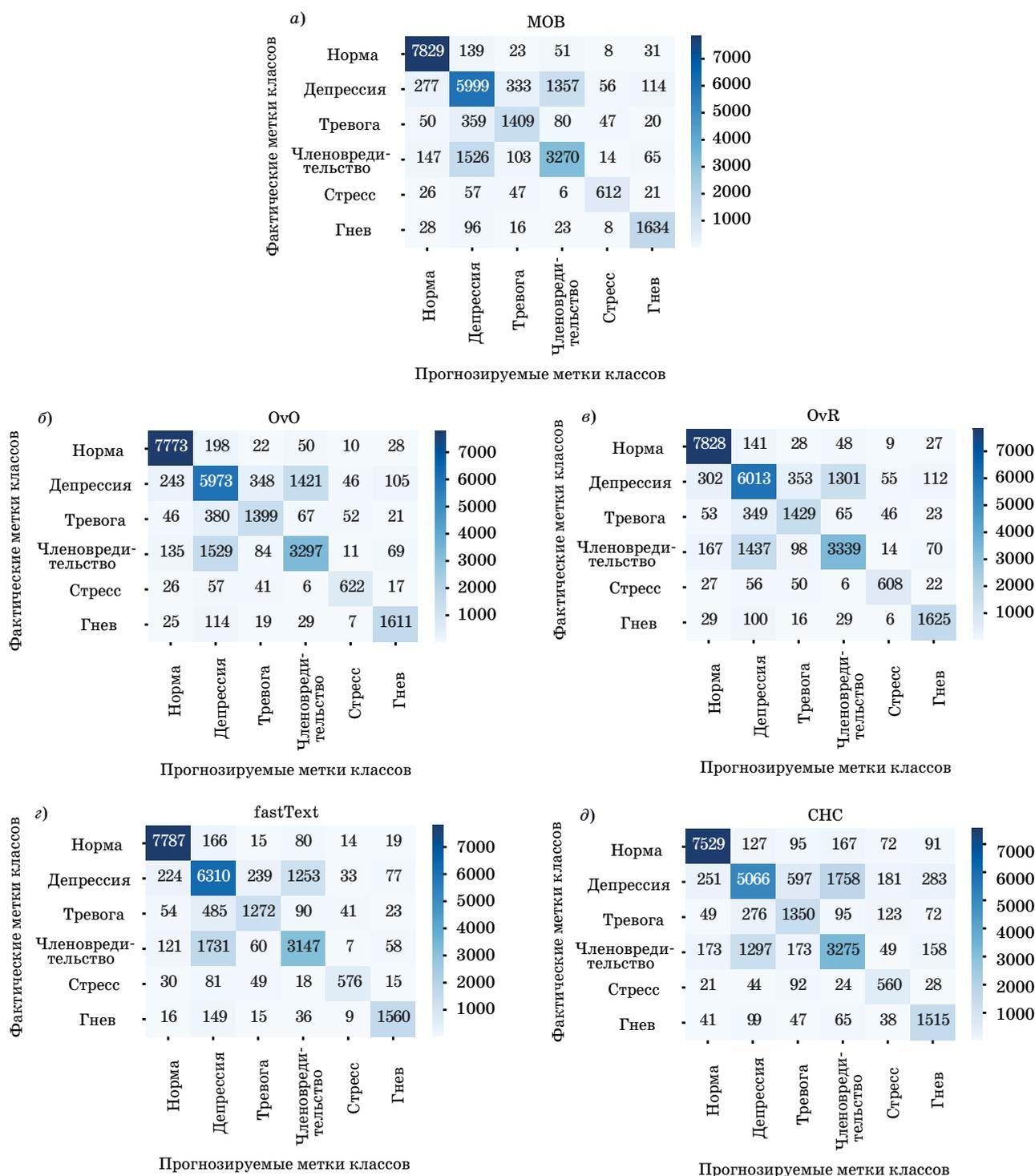
Обобщение полученных результатов представлено в виде табл. 4, которая содержит значения



■ *Рис. 6.* Обучающий и тестовый наборы данных  
 ■ *Fig. 6.* Training and testing datasets

■ *Таблица 4.* Показатели эффективности пяти классификаторов  
 ■ *Table 4.* Effectiveness indicators of five classifiers

| Классификатор | Достоверность, % | Точность, % | Полнота, %   | F-мера, %    |
|---------------|------------------|-------------|--------------|--------------|
| MOB           | 80,19            | 79,53       | 79,37        | 79,42        |
| OvO           | 79,88            | 79,61       | 79,21        | 79,39        |
| OvR           | <b>80,53</b>     | 79,73       | <b>79,62</b> | <b>79,65</b> |
| fastText      | 79,86            | <b>80,7</b> | 77,06        | 78,68        |
| CHC           | 74,55            | 68,36       | 74,13        | 70,77        |



■ **Рис. 7.** Матрица неточностей для MOB (а); OvO (б); OvR (в); fastText (з); CNN (д)

■ **Fig. 7.** Confusion matrix for SVM (a); OvO (б); OvR (в); fastText (з); convolutional neural network (д)

достоверности (accuracy) и усредненные по шести классам значения точности (precision), полноты (recall) и F-меры (F-measure), вычисленные для выбранных пяти классификаторов.

Наилучшие результаты в терминах достоверности демонстрирует классификатор OvR.

Наибольшее значение показателя точности принадлежит классификатору fastText. В то же время максимальное значение показателя F-меры, являющегося интегральной характеристикой точности и полноты, достигается при помощи классификатора OvR. Полученные результаты

подтверждают достижение поставленной цели и доказывают, что разработанный в данном исследовании список признаков позволяет повысить показатели эффективности классификаторов, обученных для определения психических нарушений на уровне текстовых сообщений, по сравнению с признаками, автоматически формируемыми в случае классификаторов fastText и СНС.

## Заключение

В статье описана разработанная методика для определения психического состояния пользователей социальной сети Reddit. В качестве исходных данных рассмотрены текстовые сообщения, публикуемые в качестве постов и комментариев в рамках сообществ данной социальной сети. Для повышения показателей эффективности исследуемых классификаторов машинного обучения предложена двухшаговая процедура предварительной обработки текста и построено несколько наборов признаков, учитывающих эмо-

циональное настроение пользователей социальной сети на уровне публикуемых ими сообщений. Экспериментальная проверка методики осуществлялась через пятиблочную перекрестную проверку, в результате применения которой наилучшие результаты при определении психических нарушений достигнуты при помощи комбинированного классификатора, построенного на основе подхода One-vs-Rest, где в качестве базовых решателей выступают линейные машины опорных векторов. Как направление дальнейших исследований стоит отметить расширение набора данных за счет сбора данных из других социальных сетей и анализа изображений, а также построение ансамблей, сочетающих разнородные классификаторы.

## Финансовая поддержка

Работа выполнена при частичной финансовой поддержке проекта РФФИ 18-29-22034 мк и бюджетной темы 0073-2019-0002.

## Литература

- Xue Y., Li Q., Jin L., Feng L., Clifton D. A., Clifford G. D. *Detecting Adolescent Psychological Pressures from Micro-Blog*. Health Information Science. Y. Zhang, G. Yao, J. He, L. Wang, N. R. Smalheiser, X. Yin Eds. Lecture Notes in Computer Science, 2014. Vol. 8423. Pp. 83–94.
- Lin H., Jia J., Guo Q., Xue Y., Li Q., Huang J., Cai L., Feng L. User-level psychological stress detection from social media using deep neural network. *Proc. 22nd ACM International Conference on Multimedia*, 2014, pp. 507–516.
- Park M., Chiyoung C., Meeyoung C. Depressive moods of users portrayed in Twitter. *Proc. HI-KDD*, Beijing, China, ACM, 2012. [https://nyuscholars.nyu.edu/ws/files/134720119/depressive\\_moods\\_kdd.pdf](https://nyuscholars.nyu.edu/ws/files/134720119/depressive_moods_kdd.pdf) (дата обращения: 05.08.2021).
- De Choudhury M., Gamon M., Counts S., Horvitz E. Predicting depression via social media. *Proc. of the International AAAI Conference on Web and Social Media*, 2013, AAAI, vol. 7, no. 1, pp. 128–137.
- Tsugawa Sh., Kikuchi Y., Kishino F., Nakajima K., Itoh Y., Ohsaki H. Recognizing depression from twitter activity. *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 3187–3196. doi:10.1145/2702123.2702280
- Schwartz H., Eichstaedt J., Kern M., Park G., Sap M., Stillwell D., Kosinski M., Ungar L. Towards assessing changes in degree of depression through Facebook. *Proc. the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2014, pp. 118–125. doi:10.3115/v1/W14-3214
- Segalin C., Celli F., Polonio L., Kosinski M., Stillwell D., Sebe N., Cristani M., Lepri B. What your Facebook profile picture reveals about your personality? *Proc. the 25th ACM International Conference on Multimedia*, 2017, pp. 460–468. doi: 10.1145/3123266.3123331
- Reece A. G., Christopher M. D. Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 2017, no. 6, pp. 1–12.
- Gkotsis G., Oellrich A., Velupillai S., Liakata M., Hubbard T. J. P., Dobson R. J. B., Dutta R. Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific Reports*, 2017, vol. 7, pp. 1–11.
- Kim J., Lee J., Park E., Han J. A deep learning model for detecting mental illness from user content on social media. *Scientific Reports*, 2020, vol. 10, pp. 1–6. doi:10.1038/s41598-020-68764-y
- Devlin J., Chang M.-W., Lee K., Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proc. NAACL-HLT*, Minneapolis, Minnesota, 2019, pp. 4171–4186.
- Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L., Stoyanov V. *Roberta: A robustly optimized BERT pretraining approach*. Preprint, 2019. <https://arxiv.org/pdf/1907.11692.pdf> (дата обращения: 05.08.2021).
- Murarka A., Balaji R., Sushma R. *Detection and classification of mental illnesses on social media using RoBERTa*. Preprint, 2020. <https://arxiv.org/pdf/2011.11226.pdf> (дата обращения: 05.08.2021).
- Браницкий А. А., Дойникова Е. В., Котенко И. В. Использование нейросетей для прогнозирования подверженности пользователей социальных сетей

деструктивным воздействиям. *Информационно-управляющие системы*, 2020, № 1, с. 24–33. doi:10.31799/1684-8853-2020-1-24-33

15. Baumgartner J., Zannettou S., Keegan B., Squire M., Blackburn J. The Pushshift Reddit dataset. *Proc. the International AAAI Conference on Web and Social Media*, 2020, vol. 14, pp. 830–839.
16. Losada D. E., Crestani F. A test collection for research on depression and language use. *Proc. International Conference of the Cross-Language Evaluation Forum for European Languages*, Springer, Cham, 2016, pp. 28–39.
17. De Choudhury M. Role of social media in tackling challenges in mental health. *Proc. the 2nd Interna-*

*tional Workshop on Socially-Aware Multimedia*, 2013, pp. 49–52.

18. Al-Mosaiwi M., Johnstone T. In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, 2018, vol. 6, no. 4, pp. 529–542.
19. Nielsen F. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Proc. #MSM2011*, 2011. <https://arxiv.org/pdf/1103.2903v1.pdf> (дата обращения: 05.08.2021).
20. Shao J. Linear model selection by cross-validation. *Journal of the American Statistical Association*, 1993, vol. 88, no. 422, pp. 486–494.

UDC 004.056

doi:10.31799/1684-8853-2022-1-8-18

#### Determination of the mental state of users of the social network Reddit based on machine learning methods

A. A. Branitskiy<sup>a,b</sup>, PhD, Tech., Senior Researcher, [orcid.org/0000-0003-3104-0622](https://orcid.org/0000-0003-3104-0622), [branitskiy@comsec.spb.ru](mailto:branitskiy@comsec.spb.ru)

Y. D. Sharma<sup>c</sup>, Student, [orcid.org/0000-0003-2491-0167](https://orcid.org/0000-0003-2491-0167)

I. V. Kotenko<sup>a,b</sup>, Dr. Sc., Tech., Professor, [orcid.org/0000-0001-6859-7120](https://orcid.org/0000-0001-6859-7120)

E. V. Fedorchenko<sup>a,b</sup>, PhD, Tech., Senior Researcher, [orcid.org/0000-0001-6707-9153](https://orcid.org/0000-0001-6707-9153)

A. V. Krasov<sup>b</sup>, PhD, Tech., Associate Professor, [orcid.org/0000-0002-9076-6055](https://orcid.org/0000-0002-9076-6055)

I. A. Ushakov<sup>b</sup>, PhD, Tech., Associate Professor, [orcid.org/0000-0002-6988-9261](https://orcid.org/0000-0002-6988-9261)

<sup>a</sup>St. Petersburg Federal Research Center of the RAS, 39, 14 Line, V.O., 199178, Saint-Petersburg, Russian Federation

<sup>b</sup>The Bonch-Bruевич Saint-Petersburg State University of Telecommunications, 22-1, Bolshevikov Pr., 193232, Saint-Petersburg, Russian Federation

<sup>c</sup>Saint-Petersburg Electrotechnical University «LETI», 5, Prof. Popov St., 197376, Saint-Petersburg, Russian Federation

**Introduction:** Diagnosing mental illness is a complex process that includes conducting dialogue conversations, analyzing the behavior of the subject and passing specialized tests. The successful solution of this problem can be influenced by both the lack of knowledge and experience of the psychologist, and the presence of contradictory or incomplete initial data on the part of the patient. To eliminate this drawback, expert-based or intelligent systems are being developed. **Purpose:** Development of a technique for determining the mental state of social network users. **Results:** Using machine learning methods, a technique has been developed designed to determine the type of a mental state of social network users. The novelty of the proposed technique is in the usage of a two-step text preprocessing procedure and the construction of several sets of features which describe the emotional mood of social network users at the level of the messages published by them. As the initial data, we have used text messages of users of the social network Reddit. There are three stages in the technique: 1) data collection, 2) data preprocessing, 3) post labeling and feature construction. To assess the functioning of a software tool built on the basis of this technique, four indicators were used: accuracy, precision, recall, and F-measure. The best results are demonstrated with a One-vs-Rest ensemble using linear support vector machines as basic solvers. **Practical relevance:** The investigation results can be used in the construction of auxiliary systems that are aimed at supporting decision-making by psychologists in determining mental disorders.

**Keywords** — machine learning, social network, mental disorders, emotional mood, support vector machine, convolutional neural network.

**For citation:** Branitskiy A. A., Sharma Y. D., Kotenko I. V., Fedorchenko E. V., Krasov A. V., Ushakov I. A. Determination of the mental state of users of the social network Reddit based on machine learning methods. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 1, pp. 8–18 (In Russian). doi:10.31799/1684-8853-2022-1-8-18

#### Financial support

The work was supported in part by RFBR project 18-29-22034 mk and budget topic 0073-2019-0002.

#### References

1. Xue Y., Li Q., Jin L., Feng L., Clifton D. A., Clifford G. D. *Detecting Adolescent Psychological Pressures from Micro-Blog*. In: *Health Information Science*. Y. Zhang, G. Yao, J. He, L. Wang, N. R. Smalheiser, X. Yin Eds. Lecture Notes in Computer Science, 2014. Vol. 8423. Pp. 83–94.
2. Lin H., Jia J., Guo Q., Xue Y., Li Q., Huang J., Cai L., Feng L. User-level psychological stress detection from social media using deep neural network. *Proc. 22nd ACM International Conference on Multimedia*, 2014, pp. 507–516.
3. Park M., Chiyoung C., Meeyoung C. Depressive moods of users portrayed in Twitter. *Proc. HI-KDD*, Beijing, China, ACM, 2012. Available at: [https://nyuscholars.nyu.edu/ws/files/134720119/depressive\\_moods\\_kdd.pdf](https://nyuscholars.nyu.edu/ws/files/134720119/depressive_moods_kdd.pdf) (accessed 5 August 2021).
4. De Choudhury M., Gamon M., Counts S., Horvitz E. Predicting depression via social media. *Proc. of the International AAAI Conference on Web and Social Media*, 2013, AAAI, vol. 7, no. 1, pp. 128–137.

5. Tsugawa Sh., Kikuchi Y., Kishino F., Nakajima K., Itoh Y., Ohsaki H. Recognizing depression from twitter activity. *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 3187–3196. doi:10.1145/2702123.2702280
6. Schwartz H., Eichstaedt J., Kern M., Park G., Sap M., Stillwell D., Kosinski M., Ungar L. Towards assessing changes in degree of depression through Facebook. *Proc. the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2014, pp. 118–125. doi:10.3115/v1/W14-3214
7. Segalin C., Celli F., Polonio L., Kosinski M., Stillwell D., Sebe N., Cristani M., Lepri B. What your Facebook profile picture reveals about your personality? *Proc. the 25th ACM International Conference on Multimedia*, 2017, pp. 460–468. doi:10.1145/3123266.3123331
8. Reece A. G., Christopher M. D. Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 2017, no. 6, pp. 1–12.
9. Gkotsis G., Oellrich A., Velupillai S., Liakata M., Hubbard T. J. P., Dobson R. J. B., Dutta R. Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific Reports*, 2017, vol. 7, pp. 1–11.
10. Kim J., Lee J., Park E., Han J. A deep learning model for detecting mental illness from user content on social media. *Scientific Reports*, 2020, vol. 10, pp. 1–6. doi:10.1038/s41598-020-68764-y
11. Devlin J., Chang M.-W., Lee K., Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proc. NAACL-HLT*, Minneapolis, Minnesota, 2019, pp. 4171–4186.
12. Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L., Stoyanov V. *Roberta: A robustly optimized BERT pretraining approach*. Preprint, 2019. Available at: <https://arxiv.org/pdf/1907.11692.pdf> (accessed 5 August 2021).
13. Murarka A., Balaji R., Sushma R. *Detection and classification of mental illnesses on social media using RoBERTa*. Preprint, 2020. Available at: <https://arxiv.org/pdf/2011.11226.pdf> (accessed 5 August 2021).
14. Branitskiy A. A., Doynikova E. V., Kotenko I. V. Use of neural networks for forecasting of the exposure of social network users to destructive impacts. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2020, no. 1, pp. 24–33 (In Russian). doi:10.31799/1684-8853-2020-1-24-33
15. Baumgartner J., Zannettou S., Keegan B., Squire M., Blackburn J. The Pushshift Reddit dataset. *Proc. the International AAAI Conference on Web and Social Media*, 2020, vol. 14, pp. 830–839.
16. Losada D. E., Crestani F. A test collection for research on depression and language use. *Proc. International Conference of the Cross-Language Evaluation Forum for European Languages*, Springer, Cham, 2016, pp. 28–39.
17. De Choudhury M. Role of social media in tackling challenges in mental health. *Proc. the 2nd International Workshop on Socially-Aware Multimedia*, 2013, pp. 49–52.
18. Al-Mosaiwi M., Johnstone T. In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, 2018, vol. 6, no. 4, pp. 529–542.
19. Nielsen F. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Proc. #MSM2011*, 2011. Available at: <https://arxiv.org/pdf/1103.2903v1.pdf> (accessed 5 August 2021).
20. Shao J. Linear model selection by cross-validation. *Journal of the American Statistical Association*, 1993, vol. 88, no. 422, pp. 486–494.

## УВАЖАЕМЫЕ АВТОРЫ!

Научные базы данных, включая SCOPUS и Web of Science, обрабатывают данные автоматически. С одной стороны, это ускоряет процесс обработки данных, с другой — различия в транслитерации ФИО, неточные данные о месте работы, области научного знания и т. д. приводят к тому, что в базах оказывается несколько авторских страниц для одного и того же человека. В результате для всех по отдельности считаются индексы цитирования, что снижает рейтинг ученого.

Для идентификации авторов в сетях Thomson Reuters проводит регистрацию с присвоением уникального индекса (ID) для каждого из авторов научных публикаций.

Процедура получения ID бесплатна и очень проста, есть возможность провести регистрацию на 12-ти языках, включая русский (чтобы выбрать язык, кликните на зеленое поле сверху справа на стартовой странице): <https://orcid.org>