

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»

Баландина Максима Дмитриевича

выполненную на тему:

«Сегментация и классификация элементов web-страниц с помощью языковых моделей»

В квалификационной работе ставится задача классификации содержимого web-страниц, в частности выделения значимого текста и удаление «мусора», который не несет информационной нагрузки.

Первая часть работы посвящена анализу существующих подходов, в частности рассмотрены методы BTE, Web Page Cleaning with Conditional Random Fields, Web2Text, BoilerNet.

В практической части обосновывается выбор тренировочных и тестовых данных и приводится описание метода, который основан на предобученной модели DistilBERT и дальнейшей классификации векторного представления текста с помощью логистической регрессии. В заключительной части приводятся результаты, которые свидетельствуют о приемлемом качестве работы модели.

Тематика исследований несомненно актуальна, работа выполнена автором с высокой степенью самостоятельности, полученные результаты показывают хорошее владение инструментарием для обучения и применения нейронных сетей на практике.

Можно отметить следующие недостатки:

- присутствуют определенные замечания по оформлению формул и текста (например формулы на странице 9);
- выбор CleanEval в качестве основного тестового набора данных сомнителен, так как web очень быстро изменяется, набор данных 2008 года не является показательным в плане актуальности разметки;
- недостаточно четко обоснован выбор алгоритма классификации, отсутствует сравнение с другими алгоритмами бинарной классификации;
- слабо затронут вопрос производительности предложенного алгоритма;
- возможно следовало бы дообучить выбранную языковую модель для конкретной задачи на размеченных данных.

Тем не менее, считаю, что выпускная квалификационная работа удовлетворяет всем требованиям, предъявляемым к выпускным квалификационным работам бакалавра и заслуживает оценки **«отлично»**.

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



Блеканов И. С.

Научный соруководитель,
ст. преподаватель кафедры
технологии программирования



Мишенин А. Н.