

Рекомендательные системы и информационный поиск

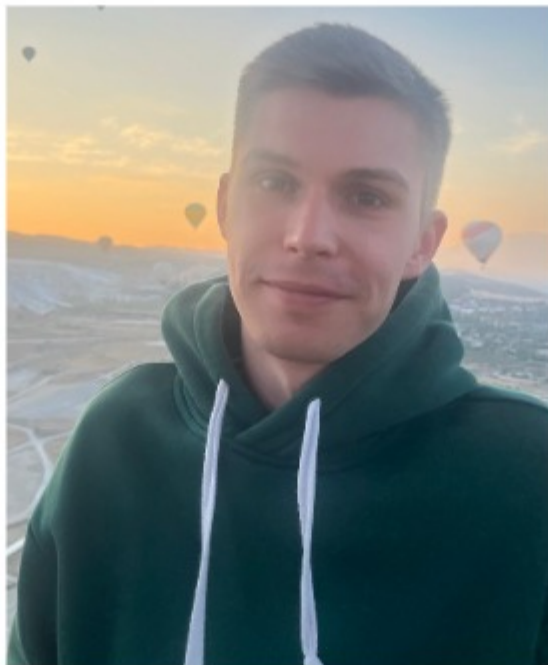
Занятие 1. Введение

Краснов Александр

Команда курса



Илья Осиновсков
@ossinovskov



Дмитрий Андросов
@androsovdy



Александр Краснов
@san9kresh

План курса

| | |
|---|--|
| Введение, рекомендательные системы | Про рекомендации и поиск, задачи рекомендаций, метрики рекомендаций, простые алгоритмы рекомендаций |
| Introduction to IR | Введение в информационный поиск, обработка текста, обратный индекс, представление поискового запроса, создание базовой поисковой системы |
| Ranked IR and search quality | Поиск и ранжирование, TF-IDF и BM25, векторное представление текста, оценка качества поиска |
| Методы основанные на статистиках | Ассоциативные правила, коллаборативные подходы, SVD, ALS |
| Контентные рекомендации | Контентные рекомендации, сжатые представления контента, гибридные рекомендации, LightFM, DeepFM, использование ANN/KNN |
| LTR, Многоуровневые ранжирующие системы | Задача LTR, лоссы и метрики, разбор концепций многоуровневых систем ранжирования |

План курса

| | |
|---|--|
| Sequential-based подходы | SASRec, BERT4REC, разбор sota подходов |
| Neural IR and language models | Векторное представление текста, векторный поиск, популярные архитектуры нейросетей для поиска, Использование LLM для поиска и рекомендаций |
| Разнообразие. Метрики. Подходы оптимизации | Понятие разнообразия в рекомендательных системах, метрики для оценки разнообразия, алгоритмы оптимизации |
| Проблемы индустриальных рекомендательных систем | Холодный старт контента/юзера, popularity/position bias, exploration-exploitation tradeoff |
| Графовые сети | Обзор графовых сетей, алгоритмы, граф знаний |
| Search query auto completion | Задача поисковых подсказок, префиксный поиск, ранжирование подсказок, персонализация |

План курса

Дополнительные материалы

- разбор статей
- разбор дизайнов промышленных систем
- систем дизайн с нуля
- разбор тем, которые не вошли в курс

Оценивание

4-6 домашек, 10 баллов макс за каждую

- собственные реализации алгоритмов и оценка качества
- контест
- реализации алгоритмов в сервисе

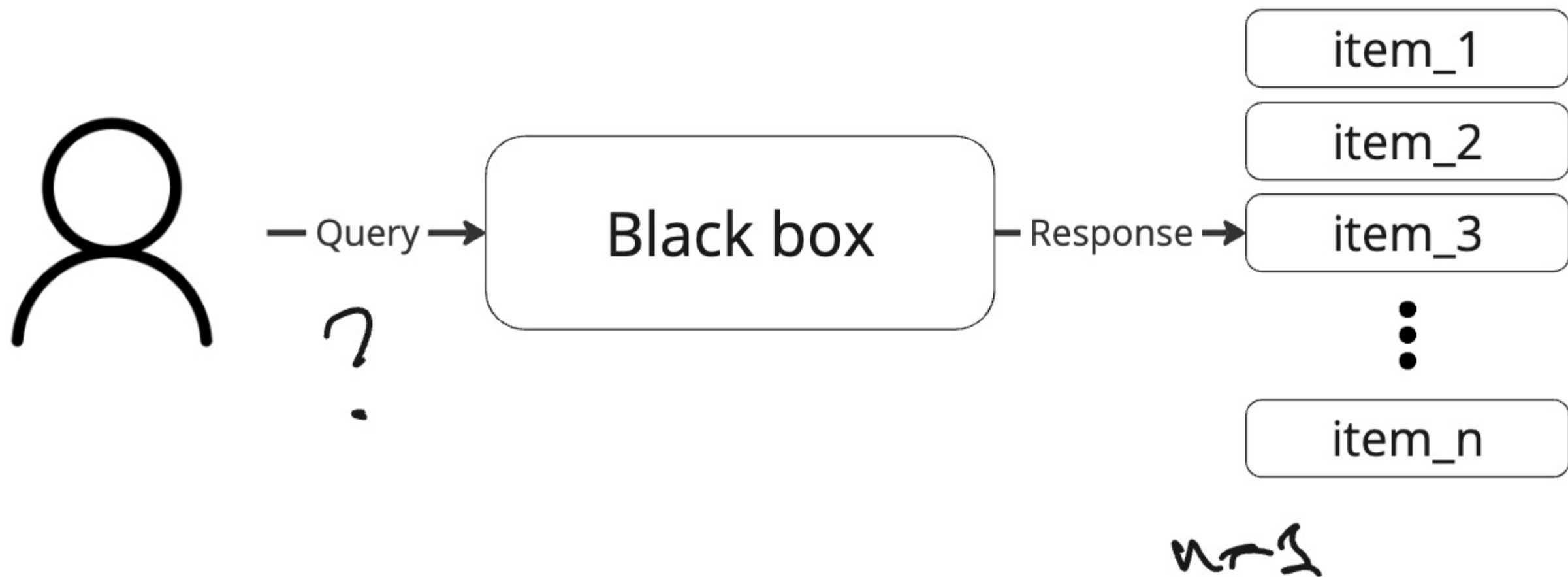
Общая оценка за курс - сумма баллов / кол-во дз

Опционально подумаем о других возможностях добрать баллы

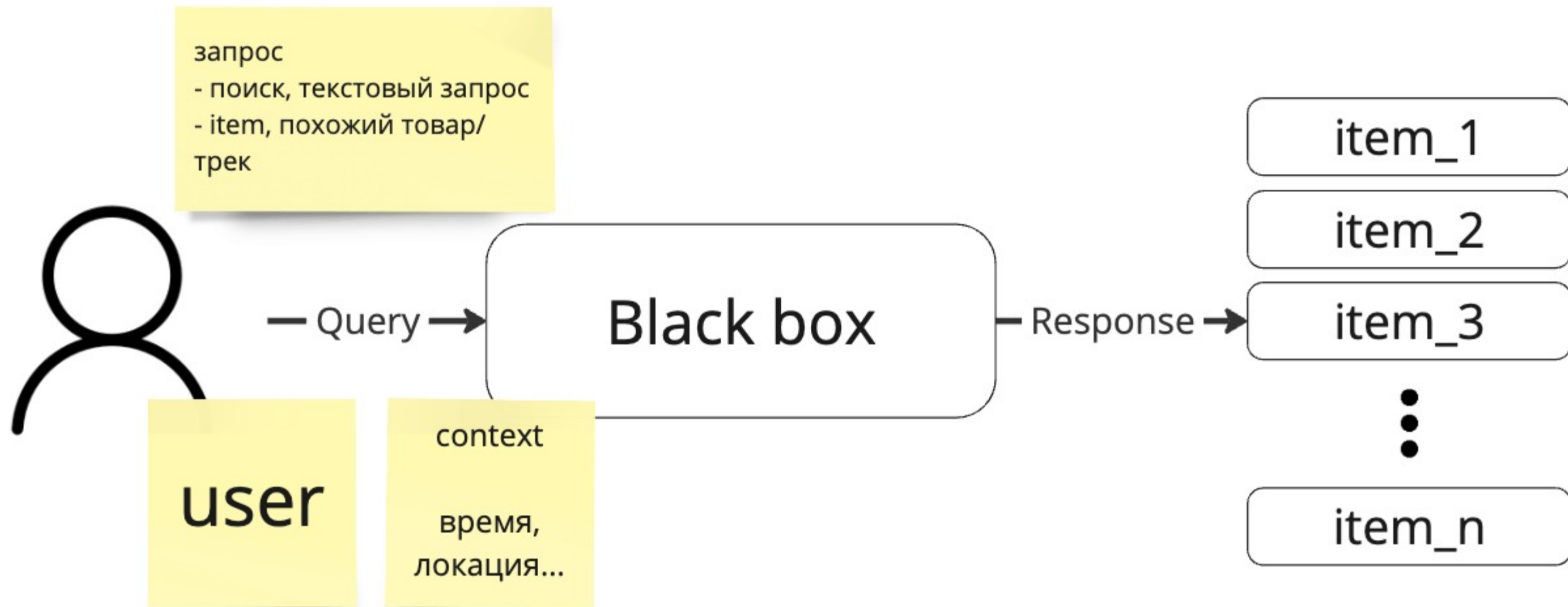
- обзор статей
- небольшой экзамен/тестирование

Вопросы

Контекст



Контекст



Контекст



— Query →

Black box

— Response →

item_1

item_2

item_3

⋮

item_n

отобразить top_n
из всей базы
документов

что
тут?

item=document

RS и IR

Рекомендательные системы (RS) — это программные инструменты и алгоритмы, предназначенные для анализа данных о пользователях, их предпочтениях и поведении, чтобы предлагать наиболее **релевантные** продукты, услуги или контент

item

Информационный поиск (IR) — это процесс поиска, отбора и предоставления **релевантной** информации в ответ на запрос пользователя

?

?

RS и IR, пересечения

- данные *item*
 - текстовые данные, метаданные, историю взаимодействий (просмотры, клики, покупки)
 - неявная (implicit) обратная связь (время просмотра документа/трека)
- ranking
 - LTR
 - формирование финального ответа (разнообразие)
- персонализация
 - учет данных пользователя для формирования ответа
- технические решения
 - обратные индексы (ключ - кандидаты)
 - фича сторы
- ML
 - векторный поиск по эмбедингам
 - схожие идеи в DL подходах (токены текста (IR) - токены действий (RS))
 - объединение контекста в общей задаче
- Метрики ранжирования

RS и IR, пересечения

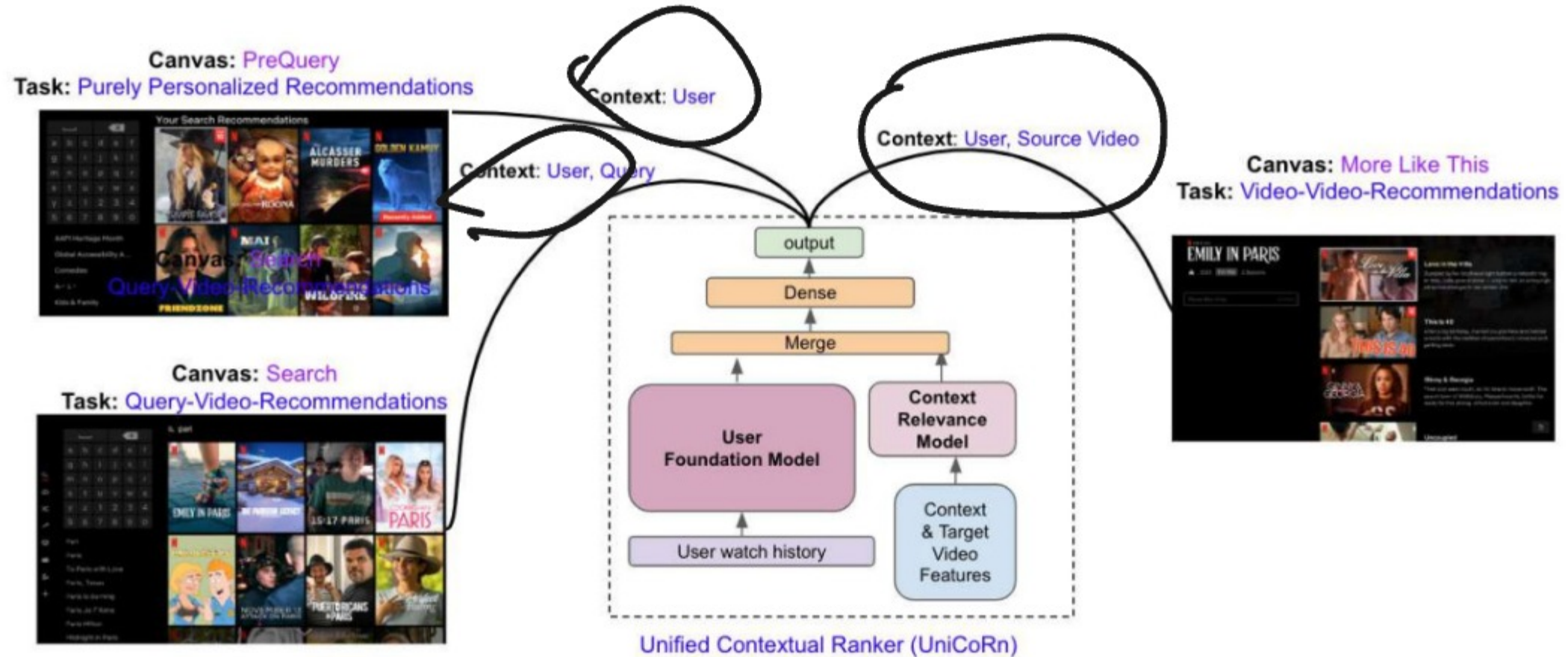
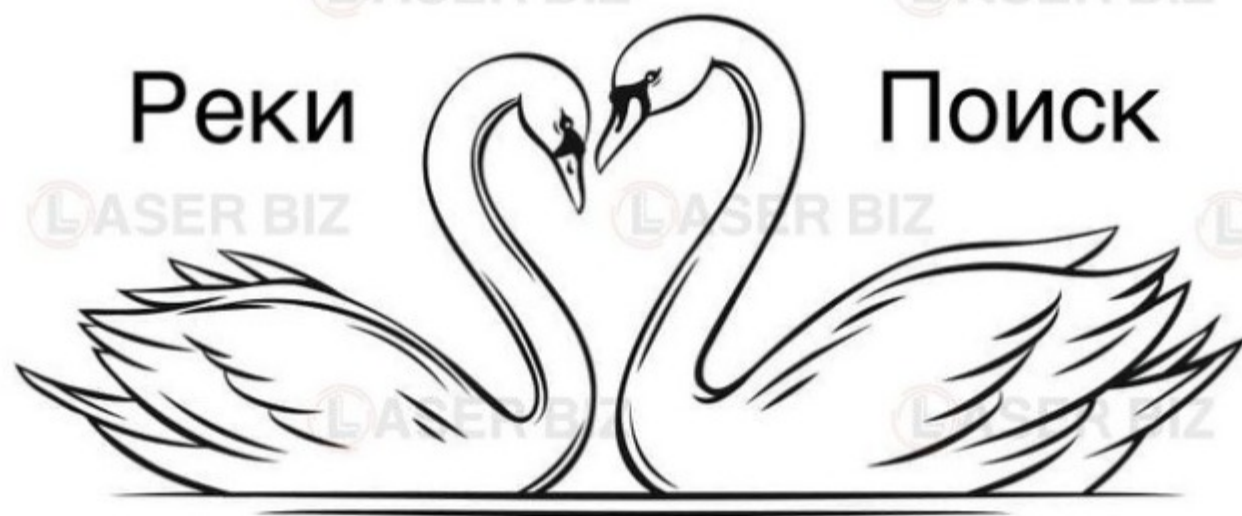


Figure 1: Unified Contextual Ranker (UniCoRn) powering multiple different search and recommendation tasks

RS и IR, отличия

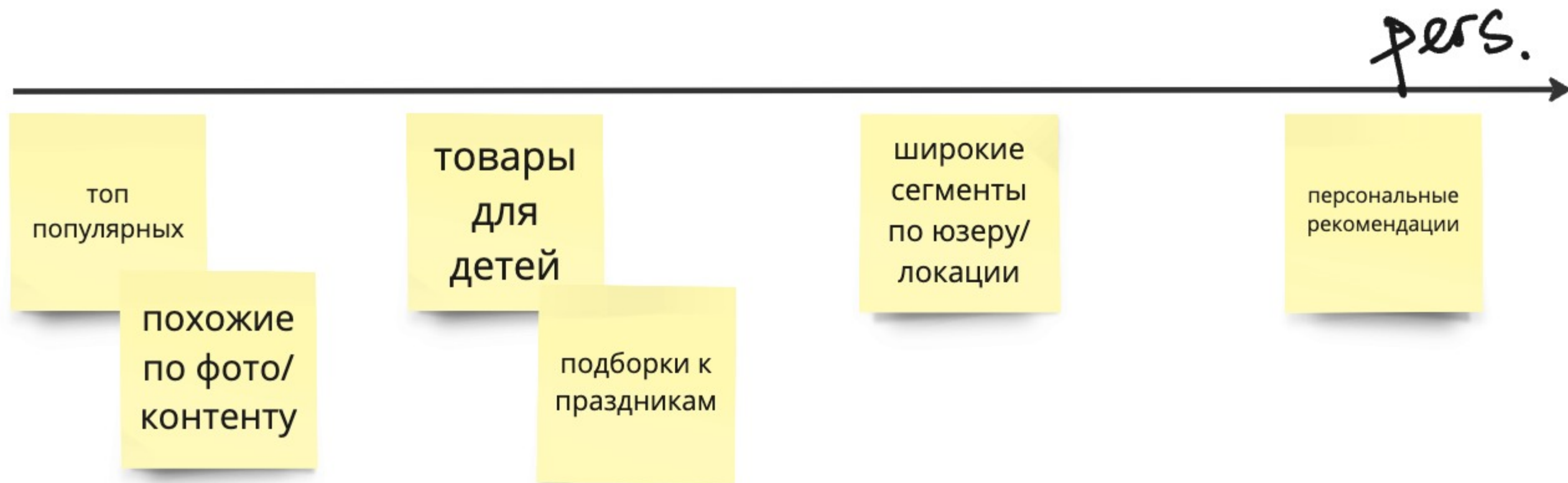
- данные
 - IR - первостепенный акцент на тексте
 - RS - данные о пользователе
- classic ML
 - IR - BM25, TF-IDF и прочее
 - RS - CF, ALS, EASE и прочее
- оценка качества
 - RS - метрики разнообразия (diversity novelty), редко используют оценку ассессоров
 - IR - текстовая релевантности - оценка от ассессоров

Вопросы



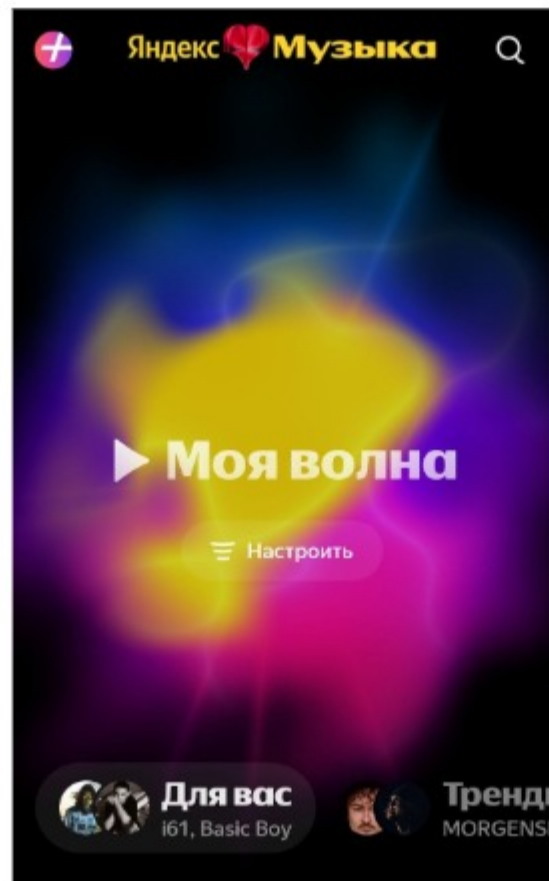
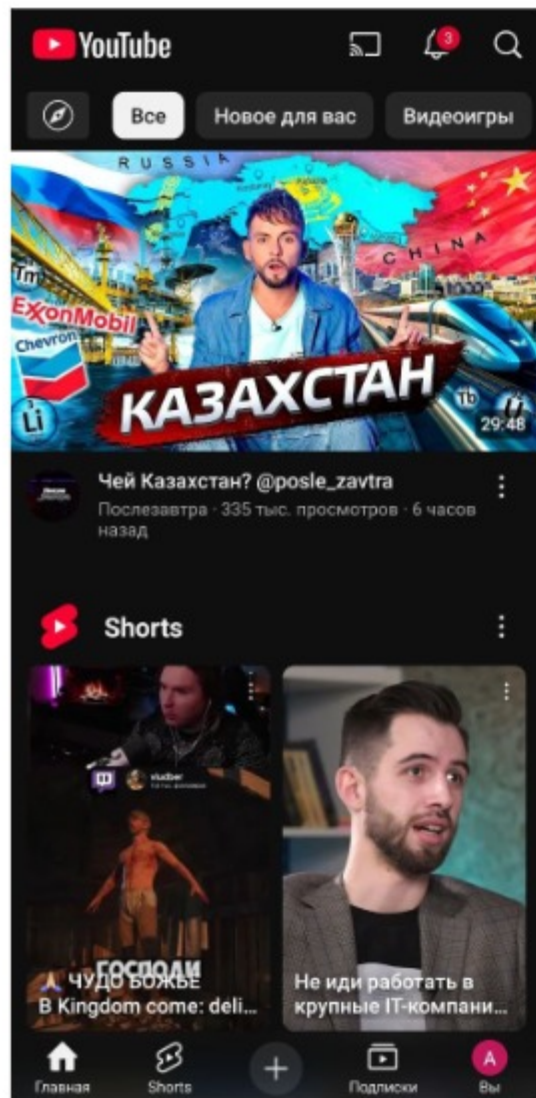
Персонализация

Персонализация — это процесс адаптации контента, услуг или товаров под индивидуальные предпочтения, интересы, поведение и характеристики конкретного пользователя.



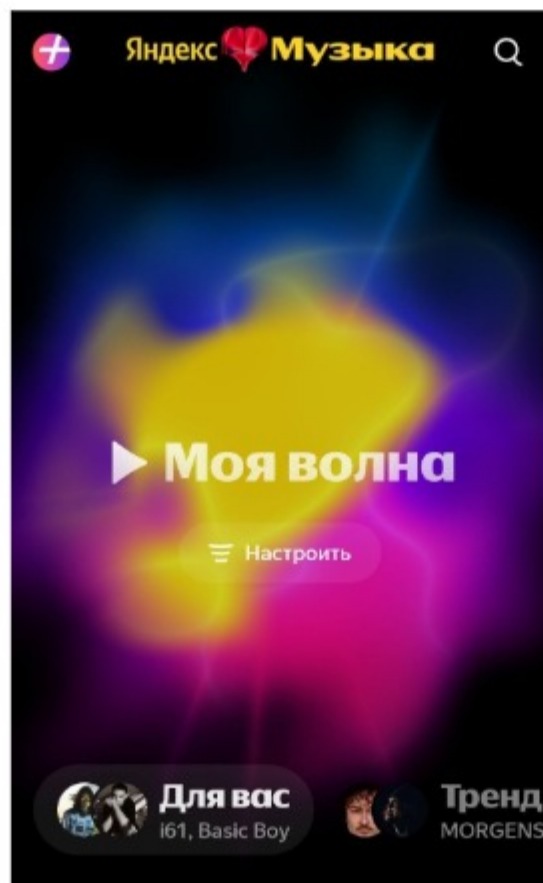
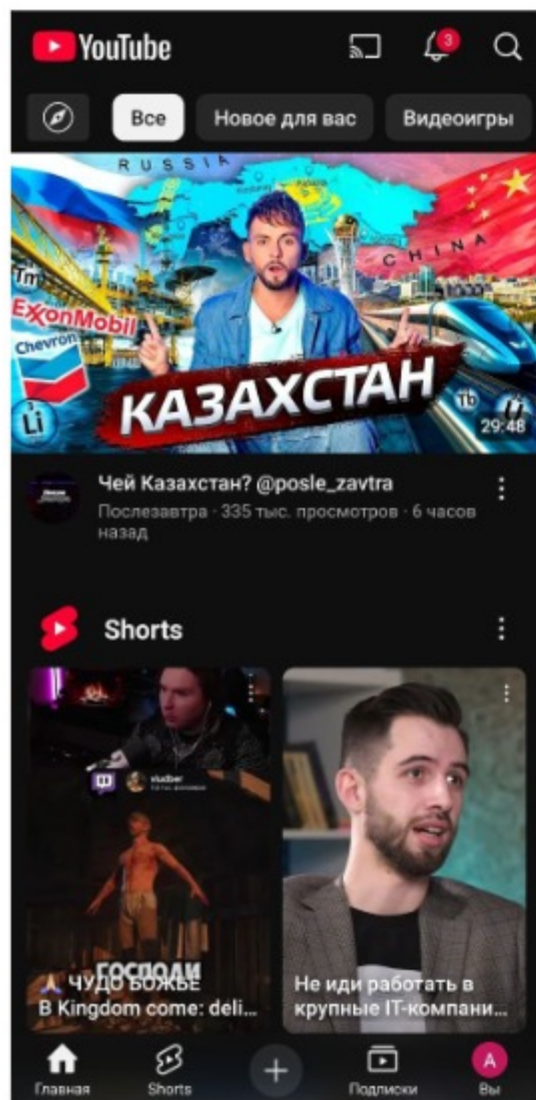
RS примеры

Контекст



зачем
рекомендации
пользователю/
компании?

RS примеры



пользователю

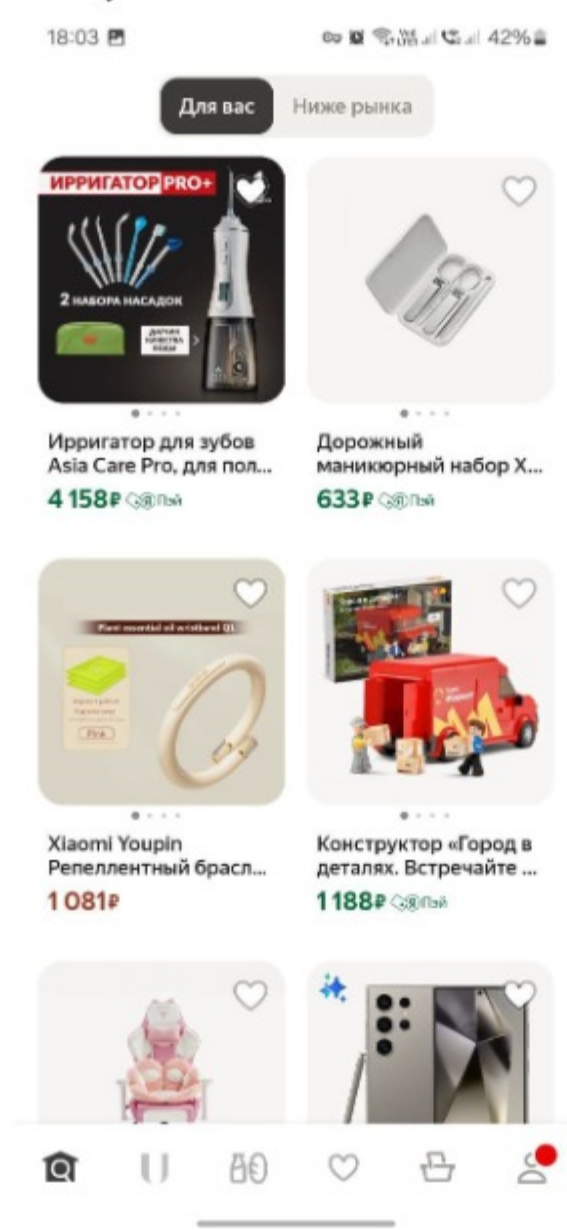
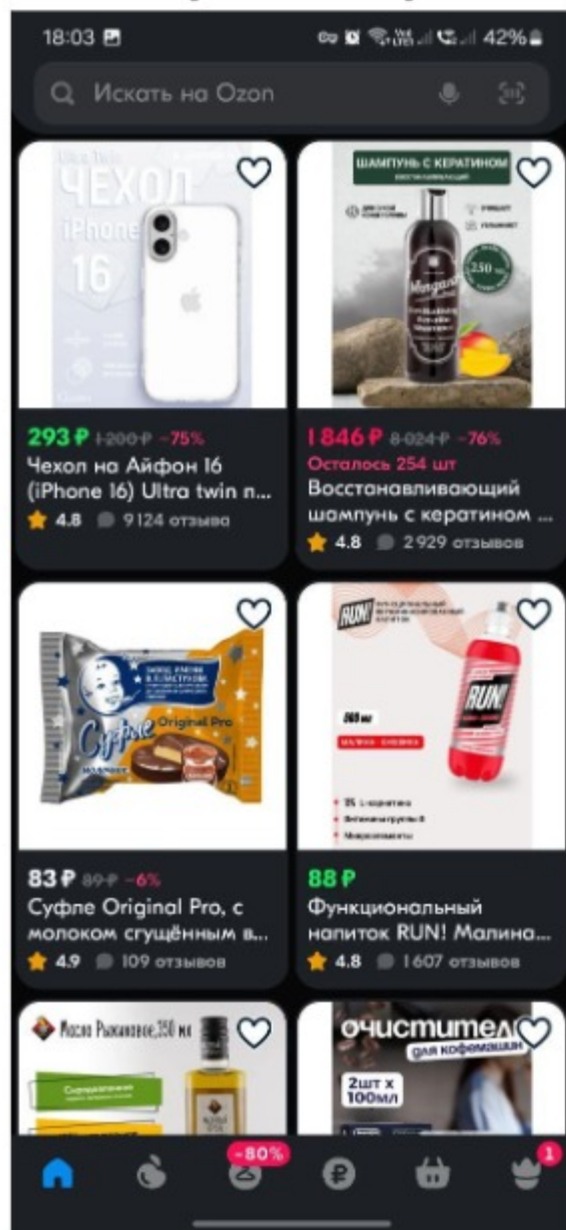
- более релевантный контент
- упрощение поиска контента
- exploration

компания

- увеличение метрик (time spent, adrev)
- увеличение вовлечённости пользователей
- поддержка авторов контента
- увеличение рекламных доходов
- упрощение поиска контента

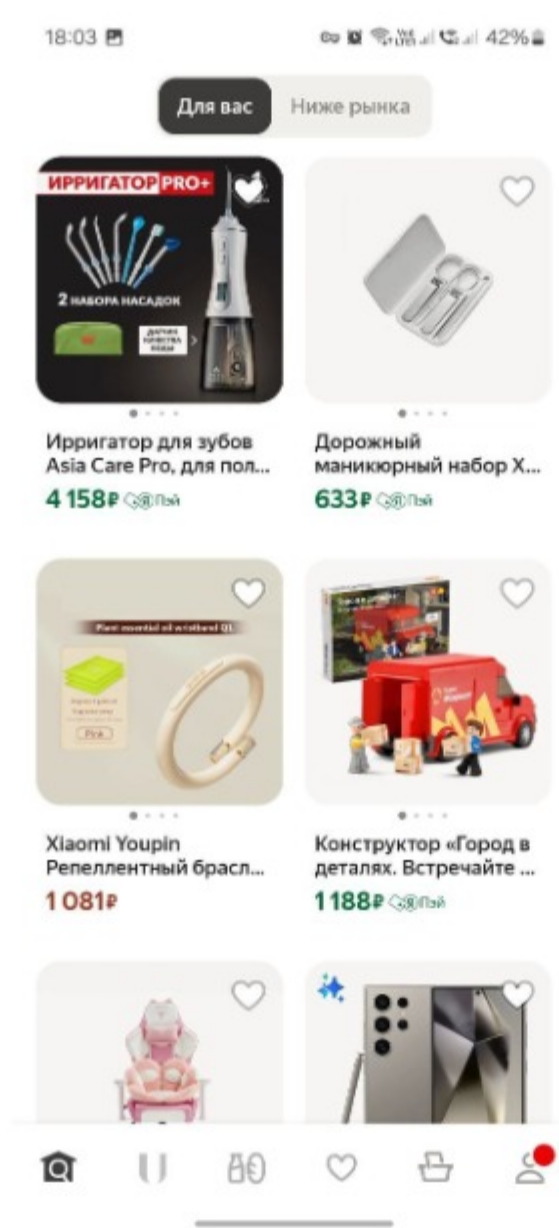
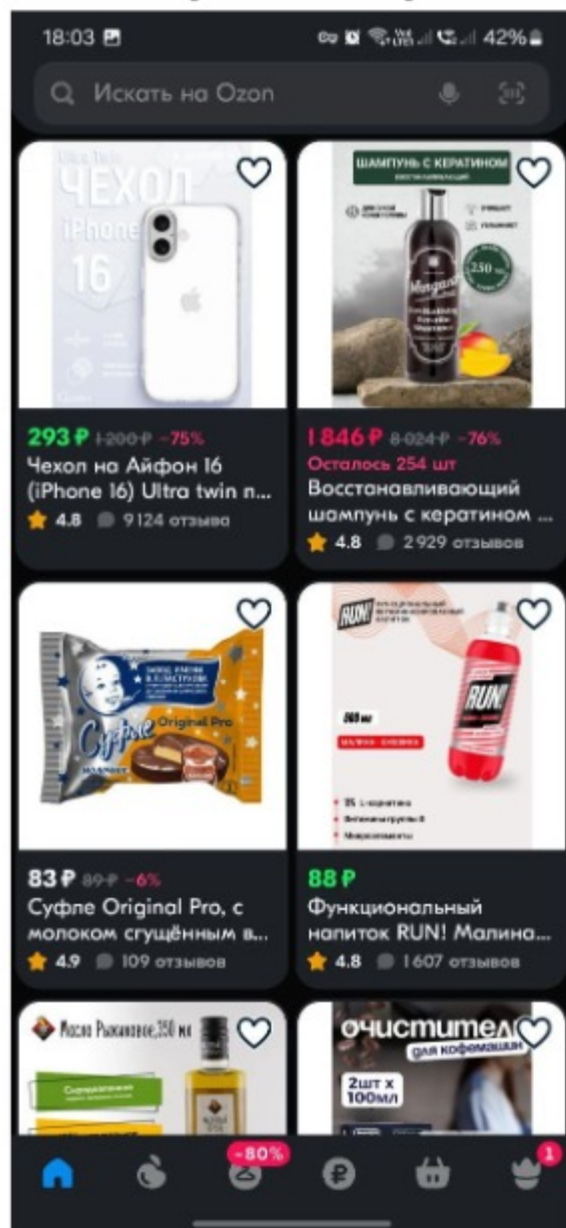
RS примеры

Маркетинг с/с/



зачем
рекомендации
пользователю/
компании?

RS примеры



пользователю

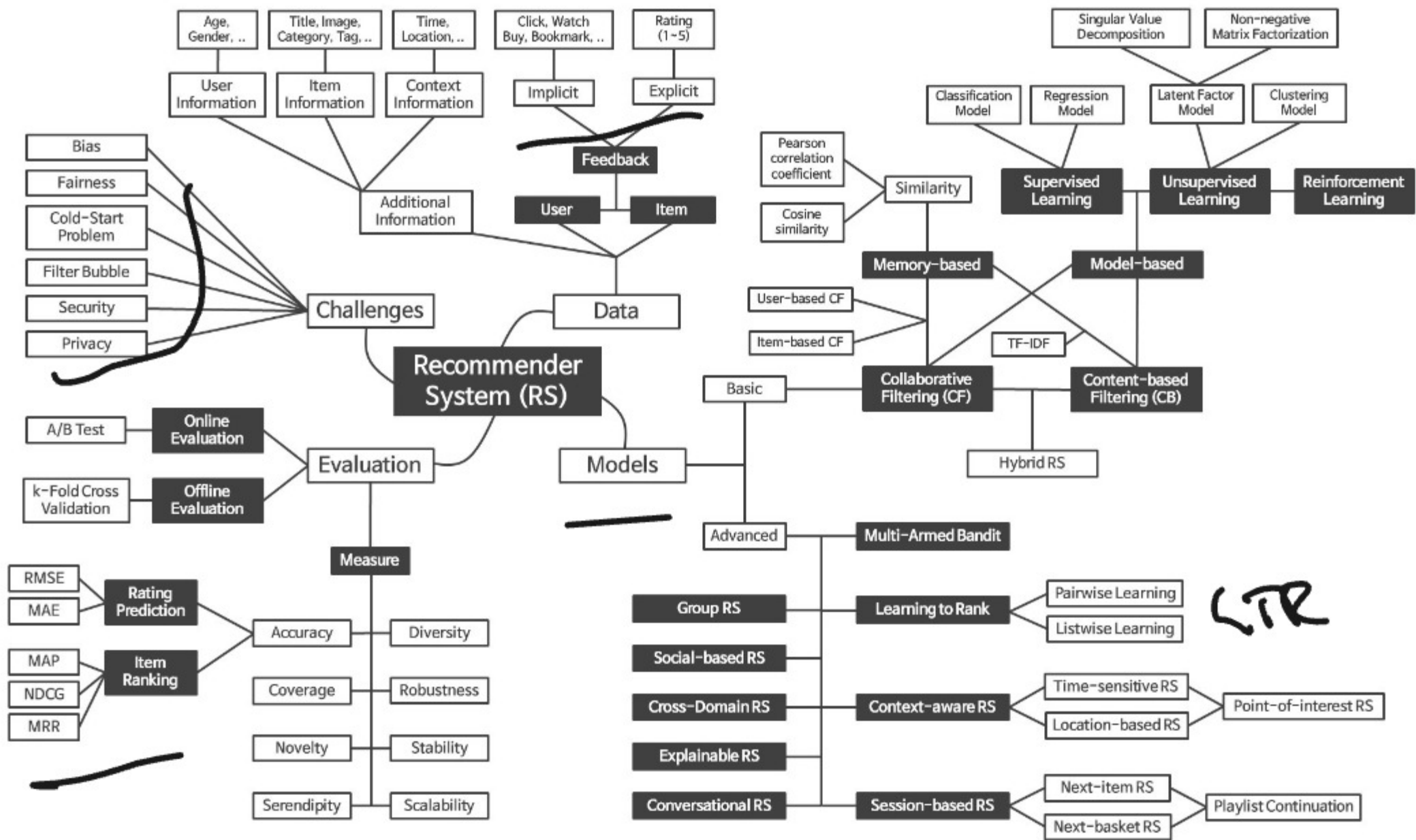
- более релевантные товары
- exploration
- выгодные предложения
- выдача, учитывающие "интересы"

компания

- увеличение метрик (gmv, adrev...)
- поддержка продавцов

RS домены

- Стриминговые платформы (Netflix, Spotify, Youtube, Я.Музыка, VK видео :), ...)
- Электронная коммерция (Ozon, WB, Amazon ...)
- Социальные сети (Instagram, X, LinkedIn ...)
- Новостные агрегаторы и медиа (Google News, Дзен ...)
- Путешествия и туризм (Booking.com, Airbnb...)
- Еда и доставка (Лавка, Uber Eats, ...)
- Игры, Искусство и дизайн, Образование и многое другое



Вопросы

Постановка задачи

$f(u, i)$



хотим упорядочить
базу документов (N)
и взять топ (n)

Релевантность

Для item'ов проставить метку {0, 1}

| | |
|--|--|
| товар на маркетплейсе/доске объявлений | Факт покупки/клика/добавления в корзину/добавления в избранное , $PDP > X \text{ sec}$ |
| контент | Полный просмотр/прослушивание, просмотр $\geq X\%$, лайк, рейтинг $> X$, share, сохранение |

Релевантность

Для item'ов проставить метку из R

| | |
|--|---|
| товар на маркетплейсе/доске объявлений | Взвешенная сумма действий с весами, gmv, оценка от ассессоров 0.1 Clicks 0.9 ORDER |
| контент | процент просмотра, общее время просмотра/ взаимодействия с контентом, оценка от ассессоров |

Метрики, Precision@K

$$\text{Precision at } K = \frac{\text{Number of relevant items in } K}{\text{Total number of items in } K}$$

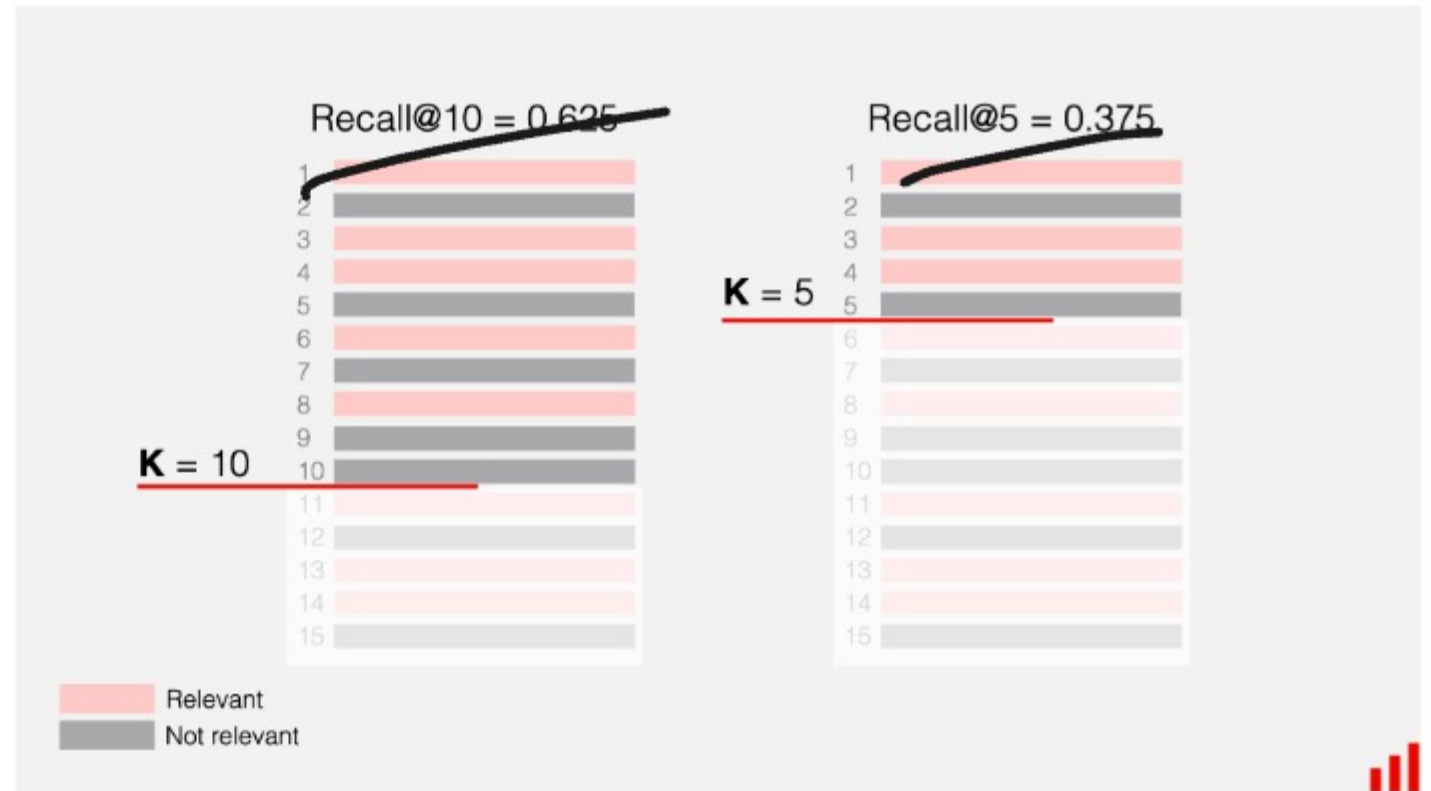


6/10

4/5

Метрики, Recall@K

$$\text{Recall at } K = \frac{\text{Number of relevant items in } K}{\text{Total number of relevant items}}$$



5/8

3/8

Метрики, Precision@K и Recall@K

| | Precision@K | Recall@K |
|--------|--|---|
| ПЛЮСЫ | <ul style="list-style-type: none">• интерпретируемость• легкость вычисления | <ul style="list-style-type: none">• интерпретируемость• используется в задачах с большим количеством релевантных объектов (кандидатогенерация)• легкость вычисления |
| МИНУСЫ | <ul style="list-style-type: none">• не учитывает порядок• есть нюанс с k (об этом на семинаре)• зависит от k | <ul style="list-style-type: none">• не учитывает порядок• зависит от k |

Метрики, AP@K Average Precision

$$AP@K = \frac{1}{N} \sum_{k=1}^K Precision(k) \times rel(k)$$

$N = \#rel$

$rel(k) = 0/1$

| Rank | Item | Precision@k |
|------|--------------|--------------|
| 1 | Relevant | $1/1 = 1$ |
| 2 | Not relevant | $1/2 = 0.5$ |
| 3 | Not relevant | $1/3 = 0.33$ |
| 4 | Relevant | $2/4 = 0.5$ |
| 5 | Relevant | $3/5 = 0.6$ |
| 6 | Not relevant | $3/6 = 0.5$ |

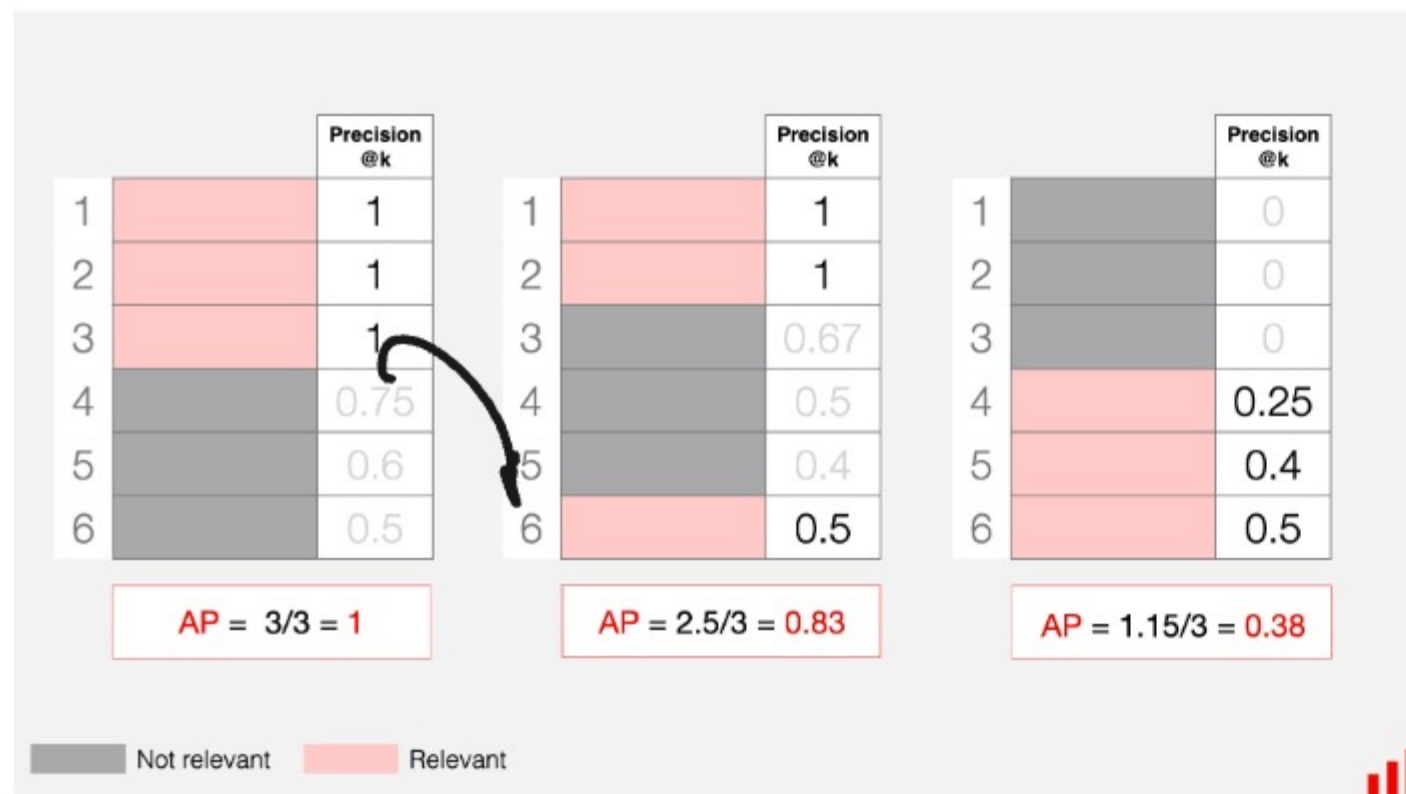
Legend: Not relevant (grey), Relevant (red)

Average precision

$$AP@6 = \frac{1 + 0.5 + 0.6}{3} = 0.7$$

too good

Метрики, MAP@K



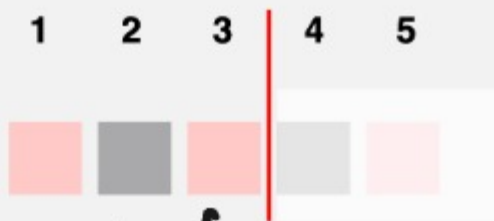
$$MAP@K = \frac{1}{U} \sum_{u=1}^U AP@K_u$$

Метрики, DCG@K

$$DCG@K = \sum_{k=1}^K \frac{rel_i}{\log_2(i+1)}$$

$rel = 0/1$
 $rel \in \mathbb{R}$

K = 3



$DCG@3 = \frac{1}{\log_2(1+1)} + \frac{0}{\log_2(1+2)} + \frac{1}{\log_2(1+3)}$

$DCG@3 = 1 + 0 + 0.5 = 1.5$

Метрики, NDCG@K

K = 3

| 1 | 2 | 3 | 4 | 5 |
|-----|-----|------|------------|------------|
| Red | Red | Gray | Light Gray | Light Pink |

$$DCG@3 = \frac{1}{\log_2(1+1)} + \frac{1}{\log_2(1+2)} + \frac{0}{\log_2(1+3)}$$
$$DCG@3 = 1 + \frac{1}{1.585} + 0 \approx 1.63$$

$$NDCG@K = \frac{DCG@K}{IDCG@K}$$

IDCG

Метрики, MAP@K и NDCG@K

| | MAP@K | NDCG@K |
|--------|--|---|
| ПЛЮСЫ | <ul style="list-style-type: none">• учёт порядка релевантных элементов• формально "баланс" между precision и recall | <ul style="list-style-type: none">• учёт порядка релевантных элементов• учет степени релевантности |
| МИНУСЫ | <ul style="list-style-type: none">• сложность вычисления• зависит от k | <ul style="list-style-type: none">• сложность вычисления• зависит от k |

Метрики, QueryAuc и Hitrate@K

QueryAUC

Classic type

$$\frac{\sum_q \sum_{i,j \in q} \sum I(a_i, a_j) \cdot w_i \cdot w_j}{\sum_q \sum_{i,j \in q} \sum w_i \cdot w_j}$$

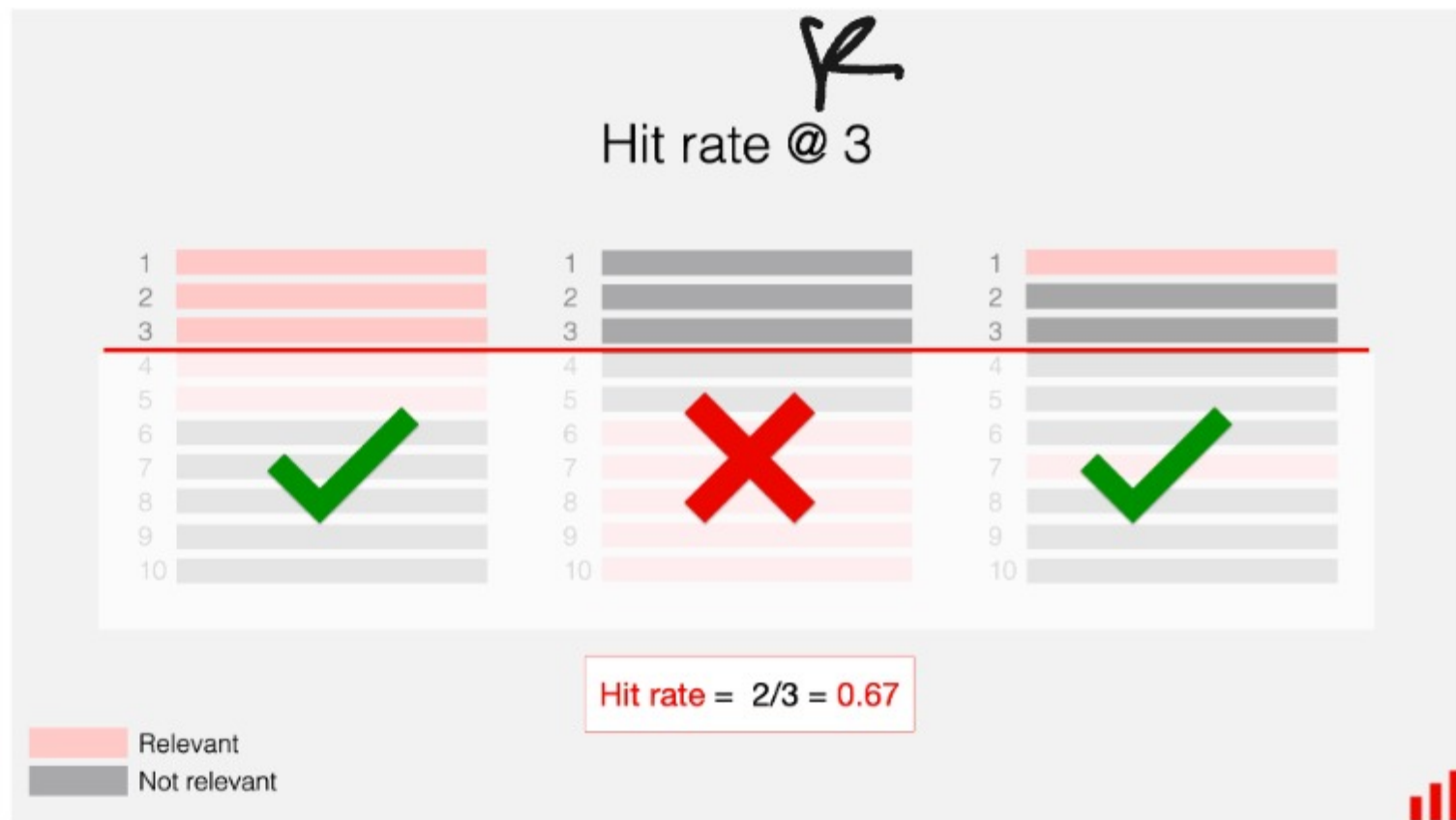
The sum is calculated on all pairs of objects (i, j) such that:

- $t_i = 0$
- $t_j = 1$

$$I(x, y) = \begin{cases} 0, & x < y \\ 0.5, & x = y \\ 1, & x > y \end{cases}$$

$$w_i = 1$$

$u_1 - \epsilon \ni u_2 - \epsilon \ni$
 $w_1 = 1 \text{ AUC}$ $w_2 = 1 \text{ AUC}$



<https://catboost.ai/docs/en/concepts/loss-functions-ranking#QueryAUC>

<https://www.evidentlyai.com/ranking-metrics/evaluating-recommender-systems#hit-rate>

2/3

Recall
NDCG@K

Вопросы

Полезные материалы

Книги

- [Practical Recommender Systems](#)
- [Recommender Systems Handbook](#) (3)
- [Personalized Machine Learning](#)
- [Building Recommendation Systems in Python and JAX](#)

Полезные материалы

Тг каналы

- https://t.me/ods_recommender_systems
- https://t.me/Recsys_IR_Travel
- <https://t.me/inforetriever>
- <https://t.me/WazowskiRecommends>
- https://t.me/knowledge_accumulator

Полезные материалы

Конференции

- ACM Recommender Systems
- ECIR (European Conference on Information Retrieval)
- SIGIR (Conference on Research and Development in Information Retrieval)
- WSDM (ACM International Conference on Web Search and Data Mining)
- WWW (The Web Conference)