

УДК 004.08

ЧАТ-БОТ ДЛЯ TELEGRAM ДЛЯ ВЫЯВЛЕНИЯ ТОКСИЧНЫХ ПОЛЬЗОВАТЕЛЕЙ

Андропова М.А. (Университет ИТМО)

Научный руководитель – доцент, кандидат технических наук, Штенников Д.Г.
(Университет ИТМО)

Введение. В мире, в котором каждый считает своим правом высказать свое мнение, прикрываясь свободой слова, требуется постоянный контроль над растущим количеством нецензурной и токсичной лексики в социальных сетях. Особенно остро эта проблема стоит в общественных чатах, например, в университетских или рабочих. Токсичные люди оскорбляют, разрушают отношения между людьми, что приводит к постоянному стрессу. Они проявляют свою агрессию, ставя окружающих в неудобные ситуации. Все это может привести к потере самоуважения и нарушению психологического здоровья. Поэтому лучше избегать контакта с токсичными людьми, чтобы предотвратить такие последствия. Процесс модерации чатов затрудняется с ростом числа участников, и своевременно блокировать токсичных пользователей получается не всегда, что приводит к спорам или даже ссорам огромного количества людей. Чтобы предотвратить это, социальным сетям нужны новые решения, которые будут неустанно поддерживать вежливую атмосферу между людьми.

Основная часть. Моим решением стал чат-бот для Telegram, который отслеживает токсичные комментарии и присваивает каждому пользователю рейтинг. Если число рейтинга какого-то конкретного пользователя упадет до определенной отметки, модераторы чата будут оповещены, чтобы они уже лично применили меры по отношению к человеку. Для распознавания токсичной лексики создана нейросеть, на вход которой подается векторизированный текст. Сверточные слои строят карту признаков на основе сообщений пользователей, выделяя признаки токсичных и позитивных сообщений. Рекуррентный слой GRU необходим для обработки полученных векторов произвольной длины. Линейный слой делает заключительное решение по определению тональности текста на основе предыдущих слоёв.

Выводы. Бот успешно справляется со своей работой, выдавая точность порядка 0.85 и иногда определяя обычное сообщение как токсичное. Решить данную проблему я смогу, создав более качественный и большой датасет, а также усложнив архитектуру нейросети.

Андропова М.А. (автор)

Подпись

Штенников Д.Г. (научный руководитель)

Подпись