Отчет по Курсовой работе №2

Тема «Разработка модели для классификации спам-сообщений на основе методов машинного обучения»

Выполнил: Козлов Эдуард гр.39

Задача классификации спам-сообщений позволяет использовать самые современные технологии машинного обучения и обработки естественного языка, такие как модель трансформер RuBert. Именно эту модель я решил дообучить на своем датасете.

RuBERT является одним из наиболее популярных и мощных вариантов модели BERT для обработки русского языка, который позволяет эффективно решать задачи классификации текстов, включая распознавание спама. Выбор данной модели обусловлен её высокой точностью на русском языке и возможностью дообучения на специфических датасетах.

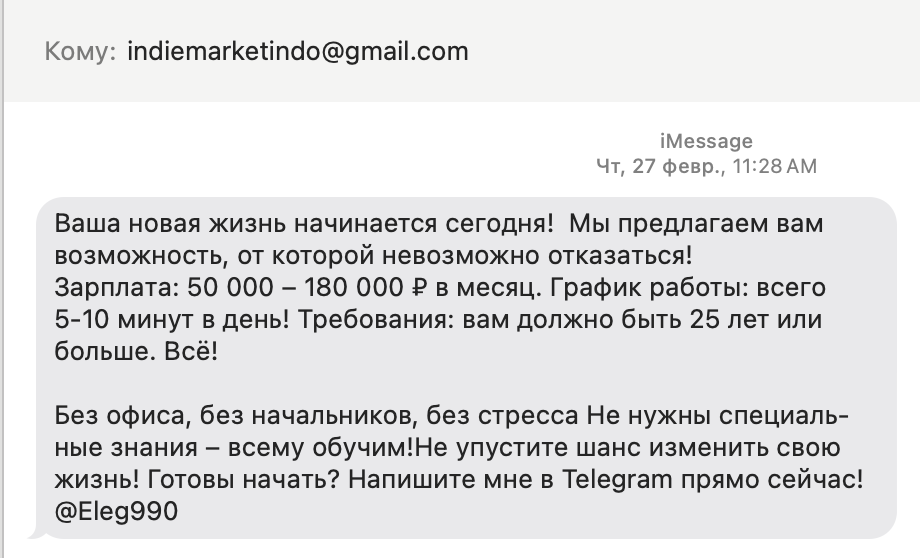
Для исследования и обучения моделей для классификации спама существуют множество доступных датасетов, которые можно использовать для обучения. Например, датасеты с реальными примерами спам-сообщений из различных источников (email, SMS, мессенджеры и т.д.). Это позволяет легко и быстро начать работу, обучая модель на реальных данных.

Изначально данные были на английском языке, что создавалось дополнительными трудностями при использовании их для анализа и обучения. Чтобы решить эту проблему, я воспользовался встроенным в Excel нейросетевым переводчиком и перевел 5000 строк данных, содержащих пары (метка, сообщение). Такой формат данных, где каждая строка содержит метку (спам/не спам) и текст сообщения, идеально подходит для задачи классификации, так как позволяет использовать метод обучения с учителем.

После настройки параметров и проведения процесса обучения, мы получили модель, способную классифицировать сообщения как спам или не спам на основе текста. Модель была обучена на 3 эпохах с размером батча 16 и коэффициентом обучения 2e-5. Модель показала точность 98% при классификации спам-сообщений.

Затем было разработано веб-приложение с использованием библиотеки Dash. В приложении реализованы следующие функциональные элементы: поле для ввода сообщения, отображение итогового результата (спам/не спам), а также сохранение истории запросов.

Для тестирования приложения использовались два реальных сообщения: одно — это спам, а другое — обычное сообщение, не являющееся спамом.



Копируем это сообщение и вставляем в наше приложение.

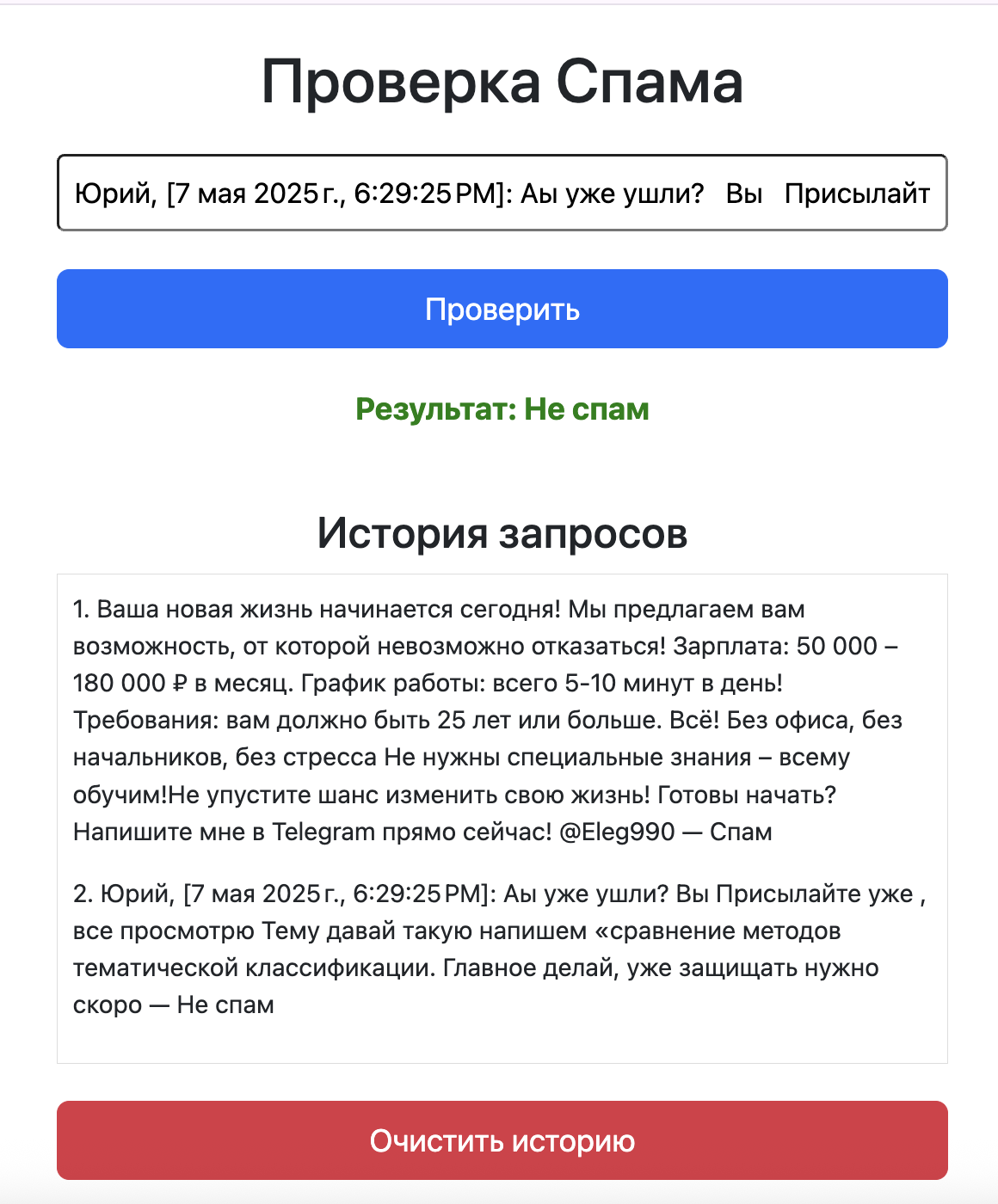
Изображение выглядит как текст, снимок экрана, Шрифт, веб-страница

Автоматически созданное описание

Видим, что модель правильно определило, что это спам. Далее берем обычное сообщение из переписки.

Изображение выглядит как текст, снимок экрана, Шрифт, дизайн

Автоматически созданное описание



Видим, что модель так же успешно определила, что сообщение не является спамом.