

ОТЗЫВ НАУЧНОГО РУКОВОДИТЕЛЯ

на выпускную квалификационную работу обучающегося
Санкт-Петербургского государственного университета
Безруковой Дарьи Михайловны

«Анализ тональности текстов новостных источников по отношению к заданному объекту»

Выпускная квалификационная работа Безруковой Дарьи Михайловны посвящена сравнению методов анализа тональности текста и их применению по отношению к текстовым комментариям, оставленным на новостных и информационных порталах, предназначенных для сравнения банковских продуктов, а также разработке технологии автоматического выявления и оценки мнений.

Тема работы, несомненно, является актуальной, так как создание программного обеспечения для решения указанной задачи позволит в автоматическом режиме отслеживать степень удовлетворенности клиентов банков обслуживанием, предлагаемыми продуктами и услугами. Для банковского сектора извлечение такой информации является критичным, в силу высокой конкуренции в этой сфере. Кроме того, это избавит потенциальных клиентов от необходимости изучения отзывов, новостей и другой текстовой информации о рассматриваемом банке, но при этом позволит принять взвешенное решение, сэкономив силы и время.

Для решения поставленной задачи Безруковой Д.М. была разработана программа на языке Python с использованием библиотек открытого доступа таких как nltk, re, gensim.models.word2vec, pymorphy2, pandas, xgboost. С использованием созданной программы задача анализа тональности текста решалась в четыре этапа: сбор данных с веб-ресурса, предварительная обработка текста, перевод текста в вещественное пространство признаков и использование методов машинного обучения для последующей классификации тональности.

При работе над магистерской диссертацией Дарьей Михайловной был рассмотрен достаточно большой объем литературы как российских, так и зарубежных авторов, с целью выбора оптимальных алгоритмов и методов, применяемых на каждом этапе решения задачи анализа тональности текста. В частности, для векторизации текста были выбраны методы BagOfWords и Word2Vec, которые впоследствии сравнивались на реальных отзывах, загруженных с портала banki.ru. В результате были получены средние показатели точности классификации тональности текстов, 67% и 57% соответственно. Полученные показатели, во-первых, объясняются зачастую несоответствием оценки отзыва и его тональности, а во-вторых, свидетельствуют о необходимости более тщательной фильтрации и предварительной обработки текста.

Основными замечаниями к работе являются небрежное оформление ВКР, наличие стилистических ошибок в тексте, а также поверхностное изложение главы 1, посвященной предварительной обработке текста.

Как научный руководитель, оцениваю выполненную работу положительно, т.к. поставленная цель была достигнута, однако в силу указанных выше замечаний, считаю, что работа заслуживает оценки «хорошо».

Кандидат физ.-мат. наук, доцент кафедры
теории систем управления
электрофизической аппаратурой СПбГУ



А. Г. Головкина