

УДК 330.59.330.43(47+57)

## ОЦЕНКА КАЧЕСТВА ЖИЗНИ НАСЕЛЕНИЯ РЕГИОНОВ РФ НА ОСНОВЕ ЦИФРОВЫХ ДАННЫХ: МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ

**Щекотин Евгений Викторович<sup>1,2</sup>,**  
evgvik1978@mail.ru

**Коварж Галина Юрьевна<sup>2</sup>,**  
kovarzhgalina@data.tsu.ru

**Гойко Вячеслав Леонидович<sup>2</sup>,**  
goiko.slava@gmail.com

**Петров Евгений Юрьевич<sup>2</sup>,**  
petrov@data.tsu.ru

**Бакулин Вячеслав Викторович<sup>2</sup>,**  
slava38710505@gmail.com

<sup>1</sup> Новосибирский государственный университет экономики и управления,  
Россия, 630099, Новосибирск, ул. Каменская, 56

<sup>2</sup> Национальный исследовательский Томский государственный университет,  
Россия, 634050, г. Томск, пр. Ленина, 36

**Щекотин Евгений Викторович**, кандидат философских наук, доцент кафедры социологии Новосибирского государственного университета экономики и управления; научный сотрудник Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета.

**Коварж Галина Юрьевна**, аналитик Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета.

**Гойко Вячеслав Леонидович**, директор Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета.

**Петров Евгений Юрьевич**, техник 2-й категории Суперкомпьютерного центра Национального исследовательского Томского государственного университета.

**Бакулин Вячеслав Викторович**, лаборант Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета.

**Актуальность.** Углубление тенденции на цифровизацию различных аспектов социальной жизни актуализирует применение цифровых технологий в социальных науках. Такие цифровые методы исследования, как машинное обучение, интеллектуальный анализ данных, большие данные и т. д., обладают огромным потенциалом для изучения и решения многих социальных проблем. Цифровые методы исследования получили широкое применение в исследованиях благополучия, качества жизни, счастья, удовлетворенности населения. **Цель исследования:** оценить качество жизни населения регионов РФ на основе данных социальных сетей и с использованием цифровых методов исследования. **Методология** основана на изучении субъективных оценок качества жизни пользователями социальной сети «ВКонтакте» в регионе своего проживания. Для реализации данной методологии использовались технологии больших данных (извлечение большого массива данных из социальной сети «ВКонтакте») и машинного обучения (анализ полученных данных – определение тональности и темы сообщений). **Результаты:** предложена методология изучения качества жизни населения регионов РФ с использованием цифровых методов и данных социальных сетей. Рассмотрены

*преимущества и недостатки представленного метода исследования. Разработан инструмент для оценки качества жизни – Индекс субъективного (не)благополучия. Проведена оценка качества жизни для 83 из 85 регионов РФ.*

**Ключевые слова:** Цифровая социология, цифровые методы, качество жизни, благополучие, регионы РФ, социальные сети, большие данные, машинное обучение.

## Введение

Оценка качества жизни и благополучия населения страны и отдельных регионов представляет огромный теоретический и практический интерес как для научных исследований, так и для органов государственной власти в связи с тем, что повышение благополучия граждан и управление качеством жизни стало рассматриваться как стратегическая цель во многих странах. Например, широкую известность получила практика Королевства Бутан, где в качестве мерила эффективности государственного управления рассматривается уровень счастья граждан страны [1]; благополучие населения как цель государственного управления также провозглашено в Великобритании [2]. В последние годы возникло глобальное движение, которое рассматривает счастье как центр публичной политики и управления на международном уровне [3]. В Российской Федерации эта тенденция также заметна. Так, в Стратегии национальной безопасности (в редакциях 2009 и 2015 гг.) [4, 5] повышение качества жизни населения включено в число долгосрочных национальных интересов. Интерес к проблемам качества жизни и благополучия населения охватывает самый широкий спектр тем, поэтому неудивительно, что изучением этих проблем занимаются исследователи, представляющие различные научные дисциплины – экономику, медицину, философию, социологию, психологию, педагогику и т. д.

В этой связи появляется острая потребность в создании методологии оценки качества жизни населения, которая позволяла бы, с одной стороны, учитывать множество компонентов общественной жизни, которые отражают реальные условия жизни людей, а с другой – реализацию такого измерения можно было осуществлять оперативно и на регулярной основе для оптимального сопровождения управленческих решений. На сегодняшний день история научного изучения благополучия и качества жизни насчитывает более полувека и за это время были предложены самые разнообразные составные элементы качества жизни, подходы к его измерению и оценке. Однако до сих пор в научном сообществе нет согласия о том, какие критерии качества жизни считать общепризнанными [6]. Например, в рамках только одной коллективной монографии «Качество жизни» под редакцией М. Нуссбаум и А. Сена можно выделить по меньшей мере четыре различных теоретико-методологических подхода к изучению качества жизни населения:

1. Теории идеалов – согласно данным концепциям, качество жизни включает в себя реализацию идеалов, к которым, например, можно отнести самореализацию [7].
2. В персонально-утилитарных теориях основное внимание сконцентрировано на удовольствии, ощущении счастья и исполнении желаний [8].
3. Теории изобилия акцентируют внимание на экономических характеристиках, например на реальных доходах, произведенных товарах и т. д. [9].
4. Теории базовых потребностей связана с тезисами Эрика Алларда, в которых говорится о том, что качество жизни достигается за счет удовлетворения потребительских потребностей, потребности общаться с людьми и потребности в бытии [10].

В самом общем виде все многообразие теорий благополучия можно разделить на две большие группы – теории объективного и субъективного благополучия в зависимости от того, какие типы данных используются для оценки благополучия в конкретной теоретической модели [11]. В случае объективных теорий благополучие измеряется прежде всего на основе данных об уровне дохода и потребления. В последнее время этот подход стал включать в себя не только экономические факторы, но и другие, неэкономические аспекты благополучия, такие как ценностные потребности, человеческие возможности, стабильность, проблемы гендера и т. д. Субъективные теории благополучия оперируют менее определенными категориями, такими как счастье, удовлетворенность жизнью и пр., т. е. благополучие измеряется на основе субъективных суждений людей о своей жизни. Данная классификация теорий благополучия является практически общепризнанной, она фиксирует различие в инструментах оценки благополучия, но является недостаточной, чтобы уловить, чем является благополучие в реальности.

### **Существующие методы измерения благополучия**

Для построения нашей методологии качества жизни населения регионов Российской Федерации были проанализированы существующие на сегодняшний день общемировые практики оценки качества жизни населения:

1. Система национальных счетов включает такие показатели, как чистый национальный доход, ВВП (внутренний валовый продукт), ВВП (валовый национальный продукт), личный и располагаемый личный доходы. Данный метод позволяет определить тенденции экономического развития регионов, измерить личный доход населения, изучить формирование доходов и расходов. Однако, система национальных счетов использует лишь экономические показатели общественной жизни [12].

2. Индекс физического качества жизни (Physical Quality of Life Index, PQLI) и его модификация включают в себя уровень младенческой смертности, ожидаемую продолжительность жизни, распространение грамотности среди взрослого населения, потребление продуктов питания. Данный метод оценки качества жизни доступен, однако он не учитывает социальные индикаторы, обезличивает человека и ориентирован на низкую степень удовлетворения естественных потребностей [13].

3. Индекс человеческого развития включает в себя образование, ожидаемую продолжительность жизни, индекс уровня жизни. Данная методика позволяет отслеживать динамику развития стран и регионов, проводить сравнительный анализ социально-экономического развития. Индекс человеческого развития доступен, но не учитывает субъективные и социальные показатели качества жизни [14].

4. Треугольный индекс благосостояния нации включает уровень развития социальной среды, экономической сферы и информационной инфраструктуры. С помощью данного индекса оценивается уровень развития описанных выше сфер общественной жизни и их сбалансированность. Однако треугольный индекс благосостояния нации требует сбора большого количества показателей, однако не учитывает политическую и духовную сферу [15].

Российские ученые также внесли свой вклад в разработку методик оценки благополучия населения:

1. Н.М. Римащевская дополняет индекс развития человеческого потенциала такими показателями, как интеллектуальный потенциал, профессионально-образовательные ресурсы, уровень жизни, социокультурная активность граждан, культурно-нравственные и духовные ценности. Предложенные данные более полно характеризуют

уровень благополучия населения, однако возникает проблема их достоверности и определения [16].

2. П.С. Мстиславский предлагает сравнивать Россию со странами, имеющими более высокий уровень жизни. Данный подход позволяет на основе опыта развития других стран определить возможные методы улучшения качества жизни населения [17].

3. А.И. Субетто использует такие индикаторы, как качество населения, качество образования, культура, доступность образования, уровень жизни и качество среды. Этот подход позволяет оценить благополучие населения на уровне как отдельного индивида и социальных групп, так и общества в целом [18].

4. С.А. Айвазян опирается в основном на официальные статистические данные различных сфер общественной жизни регионов и страны: благосостояние населения, качество населения, качество социальной сферы, природно-климатические условия, качество экологии. Однако стоит отметить, что авторами данного подхода не учитываются субъективные показатели [19].

5. Н.В. Трофимова предлагает качество жизни оценивать комплексно на основе объективных индикаторов и субъективных показателей. Автор в своей методике использует комплексную оценку качества жизни населения, в рамках которой необходимо последовательно объединить показатели в индексы (например, материальное благосостояние, социально-экономическое развитие, демографическая ситуация, условия жизни, рынок труда, экологическая ситуация, социальная напряженность). Стоит отметить, что в данном методе не учитываются субъективные оценки при расчете интегрального индекса [20].

Таким образом, существующие на сегодняшний день российские и зарубежные методики имеют как достоинства, так и недостатки. В рамках проводимого исследования при составлении модели были учтены критические суждения относительно представленных выше практик измерения благополучия. Вследствие этого был составлен агрегированный показатель, который включает в себя элементы социальной, экономической и политической сфер общественной жизни.

### **Цифровые исследования благополучия и качества жизни населения**

Обычно для изучения субъективного благополучия используются данные, полученные с помощью традиционных методов социологического исследования, хотя процессы, происходящие в обществе в связи с цифровизацией и развитием информационно-коммуникационных технологий, предоставляют новые возможности для изучения данной темы. В последнее десятилетие сложилось целое направление исследований – цифровая социология, для которой характерно активное использование «цифровых методов» [21], т. е. различных информационно-коммуникативных технологий для изучения социальных процессов и явлений (методы обработки естественного языка, алгоритмы машинного обучения и нейронные сети для обработки больших данных и др.) [22]. Цифровые методы социологических исследований также применяются в исследованиях благополучия и качества жизни населения [23–30].

Как показывает анализ исследований в этой области, источником данных в изучении благополучия и качества жизни цифровыми методами являются цифровые следы пользователей, т. е. информация о самых различных аспектах онлайн-активности пользователей в Интернете [31]. Важнейшим источником таких цифровых следов являются социальные сети. В зарубежной литературе авторы чаще всего исследуют благополучие, счастье, качество жизни, удовлетворенность жизнью на материале таких социальных сетей, как Twitter и Facebook, также исследователи из Китая используют данные

китайской социальной сети Sina Weibo [32]. В данном исследовании мы использовали данные крупнейшей российской социальной сети «ВКонтакте». Вообще все исследования на стыке изучения социальных сетей и проблем благополучия можно условно разделить на три большие группы. Первая группа исследований связана с применением новых информационных технологий для анализа и обработки данных, которые могут служить источником информации для изучения благополучия. Во второй группе социальные сети рассматриваются как новое социальное явление, которое оказывает влияние на различные аспекты благополучия людей. Для третьей группы социальные сети выступают уже как источник данных о различных аспектах благополучия и качества жизни [33].

В связи с тем, что за последние десять лет социальные сети очень прочно вошли в повседневную жизнь россиян, особенно представителей молодого и среднего возраста, эти исследования представляются чрезвычайно перспективными. Например, большая часть населения зарегистрирована в социальных сетях, где они выражают свое мнение и отношение к тем или иным событиям страны и региона, говорят о своих личных переживаниях и демонстрируют социальное поведение, которые могут скрывать в реальной жизни. В целом социальные сети являются тем пространством, в котором пользователи ведут себя более открыто и откровенно в сравнении с оффлайн-реальностью. Стоит отметить, что социальные медиа как источник данных обладают своими достоинствами и недостатками в сравнении с опросными методами. К достоинствам можно отнести:

1. Оперативный сбор данных не только в целом по России, но и отдельно по регионам, муниципальным образования и отдельным населенным пунктам.
2. Данные можно детализировать во времени с точностью до одного дня, что позволяет изучать колебание настроений пользователей в различные временные периоды.
3. Пользователи реагируют на актуальные, злободневные события, и делятся переживаниями, которые заботят их в текущий момент времени. Это позволяет фиксировать настроение и мнение пользователей в тот момент, когда эти переживания максимально актуальны для них.
4. Поскольку пользователи добровольно и самостоятельно выражают свое мнение, без посредничества интервьюеров и использования заранее сконструированных опросных листов, то это позволяет получить данные о тех аспектах благополучия, которые исследователи могли упустить на стадии планирования опроса.
5. Пользователи более открыто делятся своим мнением и выражают свое одобрение/неодобрение в связи с «анонимностью» в социальных сетях.
6. Суждения пользователей в социальных сетях более разнообразны в сравнении с вариантами ответов в опросных листах, поэтому и данные о различных социальных индикаторах благополучия являются более разнообразными.
7. Относительно низкие материальные и временные затраты на проведение исследования.
8. За счет больших объемов данных, которые анализируются в процессе исследования, достигается высокая точность оценок для той части населения, которая активно пользуется социальными сетями («цифровое население» региона).

К недостаткам использования социальных сетей как источника информации следует отнести:

1. Выборочная совокупность не в полной мере отражает генеральную совокупность, имеется смещение выборки.
2. Нетрадиционный язык общения пользователей может стать препятствием при использовании машинного обучения.



3. Технические трудности при сборе данных.

4. Пользователи склонны более активно выражать скорее негативное отношение, чем позитивное, поэтому наблюдается смещение оценок в сторону негативных мнений. Таким образом, согласно представленным выше аргументам, по нашему мнению, анализ поведения людей в социальных сетях предоставляет возможность получить достаточно полную информацию о пользователях для оценки субъективного благополучия. Необходимо подчеркнуть: в данной работе мы придерживаемся тезиса Дж. ван Дайка [34], что социальные сети воспроизводят те же процессы и явления, которые происходят в реальном мире, однако существенно их усиливая и интенсифицируя. Вследствие этого изучение социальных медиа предоставляет возможность получить вполне полное представление об оценке людьми условий своей жизни.

### **Методология исследования**

Методология проводимого исследования включает в себя реализацию нескольких этапов:

- 1) разработка модели субъективного благополучия;
- 2) отбор сообществ и публичных страниц в социальной сети «ВКонтакте»;
- 3) классификация текстового контента отобранных региональных сообществ;
- 4) автоматическая очистка и классификация публикаций отобранных сообществ;
- 5) построение индекса субъективного (не)благополучия для регионов России.

### *Модель субъективного благополучия*

Выше мы рассмотрели некоторые подходы к оценке качества жизни населения. В этих подходах используются различные компоненты и показатели качества жизни, но чаще всего авторы выделяют такие, как уровень жизни, уровень развития социальной инфраструктуры, состояние здоровья людей, экология и другие показатели [35]. Не все из этих показателей могут быть использованы в нашей модели субъективного благополучия, т. к. специфика изучения социальных медиа предполагает, что некоторые аспекты жизнедеятельности человека проблематично зафиксировать на основании данных социальной сети. Например, уровень жизни, один из самых популярных индикаторов благополучия невозможно определить с помощью средств используемого нами метода. Также весьма затруднительно определить удовлетворенность пользователя своей семейной жизнью, зато хорошо фиксируются субъективные оценки состояния социальной инфраструктуры региона. В модель субъективного благополучия мы включили такие показатели, как эмоциональное самочувствие (психологическое субъективное благополучие) и оценку населением политических процессов, т. к. уровень свободы в обществе выступает условием качества жизни и благополучия людей [36].

В итоге мы включили в модель качества три группы показателей:

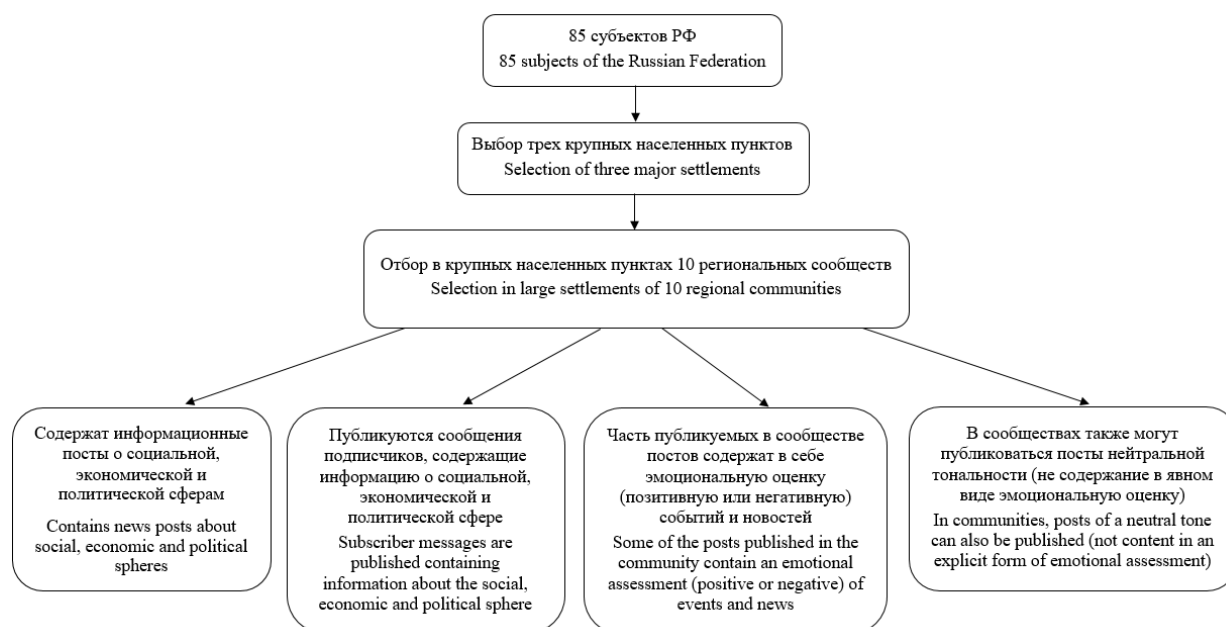
- 1) показатели социальной сферы: образование, ЖКХ, здравоохранение, инфраструктура, безопасность, экология, отношения между людьми. Сюда же мы отнесли оценку людьми своего общего эмоционального состояния;
- 2) показатели экономической сферы: работа, товары, налоги, кредитование и предпринимательство, социальная поддержка от государства;
- 3) показатели политической сферы: свобода СМИ, протестный потенциал в регионе, свобода выборов, отношение к власти, политические решения, внутренняя политика.

Более подробное описание индикаторов благополучия представлено в работе [33].

*Отбор региональных сообществ*

Далее, после выявления показателей субъективного благополучия, необходимо решить задачу определения источников данных. В качестве таких источников мы использовали сообщества пользователей в социальной сети «ВКонтакте» (рис. 1). Предварительно в каждом субъекте РФ были выбраны три наиболее крупных по числу жителей, населенных пункта. По каждому из них вручную в социальной сети «ВКонтакте» отбирались региональные сообщества – группы и публичные страницы, в которых содержатся:

- информация о социальной, экономической и политической жизни города/региона;
- открыт доступ для публикации постов от подписчиков;
- публикации о социальной, экономической и политической сферах с эмоциональным окрасом.

**Рис. 1.** Процесс отбора региональных сообществ**Fig. 1.** Selection of regional communities

Стоит отметить, что в процессе ручного отбора региональных сообществ исключались следующие типы групп:

- 1) интернет-магазины и коммерческие группы разной тематической направленности;
- 2) группы, содержащие в себе информацию об известных персонах;
- 3) сообщества, информирующие о культурных и спортивных событиях;
- 4) официальные страницы публичных заведений;
- 5) сообщества, являющиеся компаниями по доставке еды;
- 6) группы о здоровом образе жизни, фитнесе, спортивном питании, диетах и т. д.;
- 7) сообщества с предложениями по обмену вещами и передачи их даром;
- 8) группы с рассказами, личными историями и вопросами людей;
- 9) сообщества с публикациями предложений знакомства;
- 10) сообщества, в которых размещаются вакансии и предложения работы.

В результате в 83 регионах было выбрано 258 крупных населенных пункта, в которых вручную отобрано более двух тысяч региональных сообществ. Нам не удалось собрать достаточно надежные данные о сообществах в таких регионах, как Чеченская республика и Республика Мордовия. Из данных групп был выгружен текстовый контент за 2018 год, который подвергся в дальнейшем очистке и классификации по категориям и тональностям сообщений. В итоге было проанализировано более 3,3 миллиона сообщений, в отношении которых онлайн-активность проявляли почти 28,5 миллионов пользователей (оценили сообщение – поставили лайк, прокомментировали или сделали репост сообщения).

*Очистка и классификация текстового контента с помощью методов машинного обучения и методов обработки данных*

Анализ социальных сетей предполагает обработку больших массивов данных, что, в свою очередь, требует определенных временных затрат на этапе подготовки данных и обучения алгоритмов. В ходе проводимого исследования было необходимо выгрузить текстовый контент отобранных региональных сообществ, очистить их от «мусорных» сообщений, разметить их по категориям и тональностям. Для выгрузки публикаций использовалась платформа по сбору и анализу данных социальных медиа [www.opendata.university](http://www.opendata.university), разработанная сотрудниками Центра прикладного анализа больших данных НИ ТГУ (Национальный исследовательский Томский государственный университет), для выполнения других технических задач использовались методы машинного обучения и обработки данных.

Предварительная обработка данных является важнейшим этапом процесса интеллектуального анализа данных. При отсутствии предобработки данных увеличивается вероятность ошибочного результата работы модели машинного обучения, поэтому предобработке данных уделяется большое внимание в процессе реализации алгоритма. Первый этап предобработки данных – очистка текста от знаков пунктуации, различных тегов, эмодзи и нерелевантных слов. Указанные задачи реализованы с помощью регулярных выражений. После их применения все слова переводились в нижний регистр. Необходимо также удалить стоп-слова, к которым относятся предлоги, причастия, междометия, цифры, частицы и т. п., для того чтобы очистить текст от слов, которые встречаются очень часто, но не несут в себе смысла. Стоп-слова хранятся в уже готовом словаре. При нахождении в тексте слова, которое содержится в словаре, оно удаляется из общего массива текстового контента. В словари также можно добавлять свои слова, которые в дальнейшем будет необходимо удалить.

После этого, выполнен стемминг – отсечение от слова окончаний и суффиксов, чтобы оставшаяся часть, называемая stem, была одинаковой для всех грамматических форм слова. В данном случае использовалась реализация стемминга из библиотеки Natural Language Toolkit (nltk) [37]. Данные приводились к векторному виду с помощью метода TF-IDF (от англ. *TF* – term frequency, *IDF* – inverse document frequency) [38] – статическая мера, используемая для определения важности слова в документе, который является одним из элементов коллекции документов. Вес некоторого слова пропорционален количеству употребления данного текста в документе и обратно пропорционален частоте употребления слова в других файлах коллекции.

TF (term frequency – частота слова) — отношение числа вхождения некоторого слова к общему количеству слов документа. Важность слова  $t_i$  в пределах отдельного документа оценивается следующим образом:



$$TF = \frac{n_i}{\sum_k n_k},$$

где  $n_i$  – число вхождений слова в документ;  $\sum_k n_k$  – общее число слов в данном документе.

IDF (inverse document frequency – обратная частота документа) – инверсия частоты, с которой некоторое слово встречается в документах коллекции. Учёт IDF уменьшает вес широкоупотребительных слов. Для каждого уникального слова в пределах конкретной коллекции документов существует только одно значение IDF.

$$IDF = \log \frac{|D|}{|d_i \supset t_i|},$$

где  $|D|$  – количество документов в корпусе;  $|d_i \supset t_i|$  – количество документов, в которых встречается  $t_i$  (когда  $t_i$  не равен 0).

Использована реализация TF-IDF из библиотеки scikit-learn [39].

Достоинство данного метода состоит в том, что слова, наименее важные для всех документов в целом, например предлоги или междометия, получают очень низкий вес TF-IDF (т. к. часто встречаются во всех документах), а важные – высокий. К недостаткам метода относится его трудоемкость, достаточно медленная работа для больших коллекций текста.

На сегодняшний день для решения задач классификации существует множество различных методов машинного обучения, каждый из которых обладает рядом преимуществ и недостатков. Разработанное решение задачи классификации основано на градиентном бустинге (от англ. *gradient boosting*) [40].

Градиентный бустинг – один из ансамблевых методов машинного обучения, который используется для решения задач классификации и регрессии. Суть бустинга заключается в последовательном построении композиции алгоритмов машинного обучения, когда каждый следующий алгоритм стремится компенсировать недостатки предыдущего. Данный метод основан на построении модели предсказания в форме ансамбля слабых предсказывающих моделей, в данном случае деревьев решений. Другими словами, ансамблевые методы основаны на том, что объединение нескольких моделей часто может привести к созданию гораздо более мощной и качественной модели.

Целью любого алгоритма обучения с учителем является определение и минимизация функций потерь. В качестве функции потерь будет выступать ошибка – multi log-loss (от англ. *logistic loss*):

$$L = -\frac{1}{N} \sum_{i=1}^N \times \sum_{j=1}^M y_{i,j} \log(p_{i,j}),$$

где  $N$  – количество постов в базе данных;  $M$  – количество классов;  $y_{i,j}$  – равен 1, если экземпляр  $x_i$  принадлежит к классу  $j$ , иначе – 0;  $p_{i,j}$  – вероятность принадлежности экземпляра  $x_i$  к классу  $j$ .

К достоинствам выбранного алгоритма относятся:

1. Возможности для вариаций различных функций потерь, что позволяет решать как задачи классификации, так и задачи регрессии. К тому же возможность выбора произвольной функции потерь позволяет акцентировать внимание на особенностях, используемых в задаче данных.

2. Возможность рассмотрение любого семейства базовых алгоритмов, что открывает широкий спектр инструментов для учета специфики конкретной задачи. Бустинг над решающими деревьями считается одним из наиболее эффективных вариантов, учитывая то, что решающие деревья, в свою очередь, также используют базовые алгоритмы.

мы (например, пороговые, линейные и т. п.), в результате получается огромное количество вариантов для настройки конечного алгоритма.

3. Простота выполнения математических и алгоритмических оптимизаций, благодаря достаточной простоте метода и четкому математическому обоснованию, позволяет производить ряд оптимизаций, которые могут заметно ускорить работу алгоритма.

Недостатки:

1. Трудоемкость метода и достаточно медленная скорость работы. Однако данный минус в последнее время достаточно сильно нивелировался за счет появления эффективных реализаций, в том числе использующих графические карты для расчетов.

2. Без дополнительных модификаций и правильных методик обучения, он имеет свойство полностью подстраиваться под данные, в том числе под ошибки и выбросы в них, тем самым сильно подвергаясь переобучению.

3. Результаты работы бустинга сложно интерпретируемы, особенно если в композицию входят десятки алгоритмов.

В разработанном решении использовалась реализация градиентного бустинга из библиотеки LightGBM [41]. Данная реализация обладает более быстрой скоростью обучения и высокой эффективностью, а также использует меньше оперативной памяти. При этом обладает достаточно хорошей точностью и возможностью обрабатывать крупные массивы данных.

Помимо анализа информации о тексте постов, были использованы и другие данные, в том числе количество комментариев, лайков, репостов, просмотров, слов. Все перечисленные величины сканированы в зависимости от средней их величины внутри сообщества, которому принадлежит данный пост. Таким образом, в данных величинах содержится значимая информация о том, насколько пользователей заинтересовал тот или иной пост внутри сообщества. Сканированное количество слов приносит информацию о том, насколько характерна для сообщества данная длина текста. На основе описанных данных о посте и векторном представлении текста, полученном с помощью метода TF-IDF, проводилось построение классификатора градиентного бустинга.

**Таблица 1. Матрица несоответствий для классификатора категорий**  
**Table 1. Matrix of inconsistencies for the classifier of categories**

| Истинный класс<br>True class | Предсказанный класс/Predicted class |    |    |    |    |    |    |    |    |   |    |    |    |    |    |    |    |    |    |    |    |    |
|------------------------------|-------------------------------------|----|----|----|----|----|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|
|                              |                                     | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8 | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|                              | 0                                   | 86 | 0  | 0  | 0  | 1  | 5  | 0  | 0  | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 3  |
|                              | 1                                   | 32 | 28 | 0  | 0  | 9  | 9  | 0  | 0  | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 2  | 0  | 14 |
|                              | 2                                   | 19 | 0  | 53 | 0  | 7  | 10 | 1  | 0  | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 4  |
|                              | 3                                   | 28 | 0  | 0  | 34 | 3  | 19 | 0  | 0  | 0 | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 1  | 8  |
|                              | 4                                   | 18 | 0  | 0  | 0  | 63 | 7  | 0  | 0  | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 5  |
|                              | 5                                   | 10 | 0  | 0  | 0  | 1  | 84 | 0  | 0  | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  |
|                              | 6                                   | 25 | 0  | 2  | 0  | 8  | 16 | 28 | 0  | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 16 |
|                              | 7                                   | 28 | 0  | 0  | 0  | 2  | 26 | 0  | 15 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 26 |
|                              | 8                                   | 64 | 0  | 0  | 0  | 0  | 3  | 1  | 1  | 5 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 25 |
|                              | 9                                   | 24 | 0  | 0  | 1  | 2  | 6  | 0  | 0  | 0 | 42 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 2  | 3  | 13 |
|                              | 10                                  | 30 | 0  | 0  | 0  | 12 | 7  | 0  | 0  | 0 | 0  | 31 | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 12 |
|                              | 11                                  | 16 | 0  | 1  | 0  | 3  | 11 | 0  | 0  | 0 | 1  | 5  | 21 | 1  | 3  | 0  | 0  | 0  | 0  | 16 | 8  | 9  |
|                              | 12                                  | 26 | 0  | 1  | 0  | 8  | 7  | 0  | 0  | 0 | 0  | 2  | 0  | 20 | 3  | 0  | 0  | 0  | 0  | 6  | 5  | 13 |
|                              | 13                                  | 26 | 0  | 0  | 0  | 3  | 11 | 0  | 0  | 0 | 0  | 0  | 0  | 2  | 35 | 0  | 0  | 0  | 0  | 8  | 2  | 8  |
|                              | 14                                  | 32 | 0  | 0  | 0  | 10 | 24 | 0  | 0  | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 10 | 8  | 13 |
|                              | 15                                  | 31 | 0  | 0  | 0  | 7  | 12 | 0  | 0  | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 28 | 4  | 0  | 2  | 2  | 6  |
|                              | 16                                  | 23 | 0  | 0  | 0  | 1  | 3  | 0  | 0  | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 59 | 1  | 0  | 6  | 3  |
|                              | 17                                  | 50 | 0  | 1  | 1  | 6  | 7  | 0  | 0  | 0 | 0  | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 6  | 2  | 10 | 10 |
|                              | 18                                  | 24 | 0  | 0  | 0  | 10 | 11 | 1  | 0  | 0 | 0  | 0  | 0  | 0  | 2  | 0  | 0  | 0  | 0  | 28 | 8  | 8  |
|                              | 19                                  | 26 | 0  | 1  | 0  | 13 | 11 | 0  | 0  | 0 | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0  | 5  | 31 | 5  |
|                              | 20                                  | 26 | 0  | 0  | 0  | 1  | 3  | 0  | 0  | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 65 |

Была проведена кросс-валидация с 10 разбиениями выборки на обучающую и проверочную с целью определить точность построенных моделей. Точность классификатора категорий составила 73 %. Точность классификатора тональности составила 77,5 %. Ниже приведена матрица несоответствий (confusion matrix) для каждого классификатора. Значения выражены в процентах от количества экземпляров истинной категории (табл. 1 и 2).

**Таблица 2. Матрица несоответствий для классификатора тональностей**  
**Table 2. Matrix of inconsistencies for the tonality classifier**

| Класс                        | Предсказанный класс<br>Predicted class |                        |                        |                        |
|------------------------------|--|------------------------|------------------------|------------------------|
|                              | Тональность                            | Нейтральные<br>Neutral | Позитивные<br>Positive | Негативные<br>Negative |
| Истинный класс<br>True class | Нейтральные<br>Neutral                 | 42                     | 18                     | 39                     |
|                              | Позитивные<br>Positive                 | 13                     | 49                     | 36                     |
|                              | Негативные<br>Negative                 | 3                      | 4                      | 92                     |

### Индекс субъективного (не)благополучия регионов России

В ходе выполнения следующего этапа исследования был построен Индекс субъективного (не)благополучия. Поскольку достаточно хорошо известно, что пользователи социальных медиа склонны активнее реагировать на сообщения негативной тональности [42, 43], то субъективная оценка фокусируется на сообщениях, содержащих отрицательную информацию. Таким образом, наше исследование дает оценку скорее субъективному восприятию «неблагополучия» в регионе. Учитывая это обстоятельство, мы решили назвать наш инструмент «индекс субъективного неблагополучия» (ИСНБ).

ИСНБ рассчитывался по формуле

$$I_{kjt} = \frac{A_{kjt}}{B_k},$$

где  $I_{kjt}$  – это величина Индекса субъективного неблагополучия для региона ( $k$ ) по данному показателю субъективного благополучия ( $j$ ) для определенной тональности сообщения  $t$ ;  $A_{kjt}$  – показатель онлайн-активности в данном регионе по определенному показателю субъективного благополучия, рассчитывается по формуле

$$A_{kjt} = L + 2C + 5R,$$

где  $L$  – сумма лайков, которую собрали сообщения, относящиеся к определенному показателю субъективного благополучия, в данном регионе в определенной тональности;  $C$  – количество комментариев, которые собрали сообщения, относящиеся к определенному показателю субъективного благополучия, в данном регионе.

Мы приравнивали каждый комментарий к двум лайкам, т. к., по нашему мнению, это активное действие пользователя свидетельствует об особой актуальности тематики поста для комментатора. В этом случае лайк – это пассивная форма выражения поддержки данного сообщения.

$R$  – количество репостов сообщений, относящихся к определенному показателю субъективного благополучия, в данном регионе. Мы приравнивали репост к пяти лайкам, т. к., по нашему мнению, репост свидетельствует о полной и активной поддержке

пользователем данной публикации. Такое действие означает, что пользователь не просто выражает свое согласие с сообщением, но и публично демонстрирует среди своих друзей и подписчиков свою солидарность. В сравнении с различными формами онлайн-активности репост свидетельствует о наибольшей актуальности темы данного поста для пользователя.

$B_k$  – общее количество подписчиков во всех отобранных сообществах конкретного региона. Это показатель позволяет нам оценить относительную величину онлайн-активности для данного субъекта Российской Федерации.

$k$  – количество регионов (от 1 до 83).

$j$  – категория (тематика) сообщения, т. е. компонент благополучия, который мы включили в модель субъективного благополучия (от 1 до 19).

$t$  – тональность сообщения (0, 1 или 2).

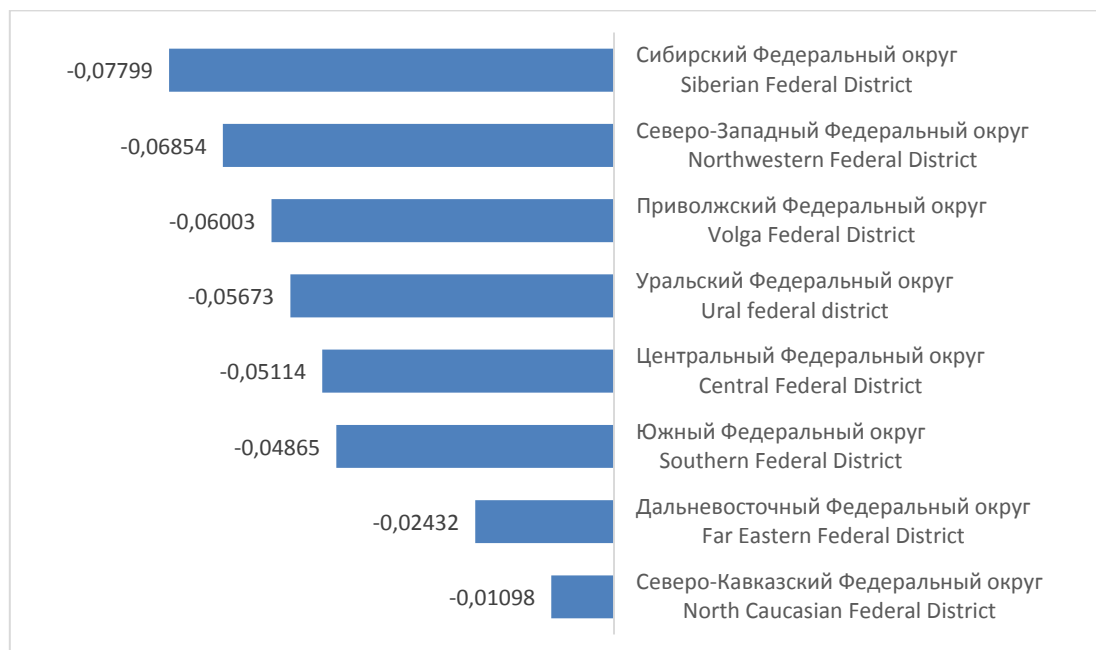
Таким образом,  $I_{kjt}$  показывает интенсивности обсуждения той или иной тематики в отобранных нами сообществах в данном регионе. Он позволяет зафиксировать актуальность и остроту для жителей региона данной темы.  $I_{kjt}$  рассчитывался по каждой тональности, т. е. для каждого региона ИСНБ имеет три значения – по положительной, негативной и нейтральной тональностям. Тональность определяется как эмоциональная оценка. Позитивная тональность поста – это эмоциональная окраска публикации, в которой прослеживается одобрение или положительная оценка того или иного показателя. Негативная тональность поста – это эмоциональная окраска сообщения, в котором наблюдаются возмущение, неприятие или негативная оценка подвергнутых анализу категорий. Нейтральная тональность поста – это отсутствие эмоциональной окраски публикации, которая является безоценочной и носит только информационный характер (информирует пользователей социальной сети «ВКонтакте» о событиях, связанных с той или иной темой социальной, экономической и политической сферой).

В рамках данного исследования был калькулирован индекс по каждому показателю и каждой тональности ежемесячно, далее рассчитано среднеемесячное значение по каждой категории и каждой тональности. Путем вычитания среднеемесячных значений индекса для положительной и отрицательной тональности была определена средняя величина индекса субъективного благополучия ИСНБ<sub>kj</sub> для каждого региона по конкретной тематике/показателю благополучия. Сообщения нейтральной тональности не использовались при расчете индекса субъективного благополучия. Итоговое значение индекса субъективного (не)благополучия ИСНБ<sub>ит</sub> рассчитывалось как сумма индекса по всем тематикам/показателям благополучия для каждого региона [22].

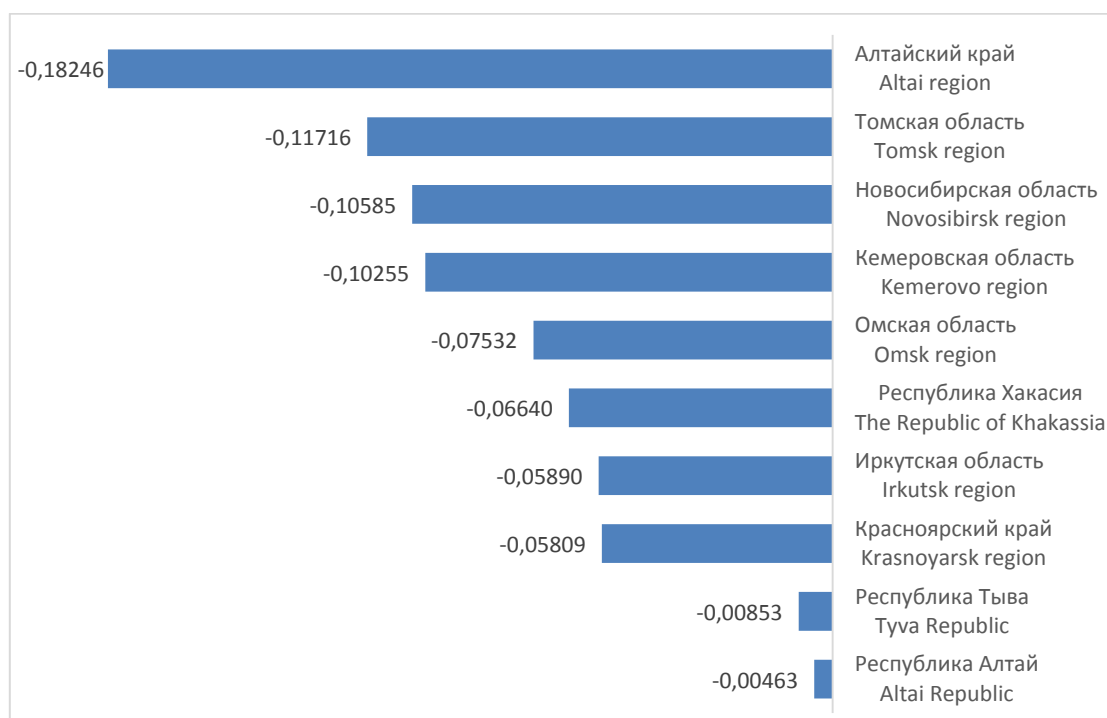
### Результаты

Таким образом, по итогам расчета индекса субъективного (не)благополучия были получены результаты по каждому субъекту РФ. Ниже представлены усредненные данные по федеральным округам (среднее значение индекса по регионам федерального округа – рассчитывается путем сложения значения итогового индекса субъективного (не)благополучия по всем регионам данного федерального округа и деления полученной суммы на количество регионов в федеральном округе) (рис. 2). На рис. 3 и 4 представлены результаты по регионам Сибирского и Северо-Кавказского федеральных округов.

На представленном выше графике можно отметить, что наибольшее отрицательное значение индекса субъективного благополучия наблюдается в Сибирском Федеральном округе.



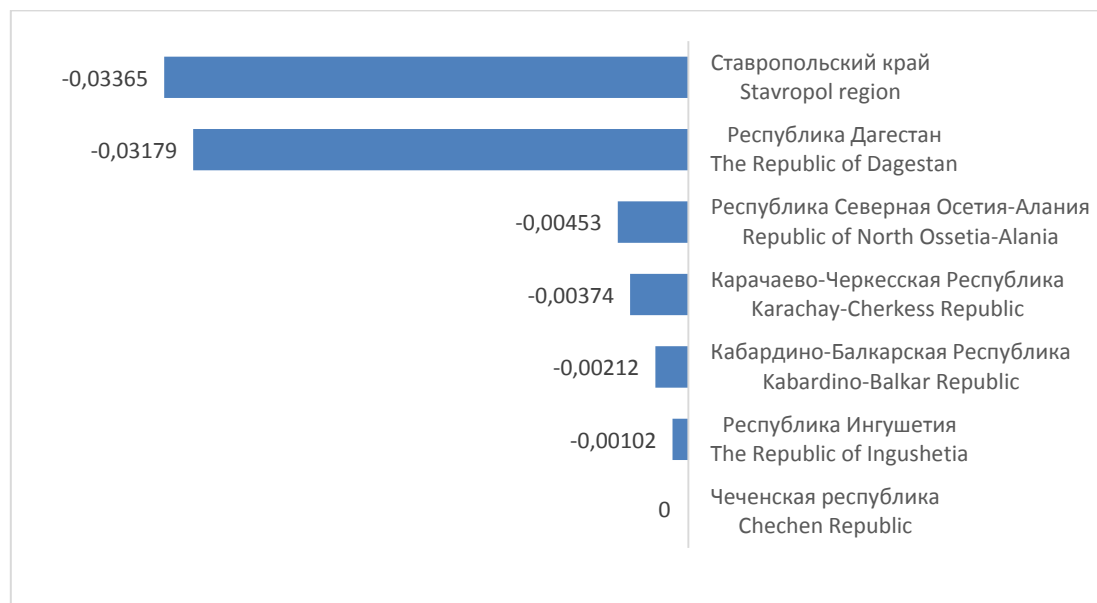
**Рис. 2. Индексы субъективного благополучия по Федеральным округам**  
**Fig. 2. Subjective well-being indices in the Federal Districts**



**Рис. 3. Индексы субъективного благополучия в Сибирском федеральном округе**  
**Fig. 3. Subjective well-being indices in the Siberian Federal District**

Следует акцентировать внимание на том, что Алтайском крае наиболее отрицательное значение индекса субъективного благополучия, в то время как в Республике Алтай и Республике Тыва субъективные оценки благополучия более позитивные. Среди всех федеральных округов самое высокое значение индекса субъективного благополучия имеет Северо-Кавказский.





**Рис. 4.** Индексы субъективного благополучия  
в Северо-Кавказском Федеральном округе

**Fig. 4.** Subjective well-being indices in the North Caucasus Federal District

Нужно заметить, что в Чеченской Республике данный показатель равен нулю, т. к. в связи со спецификой региона имеется малое количество групп в социальной сети «ВКонтакте», отражающих социальную, экономическую и политическую сферы общественной жизни данного региона.

### Заключение

В данной работе развивается новый подход к субъективной оценке благополучия с использованием данных социальных сетей и цифровых методов обработки данных. Измерение качества жизни населения на основе данных, публикуемых в социальных медиа, помогает своевременно фиксировать и реагировать на существующие социальные проблемы жителей регионов России в отличие от социологических опросов, которые требуют больше времени для сбора и анализа данных, что часто приводит к неактуальности полученных результатов. Представленный в данной статье метод имеет ряд ограничений, которые были отмечены выше, однако вместе с этим обладает преимуществами в сравнении с традиционными методами социологического исследования, к которым можно отнести быстроту, относительно низкие материальные затраты и масштабность исследования. Представляется, что использование цифровых методов изучения качества жизни, наряду с традиционными опросами, может существенно обогатить эту область научных исследований и получить больше данных о качестве жизни населения. На сегодняшний день изучение и оценка качества жизни и благополучия населения требуют использования современных инструментов и подходов, к которым можно отнести социальные сети, машинное обучение и обработку больших данных. Следовательно, можно сделать вывод о том, что созданная методика измерения субъективного благополучия актуальна в современных условиях цифровизации и развития информационно-коммуникационных технологий.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-011-00391.

## СПИСОК ЛИТЕРАТУРЫ

1. Musikanski L. Happiness in public policy // *Journal of Social Change*. – 2014. – № 6 (1). – P. 55–85.
2. Johns H., Ormerod P. Happiness, economics and public policy. – London: The Institute of Economic Affairs, 2007. – 108 p.
3. Global happiness policy report 2018. – New York: Sustainable Development Solutions Network, 2018. – 264 p.
4. О Стратегии национальной безопасности Российской Федерации до 2020 года: указ Президента Российской Федерации от 12 мая 2009 г. № 537. URL: <https://rg.ru/2009/05/19/strategia-dok.html> (дата обращения 22.10.2019).
5. О Стратегии национальной безопасности Российской Федерации: указ Президента Российской Федерации от 31 декабря 2015 года № 683. URL: <https://rg.ru/2015/12/31/nac-bezopasnost-site-dok.html> (дата обращения 22.10.2019).
6. Фофанова К.В. Методологические вопросы измерения качества жизни региона // *Регионология*. – 2016. – № 4 (97). – С. 127–139.
7. The quality of life / Eds. M. Nussbaum, S. Amartya. – Oxford: Clarendon Press, 1993. – 467 p.
8. Amartya S. Capability and well-being // *The quality of life* / Eds. M.C. Nussbaum, A. Sen. – Oxford: Clarendon Press, 1993. – P. 30–35.
9. Bliss C. Life style and the standard of living // *The quality of life* / Eds. M.C. Nussbaum, A. Sen. – Oxford: Clarendon Press, 1993. – P. 417–436.
10. Allardt E. Having, loving, being: an alternative to the Swedish model of welfare research // *The quality of life* / Eds. M.C. Nussbaum, A. Sen. – Oxford: Clarendon Press, 1993. – P. 88–94.
11. McGillivray M., Clarke M. Human well-being: concepts and measures // *Understanding human well-being* / Eds. M. McGillivray, M. Clarke. – Tokyo, New York, Paris: United Nations University Press, 2006. – P. 3–16.
12. Система национальных счетов // Федеральная служба государственной статистики. URL: [https://www.gks.ru/bgd/free/meta\\_2010/IssWWW.exe/Stg/d1996/i000310r.htm](https://www.gks.ru/bgd/free/meta_2010/IssWWW.exe/Stg/d1996/i000310r.htm) (дата обращения 17.09.2019).
13. Morris D. A physical quality of life index // *Urban Ecology*. – 1978. – V. 3. – Iss. 3. – P. 225–240.
14. Кормишкин Е.Д., Ермошина Т.В. Индекс человеческого развития в странах ЕС, аналитический обзор // *НАУКОВЕДЕНИЕ*. – 2017. – Т. 9. – № 1. URL: <http://naukovedenie.ru/PDF/36EVN117.pdf> (дата обращения 07.11.2019).
15. Зубаков В.М., Камалтдинова Р.М. Экономическая политика общего благосостояния в системе рыночных отношений. – Ульяновск: УлГТУ, 2012. – 201 с.
16. Римашевская Н.М. Качественный потенциал населения России; взгляд в XXI век // *Проблемы прогнозирования*. – 2001. – № 3. – С. 34–48.
17. Бобков В., Мстиславский П. Качество жизни: сущность и показатели // *Человек и труд*. – 1996. – № 6. – С. 76–79.
18. Субетто А.И. Качество жизни и здоровье населения // *Качество и жизнь. Сборник статей*. – СПб.: Легаси, 2010. – С. 150–170.
19. Айвазян С.А., Степанов В.С., Козлова М.И. Измерения синтетических категорий качества жизни населения региона и выявление ключевых направлений совершенствования социально-экономической политики // *Прикладная эконометрика*. – 2006. – № 2. – С. 18–84.
20. Трофимова Н.В. Интегральная оценка качества жизни населения // *Вестник РГТЭУ*. – 2010. – № 5. – С. 91–100.
21. Rogers R. Digital methods. – Cambridge, MA: MIT Press, 2013. – 274 p.
22. Щекотин Е.В. Цифровые технологии в социальных науках: предмет и метод цифровой социологии // *Социология и право*. – 2020. – № 1. – С. 49–59.
23. Yang C., Srinivasan P. Life satisfaction and the pursuit of happiness on twitter // *PLoS ONE*. – 2016. – № 11 (3). – P. e0150881. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150881> (дата обращения 07.11.2019).
24. Sabatini F., Sarracino F. Online networks and subjective well-being // *Kyklos*. – 2017. – № 70 (3). – P. 456–480.
25. Do social network sites enhance or undermine subjective well-being? / P. Verduyn, O. Ybarra, M. Resibois, J. Jonides, E. Kross // *A critical review. Social Issues and Policy Review*. – 2017. – № 11 (1). – P. 274–302.
26. Sanchez C.R., Craglia M., Bregt A.K. New data sources for social indicators: the case study of contacting politicians by Twitter // *International Journal of Digital Earth*. – 2017. – № 10 (8). – P. 829–845.
27. Clark J.L., Algae S.B., Green M.C. Social network sites and well-being: the role of social connection // *Current Directions in Psychological Science*. – 2018. – № 27 (1). – P. 32–37.
28. Well-being through the lens of the Internet / Y. Algan, F. Murtin, E. Beasley, K. Higa, C. Senik // *PLoS ONE*. – 2019. – № 14 (1). – P. e0209562. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0209562> (дата обращения 07.11.2019).

29. Bellet C., Frijters P. Big data and well-being // World Happiness Report 2019 / Eds. J. Helliwell, R. Layard, J. Sachs. – New York: Sustainable Development Solutions Network, 2019. – P. 97–122.
30. Hills T., Proto E., Sgroi D. Historical analysis of national subjective well-being using millions of digitized books // Nature: Human Behaviour. – 2019. – № 3 (12). – P. 1271–1275.
31. Флёров О.В. Цифровой след человека в Интернете: основные гуманитарные подходы // Образовательные ресурсы и технологии. – 2018. – № 4 (25). – С. 79–82.
32. Sensing Subjective Well-Being from Social Media / B. Hao, L. Li, R. Gao, A. Li, T. Zhu // Proc. of 10th International Conference Active Media Technology. – Cham, Springer. – 2014. – P. 324–336.
33. Субъективная оценка (не)благополучия населения регионов РФ на основе данных социальных сетей / Е.В. Щекотин, М.Г. Мягков, В.Л. Гойко, В.В. Кашпур, Г.Ю. Коварж // Мониторинг общественного мнения: Экономические и социальные перемены. – 2020. – № 1. – С. 78–116.
34. Van Dijk J. The network society. social aspects of new media. – London: SAGE Publ., 2006. – 300 p.
35. Гаврилова Т.В. Принципы и методы исследования качества жизни // Технологии качества жизни. – 2004. – № 4 (2). – С. 1–11.
36. Veenhoven R. Happy life-expectancy. A comprehensive measure of quality-of-life in nations // Social Indicators Research. – 1996. – № 39. – P. 1–58.
37. Natural Language Toolkit // NLTK 3.5 documentation. URL: <https://www.nltk.org/> (дата обращения 19.09.2019).
38. Interpreting TF-IDF term weights as making relevance decisions / H. Wu, R. Luk, K. Wong, K. Kwok // ACM Transactions on Information Systems. – 2008 – 26 (3). Article 13. URL: <https://dl.acm.org/doi/10.1145/1361684.1361686> (дата обращения 19.09.2019).
39. Feature extraction // Scikit learn. URL: [https://scikit-learn.org/stable/modules/feature\\_extraction.html](https://scikit-learn.org/stable/modules/feature_extraction.html) (дата обращения 21.09.2019).
40. Gradient Boosting from scratch // ML review. URL: <https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d> (дата обращения 03.10.2019).
41. Welcome to LightGBM's documentation! // LightGBM. URL: <https://lightgbm.readthedocs.io/en/latest/> (дата обращения 27.09.2019).
42. Trussler M., Soroka S. Consumer demand for cynical and negative news frames // The International Journal of Press. Politics. – 2014 – V. 19. – № 3. – P. 360–379.
43. Liebrecht C., Hustinx L., Van Mulken M. The relative power of negativity: the influence of language intensity on perceived strength // Journal of Language and Social Psychology. – 2019 – V. 38. – № 2. – P. 170–193.

*Поступила 07.06.2020 г.*

UDC 330.59.330.43(47+57)

## ASSESSMENT OF THE POPULATION LIFE QUALITY IN RUSSIAN REGIONS BASED ON DIGITAL DATA: METHODOLOGICAL ASPECTS

**Eugenii V. Shchekotin**<sup>1,2</sup>,  
evgvik1978@mail.ru

**Galina Yu. Kovarzh**<sup>2</sup>,  
kovarzhgalina@data.tsu.ru

**Vyacheslav L. Goiko**<sup>2</sup>,  
goiko.slava@gmail.com

**Evgeny Yu. Petrov**<sup>2</sup>,  
petrov@data.tsu.ru

**Vyacheslav V. Bakulin**<sup>2</sup>,  
slava38710505@gmail.com

<sup>1</sup> Novosibirsk State University of Economics and Management,  
56, Kamenskaya street, Novosibirsk, 630099, Russia.

<sup>2</sup> National Research Tomsk State University,  
36, Lenin avenue, Tomsk, 634050 Russia.

**Eugenii V. Shchekotin**, Cand. Sc., associate professor, Novosibirsk State University of Economics and Management; researcher, National Research Tomsk State University.

**Galina Yu. Kovarzh**, analyst, National Research Tomsk State University.

**Vyacheslav L. Goiko**, head of the Center of Applied Big Data Analysis, National Research Tomsk State University.

**Evgeny Yu. Petrov**, 2<sup>nd</sup> category technician, National Research Tomsk State University.

**Vyacheslav V. Bakulin**, laboratory assistant, National Research Tomsk State University.

**Relevance.** Deepening the trend towards digitalization of various aspects of social life actualizes the use of digital technologies in the social sciences. Such digital research methods as machine learning, data mining, big data, etc. open up a huge potential for studying and solving many social problems. Digital research methods are also widely used in studies of well-being, quality of life, happiness, and life satisfaction. **The main aim** of the study is to assess the population life quality in Russian regions based on social network data and using digital research methods. **Methods.** The research methodology is based on the study of subjective assessments of the quality of life by users of the social network Vkontakte in their region of residence. To implement this methodology, we used big data technologies (extracting a large array of data from a social network in Vkontakte) and machine learning (analyzing the data obtained – determining the tone and subject of messages). **Results.** A methodology for studying the quality of life of the population of Russian regions using digital methods and social network data is proposed. The advantages and disadvantages of the presented research method are considered. We have developed a tool for assessing the quality of life – the index of subjective (non-) well-being. The quality of life was assessed for 83 out of 85 regions of the Russian Federation.

**Key words:** Digital sociology, digital methods, quality of life, well-being, regions of the Russian Federation, social networks, big data, machine learning.

The research was financially supported by the RFBR within the scientific project no. 20-011-00391.

## REFERENCES

1. Musikanski L. Happiness in public policy. *Journal of Social Change*, 2014, no. 6 (1), pp. 55–85.
2. Johns H., Ormerod P. *Happiness, economics and public policy*. London, The Institute of Economic Affairs, 2007. 108 p.
3. *Global happiness policy report 2018*. New York, Sustainable Development Solutions Network, 2018. 264 p.
4. *O Strategii natsionalnoy bezopasnosti Rossiyskoy Federatsii do 2020 goda* [On the National Security Strategy of the Russian Federation until 2020]. Ukaz Prezidenta Rossiyskoy Federatsii ot 12 maya 2009 g. no. 537 [Decree of the President of the Russian Federation of May 12, 2009 N 537]. Available at: <https://rg.ru/2009/05/19/strategia-dok.html> (accessed 22 October 2019).
5. *O Strategii natsionalnoy bezopasnosti Rossiyskoy Federatsii* [On the National Security Strategy of the Russian Federation]. Ukaz Prezidenta Rossiyskoy Federatsii ot 31 dekabrya 2015 goda no. 683 [Decree of the President of the Russian Federation of December 31, 2015 N 683]. Available at: <https://rg.ru/2015/12/31/nac-bezopasnost-site-dok.html> (accessed 22 October 2019).
6. Fofanova K.V. Metodologicheskie voprosy izmereniya kachestva zhizni regiona [Methodological issues of measuring the quality of life in the region]. *Regionologiya*, 2016, no. 4 (97), pp. 127–139.
7. *The Quality of Life*. Eds. M. Nussbaum, S. Amartya. Oxford, Clarendon Press, 1993. 467 p.
8. Amartya S. Capability and well-being. *The quality of life*. Eds. M.C. Nussbaum, A. Sen. Oxford, Clarendon Press, 1993. pp. 30–35.
9. Bliss C. Life style and the standard of living. *The quality of life*. Eds. M.C. Nussbaum, A. Sen. Oxford, Clarendon Press, 1993. pp. 417–436.
10. Allardt E. Having, loving, being: an alternative to the Swedish model of welfare research. *The quality of life*. Eds. M.C. Nussbaum, A. Sen. Oxford, Clarendon Press, 1993. pp. 88–94.
11. McGillivray M., Clarke M. Human well-being: concepts and measures. *Understanding human well-being*. Eds. M. McGillivray, M. Clarke. Tokyo, New York, Paris, United Nations University Press, 2006. pp. 3–16.
12. *Sistema natsionalnykh schetov* [System of national accounts]. Available at: [https://www.gks.ru/bgd/free/meta\\_2010/IssWWW.exe/Stg/d1996/i000310r.htm](https://www.gks.ru/bgd/free/meta_2010/IssWWW.exe/Stg/d1996/i000310r.htm) (accessed 17 September 2019).
13. Morris D. A physical quality of life index. *Urban Ecology*, 1978, vol. 3, Iss. 3, pp. 225–240.
14. Kormishkin E.D., Ermoshina T.V. Indeks chelovecheskogo razvitiya v stranakh ES, analiticheskiy obzor [Human Development Index in the EU countries, analytical overview]. *NAUKOVEDENIE*, 2017, vol. 9, no. 1. Available at: <http://naukovedenie.ru/PDF/36EVN117.pdf> (accessed 7 November 2019).
15. Zubakov V.M., Kamaltdinova R.M. Ekonomicheskaya politika obshchego blagosostoyaniya v sisteme rynochnykh otnosheniy [Economic policy of general welfare in the system of market relations]. Ulyanovsk, UIGTU Publ., 2012. 201 p.
16. Rimashevskaya N.M. Kachestvennyy potentsial naseleniya Rossii: vzglyad v XXI vek [The qualitative potential of the population of Russia; look into the XXI century]. *Problemy prognozirovaniya*, 2001, no. 3, pp. 34–48.
17. Bobkov V., Mstislavskiy P. Kachestvo zhizni: sushchnost i pokazateli [Quality of life: essence and indicators]. *Chelovek i trud*, 1996, no. 6, pp. 76–79.
18. Subetto A.I. Kachestvo zhizni i zdorove naseleniya [Quality of life and health of the population]. *Kachestvo i zhizn. Sbornik statey* [Quality and life. Digest of articles]. Ed. by V.V. Okrepilova. St. Petersburg, Legasi Publ., 2010. pp. 150–170.
19. Ayvazyan S.A., Stepanov V.S., Kozlova M.I. Izmereniya sinteticheskikh kategoriy kachestva zhizni naseleniya regiona i vyyavleniye klyuchevykh napravleniy sovershenstvovaniya sotsialno-ekonomicheskoy politiki [Measuring synthetic categories of the quality of life of the region's population and identifying key areas for improving socio-economic policy]. *Prikladnaya ekonometrika*, 2006, no. 2, pp. 18–84.
20. Trofimova N.V. Integralnaya otsenka kachestva zhizni naseleniya [Integral assessment of the quality of life of the population]. *Vestnik RGT-EU*, 2010, no. 5, pp. 91–100.
21. Rogers R. *Digital methods*. Cambridge, MA, MIT Press, 2013. 274 p.
22. Shchekotin E.V. Tsifrovye tekhnologii v sotsialnykh naukakh: predmet i metod tsifrovoy sotsiologii [Digital technologies in the social sciences: the subject and method of digital sociology]. *Sotsiologiya i pravo*, 2020, no. 1, pp. 49–59.
23. Yang C., Srinivasan P. Life satisfaction and the pursuit of happiness on twitter. *PLoS ONE*, 2016, no. 11 (3), pp. e0150881. Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150881> (accessed 07 November 2019).
24. Sabatini F., Sarracino F. Online networks and subjective well-being. *Kyklos*, 2017, no. 70 (3), pp. 456–480.
25. Verduyn P., Ybarra O., Resibois M., Jonides J., Kross E. Do social network sites enhance or undermine subjective well-being? *A critical review. Social Issues and Policy Review*, 2017, no. 11 (1), pp. 274–302.



26. Sanchez C.R., Craglia M., Bregt A.K. New data sources for social indicators: the case study of contacting politicians by Twitter. *International Journal of Digital Earth*, 2017, no. 10 (8), pp. 829–845.
27. Clark J.L., Algoe S.B., Green M.C. Social network sites and well-being: the role of social connection. *Current Directions in Psychological Science*, 2018, no. 27 (1), pp. 32–37.
28. Algan Y., Murtin F., Beasley E., Higa K., Senik C. Well-being through the lens of the Internet. *PLoS ONE*, 2019, no. 14 (1), pp. e0209562. Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0209562> (accessed 07 November 2019).
29. Bellet C., Frijters P. Big data and well-being. *World Happiness Report 2019*. Eds. J. Helliwell, R. Layard, J. Sachs. New York, Sustainable Development Solutions Network, 2019. pp. 97–122.
30. Hills T., Proto E., Sgroi D. Historical analysis of national subjective well-being using millions of digitized books. *Nature: Human Behaviour*, 2019, no. 3 (12), pp. 1271–1275.
31. Flerov O.V. Tsifrovoy sled cheloveka v Internete: osnovnye gumanitarnye podkhody [The digital footprint of a person on the Internet: basic humanitarian approaches]. *Obrazovatelnye resursy i tekhnologii*, 2018, no. 4 (25), pp. 79–82.
32. Hao B., Li L., Gao R., Li A., Zhu T. Sensing subjective well-being from social media. *Proc. of 10<sup>th</sup> International Conference Active Media Technology*. Cham, Springer, 2014. pp. 324–336.
33. Shchekotin E.V., Myagkov M.G., Goyko V.L., Kashpur V.V., Kovarzh G.Yu. Subektivnaya otsenka (ne)blagopoluchiya naseleniya regionov RF na osnove dannykh sotsialnykh setey [Subjective assessment of the (non) well-being of the population of the regions of the Russian Federation based on data from social networks]. *Monitoring obshchestvennogo mneniya: Ekonomicheskie i sotsialnye peremeny*, 2020, no. 1, pp. 78–116.
34. Van Dijk J. *The network society. social aspects of new media*. London, SAGE Publ., 2006. 300 p.
35. Gavrilova T.V. Printsipy i metody issledovaniya kachestva zhizni [Principles and methods of researching the quality of life]. *Tekhnologii kachestva zhizni*, 2004, no. 4 (2), pp. 1–11.
36. Veenhoven R. Happy life-expectancy. A comprehensive measure of quality-of-life in nations. *Social Indicators Research*, 1996, no. 39, pp. 1–58.
37. *Natural language toolkit*. NLTK 3.5 documentation. Available at: <https://www.nltk.org/> (accessed 19 September 2019).
38. Wu H., Luk R., Wong K., Kwok K. Interpreting TF-IDF term weights as making relevance decisions. *ACM Transactions on Information Systems*, 2008, no. 26 (3), Article 13. Available at: <https://dl.acm.org/doi/10.1145/1361684.1361686> (accessed 19 September 2019).
39. Feature extraction. *Scikit learn*. Available at: [https://scikit-learn.org/stable/modules/feature\\_extraction.html](https://scikit-learn.org/stable/modules/feature_extraction.html) (accessed 21 September 2019).
40. Gradient Boosting from scrat-sh. *ML review*. Available at: <https://medium.com/mlreview/gradient-boosting-from-scrat-sh-1e317ae4587d> (accessed 03 October 2019).
41. Welcome to LightGBM's documentation! *LightGBM*. Available at: <https://lightgbm.readthedocs.io/en/latest/> (accessed 27 September 2019).
42. Trussler M., Soroka S. Consumer demand for cynical and negative news frames. *The International Journal of Press. Politics*, 2014, vol. 19, no. 3, pp. 360–379.
43. Liebrecht C., Hustinx L., Van Mulken M. The relative power of negativity: the influence of language intensity on perceived strength. *Journal of Language and Social Psychology*, 2019, vol. 38, no. 2, pp. 170–193.

Received: 7 June 2020.