**Министерство науки и высшего образования Российской Федерации**
**Федеральное государственное бюджетное образовательное учреждение высшего образования**
**«Московский государственный технический университет имени Н.Э. Баумана**
**(национальный исследовательский университет)»**
**(МГТУ им. Н.Э. Баумана)**

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

# РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

## *К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ*

## *НА ТЕМУ:*

«Обзор методов анализа тональности естественно-языковых текстов»

Студент: ИУ7-53Б _____ М. Д. Маслова
(группа) (подпись, дата) (И. О. Фамилия)

Руководитель: _____ А. А. Оленев
(подпись, дата) (И. О. Фамилия)

*2022 г.*

# РЕФЕРАТ

Расчетно-пояснительная записка 11 с., 0 рис., 0 табл., 5 источн., 4 прил.
АНАЛИЗ ТОНАЛЬНОСТИ

# СОДЕРЖАНИЕ

# ВВЕДЕНИЕ

# 1 Анализ предметной области

## 1.1 Актуальность задачи

В современном мире огромную роль в жизни каждого человека играет Интернет. Люди общаются в социальных сетях, ведут блоги, оставляют отзывы о товарах, услугах, фильмах, книгах и т. п. За счет этого в открытом доступе находится огромный объем данных, который позволяет проводить точные анализы для решения каких-либо задач.

Большая часть накопленных данных представлена виде текстовой информации, поэтому становится актуальной задача анализа текстов на естественном языке. [1] Одной из этих задач является анализ тональности и сентимент-анализ. За счет того, что такой анализ может быть проведен для текста, написанного на любую тему, его применение возможно во многих сферах:

– мониторинг общественного мнения [2] относительно товаров и услуг, в том числе в режиме реального времени, с целью определения их достоинств и недостатков с точки зрения покупателей и улучшения их характеристик [3];

– анализ политических и социальных взглядов пользователей (например, влияние мер, предпринятых для борьбы с вирусом COVID-19, на жизнь людей) [3];

– исследование рынка и прогнозирование цен на акции [3];

– выявление случаев эмоционального насилия и пресечение противоправных действий [4].

Решение описанных задач требует анализ большого количества текстов, что делает невозможным их ручную обработку. Также при оценке тональности текста человеком трудно соблюсти критерии этой оценки. Таким образом, возникает необходимость в автоматизированных системах анализа.

При этом в отличие от традиционной обработки текста в анализе тональности незначительные вариации между двумя элементами текста существенно меняют смысл (например, добавление частицы "не"). Обработку естественного языка затрудняет обильное использования носителями средств выразительности и переносных значений слов и фраз. Также основной из проблем сентимент-анализа является разная окраска одного и того слова в текстах на различные тематики: слово, которое считается положительным в одной, в то же время считается отрицаельным в другой.

С учетом широкого применения анализ тональности и описанных сложностей, возникает необходимость в формализации поставленной проблемы и разработки методов для её решения.

## 1.2 Основные определения

**Анализ тональности текста** (sentiment analysis) – область компьютерной лингвистики, ориентированная на извлечение из текстов субъективных мнений и эмоций. **Тональность** – это мнение, отношение и эмоции автора по отношению к объекту, о котором говорится в тексте. Чаще всего под задачей анализа тональности текста понимают определение текста к одному из двух классов: "положительный" или "отрицательный". В некоторых случаях добавляют третий класс "нейтральных" текстов.[5]

В настоящее время выделяют три основных подхода к определению тональности текста:

– лексический анализ;

Абзац про него

– методы машинного обучения;

Абзац про него

– гибридные методы;

Абзац про него

Несмотря на различные подходы к решению задачи анализа тональности, во всех подходах требуется предварительная обработка текста, основными этапами которой являются:

единый регистр

удаление пунктуации

лемматизация/стемминг

удаление стоп слов

удаление шума (хэш-тэгов, ссылок и тд)

## 1.3 Формализация задачи

В данной работе ставится задача анализа методов определения принадлежности заданного естественно-языкового текста к одному из двух классов:

– положительный;

– отрицательный.

При этом определяется лишь **факт** принадлежности тому или иному классу, и оценка вероятности отношения текста к каждому классу не проводится.

Такая задача рассматривается во многих статьях, которые использовались при написании данной работы, что упрощает задачу сравнения методов ее решения и способствует получению объективных результатов оценки.

# 2 Описание существующих решений

## 2.1 Методы основанные на лексике

Надо найти

## 2.2 Методы машинного обучения

### 2.2.1 Наивный Байес

### 2.2.2 Логическая регрессия

### 2.2.3 k ближайших соседей

### 2.2.4 что-то там про лес и деревья ;)

### 2.2.5 Нейронки

## 2.3 Гибридные

Общее описание

# 3 Классификация существующих решений

## 3.1 Технология метода

machine learning
lexicon approach
hybrid

## 3.2 Уровни

Здесь небходимо пояснить за семантические связи. "Еда вкусная, но обслуживание так себе". В целом – скорее всего нейстральный, но по аспектам: о еде: положительно, обобслуживании: отрицательно.

document
sentence
approach

## 3.3 Скорость

## 3.4 Данные/память

## 3.5 Точность

## 3.6 Время разработки

## 3.7 По предварительной обработке данных

# 4 Заключение

В ходе данной работы было выявлено:

– преобладание методов машинного обучения в данной сфере за счет ...;

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ


1. *Богданов А. Л.*, *Дуля И. С.* Сентимент-анализ коротких русскоязычных текстов в социальных медиа // Вестн. Том. гос. ун-та. Экономика. — 2019. — № 47. — С. 220—241. — URL: https://cyberleninka.ru/article/n/sentiment-analiz-korotkih-russkoyazychnyh-tekstov-v-sotsialnyh-media (дата обращения: 15.12.2021).

2. *Майорова Е. В.* О сентимент-анализе и перспективах его применения // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 6, Языкознание: Реферативный журнал. — 2020. — № 4. — С. 78—87. — URL: https://cyberleninka.ru/article/n/o-sentiment-analize-i-perspektivah-ego-primeneniya (дата обращения: 15.12.2021).

3. *Sharma A.* Natural Language Processing and Sentiment Analysis // International Research Journal of Computer Science. — 2021. — Т. 8. — С. 237—242. — URL: https://www.researchgate.net/publication/355927843_NATURAL_LANGUAGE_PROCESSING_AND_SENTIMENT_ANALYSIS (дата обращения: 15.12.2021).

4. *Колмогорова А. В.* Использование текстов жанра «Интернет-откровение» в контексте решения задач сентимент-анализа // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. — 2019. — № 3. — С. 71—82. — URL: https://cyberleninka.ru/article/n/ispolzovanie-tekstov-zhanra-internet-otkrovenie-v-kontekste-resheniya-zadach-sentiment-analiza (дата обращения: 15.12.2021).

5. *Самигулин Т. Р.*, *Джурабаев А. Э. У.* Анализ тональности текста методами машинного обучения // Научный результат. Информационные технологии. — 2021. — № 1. — URL: https://cyberleninka.ru/article/n/analiz-tonalnosti-teksta-metodami-mashinnogo-obucheniya (дата обращения: 15.12.2021).