

# анализ тональности для азербайджанского.

Григорьев Севастьян  
4 курс ОТиПЛ

# Постановка цели и задачи

Анализ тональности.

Датасет: [https://huggingface.co/datasets/DGurgurov/azerbaijani\\_sa/viewer/default/test?views%5B%5D=test](https://huggingface.co/datasets/DGurgurov/azerbaijani_sa/viewer/default/test?views%5B%5D=test)

Данные были собраны из различных источников, таких как социальные сети, обзоры.

его набор данных содержит набор данных для анализа настроений из Local Doc (2024)

# Enhancing GloVe Embeddings for Low-resource Languages with Graph Knowledge

Данные были использованы для проекта по:

<https://github.com/pyRis/retrofitting-embeddings-lrls?tab=readme-ov-file>

Этот проект направлен на усовершенствование эмбедингов GloVe для языков с ограниченными ресурсами за счет использования знаний о графах, а также на создание централизованного хранилища с предварительно подготовленными статическими эмбедингами для различных языков.

# Моя работа

Модель: `ai-forever/mGPT-1.3B-azerbaijan`

<https://hf.qhduan.com/ai-forever/mGPT-1.3B-azerbaijan>

2 метода: Метод 1: few-shot генерация

Метод 2: использование перплексии

Работа проводилась на 50 случайных предложениях.

## Метод 1: few-shot генерация

Задавался промпт на азербайджанском подобного рода:

Тональность — это эмоциональное отношение автора высказывания к некоторому объекту, выраженное в тексте. Она может быть положительной или отрицательной. Положительная тональность выражается с помощью метки '1'. Отрицательная тональность выражается с помощью метки '0'.

Ниже приведены примеры классификации текстов по тональности:

Текст: Этот фильм был ужасен.

Тональность: 0 (отрицательная)

Текст: Мне нравится эта книга.

Тональность: 1 (положительная)

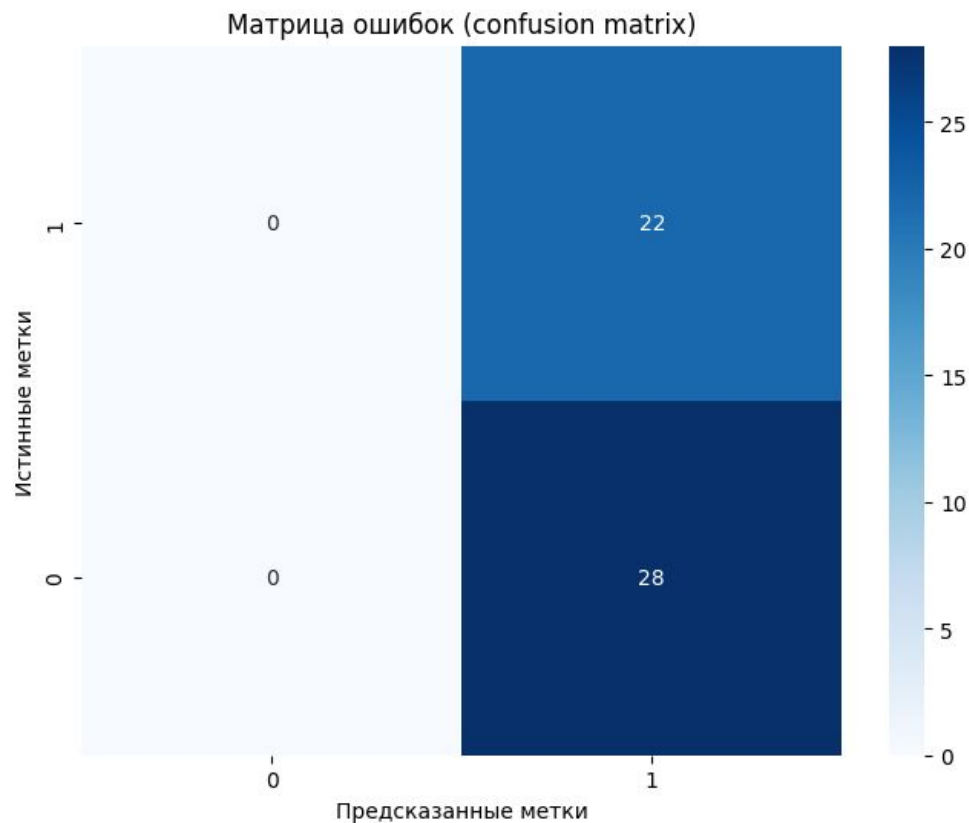
Текст: Я ненавижу это место.

Тональность: 0 (отрицательная)

Текст: {text}

Тональность:

# Результаты ( матрица ошибок)



Предсказывается только  
положительная тональность

# Метрики

**точность** — 0,56;

**Полнота** — 1,00;

**F1** — 0,72.

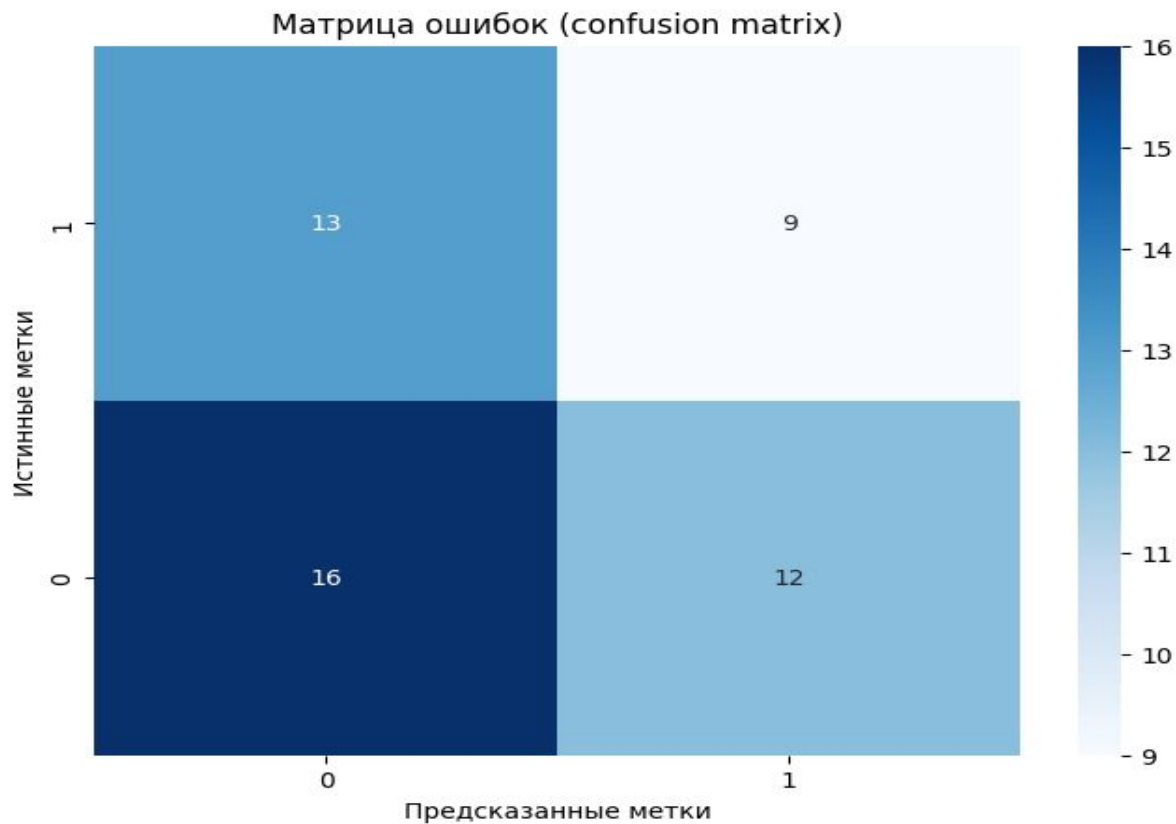
## Метод 2: использование перплексии

В теории информации, перплексия – мера того, насколько хорошо распределение вероятностей предсказывает выборку. Говоря простым языком, перплексия – мера "удивлённости" модели.

На вход модели мы подаём затравку в формате <метка тональности> + <отзыв>. Далее мы считаем перплексию каждого из них. Выбираем наименьшую перплексию из двух и присваиваем метку отзыву.



# Результаты( матрица ошибок)



# Метрики

**точность** —0,57;

**Полнота** —0,43;

**F1** —0,49.

Применение предобученной модели Bert к этим же 50 предложениям.

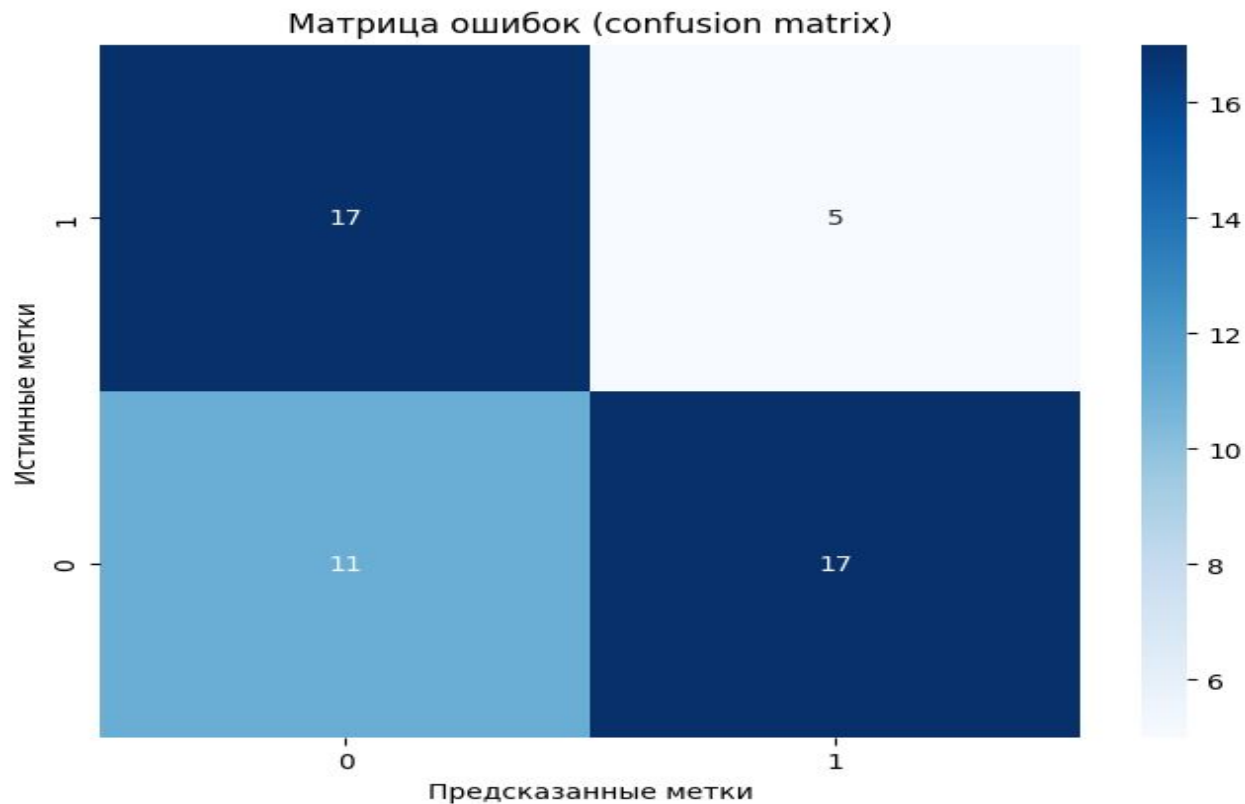
Модель: DGurgurov/xlm-r\_azerbaijani\_sentiment

Ссылка на модель:

[https://huggingface.co/DGurgurov/xlm-r\\_azerbaijani\\_sentiment](https://huggingface.co/DGurgurov/xlm-r_azerbaijani_sentiment)

Fine-tuning Dataset: DGurgurov/azerbaijani\_sa ( датасет, который взял я)

# Результаты ( матрица ошибок)



## Метрики:

**точность** —0,77;

**Полнота** —0,61;

**F1** —0,68.