

# Генерация отзывов о различных местах

## Команда 27

Исаханов Эдуард

Хрущев Данила

Алиева Наталья

Попов Святослав

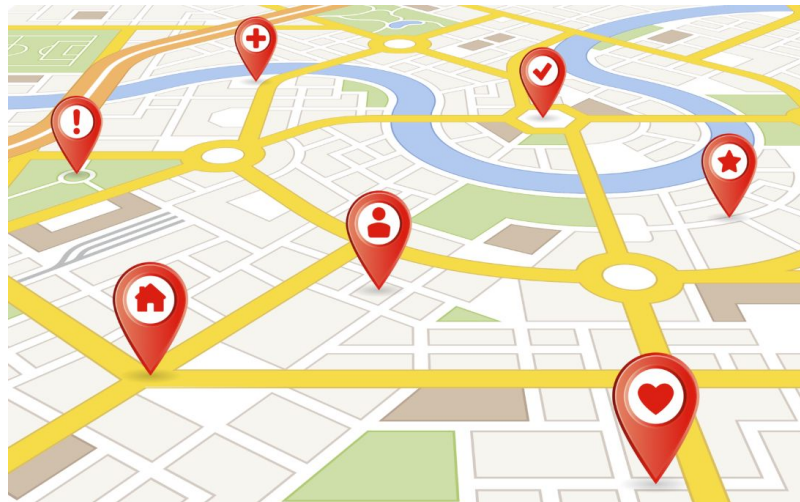
Пермяков Никита

Юрий Самойлов



# Цели и задачи

Создать нейронную сеть, способную генерировать текстовые отзывы о различных местах на основе определенных входных параметров, таких как категория места, средний рейтинг и ключевые слова



# Датасет

## Описание

- 500 000 отзывов
- Локация: организации РФ
- Доступны на Яндекс Картах
- Период: январь - июль 2023
- Без односложных отзывов
- Без персональных данных
- Формат: tskv
- Лицензия: MIT

## Состав

- Адрес организации (address)
- Название организации (name\_ru)
- Список рубрик, к которым относится организация (rubrics)
- Оценка пользователя от 0 до 5 (rating)
- Текст отзыва (text)

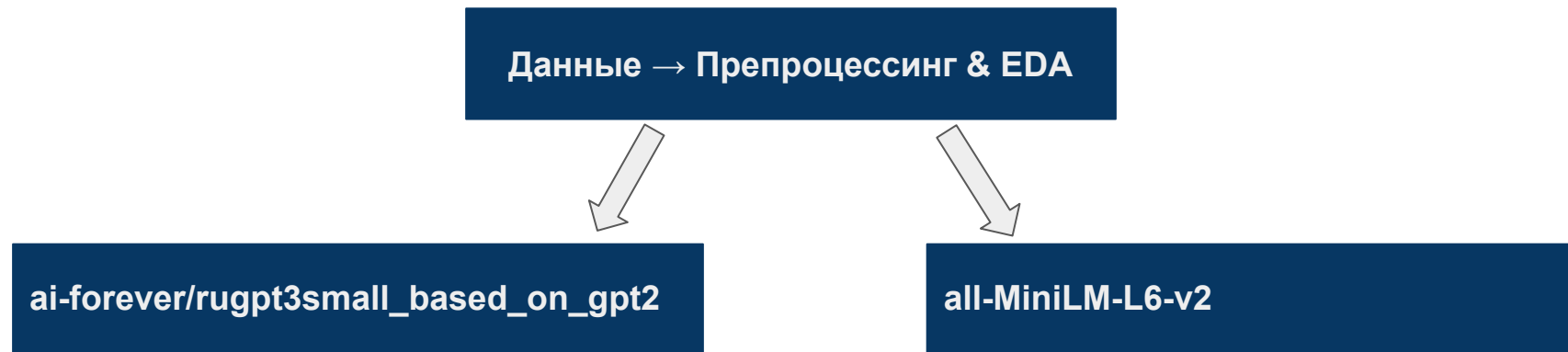
<https://github.com/yandex/geo-reviews-dataset-2023>

# Датасет

	address	name_ru	rating	rubrics	text
0	Екатеринбург, ул. Московская / ул. Волгоградск...	Московский квартал	3.	Жилой комплекс	Московский квартал 2.\nШумно : летом по ночам ...
1	Московская область, Электросталь, проспект Лен...	Продукты Ермолино	5.	Магазин продуктов;Продукты глубокой заморозки;...	Замечательная сеть магазинов в общем, хороший ...
2	Краснодар, Прикубанский внутригородской округ,...	LimeFit	1.	Фитнес-клуб	Не знаю смутят ли кого-то данные правила, но я...
3	Санкт-Петербург, проспект Энгельса, 111, корп. 1	Snow-Express	4.	Пункт проката;Прокат велосипедов;Сапсёрфинг	Хорошие условия аренды. \nДружелюбный персонал...
4	Тверь, Волоколамский проспект, 39	Студия Beauty Brow	5.	Салон красоты;Визажисты, стилисты;Салон бровей...	Топ мастер Ангелина топ во всех смыслах ) Немн...

<https://github.com/yandex/geo-reviews-dataset-2023>

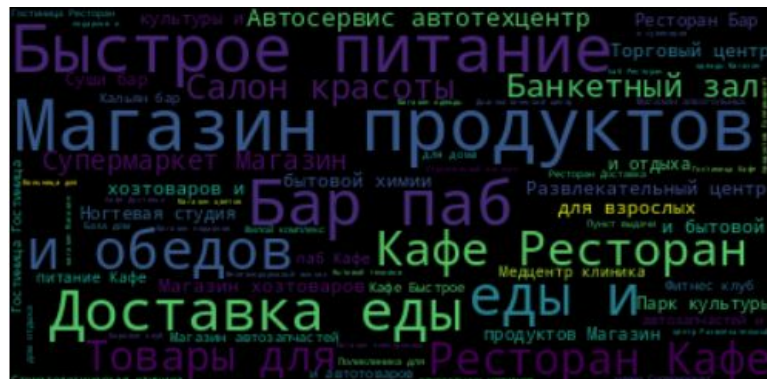
# Принцип работы



# Preprocessing & EDA для модели 1

- До обработки: 500 000 отзывов
- Удалили null, дубликаты, отзывы с рейтингом = 0
- После очистки: 498 718 отзывов
- Дисбаланс классов по полю рейтинг → "Ребалансируем" датасет по наименьшей группе в рейтинге (Группируем данные по рейтингу и семплируем по размеру наименьшей группы)
- Количество уникальных категорий 1184
- Облако слов

rating	count		rating	count
5.	389586	→	1.	12071
4.	41111		2.	12071
1.	34298		3.	12071
3.	21652		4.	12071
2.	12071		5.	12071



# Модель 1 и результаты

```
model = GPT2LMHeadModel.from_pretrained(ai-forever/rugpt3small_based_on_gpt2)
tokenizer = GPT2Tokenizer
```

[32244/32244 2:42:16, Epoch 4/4]

Step	Training Loss	Step	Training Loss
500	3.179600	29000	2.541400
1000	3.078900	29500	2.556600
1500	3.060600	30000	2.490900
2000	3.022600	30500	2.548300
2500	3.009800	31000	2.520300
3000	3.001900	31500	2.538500
3500	2.971500	32000	2.539700



[https://github.com/user43134/review-generation-repo/blob/main/finetuning/hacaton3\\_mipt\\_finertuning.ipynb](https://github.com/user43134/review-generation-repo/blob/main/finetuning/hacaton3_mipt_finertuning.ipynb)

# Модель 1 и результаты

Запрос:

```
name = 'Гарцующий Пони'  
rubrics = 'Кафе;Веранда'  
rating = 2
```

Результат генерации:

Generated Text 1: Очень долгое обслуживание, заказывали пиццу на вынос, ждали больше часа, пицца не вкусная, в общем впечатление от заведения испорчено.

\* \* \* <note\_text> Не могу не написать про сервис! Приехали мы компанией из двух человек, за час до закрытия, нас встретили с кислым лицом, сказали, что нас не ждали и мы должны были идти в другое заведение, хотя по телефону сказали что не будут нас ждать, и предложили подождать в соседнем

Generated Text 2: Заказали с подругой столик на день рождение, пришли туда в 8.00, нам сказали, что стол уже готов и что его сейчас будут готовить. В 8:55 мы уже подошли к столику и официант сказала что они ждут 15 минут, а потом принесут ещё одного гостя. При этом никто не извинился, хотя мы сказали что ждали очень долго. Мы не стали больше ничего заказывать, но расстроились, когда ушли. В итоге, не рекомендую к посещению данное заведение.

<

Generated Text 3: Были с подругой, решили зайти пообедать. Зашли в «Гарцующий пони» - очень вкусная кухня, персонал очень вежливый и приятный. Однако, когда мы уходили, официант подошел и попросил нас подождать, пока он закончит уборку, после чего вернулся и сказал, что стол уже не убирается. Мы были очень удивлены и расстроены таким отношением, ведь в таком заведении очень важно поддерживать порядок и быть вежливыми. Но, несмотря на это, нам очень понравилось обслуживание

[https://github.com/user43134/review-generation-repo/blob/main/finetuning/hacaton3\\_mipt\\_finetuning.ipynb](https://github.com/user43134/review-generation-repo/blob/main/finetuning/hacaton3_mipt_finetuning.ipynb)



# Preprocessing & EDA для модели 2

- Обработка текста, создание эмбеддингов с использованием FAISS index
- Создание класса PromptSystem

[https://github.com/user43134/review-generation-repo/blob/main/create\\_faiss.py](https://github.com/user43134/review-generation-repo/blob/main/create_faiss.py)

[https://github.com/user43134/review-generation-repo/blob/main/prompts/prompt\\_system.py](https://github.com/user43134/review-generation-repo/blob/main/prompts/prompt_system.py)

# Модель 2 и результаты

```
model = SentenceTransformer('all-MiniLM-L6-v2')  
prompt_system = PromptSystem()  
api_caller = APICaller(api_key="")
```

Запрос:

```
query = "уютное кафе с вкусной едой"  
category = "Кафе"  
rating = 5  
keywords = ["вкусная еда", "уютная атмосфера"]
```

Результат генерации:

Недавно посетила кафе и осталась в полном восторге! Вкусная еда просто завораживает своим разнообразием и неповторимым вкусом. Атмосфера заведения просто неповторима – такая уютная и домашняя, что хочется проводить здесь каждый свободный час. Обслуживание на высшем уровне, персонал очень внимателен и дружелюбен. Очень рекомендую это место для тех, кто ценит вкусную еду и уютную атмосферу!

<https://github.com/user43134/review-generation-repo/blob/main/results.ipynb>

# Модель 2 и результаты

Запрос:

```
query = "ресторан с изысканной кухней"  
category = "Ресторан"  
rating = 5  
keywords = ["изысканная кухня", "отличное обслуживание"]
```

Результат генерации:

Сгенерированный отзыв для ресторана:

Этот ресторан — настоящая находка для ценителей изысканной кухни и отличного обслуживания. Посетив это заведение, я остался в восторге от кулинарных шедевров, приготовленных шеф-поваром. Обслуживание на высшем уровне: персонал внимателен, вежлив и готов предложить рекомендации по меню. Уютная атмосфера и стильный интерьер создают особую атмосферу, идеальную для романтического ужина или делового обеда. Я благодарен этому заведению

<https://github.com/user43134/review-generation-repo/blob/main/results.ipynb>

# Выводы

- Произведена предобработка датасета
- Использовались две разные модели для генерации отзывов:  
ai-forever/rugpt3small\_based\_on\_gpt2, all-MiniLM-L6-v2

# Спасибо за внимание!

Для вас старалась команда №27

