



ФИНАНСОВЫЙ
УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

Рекомендательные системы и коллаборативная фильтрация

доцент, канд. техн. наук Семенов Роман
Владимирович

Балльно-рейтинговая система по дисциплине



ФИНАНСОВЫЙ
УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

№ п/п	Вид отчетности		Баллы
1.	Работа в семестре – всего, в т.ч.:		40
1.1	Работа в первом модуле – всего, в т.ч.		20
1.1.1	Построение простой рекомендательной системы на основе популярности товаров (общий пример Kaggle). ИНДЗ: на датасете по выбору. Оценка рекомендательных систем (1- 2 семинар)	3	3
1.1.2	Реализация User-Item коллаборативной фильтрации на примере набора данных MovieLens. ИНДЗ: на датасете по выбору. Оценка рекомендательных систем (3-4 семинар)	4	4
1.1.3	Реализация SVD, на примере набора данных MovieLens, с использованием библиотеки Surprise. ИНДЗ: реализация гибридной рек. системы SVD + коллаб. фильтрац. Оценка рекомендательных систем (5 – 6 семинар)	4	4
1.1.4	Реализация KNN с коллаб. фильтрацией, на примере набора данных MovieLens, с использованием библиотеки Surprise. ИНДЗ на датасете по выбору. Оценка рекомендательных систем (7 - 8 семинар)	4	4
1.1.5	Посещение лекций и семинаров	0,4	5,0
1.2	Работа во втором модуле– всего, в т.ч.	1	20
1.2.1	Построение рекомендаций на основе неявных данных. Реализация гибридной рек. системы на примере набора данных MovieLens, с использованием библиотеки LightFM. Оценка рекомендательной системы (9 семинар)	2	2
1.2.2	Реализация гибридной рек. системы (общий пример Kaggle), с использованием библиотеки LightFM. Оценка рекомендательной системы (10 семинар)	2	2
1.2.3	Реализация гибридной рек. системы с использованием библиотеки LightFM. ИНДЗ: на датасете по выбору. Оценка рекомендательной системы (11 - 13 семинар)	4	4
1.2.4	Реализация рек. систем на основе GNN (Graph Neural Network), LightGCN. ИНДЗ: на датасете по выбору. Оценка рекомендательной системы (14 - 16 семинар)	4	4
1.2.5	Опрос/дискуссия на семинарах/ тест электронный (17 семинар)	0,5	3,0
1.2.6	Посещение лекций и семинаров	0,4	5,0
2.	Экзамен	1	60
	Итого:		100

Рассмотрим архив BasketProducts.zip с датасетом о наборе товаров различных категорий и проведем исследовательский анализ данных (EDA) по нему.

Набор данных содержит следующие 10 столбцов:

1. индекс
2. product – название продукта
3. category – категория продукта
4. sub_category – подкатегория продукта
5. brand - торговая марка продукта
6. sale_price – цена, по которой товар продается на сайте
7. market_price – рыночная цена продукта
8. type – тип, к которому относится продукт
9. rating – совокупная оценка продукта (от 1 до 5) покупателями
10. description – описание продукта

Затем провести первичный анализ данных – в частности, распределение товаров по категориям и подкатегориям.

Далее (вариативная часть) узнать наиболее популярные бренды, наиболее популярные типы товаров (в смысле часто встречающиеся).

Затем провести оценку по рейтингу всех товаров (сколько товаров с каждой оценкой, построить гистограмму распределения).

Затем установить, существует ли связь между скидкой на товар его рейтингом (как в общем, так и по категориям).

Затем можно построить логику рекомендаций по наиболее популярному продукту в конкретной категории или продукту с наивысшим рейтингом