

СЕРВИС ДЛЯ АНАЛИЗА И КЛАССИФИКАЦИИ МЕДИЦИНСКИХ ЗАПИСЕЙ ДЛЯ ТЕКСТОВЫХ ОБРАЩЕНИЙ

Прудникова Анастасия Алексеевна

Студент

Московский авиационный институт (национальный исследовательский университет), Москва, Россия

E-mail: nprudnikova2403@gmail.com

Научный руководитель — Луккин Владимир Николаевич

Для обработки медицинских записей требуется оптимизированная система извлечения информации, позволяющая эффективно обрабатывать естественные языки. При решении задач с их использованием зачастую применяются NLP-модели[1]. Настоящая работа направлена на разработку подобного доступного инструмента для классификации текстовых обращений, который основан на современных технологиях машинного обучения и обработки текста.

Результат работы — сервис, который позволяет медицинским учреждениям эффективно анализировать обращения, классифицировать их по симптомам и направлениям медицинской деятельности и предоставлять пользователям возможность оперативно записываться к нужным специалистам.

Для осуществления записи пользователь обращается к сервису, перечисляя симптомы. Алгоритм обрабатывает данные (токенизация, лемматизация и удаление стоп-слов). Далее текст передается в NLP-модель для выделения и сопоставления симптомов с учетом предыстории пациента (если она имеется в базе). На основе полученной информации пациенту предоставляется выбор из перечня доступных специалистов для записи.

Реализация проекта включала несколько этапов: сбор и предобработка данных, разработка базы данных, обучение модели и создание пользовательского интерфейса. В рамках исследования была использована NLP модель, обученная на открытых медицинских датасетах. На первом этапе был сформирован корпус текстовых данных, включающий симптомы, диагнозы и данные о врачах. Данные прошли очистку, нормализацию и аннотирование. Для хранения информации была разработана база данных на основе PostgreSQL, обеспечивающая быстрый доступ к медицинским записям и поддерживающая масштабируемость системы.

Система успешно классифицирует текстовые обращения, определяя ключевые симптомы и предлагая наиболее подходящего специалиста. Тестирование показало высокую точность модели: F1-score достиг свыше 86%, что свидетельствует о высокой степени соответствия рекомендаций реальным медицинским ситуациям.

Проект продемонстрировал эффективность применения глубокого обучения и обработки естественного языка в сфере медицины, обеспечивая точность и оперативность маршрутизации пациентов. В дальнейшем возможна доработка модели, реализация на других платформах, а также расширение базы данных и внедрение поддержки других языков.

Литература

1. Devlin, J., Chang, M.-W., Lee, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. // Association for Computational Linguistics, Minneapolis, Minnesota, 2019, P.,4174–4175.