

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Дисциплина: «Анализ данных»

Домашнее задание на тему:
«Лабораторная работа №17»

Выполнил: Осипов Лев,
студент группы БПИ21 (1).

СОДЕРЖАНИЕ

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ.....	3
Задание 1	3
Задание 2	3
Задание 3	3
СПИСОК ЛИТЕРАТУРЫ	4

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Задание 1

Несмотря на то, что в Netflix пользователей больше, чем продуктов (и поэтому казалось логичнее взять именно user-based), больше смысла имеет брать item-based подход, так как данные о товарах с большим количеством пользователей будет информативнее. Скорее у товаров будет больше информации о пользователях (такой-то товар купили такие-то пользователи), а не у пользователей будет много информации о товарах (такой-то пользователей предпочитает такие-то товары).

Задание 2

Здесь работает та же логика: у товаров будет больше информации о пользователях (такой-то товар купили такие-то пользователи), а не у пользователей будет много информации о товарах (такой-то пользователей предпочитает такие-то товары), поэтому тоже лучше использовать item-based подход.

Задание 3

Если у фильма нет оценки, можно брать какие-либо его атрибуты. Предположим, жанр, актеры, год и страна выпуска. Соответственно, предлагать его пользователям, ставившим высокие оценки фильмам с похожей атрибутикой.

Если нужно расширить диапазон (в случае, если похожесть наблюдается низкая), можно также брать этих пользователей (высоко оценивших похожие, хоть и слабо, фильмы), брать уже их параметры (к примеру, пол, возраст) и предлагать этот новый фильм всем пользователям с такими или похожими параметрами. Однако данный подход может быть совсем неточен. Возможно, его следует использовать лишь в крайних случаях для расширения списка.

СПИСОК ЛИТЕРАТУРЫ

- 1) **Анализ данных (Программная инженерия) –**
[http://wiki.cs.hse.ru/Анализ_данных_\(Программная_инженерия\)](http://wiki.cs.hse.ru/Анализ_данных_(Программная_инженерия))