

Домашнее задание 4

Глубинное обучение в анализе графовых данных.

1 Задание 1. Проектирование рекомендательной системы (6 баллов)

Для использования вам предлагается изучить датасет <https://cseweb.ucsd.edu/~jmcauley/datasets.html#twitch>

Это датасет пользователей стриминговой платформы Twitch, для каждого юзера даны данные по тем стримам, которые он смотрел, с указанием отметок времени начала просмотра и конца. Вам предлагается разработать концепт рекомендательной системы, которая предсказывает, какие стримы будут интересны определенному юзеру.

Расскажите, какой подход вы собираетесь использовать, какие метрики качества вам помогут оценить работу придуманного решения, как стоит разбивать датасет, какие проблемы могут возникнуть с данными и с обучением. Обоснуйте выбор архитектуры/решения, опишите плюсы и минусы.

Если выбрано не нейросетевое решение, то максимальное количество баллов за задание составляет 4.

Бонус 1 балл. Проведите подробный EDA (анализ датасета)

2 Задание 2. Реализация. (5 баллов)

Реализуйте придуманное вами решение.

В случае если в первом задании решение нейросетевое, вы все равно можете на данном этапе реализовать решение каким-то другим методом, при этом предоставив его описание.