



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:	<u>ИУ7-53Б</u> (группа)	_____ (подпись, дата)	<u>М. Д. Маслова</u> (И. О. Фамилия)
Руководитель:		_____ (подпись, дата)	<u>А. А. Оленев</u> (И. О. Фамилия)

2022 г.

## **РЕФЕРАТ**

Расчетно-пояснительная записка 12 с., 0 рис., 0 табл., 19 источн., 4 прил.  
**АНАЛИЗ ТОНАЛЬНОСТИ**

# СОДЕРЖАНИЕ

<b>РЕФЕРАТ</b>	<b>2</b>
<b>ВВЕДЕНИЕ</b>	<b>4</b>
<b>1 Анализ предметной области</b>	<b>5</b>
1.1 Основные определения . . . . .	5
1.2 Важность/актуальность . . . . .	5
1.3 Формализация задачи . . . . .	5
1.4 Препроцессинг . . . . .	5
<b>2 Описание существующих решений</b>	<b>6</b>
2.1 Методы основанные на лексике . . . . .	6
2.2 Методы машинного обучения . . . . .	6
2.2.1 Наивный Байес . . . . .	6
2.2.2 Логическая регрессия . . . . .	6
2.2.3 к ближайших соседей . . . . .	6
2.2.4 что-то там про лес и деревья ;) . . . . .	6
2.2.5 Нейронки . . . . .	6
2.3 Гибридные . . . . .	6
<b>3 Классификация существующих решений</b>	<b>7</b>
3.1 Технология метода . . . . .	7
3.2 Уровни . . . . .	7
3.3 От тематики??? . . . . .	7
3.4 Скорость . . . . .	7
3.5 Данные/память . . . . .	7
3.6 Точность . . . . .	7
3.7 Время разработки . . . . .	7
3.8 Язык??? . . . . .	7
3.9 По предварительной обработке данных . . . . .	7
3.10 Влияние частей речи??? . . . . .	7
<b>4 Заключение</b>	<b>8</b>



# **ВВЕДЕНИЕ**

# **1 Анализ предметной области**

## **1.1 Основные определения**

Тональность

Анализ тональности

Предобработка

Корпус

Словарь

Позитивный/негативный/нейтральный

## **1.2 Важность/актуальность**

-> Интернет жизнь

-> можно узнать мнение народа ;) о товаре

-> отношение к ситуации (политической, экономической, любой другой, настроение масс, так сказать)

-> регулировка комментариев (удаление нецензурных комментариев, оскорблений, блокировка пользователей и тд)

## **1.3 Формализация задачи**

Три класса текста. Задача в определении текста к одному из классов (без вероятности).

## **1.4 Препроцессинг**

единый регистр

удаление пунктуации

лемматизация/стемминг

удаление стоп слов

удаление шума (хэш-тэгов, ссылок и тд)

## **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**

document

sentence

approach

#### **3.3 От тематики???**

#### **3.4 Скорость**

#### **3.5 Данные/память**

#### **3.6 Точность**

#### **3.7 Время разработки**

#### **3.8 Язык???**

#### **3.9 По предварительной обработке данных**

#### **3.10 Влияние частей речи???**



## **4 Заключение**

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *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-analyze-i-perspektivah-ego-primeneniya> (дата обращения: 15.12.2021).
2. *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).
3. *Семина Т. А.* Анализ тональности текста: современные подходы и существующие проблемы // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 6, Языкознание: Реферативный журнал. — 2020. — № 4. — С. 47—64. — URL: <https://cyberleninka.ru/article/n/analiz-tonalnosti-teksta-sovremennye-podhody-i-suschestvuyuschie-problemy> (дата обращения: 15.12.2021).
4. *Khalil E., El Houby E., Mohamed H.* Sentiment Analysis Tasks and Approaches // International Journal of Computer Science and Information Security. — 2021. — Т. 19. — URL: [https://www.researchgate.net/publication/356068600\\_Sentiment\\_Analysis\\_Tasks\\_and\\_Approaches](https://www.researchgate.net/publication/356068600_Sentiment_Analysis_Tasks_and_Approaches) (дата обращения: 15.12.2021).
5. *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](https://www.researchgate.net/publication/355927843_NATURAL_LANGUAGE_PROCESSING_AND_SENTIMENT_ANALYSIS) (дата обращения: 15.12.2021).
6. *Pathak A., Sharma S., Pandey R.* A Methodological Survey on Sentiment Analysis Techniques and Their Applications in Opinion Mining // International

- Journal of Emerging Trends in Engineering and Development. — 2021. — Т. 1. — С. 37–45. — URL: [https://www.researchgate.net/publication/349154400\\_A\\_METHODOLOGICAL\\_SURVEY\\_ON\\_SENTIMENT\\_ANALYSIS\\_TECHNIQUES\\_AND\\_THEIR\\_APPLICATIONS\\_IN\\_OPINION\\_MINING](https://www.researchgate.net/publication/349154400_A_METHODOLOGICAL_SURVEY_ON_SENTIMENT_ANALYSIS_TECHNIQUES_AND_THEIR_APPLICATIONS_IN_OPINION_MINING) (дата обращения: 15.12.2021).
7. *А. Л. Богданов И. С. Д.* 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).
  8. Sentiment Analysis for Mining Texts and Social Networks Data: Methods and Tools / C. Zucco [и др.] // WIREs Data Mining and Knowledge Discovery. — 2020. — Т. 10, № 1. — URL: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1333> (дата обращения: 15.12.2021).
  9. Opinion mining and sentiment analysis / R. K. Bakshi [и др.] // 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). — 2016. — С. 452–455. — URL: <https://www.semanticscholar.org/paper/Opinion-mining-and-sentiment-analysis-Bakshi-Kaur/87ee57c4915c5f7c5c7b9486df0c8af7e31bd747> (дата обращения: 15.12.2021).
  10. *Т. Р. Самигулин А. Э. У. Д.* 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).
  11. *Пескишева Т. А.* Методы анализа тональности текстов на естественном языке // Общество. Наука. Инновации (НПК-2017). — 2017. — С. 1730–1742. — URL: (дата обращения: 15.12.2021).
  12. *Berrar D.* Bayes' Theorem and Naive Bayes Classifier. — 2018. — URL: [https://www.researchgate.net/publication/324933572\\_Bayes%27\\_Theorem\\_and\\_Naive\\_Bayes\\_Classifier](https://www.researchgate.net/publication/324933572_Bayes%27_Theorem_and_Naive_Bayes_Classifier) (дата обращения: 15.12.2021).

13. Sentiment analysis using logistic regression algorithm / Y. Jaswanth [и др.] // Т. 7. — 2020. — URL: <https://www.semanticscholar.org/paper/Sentiment-analysis-using-logistic-regression-Jaswanth-Kumar/1af4aaa6670a8bf62460ef69476ead4f984993af> (дата обращения: 15.12.2021).
14. Zvonarev A., Bilyi A. A Comparison of Machine Learning Methods of Sentiment Analysis Based on Russian Language Twitter Data // MICSECS. — 2019. — URL: <https://www.semanticscholar.org/paper/A-Comparison-of-Machine-Learning-Methods-of-Based-Zvonarev-Bilyi/3f1a179dab449e32677aac22605e37ee1ed05f74> (дата обращения: 15.12.2021).
15. Rajput A. Natural Language Processing, Sentiment Analysis and Clinical Analytics. — 2019. — URL: [https://www.researchgate.net/publication/330871275\\_Natural\\_Language\\_Processing\\_Sentiment\\_Analysis\\_and\\_Clinical\\_Analytics](https://www.researchgate.net/publication/330871275_Natural_Language_Processing_Sentiment_Analysis_and_Clinical_Analytics) (дата обращения: 16.12.2021).
16. Research on Sentiment Dictionary Based on Sentiment Analysis in News Domain / Y. Gao [и др.] // 2021 7th IEEE Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS). — 2021. — С. 117–122. — URL: <https://www.semanticscholar.org/paper/Research-on-Sentiment-Dictionary-Based-on-Sentiment-Gao-Su/3b6e3c91c59b7654c9b701d84e879334f4427a27> (дата обращения: 16.12.2021).
17. Sadia A., Khan F. K., Bashir F. An Overview of Lexicon-Based Approach For Sentiment Analysis // — 2018. — URL: <https://www.semanticscholar.org/paper/An-Overview-of-Lexicon-Based-Approach-For-Sentiment-Sadia-Khan/e53033c31e6ee88bad1cb3da4b122be60a53d4d5> (дата обращения: 16.12.2021).

18. Sentiment Analysis using Lexicon based Approach / M. Pandey [и др.] // . — 2019. — URL: <https://www.semanticscholar.org/paper/Sentiment-Analysis-using-Lexicon-based-Approach-Pandey-Williams/dd26e84dd90a1ed3db58980fcaa8bebfa885b964> (дата обращения: 16.12.2021).
19. *Cindo M., Rini D. P., Ermatita E.* Sentiment Analysis on Twitter by Using Maximum Entropy and Support Vector Machine Method // . — 2020. — URL: <https://www.semanticscholar.org/paper/SENTIMENT-ANALYSIS-ON-TWITTER-BY-USING-MAXIMUM-AND-Cindo-Rini/fe502d0ad0d971435b51b5b8f2b8aafe4183792c> (дата обращения: 16.12.2021).