

DOI: 10.26794/3033-7097-2025-1-4-16-25
УДК 004.89(045)

Анализ тональности пользовательского текста методами машинного обучения

Е.А. Горбунова^а, Р.А. Кочкаров^б, Э.А. Окунева^с^а ООО «Лаборатория систем автоматизации процессов», Санкт-Петербург, Российская Федерация;
^{б,с} Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

АННОТАЦИЯ

В статье рассматривается применение методов машинного обучения для анализа тональности текстов, опубликованных пользователями социальной сети ВКонтакте. Это дает возможность в режиме реального времени отслеживать и анализировать настроения миллионов пользователей, что способствует оперативному принятию решений и прогнозированию социальных процессов. В рамках исследования был реализован сбор текстовых данных с использованием VK API, включающих посты и комментарии пользователей. Проведена предобработка текстов: очистка, лемматизация, удаление стоп-слов и векторизация методом TF-IDF. Для классификации эмоциональной окраски были протестированы модели: логистическая регрессия, случайный лес, наивный байесовский классификатор, а также нейросетевые архитектуры LSTM и Transformers (RuBERT). Наивный байесовский классификатор показал наилучшие результаты по метрике полноты и сбалансированности по другим метрикам. Согласно результатам анализа, большинство текстов пользователей имеют нейтральную или положительную тональность, и лишь незначительная часть — негативную. Представлены визуализации и статистика распределения тональности. **Работа демонстрирует** эффективность применения классических методов машинного обучения для обработки и анализа текстов в русскоязычных социальных сетях.

Ключевые слова: анализ тональности; машинное обучение; социальные сети; ВКонтакте; обработка естественного языка; TF-IDF; байесовский классификатор; сентимент-анализ; посты; комментарии

Для цитирования: Горбунова Е.А., Кочкаров Р.А., Окунева Э.А. Анализ тональности пользовательского текста методами машинного обучения. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(4):16-25. DOI: 10.26794/3033-7097-2025-1-4-16-25

ORIGINAL PAPER

Sentiment Analysis of User Texts with Machine Learning Methods

Е.А. Gorbunova^а, R.A. Kochkarov^б, E.A. Okuneva^с^а Laboratory of Process Automation Systems Limited Liability Co., Saint Petersburg, Russian Federation;
^{б,с} Financial University under the Government of the Russian Federation, Moscow, Russian Federation

ABSTRACT

This paper explores the application of machine learning methods for sentiment analysis of user-generated texts in the Russian social network VKontakte. The sentiments of millions of users could be monitored and analyzed in real time, that facilitates prompt decision making and forecasting of social processes. Textual data, including posts and comments, were collected via the VK API. The preprocessing pipeline involved text cleaning, lemmatization, stop-word removal, and TF-IDF vectorization. Several classification models were tested, including logistic regression, random forest, and naïve Bayes, as well as deep learning models such as LSTM and Transformers (RuBERT). The naïve Bayes classifier demonstrated the best performance in terms of recall and overall metric balance. Sentiment analysis results revealed that the majority of user texts were neutral or positive, with only a small portion being negative. The paper includes visualizations and statistical summaries of sentiment distribution. **The study confirms** the effectiveness of classical machine learning methods for processing and analyzing textual data in Russian social networks.

Keywords: sentiment analysis; machine learning; social networks; VKontakte; natural language processing; TF-IDF; naïve Bayes; user-generated content; posts; comments

For citation: Gorbunova E.A., Kochkarov R.A., Okuneva E.A. Sentiment analysis of user texts with machine learning methods. *Digital Solutions and Artificial Intelligence Technologies*. 2025;1(4):16-25. DOI: 10.26794/3033-7097-2025-1-4-16-25

© Горбунова Е.А., Кочкаров Р.А., Окунева Э.А., 2025



ВВЕДЕНИЕ

Современные социальные сети стали важной площадкой для выражения мнений, обмена информацией и формирования общественного мнения. Миллионы пользователей ежедневно публикуют тексты, комментарии, делятся эмоциями и впечатлениями, создавая огромные объемы неструктурированных данных. Анализ тональности (sentiment analysis) — автоматическое определение эмоциональной окраски текстов — приобретает ключевое значение для понимания общественных настроений, мониторинга репутации брендов, оценки эффективности маркетинговых кампаний и выявления трендов.

Особенно актуальным анализ тональности становится в контексте социальных сетей, таких как ВКонтакте, которые являются крупнейшими платформами в русскоязычном сегменте интернета. Здесь исследователи и компании получают возможность в режиме реального времени отслеживать и анализировать настроения миллионов пользователей, что способствует оперативному принятию решений и прогнозированию социальных процессов [1].

Однако специфика русского языка, включая его морфологическую сложность и богатство форм слов, предъявляет особые требования к методам обработки и анализа текстов. Современные технологии машинного обучения и глубокого обучения, такие как наивный байесовский классификатор, рекуррентные нейронные сети (LSTM) и трансформеры (RuBERT), активно применяются для повышения точности определения тональности текстов на русском языке [2].

Таким образом, актуальность анализа тональности в социальных сетях обусловлена не только ростом объема пользовательских данных, но и необходимостью создания эффективных инструментов для их автоматической обработки и интерпретации, что имеет важное значение для бизнеса, науки и социальной сферы [3].

Цель работы — выявление эмоциональной окраски текстов пользователей сети ВКонтакте с помощью методов анализа тональности.

МЕТОДЫ ИССЛЕДОВАНИЯ

Исследование проводилось в несколько этапов:

- сбор данных из сети ВКонтакте (посты и комментарии) через API;
- предобработка текста (очистка, токенизация, лемматизация);
- векторизация методом TF-IDF;
- обучение и сравнение моделей (логистическая регрессия, случайный лес, наивный байес, LSTM, RuBERT);
- оценка моделей по метрикам (Accuracy, Precision, Recall, F1-score);

- анализ распределения тональности (позитивные, нейтральные, негативные тексты).

В рамках данного исследования для анализа тональности текстов из социальной сети ВКонтакте применялись современные методы обработки естественного языка (NLP) и машинного обучения [4]. Основная цель состояла в автоматическом определении эмоциональной окраски текстов — положительной, нейтральной или отрицательной.

Первым этапом являлась предобработка данных, которая включала очистку текста от пунктуации, специальных символов и URL, токенизацию, удаление русскоязычных стоп-слов и лемматизацию [5]. Эти шаги способствуют снижению шума в данных и повышают качество последующего анализа. Для лемматизации использовались специализированные инструменты, учитывающие особенности русского языка.

Для преобразования текстовых данных в числовой формат была применена модель TF-IDF (Term Frequency — Inverse Document Frequency), которая позволяет учитывать не только частоту слов, но и их значимость относительно всего корпуса текстов [6]. Этот подход улучшает различимость ключевых терминов в контексте тональности.

Далее были обучены и сравнены несколько моделей машинного обучения: логистическая регрессия, случайный лес и наивный байесовский классификатор [7]. Кроме классических моделей, были протестированы модели глубокого обучения — рекуррентная нейронная сеть LSTM и трансформерная модель RuBERT, специально адаптированная для русского языка. Выбор таких моделей обусловлен их способностью выявлять сложные зависимости в текстах и учитывать контекст при анализе тональности.

Оценка эффективности моделей проводилась по стандартным метрикам качества классификации — точности (Accuracy), полноте (Recall), точности классификации (Precision) и F1-мере. Анализ показал, что наивный байесовский классификатор обеспечил лучший баланс между полнотой и точностью, что важно для задач сентимент-анализа.

Таким образом, использование комплексного подхода — от тщательной предобработки до применения различных моделей машинного и глубокого обучения — позволило получить надежные результаты по определению эмоциональной окраски текстов пользователей ВКонтакте.

ОПИСАНИЕ СОЦИАЛЬНОЙ СЕТИ ВКОНТАКТЕ КАК ОБЪЕКТА ИССЛЕДОВАНИЯ

Социальная сеть ВКонтакте (VK) является одной из крупнейших и наиболее популярных плат-

форм для онлайн-коммуникации в России и странах СНГ. Запущенная в 2006 г., она предоставляет пользователям широкий спектр возможностей для общения, обмена контентом, создания сообществ и проведения маркетинговых кампаний. По состоянию на 2024 г., количество активных пользователей сети ВКонтакте превышает 100 млн, что делает ее уникальным объектом для социологических и лингвистических исследований.

Платформа ВКонтакте сочетает в себе функции социальной сети, медиаплатформы и инструмента для организации сообществ по интересам. Пользователи могут создавать и публиковать посты, комментировать и ставить лайки, делиться фотографиями, аудио- и видеоматериалами¹. Такая многообразная активность обеспечивает богатый и разнообразный корпус данных для анализа.

Особенностью ВКонтакте является относительно открытый доступ к данным через публичный API, что позволяет исследователям собирать и анализировать значительные объемы информации, сохраняя при этом конфиденциальность пользователей². API ВКонтакте предоставляет возможности для получения данных о пользователях, их постах, комментариях, реакциях и взаимодействиях, что важно для комплексного анализа социальных процессов и настроений.

Для анализа эмоциональной окраски текстов сеть ВКонтакте представляет интерес благодаря своей русскоязычной аудитории, разнообразию тем и стилю общения [8]. Платформа отражает как повседневные коммуникации, так и реакции на социальные, политические и культурные события, что делает ее идеальной площадкой для изучения динамики общественного мнения и эмоциональных реакций.

Таким образом, социальная сеть ВКонтакте — это многогранная среда с огромным массивом текстовой информации, которая служит надежной основой для исследования тональности и настроений пользователей в современном цифровом пространстве.

ИСПОЛЬЗОВАНИЕ VK API ДЛЯ СБОРА ДАННЫХ И ОПИСАНИЕ ДАТАФРЕЙМОВ

Для сбора данных из социальной сети ВКонтакте в рамках данного исследования использовался официальный программный интерфейс приложения — VK API. Этот инструмент предоставляет доступ к широкому спектру информации о пользователях, их активностях, постах и комментариях, что делает его незаменимым для анализа социальных сетей.

¹ Что такое ВК. Вконтакте. URL: <https://vk.com/company/ru/company/about/>

² API/VK для разработчиков. Вконтакте. URL: <https://dev.vk.com/ru/reference>

Таблица 1 / Table 1

Данные о пользователях социальной сети Вконтакте / VKontakte Social Network User Data

№	Обозначение / Designation	Данные / Data
1	id	Идентификационный номер пользователя
2	bdate	Дата рождения
3	city_title	Город
4	country_title	Страна
5	first_name	Имя
6	last_name	Фамилия

Источник / Source: составлено авторами / Compiled by the authors.

Таблица 2 / Table 2

Данные о постах пользователей социальной сети Вконтакте / User Post Data from VK Social Network

№	Обозначение / Designation	Данные / Data
1	user_id	Идентификационный номер пользователя
2	post_id	Идентификационный номер публикации пользователя
3	post_text	Текст поста
4	date_formatted	Дата публикации в формате ДД.ММ.ГГГГ
5	time	Время публикации в формате ЧЧ: ММ: СС
6	likes	Количество лайков
7	comments	Количество комментариев
8	views	Количество просмотров
9	photo_url	URL фотографии в посте, если есть
10	video_link	Ссылка на видео в посте, если есть
11	audio_link	Ссылка на аудио в посте, если есть

Источник / Source: составлено авторами / Compiled by the authors.

В работе применялись ключевые методы API, такие как:

- `users.search` — метод для поиска пользователей по заданным параметрам (например, возраст, пол, город), что позволило сформировать репрезентативную выборку пользователей для анализа;
- `users.get` — предоставлял подробную информацию о конкретных пользователях, включая их демографические данные и профильные атрибуты;
- `wall.get` — метод, используемый для извлечения постов с пользовательских стен, включая тексты публикаций, время создания, количество лайков, комментариев и просмотров;
- `likes.getList` и `comments.getList` — методы для получения информации о пользователях, которые взаимодействовали с постами (ставили лайки, оставляли комментарии)³.

В результате использования методов VK API для сбора данных была сформирована серия датафреймов, представляющих собой структурированные таблицы с информацией о пользователях, их публикациях, а также взаимодействиях с этими публикациями (лайки и комментарии). Каждый из датафреймов отражает определенный аспект активности в социальной сети ВКонтакте и является основой для последующего анализа.

Данные датафрейма `df_all_users` — информация о пользователях в социальной сети ВКонтакте содержит 14 711 уникальных пользователей социальной сети (табл. 1).

Для сбора данных о постах пользователей были применены методы: `users.get` и `wall.get`.

Данные датафрейма `df_all_posts` — информация о пяти последних постах пользователей в социальной сети ВКонтакте, содержат 55 589 уникальных идентификаторов постов пользователей (табл. 2).

Данные датафрейма `df_likes` — информация о людях, поставивших лайк на посты, содержат 1 191 231 записей (табл. 3).

Данные датафрейма `df_comments` содержат информацию о людях, которые оставили комментарий под постом другого пользователя, и насчитывают 23 467 записей (табл. 4).

В процессе сбора данных строго учитывался ряд ограничений, установленных платформой ВКонтакте для обеспечения стабильной работы сервиса и защиты данных пользователей.

- Ограничение по количеству запросов: API ВКонтакте накладывает лимит на количество запросов в единицу времени (например, не более трех запросов в секунду для одного токена). Для

Данные о взаимных лайках пользователей социальной сети / Mutual Likes Data between Social Network Users

№	Обозначение / Designation	Данные / Data
1	owner_id	Идентификационный номер пользователя, кому принадлежит пост
2	post_id	Идентификационный номер поста
3	liker_id	Идентификационный номер пользователя, который поставил лайк посту

Источник / Source: составлено авторами / Compiled by the authors.

Таблица 4 / Table 4

Данные о комментариях пользователей социальной сети / Social Media User Comments Data

№	Обозначение / Designation	Данные / Data
1	owner_id	Идентификационный номер пользователя, кому принадлежит пост
2	post_id	Идентификационный номер поста
3	commenter_id	Идентификационный номер пользователя, который оставил комментарий под постом
4	comment_text	Текст комментария
5	likes_count	Количество лайков данного комментария

Источник / Source: составлено авторами / Compiled by the authors.

предотвращения превышения лимита запросы были организованы с использованием очередей и задержек, что обеспечило плавное и корректное взаимодействие с сервером.

- Пагинация данных: для получения больших объемов данных применялась страничная загрузка с использованием параметров `offset` и `count`. Это позволяло эффективно обходить ограничение на максимальное количество записей, возвращаемых за один запрос.

- Этические и правовые нормы: соблюдались правила обработки персональных данных, уста-

³ Использование VK API. ВКонтакте. URL: <https://dev.vk.com/ru/api/overview>

новленные законодательством и политикой сети ВКонтакте. Данные пользователей обрабатывались анонимно и использовались исключительно в агрегированном виде с целью проведения научного анализа.

- Обновление и мониторинг API: в процессе работы регулярно отслеживались изменения в документации VK API и корректировались методы и параметры запросов для поддержания актуальности и стабильности сбора данных.

Таким образом, благодаря четкой организации процесса и соблюдению технических и этических норм был обеспечен качественный и безопасный сбор данных, необходимых для дальнейшего анализа тональности и поведения пользователей социальной сети.

ПРЕДОБРАБОТКА ДАННЫХ

Перед применением моделей машинного обучения для анализа тональности текстов необходимо провести качественную предобработку данных. Целью данного этапа является очистка текста от лишней информации, приведение его к унифицированному виду и подготовка к векторному представлению, пригодному для машинной обработки.

Первым шагом в предобработке текстов является их очистка от шумов, не несущих семантической нагрузки. Для этого были выполнены следующие действия: удаление знаков пунктуации, специальных символов, гиперссылок с использованием регулярных выражений, чисел и эмодзи, если они не несли значимой информации [9].

Все тексты были приведены к нижнему регистру, что позволяет унифицировать слова при последующей обработке и избежать дублирования лексем, различающихся только по регистру.

Следующим этапом является токенизация — процесс разбиения текста на отдельные слова (токены). Для этого использовались стандартные инструменты библиотеки nltk [10].

После токенизации из текстов были удалены стоп-слова — наиболее частотные слова в русском языке, не несущие значимой смысловой нагрузки. Удаление стоп-слов позволило сократить количество признаков в векторном пространстве и повысить значимость оставшихся слов.

Для нормализации форм слов применялась лемматизация — приведение слов к их начальной (словарной) форме. Основным инструментом для этого выступила библиотека rymorphu2, ориентированная на морфологический анализ русского языка [11].

Лемматизация позволяет сократить размер словаря признаков и устранить избыточность, вызванную грамматическими формами одного и того же слова.

Результатом предобработки стало создание новой колонки `cleaned_text` в каждом из датафреймов.

Для приведения текстов к числовому формату, пригодному для обучения моделей, применялась модель TF-IDF [12]. Этот метод позволяет оценить важность каждого слова в документе относительно корпуса текстов.

В результате применения TF-IDF каждый текст был представлен в виде разреженного вектора, где каждый элемент соответствует весу конкретного термина. Эти векторы стали входными признаками для дальнейшего обучения моделей классификации.

АНАЛИЗ ТОНАЛЬНОСТИ ТЕКСТА

Анализ тональности текстов — важный этап обработки пользовательского контента, позволяющий определить эмоциональную окраску сообщений. В настоящей работе были рассмотрены и обучены как классические алгоритмы машинного обучения, так и современные методы глубокого обучения [13, 14]. В данном разделе описывается архитектура использованных моделей, методы оценки и выбор финального решения на основе сравнительного анализа.

Для определения наиболее эффективной модели было проведено сравнение пяти алгоритмов: логистической регрессии, случайного леса, наивного байесовского классификатора, LSTM, Transformers (табл. 5) [15, 16].

Сравнение производилось на основе таких метрик, как точность (Accuracy), точность классификации (Precision), полнота (Recall), F1-мера (F1-Score) и площадь под кривой ошибок (ROC-AUC). В результате анализа наилучшие характеристики продемонстрировал наивный байесовский классификатор.

Он был выбран в качестве основной модели по следующим причинам:

- лучшая полнота (Recall) — важна в задачах, где необходимо минимизировать пропуск определенного класса (например, негативных комментариев);
- сбалансированное поведение по точности и F1-мере, сопоставимое с другими моделями;
- простота и высокая скорость обучения;
- стабильность при обработке разреженных данных [17].

Выбранная модель была применена к очищенным текстам постов из датасета `df_all_posts`. В результате каждому посту была присвоена оценка его эмоциональной окраски, которая добавлялась в датасет как новая колонка «Индекс эмоциональной оценки». Для дальнейшего анализа был рассчитан средний индекс эмоциональной оценки для каждого пользователя на основе его последних

**Сравнительная характеристика моделей машинного обучения /
Comparative Characteristics of Machine Learning Models**

Модель / Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Логистическая регрессия	0,73	0,73	0,73	0,73	0,81
Случайный лес	0,72	0,72	0,74	0,73	0,80
Наивный байесовский классификатор	0,73	0,74	0,72	0,73	0,81
LSTM	0,69	0,69	0,71	0,70	0,77
Transformers	0,71	0,72	0,70	0,71	0,78

Источник / Source: составлено авторами / Compiled by the authors.

пяти постов, результаты расчетов были добавлены в датасет df_all_users. Это позволило выявить общие тренды в эмоциональном состоянии пользователей и оценить динамику изменения настроения в социальной сети (рис. 1).

Визуализация эмоциональной окраски текста постов пользователей помогла проанализировать данные, по которым можно сделать вывод, что 11 885 постов содержат нейтральную окраску текста, 7 249 — позитивную и 655 — негативную.

Обученная модель наивного Байеса была применена к векторизованным комментариям для их классификации на положительные, негативные

и нейтральные. Результаты классификации были визуализированы для наглядности распределения эмоциональных окрасок [18]. Это позволило оценить общее настроение и реакции пользователей (рис. 2).

В результате анализа визуализации эмоциональной окраски текста комментариев можно сделать вывод, что 54% комментариев имеют нейтральную окраску текста, 42% комментариев — положительную и 4% — негативную.

Применение наивного байесовского классификатора к комментариям демонстрирует его эффективность в определении эмоциональных

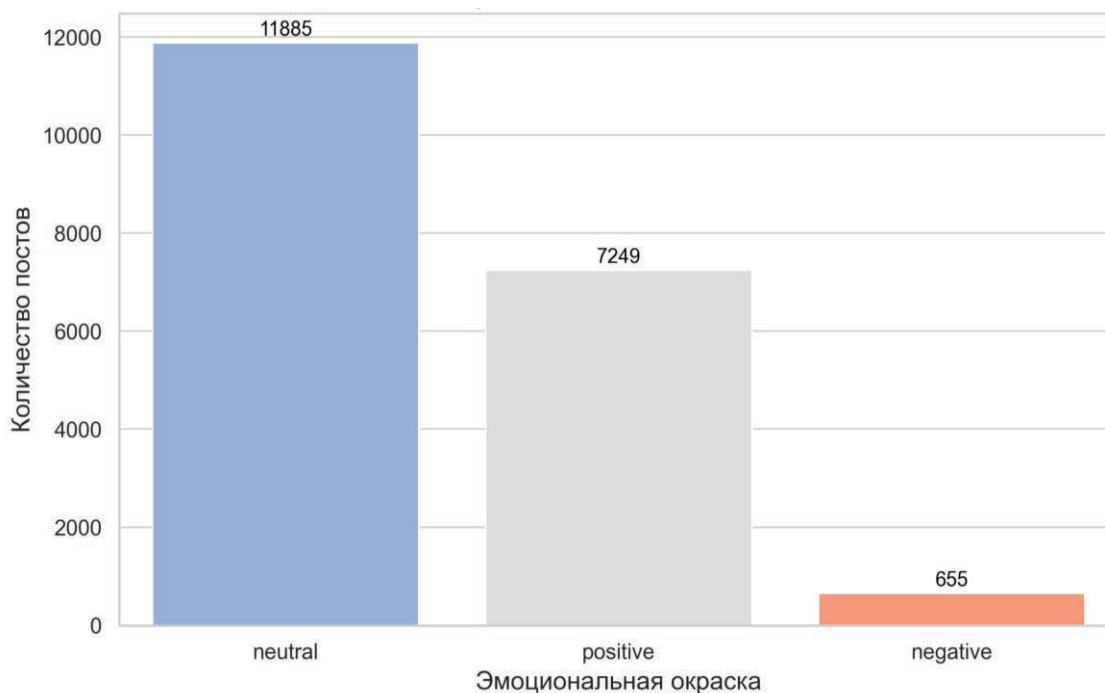


Рис. 1 / Fig. 1. Эмоциональная окраска текста пользователей / Sentiment of User-generated Text

Источник / Source: составлено авторами / Compiled by the authors.

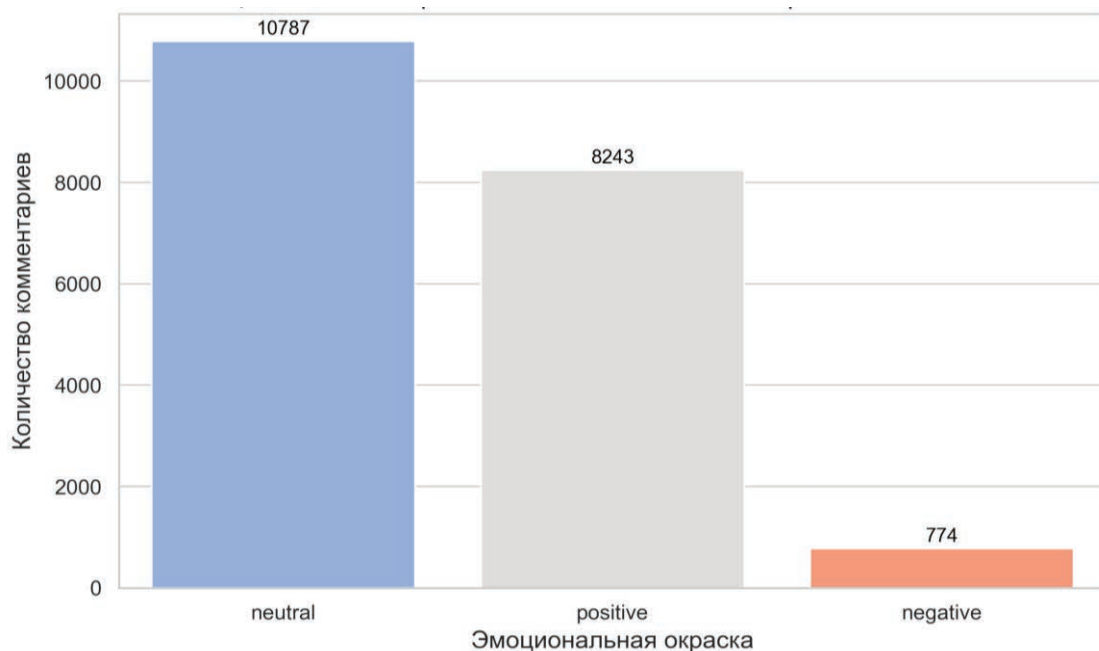


Рис. 2 / Fig. 2. Эмоциональная окраска текста всех комментариев пользователей / Sentiment Analysis of All User Comments

Источник / Source: составлено авторами / Compiled by the authors.

тенденций. Полученные данные могут служить основой для стратегий взаимодействия и контент-планирования.

ВЫВОДЫ

Результаты автоматизированного анализа тональности текстов позволяют не только оценить общее настроение пользователей социальной сети, но и выявить потенциальные поведенческие и психологические особенности их взаимодействия в цифровом пространстве.

На основе классификации тональности, проведенной с использованием модели наивного байесовского классификатора, были получены следующие распределения по эмоциональной окраске текстов.

Посты пользователей (всего 21 989):

- нейтральные — 11 885 постов (54%);
- положительные — 7 249 постов (42%);
- негативные — 655 постов (4%).

Комментарии пользователей (всего 17 904):

- нейтральные — 10 787 комментариев (60%);
- положительные — 8 243 комментария (37%);
- негативные — 774 комментария (3%).

Такая картина позволяет говорить о доминировании нейтрального и позитивного контента в обоих типах пользовательских сообщений.

- Пользователи социальной сети ВКонтакте в большей степени склонны к нейтральной или положительной тональности как в публикациях, так и в комментариях.

- Доля негативных сообщений во всех типах контента не превышает 4%, что может свидетельствовать о контролируемом 62-м уровне эмоциональной агрессии и высокой социальной нормативности в выражении мыслей.

- Посты оказываются чуть более позитивными, чем комментарии, где наблюдается большая доля нейтрального высказывания — это может указывать на более сдержанный стиль коммуникации при обсуждении чужих сообщений.

Таким образом, проведенный анализ не только выявил структуру эмоциональной окраски контента, но и открыл перспективы для дальнейших междисциплинарных исследований — в области социологии, психологии и цифровой лингвистики.

СПИСОК ИСТОЧНИКОВ

1. Rodríguez-Ibáñez M., Casanez-Ventura F., Castejón-Mateos F., Cuenca-Jiménez P.-M. A review on sentiment analysis from social media platforms. *Expert Systems with Applications*. 2023;223:119862. DOI: 10.1016/j.eswa.2023.119862
2. Wankhade M., Rao A.C.S., & Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*. 2022;55(7):5731–5780. DOI: 10.1007/s10462-022-10144-1

3. Cortis K., Davis B. Over a Decade of Social Opinion Mining: A Systematic Review. *Artificial Intelligence Review*. 2021;54(1):4873–4965. DOI: 10.1007/s10462-021-10030-2
4. Mutanov G., Karyukin A., Mamykova G. Multi-Class Sentiment Analysis of Social Media Data with Machine Learning Algorithms. *Computers, Materials & Continua*. 2021;69(1):913–930. DOI: 10.32604/cmc.2021.017827
5. Salman I.K., Feizi Derakhshi M.R., Pashazadeh S., Asadpour M. A Comprehensive Review of Visual-Textual Sentiment Analysis from Social Media Networks. *ArXiv preprint*. 2022;arXiv:2207.02160. DOI: 10.48550/arXiv.2207.02160
6. Zhou, H. Research of text classification based on TF-IDF and CNN-LSTM. *Journal of Physics: Conference Series*. 2022;2171:012021. DOI: 10.1088/1742-6596/2171/1/012021
7. Oliveira D.F., Nogueira A., Brito M. Performance comparison of machine learning algorithms in classifying information technologies incident tickets. *AI*. 2022;3(3):601–622. DOI: 10.3390/ai3030035
8. Smetanin S. The applications of sentiment analysis for Russian language texts: current challenges and future perspectives. *IEEE Access*. 2020;8:110693–110719. DOI: 10.1109/ACCESS.2020.3002215
9. Braga M., Milanese G.C., Pasi G. Investigating large language models' linguistic abilities for text preprocessing. *arXiv preprint*. 2025;arXiv:2510.11482. DOI: 10.48550/arXiv.2510.11482
10. Feng J.H., Mohaghegh M. Hybrid model of data augmentation methods for text classification task. Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC 3K 2021). 2021:194–197. DOI: 10.5220/0010688500003064
11. Гадасин Д.В., Пак Е.В., Коровушкина В.М., Мелькова Е.К. Предобработка текстовой информации на основе термов естественного языка. *REDS: Телекоммуникационные устройства и системы*. 2022;1:4–12. URL: <https://www.elibrary.ru/pdgavp>
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. *arXiv preprint*. 2019;arXiv:1907.11692. DOI: 10.48550/arXiv.1907.11692
13. Щекотин Е.В., Гойко В.Л., Басина П.А., Бакулин В.В. Использование машинного обучения для изучения качества жизни населения: методологические аспекты. *Цифровая социология*. 2022;5(1):87–97. DOI: 10.26425/2658-347X-2022-5-1-87-97
14. Гальченко Ю.В., Нестеров С.А. Классификация текстов по тональности ML-методами. Системный анализ в проектировании и управлении. Сборник научных трудов XXVI Международной научно-практической конференции. В 3 ч. Ч. 3. Санкт-Петербург, 13–14 октября 2023 г. СПб.: ПОЛИТЕХ-ПРЕСС. 2023;26(3):369–378. DOI: 10.18720/SPBPU/2/id23-501
15. Мезенев К.А., Бадрызлова Ю.Г. Анализ эмоциональной тональности русскоязычных текстов с цифровыми методами. НИУ ВШЭ, магистерская диссертация. Москва, 2025. URL: <https://www.hse.ru/edu/vkr/1055012487>
16. Катермина Т.С., Тагиров К.М., Тагиров Т.М. Элементы ИИ в анализе текстов: LSTM-приложение к Вконтакте. *Computational Nanotechnology*. 2022;9(2):35–44. DOI: 10.33693/2313-223X-2022-9-2-35-44.
17. Челышев Э.А., Оцоков Ш.А., Раскатова М.В., Щёголев П. Сравнение методов классификации русскоязычных новостных текстов с использованием алгоритмов машинного обучения. *Вестник кибернетики*. 2022;1(45):63–71. DOI: 10.34822/1999-7604-2022-1-63-71.
18. Ивахин Д.Е., Андиева Е.Ю. Автоматический анализ текста для выявления профессиональных навыков: гибридный подход на основе TF-IDF и нейросетевых эмбедингов. *Вестник науки*. 2025;4(85-2):685–692. URL: <https://www.вестник-науки.пф/article/22263>

REFERENCES

1. Rodríguez-Ibáñez M., Casanez-Ventura F., Castejón-Mateos F., Cuenca-Jiménez P.-M. A review on sentiment analysis from social media platforms. *Expert Systems with Applications*. 2023;223:119862. DOI: 10.1016/j.eswa.2023.119862
2. Wankhade M., Rao A.C.S., & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*. 2022;55(7):5731–5780. DOI: 10.1007/s10462-022-10144-1
3. Cortis K., Davis B. Over a Decade of Social Opinion Mining: A Systematic Review. *Artificial Intelligence Review*. 2021;54(6):4873–4965. DOI: 10.1007/s10462-021-10030-2
4. Mutanov G., Karyukin A., Mamykova G. Multi-Class Sentiment Analysis of Social Media Data with Machine Learning Algorithms. *Computers, Materials & Continua*. 2021;69(1):913–930. DOI: 10.32604/cmc.2021.017827
5. Salman I.K., Feizi Derakhshi M.R., Pashazadeh S., Asadpour M. A Comprehensive Review of Visual-Textual Sentiment Analysis from Social Media Networks. *ArXiv preprint*. 2022;arXiv:2207.02160. DOI: 10.48550/arXiv.2207.02160

6. Zhou, H. Research of text classification based on TF-IDF and CNN-LSTM. *Journal of Physics: Conference Series*. 2022;2171:012021. DOI: 10.1088/1742-6596/2171/1/012021
7. Oliveira D.F., Nogueira A., Brito M. Performance comparison of machine learning algorithms in classifying information technologies incident tickets. *AI*. 2022;3(3):601–622. DOI: 10.3390/ai3030035
8. Smetanin, S. The applications of sentiment analysis for Russian language texts: current challenges and future perspectives. *IEEE Access*. 2020;8:110693–110719. DOI: 10.1109/ACCESS.2020.3002215
9. Braga M., Milanese G.C., Pasi G. Investigating large language models' linguistic abilities for text preprocessing. *arXiv preprint*. 2025;arXiv:2510.11482. DOI: 10.48550/arXiv.2510.11482
10. Feng J.H., Mohaghegh M. Hybrid model of data augmentation methods for text classification task. In Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC 3K 2021). 2021:194–197. DOI: 10.5220/0010688500003064
11. Gadasin D.V., Pak E.V., Korovushkina V.M., Melkova E.K. Natural Language Term-Based Text Information Preprocessing. *REDS: Telecommunication Devices and Systems*. 2022;1:4–12. URL: <https://www.elibrary.ru/pdgavp> (In Russ.).
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. *arXiv preprint*. 2019; arXiv:1907.11692. DOI: 10.48550/arXiv.1907.11692
13. Shchekotin E.V., Goiko V.L., Basina P.A., Bakulin V.V. Using machine learning to study the population life quality: methodological aspects. *Digital Sociology*. 2022;5(1):87–97. (In Russ.). DOI: 10.26425/2658-347X-2022-5-1-87-97
14. Galchenko Yu.V., Nesterov S.A. Sentiment analysis with machine learning methods. Systems Analysis in Design and Management. Proc. Of the XXVI International scientific conference, St. Petersburg, October 13–14, 2023. St. Petersburg: Politekh-Press; 2023;3:369–378. (In Russ.). DOI: 10.18720/SPBPU/2/id23-501
15. Mezenev K.A., Badryzlova Yu.G. Sentiment Analysis of Russian-Language Texts Using Digital Methods. Master's Thesis. National Research University Higher School of Economics (HSE), Moscow. 2025. URL: <https://www.hse.ru/edu/vkr/1055012487> (In Russ.).
16. Katermina T.S., Tagirov K.M., Tagirov T.M. Elements of artificial intelligence in solving problems of text analysis. *Computational Nanotechnology*. 2022;9(2):35–44. (In Russ.). DOI: 10.33693/2313-223X-2022-9-2-35-44
17. Chelyshev E.A., Otsokov Sh.A., Raskatova M.V., Shchegolev P. Comparing classification methods for news texts in russian using machine learning algorithms. *Proceedings of Cybernetics*. 2022;1(45):63–71. (In Russ.). DOI: 10.34822/1999-7604-2022-1-63-71
18. Ivakhin D.E., Andieva E. Yu. Automatic text analysis for identifying professional skills: a hybrid approach based on TF-IDF and neural network embeddings. *Bulletin of Science*. 2025;4(85):685–692. (In Russ.). URL: <https://www.vestnik-nauki.com/article/22263>

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Екатерина Александровна Горбунова — старший разработчик программного обеспечения ООО «Лаборатория систем автоматизации процессов», Санкт-Петербург, Российская Федерация

Ekaterina A. Gorbunova — Senior Software Developer, LLC “Laboratory of Process Automation Systems”, Saint Petersburg, Russian Federation
<https://orcid.org/0009-0009-5748-9031>
kateswep@mail.ru

Расул Ахматович Кочкаров — кандидат экономических наук, доцент кафедры искусственного интеллекта факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Rasul A. Kochkarov — Cand. Sci. (Econ.), Assoc. Prof. of Artificial Intelligence Department, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation
<https://orcid.org/0000-0003-3186-3901>
rkochkarov@fa.ru

Эвелина Александровна Окунева — ассистент кафедры математики и анализа данных факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация



Evelina A. Okuneva — Assistant of the Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0009-0006-4385-4462>

Автор для корреспонденции / Corresponding author:

eaokuneva@fa.ru

Заявленный вклад авторов:

Е.А. Горбунова — разработка общей концепции статьи, разработка архитектуры эксперимента, интерпретация результатов анализа тональности и поведенческих паттернов пользователей, сбор и анализ данных, визуализация результатов.

Р.А. Кочкаров — разработка методологии, общая редакция статьи.

Э.А. Окунева — разделы введение и выводы, список литературы.

Authors' declared contributions:

E.A. Gorbunova — development of the general concept of the article, design of the experimental architecture, interpretation of sentiment analysis results and user behavioral patterns, data collection and analysis, visualization of results.

R.A. Kochkarov — development of methodology, overall editing of the article.

E.A. Okuneva — writing the Introduction and Conclusion sections, compiling the reference list.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила в редакцию 13.10.2025; принята к публикации 24.11.2025.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was submitted on 13.10.2025; accepted for publication on 24.11.2025.

The authors read and approved the final version of the manuscript.