2

# СОДЕРЖАНИЕ

**ВВЕДЕНИЕ 3**

**Глава 1. Погружение в проект 0**

1.1 Выбор темы практики 0

1.2 Описание проекта 0

1.3 Задание на практику 0

**Глава 2. Реализация приложения 0**

2.1 Реализация1 0

2.2 РеализацияN 0

**Глава 3. Описание результатов работы 0**

3.1 Результат1 0

3.2 РезультатN 0

**Глава 4. Описание командной работы 0**

**ЗАКЛЮЧЕНИЕ 0**

**СПИСОК ЛИТЕРАТУРЫ 0**

**ПРИЛОЖЕНИЕ 0**

3.1 Презентация проекта на студенческой научной конференции 0

3.3 Код методов 0

|  |  |  |
| --- | --- | --- |
| Ответственный за соответствие содержания и оформления текста отчёта предъявляемым требованиям | (подпись) | Ступников Андрей Анатольевич |

3

# ВВЕДЕНИЕ

Тип практики: технологическая (проектно-технологическая)

Дата начала: 26.02.2024

Дата окончания: 26.06.2024

Цель практики – получение первичных профессиональных умений и навыков, в том числе первичных умений и навыков научно-исследовательской деятельности, обучающихся по направлению подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем (уровень бакалавриата).

Задачи учебной практики:

* Закрепление теоретических и практических знаний, умений и навыков, полученных на первом и втором курсах обучения;
* Выполнение заданий, предусмотренных программой практики и назначенных руководителем;
* Освоение отдельных компьютерных программ и информационных систем, используемых в профессиональной деятельности;
* Получение навыков работы с периодическими, реферативными и справочными информационными изданиями по прикладной математике и информационным системам;
* Подготовка к осознанному и углубленному изучению дисциплин старших курсов;
* Подготовка и защита в установленный срок отчета по практике.

В результате выполнения технологической (проектно-технологической) учебной практики студент должен обладать следующими компетенциями:

* Анализ и обработка больших данных
* Machine learning
* Создание веб-приложения с помощью фреймворка Flask

4

**Глава 1. Погружение в проект**

**1.1 Выбор темы практики**

Участники нашей команды ранее были знакомы и выполняли совместную работу в рамках хакатона. Нам была интересна сфера машинного и, в частности, глубокого обучения.

Данный проект основывается на работе с текстом (NLP) – данная тема является одной из наиболее актуальных в сфере машинного обучения. Задача проекта была ясна. Мы владели навыками алгоритмического и объектно-ориентированного программирования, однако мы не владели компетенциями машинного обучения.

**1.2. Описание проекта**

Связность текста – это признак текста, при котором следующее предложение строится на базе предыдущего с помощью языковых средств [1]. Связность также является важным содержательным критерием, от которого зависит качество его реферирования [2] и сложность восприятия читателем [3]. Инструменты для оценки данного критерия могут найти применение в области обработки естественного языка, в частности при анализе сложности восприятия текста читателем.

Существующие сервисы для английского языка используют различные подходы, включающие семантические сети (WordNet), дистрибутивную семантику (LSA, word2vec), эвристические подходы подсчета повторяющихся слов [4-5]. При этом сервисы для русского языка с открытым описанием их принципа работы используют только последний подход, не учитывая словосочетания производных слов, родовидовые и ссылочные отношения, которые также являются средствами связности [6].

Цель работы – разработать веб-приложение для оценки локальной связности перевода на русский язык с помощью модели машинного обучения, учитывающей синтаксические, лексические и морфологические средства связи.

Задачи проекта:

* Сформировать и разметить датасет на основе 3503 текстов публицистической и информационной направленности корпуса несовершенных переводов [1].
* Провести сравнительный анализ регрессионных моделей машинного обучения
* Внедрить модель для оценки локальной связности текста.
* Создать приложение, предоставляющее пользователю оценку связности введённого им текста.