

Разработка математической модели процессов исследования обращения граждан

Е. Н. Ванчикова¹, А. Н. Тимофеев², Н. Б. Садуев¹, Е. О. Ванзатова¹

¹Бурятская государственная сельскохозяйственная академия имени В. Р. Филиппова
²ООО «СибДиджитал»

Аннотация: В настоящее время граждане активно участвуют в общественных обсуждениях, влияющих на жизнь общества. С развитием цифровизации пользователи всё чаще используют социальные сети и платформы обратной связи (далее – ПОС) для высказывания мнений по различным сферам деятельности. В связи с этим возникает необходимость анализа общественных настроений с помощью современных аналитических инструментов с целью выявления потенциальных проблем на ранних стадиях. В статье описана разработка математической модели для анализа общественных настроений на основе текстовых сообщений, опубликованных в социальных сетях и других источниках. Модель включает в себя этапы предобработки данных, кластеризации обращений, именовании кластеров с использованием языковых моделей и обобщения выявленных проблем. Приведены результаты экспериментальной проверки модели на реальных данных обращений граждан Республики Бурятия, демонстрирующие ее практическую применимость и оценку точности.

Ключевые слова: математическая модель, машинное обучение, нейронные сети, обращения граждан, анализ общественных настроений.

Для цитирования: Ванчикова Е. Н., Тимофеев А. Н., Садуев Н. Б., Ванзатова Е. О. Разработка математической модели процессов исследования обращения граждан // Вестник СибГУТИ. 2026. Т. 20, № 1. С. 39–56. <https://doi.org/10.55648/1998-6920-2026-20-1-39-56>.

Финансирование: Работа выполнена при финансовой поддержке республиканского бюджета Республики Бурятия, проект №25-28-20352 «Применение методов машинного обучения для оценки эффективности реализации проектов по программе «Комплексное развитие сельских территорий на основе мониторинга изменений актуальных проблем сельских жителей» и НИР «Построение математической модели для исследования общественных настроений их текстовых сообщений, опубликованных в социальных сетях на основе цифровых следов жителей Республики Бурятия».



Контент доступен под лицензией
Creative Commons Attribution 4.0
License

© Ванчикова Е. Н., Тимофеев А. Н.,
Садуев Н. Б., Ванзатова Е. О., 2026

Статья поступила в редакцию 09.12.2025;
переработанный вариант – 11.12.2025;
принята к публикации 13.01.2026.

1. Введение

Современное общество активно использует цифровые технологии для выражения своих мнений, жалоб и предложений, особенно через социальные сети и официальные платформы обращений [19]. Это создает огромный объем текстовых данных, которые могут быть использованы для анализа общественных настроений и выявления актуальных проблем. Однако обработка таких данных вручную становится неэффективной из-за их масштаба и слож-

ности. Резкое увеличение объема обращений граждан потребовало проведения постоянных исследований по автоматической классификации обращений граждан из-за затрат времени, усилий и ошибок неправильной классификации, связанных с ручной классификацией [10]. В связи с этим возникает необходимость в разработке автоматизированных систем, основанных на технологиях искусственного интеллекта, способных анализировать текстовые сообщения, классифицировать их, прогнозировать тенденции и обобщать ключевые проблемы.

Актуальность исследования обусловлена необходимостью повышения качества взаимодействия между гражданами и государственными структурами, а также улучшения оперативности реагирования на возникающие проблемы. Республика Бурятия, как регион с уникальными социально-экономическими условиями, представляет собой важный объект для изучения общественных настроений через анализ текстовых сообщений.

Цель работы заключается в разработке и исследовании математической модели для анализа общественных настроений на основе текстовых сообщений, опубликованных в социальных сетях и других источниках.

Задачи исследования включают:

1. анализ и подготовку предоставленных датасетов текстовых обращений граждан за исторический период (2023–2025 гг.);
2. выбор методов и технологий искусственного интеллекта для решения задач кластеризации, классификации, прогнозирования и обобщения обращений граждан;
3. построение математической модели для исследования общественных настроений на основе текстовых сообщений;
4. экспериментальная проверка модели, оценка ее эффективности и сравнение с существующими подходами.

Объект исследования – текстовые сообщения граждан, опубликованные в социальных сетях и других источниках обращений (ИМ, ПОС, БурятИнфо).

Предмет исследования – методы и средства кластеризации, прогнозирования и обобщения обращений граждан на основе технологий искусственного интеллекта.

Научная новизна работы заключается в разработке комплексной математической модели, объединяющей методы предобработки текстов, кластеризации, именования кластеров с использованием языковых моделей и генерации обобщенных формулировок проблем.

Практическая значимость исследования заключается в возможности применения разработанной модели для анализа обращений граждан в Республике Бурятия, что позволит повысить эффективность взаимодействия между гражданами и органами власти, а также улучшить качество предоставляемых услуг.

Таким образом, данное исследование направлено на создание инструмента для автоматизированного анализа общественных настроений, что имеет важное значение для повышения качества жизни населения и оперативности реагирования на возникающие проблемы.

2. Описание проблематики и постановка задачи

2.1. Проблематика исследования и решаемые задачи

В рамках исследования рассматривается построение математической модели для анализа общественных настроений на основе текстовых сообщений, опубликованных в социальных сетях жителями Республики Бурятия. Для этого предполагается:

1. разработка программного комплекса, обеспечивающего автоматический сбор и загрузку данных источников обращений граждан, а также их последующую обработку, включая очистку, нормализацию и интеграцию в единую информационную базу;
2. разработка алгоритма выявления новых признаков обращений на основе исходных данных. Это позволит более точно анализировать настроения населения путем расширения существующего классификатора подкатегорий обращения граждан. Методы машинного

обучения, включая алгоритмы кластеризации и классификации, будут использованы для выявления групп жителей с похожими проблемами;

3. разработка метода прогнозирования тематик обращения граждан, что может позволить улучшить качество оказываемых услуг;

4. автоматизация процесса обобщения содержания тестовых обращений граждан для генерации общей проблемы в виде текста на естественном языке на основе применения генеративных технологий запросов через API запросы к языковой модели.

Все описанные выше задачи напрямую связаны с применением технологий искусственного интеллекта, которые более подробно рассмотрим далее.

2.2. Применение искусственного интеллекта для решения поставленных задач

Создание классификатора обращений граждан на основе искусственного интеллекта (ИИ) представляет собой важный шаг к повышению эффективности и качества обслуживания. ИИ, используя алгоритмы машинного обучения, способен адаптироваться к новым данным и выявлять сложные паттерны в обращениях. Это позволяет не только точно классифицировать запросы, но и учитывать контекст каждого обращения, что значительно улучшает понимание намерений граждан. Методы обработки естественного языка (NLP) обеспечивают возможность анализа текстовых обращений, извлекая ключевую информацию и автоматически распределяя их по категориям.

Кроме того, ИИ позволяет обрабатывать большие объемы данных, что делает систему масштабируемой и способной выявлять тренды и аномалии в обращениях. Постоянное обучение моделей на новых данных способствует повышению их актуальности и точности. Использование предсказательной аналитики помогает не только в выявлении потребностей граждан, но и в создании персонализированных рекомендаций, что делает взаимодействие более целенаправленным и эффективным.

Наконец, интеграция ИИ-классификатора с другими системами управления обращениями обеспечивает упрощение взаимодействия и обмен данными между различными государственными и частными структурами. Это создает единое информационное пространство, позволяющее оперативно реагировать на возникающие проблемы и повышать уровень удовлетворенности граждан. В итоге, использование ИИ в классификации обращений граждан не только оптимизирует процессы, но и способствует улучшению качества предоставляемых услуг.

Для решения поставленных задач был проведен анализ публикаций, который показал достаточно большой научный интерес к проблематике данного исследования.

Статья [18] посвящена анализу эмоционального состояния пользователей социальных сетей и применению современных методов обработки текстовых данных. Основное внимание уделяется сравнению традиционных подходов к анализу текста с методами машинного обучения, такими как LSTM и BERT.

В статье «Кластеризация сообществ социальной сети «ВКонтакте» [15] исследуют применение методов кластерного анализа для анализа социальных сетей. В частности, определено оптимальное количество кластеров (от трех до шести) на основе анализа данных, на основе которых выявлены существенные различия в характеристиках участников различных групп, что позволяет лучше понять динамику политизации в социальных сетях. В качестве исходных данных применен датасет в количестве тридцати восьми групп с различной тематикой.

В статье [14] представлен усовершенствованный алгоритм кластеризации текстов общественного мнения в условиях чрезвычайных ситуаций, основанный на глубоком обучении и учитывающий семантическое сходство на уровне предложений. Метод включает оптимизацию хранения данных через гетерогенный ориентированный граф, реконструкцию признакового пространства и динамическое управление семантическими признаками. Интеграция

нечёткой кластеризации (FCM) с глубоким обучением позволила улучшить сходимость (снижение ошибки на 18,7%), скорость обработки (на 32%) и точность (рост F1-меры на 15,2%).

Это позволяет сделать вывод о том, что кластеризация позволяет не только структурировать данные о сообществах, но и выявлять тенденции. Поэтому для решения задачи, связанной с расширением классификатора обращений, можно использовать на первоначальном этапе кластеризацию обращений, для их структурирования и группировки по будущим новым наименованиям подкатегорий.

После получения массива кластеров необходимо выполнить процедуру расширения классификатора новыми подкатегориями. Для этого можно воспользоваться языковыми моделями.

Так, статья [3] посвящена автоматизации классификации обращений граждан на горячую линию Президента РФ с использованием сверточной нейронной сети (CNN). На основе 7000 обращений применены методы предобработки текста (фильтрация, удаление стоп-слов). Для обучения использованы предобученные модели NL BERT и sciBERT. Результаты показывают точность классификации по 15 темам на уровне 86%. Метод рекомендован для внедрения в практику обработки обращений. В статье используются предобученные языковые модели BERT (NL BERT и sciBERT) для обучения классификатора. Эти модели относятся к классу крупных языковых моделей (LLM) и применяются для семантического анализа и категоризации текста.

Исследование [7] посвящено применению сверточных нейронных сетей (CNN) для классификации текстов. На основе экспериментов с большими текстовыми данными показано, что CNN превосходят традиционные методы по точности. Работа анализирует ключевые подходы адаптации архитектур CNN, изначально разработанных для обработки изображений, к задачам работы с текстом. Также в статье предлагается использовать для повышения точности классификации текстов API к известным LLM моделям.

В более современных работах отмечается тренд на использование генеративных LLM не только для классификации, но и для семантического анализа, суммирования и извлечения тем из текстовых корпусов. В частности, подход zero/few-shot обучения позволяет адаптировать LLM к новым доменам без обширной дообучки, что актуально для задач анализа обращений граждан с постоянно эволюционирующей тематикой.

В статье [9] проведен анализ настроений в китайских отзывах в социальных сетях с помощью модели сети с длительной кратковременной памятью (LSTM). Точность предложенной модели достигла приблизительно 92%.

В статье [8] авторы предложили инструмент моделирования распространения влияния на основе LLM под названием LLM-AIDSim, где применили методику моделирования распространения влияния к модели пользовательского агента, управляемой LLM, которая использует реальные социальные сети для прогнозирования поведения пользователей в отношении новостей, идей и тем. Такой подход не только позволяет предвидеть реакцию общественности, но и помогает оптимизировать сообщения, чтобы они лучше соответствовали желаемым результатам.

В ходе исследования [13] была создана система анализа настроений, основанная на машинном обучении, для улучшения мониторинга общественного мнения в социальных сетях в режиме реального времени. Для сложной идентификации настроений предлагаемый подход сочетает в себе передовые архитектуры на основе трансформаторов (DistilBERT, RoBERTa) с традиционными моделями машинного обучения (логистическая регрессия, SVM, наивный Байесовский подход). Система достигла точности до 80-85%, используя трансформаторные модели в реальных сценариях, после тестирования с использованием как методов глубокого обучения, так и стандартных процессов машинного обучения на основе аннотированных наборов данных социальных сетей.

В статье [18] рассмотрена модель использования скелетных структур на базе синтаксической разметки для предобработки корпусов текстов перед передачей в нейросетевые модели машинного перевода с целью повышения качества их работы, реализованная с помощью ча-

стеречной и синтаксической разметок корпусов текстов, использующих языковую модель, с использованием сети BERT и набора правил. Использование предложенной модели совместно с языковыми моделями для машинного перевода Google и Яндекс позволило увеличить качество машинного перевода на 0.1-0.23 пункта по метрикам BLEU и TER с языковыми парами русский-английский и немецкий-английский.

2.3. Описание и подготовка предоставленных датасетов

В качестве исходных данных для классификатора обращения граждан были использованы данные от ИМ, ПОС и БурятИнфо за период 2023-2025 гг.

Данный набор данных представляет собой таблицы, содержащие информацию о различных категориях обращений граждан. В таблицах есть несколько ключевых колонок, включая «Укрупненные категории» и «Подкатегория», которые позволяют структурировать данные и выделять основные темы обращений. Это поможет алгоритму ИИ лучше понимать контекст запросов и классифицировать их по заданным категориям.

В структуре данных выделяются укрупненные категории, подкатегории и маппинг категорий обращений граждан, поступающих в ИМ, ПОС и БурятИнфо:

1. Укрупненные категории – это основные группы, к которым относятся обращения. Например, категории включают «Безопасность», «Благоустройство», «Госуслуги», «ЖКХ» и другие (табл. 1). Каждая категория охватывает широкий спектр вопросов, что позволяет классификатору обрабатывать разнообразные запросы граждан.

2. Подкатегории – это детализация укрупненных категорий, для уточнения типа обращения (табл. 1). Например, в категории «Безопасность» могут быть подкатегории, связанные с Паводками, Пожарами и Роспотребнадзором. Это детализирует данные и поможет алгоритму ИИ более точно определить, к какой именно проблеме относится конкретное обращение. Подкатегории могут быть в дальнейшем использованы для обобщения проблем и выявления наиболее критичной в заданный период времени.

3. Маппинг категорий позволяет создать единый общий классификатор для дальнейшего обобщенного статистического анализа обращений граждан в пределах муниципальных образований и Республики Бурятия в целом.

Таблица 1. Укрупненные категории и подкатегории обращений

№	Укрупненная категория	Подкатегории
1	Здравоохранение	'Здравоохранение. COVID', Медицина, Использование COVID-сертификата, Обращение по проблеме вакцинации или лечения от Коронавируса, Коронавирус COVID-19
2	Образование	Горячее питание для младшеклассников, Образование, Сведения об образовании и достижениях на портале Госуслуг
3	Соц. защита	Соцобслуживание и защита, Выплаты детям от 3 до 7 лет, Выплаты и пособия, Реабилитация и абилитация инвалидов и детей-инвалидов, Снижение бедности, Социальное обслуживание и защита, Опек, дети-сироты
4	Иное	-
5	Энергетика	'Энергетика. Теплоснабжение', 'Энергетика. Электроснабжение', Государственные услуги в сфере энергетики, Газовое оборудование (частный сектор), Ресурсоснабжение. Теплоснабжение
6	Дороги	Автомобильные дороги, Обочины дорог
7	Госуслуги	МФЦ, МОИ документы, Госимущество
8	Транспорт	Авиабилеты, ОбщТранспорт, ЖелТранспорт
9	Благоустройство	Благоустройство, Благоустройство общественных пространств, Дворы и территории общего пользования, Зеленые насаждения, Парки культуры и отдыха, Придомовая территория
10	Безопасность	'Безопасность. КРС', 'Безопасность. Паводки', 'Безопасность. Пожары',

№	Укрупненная категория	Подкатегории
		'Безопасность. Роспотребнадзор', 'Безопасность. Собаки'
11	ЖКХ	'ЖКХ. Водоснабжение'
12	Мусор	Мусор/Свалки/ТКО, Выбросы вредных веществ, Мусор, Обращение с отходами, Обращение с твердыми коммунальными отходами, Сбор и уничтожение биологических отходов, Стихийные свалки
13	СВО, мобилизация	Военная служба, Обращения военнослужащих мобилизованных и их семей, Проблемы приграничных и новых территорий, Ветераны боевых действий, Обжалование решений по воинскому учету
14	Экология, недропользование, лесное хозяйство	Животный и растительный мир, Информация о загрязнении окружающей среды, Иные вопросы в сфере экологии, Красная книга и особо охраняемые природные территории, Природа, Экология, Государственные услуги в сфере недропользования, Лесное хозяйство, Недропользование, Сохранение лесов, Уголь, Водный объект, Сохранение лесов, Экологическая экспертиза объекта
15	Культура	Государственные услуги в сфере культуры, Культура, Пушкинская карта, Библиотека, Дом культуры, Музей, Театры и концертные организации
16	Труд и занятость	Трудовые права, Охрана труда, Служба занятости
17	Связь и информационные системы	Сотовая связь, Связь и телевидение, Вопросы в сфере IT и связи
18	Физ. культура и спорт	Учреждение допобразования (спорт), Государственные услуги в сфере физ. культуры и спорта, Иные вопросы в сфере физкультуры и спорта, Спортивное мероприятие, Спортивные объекты, Физическая культура и спорт, Комплекс ГТО
19	Экономика	'Экономика, промышленность, торговля, предпринимательство', Экономика и бизнес, Государственная поддержка, Государственная поддержка бизнеса, Государственное управление, Государственные услуги в сфере промышленности и торговли, Иные вопросы в сфере промышленности и торговли, Меры поддержки, Общественное питание, Промышленность, Торговля, Экономика и бизнес, Экономическое сотрудничество, Международная торговля
20	С/х и продовольствие	Производство с/х продукции, Земля для с/х, Сельское хозяйство и охота
21	Строительство	Долевое строительство, Жилищная политика, Нарушения при строительстве, Правила землепользования и застройки территории, Строительство школ, детских садов
22	Общ-полит. Вопросы общие	Ветераны Великой Отечественной войны, Внутренняя политика, Межнациональные отношения
23	Туризм	Государственные услуги в сфере туризма, Туристический кешбэк
24	Категории существующие	<i>Новые подкатегории</i>

Для разработки классификатора обращений граждан на основе ИИ, данные из этого набора могут быть использованы для обучения модели. Алгоритмы машинного обучения, такие как Naïve Bayes, Decision Trees или более сложные нейронные сети, могут быть обучены на этих данных для автоматической классификации новых обращений.

Набор данных может быть использован для создания классификатора обращений граждан.

Одной из задач исследования является расширение существующего классификатора обращения граждан в части создания новых подкатегорий внутри укрупненных категорий. В качестве основных методов искусственного интеллекта можно использовать методы кластеризации обращений граждан для категоризации обращений, а также метод классификации полученных категорий через API к LLM для именованя кластеров.

2.4. Предварительный анализ данных для обучения моделей машинного обучения и определения структуры базы данных

Предварительный анализ данных сделан по источникам обращений: ИМ, ПОС и Бурят Инфо (табл.2).

Таблица 2. Предварительный анализ данных по источникам обращений

№	Источник	Анализ структуры данных	Комментарий
1	ИМ	Есть дата и текст обращения. Можно определить адресную часть. Указана категория и подкатегория.	Наиболее подходит для обучения ММО
2	ПОС	Есть дата и обобщенная проблематика оператором (текста обращения нет). Место инцидента непонятно. Указана категория и подкатегория.	Не подходит для обучения
3	БурятИнфо	Есть дата и текст обращения. Можно определить адресную часть: район Бурятии. Указана категория.	Подходит для обучения ММО

Структура данных от ИМ приведена в таблице 3.

Таблица 3. Структура данных ИМ

Поле	Тип данных	Описание
ID	Целое число	Уникальный идентификатор инцидента
Номер инцидента	Строка	Номер, присвоенный инциденту
Отдел	Строка	Отдел, к которому относится инцидент
Исполнитель	Строка	Имя или идентификатор исполнителя инцидента
Общее время первого ответа	Время	Общее время, затраченное на первый ответ
Общий текст первого ответа	Текст	Текст первого ответа на инцидент
Время с даты создания инцидента до времени первого ответа	Время	Время от создания инцидента до первого ответа
Время с даты создания инцидента до времени первого ответа, рабочее время	Время	Время, учитывающее только рабочие часы
Время назначения последнего исполнителя	Время	Время, когда был назначен последний исполнитель
Время 1го ответа ПИ	Время	Время первого ответа от исполнителя
Первая реакция с момента назн. ПИ	Время	Время, прошедшее с момента назначения исполнителя до первого ответа.
Уложились или нет после назн. ПИ	Логический	Флаг, указывающий, уложились ли в сроки после назначения исполнителя.
Время с начала создания инцидента до окончания	Время	Общее время от создания инцидента до его закрытия.
Время отправки первого ответа ПИ на согласование	Время	Время, когда первый ответ был отправлен на согласование.
Время, которое первый ответ ПИ находился на согласовании	Время	Время, в течение которого ответ находился на согласовании.
Текущий шаг инцидента	Строка	Текущий статус или шаг инцидента в процессе обработки.
Дата создания	Дата	Дата, когда был создан инцидент.
Дата окончания	Дата	Дата, когда инцидент был закрыт.
Группа тем	Строка	Группа, к которой относится тема инцидента.
Тема	Строка	Конкретная тема инцидента.
Регион	Строка	Регион, в котором произошел инцидент.
Муниципалитет	Строка	Муниципалитет, к которому относится инцидент.
Населенный пункт	Строка	Населенный пункт, где произошел инцидент.

Улица	Строка	Улица, на которой произошел инцидент.
Дом	Строка	Номер дома, где произошел инцидент.
Объект старый	Строка	Информация о старом объекте, связанная с инцидентом.
Объект	Строка	Информация о новом объекте, связанная с инцидентом.
Тип инцидента	Строка	Классификация типа инцидента.
Итог	Текст	Итоговое решение по инциденту.
Есть медиафайл в ответе	Логический	Флаг, указывающий, есть ли медиафайл в ответе.
Есть медиафайл в сообщении	Логический	Флаг, указывающий, есть ли медиафайл в сообщении.
Урл поста	Строка	Ссылка на пост, связанный с инцидентом.
Урл блога	Строка	Ссылка на блог, связанный с инцидентом.
Текст инцидента	Текст	Подробное описание инцидента.
1ый ответ ПИ	Текст	Первый ответ исполнителя на инцидент.
2ой ответ ПИ	Текст	Второй ответ исполнителя на инцидент.
3ий ответ ПИ	Текст	Третий ответ исполнителя на инцидент.
Причина закрытия	Текст	Причина, по которой инцидент был закрыт.
Аудитория	Строка	Целевая аудитория, к которой относится инцидент.
Вовлеченность	Строка	Уровень вовлеченности пользователей в инцидент.
URL автора	Строка	Ссылка на профиль автора инцидента.
Аккаунт публикации ответа	Строка	Аккаунт, через который был опубликован ответ.
Теги	Список	Теги, связанные с инцидентом.
Примечание	Текст	Дополнительные примечания по инциденту.
Время попадания на шаг оценка удовлетворенности	Время	Время, когда инцидент попал на шаг оценки удовлетворенности.
Опрос отправлен	Логический	Флаг, указывающий, был ли отправлен опрос по удовлетворенности.

Структура данных от БИ приведена в таблице 4.

Таблица 4. Структура данных БИ

Поле	Тип данных	Описание
ID	Целое число	Уникальный идентификатор обращения.
Дата создания	Дата	Дата, когда было создано обращение.
Район	Строка	Район, в котором было сделано обращение.
Категория	Строка	Категория обращения (например, жалоба, предложение, запрос и т.д.).
Обращение	Текст	Подробное описание обращения.

Структура данных от ПОС приведена в таблице 5.

Таблица 5. Структура данных ПОС

Поле	Тип данных	Описание
Номер	Целое число	Уникальный номер сообщения.
Номер ЕПГУ	Строка	Уникальный номер, присвоенный в Едином портале государственных услуг (ЕПГУ).
Источник	Строка	Источник обращения (например, сайт, телефон, личное обращение и т.д.).
Верхнеуровневый ЛКО	Строка	Название региона
Категория	Строка	Основная категория обращения (например, жалоба, запрос, предложение и т.д.).
Подкатегория	Строка	Более детальная подкатегория обращения.
Факт	Текст	Описание фактической ситуации, связанной с обращением.

<i>Организация, в которую поступило сообщение</i>	<i>Строка</i>	<i>Название организации, которая приняла обращение.</i>
Организация, в которой находится сообщение	Строка	Название организации, где в данный момент находится обращение.
<i>Дата поступления</i>	<i>Дата</i>	<i>Дата, когда сообщение было зарегистрировано.</i>
Дата планируемого завершения работ	Дата	Дата, к которой планируется завершение работы по обращению.
Дата фактического завершения работ	Дата	Дата, когда работа по обращению была фактически завершена.
Стадия	Строка	Текущая стадия обработки обращения (например, в работе, завершено, на рассмотрении).
Статус	Строка	Статус обращения (например, открыто, закрыто, ожидает ответа и т.д.).
Просрочено	Логическое значение	Указывает, просрочено ли обращение (да/нет).
Фаст-трек	Логическое значение	Указывает, является ли обращение фаст-треком (да/нет).
ФЗ	Строка	Указание на федеральный закон, относящийся к обращению.
Тип решения	Строка	Тип решения, принятого по обращению (например, положительное, отрицательное).
Направлено по email в ФОИВ, не подключенный к ПОС	Логическое значение	Указывает, было ли сообщение направлено по email в федеральный орган исполнительной власти.
Оценка ответа заявителем	Оценка (1-5)	Оценка, выставленная заявителем в ответ на обработку обращения.
Повторное рассмотрение	Логическое значение	Указывает, было ли обращение направлено на повторное рассмотрение (да/нет).
Организация координатора	Строка	Название организации, назначенной координатором обращения.
ФИО координатора	Строка	Фамилия, имя и отчество координатора.
Организация исполнителя	Строка	Название организации, ответственной за выполнение работы по обращению.
ФИО исполнителя	Строка	Фамилия, имя и отчество исполнителя.
Организация руководителя	Строка	Название организации, в которой работает руководитель.
ФИО руководителя	Строка	Фамилия, имя и отчество руководителя.

Анализ предоставленных данных позволил определить параметры, которые необходимо занести в базу данных для дальнейшего использования: Уникальный идентификатор, Дата, Код укрупненной категории, Источник данных (ИМ, БИ, ПОС), Код источника данных, текст обращения, Адрес.

Эти данные будут использоваться для кластеризации и формирования новых подкатегорий обращения граждан.

3. Математическая модель процессов исследования обращения граждан

На основе детального описания в предыдущей главе можно предложить комплексную математическую модель, охватывающую все этапы исследования. Модель будет состоять из нескольких взаимосвязанных блоков.

Пусть:

D – исходный набор данных (корпус обращений);

d_i – i -ое текстовое обращение в наборе данных;

C – множество существующих укрупненных категорий;

S – множество существующих подкатегорий;

S^* – множество новых (выявляемых) подкатегорий;

S^{**} – множество обобщений по категориям обращений;

P – множество прогнозных значений укрупненных категорий обращений на указанную дату.

Блок 1. Модель предобработки текста и векторного представления

Прежде чем применять алгоритмы ИИ, тестовые сообщения граждан должны быть преобразованы в числовой формат.

Токенизация и очистка: каждое обращение d_i преобразуется в последовательность токенов (слов, подслов) после применения следующих процедур: удаление стоп-слов, приведение к нижнему регистру, лемматизация.

$$d_i \rightarrow (t_1, t_2, \dots, t_n), \quad t_j - j\text{-ый токен.}$$

Векторизация: последовательность токенов преобразуется в числовой вектор v_i посредством построения контекстуальных эмбедингов (модели BERT):

$$v_i = \text{BERT}(d_i), \text{ где}$$

v_i – вектор [CLS]-токена или усредненный вектор выходов модели, который лучше всего captures контекст всего обращения. Таким образом, получаем новое отображение:

$$d_i \rightarrow v_i, \text{ где}$$

e – размерность пространства эмбедингов.

Блок 2. Модель кластеризации для выявления новых подкатегорий классификатора обращения граждан

Цель – разбить множество векторных представлений обращений $\{v_i\}$ на K непересекающихся кластеров, где K может быть известно заранее или найдено автоматически.

На примере модели KMeans задача кластеризации – найти такие K центроидов $\{\mu_1, \mu_2, \dots, \mu_k\}$, чтобы минимизировать сумму квадратов расстояний внутри кластеров:

$$J = \sum_{i=1}^N \sum_{k=1}^K r_{ik} * (v_i - \mu_k)^2, \text{ где}$$

$r_{ik} = 1$, если v_i принадлежит кластеру k , и 0 – иначе.

В результате каждому обращению d_i присваивается метка кластера $l_i \in \{0, 1, \dots, K\}$. Эти кластеры являются прототипами новых подкатегорий S^* .

Блок 3. Модель именованя кластеров с помощью LLM

Цель – классификация полученных кластеров, т.е. генерация уникальных наименований каждого кластера k . Это относится к задаче генерации текста на основе набора документов с применением LLM.

В качестве исходных данных выступает множество исходных текстовых обращений d_i , объединенных с наименованием укрупненной категории c_i и наименованием подкатегории s_i , указанных для d_i текста обращения в исходном датасете:

$$\{d_i + c_i + s_i | l_i = k\}, \text{ принадлежащих кластеру } k.$$

Функция (с использованием LLM через API) может быть описана в виде:

$$\text{Name}_k = \text{LLM}(\text{Prompt1}, \{d_i | l_i = k\}), \text{ где}$$

Prompt1 – это текстовый шаблон: «Нужно разработать классификатор обращений в социальные службы. Для этого на основе текстов обращений была проведена их кластеризация и получены кластеры. Теперь для текстов обращений на основе которых сформированы кластеры сгенерируй краткое название категорий на русском. Учти существующие категории: 'Безопасность', 'Безопасность. КРС', 'Безопасность. Паводки', 'Безопасность. Пожары', 'Безопасность. Роспотребнадзор', 'Безопасность. Собаки', 'Благоустройство', 'Госуслуги', 'Дороги', 'ЖКХ', 'ЖКХ. Водоснабжение', 'Здравоохранение', 'Здравоохранение. COVID', 'Иное', 'Культура', 'Мусор', 'Образование', 'Общ- полит. вопросы', 'Общ- полит. вопросы. СВО, мобилизация', 'С/х и продовольствие', 'Соц. защита', 'Строительство', 'Физ. культура и спорт', 'Транспорт', 'Туризм', 'Экология, недропользование, лесное хозяйство', 'Экономика, промышленность, торговля, предпринимательство', 'Энергетика', 'Энергетика. Теплоснабжение',

'Энергетика. Электроснабжение' Разрешается добавлять новые категории к уже существующим или через символ разделитель "." прописывать подкатегории для общих крупных категорий. Нужно создать идеальный список категорий для работы классификатора».

В результате формируется множество новых подкатегорий:

$$S^* = \{Name_1, \dots, Name_k\}, \text{ где}$$

$Name_i$ – наименование подкатегории классификатора.

Блок 4. Модель обобщения проблемы для каждого кластера

Цель – для кластера обращений $G = \{d_i | l_i = k\}$ сгенерировать связный итоговый текст T , обобщающий проблему.

В качестве исходных данных выступает множество исходных текстовых обращений d_i i -го кластера.

Генеративная модель (LLM через API) может быть описана, как:

$$T = \text{LLM}(\text{Prompt2}, G), \text{ где}$$

Prompt2 – сложный шаблон, например: «Ты аналитик. Проанализируй следующие жалобы граждан: [ТЕКСТЫ]. Сгенерируй сводный отчет на русском языке, который включает: 1) Суть общей проблемы. 2) Основные причины, если они указаны. 3) Типичные локации (населенные пункты, улицы). 4) Предложения по решению. Используй официально-деловой стиль».

В результате формируется множество обобщений:

$$S^{**} = \{T_1, \dots, T_k\}, \text{ где}$$

T_i – текстовое краткое описание проблемы с указанием некоторых деталей при их наличии в d_i обращениях (причина, адрес, организация, персонала и т.п.).

Блок 5. Модель прогнозирования динамики обращений

Цель – спрогнозировать количество обращений $Y_i(t)$ по укрупненным категориям C в будущий момент времени t .

В качестве исходных данных выступает временной ряд:

Y_1, Y_2, \dots, Y_C , где Y_C – количество обращений по каждой укрупнённой категории за 2023-2025 гг. После обучения прогнозной модели и ее применения можно получить множество прогнозируемых значений:

$$P = \{p_1, p_2, \dots, p_k\}, \text{ где}$$

p_i – прогнозируемое количество обращений граждан по множеству укрупненных обращений C .

Последовательность действий по реализации математической модели приведена на рисунке 1.



Рис. 1. Алгоритм работы модели исследования обращений граждан

Предложенная математическая модель является комплексной и каскадной. Выход одного блока (например, векторные представления из Блока 1) становится входом для другого (Блок 2 или 4). Ядром всей системы являются векторные представления текста (эмбеддинги) и генеративные языковые модели (LLM), которые используются как для анализа (кластеризация, классификация), так и для синтеза текста (именование кластеров, генерация обобщений). Эта модель полностью соответствует поставленным в исследовании задачам и описанному алгоритму работы.

4. Эксперименты и результаты

Для проверки работоспособности и оценки эффективности предложенной математической модели был проведен эксперимент на реальных данных обращений граждан, собранных из источника «ИМ» (Инцидент менеджмент). В эксперименте использовалась выборка, содержащая 10 000 текстовых обращений с размеченными категориями.

Эксперимент проводился в соответствии с предложенным алгоритмом (Рис. 1).

На этапе предобработки и векторизации для текстов обращений применялась лемматизация (с помощью `rumorphy2`) и удаление стоп-слов. Векторные представления генерировались с помощью предобученной русскоязычной модели `sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2`, создающей эмбеддинги размерностью 384.

Далее для кластеризации обращений применялся алгоритм K-Means на векторизованных данных. Оптимальное количество кластеров ($K=45$) было определено методом локтя (Elbow method) и анализом силуэтного коэффициента.

Для генерации названий новых подкатегорий использовался API к крупной языковой модели YaLM 2.0 (13B). В промпте передавались примеры текстов из каждого кластера и инструкция, аналогичная описанной в Блоке 3 модели.

По количеству обращений были выбраны 5 наиболее крупных кластеров, для которых с помощью того же API к YaLM 2.0 генерировались сводные отчеты по шаблону из Блока 4.

Для наиболее частой категории «ЖКХ» на основе ежемесячных агрегатов был построена и обучена модель SARIMA (сезонное авторегрессионное интегрированное скользящее среднее) с целью прогнозирования на 3 месяца вперед.

Оценка результатов проводилась по следующим критериям:

- Качество кластеризации: Средний силуэтный коэффициент составил 0.52, что указывает на удовлетворительное разделение данных. Визуализация методом UMAP показала четкие сгущения точек, соответствующие тематическим группам.
- Именованние кластеров: Экспертная оценка сгенерированных LLM названий для 45 кластеров показала, что 78% названий были признаны релевантными и информативными (например, «Незаконная установка рекламных конструкций», «Проблемы с вывозом ТКО в частном секторе», «Низкое качество питьевой воды в многоквартирных домах»). Около 15% потребовали незначительной редакторской правки, и 7% были отклонены как слишком общие.
- Качество обобщения: Сгенерированные текстовые отчеты по кластерам были оценены двумя независимыми экспертами по 5-балльной шкале по критериям: полнота, связность, полезность для принятия решений. Средняя оценка составила 4.1 балла. Пример обобщения для кластера по проблемам дорог: «Основная проблема – разрушение дорожного полотна на улицах Советская и Трудовая в г. Улан-Удэ после весеннего паводка. Обращения указывают на образование ям, что создает аварийные ситуации. Граждане требуют проведения ямочного ремонта. Причина – износ покрытия и воздействие тяжелого транспорта».
- Точность прогнозирования: Прогнозная модель SARIMA для категории «ЖКХ» показала среднюю абсолютную процентную ошибку (MAPE) на тестовой выборке в 12.3%, что является приемлемым результатом для социальных данных.
- Сравнение с существующими подходами:
 - Против подхода на основе чистой кластеризации (K-Means + ручное именованние): Предложенная модель с LLM сократила время на категоризацию и именованние новых тем на ~85%, устранив субъективный фактор при ручной интерпретации кластеров.
 - Против подхода с классификатором на основе BERT (дообученного на размеченных данных, как в [3]): Предложенная модель не требует предварительной разметки для выявления новых тем, что является ключевым преимуществом в условиях появления ранее неизвестных проблем. Однако для окончательной классификации известных категорий гибридный подход (наша кластеризация + дообученный BERT для финального отнесения) может показать лучшие результаты, что является темой дальнейших исследований.

Отметим также некоторые ограничения модели.

Во-первых, кластеризация чувствительна к качеству эмбедингов, которое часто страдает на коротких неформальных текстах, характерных для социальных сетей. Во-вторых, новые уникальные проблемы в единичных обращениях могут оставаться незамеченными. В третьих, процесс генерации наименований и обобщений для сотен кластеров является ресурсоемким и может вызывать задержки, что делает необходимым внедрение механизмов кэширования. И наконец, существует риск, что LLM сгенерирует правдоподобные, но фактологически неточные детали в обобщениях, особенно при работе с противоречивыми или скудными исходными данными. Это требует обязательной дополнительной проверки критически важной информации, такой как адреса или финансовые суммы.

Заключение

Таким образом, в статье описана комплексная математическая модель исследования обращений граждан, состоящая из нескольких взаимосвязанных блоков, включая преобразование текстов, кластеризацию, именованние кластеров с помощью языковых моделей, генерацию текста обобщенной проблемы по каждому кластеру и прогнозирование динамики обращений.

Модель начинается с преобразования текстовых обращений в векторное представление, что позволяет выявить скрытые паттерны через кластеризацию, а затем генерировать уникальные названия для кластеров.

Автоматизация процесса обобщения содержания тестовых обращений граждан для генерации общей проблемы в виде текста на естественном языке на основе применения генеративных технологий запросов через API помогает сформулировать краткую формулировку проблемы, а прогнозирование динамики обращений позволяет предсказывать будущие тенденции.

Каскадный характер модели подчеркивает взаимозависимость блоков и их соответствие задачам исследования, что делает её эффективным инструментом для анализа и классификации обращений граждан.

Предложенная модель демонстрирует высокую практическую применимость для автоматического мониторинга и структурирования обращений граждан. Её ключевое преимущество – способность адаптивно выявлять и формулировать новые темы без постоянного переобучения классификатора, что актуально для динамичной социальной среды.

Литература

1. *Palumbo G., Carneiro D., Guimares M. et al.* Algorithm Recommendation and Performance Prediction Using Meta-Learning // *International Journal of Neural Systems*. 2023. Vol. 33, No. 03. DOI: 10.1142/s0129065723500119.
2. *Brown T. B. et al.* Language Models are Few-Shot Learners // *Advances in Neural Information Processing Systems (NeurIPS)*. 2020. Vol. 33. P. 1877–1901.
3. *Bunova E. V., Serova V. S.* Methodology for solving problems of classification of appeals/requests of citizens to the "hotline" of the President of the Russian Federation // *Вестник ЮУрГУ. Серия: Компьютерные технологии, управление, радиоэлектроника*. 2022. № 2. С. 68–79.
4. *Crossley N.* Social Networks of Meaning and Communication // *American Journal of Sociology*. 2023. Vol. 128, No. 6. P. 1849–1851. DOI: 10.1086/723877.
5. *Makarov I. S., Bagantsova E. R., Iashin P. A. et al.* Development of and research on machine learning algorithms for solving the classification problem in Twitter publications // *Computer Research and Modeling*. 2023. Vol. 15, No. 1. P. 185–195. DOI: 10.20537/2076-7633-2023-15-1-185-195.
6. *Devlin J., Chang M.-W., Lee K., Toutanova K.* BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. 2019. Vol. 1. P. 4171–4186. DOI: 10.18653/v1/N19-1423.
7. *Muhamediyeva D. K., Abdurakhmanova N. N., Mirzayeva N. S.* Using Conventional Neural Networks for the Problem of Text Classification // *Central Asian Academic Journal of Scientific Research*. 2021. Vol. 1, No. 1. P. 64–71.
8. *Zhang L., Hu Yu., Li W. et al.* LLM-AIDSim: LLM-Enhanced Agent-Based Influence Diffusion Simulation in Social Networks // *Systems*. 2025. Vol. 13, No. 1. P. 29. DOI: 10.3390/systems13010029.
9. *Lukai Wang, Lei Wang.* A Case Study of Chinese Sentiment Analysis of Social Media Reviews Based on LSTM // *SHS Web of Conferences*. 2023. Vol. 157. 04012. DOI: 10.1051/shsconf/202315704012.
10. *Kim N., Hong S.* Automatic classification of citizen requests for transportation using deep learning: Case study from Boston city // *Information Processing & Management*. 2021. Vol. 58, Issue 1. 102410. DOI: 10.1016/j.ipm.2020.102410.
11. *Narayan R., Samanta P.* A Machine Learning Approach for Sentiment Analysis Using Social Media Posts // *International Journal of Information Technology and Computer Science*. 2024. Vol. 16, No. 5. P. 23–35. DOI: 10.5815/ijitcs.2024.05.02.

12. Cronin R. M., Fabbri D., Denny J. C. et al. A comparison of rule-based and machine learning approaches for classifying patient portal messages // *International Journal of Medical Informatics*. 2017. Vol. 105. P. 110–120. DOI: 10.1016/j.ijmedinf.2017.06.004.
13. Tolebay A. N. Sentiment Analysis of Texts from Social Networks Based on Machine Learning Methods for Monitoring Public Sentiment // *Bulletin of Science*. 2025. No. 4 (85). P. 123–134.
14. Yang Y. F., Zhou H., Zhou B. Data clustering based on the modified relaxation Cheeger cut model // *Computational and Applied Mathematics*. 2022. Vol. 41, No. 1. P. 55–62.
15. Анохина И. Ю., Рощина Е. В. Кластеризация сообществ социальной сети "ВКонтакте" // *Информатика и кибернетика*. 2018. № 2(12). С. 34–42.
16. Евдошенко О. И., Евдошенко Н. И. Разработка модели автоматического классификатора обращений граждан // *Прикаспийский журнал: управление и высокие технологии*. 2024. № 2(66). С. 56–68.
17. Мартышкин А. И., Бритов В. С., Дроздова Ю. Р. Возможности выявления эмоционального состояния пользователей социальных сетей с использованием методов машинного обучения // *Современные информационные технологии*. 2024. № 40(40). С. 25–27.
18. Мыльникова А. В., Мыльников Л. А. Языковые модели для предобработки текстов в машинном переводе // *Научно-техническая информация. Серия 2: Информационные процессы и системы*. 2025. № 7. С. 32–44. DOI: 10.36535/0548-0027-2025-07-3.
19. Наумова Н. В., Абрамков А. О. К вопросу об использовании искусственного интеллекта в сфере работы с обращениями граждан // *Гуманитарные, социально-экономические и общественные науки*. 2022. № 9. С. 112–118.
20. Овездурдыева И. К., Гараджаева Д. Я. Анализ социальных сетей с помощью искусственного интеллекта: современные подходы и перспективы // *Наука и мировоззрение*. 2025. № 55. С. 45–52.
21. Saifulla S., Drezhevski R., Dwiyanto F. A., Aribowo A. S., Fauzia Y. Analysis of sentiment using a feature-based machine learning approach to detect anxiety // *Computational Science – ICCS 2023* / ed. J. Mikishka, C. de Mulatier, M. Pashinsky et al. Cham: Springer, 2023. P. 123–135. (Lecture Notes in Computer Science; Vol. 14074). DOI: 10.1007/978-3-031-36021-3_38.
22. Тумбинская М. В., Галиев Р. А. Идентификация фейк-новостей с помощью веб-ресурса на основе нейронных сетей // *Программные продукты и системы*. 2023. Т. 36, № 4. С. 590–599. DOI: 10.15827/0236-235X.142.590-599.
23. Улизько М. В., Ватъян А. С., Гусарова Н. Ф., Добренко Н. В. Применение больших языковых моделей для юридических экспертиз // *ЭПИ*. 2025. № 1. С. 78–89.
24. Ярушкина Н. Г., Мошкин В. С., Константинов А. А. Применение языковых моделей word2vec и bert в задаче сентимент-анализа текстовых сообщений социальных сетей // *Автоматизация процессов управления*. 2020. № 3(61). С. 60–69. DOI: 10.35752/1991-2927-2020-3-61-60-69.

Ванчикова Елена Николаевна

д. э. н., профессор кафедры менеджмента, Бурятская ГСХА (БГСХА, 670010, Улан-Удэ, ул. Пушкина, д. 8), тел. +7 3012 46 26 11, e-mail: evanch@mail.ru. ORCID ID: 0000-0002-6206-0178.

Тимофеев Александр Николаевич

генеральный директор ООО «СибДиджитал», (670046, Улан-Удэ, ул. Красной звезды 23, оф. 18), тел. +7 902 1 632 777, e-mail: tan@sibdigital.net. ORCID ID: 0009-0006-1324-5535.

Садуев Нима Батодоржиевич

к. ф.-м. н., доцент кафедры информатики и информационных технологий в экономике, Бурятская ГСХА (БГСХА, 670010, Улан-Удэ, ул. Пушкина, д. 8), тел. +7 914 8447 906, e-mail: saduev@yandex.ru. ORCID ID: 0009-0000-7308-7045.

Ванзатова Елена Очировна

к. э. н., доцент кафедры информатики и информационных технологий в экономике, Бурятская ГСХА (БГСХА, 670010, Улан-Удэ, ул. Пушкина, д. 8), тел. +7 902 4512 515, e-mail: vanzatova.elena@yandex.ru. ORCID ID: 0000-0002-0887-3362.

Авторы прочитали и одобрили окончательный вариант рукописи.

Авторы заявляют об отсутствии конфликта интересов.

Вклад соавторов: Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.

Development of a mathematical model for studying citizens' appeals

Elena N. Vanchikova¹, Alexander N. Timofeev², Nima B. Saduev¹, Elena O. Vanzatova¹

¹Buryat State Agricultural Academy named after V.R. Filippova, Ulan-Ude, Russia

²SibDigital LLC

Abstract: Currently, citizens actively participate in public discussions that affect the life of society. With the development of digitalization, users are increasingly using social networks and feedback platforms (FBPs) to express their opinions on various areas of activity. In this regard, there is a need to analyze public sentiment using modern analytical tools in order to identify potential problems at an early stage. This article describes the development of a mathematical model for analyzing public sentiment based on text messages published on social networks and other sources. The model includes data preprocessing, clustering of requests, naming of clusters using language models, and summarization of identified problems.

Keywords: mathematical model, machine learning, neural networks, public appeals, and public sentiment analysis.

For citation: Vanchikova E. N., Timofeev A. N., Saduev N. B., Vanzatova E. O. Development of a mathematical model for studying citizens' appeals [Paper Preparation Manual for Vestnik SibGUTI]. *Vestnik SibGUTI*, 2026, vol. 20, no. 1, pp. 39-56. <https://doi.org/10.55648/1998-6920-2026-20-1-39-56>.



Content is available under the license
Creative Commons Attribution 4.0
License

© Vanchikova E. N., Timofeev A. N.,
Saduev N. B., Vanzatova E. O., 2026

The article was submitted: 09.12.2025;
revised version: 11.12.2025;
accepted for publication 13.01.2026.

References

1. Palumbo G., Carneiro D., Guimares M., et al. Algorithm Recommendation and Performance Prediction Using Meta-Learning. *International Journal of Neural Systems*, 2023, vol. 33, no. 3. DOI: 10.1142/s0129065723500119.

2. Brown T. B., et al. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020, vol. 33, pp. 1877–1901.
3. Bunova E. V., Serova V. S. Methodology for solving problems of classification of appeals/requests of citizens to the "hotline" of the President of the Russian Federation. *Vestnik YuUrGU. Seriya: Komp'yuternye tekhnologii, upravlenie, radioelektronika*, 2022, no. 2, pp. 68–79.
4. Crossley N. Social Networks of Meaning and Communication. *American Journal of Sociology*, 2023, vol. 128, no. 6, pp. 1849–1851. DOI: 10.1086/723877.
5. Makarov I. S., Bagantsova E. R., Iashin P. A., et al. Development of and research on machine learning algorithms for solving the classification problem in Twitter publications. *Computer Research and Modeling*, 2023, vol. 15, no. 1, pp. 185–195. DOI: 10.20537/2076-7633-2023-15-1-185-195.
6. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, vol. 1, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.
7. Muhamediyeva D. K., Abdurakhmanova N. N., Mirzayeva N. S. Using Conventional Neural Networks for the Problem of Text Classification. *Central Asian Academic Journal of Scientific Research*, 2021, vol. 1, no. 1, pp. 64–71.
8. Zhang L., Hu Yu., Li W., et al. LLM-AIDSim: LLM-Enhanced Agent-Based Influence Diffusion Simulation in Social Networks. *Systems*, 2025, vol. 13, no. 1, pp. 29. DOI: 10.3390/systems13010029.
9. Lukai Wang, Lei Wang. A Case Study of Chinese Sentiment Analysis of Social Media Reviews Based on LSTM. *SHS Web of Conferences*, 2023, vol. 157, 04012. DOI: 10.1051/shsconf/202315704012.
10. Kim N., Hong S. Automatic classification of citizen requests for transportation using deep learning: Case study from Boston city. *Information Processing & Management*, 2021, vol. 58, no. 1, 102410. DOI: 10.1016/j.ipm.2020.102410.
11. Narayan R., Samanta P. A Machine Learning Approach for Sentiment Analysis Using Social Media Posts. *International Journal of Information Technology and Computer Science*, 2024, vol. 16, no. 5, pp. 23–35. DOI: 10.5815/ijitcs.2024.05.02.
12. Cronin R. M., Fabbri D., Denny J. C., et al. A comparison of rule-based and machine learning approaches for classifying patient portal messages. *International Journal of Medical Informatics*, 2017, vol. 105, pp. 110–120. DOI: 10.1016/j.ijmedinf.2017.06.004.
13. Tolebay A. N. Sentiment Analysis of Texts from Social Networks Based on Machine Learning Methods for Monitoring Public Sentiment. *Bulletin of Science*, 2025, no. 4 (85), pp. 123–134.
14. Yang Y. F., Zhou H., Zhou B. Data clustering based on the modified relaxation Cheeger cut model. *Computational and Applied Mathematics*, 2022, vol. 41, no. 1, pp. 55–62.
15. Anokhina I. Yu., Roshchina E. V. Klasterizatsiya soobshchestv sotsial'noi seti "VKontakte" [Clustering of VKontakte Social Network Communities]. *Informatika i kibernetika*, 2018, no. 2, pp. 34–42.
16. Evdoshenko O. I., Evdoshenko N. I. Razrabotka modeli avtomaticheskogo klassifikatora obrashchenii grazhdan [Development of an Automatic Classifier Model for Citizens' Appeals]. *Prikaspiiskii zhurnal: upravlenie i vysokie tekhnologii*, 2024, no. 2, pp. 56–68.
17. Martyshkin A. I., Britov V. S., Drozdova Yu. R. Vozmozhnosti vyyavleniya emotsional'nogo sostoyaniya pol'zovatelei sotsial'nykh setei s ispol'zovaniem metodov mashinnogo obucheniya [Opportunities for Identifying the Emotional State of Social Media Users Using Machine Learning Methods]. *Sovremennye informatsionnye tekhnologii*, 2024, no. 40, pp. 25–27.
18. Myl'nikova A. V., Myl'nikov L. A. Yazykovye modeli dlya predobrabotki tekstov v mashinnom perevode [Language Models for Text Preprocessing in Machine Translation]. *Nauchno-tekhnicheskaya informatsiya. Seriya 2: Informatsionnye protsessy i sistemy, 2025, no. 7, pp. 32–44. DOI: 10.36535/0548-0027-2025-07-3.
19. Naumova N. V., Abramkov A. O. K voprosu ob ispol'zovanii iskusstvennogo intellekta v sfere raboty s obrashcheniyami grazhdan [On the Use of Artificial Intelligence in Working with Citizens' Appeals]. *Gumanitarnye, sotsial'no-ekonomicheskie i obshchestvennye nauki*, 2022, no. 9, pp. 112–118. Available at: <https://cyberleninka.ru/article/n/k-voprosu-ob-ispolzovanii-iskusstvennogo-intellekta-v-sfere-raboty-s-obrashcheniyami-grazhdan> (accessed: 25.10.2025).

20. Ovezdurdyeva I. K., Garadzhaeva D. Ya. Analiz sotsial'nykh setei s pomoshch'yu iskusstvennogo intellekta: sovremennye podkhody i perspektivy [Analysis of Social Networks Using Artificial Intelligence: Modern Approaches and Prospects]. *Nauka i mirovozzrenie*, 2025, no. 55, pp. 45–52. Available at: <https://cyberleninka.ru/article/n/analiz-sotsialnyh-setey-s-pomoschyu-iskusstvennogo-intellekta-sovremennye-podhody-i-perspektivy> (accessed: 18.11.2025).
21. Saifulla S., Drezhevski R., Dwiyanto F. A., Aribowo A. S., Fauzia Y. Analysis of sentiment using a feature-based machine learning approach to detect anxiety. In: Mikishka J., de Mulatier C., Pashinsky M., et al., eds. *Computational Science – ICCS 2023*. Cham, Springer, 2023, pp. 123–135. (Lecture Notes in Computer Science; Vol. 14074). DOI: 10.1007/978-3-031-36021-3_38.
22. Tumbinskaya M. V., Galiev R. A. Identifikatsiya feik-novostei s pomoshch'yu veb-resursa na osnove neironnykh setei [Identification of Fake News Using a Web Resource Based on Neural Networks]. *Programmnye produkty i sistemy*, 2023, vol. 36, no. 4, pp. 590–599. DOI: 10.15827/0236-235X.142.590-599.
23. Uliz'ko M. V., Vat'yan A. S., Gusarova N. F., Dobrenko N. V. Primenenie bol'shikh yazykovykh modelei dlya yuridicheskikh ekspertiz [The Use of Large Language Models for Legal Expertise]. *EPI*, 2025, no. 1, pp. 78–89. Available at: <https://cyberleninka.ru/article/n/primenenie-bolshih-yazykovykh-modeley-dlya-yuridicheskikh-ekspertiz> (accessed: 18.10.2025).
24. Yarushkina N. G., Moshkin V. S., Konstantinov A. A. Primenenie yazykovykh modelei word2vec i bert v zadache sentiment-analiza tekstovykh soobshchenii sotsial'nykh setei [Application of the word2vec and bert Language Models in the Task of Sentiment Analysis of Text Messages on Social Networks]. *Avtomatizatsiya protsessov upravleniya*, 2020, no. 3, pp. 60–69. DOI: 10.35752/1991-2927-2020-3-61-60-69.

Elena N. Vanchikova

Dr of Economics, Professor of the Department of Management, Buryat State Agricultural Academy (BSCA, 670010, Ulan-Ude, Pushkin St., 8), tel. +7 3012 46 26 11, e-mail: evanch@mail.ru. ORCID ID: 0000-0002-6206-0178

Alexander N. Timofeev

Sibdigital LLC, General Director (23 Krasnaya Zvezda str., office 18, Ulan-Ude, 670046), tel. +7 902 1 632 777, e-mail: tan@sibdigital.net. ORCID ID: 0009-0006-1324-5535

Nima B.Saduev

PhD, Associate Professor of the Department of Informatics and Information Technologies in Economics, Buryat State Agricultural Academy (BSCA, 670010, Ulan-Ude, Pushkin St., 8), tel. +7 914 8447 906, e-mail: saduev@yandex.ru. ORCID ID: 0009-0000-7308-7045

Elena O. Vanzatova

Candidate of Economics, Associate Professor of the Department of Informatics and Information Technologies in Economics, Buryat State Agricultural Academy (BSCA, 670010, Ulan-Ude, Pushkin St., 8), tel. +7 902 4512 515, e-mail: vanzatova.elena@yandex.ru. ORCID ID: 0000-0002-0887-3362.