

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра математического моделирования и анализа данных

Аннотация к магистерской диссертации

**«Построение вопросно-ответной системы средствами машинного обучения
и прикладного семантического анализа текстовых данных»**

Антоненко Александр Олегович

Научный руководитель – доктор технических наук, доцент
Недзьведь А. М.

Минск, 2022

РЕФЕРАТ

Магистерская диссертация, 80 с., 16 рис., 8 таблиц, 18 источников.

Ключевые слова: ЕСТЕСТВЕННАЯ ОБРАБОТКА ЯЗЫКА, МАШИННОЕ ОБУЧЕНИЕ, РЕГРЕССИЯ, РЕГУЛЯРИЗАЦИЯ, КЛАССИФИКАЦИЯ, ТРАНСФОРМЕРЫ, ПОЛНОТА, F-МЕРА, ТОЧНОСТЬ, БУСТИНГ, БЭГГИНГ, СТЭКИНГ, СЕМАНТИЧЕСКИЙ АНАЛИЗ.

Объект исследования – методы обработки естественного языка, классические и ансамблевые методы машинного обучения и их модернизации для задач классификации.

Цель работы – построение вопросно-ответной системы средствами машинного обучения и прикладного семантического анализа текстовых данных.

Методы исследования – статистические методы обработки текстов, методы теории вероятностей и математической статистики, а так же современные методы машинного обучения.

В результате построена собственная вопросно-ответная система на основе композиции моделей машинного обучения, проведено поэтапное исследование, с последующим моделированием и сравнением качества работы всех изученных классических и ансамблевых методов машинного обучения, используемых в задачах классификации текстовых данных.

ABSTRACT

Master's thesis, 80 p., 16 fig., 8 tables, 18 sources.

Keywords: NATURAL LANGUAGE PROCESSING, MACHINE LEARNING, REGRESSION, REGULARIZATION, CLASSIFICATION, TRANSFORMERS, COMPLETENESS, F-MEASURE, ACCURACY, BOOSTING, BAGGING, STACKING, SEMANTIC ANALYSIS.

The object of study – natural language processing methods, classical and ensemble machine learning methods and their modernization for classification problems.

The purpose of the work – building a question-and-answer system by means of machine learning and applied semantic analysis of text data.

Research methods – statistical methods of text processing, methods of probability theory and mathematical statistics, as well as modern methods of machine learning.

As a result of work own proprietary question-and-answer system was built based on the composition of machine learning models, a step-by-step study was conducted, followed by modeling and comparing the performance of all studied classical and ensemble machine learning methods used in text data classification tasks.