



РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ НА ПРАКТИКЕ

СЕЛЕЗНЕВ АРТЕМ



 @SeleznevArtem

 /seleznev-artem

 /NameArtem

CVM



Только для вас

Персональные предложения



До 10%
бонусами!

Подробнее →



Календарь
скидок в
Магнит Се...

Подробнее →





Mobile App



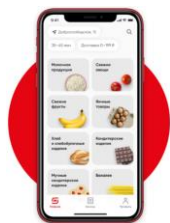
SMS



E-mail



> 60 млн. человек



Mobile App



SMS



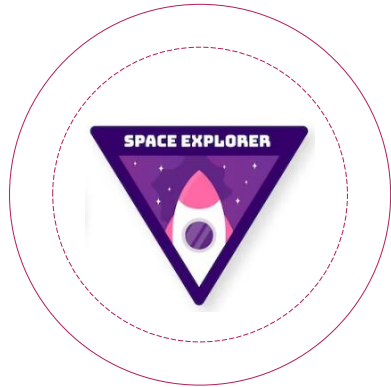
E-mail



> 60 млн. человек

Массовое промо
Рекомендации

ПРОГРАММА БУТКЭМПА

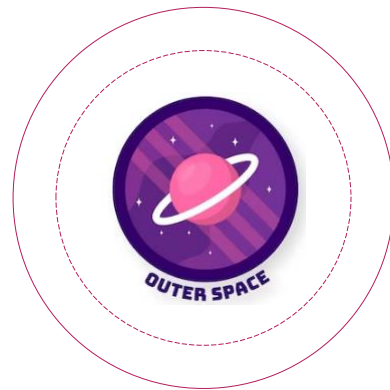


День 1

Основы рекомендаций

Постановка задачи

Базовые алгоритмы

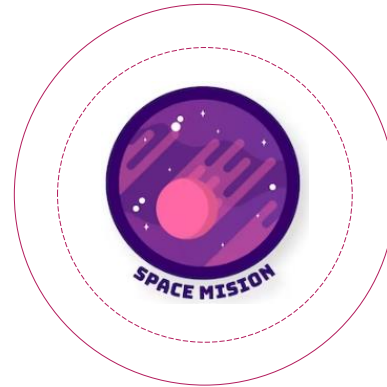


День 2

Бустинги и рекомендации

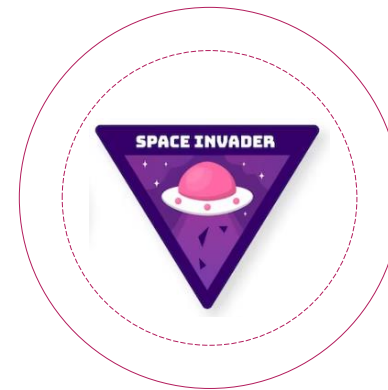
Задачи ранжирования

Валидация и метрики



Запуск соревнования Kaggle

Боремся за метрику



День 3

Проблемы в рекомендациях

Больше факторов



Финальный проект

Создать готовое решение

Проект проходи
«продуктовую» приёмку

ПРОГРАММА БУТКЭМПА



Запуск соревнования
Kaggle

Popularity-Based Jokes Recommendation

Notebook Input Output Logs Comments (4)

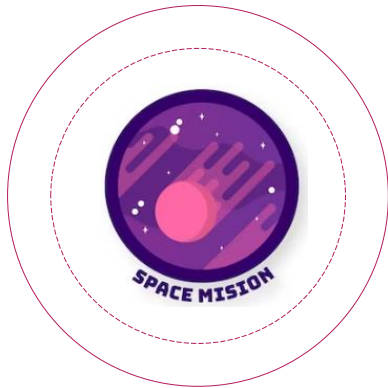
Detail Compact Column

About this file

List of jokes

▲ 1: A man visits th... ≡	▲ I have bad news f... ≡	▲ Well, thank God I ... ≡
142 unique values	[null] 22% "Ah" 1% Other (109) 77%	[null] 30% "No" 1% Other (98) 69%
2: This couple had an excellent relationship going until one day he came home from work to find his ...	"That's an awfully big word for a ten year old."	
3: Q. What's 200 feet long and has 4 teeth? A. The front row at a Willie Nelson concert.		
4: Q. What's the difference between a man and a toilet? A. A toilet doesn't follow you around after ...		

ПРОГРАММА БУТКЭМПА



Запуск соревнования
Kaggle

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$



Лучший результат

ПРОГРАММА БУТКЭМПА



Финальный проект

Создать готовое решение

Проект проходи
«продуктовую» приёмку

$$MAP@k = \frac{1}{|Users|} \sum_{u \in Users} AP@k(u)$$

Проверка по RMQ:

- Соответствие гипотезе
- Финансовая оценка на A/B тесте с BaseLine
- Обработка не популярных предметов (покрытие рекомендациям)
- Средняя частота взаимных совпадений (ARHR)
- Новизна
- Интуитивность (Serendipity)

ПРОГРАММА БУТКЭМПА

 /NameArtem/recommendation_in_practice_tutorial

ПРОГРАММА БУТКЭМПА (МАРКЕРЫ)



Владелец продукта



Junior



Senior



Пользователь

ПРОГРАММА БУТКЭМПА (МАРКЕРЫ)



Владелец продукта



Junior



Пользователь



Senior

Я создаю гипотезы

Я работаю с метриками и очень data-driven



Владелец продукта

ПРОГРАММА БУТКЭМПА (МАРКЕРЫ)



Владелец продукта



Junior



Пользователь



Senior

Меня окружает мир рекомендаций
Является ли это моим настоящим выбором?



Пользователь

ПРОГРАММА БУТКЭМПА (МАРКЕРЫ)



Владелец продукта



Junior



Senior



Пользователь

Я всё могу!



Junior

Я всё могу?



Senior



ДЕНЬ 1

РЕКОМЕНДАЦИИ ВЕЗДЕ



Ваши первые рекомендации



Первые безответственные рекомендации



Что-то было не рекомендацией



РЕКОМЕНДАЦИИ ВЕЗДЕ



**Ваши первые
рекомендации**



**Первые
безответственные
рекомендации**



**Таргетировано,
но ещё не рекомендации**



РЕКОМЕНДАЦИИ ВЕЗДЕ



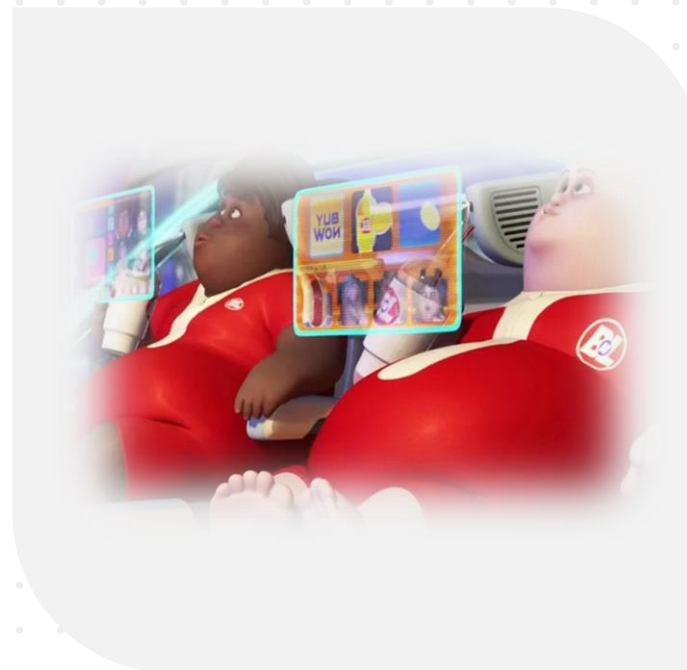
Ваши первые рекомендации



Первые безответственные рекомендации



ВАЛЛ-И плохое будущее



МЫ В ОТВЕТЕ ЗА ТЕХ, КОМУ ПОРЕКОМЕНДОВАЛИ



МЫ В ОТВЕТЕ ЗА ТЕХ, КОМУ ПОРЕКОМЕНДОВАЛИ

Что это означает?
Они будут не самые релевантные подходящие?



Мы должны знать больше о покупателях...



ЗНАНИЯ ДО РЕКОМЕНДАЦИЙ

Жизненный цикл клиента в системе



ЗНАНИЯ ДО РЕКОМЕНДАЦИЙ

Жизненный цикл клиента в системе



Миссии клиента в системе

UID	Жанр / Время			
	Комедии	Ужасы	Спорт	Мультфильмы
1	0	1 / 93	5 / 500	0
2	4 / 374	2 / 180	0	6 / 540
3	9 / 810		0	45 / 1050

UID	Миссия
1	Просмотр футбола
2	Семейный
3	Мультки в обеденный перерыв

Но мне в ленту рекомендаций попадают фильмы
или товары, которые я уже смотрел / купил.



Это называется продвижение.
Не зависимо от того, что мы используем лента
должна «зарабатывать»



Но моя метрика...



ТАКСОНОМИЯ РЕКОМЕНДАЦИЙ

Введите запрос

The Midnight - Crystalline

NewRetroWave

1,26 млн подписчиков

Вы подписаны

35 тыс.

681

Поделиться

Создать клип

3,5 млн просмотров · 5 лет назад

New Retro Wave + The Midnight · Crystalline

The Midnight are back with a superb new single! The track oozes greatness. Listen up and enjoy! :)

Ещё

1 568 комментариев

Упорядочить

Введите комментарий

Закреплено пользователем NewRetroWave

@NewRetroWave

5 лет назад (изменено)

The Midnight are back with a superb new single! These guys are set to take over the scene I tell ya. lol. The track oozes greatness. Listen up and enjoy! :) OUT NOW!
<https://themidnight.bandcamp.com/track/crystalline>

591

Ответить

19 ответов

3:12

NewRetroWave

5:13

The Midnight - Days of Thunder

The Midnight

5:27

The Midnight - Sunset

Retrowave Cinema

4:23

GUNSHIP - Ghost (feat. @powerloveaudio) [Official...

GUNSHIPMUSIC

4:38

The Midnight - Lost Boy

Retrowave Cinema

5:20

The Midnight - Light Years (feat. Nikki Flores)

Neon Nights

4:32

FM-84 - Running In The Night (feat. Ollie Wide) - [OFFICIAL...

NewRetroWave

< oWave

Для вас

Похожий контент

>

1:19:52

Female Vocal Trance | The Voices Of Angels #19

Aurora

597 тыс. просмотров · 2 года...

46:52

T R I P + D R I V E [synthwave - chillwave - retrowave mix] [rick...

admirablemice

29 тыс. просмотров · 3 недели...

4:31

REGRETS - 2023

DRORELIMELECH120V

4 просмотра · 42 минуты назад

Новинка

58:19

Galactic Waves – A Downtempo Chillwave Mix [Chill - Relax -...

Chill Pulse Music

181 тыс. просмотров · 1 месяц...

The Midnight - Days of Thunder/ Gloria

NewRetroWave

ТАКСОНОМИЯ РЕКОМЕНДАЦИЙ

Введите запрос

The Midnight - Crystalline

NewRetroWave 1,26 млн подписчиков

Вы подписаны

35 тыс. 681

Поделиться

Создать клип

3,5 млн просмотров 5 лет назад

New Retro Wave + The Midnight Crystalline

The Midnight are back with a superb new single! The track oozes greatness. Listen up and enjoy! :)

Ещё

1 568 комментариев

Упорядочить

Введите комментарий

Закреплено пользователем NewRetroWave

@NewRetroWave

 5 лет назад (изменено)

The Midnight are back with a superb new single! These guys are set to take over the scene I tell ya. lol. The track oozes greatness. Listen up and enjoy! :) OUT NOW!

<https://themidnight.bandcamp.com/track/crystalline>

591 19 ответов

oWave

Для вас

Похожий контент

Femalé Vocal Trance | The Voices Of Angels #19 Aurora 597 тыс. просмотров · 2 года...
1:19:52

TRIP + DRIVE [synthwave - chillwave - retrowave mix] [rick... admirablemeice 29 тыс. просмотров · 3 недели...
46:52

REGRETS - 2023 DRORELEMECH120V 4 просмотра · 42 минуты назад Новинка
4:31

Galactic Waves - A Downtempo Chillwave Mix [Chill - Relax - ... Chill Pulse Music 181 тыс. просмотров · 1 месяц...
58:19

The Midnight - Days of Thunder/ Gloria NewRetroWave

Специализация рекомендаций

Задача и контекст

NewRetroWave
1.26 млн подписчиков

 Вы подписаны

 35 тыс.

681

Поделиться

 Создать клип

New Retro Wave + The Midnight Crystalline

The Midnight are back with a superb new single! The track oozes greatness. Listen up and enjoy! :D Ещё

1 568 комментариев

≡ Упорядочить

Введите комментарий

Закреплено пользователем NewRetroWave








@NewRetroWave · 5 лет назад (изменено)

The Midnight are back with a superb new single! These guys are set to take over the scene I tell ya. lol. The track oozes greatness. Listen up and enjoy! :)
OUT NOW!

<https://themiidnight.bandcamp.com/track/crystalline>

591 Ответить

▼ 19 ответов

-  **NewRetroWave**
-  **The Midnight - Days of Thunder**
The Midnight
-  **The Midnight - Sunset**
Retrowave Cinema
-  **GUNSHIP - Ghost (feat. @powerloveaudio) [Official...]**
GUNSHIPMUSIC
-  **The Midnight - Lost Boy**
Retrowave Cinema
-  **The Midnight - Light Years (feat. Nikki Flores)**
Neon Nights
-  **FM-84 - Running In The Night (feat. Ollie Wilde) - [OFFICIAL...]**
NewRetroWave

< **Wave** **Для вас** Похожий контент >

Female Vocal Trance | The
Voices Of Angels #19
Aurora ✓
597 тыс. просмотров · 2 года...

TRIP+DRIVE [synthwave - chillwave - retrowave mix] [rick...
admirablemice
29 тыс. просмотров • 3 недели...

REGRETS - 2023
DRORELIMELECH120V
4 просмотра · 42 минуты назад
Новинка

Galactic Waves – A Downtempo
Chillwave Mix [Chill - Relax ...
Chill Pulse Music
181 тыс. просмотров • 1 месяц...

 The Midnight - Days of Thunder/ Gloria
NewRetroWave


Персональные

Не персональные, но самый популярный

ТАКСОНОМИЯ РЕКОМЕНДАЦИЙ

Контекст и интерфейс создают успех!

Введите запрос



The Midnight - Crystalline

NewRetroWave

1,26 млн подписчиков

Вы подписаны

35 тыс.

681

Поделиться

Создать клип

3,5 млн просмотров

5 лет назад

New Retro Wave + The Midnight

Crystalline

The Midnight are back with a superb new single! The track oozes greatness. Listen up and enjoy! :)

Ещё

1 568 комментариев

Упорядочить

Введите комментарий

Закреплено пользователем NewRetroWave

@NewRetroWave

5 лет назад (изменено)

The Midnight are back with a superb new single! These guys are set to take over the scene I tell ya. lol. The track oozes greatness. Listen up and enjoy! :)

OUT NOW!

<https://themidnight.bandcamp.com/track/crystalline>

591

Ответить

19 ответов


Релевантный контекст

Не релевантный контекст


oWave

Для вас


Похожий контент




Female Vocal Trance | The Voices Of Angels #19
Aurora
597 тыс. просмотров · 2 года...




TRIP + DRIVE [synthwave - chillwave - retrowave mix] [rick...]
admirablemeice
29 тыс. просмотров · 3 недели...



REGRETS - 2023
DRORELIMELECH120V
4 просмотра · 42 минуты назад
Новинка




Galactic Waves - A Downtempo Chillwave Mix [Chill - Relax -...]
Chill Pulse Music
181 тыс. просмотров · 1 месяц...



The Midnight - Days of Thunder/ Gloria
NewRetroWave

ТАКСОНОМИЯ РЕКОМЕНДАЦИЙ

Введите запрос



The Midnight - Crystalline

NewRetroWave

1,26 млн подписчиков

Вы подписаны

35 тыс.

681

Поделиться

Создать клип

3,5 млн просмотров

5 лет назад

New Retro Wave + The Midnight

Crystalline

The Midnight are back with a superb new single! The track oozes greatness. Listen up and enjoy! :)

Ещё

1 568 комментариев

Упорядочить

Введите комментарий

Закреплено пользователем NewRetroWave

@NewRetroWave

5 лет назад (изменено)

The Midnight are back with a superb new single! These guys are set to take over the scene I tell ya. lol. The track oozes greatness. Listen up and enjoy! :)


OUT NOW!

<https://themidnight.bandcamp.com/track/crystalline>

591

Ответить


19 ответов



Wave

Для вас


Похожий контент



Female Vocal Trance | The Voices Of Angels #19

Aurora


597 тыс. просмотров · 2 года...



TRIP + DRIVE [synthwave - chillwave - retrowave mix] [rick...]

admirablemice

29 тыс. просмотров · 3 недели...




REGRETS - 2023

DRORELIMELECH120V

4 просмотра · 42 минуты назад


Новинка



Galactic Waves – A Downtempo Chillwave Mix [Chill - Relax - ...]

Chill Pulse Music

181 тыс. просмотров · 1 месяц...



The Midnight - Days of Thunder/ Gloria

NewRetroWave

Агрегатор рекламы
Это уже не рекомендация



ИНФОРМАЦИЯ О ПОЛЬЗОВАТЕЛЯХ

У меня есть данные, но как на их основе
строить рекомендации?



~~Красная или синяя, Neo~~
CSR или CSC матрица



ПРЕДСТАВЛЕНИЕ ДАННЫХ

UID	пол	возраст	регион	кол-во скроринга	кол-во просмотров	рекомендация	рейтинг
1	М	23	М	0,11	1	43	3
2	Ж	25	М	0,43	2	11	4
3	Ж	18	С	0,95	1	7	1



UID	43	11	7
1	3	?	?
2	?	4	?
3	?	?	1

UID	IID	Rating
1	43	3
2	11	4
3	7	1

ПРЕДСТАВЛЕНИЕ ДАННЫХ

Compressed Sparse Column Matrix

UID	43	11	7
1	3	?	?
2	?	4	?
3	?	?	1

Кумулята не 0 в столбце: 0, 1, 2
Строка: 0, 1, 2
Значение: 3, 4, 1
Размерность: 3x3

Compressed Sparse Row Matrix

UID	43	11	7
1	3	?	?
2	?	4	?
3	?	?	1

Кумулята не 0 в строке: 0, 1, 2
Столбец: 0, 1, 2
Значение: 3, 4, 1
Размерность: 3x3

ПРЕДСТАВЛЕНИЕ ДАННЫХ

Compressed Sparse Column Matrix

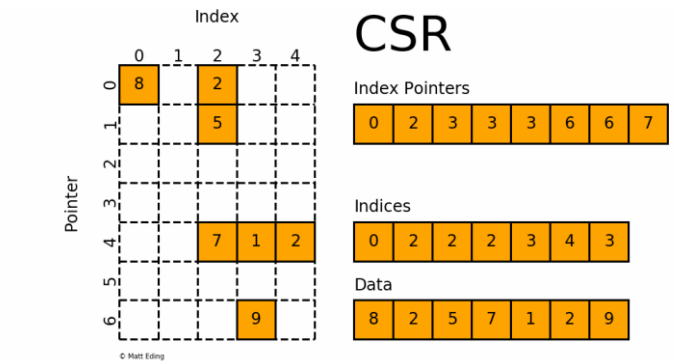
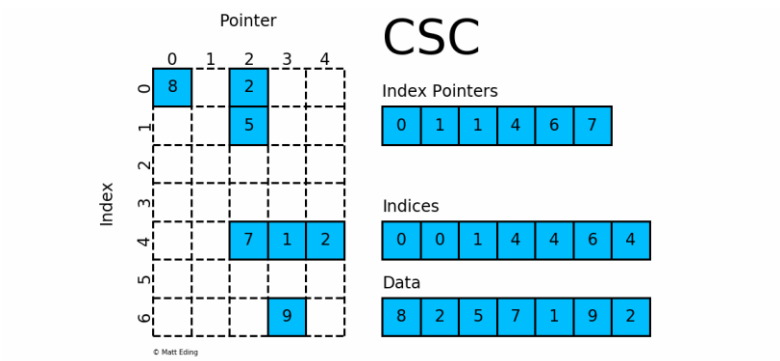
UID	43	11	7
1	3	?	?
2	?	?	?
3	?	4	1

Кумулята не 0 в столбце: 0, 1, 2
Строка: 0, 0, 2
Значение: 3, 4, 1
Размерность: 3x3

Compressed Sparse Row Matrix

UID	43	11	7
1	3	?	?
2	?	?	4
3	?	?	1

Кумулята не 0 в строке: 0, 1, 2
Столбец: 0, 0, 2
Значение: 3, 4, 1
Размерность: 3x3



ПРЕДСТАВЛЕНИЕ ДАННЫХ

Тип	Пустой вид
CSC	Matrix((M, N), dtype)
CSR	Matrix((M, N), dtype)
BSR	Matrix((M, N), blocksize = (R,C), dtype)
COO	Matrix((M, N), dtype)
DOK	Matrix((M, N), dtype)
LIL	Matrix((M, N), dtype)
DIA	Matrix((M, N), dtype)

ИНФОРМАЦИЯ О ПОЛЬЗОВАТЕЛЕ

Так, а что мы там будем хранить?
Мы же знаем о пользователе всё?



ИНФОРМАЦИЯ О ПОЛЬЗОВАТЕЛЕ. АРХЕТИПЫ



Бесстрашный

Ужасы, триллеры, фанфик

Всегда смотрю самые жесткие фильмы, но не против фильмов на разные любительские тематики

В системе: 150 дней
Покупок: 21

Семейный, с особенностями

Комедии, детское, семейное

В основном включаю мультки ребенку, но иногда смотрю комедии 90х

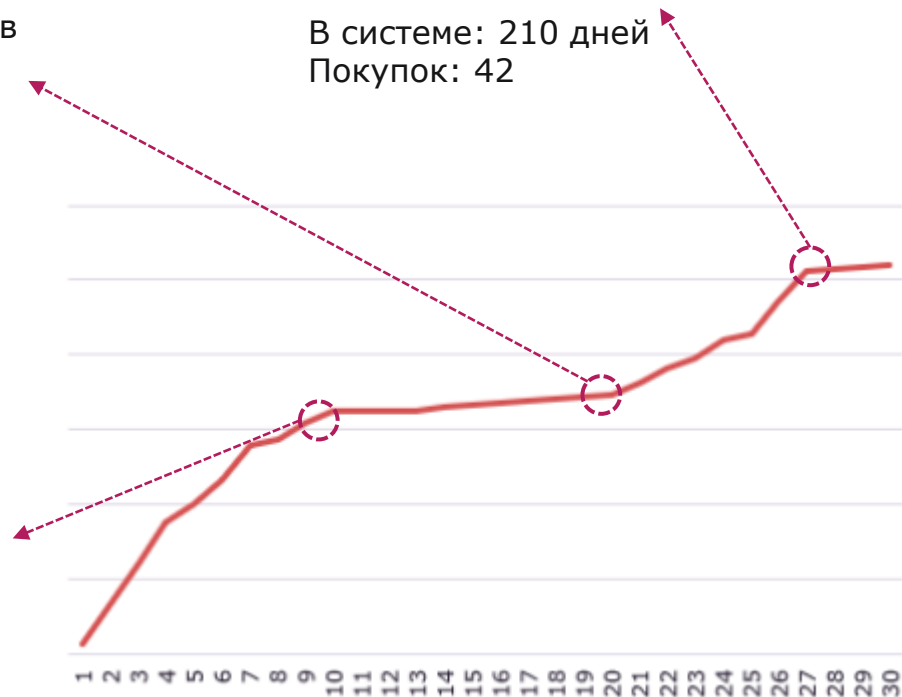
В системе: 90 дней
Покупок: 3

Успешный успех

Драмы, экшн, документальные

Смотрю фильмы на основе разных событий. Любую движуху.

В системе: 210 дней
Покупок: 42



ИНФОРМАЦИЯ О ПОЛЬЗОВАТЕЛЕ. АРХЕТИПЫ

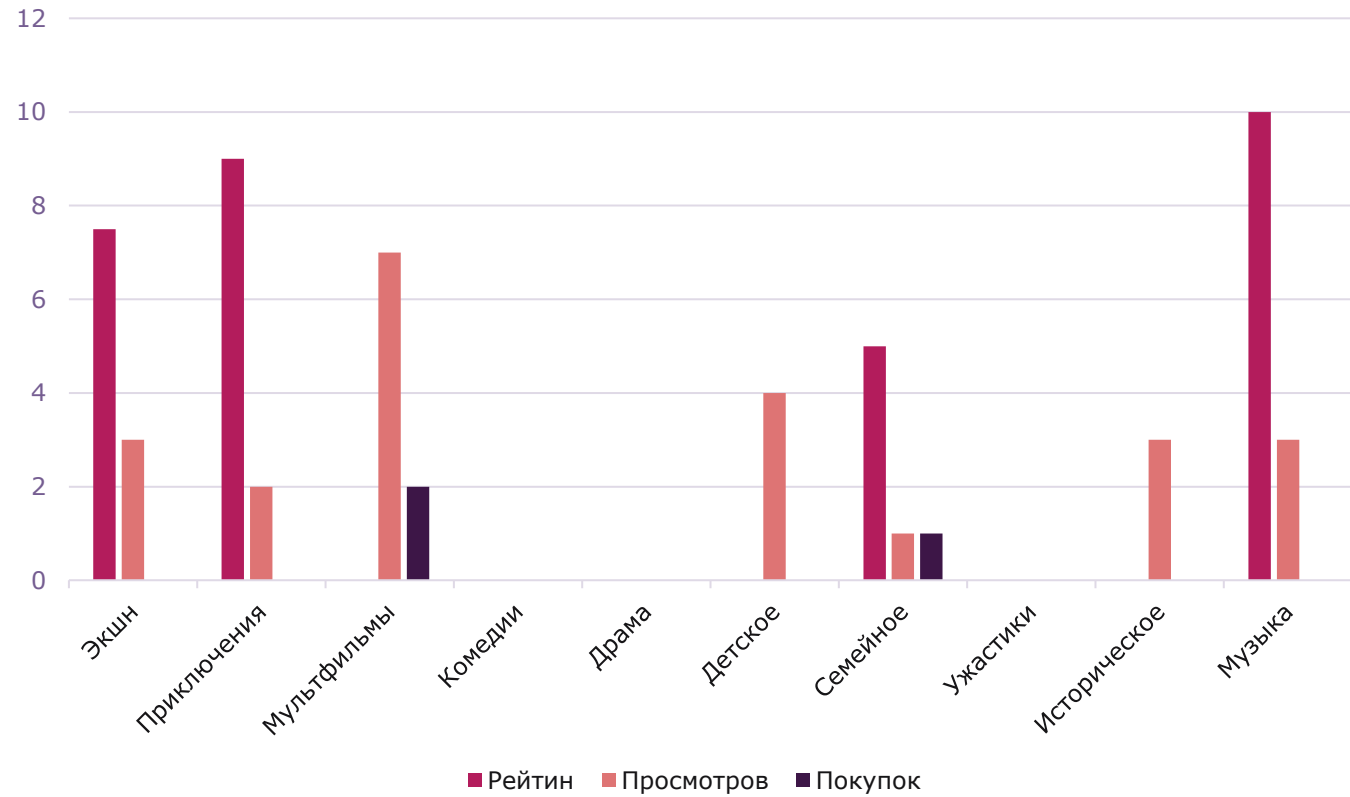


Семейный, с особенностями

Комедии, детское, семейное

В основном включаю мультики
ребенку, но иногда смотрю
комедии 90х

В системе: 90 дней
Покупок: 3



ИНФОРМАЦИЯ О ПОЛЬЗОВАТЕЛЕ



Действие в системе

Просмотр объекта (задержка внимания)

Переход на объект

Добавление в корзину / отметка объекта

Покупка

Оценка / рейтинг

Рекомендация

Тайминг проведенный с объектом

Тайминг в системе

Оставленная корзина

Поисковые запросы в системе

Действия с объектом

(чтение, перемотка, воспроизведение)

Предложите выбор

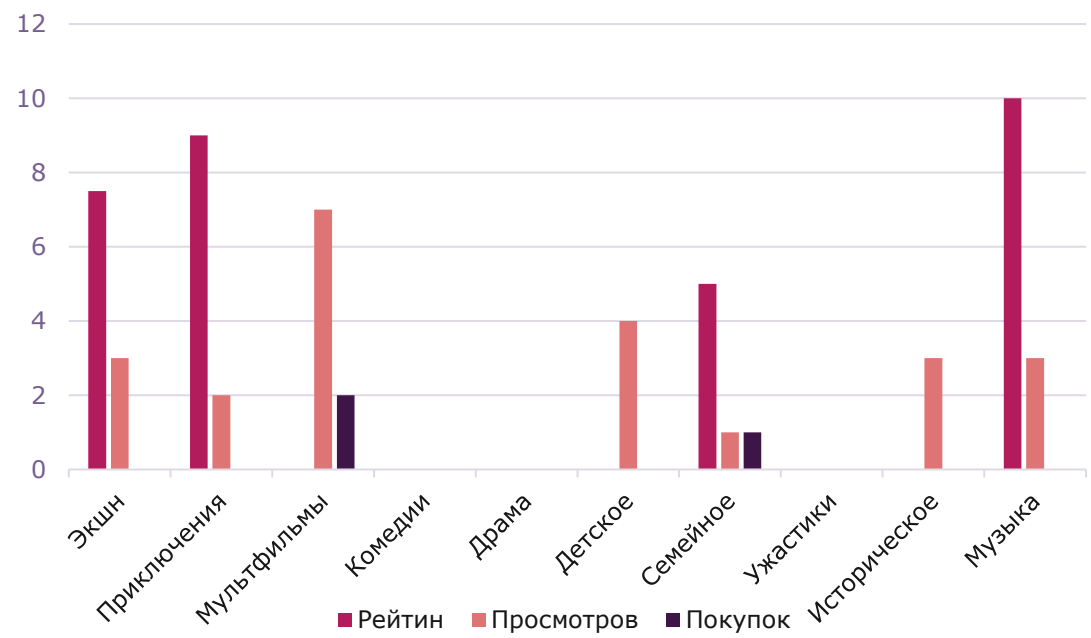
(корректировка по жанрам и т.д.)

ИНФОРМАЦИЯ О ПОЛЬЗОВАТЕЛЕ

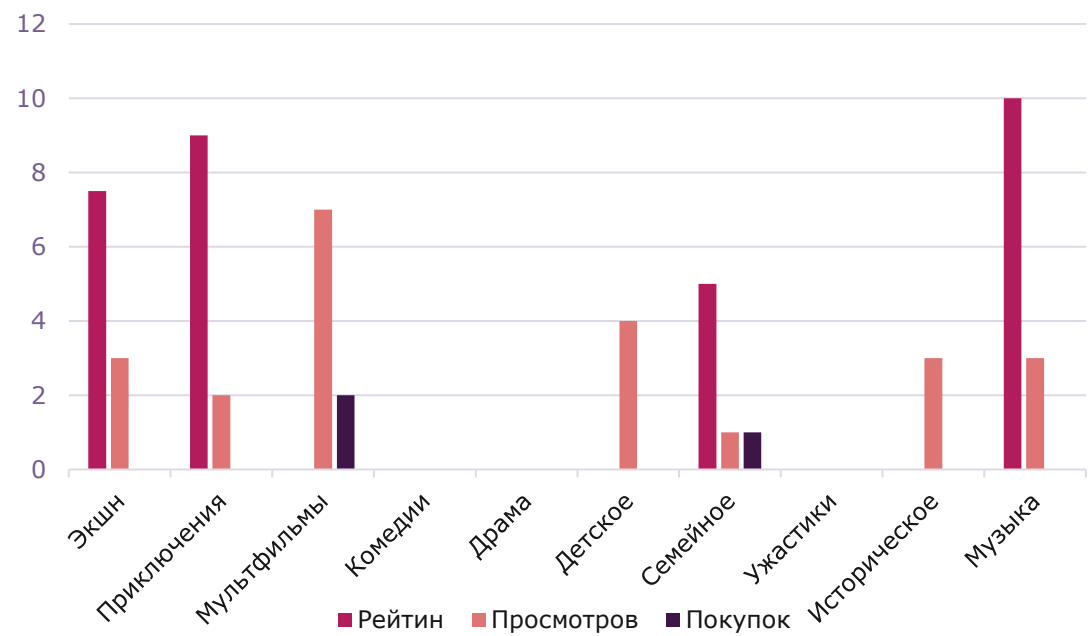


Действие в системе	Использование
Просмотр объекта (задержка внимания)	Фильтр
Переход на объект	Фильтр
Добавление в корзину / отметка объекта	Фильтр
Покупка	Рекомендация
Оценка / рейтинг	Рекомендация
Рекомендация	Рекомендация
Тайминг проведенный с объектом	Фильтр
Тайминг в системе	Фильтр
Оставленная корзина	Фильтр
Поисковые запросы в системе	Рекомендация
Действия с объектом (чтение, перемотка, воспроизведение)	Фильтр
Предложите выбор (корректировка по жанрам и т.д.)	Рекомендация

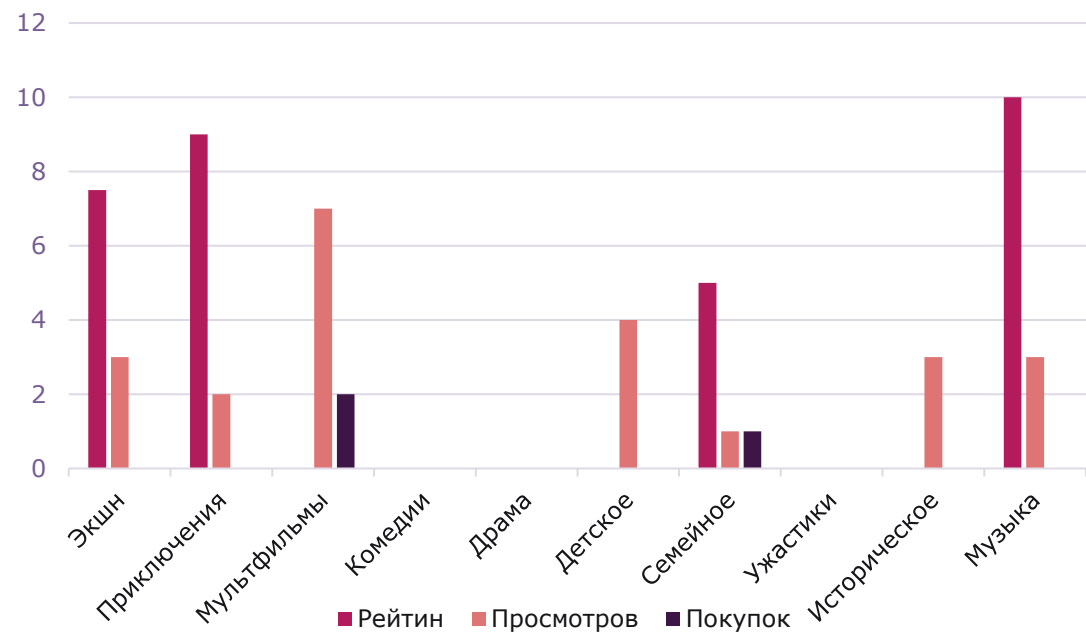
ЧТО ЕСТЬ ОЦЕНКА?



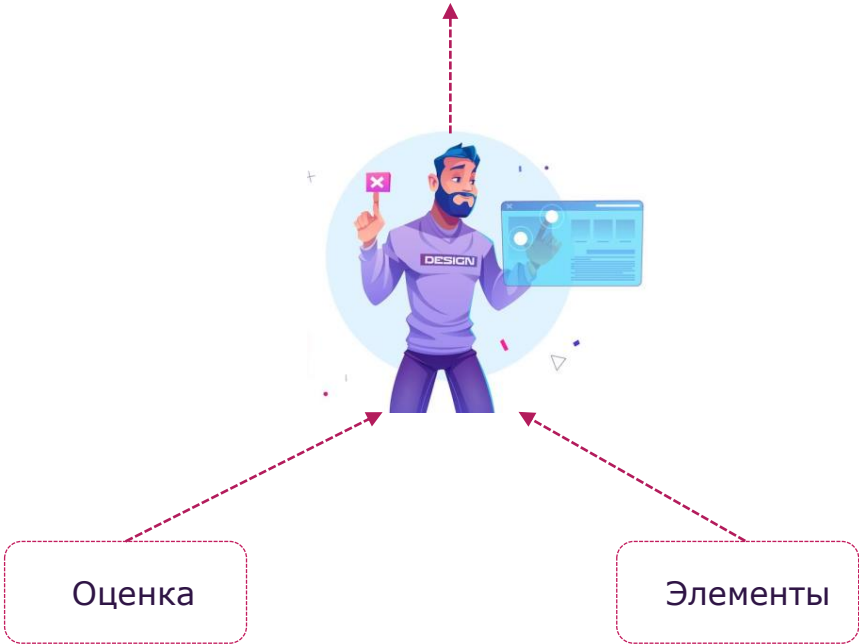
ЧТО ЕСТЬ ОЦЕНКА?



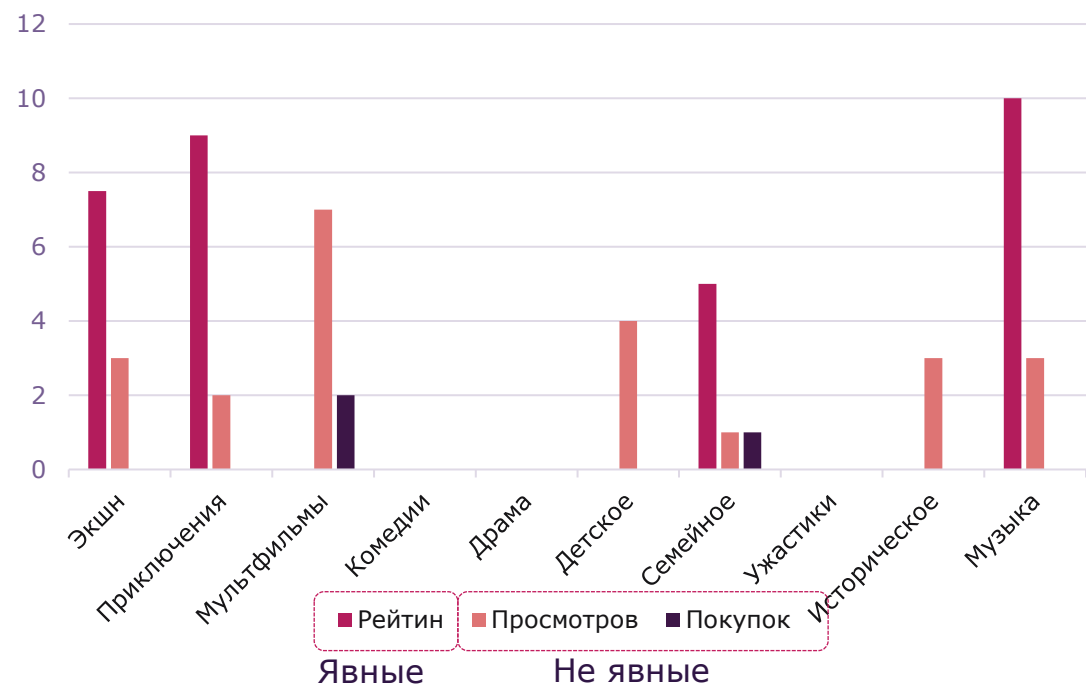
ЧТО ЕСТЬ ОЦЕНКА?



UID	43	11	7
1	3	?	?
2	?	4	?
3	?	?	1



ЧТО ЕСТЬ ОЦЕНКА?



UID	43	11	7
1	3	?	?
2	?	4	?
3	?	?	1



Выбираем тот тип данных, который дает больше информации

ЧТО ЕСТЬ ОЦЕНКА



UID	43	11	7
1	1	0	0
2	0	1	0
3	0	0	1

Купил или нет. Двоичная матрица.

Черно-белая матрица =)

UID	43	11	7
1	10/N	0	0
2	0	77/N	0
3	0	0	3/N

Нормализованные на кол-во значения

UID	43	11	7
1	A1	0	0
2	0	A2	0
3	0	0	A3

Значение на основе функции:
рейтинг – $1 / (a + 2)^b$

ЧТО ЕСТЬ ОЦЕНКА (НЕ ЯВНАЯ ОЦЕНКА)



Действие	Оценка
Покупка	Высшая оценка
Просмотр и долгая работа со страницей	Хорошая оценка
Возвращение к просмотру	Удовлетворительно
Один просмотр	Плохо

$$IS = (w1*событие1) + ... + (wn*событиен)$$

ЧТО ЕСТЬ ОЦЕНКА (НЕ ЯВНАЯ ОЦЕНКА)



Действие	Оценка
Покупка	Высшая оценка
Просмотр и долгая работа со страницей	Хорошая оценка
Возвращение к просмотру	Удовлетворительно
Один просмотр	Плохо

UID	43	11	7
1	1	IS	IS
2	IS	1	IS
3	IS	IS	1

$IS = (w1*событие1) +...+ (wn*событиен)$

ЧТО ЕСТЬ ОЦЕНКА. РЕДКИ ЭЛЕМЕНТЫ БОЛЕЕ ЦЕННЫ

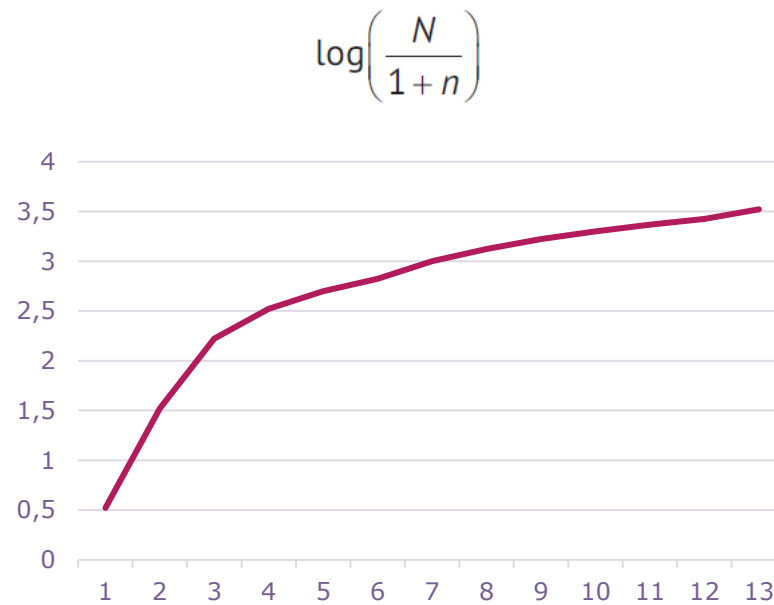
У нас 1.000.000 аудитория, 100 оценок на элементе для нас не важны



Это могут быть редкие элементы, которые очень значимы



ЧТО ЕСТЬ ОЦЕНКА. РЕДКИ ЭЛЕМЕНТЫ БОЛЕЕ ЦЕННЫ



$$R_{i,u} \times \log\left(\frac{N}{1+n}\right)$$

R – рейтинг
N – кол-во пользователей
n – кол-во оценок элемента

Наша система должна рекомендовать сразу.
Нет времени ждать «знаний» о пользователе



Наверное, я решу это проблему
как «холодный старт»



Но проще сделать не персональные...



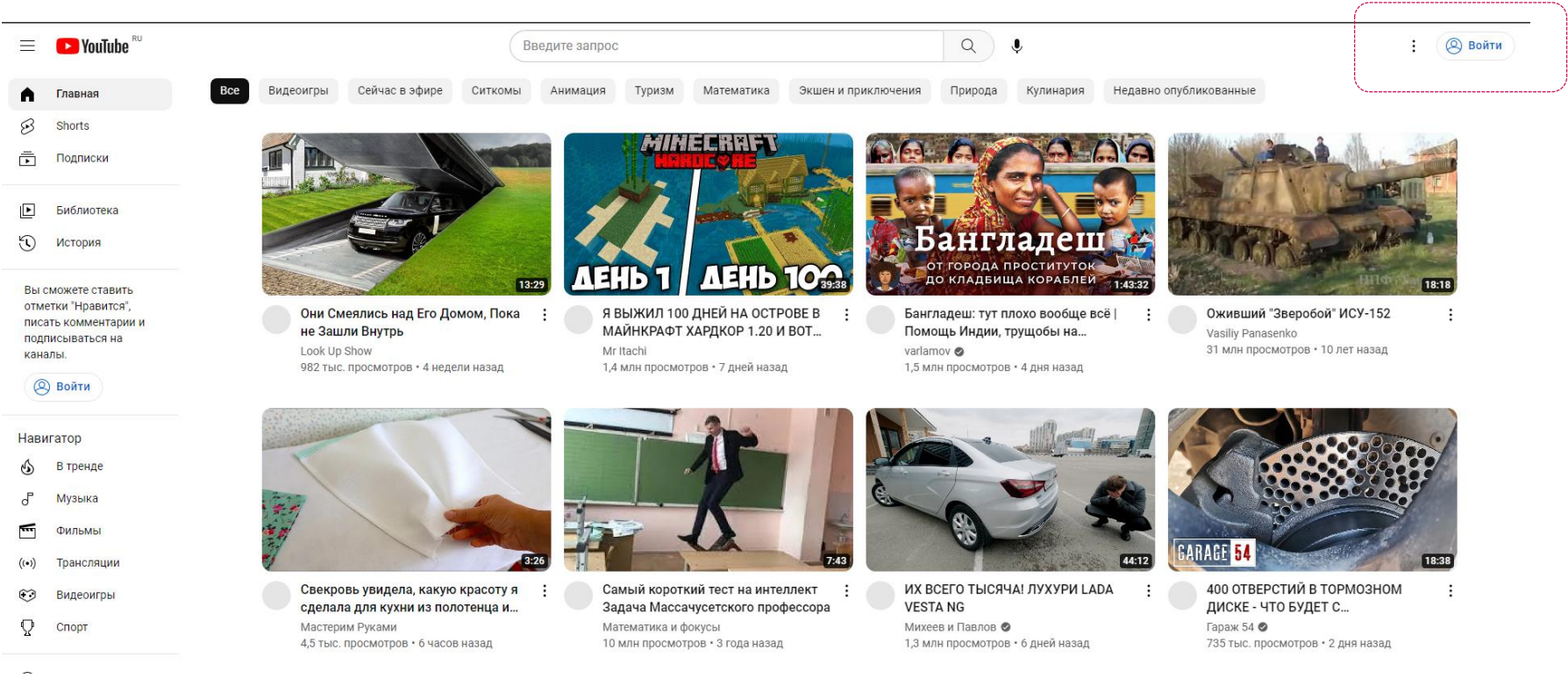


НЕ ПЕРСОНАЛЬНЫЕ РЕКОМЕНДАЦИИ

НЕ ПЕРСОНАЛЬНЫЕ РЕКОМЕНДАЦИИ

Реклама
– попытка убедить

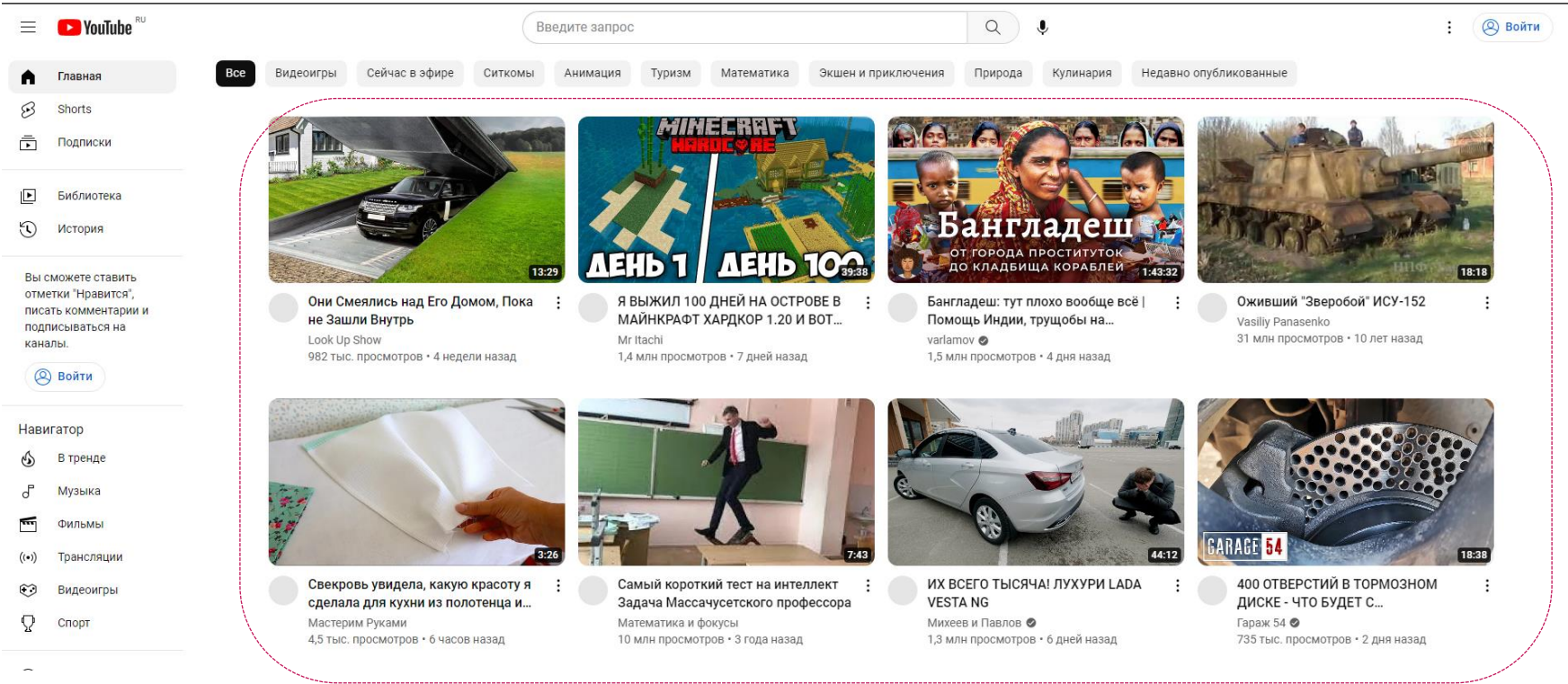
Рекомендация
– то, что хочет пользователь



НЕ ПЕРСОНАЛЬНЫЕ РЕКОМЕНДАЦИИ

Демонстрация контента:

- Топ N (по региону)
- Среднее / медиана по рейтингу
- «Выбор редакции»
- Купили / посмотрели сегодня
- Набор ассоциативных правил



НЕ ПЕРСОНАЛЬНЫЕ РЕКОМЕНДАЦИИ

Демонстрация контента:

Топ N (по региону)

Среднее / медиана по рейтингу

«Выбор редакции»

Купили / посмотрели сегодня

Набор ассоциативных правил

Рассмотрим на примере



**ВЕРНЕМСЯ К ТЕМЕ, КОГДА БУДЕМ РЕШАТЬ
«ПРОБЛЕМЫ РЕКОМЕНДАЦИЙ»**



ПРОСТЫЕ РЕКОМЕНДАЦИИ

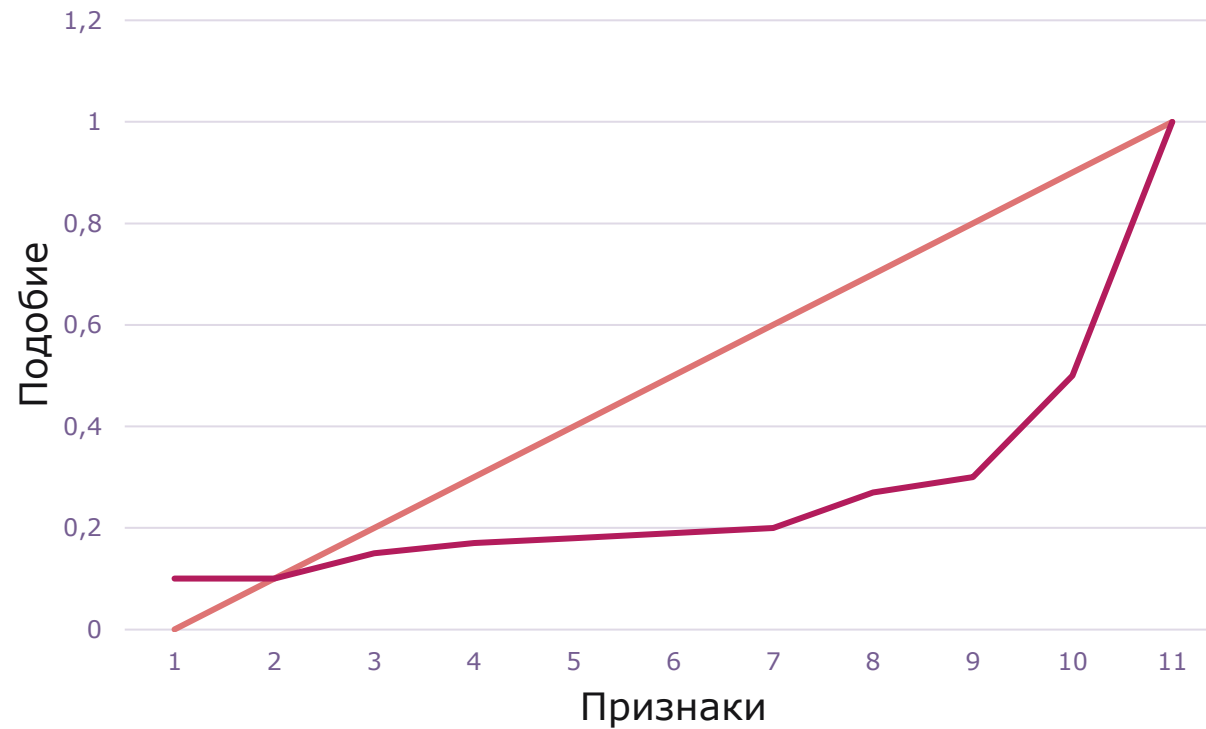
А что такое, это ваше «сходство»?



СХОДСТВО И ПОДОБИЕ

$SIM(i, j) = 1$ - идентичны

$SIM(i, j) = 0$ – не имеют общего



СХОДСТВО И ПОДОБИЕ

Для разных данных, разное подобие

Тип	Пример	Подобие
Количество (лайки, покупки)	Uid поставил лайк. Uid купил этот товар	Коэф. Жаккара
Бинарные данные	UiD не понравился фильм	Коэф. Жаккара
Количественные данные	UiD поставил 7/10 оценку	Коэф.Пирсона, Косинус, L-нормы

$$сходство_{Жаккара}(i, j) = \frac{\#пользователи, купившие оба товара}{\#пользователи, которые купили либо i, либо j}$$

СХОДСТВО И ПОДОБИЕ

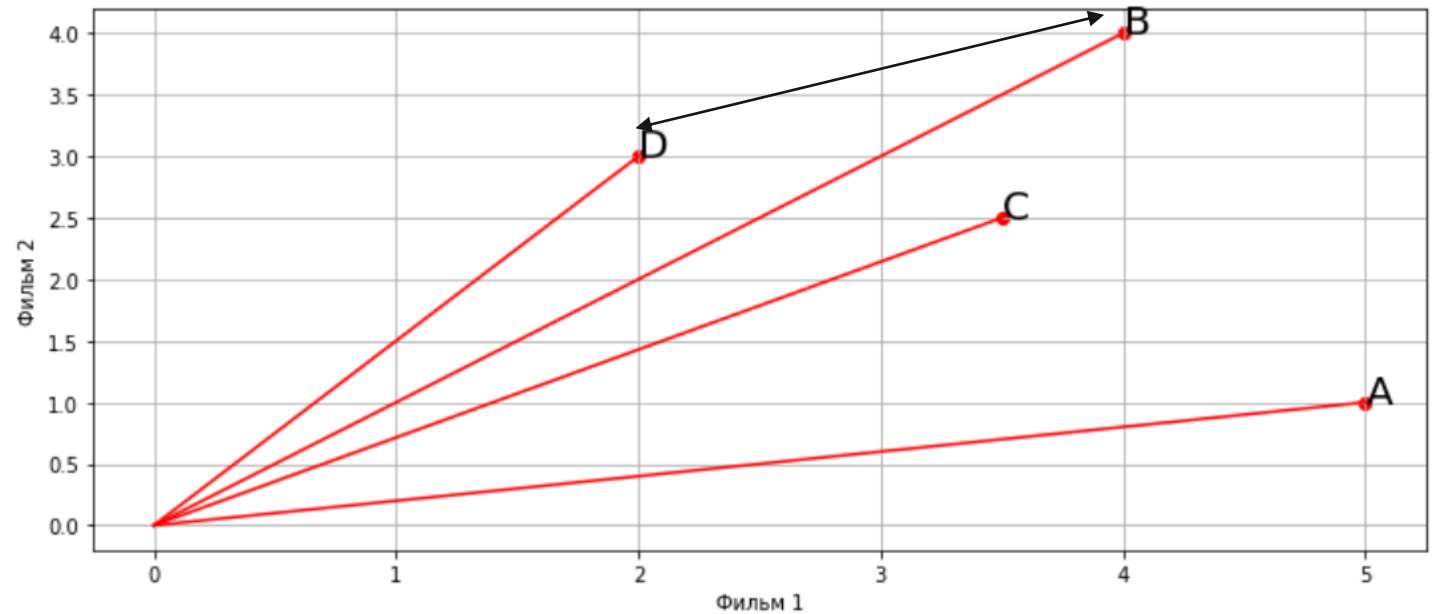
Эвклидово расстояние (L2-norm)

$$d(x, y) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$



RMSE

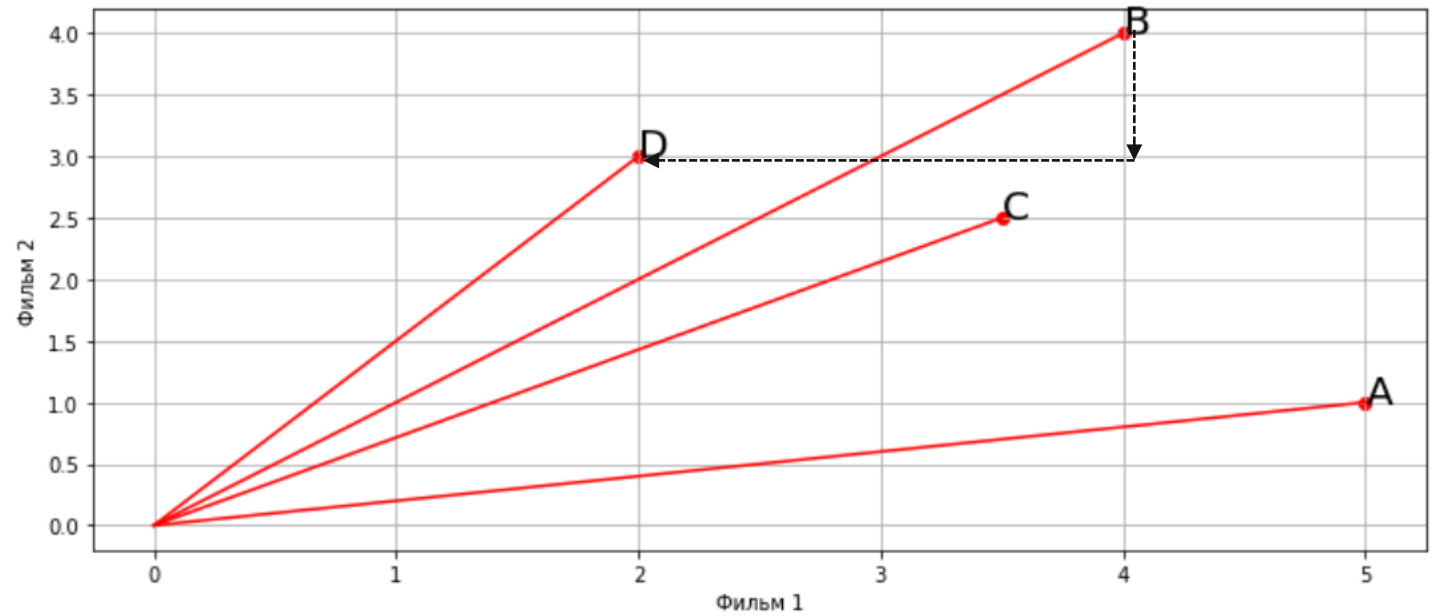
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2}$$



СХОДСТВО И ПОДОБИЕ

Манхэттен (L1-norm)

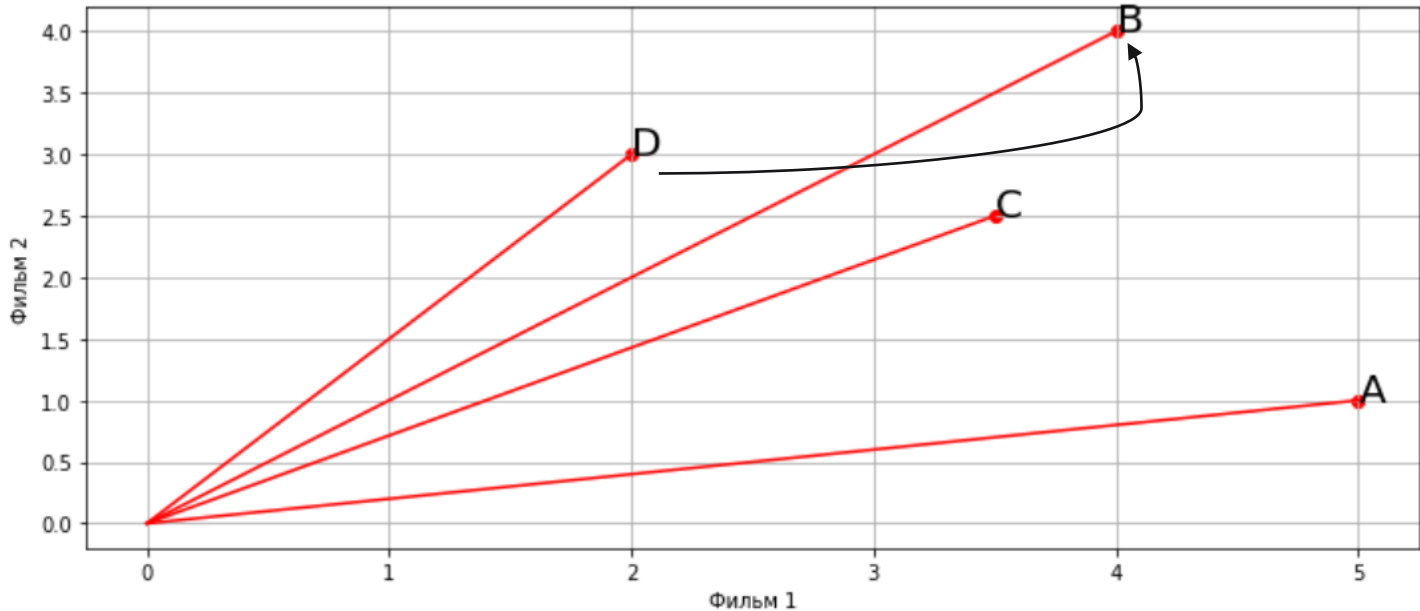
$$|x_1 - x_2| + |y_1 - y_2|$$



СХОДСТВО И ПОДОБИЕ

Минковский

$$\sqrt[p]{(x_1 - y_1)^p + (x_2 - y_2)^p + \dots + (x_N - y_N)^p}$$

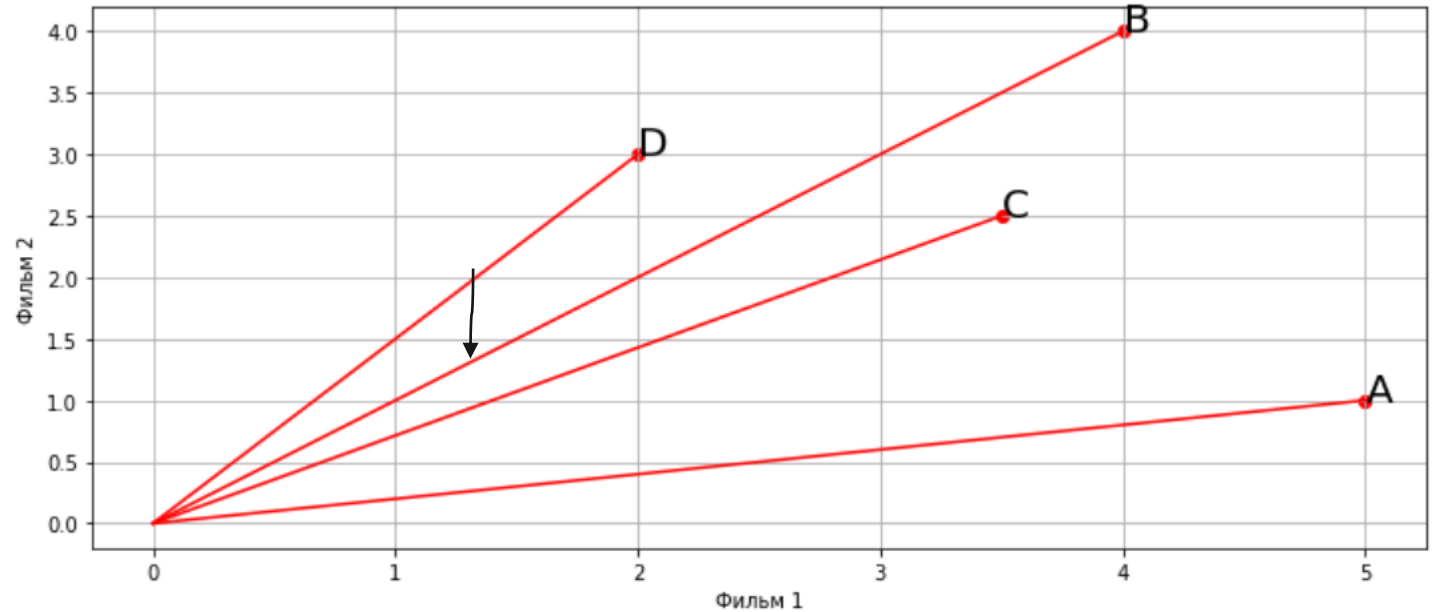


СХОДСТВО И ПОДОБИЕ

Косинусное (Отиаи)

$$\text{Cosine Similarity (A,B)} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

$$K = \frac{|A \cap B|}{\sqrt{|A| \cdot |B|}}$$



СХОДСТВО И ПОДОБИЕ

$SIM(i, j) = 1$ - идентичны

$SIM(i, j) = 0$ – не имеют общего

Рассмотрим на примере



Хорошо, а в другой команде используют
какие-то «фильтрации»



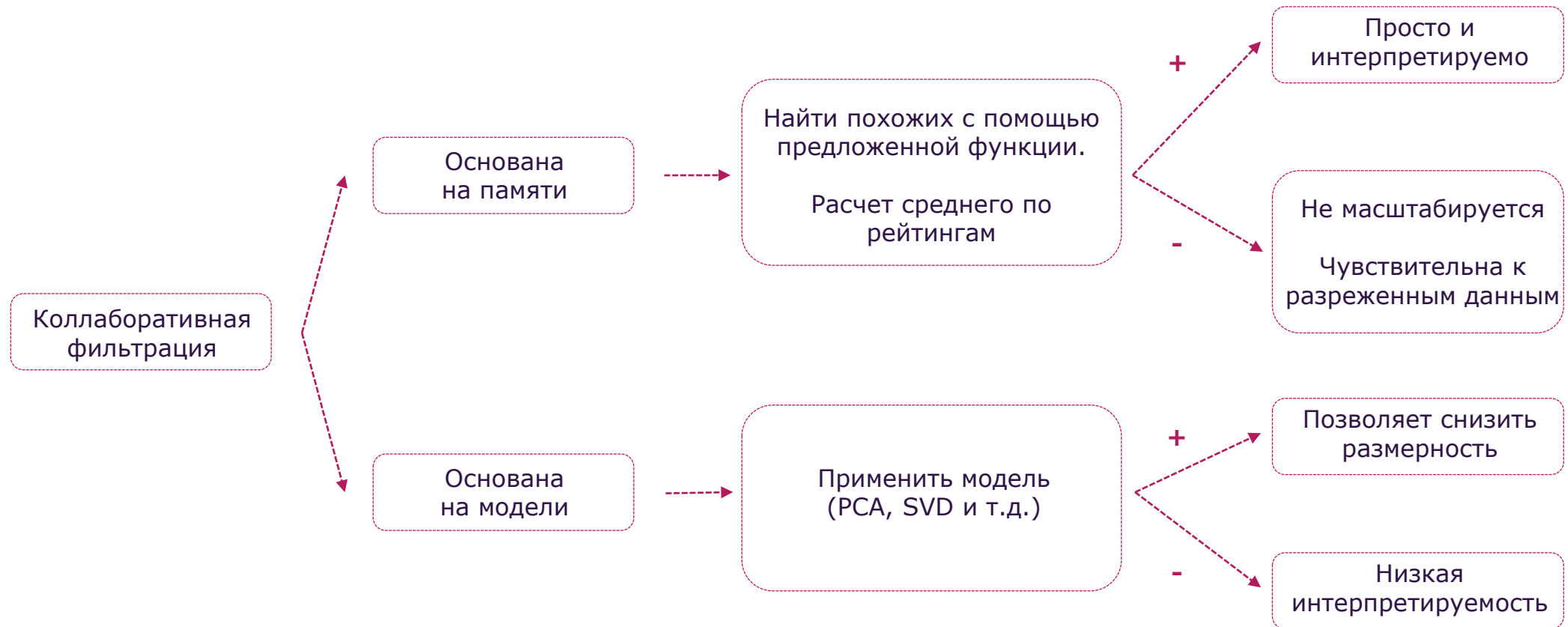
Есть фильтрации.

При их использовании:

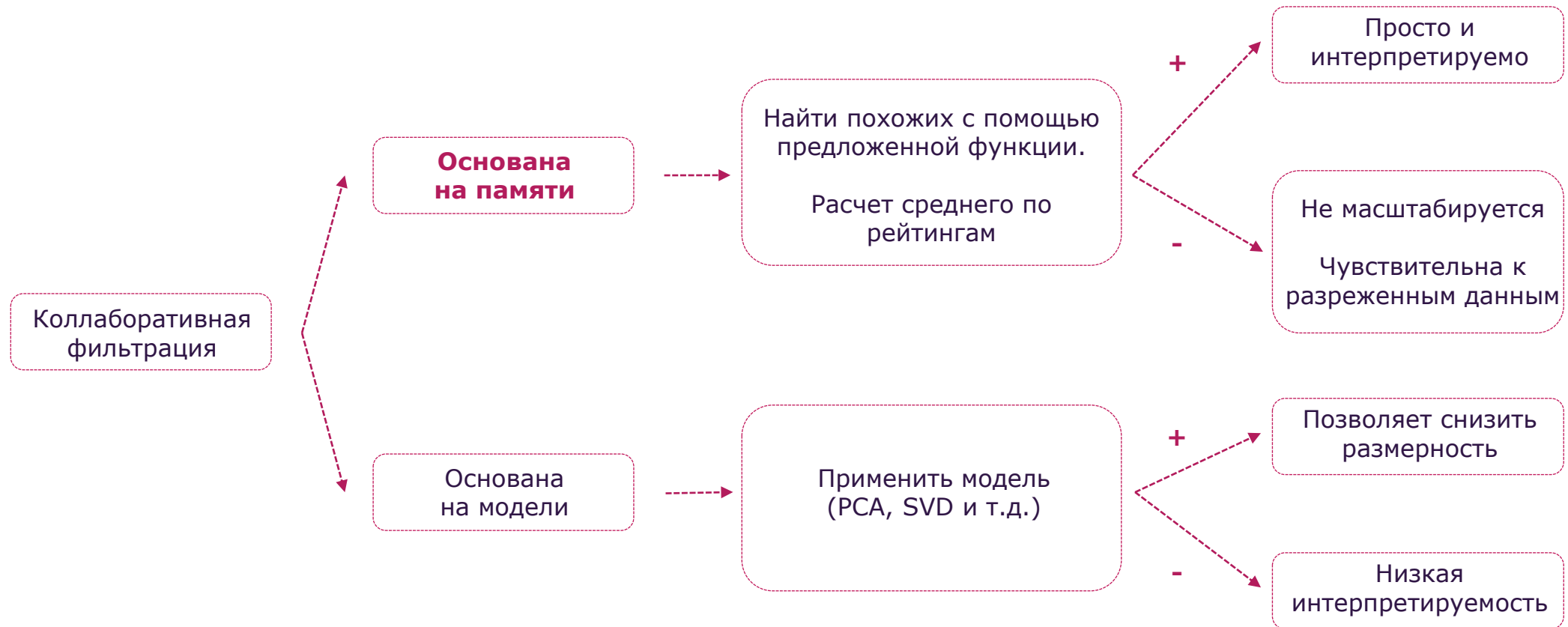
- если контент без оценки – он не рекомендуется;
- если у пользователя нет перекрывающихся оценок, то на него будут плохие рекомендации.



КОЛЛАБОРАТИВНАЯ ФИЛЬТРАЦИЯ



КОЛЛАБОРАТИВНАЯ ФИЛЬТРАЦИЯ



СОВМЕСТНАЯ ФИЛЬТРАЦИЯ

UID	43	11	7
1	3	-	1
2	-	4	-
3	-	-	1

СОВМЕСТНАЯ ФИЛЬТРАЦИЯ

UID	43	11	7
1	3	-	1
2	-	4	-
3	-	-	1



Вычисляем сходство между
пользователем (UID1) и остальным

UID	1
1	1
2	0
3	0.5

СОВМЕСТНАЯ ФИЛЬТРАЦИЯ

UID	43	11	7
1	3	-	1
2	-	4	-
3	-	-	1

----->
Вычисляем сходство между
пользователем (UID1) и остальным

UID	1
1	1
2	0
3	0.5

----->
Сортируем по результату

СОВМЕСТНАЯ ФИЛЬТРАЦИЯ

UID	43	11	7
1	3	-	1
2	-	4	-
3	-	-	1

----->
Вычисляем сходство между
пользователем (UID1) и остальным

UID	1
1	1
2	0
3	0.5

----->
Сортируем по результату

UID	1
1	1
3	0.5
2	0

СОВМЕСТНАЯ ФИЛЬТРАЦИЯ

UID	43	11	7
1	3	-	1
2	-	4	-
3	-	-	1

Вычисляем сходство между пользователем (UID1) и остальным

UID	1
1	1
2	0
3	0.5

Сортируем по результату

UID	1
1	1
3	0.5
2	0

Выбираем диапазон ближайших,
Вычисляем прогноз оценки

Топ N (количество)

Кластеризация

Установить пороговое значение (качество)

R = среднее по Топ N

СОВМЕСТНАЯ ФИЛЬТРАЦИЯ

UID	43	11	7
1	3	-	1
2	-	4	-
3	-	-	1

Вычисляем сходство между
пользователем (UID1) и остальным

UID	1
3	1
2	0
1	0.5

Сортируем по результату

UID	1
3	1
1	0.5
2	0

Выбираем диапазон ближайших,
Вычисляем прогноз оценки

R = 3

СОВМЕСТНАЯ ФИЛЬТРАЦИЯ

UID	43	11	7
1	3	-	1
2	-	4	-
3	-	-	1

----->
Вычисляем сходство между
пользователем (UID1) и остальным

UID	1
1	1
2	0
3	0.5

----->
Сортируем по результату

UID	1
1	1
2	0.5
3	0

----->
Выбираем диапазон ближайших,
Вычисляем прогноз оценки

R = 4

СОВМЕСТНАЯ ФИЛЬТРАЦИЯ. КАК AMAZON

Для каждого товара в каталоге

Для каждого клиента, который купил элемент

Для каждого товара, который купил клиент

Записать приобретение клиентом товаров

Для каждого товара

Вычислить сходство между товарами

UID	43	11	7
1	3	-	1
2	-	4	3
3	-	-	1

СОВМЕСТНАЯ ФИЛЬТРАЦИЯ. КАК AMAZON

UID	43	11	7
1	3	-	1
2	-	4	3
3	-	-	1

IID	
43	7
11	7
7	

$$K = \frac{|A \cap B|}{\sqrt{|A| \cdot |B|}}$$

UID	43	11	7
1	K	-	K
2	-	K	K
3	-	-	K

	43	11	7
43	1	-	-
11	-	1	-
7	-	-	1

А как появляется предикт, а то тут только средние значения везде, да корреляции?



Да, как?



ПОСТРОЕНИЕ ПРОГНОЗА

Регрессия

Классификация

Фильтрация по элементам

ПОСТРОЕНИЕ ПРОГНОЗА. ЭЛЕМЕНТ - ЭЛЕМЕНТ

Predict(u, i)

$$Pred(u, i) = \bar{r}_u + \frac{\sum_{j \in S_i} (sim(i, j) \times r_{u, j})}{\sum_{j \in S_i} sim(i, j)}$$

\bar{r}_u – средняя оценка пользователя

r_{uj} – оценка элемента j пользователя

S_i – набор элементов по Топ N (или после другого отбора)

$sim(i, j)$ – сходство между элементами

Рассмотрим на примере





ССЫЛКИ С ДОП.МАТЕРИАЛОМ

ДОПОЛНИТЕЛЬНЫЙ МАТЕРИАЛ

