

О Т З Ы В

на выпускную квалификационную работу бакалавра IV курса
по направлению 01.03.02 «Прикладная математика и информатика»
основной образовательной программы СВ.5005
«Прикладная математика, фундаментальная информатика и программирование»

ШЕПЕЛЕВА Павла Леонидовича,

выполненную на тему:

«Методы анализа тональности отзывов пользователей ресторанов»

Выпускная квалификационная работа Шепелева П.Л. посвящена такому актуальному направлению в области обработки естественных языков, как анализ тональности опубликованных пользователями текстов, который имеет широкое применение в рекомендательных системах, сервисах для анализа работы бизнеса и маркетинга.

Целью выпускной квалификационной работы Шепелева П.Л. являлась разработка метода автоматического определения тональности русскоязычных отзывов посетителей ресторанов с точностью (F-мерой) не менее 60%. Для достижения поставленной цели задача была сведена к трех-классовой классификации с выделением позитивных, негативных или нейтральных отзывов, а также был собран корпус русскоязычных отзывов посетителей ресторанов для апробации и оценки качества предлагаемых подходов.

Кроме того, студентом были рассмотрены и реализованы различные способы предобработки текстов. В качестве признаков рассматривались различные комбинации n-грамм. Был проведен обзор и реализация следующих алгоритмов машинного обучения: наивный байесовский классификатор, логистическая регрессия, AdaBoost классификатор, метод стохастического градиента и метод опорных векторов. Для решения проблемы несбалансированных классов данных использовался алгоритм SMOTE. После проверки на кросс-валидации лучшее качество классификации показал алгоритм логистическая регрессия. Качество базового алгоритма классификации было улучшено с помощью добавления биграмм и применения алгоритма SMOTE к обучающему множеству.

К недостаткам выпускной квалификационной работы Власова И.О. можно отнести следующие:

1. Предложенный студентом подход на собранной тестовой выборке (с учетом решения проблемы несбалансированности классов в ней) имеет оценку качества $F\text{-score} = 0.64$, что является крайне низким значением для этой задачи. Конкурентно способным значением в условиях специфичности и зашумленности пользовательских данных является значение $F\text{-score}$ не ниже 0.75.
2. Автор не сравнивает решение с подходами на основе глубинного обучения, transfer learning и предобученных трансформеров типа BERT, XLNet и т.п., которые за последние 2 года показали прирост качества (по сравнению с классическими методами) на 18-35% по 18 задачам обработки естественного языка, включая задачу тональности тестов.

Указанные замечания не влияют на общую положительную оценку.

На основании вышеизложенного, с учетом замечаний, считаю, что выпускная квалификационная работа Шепелева П.Л. на тему «Методы анализа тональности отзывов пользователей ресторанов» выполнена в соответствии с предъявляемыми требованиями и заслуживает оценку «хорошо». Однако при этом допускаю, что со стороны комиссии может быть дополнительно учтен объем и качество защиты работы, что в конечном итоге может изменить оценку в положительную сторону.

Научный руководитель
Заведующий Кафедрой
технологии программирования
доцент, к.т.н.



Блеканов И.С.

28 мая 2020 г.