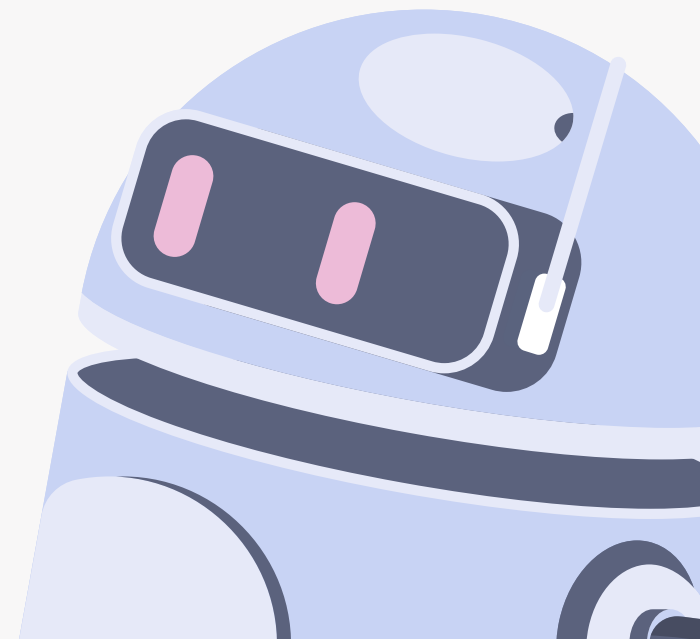
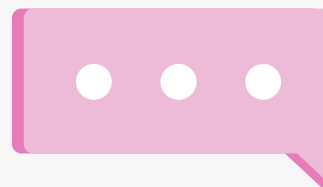
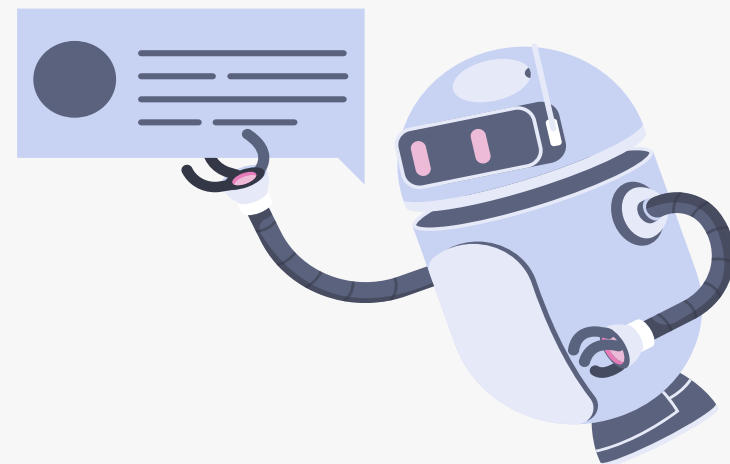


Анализ тональности русскоязычного текста



Наша команда



Маринич Артём



Data scientist
marinich_art@mail.ru

Панфёров Владислав



Fullstack-разработчик
panferov.public@gmail.com

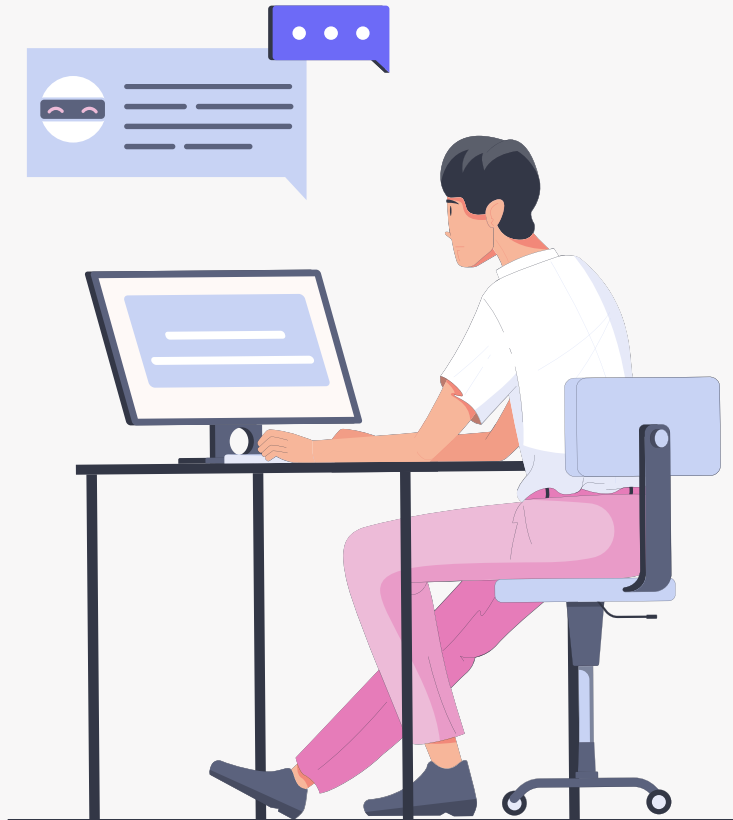
Маркин Антон



Разработчик ботов
antonimarki2006@gmail.com

Цели участия в хакатоне

1. Популяризация тематики применения искусственного интеллекта;
2. Создание общедоступного ПО на микросервисной архитектуре для решения практических задач;
3. Развитие компетенций в разработке ПО в сжатые сроки, используя технологии искусственного интеллекта.



Наше решение

Основные используемые модели/библиотеки:

- **Dostoevsky**
- **Natasha**
- **Rubert**

Была использована предобученная модель RuBert для извлечения признаков (features), голова модели была обучена на наших данных.

При использовании модели Dostoevsky удалось добиться значения метрики f1-score 0.809 на более, чем 226 тыс. тестовых сообщениях.

Наш продукт **уже** можно использовать для: определения реакции населения на новости/события;
Реакция пользователей на обновление приложения.

```
for row in tqdm(df_res.itertuples()):
    text = []
    text.append(row.text)
    results = dostoevsky_model.predict(text, k=5)
    del results[0]['skip']
    del results[0]['speech']
    results[0][0] = results[0].pop("neutral")
    results[0][-1] = results[0].pop("negative")
    results[0][1] = results[0].pop('positive')
    ordered_dict = dict(sorted(results[0].items(), reverse=True, key=lambda x: x[1]))
    first_key = next(iter(ordered_dict))
    df_res.iloc[row.Index, 2] = first_key
```

226834it [08:25, 448.78it/s]

```
f1_score(df_res['is_positive'], df_res['predictions'])
```

0.8089100218907422

Датасет:

<https://www.kaggle.com/datasets/thorinhood/russian-twitter-sentiment>

Наше решение Rest API

Основные используемые модели/библиотеки:

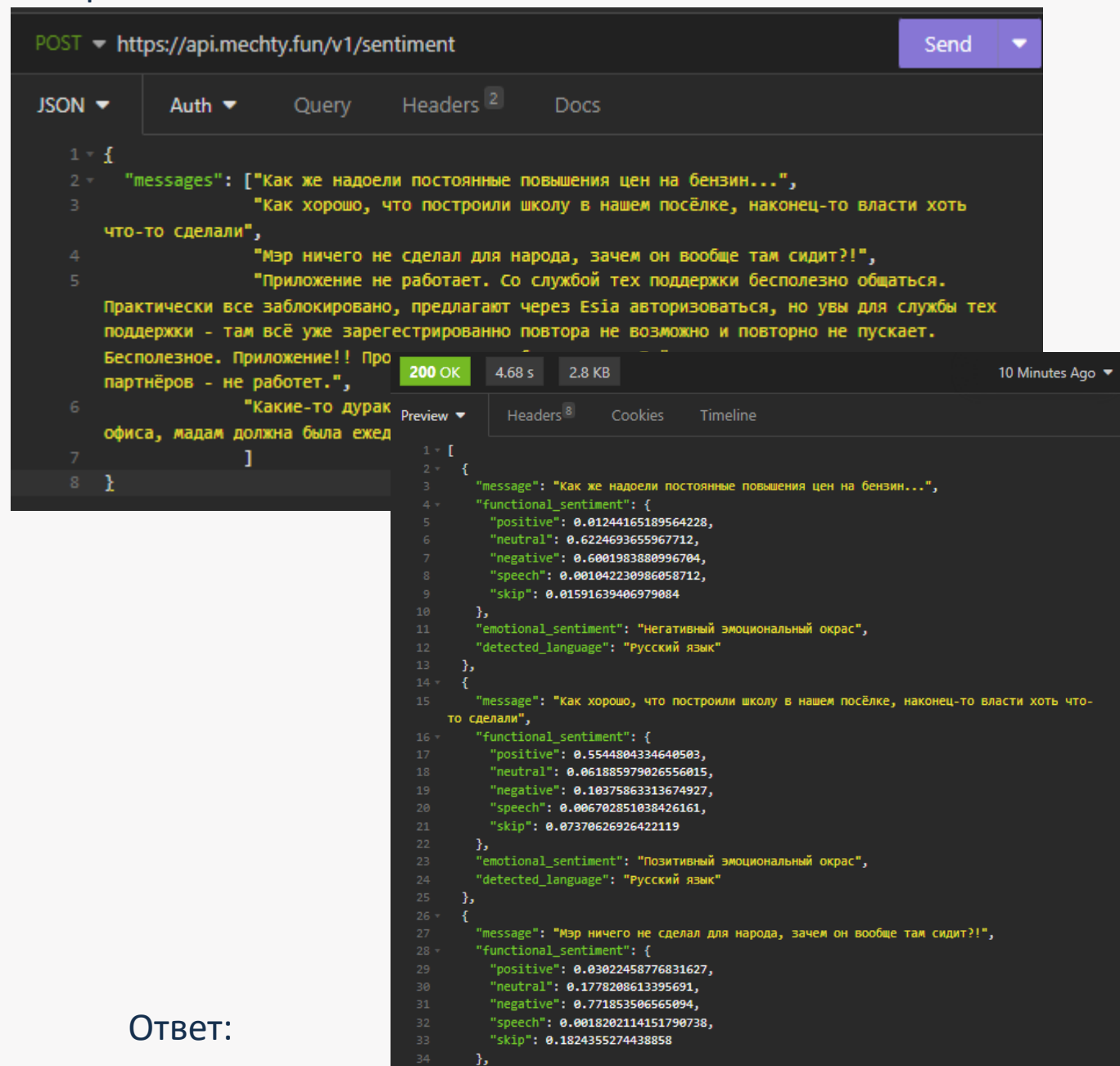
- Dostoevsky
- Natasha
- Rubert

Была использована предобученная модель RuBert для извлечения признаков (features), голова модели была обучена на наших данных.

При использовании модели Dostoevsky удалось добиться значения метрики f1-score 0.809 на более, чем 226 тыс. тестовых сообщениях.

Наш продукт **уже** можно использовать для: определения реакции населения на новости/события;
Реакция пользователей на обновление приложения.

Запрос:



The screenshot displays a REST client interface with a POST request to `https://api.mechty.fun/v1/sentiment`. The request body is a JSON array of three messages. The response is a JSON array of three sentiment analysis results, each containing functional and emotional sentiment scores, the detected language, and a skip flag.

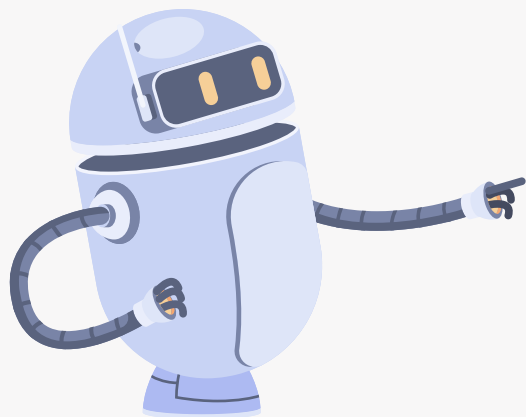
```
POST https://api.mechty.fun/v1/sentiment

{
  "messages": [
    "Как же надоели постоянные повышения цен на бензин...",
    "Как хорошо, что построили школу в нашем посёлке, наконец-то власти хоть что-то сделали",
    "Мэр ничего не сделал для народа, зачем он вообще там сидит?!",
    "Приложение не работает. Со службой тех поддержки бесполезно общаться. Практически все заблокировано, предлагают через Esia авторизоваться, но увы для службы тех поддержки - там всё уже зарегистрированно повтора не возможно и повторно не пускает. Бесполезное. Приложение!! Про партнёров - не работает.",
    "Какие-то дурак офиса, мадам должна была ежед"
  ]
}
```

```
{
  "message": "Как же надоели постоянные повышения цен на бензин...",
  "functional_sentiment": {
    "positive": 0.01244165189564228,
    "neutral": 0.6224693655967712,
    "negative": 0.6001983880996704,
    "speech": 0.001042230986058712,
    "skip": 0.01591639406979084
  },
  "emotional_sentiment": "Негативный эмоциональный окрас",
  "detected_language": "Русский язык"
},
{
  "message": "Как хорошо, что построили школу в нашем посёлке, наконец-то власти хоть что-то сделали",
  "functional_sentiment": {
    "positive": 0.5544804334640503,
    "neutral": 0.061885979026556015,
    "negative": 0.10375863313674927,
    "speech": 0.006702851038426161,
    "skip": 0.07370626926422119
  },
  "emotional_sentiment": "Позитивный эмоциональный окрас",
  "detected_language": "Русский язык"
},
{
  "message": "Мэр ничего не сделал для народа, зачем он вообще там сидит?!",
  "functional_sentiment": {
    "positive": 0.03022458776831627,
    "neutral": 0.1778208613395691,
    "negative": 0.771853506565094,
    "speech": 0.0018202114151790738,
    "skip": 0.1824355274438858
  },
  "emotional_sentiment": "Негативный эмоциональный окрас",
  "detected_language": "Русский язык"
}
```

Ответ:

Наше решение telegram-bot



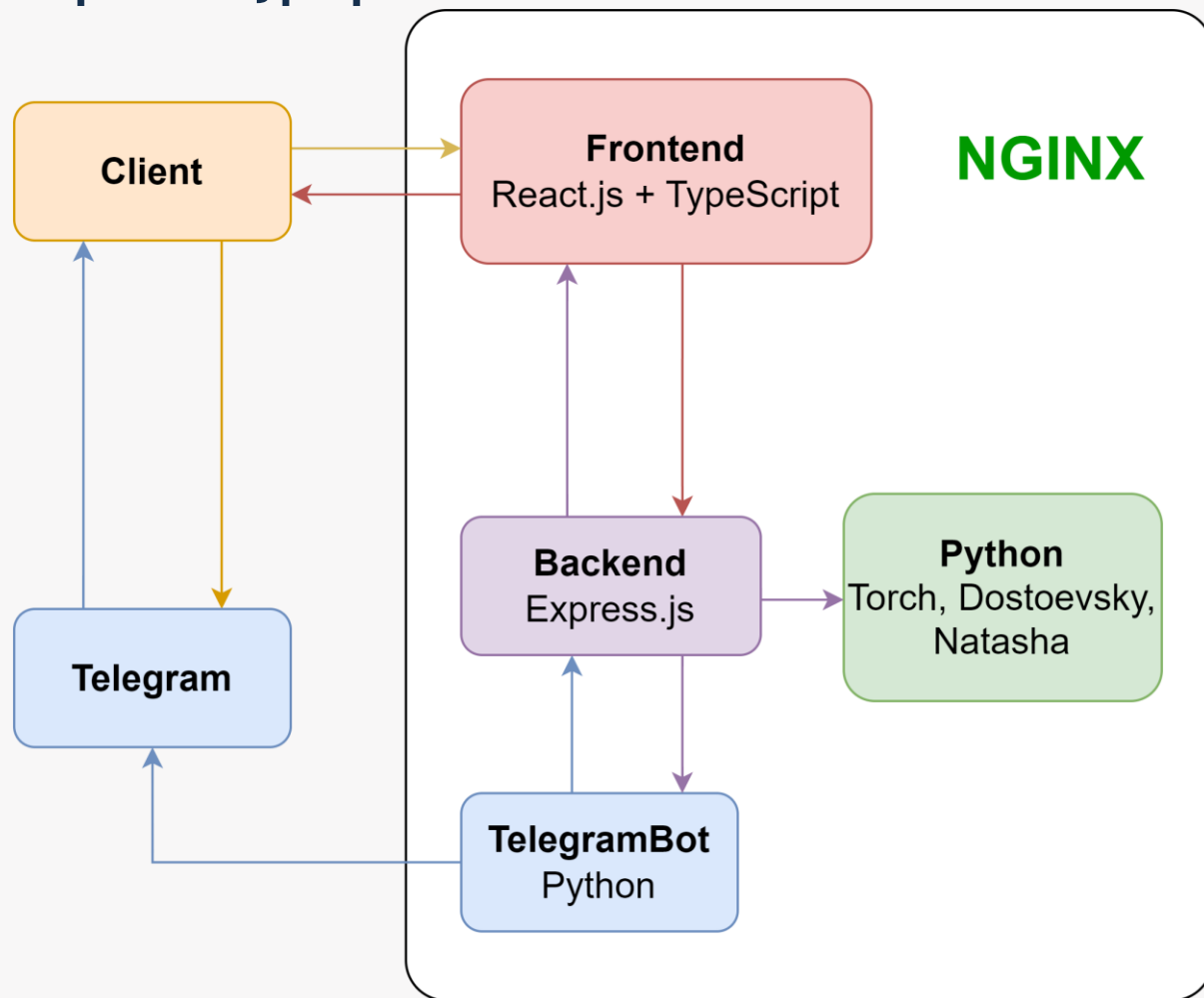
Наше решение telegram-bot

@mechty_ostrova_bot
(https://t.me/mechty_ostrova_bot)



Наше решение web-сервис

Архитектура решения:



<https://mechty.fun/>

Результаты работы

<https://github.com/vp-pro/hackathon.sakh.12.2023-telegram>

<https://github.com/vp-pro/hackathon.sakh.12.2023-frontend>

<https://github.com/vp-pro/hackathon.sakh.12.2023-sentiment>

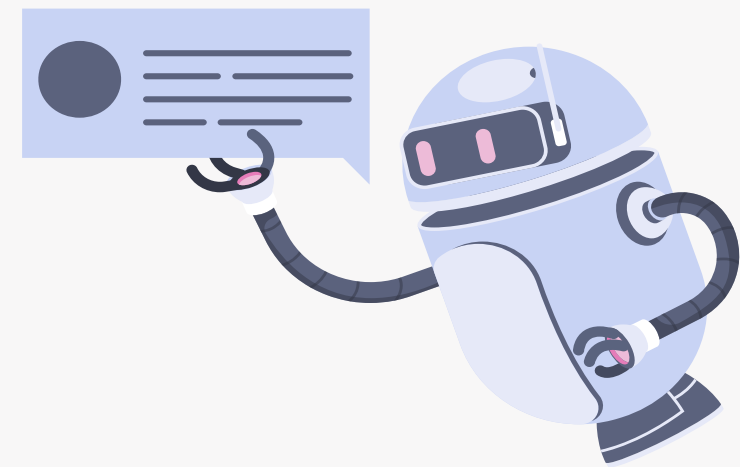
<https://github.com/vp-pro/hackathon.sakh.12.2023-backend>

[Документация проекта](#)

Идеи, которые можно будет реализовать

1. Определение тональности и темы новостей;
2. Парсинг новостей и комментариев из пабликов в социальных сетях;
3. Предоставление отчётности за выбранный период;
4. Определение людей, мест, организаций, о которых идёт речь в тексте.

Анализ тональности русскоязычного текста



Маринич Артём



Data scientist
marinich_art@mail.ru

Панфёров Владислав



Fullstack-разработчик
panferov.public@gmail.com

Маркин Антон



Разработчик ботов
antonimarki2006@gmail.com