

УДК 631.1,62-5

UDC 631.1,62-5

5.2.2. Математические, статистические и инструментальные методы экономики (физикоматематические науки, экономические науки)

5.2.2. Mathematical, statistical and instrumental methods of economics (physical and mathematical sciences, economic sciences)

**ИССЛЕДОВАНИЕ И РАЗРАБОТКА СИСТЕМЫ ДЛЯ АНАЛИЗА И КЛАССИФИКАЦИИ ОБРАЩЕНИЙ КОНТАКТНОГО ЦЕНТРА**

**RESEARCH AND DEVELOPMENT OF THE SYSTEM FOR ANALYZING AND CLASSIFICATION CONTACT CENTER CASES**

Мурлина Владислава Анатольевна  
Кандидат технических наук, доцент кафедры информационных систем и программирования  
РИНЦ-SCIENCE INDEX SPIN-код: 1256-1843  
*ФГБОУ ВО “Кубанский государственный технологический университет”, 350020, улица Московская, 2, Краснодар, Россия*

Murlina Vladislava Anatolevna  
Cand.Tech.Sci., Associate Professor of the Department of Information Systems and Programming  
RSCI-SCIENCE INDEX SPIN-code: 1256-1843  
*Kuban State Technological University, 350020, Moskovskaya, 2, Krasnodar, Russia*

Ковалева Гульнара Фаритовна  
студентка кафедры информационных систем и программирования  
g.kovaleva0909@gmail.com  
*ФГБОУ ВО “Кубанский государственный технологический университет”, 350020, улица Московская, 2, Краснодар, Россия*

Kovaleva Gulnara Faritovna  
student of the Department of Information Systems and Programming  
g.kovaleva0909@gmail.com  
*Kuban State Technological University, 350020, Moskovskaya, 2, Krasnodar, Russia*

Цель исследования заключается в создании функциональной и надёжной системы управления обращениями, предназначенной для оптимизации процессов обработки обращений, взаимодействия с клиентами и партнёрами, а также повышения эффективности работы сотрудников. В рамках системы будут реализованы следующие ключевые возможности: создание и маршрутизация обращений с учётом их темы, отдела и источника поступления; отслеживание статусов и выполнения обращений с возможностью автоматического обновления; интеграция с внешними каналами связи, такими как Telegram, WhatsApp и Email; поддержка комментариев сотрудников и администраторов для улучшения коммуникации внутри системы; гибкая настройка справочников отделов и тем обращений для повышения адаптивности системы

The purpose of the research is to develop a functional and reliable case management system designed to optimize the processes of handling inquiries, interacting with clients and partners, and improving employee efficiency. The system will implement the following key features: creation and routing of cases based on their topic, department, and source of origin; tracking the status and resolution of cases with the capability for automatic updates; integration with external communication channels such as Telegram, WhatsApp, and Email; support for employee and administrator comments to enhance internal communication within the system; flexible configuration of department and case topic directories to improve the system’s adaptability

Ключевые слова: КОНТАКТНЫЙ ЦЕНТР, АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ, ИНТЕГРАЦИЯ, ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА, NLP, МАРШРУТИЗАЦИЯ ОБРАЩЕНИЙ, МАШИННОЕ ОБУЧЕНИЕ

Keywords: CONTACT CENTER, AUTOMATIC CLASSIFICATION, INTEGRATION, NATURAL LANGUAGE PROCESSING, NLP, CASE ROUTING, MACHINE LEARNING

<http://dx.doi.org/10.21515/1990-4665-205-039>

<http://ej.kubagro.ru/2025/01/pdf/39.pdf>

## **Введение**

Современные контактные центры играют ключевую роль в обеспечении высокого уровня обслуживания клиентов. Однако с ростом объёма входящих обращений возрастает сложность их обработки, что создаёт дополнительную нагрузку на сотрудников и снижает общую эффективность работы. В условиях цифровизации и растущих ожиданий клиентов, автоматизация анализа и классификации обращений становится одной из приоритетных задач. Настоящая статья посвящена разработке системы, которая позволяет автоматизировать обработку заявок, используя технологии машинного обучения и интеграционные решения.

**Постановка проблемы.** Основной проблемой обработки обращений в контактных центрах является высокая нагрузка на операторов, вызванная большим числом однотипных задач. Сотрудникам приходится вручную распределять обращения, что замедляет процесс обработки и увеличивает вероятность ошибок. Кроме того, использование различных каналов коммуникации, таких как электронная почта и мессенджеры, создаёт дополнительную сложность в управлении данными. В то же время клиенты ожидают оперативного ответа на свои запросы, что требует от системы контактного центра высокой производительности и точности.

Решение этой проблемы требует создания системы, которая автоматически классифицирует обращения и распределяет их по соответствующим отделам. Такая система должна учитывать специфику текстовых данных, поступающих из разных источников, и обеспечивать их надёжную обработку. Кроме того, она должна быть устойчивой к

увеличению объёма данных и поддерживать гибкость интеграции с новыми каналами связи.

**Цель исследования** заключается в разработке системы, обеспечивающей эффективный анализ и классификацию обращений в контактных центрах. Такая система должна уменьшить нагрузку на операторов, ускорить обработку запросов и повысить точность распределения обращений по отделам. В основе работы системы лежит алгоритм машинного обучения — случайный лес, который обеспечивает высокую точность классификации текстовых данных.

**Объектом исследования** является система управления обращениями, включающая процессы обработки обращений, маршрутизации, взаимодействия с клиентами, партнёрами и сотрудниками, а также инструменты автоматизации и предиктивного анализа для оптимизации работы.

Для создания эффективной системы анализа и классификации обращений необходимо удовлетворить следующие ключевые требования. Во-первых, система должна поддерживать интеграцию с различными каналами связи, включая электронную почту и мессенджеры (WhatsApp, Telegram). Во-вторых, требуется обеспечить высокую точность классификации на основе исторических данных и возможность обработки низкокачественных текстов, таких как сообщения с опечатками или неформальными выражениями. В-третьих, архитектура системы должна быть масштабируемой, чтобы поддерживать рост объёма обращений без потери производительности. Наконец, система должна предоставлять удобный веб-интерфейс для работы сотрудников с обращениями.

### **Анализ существующих решений**

Для решения задачи анализа и классификации обращений в контактных центрах активно используются технологии машинного обучения и обработки естественного языка (NLP).

Классификация текстовых обращений относится к задачам машинного обучения с учителем, где модель обучается на примерах и впоследствии классифицирует новые данные, опираясь на выявленные зависимости. Алгоритмы классификации используются для задач типа «да/нет» (двоичная классификация) или распределения по категориям (мультиклассовая классификация). Первый этап работы включает преобразование текстов в пригодный для алгоритмов вид, обычно векторы в пространстве признаков. Для этого используется предобработка текста и уменьшение пространства признаков.

Предобработка текста включает в себя такие методы:

- лексический анализ - удаление специальных символов (знаков препинания, пробелов), чтобы минимизировать шум в данных;
- токенизация - разделение текста на отдельные слова или токены, создавая основу для дальнейшего анализа;
- стеммитизация - приведение слов к основе, уменьшая морфологические формы (например, "покупать" → "покуп");
- лемматизация - приведение слов к их начальной форме с учетом контекста и словаря (например, "покупал" → "покупать");
- удаление стоп-слов - исключение часто встречающихся, но нерелевантных слов (например, "и", "в", "на");
- индексация документов - преобразование текста в числовые представления. Для этого чаще всего используют метод «мешка слов» (Bag-of-Words) - представление текста как вектора частот слов.

Снижение размерности пространства признаков помогает уменьшить вычислительные затраты и улучшить производительность моделей. Для этого используют алгоритмы:

- TF-IDF (Term Frequency-Inverse Document Frequency) - оценивает важность слов, снижая влияние часто встречающихся, но неинформативных терминов;

- Word2Vec - обучает плотные векторные представления слов, отражая их семантическую близость;
- PCA (Principal Component Analysis) - метод главных компонент, снижающий размерность путем выявления наиболее значимых осей векторного пространства;
- LSA (Latent Semantic Analysis) - выделяет скрытые семантические структуры в тексте, часто используется для работы с матрицей TF-IDF.

Второй этап — обучение модели на примерах для построения классифицирующей функции. Основные методы машинного обучения, используемые для задач классификации текста:

- наивный байесовский классификатор - рассчитывает вероятность принадлежности текста к классу на основе теоремы Байеса. Простой и эффективный метод для текстовых данных;
- метод опорных векторов (SVM) - строит гиперплоскость, разделяющую классы. Эффективен для высокоразмерных данных;
- логистическая регрессия - линейный метод классификации, который хорошо работает с бинарными задачами;
- дерево решений - иерархическая структура, где узлы соответствуют признакам, а листья — классам. Легко интерпретируется;
- случайный лес - комбинация деревьев решений, обеспечивающая устойчивость к шуму;
- метод k ближайших соседей (k-NN) - сравнивает новые данные с ближайшими в обучающей выборке;
- искусственные нейронные сети (ANN) - универсальный подход, способный находить сложные зависимости в данных, но требующий больших объемов данных и вычислительных ресурсов [1].

Одна из существующих систем для анализа и классификации обращений, предложенная К. Poczeta, ориентирована на автоматическое

определение одной или нескольких категорий сообщений. Используется гибридный подход: векторизация текстов (TF-IDF, Bag of N-grams), снижение размерности (PCA, LSA) и модель нейронной сети многослойный перцептрон. Преимущество заключается в способности системы обрабатывать сложные случаи, но высокая вычислительная сложность и зависимость от качества данных являются недостатками [2].

Другое решение для автоматической классификации заявок, предложенное Al-Hawari F., Barham H., использует метод опорных векторов (SVM) и TF-IDF. Особенностью является работа с текстами, включающими заголовки, описания и комментарии. Точность модели увеличена до 81,4% благодаря тщательной предобработке данных. Однако сложность предобработки и чувствительность к объемам данных остаются проблемами [3].

Система классификации обращений абонентов в колл-центр телекоммуникационной компании, созданная Жарковым М. С., использует стохастический градиентный спуск (SGD) и TF-IDF. SGD — оптимизационный метод, используемый для настройки параметров модели. Вместо обработки всего набора данных на каждом шаге, обновления параметров производятся по одной случайной выборке, что снижает вычислительную сложность и ускоряет обучение. Для обработки русскоязычных данных применяются транскрибация и предобработка текстов. Достоинством является адаптация к особенностям языка, а недостатками — несбалансированность данных и сложность оптимизации [4].

Следующее решение - модель KU-LR (Key Utterance labels and Logistic Regression), предложенная Liu Y. et al., использует логистическую регрессию для выделения ключевых реплик в текстах, что улучшает производительность при ограниченных данных. Метод эффективен для

малых выборок, но менее конкурентоспособен при больших объемах данных [5].

Другое существующее решение – система AI-ICS (Incident Classification System), основанная на модели LaBSE. LaBSE (Language-agnostic BERT Sentence Embedding) — это модель глубокого обучения от Google, предназначенная для создания языково-независимых эмбедингов предложений. Она основана на архитектуре большой языковой модели BERT (Bidirectional Encoder Representations from Transformers) и обучена на множестве языков (около 100), что позволяет эффективно сравнивать предложения на разных языках. Система предназначена для обработки многоязычных данных и использует эмбединги текста для классификации обращений. Точность составляет 85%, однако система зависима от качества исходных данных, а доля ручной проверки остается высокой [6].

Следующее решение, разработанное компанией Accenture для автоматизации бизнес-процессов колл-центра компании 5Post, основано на модели BERT, дообученной для классификации и извлечения сущностей. BERT основывается на архитектуре трансформеров и достигает высокой точности в NLP-задачах. Система позволяет сократить до 74% временных затрат на обработку обращений. Преимущества включают высокую точность и возможность извлечения сущностей, но высокие вычислительные затраты и сложность разметки данных остаются проблемами [7].

Существующие решения для автоматизации обработки обращений показывают высокую точность, однако требуют значительных вычислительных ресурсов и большого объема обучающих данных. В то же время, алгоритмы на основе случайного леса обеспечивают баланс между производительностью и точностью, что делает их оптимальным выбором для корпоративных приложений. Также многие существующие решения не

учитывают специфику текстов, поступающих из мессенджеров и социальных сетей, что снижает их точность в реальных условиях.

Разработанная система учитывает эти недостатки. Она использует модуль интеграции для нормализации входящих данных и модуль классификации, оптимизированный для работы с текстами из различных источников. Это позволяет повысить точность классификации и сократить время обработки заявок.

### **Исследовательская часть**

Система анализа и классификации обращений состоит из четырёх основных компонентов: модуля интеграции, модуля классификации, базы данных и веб-приложения. Каждый из компонентов играет важную роль в обеспечении общей функциональности и производительности системы. Модуль интеграции отвечает за сбор данных из различных источников и нормализацию текстов. Модуль классификации использует алгоритм случайного леса для распределения обращений по отделам. Веб-приложение предоставляет пользователям доступ к базе данных для обработки заявок, а RabbitMQ обеспечивает устойчивость системы и балансировку нагрузки.

Работа системы начинается с приёма входящих данных модулем интеграции, который осуществляет нормализацию текста и передаёт сообщения в очередь RabbitMQ. RabbitMQ – это распределённый и горизонтально масштабируемый брокер сообщений, который обеспечивает надёжную доставку сообщений. Далее модуль классификации извлекает данные из очереди и применяет алгоритм случайного леса для определения соответствующего отдела. После классификации результаты сохраняются в базе данных и становятся доступными для обработки через веб-приложение. Такой подход позволяет обеспечить чёткое разделение ответственности между компонентами системы, что способствует её

масштабируемости, производительности и удобству дальнейшего развития. Алгоритм работы системы представлен на рисунке 1.

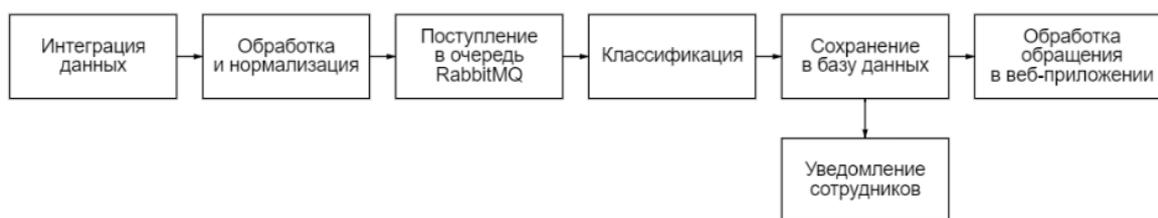


Рисунок 1 - Алгоритм работы системы

Модуль интеграции отвечает за приём и обработку данных из различных каналов связи. Для работы с электронной почтой используются протоколы IMAP и SMTP, что позволяет автоматически извлекать сообщения и их вложения из корпоративных почтовых серверов. Для интеграции с мессенджерами WhatsApp и Telegram используются официальные API, предоставляемые платформами. Каждое входящее сообщение проходит процесс нормализации, включающий удаление лишних символов, токенизацию текста, лемматизацию и удаление стоп-слов. Эти шаги выполняются с использованием библиотек Python, таких как nltk и rutorphy3. Нормализованные данные передаются в очередь RabbitMQ, которая обеспечивает их дальнейшую обработку.

Модуль классификации построен на базе алгоритма случайного леса, предоставляемого библиотекой Scikit-learn. Этот алгоритм был выбран за его высокую точность и устойчивость к переобучению. Для обучения модели был использован датасет, содержащий 3 000 обращений, в котором каждая заявка была помечена соответствующим отделом. Перед обучением данные проходили этапы предварительной обработки: очистку текста, токенизацию, лемматизацию и преобразование в численное представление с использованием TF-IDF. Для построения модели использовались 100 деревьев решений с фиксированным случайным состоянием (`random_state=42`), чтобы обеспечить воспроизводимость

результатов. После обучения модель была протестирована на тестовой выборке, составляющей 30% от общего объёма данных. Результаты тестирования показали, что точность классификации составляет 89%, а F1-мера достигает 0.85. Алгоритм обучения модели классификации представлен на рисунке 2.

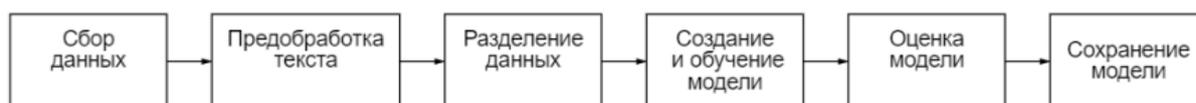


Рисунок 2 - Алгоритм обучения модели классификации

Для хранения данных используется база SQLite, интегрированная с Django ORM. Основная таблица "Ticket", предназначенная для хранения данных об обращениях, содержит поля для описания темы, деталей, привязки к отделу, исполнителей и источника поступления, а также для отслеживания статуса и других характеристик заявки. База данных включает также справочники отделов и тем обращений, данные о сотрудниках, клиентах, ресторанах-партнёрах, а также комментарии, добавляемые к заявкам. Система поддерживает учёт отправителей и связана с моделями клиентов и партнёров. Такая структура обеспечивает удобство хранения данных и возможность выполнения сложных аналитических запросов. Ниже представлена схема основной базы данных (Рис. 3).

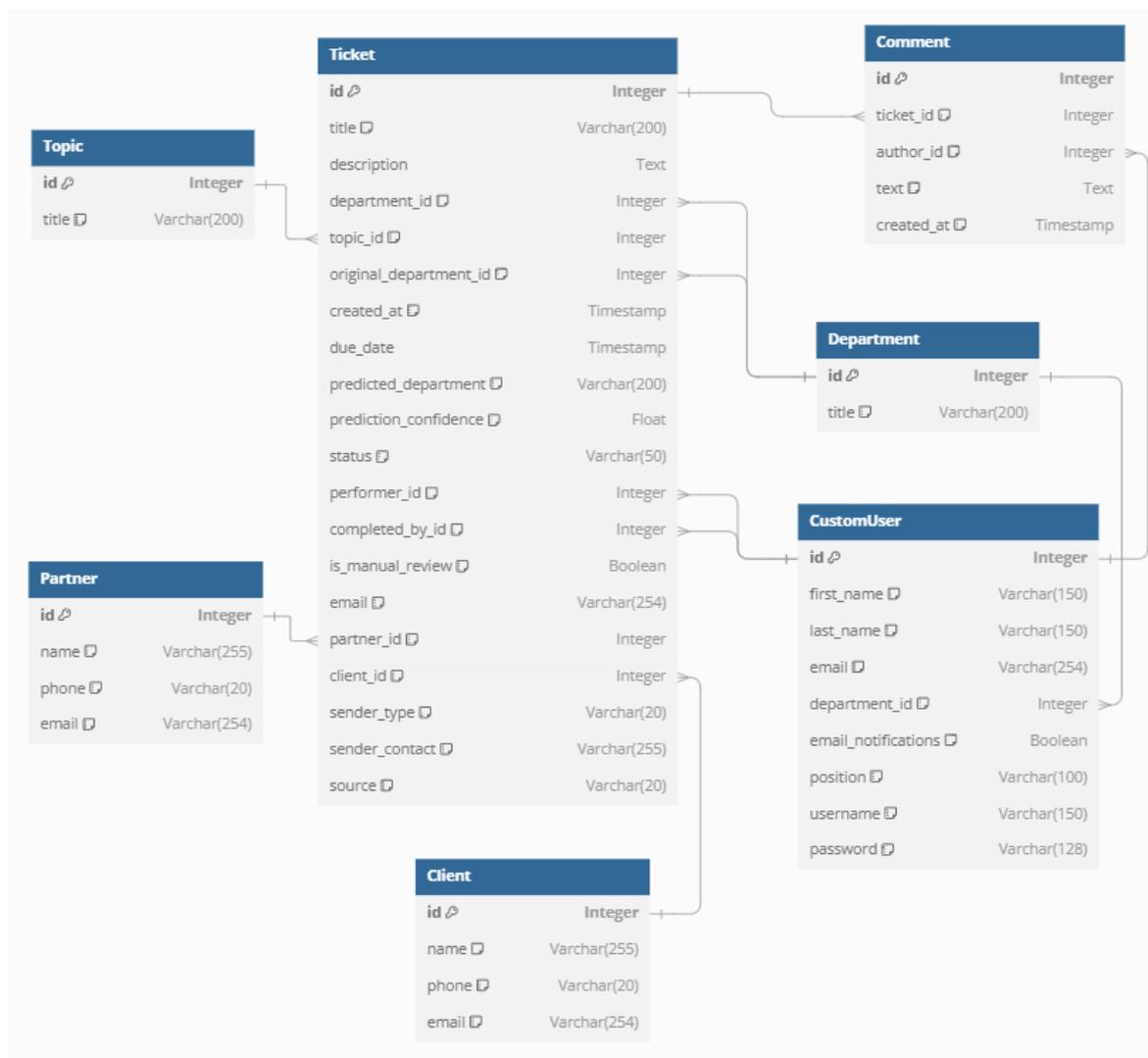


Рисунок 3 - Схема основной базы данных

Веб-приложение разработано с использованием фреймворка Django. Оно предоставляет пользователям удобный интерфейс для работы с заявками, включая их просмотр, редактирование, добавление комментариев и изменение статуса. Приложение поддерживает три роли пользователей: сотрудник, контакт-центр и администратор. Администратор имеет доступ к управлению пользователями и настройке системы, включая параметры интеграции и модели машинного обучения. Веб-приложение также включает аналитический модуль, позволяющий генерировать отчёты о работе контактного центра, такие как распределение заявок по отделам, время обработки заявок и эффективность сотрудников.

Для обеспечения устойчивости и масштабируемости системы используются очереди сообщений RabbitMQ. Это решение позволяет равномерно распределять нагрузку между компонентами системы, что особенно важно при обработке большого объёма заявок.

В ходе экспериментов было развернуто тестовое окружение с RabbitMQ и Django-приложением. На тестовых данных, представляющих выборку обращений из различных каналов (электронная почта, WhatsApp, Telegram), среднее время обработки каждого сообщения оказалось достаточно малым для оперативного взаимодействия. Точность работы алгоритма «Случайный лес» при классификации обращений на уровне 85–90% позволила существенно сократить время, затрачиваемое на ручную сортировку заявок.

Также была проверена производительность системы при увеличении объёма данных. Использование RabbitMQ позволило равномерно распределить нагрузку, что обеспечило стабильную работу при обработке до 5000 заявок в минуту. Это подтверждает устойчивость и масштабируемость архитектуры.

Практическое использование системы показало, что автоматизация обработки обращений позволяет снизить нагрузку на операторов на 30% и ускорить обработку запросов на 40%. Это особенно важно в условиях высоких нагрузок, когда требуется оперативное распределение заявок.

**Результаты исследования** подтверждают, что использование современных технологий и структурированной архитектуры базы данных позволяет создать эффективную систему управления обращениями, способствующую оптимизации процессов обработки заявок и повышению качества взаимодействия между пользователями, сотрудниками и партнёрами.

Система обеспечивает интеграцию с различными каналами связи, нормализацию данных, точную классификацию заявок и удобный

интерфейс для работы сотрудников. Использование алгоритма Random Forest позволило достичь высокой точности классификации при умеренных вычислительных затратах.

Система показала высокую адаптивность к изменениям в бизнес-процессах, обеспечив удобство настройки справочников отделов и тем обращения, а также интеграцию с каналами связи, такими как Telegram, WhatsApp и Email.

Полученные результаты демонстрируют, что разработанная система предоставляет функциональное и удобное решение для управления обращениями, оптимизируя бизнес-процессы, улучшая взаимодействие с клиентами и повышая эффективность сотрудников. Эта система может быть успешно внедрена в организациях с широким спектром задач по обработке и маршрутизации обращений, поддерживая современные подходы к цифровизации процессов.

#### ЛИТЕРАТУРА

1. Большакова Е. И. и др. Автоматическая обработка текстов на естественном языке и анализ данных //М.: Изд-во НИУ ВШЭ. – 2017.
2. Poczeta K. et al. A multi-label text message classification method designed for applications in call/contact centre systems //Applied Soft Computing. – 2023. – Т. 145. – С. 110562.
3. Al-Hawari F., Barham H. A machine learning based help desk system for IT service management //Journal of King Saud University-Computer and Information Sciences. – 2021. – Т. 33. – №. 6. – С. 702-718.
4. Жарков М. С. Разработка системы классификации обращений абонентов в колл-центр ГК "Орион телеком" : дис. – Сибирский федеральный университет, 2022.
5. Liu Y. et al. Improving the classification of call center service dialogue with key utterances //Wireless Networks. – 2021. – Т. 27. – №. 5. – С. 3395-3406.
6. Колосовская О. Маршрутизация обращений: автоматизация в ИТ-поддержке с помощью ИИ и языковых моделей. 2024 г. URL: <https://habr.com/ru/companies/nlmc/articles/824126/> (дата обращения: 20.10.2024).
7. Яшневский. Н. Автоматизация обработки клиентских обращений, 2022 г. URL: <https://habr.com/ru/company/X5Group/blog/598919/> (дата обращения: 28.10.2024).

#### REFERENCES

1. Bol'shakova E. I. i dr. Avtomaticheskaja obrabotka tekstov na estestvennom jazyke i analiz dannyh //M.: Izd-vo NIU VShJe. – 2017.
2. Poczeta K. et al. A multi-label text message classification method designed for applications in call/contact centre systems //Applied Soft Computing. – 2023. – Т. 145. – С. 110562.

3. Al-Hawari F., Barham H. A machine learning based help desk system for IT service management //Journal of King Saud University-Computer and Information Sciences. – 2021. – Т. 33. – №. 6. – S. 702-718.

4. Zharkov M. S. Razrabotka sistemy klassifikacii obrashhenij abonentov v koll-centr GK "Orion telekom" : dis. – Sibirskij federal'nyj universitet, 2022.

5. Liu Y. et al. Improving the classification of call center service dialogue with key utterances //Wireless Networks. – 2021. – Т. 27. – №. 5. – S. 3395-3406.

6. Kolosovskaja O. Marshrutizacija obrashhenij: avtomatizacija v IT-podderzhke s pomoshh'ju И i jazykovyh modelej. 2024 g. URL: <https://habr.com/ru/companies/nlmc/articles/824126/> (data obrashhenija: 20.10.2024).

7. Jashnevskij. N. Avtomatizacija obrabotki klientskih obrashhenij, 2022 g. URL: <https://habr.com/ru/company/X5Group/blog/598919/> (data obrashhenija: 28.10.2024).