

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра информационных систем управления

Аннотация к дипломной работе

«Разработка системы классификации текстов»

Мацулевич Никита Андреевич

Научный руководитель – старший преподаватель кафедры ИСУ Рубашко Н. К.

Минск, 2024

РЕФЕРАТ

Дипломная работа, 39 стр., 18 иллюстр., 7 источников.

Ключевые слова: МАШИННОЕ ОБУЧЕНИЕ, КЛАССИФИКАЦИЯ ТЕКСТОВ, МОДЕЛИ-ТРАНСФОРМЕРЫ, НАИВНЫЙ БАЙЕСОВСКИЙ АЛГОРИТМ, НЕЙРОННЫЕ СЕТИ, PYTHON, TENSORFLOW, KERAS.

Объект исследования: алгоритмы классификации текстов и инструменты для их разработки.

Предмет исследования: разработка системы классификации текстов по эмоциональной тональности.

Цель работы: изучение существующих алгоритмов и методов классификации текстов и разработка на их основе системы классификации текстов.

Методы исследования: сбор релевантного набора данных, программирование на языке Python, наблюдения, тестирование, сравнение результатов.

Результат: был собран набор текстовых данных на английском языке для обучения систем классификации в виде отзывов о товарах в интернет-магазине Amazon в количестве 25 тысячи экземпляров; разработаны системы классификации текстов по эмоциональной тональности, основанные на наивном байесовском алгоритме и моделях-трансформерах; проведен сравнительный анализ эффективности разработанных систем классификации.

Области применения: разработанные системы классификации могут быть использованы для анализа эмоциональных тоналностей текстовых документов.

ABSTRACT

Graduate work, 39 pages, 18 pictures, 7 sources.

Keywords: MACHINE LEARNING, TEXT CLASSIFICATION, TRANSFORMER MODELS, NAIVE BAYESIAN ALGORITHM, NEURAL NETWORKS, PYTHON, TENSORFLOW, KERAS.

The object of the study: text classification algorithms and tools for their development.

The subject of the study: development of a text classification system based on emotional tonality.

The purpose of the work: the study of existing algorithms and methods of text classification and the development of a text classification system based on them.

Research methods: collecting a relevant dataset, programming in Python, observations, testing, and comparing results.

Result: a set of text data in English was collected for teaching classification systems in the form of product reviews in the Amazon online store in the amount of 25 thousand copies; text classification systems based on emotional tonality based on a naive Bayesian algorithm and transformer models were developed; a comparative analysis of the effectiveness of the developed classification systems was carried out.

Areas of application: the developed classification systems can be used to analyze the emotional tonalities of text documents.