

Анализ тональности отзывов с помощью библиотеки TextBlob и VaderSentiment

Анишкова Анастасия Сергеевна

Приамурский государственный университет имени Шолом-Алейхема
Студент

Аннотация

Целью данной работы является проведение анализа тональности текстовых отзывов с использованием библиотек Python — TextBlob и VADERSentiment, а также сравнение их эффективности в определении эмоциональной окраски текста. Результаты исследования могут быть использованы в образовательных целях для изучения методов анализа текста и машинного обучения.

Ключевые слова: анализ тональности, TextBlob, VaderSentiment, NLTK.

Analyzing the sentiment of reviews using the TextBlob or VaderSentiment library

Anishkova Anastasia Sergeevna

Sholom-Aleichem Priamurskiy State University
Student

Abstract

The purpose of this work is to analyze the sentiment of text reviews using Python libraries such as TextBlob and VADERSentiment, and to compare their effectiveness in determining the emotional tone of a text. The results of this research can be used for educational purposes to learn about text analysis and machine learning techniques.

Key words: tonality analysis, TextBlob, VaderSentiment, NLTK.

1 Введение

1.1 Актуальность

Анализ тональности текстовых отзывов играет важную роль в обработке естественного языка и используется для оценки эмоциональной окраски высказываний в социальных сетях, интернет-магазинах и других источниках. В условиях роста объема пользовательских данных автоматизация анализа становится особенно значимой. Для решения этой задачи применяются библиотеки, такие как TextBlob и VADER, позволяющие эффективно определять позитивную, негативную или нейтральную тональность текста.

TextBlob удобен для базового анализа и прост в использовании, однако имеет ограничения в понимании сложных лингвистических конструкций. VADER, напротив, показывает высокую точность при обработке разговорных и социальных текстов, учитывая особенности языка, такие как отрицание,

усиление и контекст. Это делает его более подходящим инструментом для анализа реальных отзывов.

Сравнение и применение данных методов позволяет выявить наиболее эффективные подходы к анализу текста, что актуально как в научных исследованиях, так и в практических задачах, связанных с обработкой больших объемов данных.

1.2 Обзор исследований

М. А. Кирина, Л. Д. Тельнина реализовала автоматическую оценку впечатлений, обучающихся методами анализа тональности (на материале отзывов на онлайн-курсы на русском и английском) [1]. Датасетом и разработкой инструмента учебной аналитики для извлечения проявлений студенческой агентности из текстов отзывов на MOOK поделилась Ю. Ю. Дюличева [2]. А. Д. Егорова, Н. В. Станкевич описала сентиментные траектории как инструмент анализа текста [3]. COVID-19: Анализ эмоциональной окраски сообщений в социальных сетях (на материале сети «Twitter») описала П. А. Денисова [4]. Э. С. Нагапетян оценила влияния настроения населения на процентные ставки по кредитам и депозитам банков РА [5].

1.3 Цель исследования

Целью данной работы является проведение анализа тональности текстовых отзывов с использованием библиотек Python — TextBlob и VaderSentiment, а также сравнение их эффективности в определении эмоциональной окраски текста.

2 Материалы и методы

Для проведения исследования использовалась среда Google Colab, язык программирования Python и библиотеки TextBlob и VaderSentiment.

3 Результаты

Для эффективного анализа тональности текста загрузим библиотеки TextBlob и VADER, обеспечивающие простой синтаксический разбор и точную оценку эмоциональной окраски сообщений, а также подключим инструментальный NLTK вместе с лексиконом VADER, содержащим широкий спектр терминов и оценочных метрик, необходимых для точного распознавания настроений в современных коммуникациях (рис.1).

```
✓ 15
OK. # Установка необходимых библиотек (если ещё не установлено)
!pip install textblob
!pip install vaderSentiment

# Загрузка корпуса для VADER (впервые)
import nltk
nltk.download('vader_lexicon')

Requirement already satisfied: textblob in /usr/local/lib/python3.11/dist-packages (0.19.0)
Requirement already satisfied: nltk<=3.9 in /usr/local/lib/python3.11/dist-packages (from textblob) (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk<=3.9->textblob) (8.2.1)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk<=3.9->textblob) (1.5.1)
Requirement already satisfied: regex<=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk<=3.9->textblob) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk<=3.9->textblob) (4.67.1)
Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl.metadata (572 bytes)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from vaderSentiment) (2.32.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (3.4.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (2.4.0)
Requirement already satisfied: certifi<=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (2025.4.26)
Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
126.0/126.0 kB 2.6 MB/s eta 0:00:00
Installing collected packages: vaderSentiment
Successfully installed vaderSentiment-3.3.2
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
True
```

Рисунок 1. Установка необходимых библиотек и данных

Выполним анализ тональности (sentiment analysis) текстовых отзывов с использованием библиотеки TextBlob (рис.2). Для этого импортируем класс TextBlob, который позволяет легко выполнять обработку и анализ текста, включая определение эмоциональной окраски. Создадим список из трёх примеров отзывов на английском языке: положительный, отрицательный, нейтральный/смешанный.

Для каждого отзыва создаётся объект TextBlob, который автоматически анализирует текст. Тональность определяется на основе значения полярности (оно находится в диапазоне от -1 до 1):

- 0 → положительный (positive);
- <0 → отрицательный (negative);
- = 0 → нейтральный (neutral).

Анализируя три текстовых отзыва на английском языке TextBlob успешно определяет эмоциональную окраску (тональность) каждого из них. Настроение первых двух отзывов — явно позитивного и явно негативного. Третий отзыв, который кажется нейтральным или слегка позитивным, получил слабо негативную оценку, потому что анализ тональности учитывает не только содержание, но и лингвистическую структуру.



```
from textblob import TextBlob

# Пример отзывов
reviews = [
    "Мне нравится этот продукт, он потрясающий!",
    "Это самый худший опыт, который у меня когда-либо был.",
    "Все в порядке, не так уж плохо.",
]

print("Анализ настроений текстовых блоков:")
for review in reviews:
    analysis = TextBlob(review)
    # polarity: от -1 (очень негативный) до 1 (очень позитивный)
    sentiment = "положительный" if analysis.sentiment.polarity > 0 else "негативный" if analysis.sentiment.polarity < 0 else "нейтральный"
    print(f"Обзор: {review}")
    print(f"Настроение: {sentiment} (полярность: {analysis.sentiment.polarity:.2f})\n")
```

Анализ настроений текстовых блоков:
Обзор: Мне нравится этот продукт, он потрясающий!
Настроение: нейтральный (Полярность: 0.00)

Обзор: Это самый худший опыт, который у меня когда-либо был.
Настроение: нейтральный (Полярность: 0.00)

Обзор: Все в порядке, не так уж плохо.
Настроение: нейтральный (Полярность: 0.00)

Рисунок 2. Анализ тональности с помощью TextBlob

Выполним анализ тональности текстовых отзывов на английском языке с использованием VADER (Valence Aware Dictionary and sEntiment Reasoner) — инструмента из библиотеки NLTK, специально разработанного для анализа социальных и разговорных текстов. Сначала импортируем модуль VADER для анализа эмоциональной окраски текста и создадим экземпляр анализатора, готовый оценивать тексты.

Для каждого отзыва вычисляется совокупный балл (compound score) — значение от -1 (максимально негативный) до +1 (максимально позитивный):

- Если оценка $\geq 0.05 \rightarrow$ отзыв считается положительным;
- Если оценка $\leq -0.05 \rightarrow$ отзыв считается негативным.

Результат анализа тональности с помощью VADER показал, что модель корректно определила настроение всех трёх отзывов, включая мягкие и разговорные формулировки, продемонстрировав высокую точность в оценке эмоциональной окраски текста.



```
from nltk.sentiment.vader import SentimentIntensityAnalyzer

# Инициализация анализатора
sia = SentimentIntensityAnalyzer()

print("Анализ настроений:")
for review in reviews:
    score = sia.polarity_scores(review)['compound']
    if score >= 0.05:
        sentiment = "позитивный"
    elif score <= -0.05:
        sentiment = "негативный"
    else:
        sentiment = "нейтральный"
    print(f"Обзор: {review}")
    print(f"Настроение: {sentiment} (Совокупный балл: {score:.2f})\n")
```

Анализ настроений:
Обзор: I love this product, it's amazing!
Настроение: позитивный (Совокупный балл: 0.85)

Обзор: This is the worst experience I've ever had.
Настроение: негативный (Совокупный балл: -0.62)

Обзор: It's okay, not too bad.
Настроение: позитивный (Совокупный балл: 0.58)

Рисунок 3. Анализ тональности с помощью VADERSentiment

4 Выводы

Проведя сравнительный анализ результатов анализа тональности текстовых отзывов с использованием TextBlob и VADER, можно сделать вывод, что VADER является более точным и надёжным инструментом для анализа эмоциональной окраски текста. Он особенно эффективен при обработке реальных пользовательских отзывов, разговорной речи и социальных комментариев, учитывая такие лингвистические особенности, как отрицание, усиление и контекст.

В то же время TextBlob, несмотря на простоту использования и удобство для базового анализа, имеет ограничения в понимании тонкостей языка, что может привести к некорректной оценке мягких или нейтральных формулировок.

Библиографический список

1. Кирина М. А., Тельнина Л. Д. Автоматическая оценка впечатлений, обучающихся методами анализа тональности (на материале отзывов на онлайн-курсы на русском и английском) //Цифровая гуманитаристика и технологии в образовании (DHTE 2022). 2022. С. 355-374.
2. Дюличева Ю. Ю. Датасет и разработка инструмента учебной аналитики для извлечения проявлений студенческой агентности из текстов отзывов на MOOK //Вопросы образования. 2024. №. 1. С. 288-313.
3. Егорова А. Д., Станкевич Н. В. Сентиментные траектории как инструмент анализа текста //Культура и текст. 2024. №. 2 (57). С. 169-180.

4. Денисова П. А. COVID-19: Анализ эмоциональной окраски сообщений в социальных сетях (на материале сети «Twitter»): магистерская диссертация: дис. б. и., 2021.
5. Нагапетян Э. С. Оценка влияния настроения населения на процентные ставки по кредитам и депозитам банков РА //Регион и мир. 2021. №. 1. С. 167-171.