Данный интересный проект содержит:

gpt\_restaurants.ipynb -– Юпитер Ноутбук, который загружает датасет от Яндекса с отзывами реальных заведений(id комментаторов обезличены). Более подробные комментарии присутствуют в markdown Затем:

1. Отсеивает только отзывы о ресторанах, очищает их, токенизирует, делает паддинг. Затем данные делятся на обучающие и тестовые.
2. Загрузка предобученной модели Bert, добавление слоев(GlobalAveragePooling1D -  производим выходной тензор размерности (768,). Это позволяет получить одно векторное представление для всей последовательности; Dropout(0.3) - случайным образом обнуляем 30% выходных значений. Это помогает предотвратить переобучение модели; Dense - выходной  слоей с одним нейроном, функция активации sigmoid. Выдает вероятность между 0 и 1)
3. Настраивание модели, добавление колбэков для наибольшей эффективности
4. Ячейка статистики
5. Инициализация БД

api.py – Создание приложения flask, которое принимает на локалхосте POST запрос на предсказание тональности отзыва(в запросе передаем тело отзыва, в ответ получаем good или bad).

Пример курл запроса на сервер - curl -X POST http://127.0.0.1:5000/predict -H "Content-Type: application/json" -d "{\"text\": \"хороший ресторан, готовят вкусно, буду рекомендовать друзьям\"}"

Получаемый ответ - {"prediction":"good"}

reviews.db – база данных sqlite. Содержит базу об отправляемых на сервер отзывах и полученных predict ответах. Инициализирована в gpt\_restaurants.ipynb, но заполнение пока не сделал, ибо не до конца устроил результат обучения(мои предположения, что предобученная модель плохо работает с русским языком, завтра переобучу)

Roadmap: работа с OpenAI API, визуализации наиболее частых слов в позитивных и негативных отзывах.