

Telegram бот для эмоционального анализа новостей

Ли Михаил Русланович
10 класс ГБОУ г. Москвы №1507

Руководитель: Русаков Алексей Михайлович

Преподаватель детского технопарка «Альтаир»

Москва, 2022

Оглавление

Оглавление	1
------------------	---

Введение	3
Цель	3
Задачи.....	3
Методы и этапы	4
Актуальность.....	4
Использованные программного решения	4
Обзор проектного решения	4
Результаты тестирования и реальное применение.....	6
Итоги и перспективы.....	7
Список литературы.....	7

Введение

Публикации привлекают внимание и лучше запоминаются, оставляя какую-то эмоцию. Данный Telegram бот поможет определить эту эмоцию.

Цель

Разработать бота, определяющего эмоциональную окраску новостей.

Задачи

- 1) Определить функционал разрабатываемого бота
- 2) Провести анализ используемых для реализации проекта технологий
- 3) Создать Telegram бота
- 4) Написать код для определения эмоциональной окраски новостей
- 5) Подключить написанный алгоритм к боту
- 6) Протестировать полученный результат, исправить ошибки
- 7) Определить перспективы развития проекта

Методы и этапы

Изучение материалов по данной теме	<ol style="list-style-type: none">1. Знакомство с имеющимися технологиями2. Опросы специалистов3. Личный опыт
Программирование	<ol style="list-style-type: none">1. Изучение языка Python2. Создание бота3. Разработка алгоритма
Внедрение	<ol style="list-style-type: none">1. Выбор хостинга

Актуальность

Публикации, которые вызывают яркие эмоции, чаще получают хороший охват и отклик аудитории. Данный бот дает возможность использовать любое мобильное или компьютерное средство для определения эмоциональной окраски новости.

Использованные программные решения

Язык программирования: Python

Среда разработки: PyCharm

Использованные библиотеки: scikit-learn, nltk, pytelegrambotapi

Обзор проектного решения

Мессенджер Telegram является одним из самых популярных мессенджеров. Он легкодоступен, как для телефонов, так и для компьютеров.

На сервере с помощью библиотеки pytelegrambotapi осуществлен доступ к боту и его функционалу через индивидуальный защищенный токен. Бот имеет команды:

/help – показать список команд (рис. 1)

/t {текст} – проанализировать новость

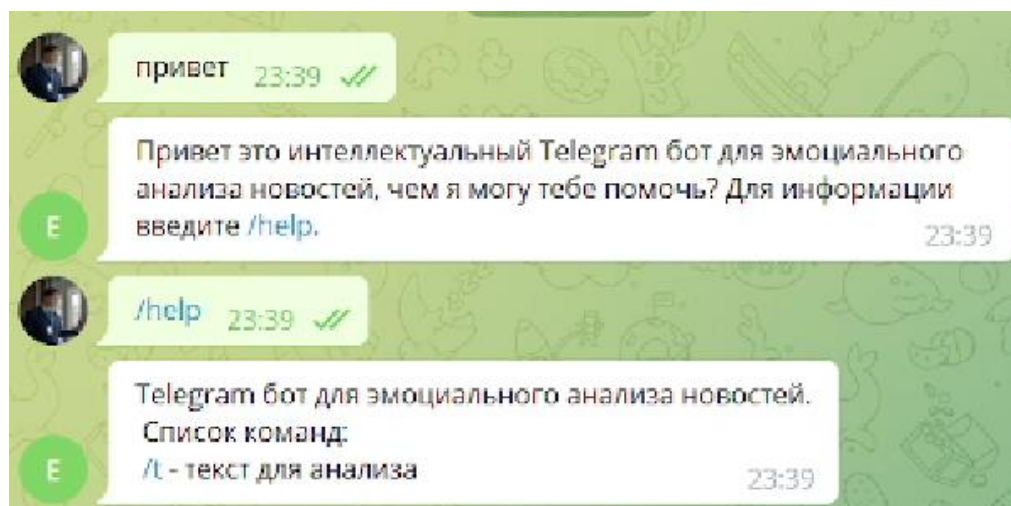


Рис. 1

Анализ текста происходит с помощью библиотеки `scikit-learn`. В этой библиотеке находится большое количество алгоритмов для задач, связанных с классификацией и машинным обучением в целом. Используя их, мы создаем модель (рис. 2) на основе подготовленных данных (рис. 3), которую мы можем использовать для эмоционального анализа текста.

```
text_clf = Pipeline([
    ('tfidf', TfidfVectorizer('ngramrange=(1,2)'),),
    ('clf', RandomForestClassifier())
])
text_clf.fit(texts, texts_labels)
stop_words = set(stopwords.words('english') + stopwords.words('russian'))
word_tokens = word_tokenize(word_list)
filtered_words = [w for w in word_tokens if not w in stop_words]
filtered_words = []
for w in word_tokens:
    if w not in stop_words:
        filtered_words.append(w)
# print(word_list)
filtered_words = [' '.join(filtered_words)]
print(filtered_words)
res = text_clf.predict(filtered_words)
return (res)
```

Рис. 2

910	Печаль	Разочарование
911	Печаль	Тоска
912	Печаль	Сочувствие
913	Печаль	Отчаяние
914	Печаль	Скорбь
915	Печаль	Сожаление
916	Печаль	Огорчение
917	Печаль	Жалость
918	Печаль	Горе
919	Страх	Страх
920	Страх	Беспокойство
921	Страх	Тревога
922	Страх	Ужас

Рис. 3

Результаты тестирования и реальное применение

Принцип работы бота един – он получает текст для анализа, использует методы и алгоритмы библиотеки `scikit-learn`, которые строят модель на основе данных. Затем бот возвращает результат пользователю (рис. 4).



Рис. 4

Итоги и перспективы

При выполнении работы был определен функционал бота, проанализированы современные методы решения проблемы. Изучены основные алгоритмы для анализа текста библиотеки scikit-learn. В итоге был создан Telegram бот, выполняющий свою основную функцию – эмоциональный анализ новостей.

Следующими направлениями разработки данного приложения могут быть:

1. Реализация поддержки различных языков
2. Выбор хостинга
3. Добавление большего количества эмоций

Список литературы

1. [Электронный ресурс] Язык программирования Python 3 для начинающих и чайников] URL: <https://pythonworld.ru/>
2. [Учебник] Шихи Д. Серия: Структуры данных в Python. Начальный курс. Изд-во ДМК-Пресс. 2021.
3. [Учебник] Джейми Чан. Серия: Библиотека программиста. Python. Быстрый старт. Изд-во Питер. 2021.
4. [Электронный ресурс] Руководство по pyTelegramBotAPI Дата обновления 01.02.2022. URL: <https://github.com/eternnoir/pyTelegramBotAPI>
5. [Электронный ресурс] Официальная документация по созданию ботов Telegram Дата обновления 01.02.2022. URL: <https://core.telegram.org/bots/api>
6. [Электронный ресурс] Библиотека NLTK - Natural Language Toolkit Дата обновления 01.02.2022. URL: <https://www.nltk.org/>
7. [Учебник] Бенджамин Б., Ребекка Б., Тони О. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. – " Издательский дом"" Питер""", 2018.