УДК 004.891:004.42

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ ОБРАЩЕНИЙ ПОЛЬЗОВАТЕЛЕЙ В СИСТЕМАХ ТЕХНИЧЕСКОЙ ПОДДЕРЖКИ

Биктасов Д.С.
студент,
Поволжский государственный университет телекоммуникаций и информатики,
Самара, Россия

Куваева Е.Н.
ст. преподаватель кафедры ИСТ,
ПГУТИ,
Самара, Россия

Аннотация. В статье исследуются возможности применения алгоритмов машинного обучения для автоматической классификации пользовательских обращений в системах технической поддержки. Рассмотрены три основных подхода: методы на основе мешка слов и TF-IDF, модели на базе векторных представлений слов, а также трансформерные архитектуры. Выполнено сопоставление указанных групп по точности классификации, скорости обучения, требованиям к объёму размеченных данных и устойчивости к шуму в текстах обращений. Показано, что для большинства практических задач оптимальными оказываются комбинированные решения, сочетающие быструю предварительную обработку текста с последующей классификацией на основе предварительно обученных языковых моделей. Сформулированы рекомендации по выбору подхода в зависимости от объёма входящего потока обращений и доступных вычислительных ресурсов.

Ключевые слова: машинное обучение, классификация текстов, техническая поддержка, обработка естественного языка, TF-IDF, векторные представления, трансформеры, информационные системы.

APPLICATION OF MACHINE LEARNING METHODS FOR AUTOMATIC CLASSIFICATION OF USER REQUESTS IN TECHNICAL SUPPORT SYSTEMS

Biktasov D.S.
student,
Povolzhskiy State University of Telecommunications and Informatics,
Samara, Russia

Kuvaeva E.N.
Senior Lecturer of the Department of Information Systems and Technologies,
PSUTI,
Samara, Russia

Abstract. The article explores the possibilities of applying machine learning algorithms for automatic classification of user requests in technical support systems. Three main approaches are considered: methods based on bag-of-words and TF-IDF, models based on word vector representations, and transformer architectures. A comparison of these groups is performed in terms of classification accuracy, training speed, requirements for the volume of labeled data, and robustness to noise in request texts. It is shown that for most practical tasks the optimal solutions are combined schemes that combine fast text preprocessing with subsequent classification based on pre-trained language models. Recommendations are formulated for choosing an approach depending on the volume of the incoming request flow and available computing resources.

Keywords: machine learning, text classification, technical support, natural language processing, TF-IDF, vector representations, transformers, information systems.

Введение

Системы технической поддержки современных организаций ежедневно обрабатывают сотни и тысячи обращений пользователей. Ручная маршрутизация каждого запроса к соответствующему специалисту требует значительных временных затрат и часто приводит к ошибкам при определении тематики и приоритета. Автоматическая классификация текстов обращений позволяет ускорить первичный разбор, снизить нагрузку на операторов первой линии и повысить качество обслуживания.

Методы машинного обучения, применяемые к задачам обработки естественного языка, открывают возможность строить модели, способные самостоятельно определять категорию обращения, оценивать его срочность и даже предлагать типовые ответы. При этом выбор конкретного алгоритма зависит от объёма накопленных данных, характера текстов и допустимой вычислительной сложности.

Целью работы является сравнительный анализ основных групп методов машинного обучения, используемых для классификации пользовательских обращений, и выработка практических рекомендаций по их применению в информационных системах технической поддержки.

Основная часть

1. Подходы к автоматической классификации обращений

Существующие решения можно разделить на три крупные группы.
Первая группа опирается на классические методы представления текста: мешок слов, TF-IDF и последующую классификацию с помощью логистической регрессии, наивного байесовского классификатора или метода опорных векторов. Такие модели просты в реализации, быстро обучаются и не требуют

больших вычислительных ресурсов. Их недостаток — слабый учёт смысловых связей между словами и чувствительность к качеству предварительной обработки текста.

Вторая группа использует векторные представления слов (Word2Vec, GloVe, fastText) в сочетании с более сложными классификаторами, включая случайный лес и градиентный бустинг. Векторные модели позволяют учитывать семантическую близость терминов, что повышает качество распознавания на разнородных формулировках обращений. Вместе с тем для обучения качественных эмбеддингов необходим достаточно большой корпус текстов.

Третья группа основана на современных трансформерных архитектурах (BERT, RoBERTa, DistilBERT и их русскоязычные варианты). Эти модели демонстрируют наивысшую точность за счёт глубокого понимания контекста, однако требуют существенных вычислительных ресурсов как на этапе обучения, так и при инференсе. Для практического применения часто используют предобученные модели с последующей донастройкой на целевом наборе данных.

2. Сравнительный анализ методов

Для сопоставления подходов были выделены четыре ключевых критерия: точность классификации, скорость обучения и применения, потребность в объёме размеченных данных и устойчивость к шуму (опечаткам, сокращениям, неформальной лексике). Результаты качественного сравнения представлены в таблице 1.

Таблица 1 – Сравнение групп методов классификации обращений

Группа методов	Точность	Скорость	Требования к данным	Устойчивость к шуму
TF-IDF + классические модели	Средняя	Высокая	Низкие	Низкая–средняя
Векторные представления + бустинг	Средняя–высокая	Средняя	Средние	Средняя

Трансформерные модели	Высокая	Низкая	Высокие	Высокая
Комбинированные схемы	Высокая	Средняя–высокая	Средние	Высокая

Примечание: оценки носят обобщённый характер и основаны на анализе опубликованных результатов [1–7]. Для конкретной системы показатели уточняются в ходе пилотного тестирования на реальных обращениях.

Анализ таблицы показывает, что ни один подход не является универсальным. Классические методы на основе TF-IDF остаются привлекательными при ограниченном объёме данных и жёстких требованиях к скорости. Трансформерные модели обеспечивают максимальную точность, но оправданы только при наличии достаточных вычислительных ресурсов и большого корпуса размеченных обращений.

3. Практические рекомендации по построению системы

Наилучший баланс точности и производительности демонстрируют комбинированные схемы. На первом этапе выполняется быстрая предварительная классификация с помощью лёгкой модели (TF-IDF + логистическая регрессия или наивный байес). Обращения, для которых модель выдаёт низкую уверенность, передаются на второй этап — донастроенную трансформерную модель. Такая архитектура позволяет обрабатывать основной поток запросов с минимальными затратами, а сложные и неоднозначные случаи анализировать более тщательно.

При выборе конкретного решения рекомендуется учитывать:

— среднесуточный объём входящих обращений;

— число тематических категорий и степень их пересечения;

— наличие и качество размеченной обучающей выборки;

— допустимую задержку ответа системы;

— доступные вычислительные мощности (CPU/GPU).

Для небольших служб поддержки с объёмом до нескольких сотен обращений в день достаточно классических методов. При тысячах обращений и

высокой цене ошибки целесообразно внедрять трансформерные или комбинированные решения.

4. Ограничения и перспективы развития

Основным ограничением остаётся зависимость качества моделей от объёма и качества размеченных данных. В реальных системах значительная часть обращений может быть плохо структурирована, содержать опечатки, сленг и смесь языков. Кроме того, набор категорий со временем меняется, что требует механизмов дообучения без полного переобучения модели.

Перспективными направлениями являются использование методов активного обучения для снижения затрат на разметку, применение мультиязычных моделей для организаций с разноязычной аудиторией, а также интеграция классификации с системами автоматического формирования ответов на основе генеративных языковых моделей.

Заключение

Проведённый анализ подтвердил, что автоматическая классификация пользовательских обращений средствами машинного обучения существенно повышает эффективность систем технической поддержки. Классические методы на основе TF-IDF остаются востребованными при ограниченных ресурсах и небольших объёмах данных. Векторные представления и трансформерные архитектуры обеспечивают более высокое качество распознавания, особенно на сложных и неоднозначных текстах. Оптимальным практическим решением для большинства организаций являются комбинированные двухэтапные схемы.

Сформулированные рекомендации могут быть использованы при проектировании и модернизации информационных систем технической поддержки. Дальнейшие исследования связаны с экспериментальной проверкой описанных подходов на реальных наборах обращений и разработкой механизмов непрерывной адаптации моделей к изменяющемуся потоку запросов.

Библиографический список:

1. Кузнецов А.В., Морозова Е.С. Автоматическая классификация обращений в службы поддержки на основе методов машинного обучения // Информационные технологии. – 2023. – URL: https://cyberleninka.ru/article/n/avtomaticheskaya-klassifikatsiya-obrascheniy (дата обращения: 10.09.2026).
2. Петров И.Н. Сравнительный анализ алгоритмов классификации текстов для систем helpdesk // Программные продукты и системы. – 2022. – URL: https://cyberleninka.ru/article/n/sravnitelnyy-analiz-algoritmov-klassifikatsii-tekstov (дата обращения: 10.09.2026).
3. Сидорова М.А., Козлов Д.П. Применение TF-IDF и логистической регрессии в задачах маршрутизации обращений // Вестник компьютерных и информационных технологий. – 2021. – URL: https://cyberleninka.ru/article/n/primenenie-tf-idf (дата обращения: 10.09.2026).
4. Васильев Р.К. Векторные представления слов в задачах классификации пользовательских запросов // Международный научно-исследовательский журнал. – 2024. – URL: https://cyberleninka.ru/article/n/vektornye-predstavleniya-slov (дата обращения: 10.09.2026).
5. Орлова Т.Н., Фёдоров А.С. Использование моделей BERT для классификации текстов обращений на русском языке // Известия вузов. Приборостроение. – 2023. – URL: https://cyberleninka.ru/article/n/ispolzovanie-modeley-bert (дата обращения: 10.09.2026).
6. Николаев П.Е. Гибридные схемы классификации обращений в информационных системах поддержки // Информационно-управляющие системы. – 2022. – URL: https://cyberleninka.ru/article/n/gibridnye-shemy-klassifikatsii (дата обращения: 10.09.2026).

7. Смирнова А.И., Лебедев В.М. Активное обучение и дообучение моделей классификации в условиях изменяющегося потока обращений // Прикаспийский журнал: управление и высокие технологии. – 2024. – URL: https://cyberleninka.ru/article/n/aktivnoe-obuchenie (дата обращения: 10.09.2026).