

Кейс:

# Центральный Банк Российской Федерации

Анализатор текстовых пресс-релизов

Стек технологий:

Python

PyTorch

TensorFlow

Pandas

NumPy

Matplotlib

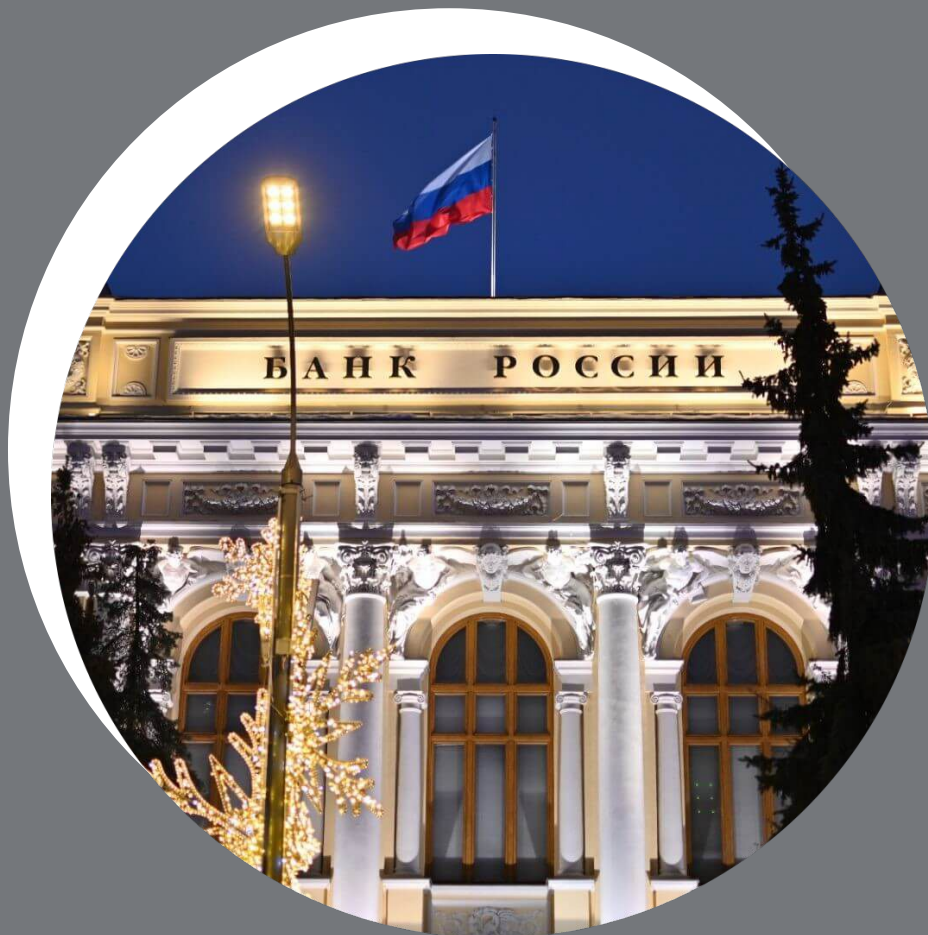
BERT

Natasha

Sklearn



@domnenko\_a\_n  
0001@list.ru



## РЕШЕНИЕ

# BERT

### Fine tuning

точная настройка  
предобученной модели,  
дообучение на своих  
данных



# +

Killer

## Features

### Text + NER

count	average	unique	name	data	loc
-------	---------	--------	------	------	-----

Concatenate



F1 Score + 2-3%

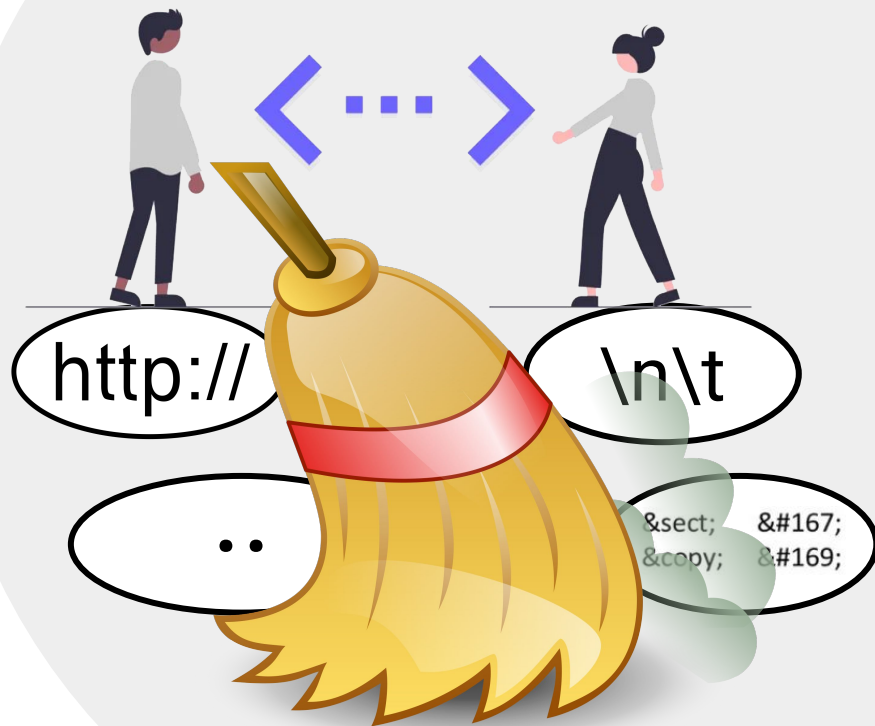
# F1 Score

## 0,7067 / 0.4861

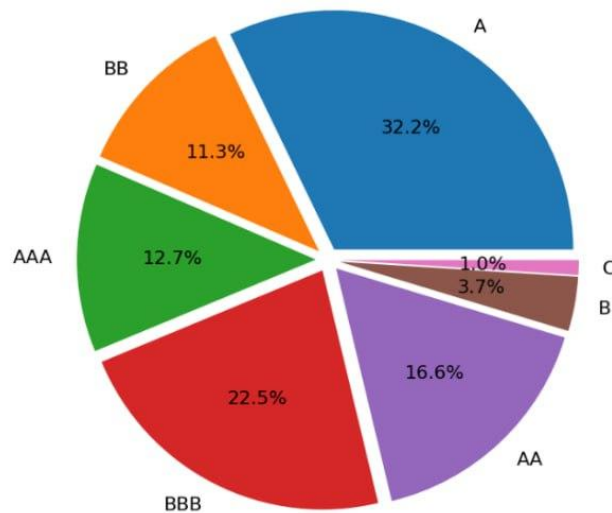
## АНАЛИЗ И ОБРАБОТКА ДАННЫХ

### Удаляем:

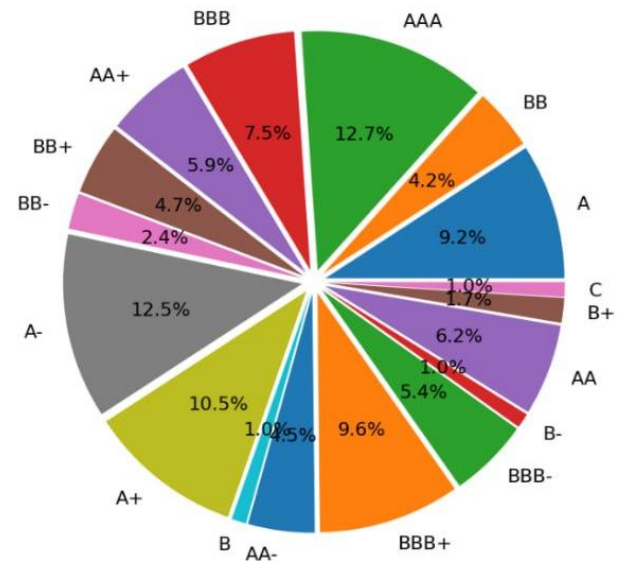
- управляющие символы
- <теги>
- ссылки `http://`
- спецсимволы
- двойные пробелы
- NER
  - ('АО «Эксперт РА', 10416),
  - ('АКРА', 3575),
  - ('Компания', 2997),
  - ('Группа', 1807)



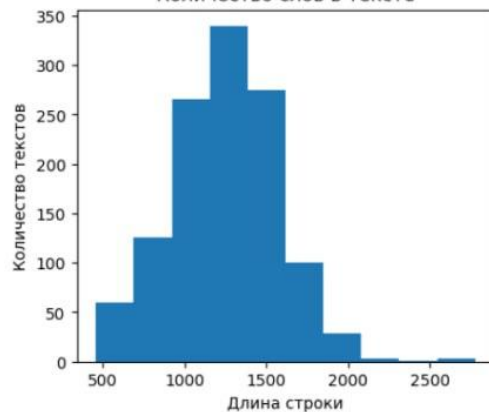
Распределение категорий



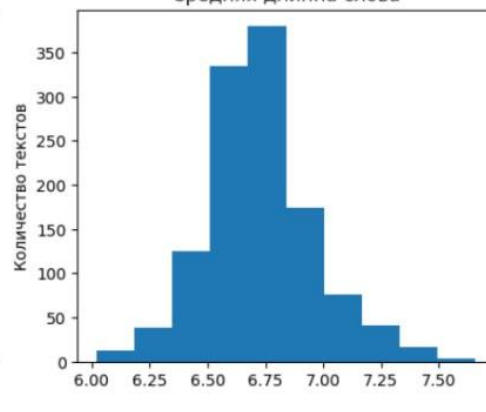
Распределение рейтингов



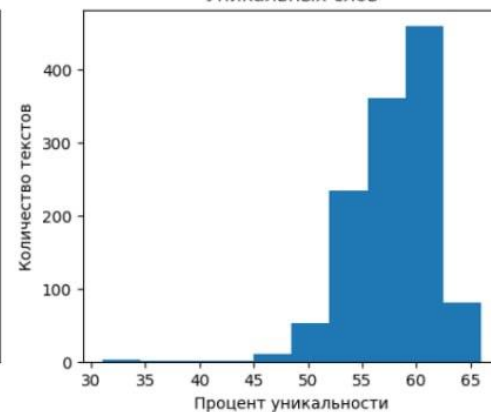
Количество слов в тексте



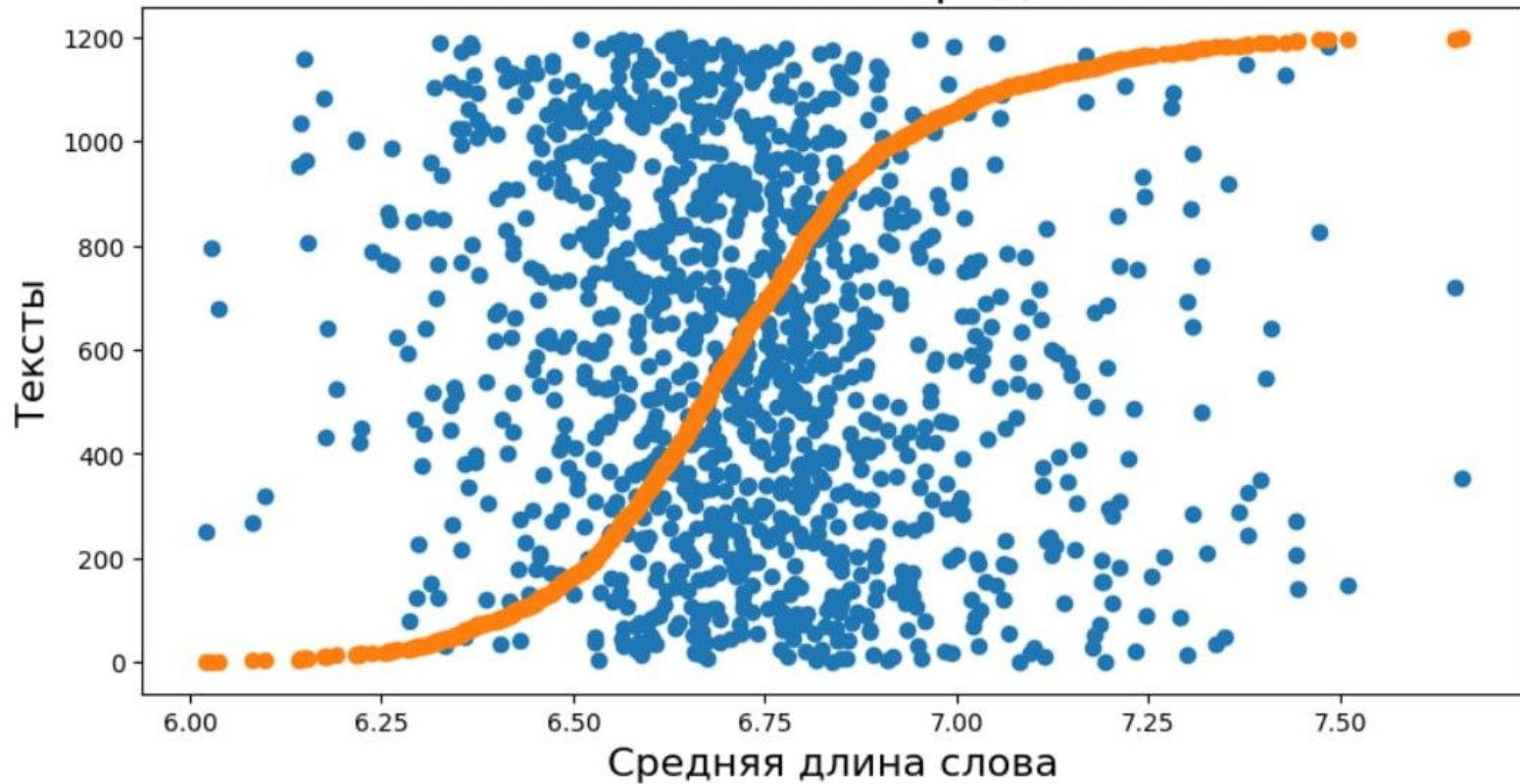
Средняя длина слова



Уникальных слов



## Количество слов в предложении



## МАСШТАБИРУЕМОСТЬ

### Работа с NER (Deerpavlov) :

- увеличение количества сущностей,
- разбиение сущностей на классы
- интерпретируемость фич
- интерпретируемость признаков с помощью шейп
- первые слова по распределению ципфа
- для сущностей находим зависимые слова в предложениях



---

### Работа с текстом

- подсчет количества иностранных слов и др.



---

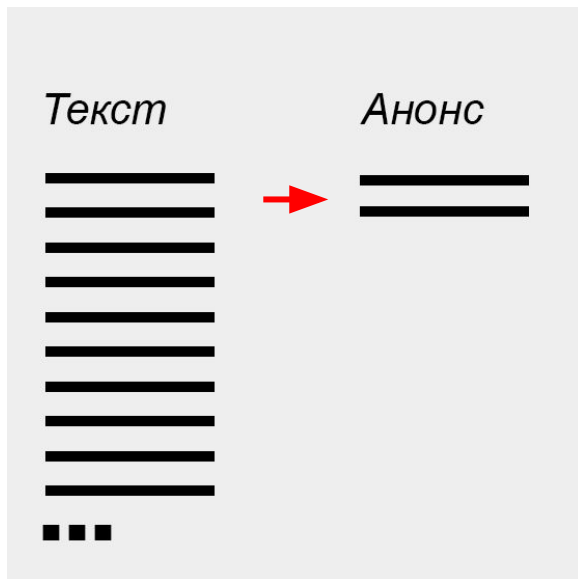
### Работа с внешними факторами:

- оценка тональности комментариев
- определение настроения инвесторов
- предсказание тренда роста рынка на основе переписки инвесторов в общедоступных чатах
- предсказательная аналитика

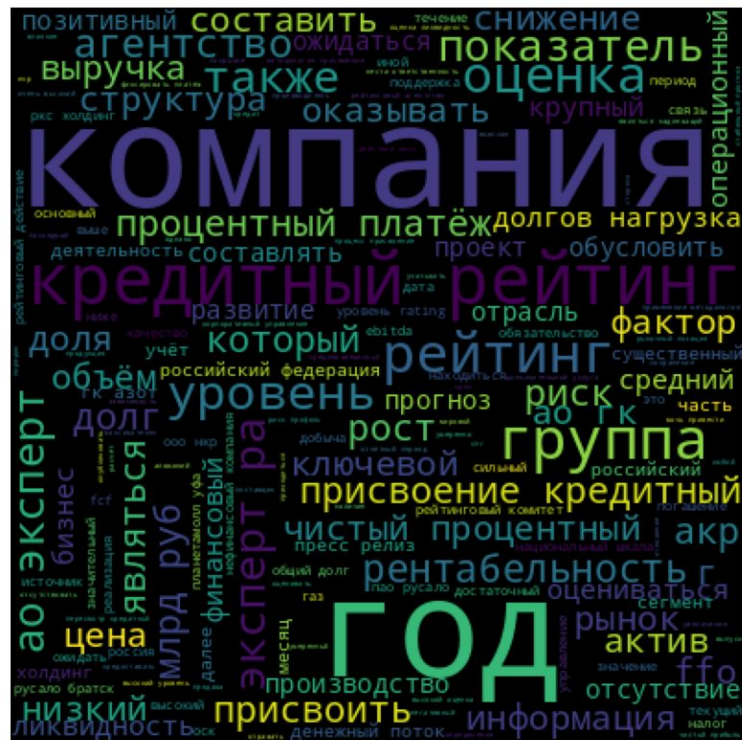




## ОЦЕНКА ВЛИЯНИЯ ТЕКСТОВЫХ ФАКТОРОВ НА РЕЙТИНГ



## Аннотация текста (суммаризация)



## Ключевые слова

## ОЦЕНКА ВЛИЯНИЯ ТЕКСТОВЫХ ФАКТОРОВ НА РЕЙТИНГ

### Именованные сущности

ООО ТК Нафтатранс Плюс | Российская Федерация | Новосибирск |  
Агентство АКРА | ПАО | ВымпелКом |  
ООО РКС Холдинг | Агентство | Петрозаводск | Благовещенск | Фонд развитие территории |  
АО ГК ДИНАМИКА |  
АО ГК Азот | Москва | Китай | Тайвань |  
АО ГК Азот | Россия |  
АО МаксимаТелеком |  
РУСАЛ | Публичное акционерное общество | РУСАЛ | Братский алюминий завод | ОСК | Холдинг |  
ПАО НОВАТЭК | Россия |  
ООО Планетамолл Уфа |



## Веб интерфейс

Логин (admin)

admin

Пароль

••••••••

Вход

Команда Союз

Загрузка данных

НКР присвоило ООО "АСПЭК-Домстрой" кредитный рейтинг «rating» со стабильным прогнозом  
10 марта 2023 года  
Структура рейтинга  
Структура БОСК  
БОСК1

Отправить

Информация

Рейтинг: **BVB+**

Команда Союз

Загрузка данных

НКР присвоило ООО "АСПЭК-Домстрой" кредитный рейтинг «rating» со стабильным прогнозом  
10 марта 2023 года  
Структура рейтинга  
Структура БОСК  
БОСК1

Отправить

Информация

Рейтинг: **BVB+**

Команда СОЮЗ

## НАША КОМАНДА



**Хуторной Борис**

Data scientist  
front-end



[@boris\\_khutornoi](https://t.me/boris_khutornoi)



**Чесников Леонид**

Data scientist  
AI-engineer



[@FatherKomm](https://t.me/FatherKomm)



**Ярулин Дамир**

Data scientist  
front-end



[@fra1ton27](https://t.me/fra1ton27)



**Василевский Иван**

Data scientist



[@Mrghost0](https://t.me/Mrghost0)



**Домненко Алексей**

Data scientist  
Full-stack



[@domnenko\\_a\\_n](https://t.me/domnenko_a_n)