

Первая рекомендательная система без пота и слёз

“Many receive advice, only the wise profit from it.” – Harper Lee

27.05.2023



Кто я?

Юлия Карамнова

ML-инженер в команде
рекомендаций МТС Диджитал



tg: @juliakup



Где используются рекомендательные системы?

Кион

ФинТех

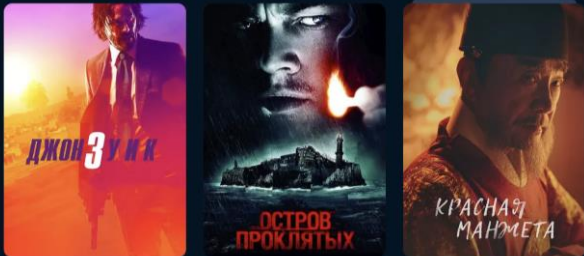
Лента новостей, постов

Ретейл

Знакомства

Реклама

Специально для вас



Джон Уик 3
★ 8,4 / Подписка


Остров проклятых
★ 8,9 / Подписка

Красная манжета
★ 9,6 / Подписка

С этим товаром покупают

Все товары Клип-кейс Стекло защитное Пленка защитная АЗУ Внешний аккумулятор Дата-кабель СЗУ Портативный зарядный блок Беспроводные наушники с микрофоном Дополнительная гарантия Защита покупки Комплексная защита покупки Беспроводные наушники

Тариф Наклейка стекла/пленки



Дата-кабель
Apple USB-Lightning 1m White (MXLY22M/A)

Зарядное устройство
Apple MagSafe White (MHXN3ZE/A)

Беспроводные наушники
Apple AirPods 2 (с зарядным футляром)...

СЗУ
Apple USB Type-C Белый MHIE32M/A

Похожие товары



Смартфон Apple iPhone 13 Pro 256Gb Небесно-голубо...



Смартфон Apple iPhone 13 Pro 128Gb Графитовый 0101...



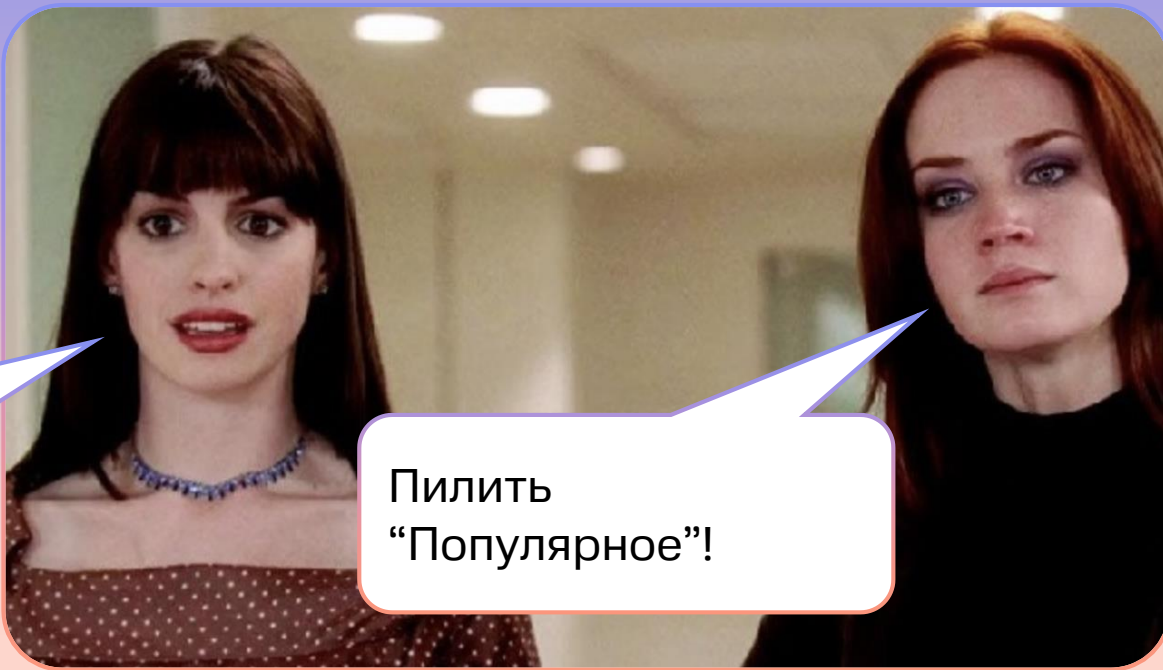
Смартфон Apple iPhone 13 Pro 128Gb Серебристый...





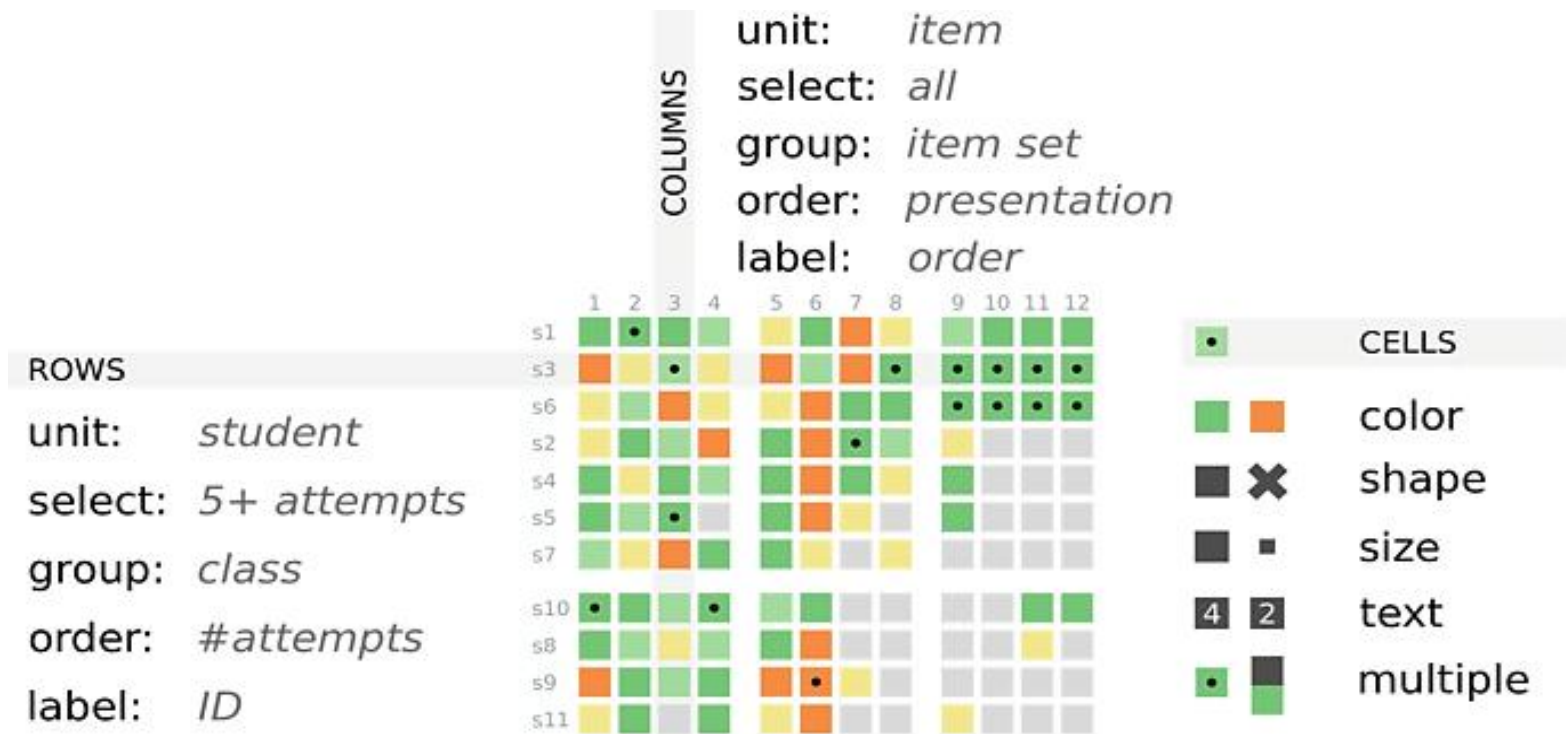
Мне нужны рекомендации в следующие
5 минут 😞

Что же делать?! 😱



Пилить
“Популярное”!

Стоп! А какие данные у нас вообще есть?



Как интерпретировать историю взаимодействий?

Виды фидбэка

- явный
- неявный

Что может стоять в значениях?

- факт взаимодействия
- рейтинг
- время удержания внимания
- и их производные



1,2K



238

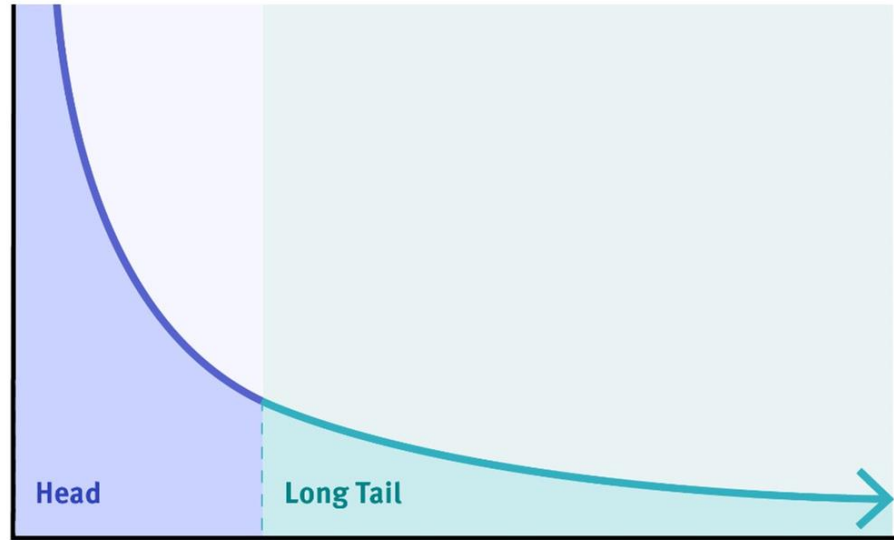


142



The Long Tail

Popularity



Head

Long Tail

Items

These items are:

- High-impact
- Popular
- Few in number
- Mainstream

These items are:

- Low-impact
- Niche
- Many in number
- Obscure

NN/g

NNGROUP.COM

Что такое “популярность”?

- число пользователей?
- число взаимодействий?
- что-то еще?

Популярное - это неперсонализированные рекомендации

К этой категории также относятся ассоциативные правила, рекомендации в зависимости от сегмента.

- + не требует иметь историю взаимодействий конкретного пользователя
- + интерпретируемость
- достоверность требует наличие достаточного объема накопленной истории
- малоэффективны и не позволяют достигаться до малых групп





Всем рекомендуется одно и то же. Мы никак не учитываем информацию о текущем просматриваемом айтеме.

Без проблем. Джарвис, посчитай content-based recommendations.



Content - based recommendations

Плюсы и минусы подхода



не требуется знаний о других пользователя,
так как зависит только от признаков айтемов



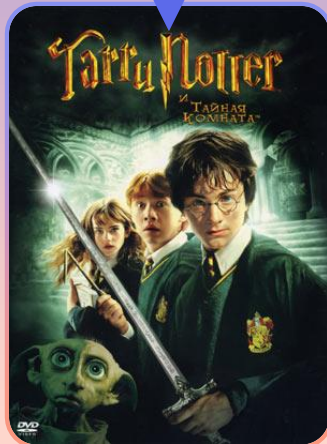
подходит для очень нишевых айтемов



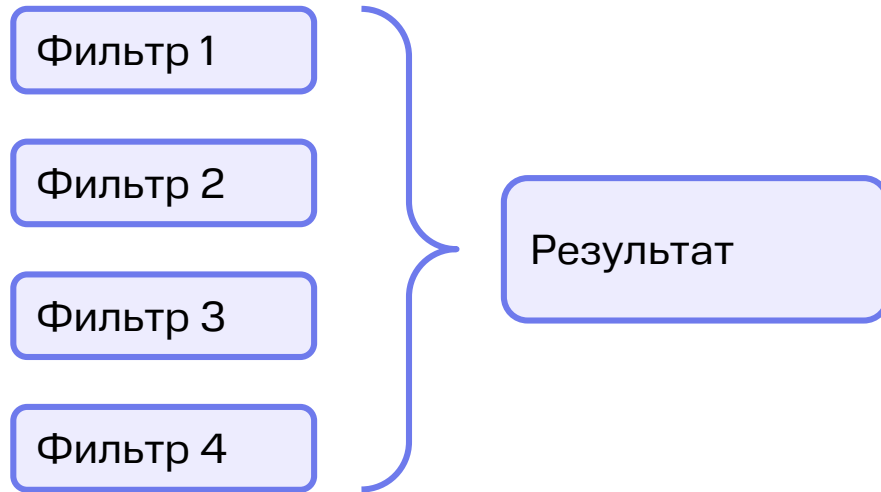
учитываются только текущие интересы
пользователя



требуется знания области и ручной генерации
фичей



Knowledge-based recommendations

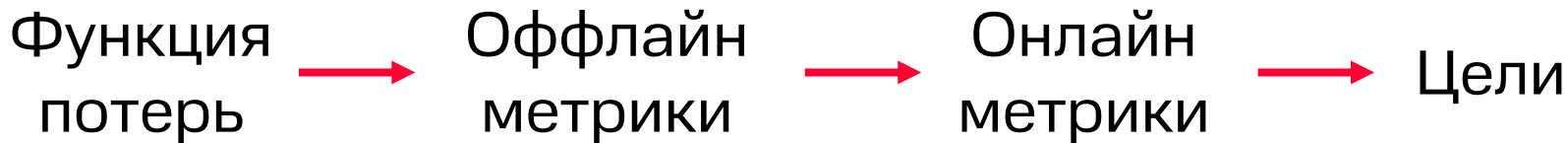


Оценка качества рекомендаций

	Офлайн	Онлайн
<i>Как измеряем</i>	На истории	На АВ-тесте
<i>Какие метрики</i>	<ul style="list-style-type: none">• Регрессия• Классификация• Ранжирование• Beyond accuracy metrics• Визуальный анализ	<p>Бизнес-метрики</p> <ul style="list-style-type: none">• клики• просмотры• лайки• покупки• подписки
<i>Что хорошо</i>	Легко и дешево	Более или менее объективно
<i>Что плохо</i>	Может не отражать действительность	Трудозатратно и дорого

Метрики

How it has started...



How it is going...



**Потолок качества
рекомендаций определяется
используемыми метриками**



Я хочу, чтобы каждый получал персонализированные рекомендации на основе своих предпочтений. Займись!

Есть! Я запущу алгоритм коллаборативной фильтрации.



Коллаборативная фильтрация



персонализация



можем рекомендовать
что-то новое
пользователям



не нужно вручную
создавать признаки



.....
требует знания области
и ручной генерации
фичей



Harry Potter



The Triplets of
Belleville



Shrek



The Dark
Knight Rises



Memento



?



“Холодное” vs “Горячее”













Мало данных

Новые или непопулярные объекты
Новые пользователи

Достаточно данных

Популярные объекты
Постоянные пользователи

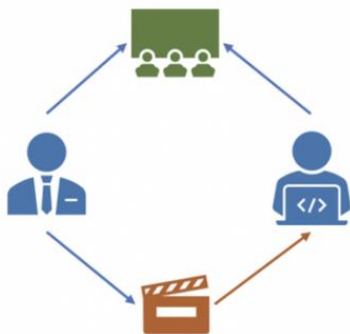


	Коллаборативная филترация	Контентные рекомендации
“Холодный” пользователь		
“Холодный” объект		
Объяснимость рекомендаций		
Простота в использовании		
Разнообразие, новизна, неожиданная релевантность рекомендаций		

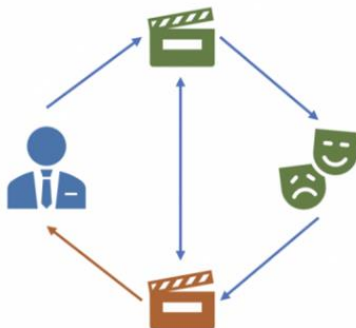


Виды рекомендательных моделей

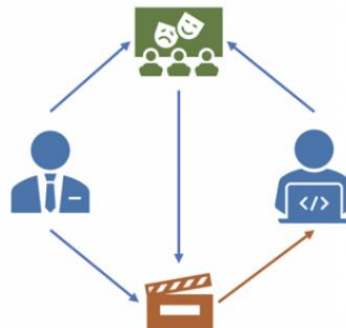
Collaborative Filtering



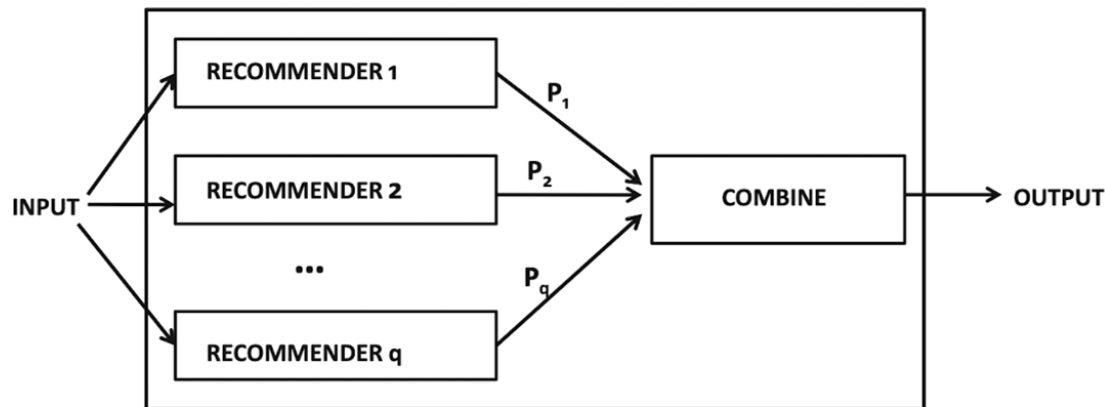
Content-based Filtering



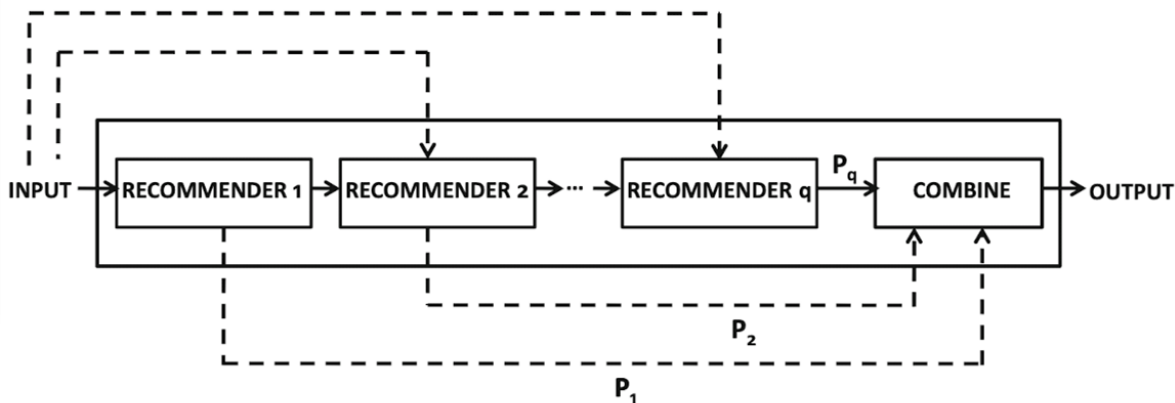
Hybrid systems



Hybrid systems



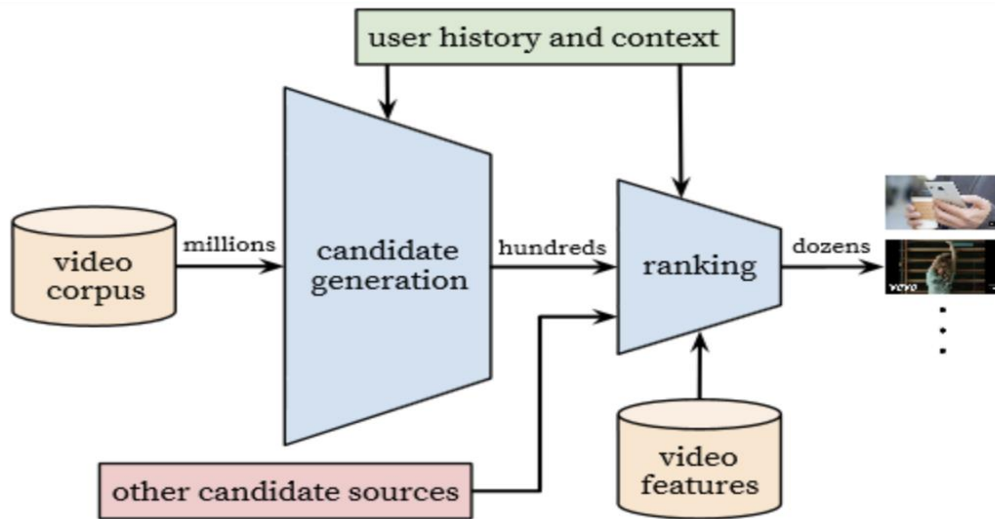
(a) Parallel design



(b) Sequential design

До этого мы сразу пытались ранжировать айтемы.

А что, если мы сначала отберем кандидатов, а потом из отранжируем - **Двухуровневые модели!**



YouTube Recommendation Model Architecture

Еще больше проблем!

- Проблема пакета
- Feedback loop
- Positional bias
- Exploration vs. Exploitation



Тренды в рекомендациях

Deep Learning

- Sequential Recommender Networks - SASRec
- Bert4Rec

Reinforcement Learning и Бандиты

- on-policy vs off-policy
- non-stationary environments
- impact of delayed rewards

Метрики

- loss engineering
- second-order optimization

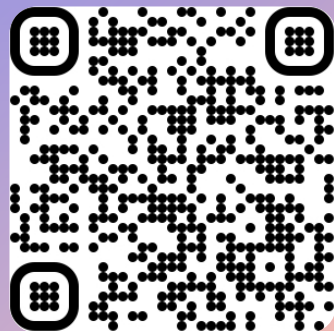
Что дальше?

Экспериментировать с
[Rectools](#)

Проходить курсы
[Your First RecSys](#) и
[Your Second RecSys](#)

Посмотреть [Microsoft Recommenders](#)

RecTools



Your First RecSys

