

Тестовое задание

Тебе даны данные одного крупного онлайн магазина и перед ним стоит классическая задача продуктового машинга. У магазина много поставщиков одинаковых товаров и все они называют их по разному, задача найти одни и те же товары с разными названиями и объединить их в группы. Смотри пример с описанием данных и с примером оценки бейзлайна.

Для каждого товара нужно выдать список товаров которые модель считает одинаковыми. При этом таких товаров не может быть более 50 для Метрикой для задачи будет среднее построчного F1.

$$F_1 = 2 \frac{precision * recall}{precision + recall} = \frac{2 * TP}{2 * TP + FN + FP}$$

Пример расчета метрики есть в ноутбуке.

Может помочь для решения задачи:

- <https://github.com/scikit-learn-contrib/metric-learn>
- <https://arxiv.org/pdf/1503.03832.pdf>
- <https://github.com/KevinMusgrave/pytorch-metric-learning>
- <https://arxiv.org/pdf/2003.08505.pdf>

Если используешь чужой код, то нужно указать ссылку откуда ты его взял.

Нейронки(если будешь их использовать) лучше писать на PyTorch