

# **Отзыв рецензента**

## **о квалификационной работе бакалавра Лыщика Андрея Игоревича**

Работа Андрея Лыщика «Анализ эмоциональной окраски рецензий к фильмам» посвящена анализу и сравнению нескольких классических классификаторов для задачи оценки тональности (сентимент-анализа). Основная заслуга работы состоит в том, что в ней систематически сравниваются разные алгоритмы классификации в применении к специально подготовленному русскоязычному набору данных (с сайта kinopoisk.ru) для анализа тональности отзывов о фильмах.

Среди сильных сторон работы можно выделить широкий спектр протестированных в ней моделей, который приводит к тому, что результаты сравнения показательны и интересны. Например, один из интересных результатов – то, что ансамбль, построенный при помощи xgboost, оказался в целом ничем не лучше, чем ансамбль, основанный на обычном голосовании, и оба они лишь немногим лучше рассмотренных вариантов метода опорных векторов. Для каждой модели была проведена значительная работа по подбору гиперпараметров и параметров признаков, что позволило сравнивать оптимальные варианты моделей. Кроме того, работа содержит объёмный и содержательный обзор области.

Можно выделить два основных недочёта работы. Первый недостаток состоит в том, что несмотря на широкий спектр рассмотренных моделей, их список всё же неполон. Так, например, автор строит распределённые представления слов и даже параграфов (doc2vec), но при этом не пытается рассматривать современные методы анализа тональности, использующие эти распределённые представления (хотя для того, чтобы обучить базовую LSTM-модель, имея готовые распределённые представления слов, достаточно запустить готовый код с [https://github.com/fchollet/keras/blob/master/examples/imdb\\_lstm.py](https://github.com/fchollet/keras/blob/master/examples/imdb_lstm.py)). Второй недочёт работы — следовало бы внимательнее отнестись к терминологии. Например, на первой же странице идёт речь об «анализе эмоциональности», в то время как в русскоязычных источниках говорят либо об «анализе тональности», либо об «анализе эмоциональной окраски». Здесь же отмечу очень большое число пунктуационных и стилистических ошибок и опечаток в работе (практически на каждой странице, слишком много, чтобы перечислить в рецензии).

Тем не менее, я считаю, что представленная бакалаврская диссертация демонстрирует высокий профессиональный уровень автора и полностью удовлетворяет требованиям к таким квалификационным работам; рекомендую присвоить автору степень бакалавра, а работу оценить на «**отлично**» и допустить к защите.

рецензент,  
кандидат физико-математических наук  
Николенко Сергей Игоревич

