

50 оттенков рекомендаций

или Как мы пытались растить выручку через персонализацию

Ибрагим Бадертдинов

О себе

1. 5 лет занимаюсь **NLP** и **RecSys**
2. **Домены:** медицина, финтех, креативные индустрии
3. Знаю **татарский** в совершенстве
4. Однажды **проводил корпоратив** для жен хоккеистов Ак Барса



Как мы делали рекомендации и советы

1. Зачем нам рекомендации?
2. Три версии нашей системы.
 1. Быстрые эвристики
 2. Колаборативная фильтрация
 3. Ранкер и сложности
3. Советы себе в прошлое



Дисклеймер

1. Цель доклада – **сэкономить время**
2. **Серебряной пули нет**
3. Тут про стартапы, поэтому было **быстро и иногда больно**

В презентации будут ссылки



Задача рекомендательной системы – это

для каждого **пользователя**, на основе анализа **его поведения и других признаков** – предложить наиболее релевантные для него (и полезные для продукта) объекты



NETFLIX

В рекомендациях много тумана



1. **Не понятны** “хорошие” метрики
2. Оффлайн тесты **не всегда коррелируют** с онлайн метриками
3. Алгоритмы из научных статей сложно сравнивать из-за разного сетапа
4. **Верхняя граница** влияния рекомендаций на целевую метрику

Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches

Widespread Flaws in Offline Evaluation of Recommender Systems

У нас было:

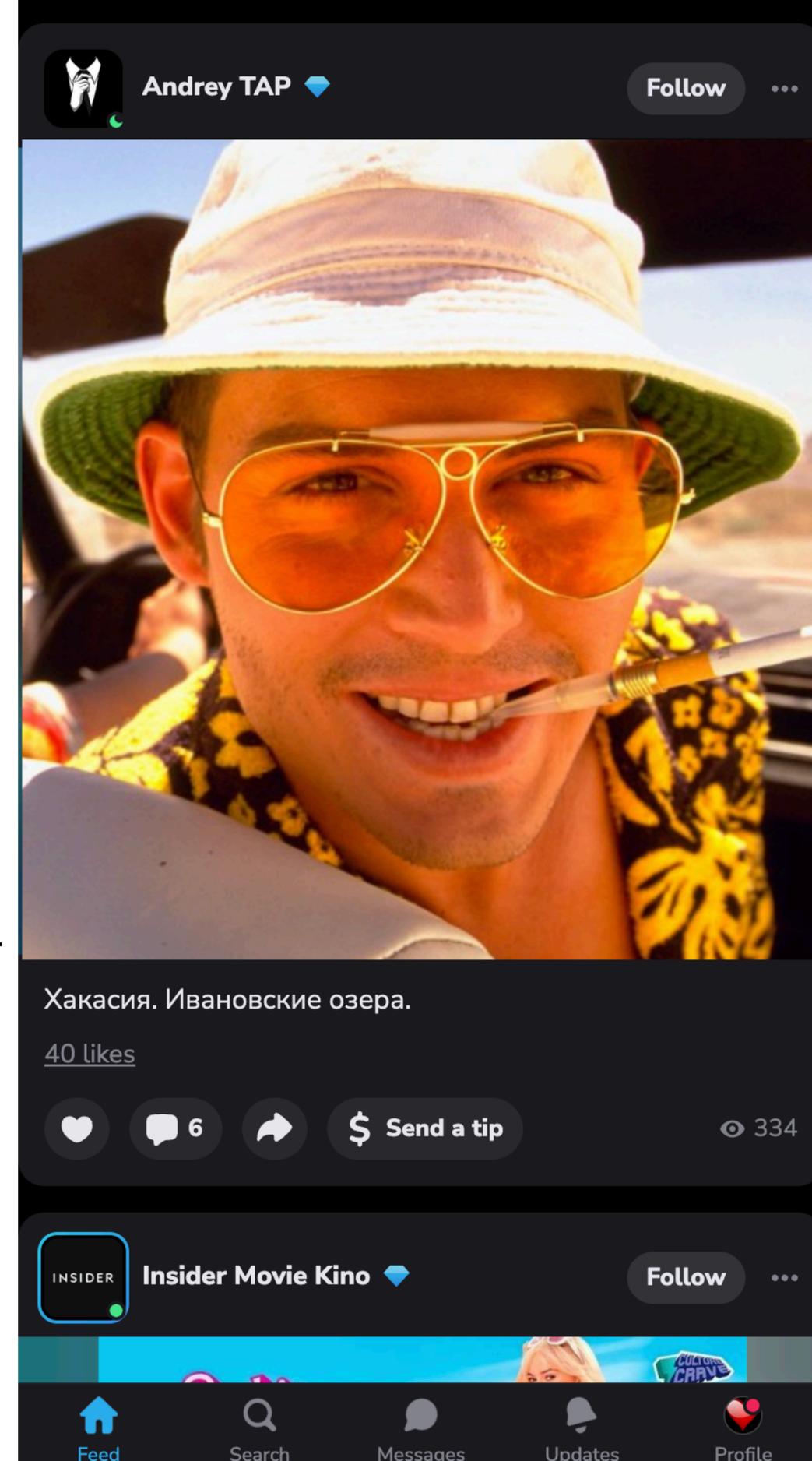
1 платформа для монетизации контента

2 модели монетизации: премиум подписки + комиссия с операций.

10+ механик соц сетей: подписка, лайки, сообщения ...

500 000+ постов в хронологической ленте

300 000+ юзеров, которые создают контент и зарабатывают



Коротко про СТЭК

Go + TypeScript



MySQL / MariaDB + Redis



Elastic



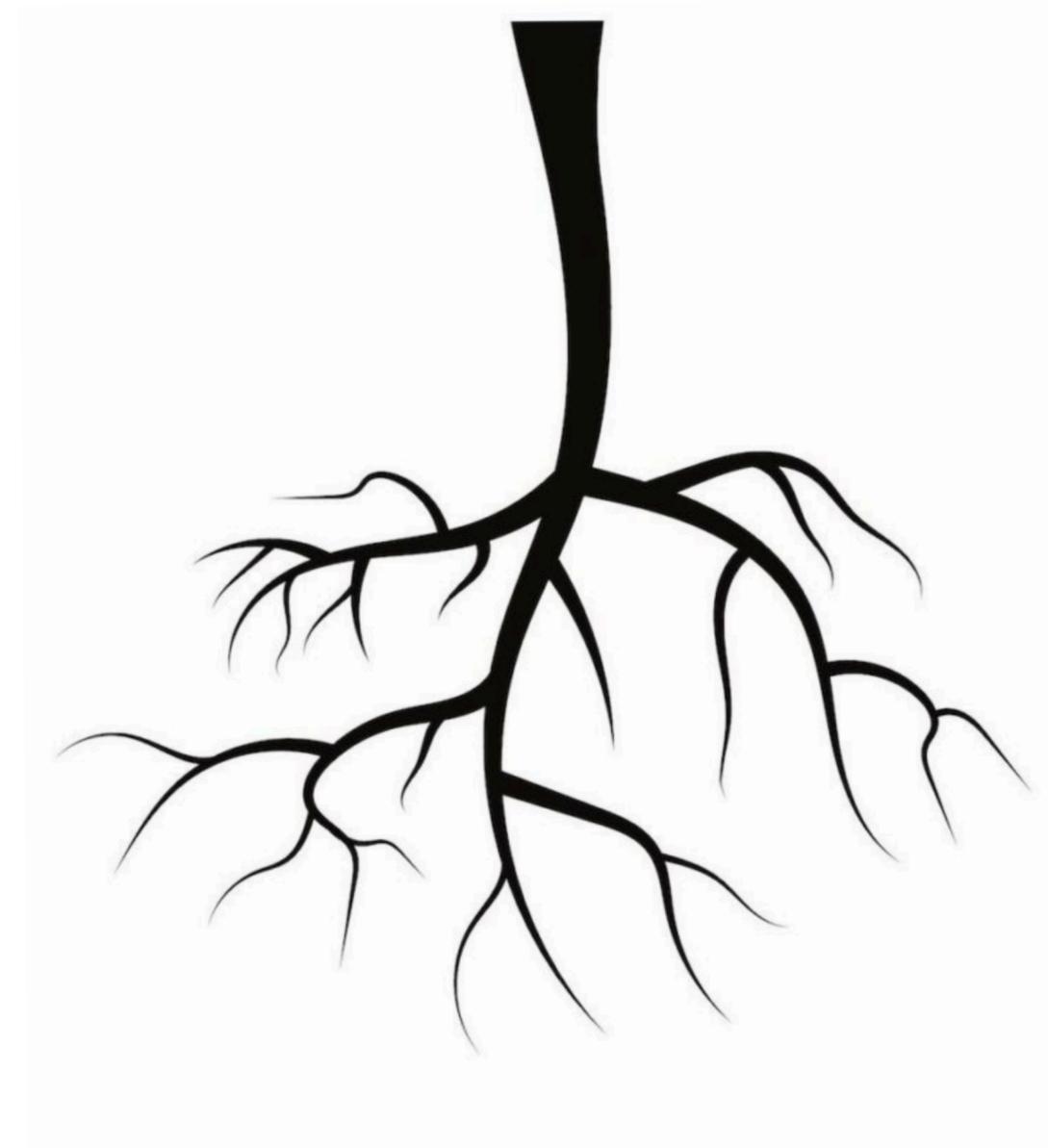
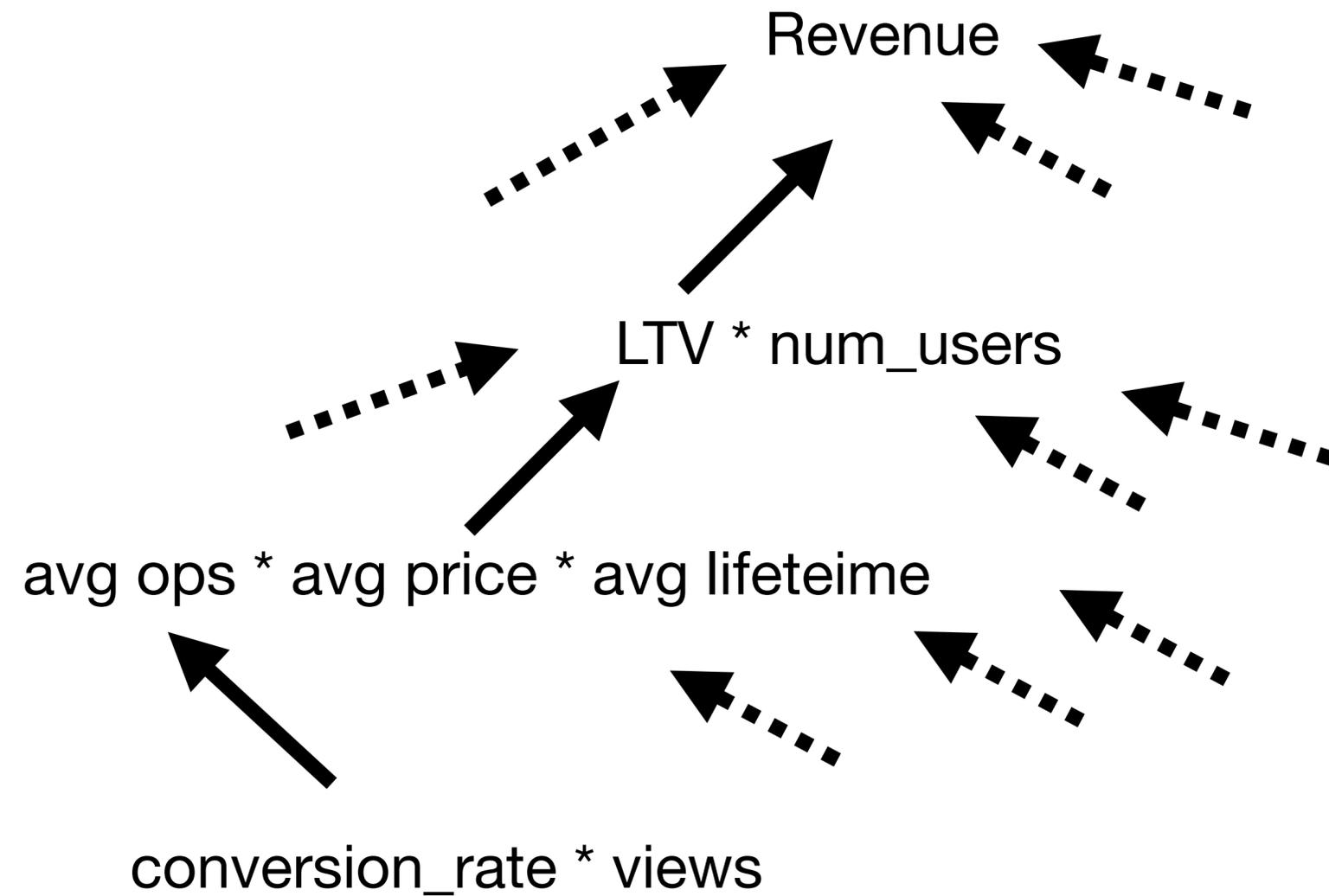
Clickhouse + Grafana



Kubernetes / Drone / Sentry / Amplitude ..

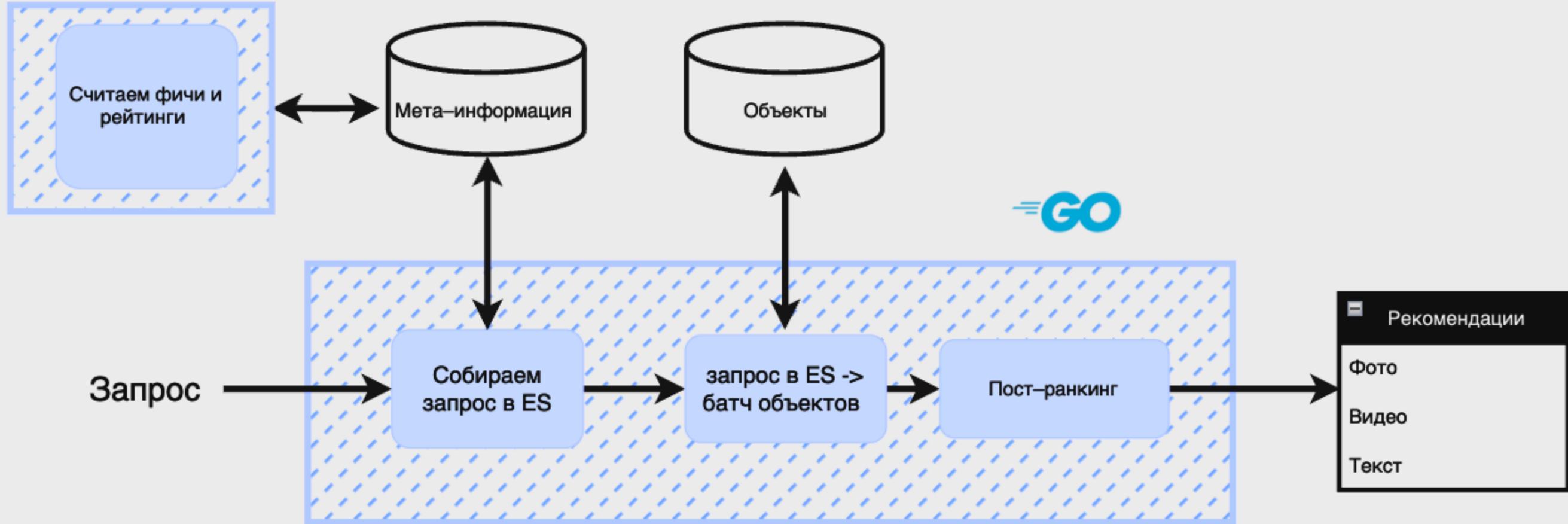


Любая метрика – состоит из других метрик





Первая версия



- must
- must not
- hash-map скоров
- Формула на "painless"

- Авторы подряд
- Тип контента подряд

Как можно считать рейтинг?

1. **Кумулятивный:** рейтинг = (положительные оценки - отрицательные оценки).
2. **Средний:** рейтинг = средняя оценка = положительные оценки / все оценки
3. **Wilson score:** для конверсий, учитываем еще ДИ
4. **Bayesian average:** сводим к среднему, пока мало инфы
5. **Dirichlet distribution sampling:** задаем стартовый рейтинг

<https://jeremykun.com/2017/03/13/bayesian-ranking-for-rated-items/>

<https://medium.com/district-data-labs/computing-a-bayesian-estimate-of-star-rating-means-651496a890ab>

<https://jerryliu.xyz/tech/2018/09/18/how-not-to-sort-by-average-rating.html>

Наш набор эвристик на старте

1. Фильтры по языку / гео (**свой язык + английский** → остальное)
2. **Баны** / Скрытые баны / Пользовательские списки (скрыть)
3. **Затухание** по времени
4. **Штрафы** для тех у кого много трафика и кого уже видел юзер.
5. **Буст** для новых авторов / топовых авторов (продажи)



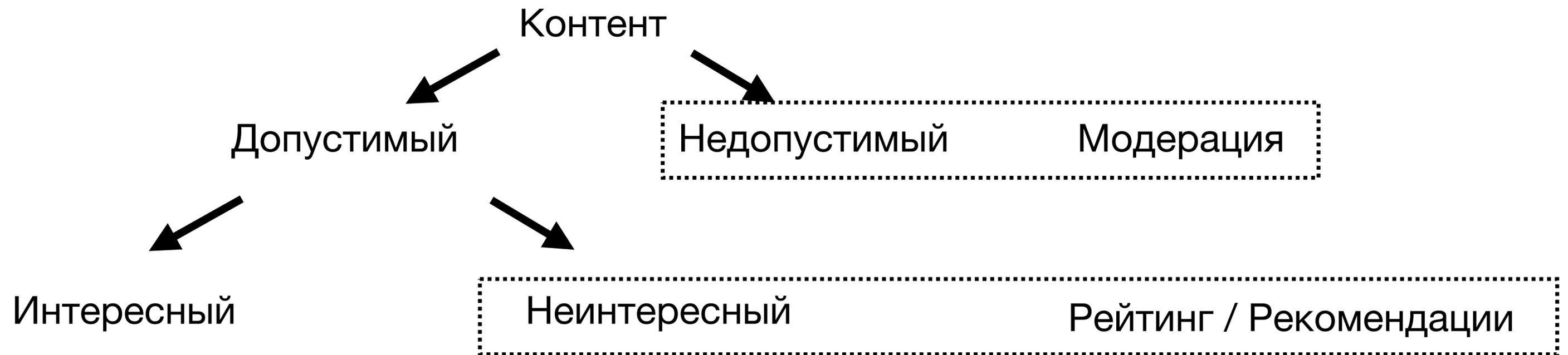
Быстрая, но не персонализированная версия

- + Завели за **2 недели**
- + В логику эластича можно заложить **много эвристик**
- + **Адекватная лента** благодаря популярному

- **нет персонализации**
- эвристики нужно подбирать **вручную**
- непонятно как **тестировать**

views per session: **+ 24%**
like-to-view rate: **+10%**

Про недопустимый и неинтересный контент

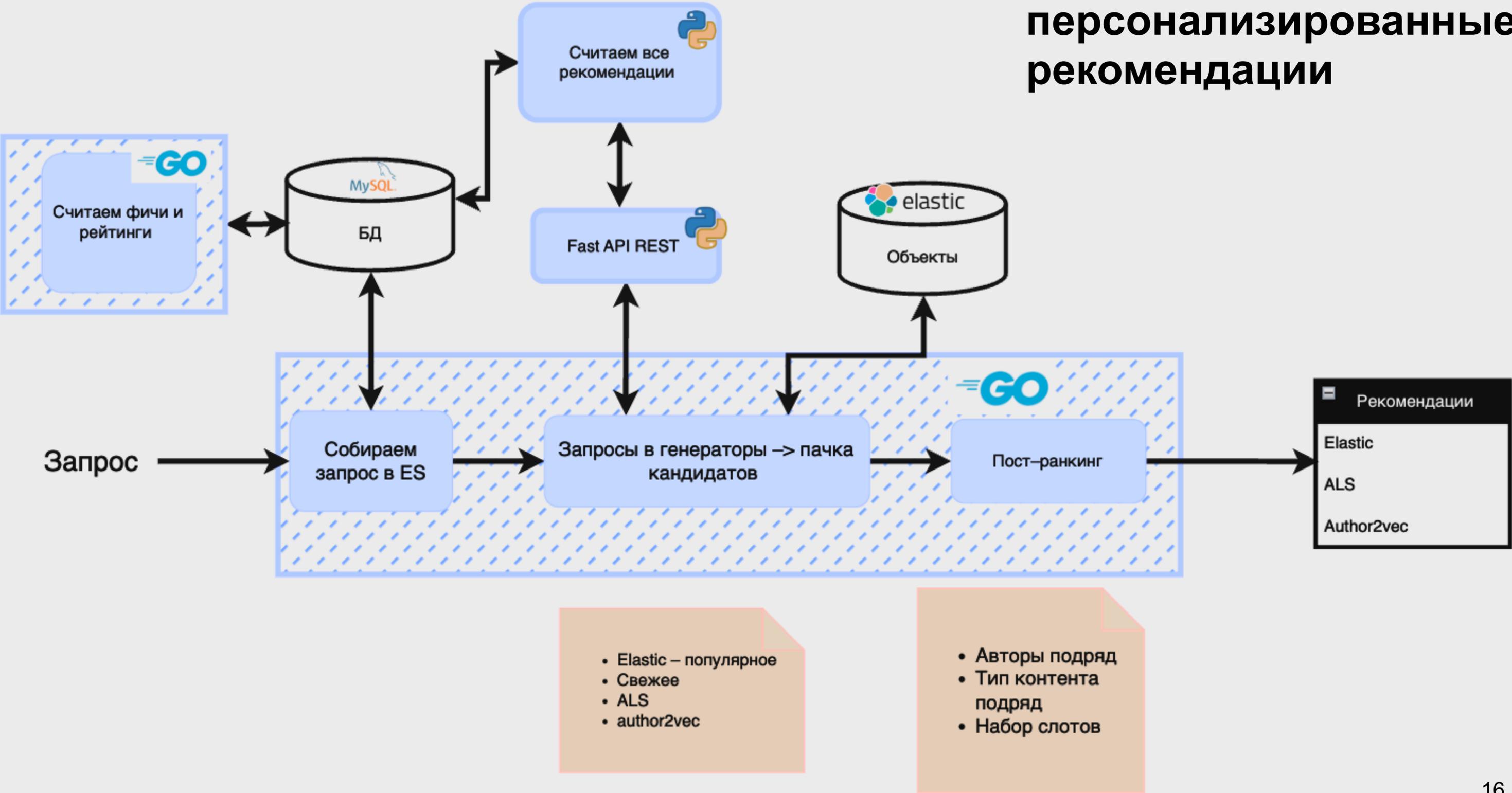


Пивотнулись в площадку для создателей контента

1. Контент можно публиковать **после верификации**
2. Стало **больше инструментов для монетизации**
3. Решили проблему с **качеством контента**
4. И авторов, и юзеров стало меньше.



Вторая версия – персонализированные рекомендации



Особенности КФ

Что брать как **взаимодействие**?

Насколько **лайки** реально “лайки”?

Фильтрация **взаимодействий** (боты → шум)

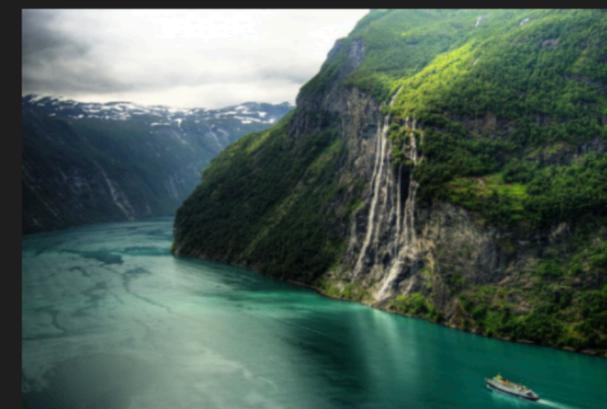
Item2item / user2item – это разные подходы

Семантические эмбединги из взаимодействий



189385

Бразилия...



158902

#природа #фото #реки
#пейзажи #топ



154490

Какое замечательное кафе
обустроенное прямо на

Плюсы и минусы второй версии

- + есть персонализация
- + не стало дольше (просто горютины и блендер)
- + больше разнообразие
- + нивелируются недостатки отдельных подходов

- в КФ нет учета времени
- **некачественные** рекомендации, когда мало взаимодействий (**user-item подход**)
- непонятно как соединять и что работает лучше

Avg-session-time: **+12%**
follow-to-view authors: **+17%**

Как тестировать и понять, что рекомендации не

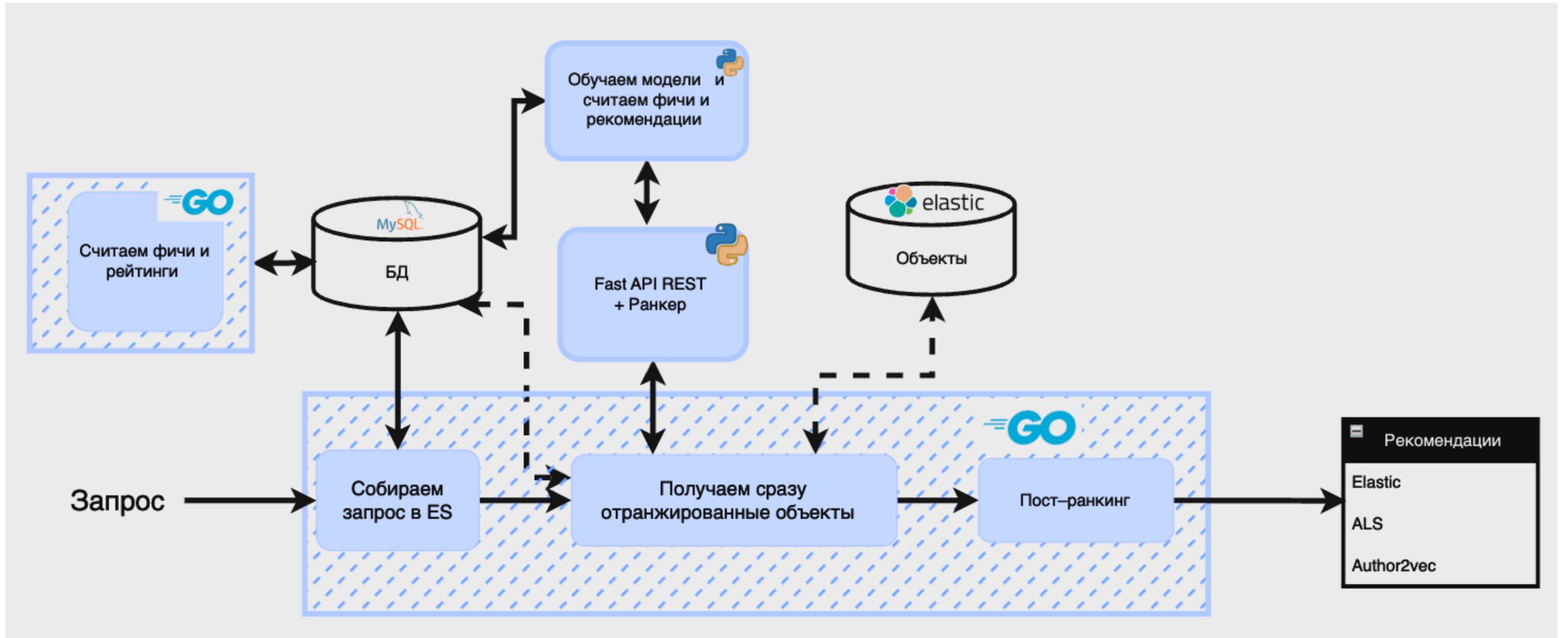
Оффлайн:

- **Метрики** Information Retrieval + value
- **Кросс–валидация**
- **Прокси метрики** (разнообразие, новизна итп)
- **Синтетическая история** и проверка на адекватность
- **Близкие ли item-вектора** внутри одной тематики?

Онлайн:

- **Смотреть глазами**
- **Прокси–метрики**
- **Логирование** (статистика рекомендаций и фичей)
- **АВ тесты**
- **Возможность** смотреть рекомендации на всех
- **Стоп-краны** и рычаги (админка)
- **Выводить доп инфу** в админских аккаунтах

Третья версия – добавим ранкер



Плюсы и минусы ранкера

- + еще больше персонализации
- + выше качество
- + можно учить на разные таргеты
- + можно посмотреть важность фичей

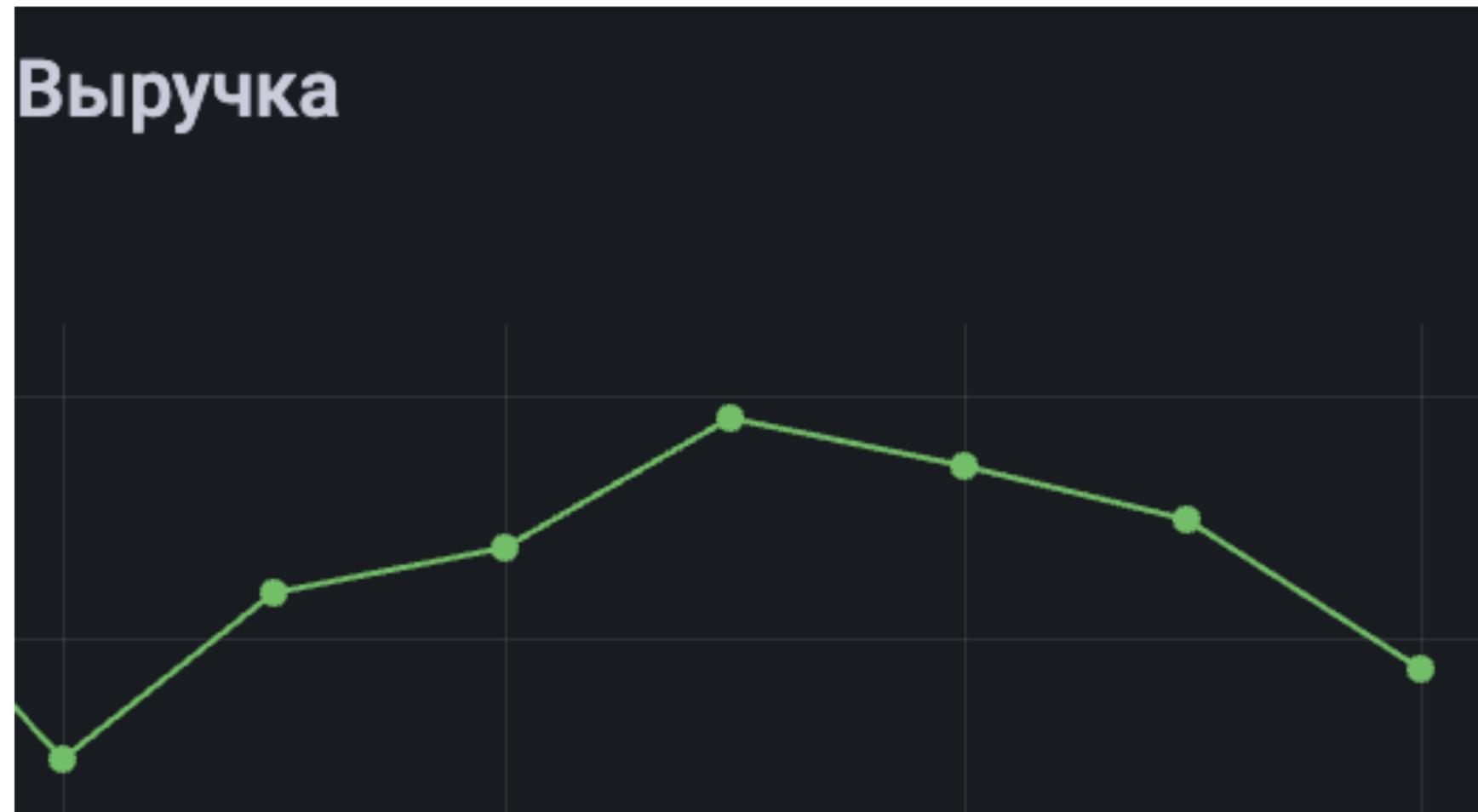
- долго считать фичи
- **однообразный** топ (по типу контента, по авторам)

Avg-session-time: **+55%**
skip-ratio **-30%**

Как росла наша выручка



Как росла наша выручка



Как росла наша выручка

Выручка

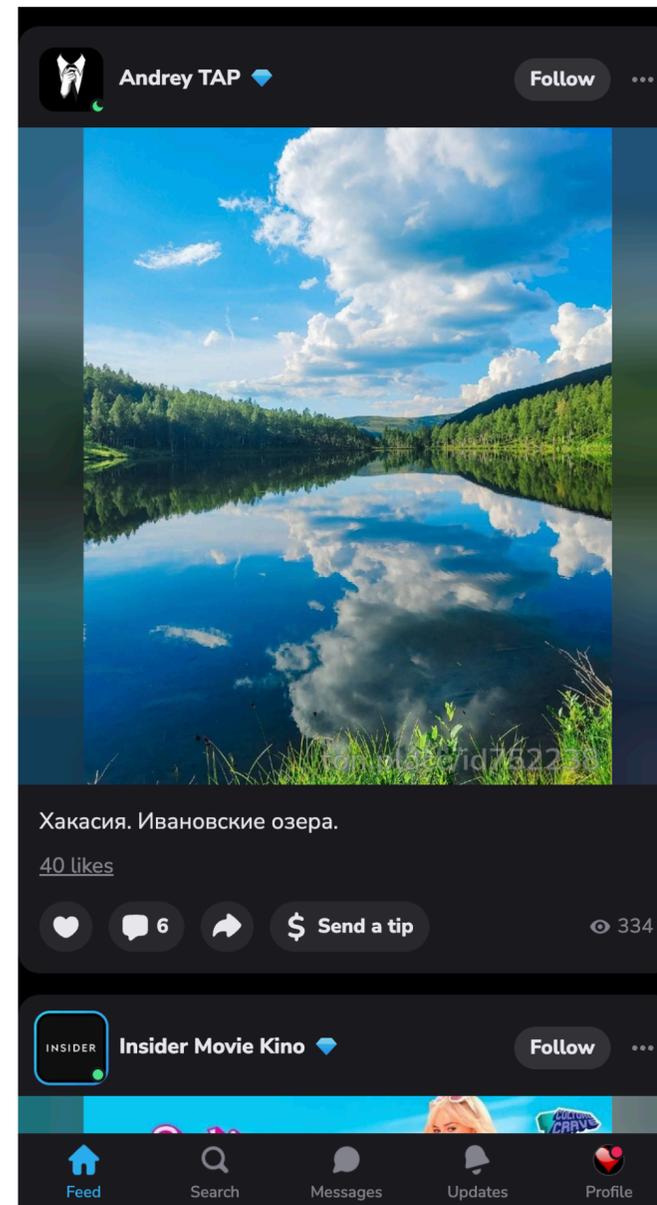


Разделили ленту на две

Лента
Оптимизируем
конверсию воронки

**Рекомендации
авторов** →
рекомендации
постов автора

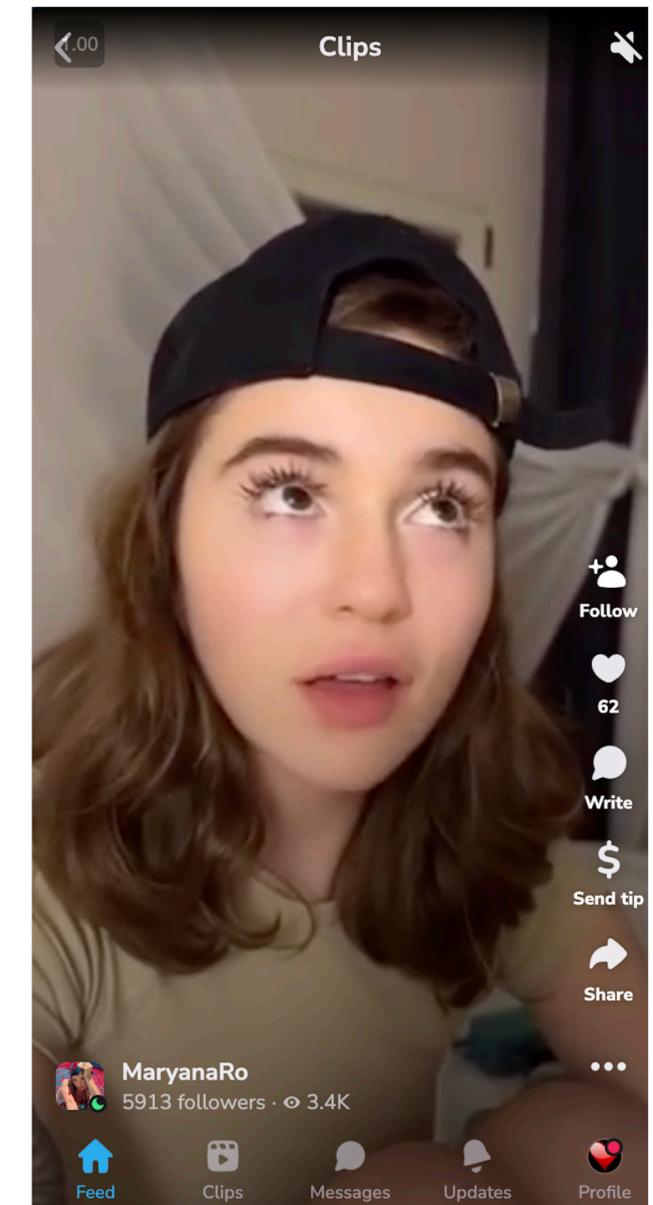
Учитываем **разные
факторы**



Короткие видео
оптимизируем
долгий просмотр

**Рекомендации
контента**

Почти нет
дополнительных
факторов



Топ 5 советов себе

- **Кради как художник:** у похожих больших продуктов (механики, интерфейсы, экраны)
- **Альтернатива ML**, это не только эвристики, но и **новые фичи** и разделы
- Важен **таргет** и его связь с **метриками**
- **Item2Item** – сила, когда >1 разных вариантов
- Узнавать, что все сломалось **из алертов**, а не от юзеров



https://t.me/profunctor_io/9009

Спасибо!

Ибрагим Бадертдинов

 [t.me/ ibragim_bad](https://t.me/ibragim_bad)



<https://t.me/c0mmit>