

УДК 004.85

**АНАЛИЗ ТОНАЛЬНОСТИ ТЕКСТОВ КЛИЕНТСКИХ ОТЗЫВОВ НА
БАНКОВСКУЮ ТЕМАТИКУ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО
ОБУЧЕНИЯ ДЛЯ ОБЕСПЕЧЕНИЯ ПРИОРИТЕЗАЦИИ ИХ РАССМОТРЕНИЯ.**

Гафарова Я.К. (Национальный исследовательский университет ИТМО), **Самоследов Г.А.**
(Национальный исследовательский университет ИТМО)

Воробьева А. А., (Национальный исследовательский университет ИТМО)

Научный руководитель – Воробьева А. А., канд. техн. наук, доцент,
(Национальный исследовательский университет ИТМО)

Аннотация: в работе рассматривается метод уменьшения времени, необходимого на выявление деструктивного информационного воздействия в текстах клиентских отзывов на банковскую тематику. Предлагается использовать машинное обучение для определения негативной коннотации текста как одного из основных факторов наличия деструктивного информационного воздействия.

В настоящее время большинство банковских организаций сталкиваются с направленным на них деструктивным информационным воздействием. Оно является фактором снижения доверия к организации, и, как следствие, ее прибыльности. Как следствие, организации заинтересованы в недопущении деструктивного информационного воздействия путем решения проблемы, содержащейся в данной информации. Поскольку главным источником информации является сеть Интернет, она подлежит мониторингу на предмет их выявления.

Главный источник деструктивного информационного воздействия в сети Интернет - доступная публично текстовая информация. Размер текстов может быть разными: большим (новости, статьи), средним (отзывы клиентов на специализированных ресурсах), маленьким (комментарии в соцсетях). Наиболее перспективными для анализа являются клиентские отзывы: в отличие от комментариев, они имеют достаточную длину; в то же время их количество намного больше, чем количество соответствующих новостных или авторских материалов.

Один из признаков деструктивного информационного воздействия в тексте – его негативная коннотация: специально обученный сотрудник организации осуществляет мониторинг определенных веб-сайтов на предмет наличия таких текстов. Однако такой подход очень трудозатратен: сотрудник просматривает отзывы по порядку вне зависимости от того, насколько быстро необходимо отреагировать: например, текст с негативной коннотацией требует более быстрого ответа, чем текст с благодарностью.

Целью данной работы является уменьшение времени, затрачиваемого на выявление деструктивного информационного воздействия в текстах клиентских отзывов на банковскую тематику.

Предлагается использовать машинное обучение для снижения трудозатрат сотрудников, выявляющих деструктивное информационное воздействие в текстах. В качестве текстового корпуса взяты отзывы с крупнейшего русскоязычного ресурса на банковскую тематику - banki.ru. Были взяты только отзывы, имеющие метку рейтинга – метка выставляется автором отзыва. Обучение модели производится на размеченных отзывах – предполагается, что отзывы с низким рейтингом имеют негативную коннотацию, а отзывы с высоким рейтингом – позитивную. Другая часть корпуса используется в качестве тестовой выборки. Классификатор определяет рейтинг отзыва от 1 до 5, после чего определенный классификатором рейтинг можно сравнить с выставленным клиентом и вычислить точность определения рейтинга. Необходимо отметить, что оценка, выставляемая пользователем, субъективна, и, например, отзыву с рейтингом «2» один эксперт может выставить оценку «1»,

а другой – «3», при этом текст отзыва будет один и тот же. Поэтому, если классификатор ошибался с рейтингом на +1 или -1, рейтинг все равно считался определенным корректно. В зависимости от рейтинга отзыву выставляется приоритет рассмотрения (чем ниже оценка, тем выше приоритет), в результате чего формируется очередь из отзывов.

В результате работы получен программный комплекс, позволяющий уменьшить затрачиваемое время на выявление деструктивного информационного воздействия в текстах клиентских отзывов на банковскую тематику. Данное программное средство отличает:

- 1) удобство работы - отображение в графическом интерфейсе очереди из отзывов в соответствии с приоритетами;
- 2) возможность определения рейтинга неразмеченных отзывов;
- 3) возможность работы с отзывами с других Интернет-ресурсов;
- 4) масштабирование – возможность подключения модулей для работы с текстами большого и малого размеров;
- 5) отсутствие затрат на программное обеспечение – используются бесплатные инструменты.
- 6) точность определения – до 80%.

В дальнейших исследованиях планируется проведение экспериментов, подтверждающих эффективность разработки. Также планируется увеличение точности определения тональности за счет использования оптимальных параметров для модели машинного обучения.

Гафарова Я.К. (автор)

Подпись

Самоследов Г.А. (автор)

Подпись

Воробьева А. А. (научный руководитель)

Подпись