

Лабораторная работа

Решите пожалуйста следующие задачи.

1. Рассчитайте метрики TF-IDF для любых 10 песен на одном языке, которые вы сами выберите.

Не забудьте, что нужно привести слова к начальной форме, убрать стоп-слова.

2. Цель этого задания - использовать предварительно обученную модель BERT для классификации тональности отзывов на фильмы.

- Скачайте [датасет](#) отзывов на фильмы. Датасет содержит текст отзыва и бинарную метку тональности (положительный/отрицательный).

- Используйте библиотеку Hugging Face для загрузки предварительно обученной модели BERT и токенизатора.

- Подготовьте данные: используйте токенизатор BERT для преобразования текстовых данных в формат, который можно подать на вход модели BERT.

- Создайте классификатор на основе BERT: это может быть модель BERT с одним линейным слоем для классификации на вершине.

- Обучите классификатор на данных обучения и оцените его производительность на данных для тестирования.