

КОМАНДА **ХАХАТОНЩИКИ**
ПРЕДСТАВЛЯЕТ ПРОЕКТ
«КОТИКИ ЛЮБЯТ БЕРТА»



НАША КОМАНДА



Бриг
Data Scientist



Маша
Стажер



Я
Капитан команды
Data Scientist

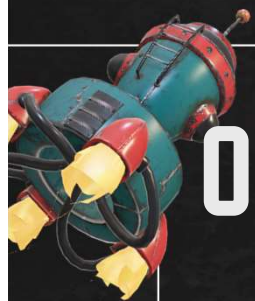


Саша
Data Scientist



Ваня
DL-engineer

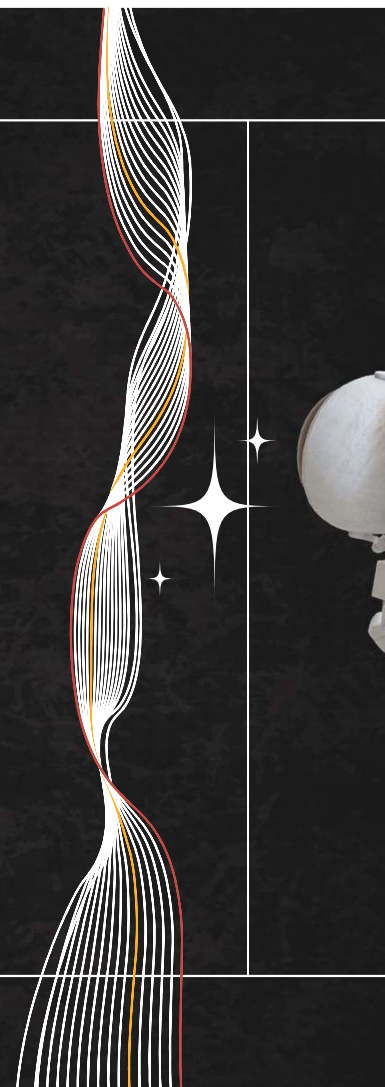
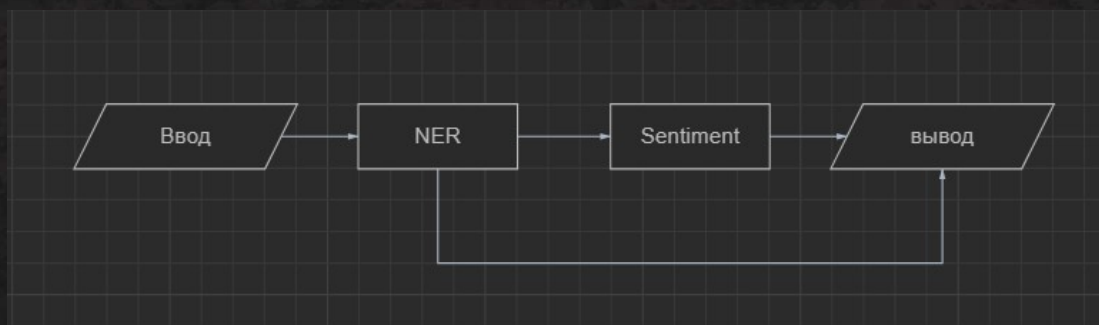




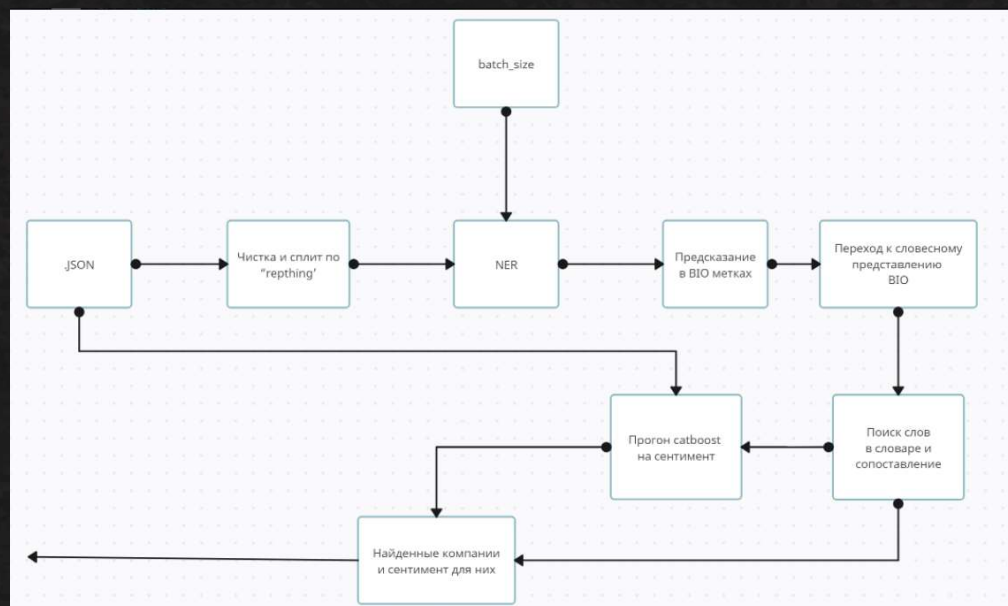
ОСНОВЫ ПОДХОДА

- × Очистка данных
- × RuBert для **NER**
- × CatBoost для **Named Entity Oriented Sentiment analysis**

× × ×
× × ×
× × ×
× × ×
× × ×
× × ×



ОСНОВЫ ПОДХОДА



XXXXX
XXXXX
XXXXX
XXXXX
XXXXX
XXXXX

НАШИ БОЛИ



- × У половины команды **выключили свет** на целый день
- × **Дисбаланс** классов при оценке сентимента
- × Неоднозначная **разметка** текста

xxx
xxx
xxx
xxx
xxx
xxx
xxx

```
data['SentimentScore'].value_counts()
```

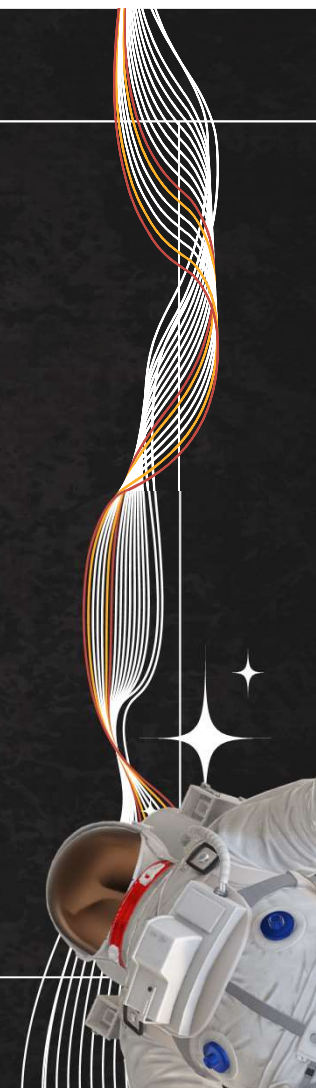
```
SentimentScore
4      3854
3      3511
2       944
5       753
0       163
1         64
Name: count, dtype: int64
```



НАШИ РЕЗУЛЬТАТЫ

- × Accuracy : **0.65**
- × F1 - score : **0.978**
- × Скорость работы на инференсе :
 - × CatBoost : **0.1 СЕК./1000 СООБЩЕНИЙ**
 - × Bert :
 - × Umozo :

xxx
xxx
xxx
xxx
xxx
xxx
xxx



НАШИ ФИШКИ

- × Интерпретация эмодзи, как текст
- × Саммаризатор сообщения с независимой оценкой от ChatGPT
- × Две текстовые фишки: одна это текст, другая это список синонимов компании
- × Подготовлен код для преобразования модели в ONNX формат, квантизации весов и инференса модели через ONNX Runtime

xxx
xxx
xxx
xxx
xxx
xxx





СПАСИБО ЗА ВНИМАНИЕ

Будем рады **ответить** на ваши вопросы

x x x
x x x
x x x
x x x
x x x
x x x
x x x

x Попов Матвей tg: [@MtPriest](#)
x Соколов Владислав tg: [@Vlad.SokolOFF](#)
x Ковалев Александр tg: [@ReallyBadDoc](#)
x Карлов Иван tg: [@hentaiengoer](#)
x Андриюшина Мария tg: [@Otter.Small](#)



CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#).



НЕОДНОЗНАЧНАЯ РАЗМЕТКА

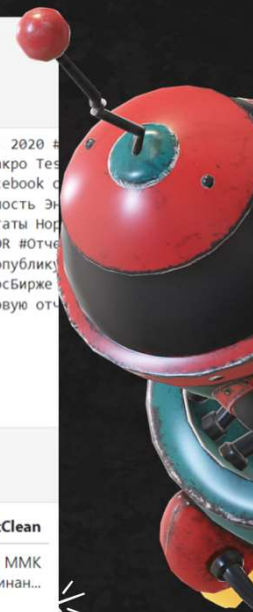
```
5]: print(jj['text'])
l = []
for i in jj['name']:
    l.append(jj['text'][i[0]:i[1]])
print('\nCompanies - ', np.array(l))
```

Фокус недели #ФН Сегодня ММК публикует финансовую отчётность за 1 кв. 2020 #MAGN #Отчетность Магнит представит операционные результаты за 1 кв. 2020 #GNT #Результаты Новатэк публикует финансовую отчётность за 1 кв. 2020 #NVTK #Отчетность ФРС США представит решение по процентной ставке #ФРС #Макро Тесла опубликует финансовую отчётность за 1 кв. 2020 #TSLA #Отчетность Microsoft опубликует финансовую отчётность за 1 кв. 2020 #MSFT #Отчетность Facebook опубликует финансовую отчётность за 1 кв. 2020 #FB #Отчетность 30 апреля Детский мир опубликует финансовую отчётность по МСФО за 2019 #DSKY #Отчетность Энел Россия опубликует финансовую отчётность по РСБУ за 1 кв. 2020 #ENRU #Отчетность Evraz представит операционные результаты за 1 кв. 2020 #EVR #Результаты Норникель опубликует производственные результаты за 1 кв. 2020 #GMKN #Результаты РусГидро представит финансовую отчётность по РСБУ за 1 кв. 2020 #HYDR #Отчетность СД НЛМК рассмотрит дивиденды за 1 кв. 2020 #NLMK #Дивиденды Apple опубликует финансовую отчётность за 1 кв. 2020 #AAPL #Отчетность Amazon опубликует финансовую отчётность за 1 кв. 2020 #AMZN #Отчетность Gilead опубликует финансовую отчётность за 1 кв. 2020 #GILD #Отчетность 1 мая Торги на МосБирже не проводятся из-за праздников #МОЕХ #График Еххон Mobil опубликует финансовую отчетность за 1 кв. 2020 #XOM #Отчетность Chevron публикует финансовую отчётность за 1 кв. 2020 #CVX #Отчетность 4 мая Акции ЦМТ последний день торгуются с дивидендом 0,56 руб. на акцию #WTCM #Дивиденды

Companies - ['GMKN' 'Норникель' 'Детский мир' 'DSKY' 'MGNT' 'MGNT' 'Магнит' 'MAGN' 'ММК' 'МосБирже' 'МОЕХ' 'Новатэк' 'NVTK' 'NLMK' 'НЛМК' 'РусГидро' 'HYDR' 'WTC' 'ЦМТ' 'Энел Россия' 'ENRU' 'вк']

```
9]: names_texts = pd.read_pickle('mentions_texts.pickle')
names_texts = clear_text(names_texts)
names_texts.loc[names_texts['MessageTextClean']==jj['text']]
```

	ChannelID	messageid	issuerid	MessageID	DateAdded	DatePosted	MessageText	IsForward	MessageTextClean
0	1197210433	5408	90	5408	2021-02-06 01:42:42	2020-04-29 07:29:01	?? Фокус недели #ФН Сегодня ????? ММК опублик...	False	Фокус недели #ФН Сегодня ММК опубликует финан...



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**.