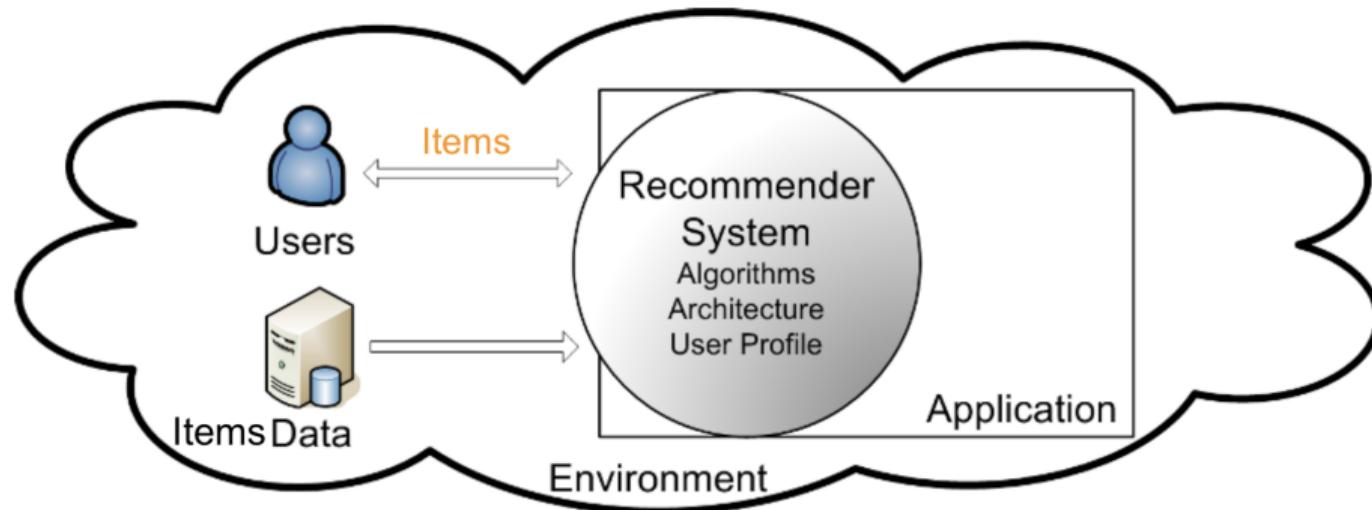


# Рекомендательные сервисы в продакшене

Николай Анохин

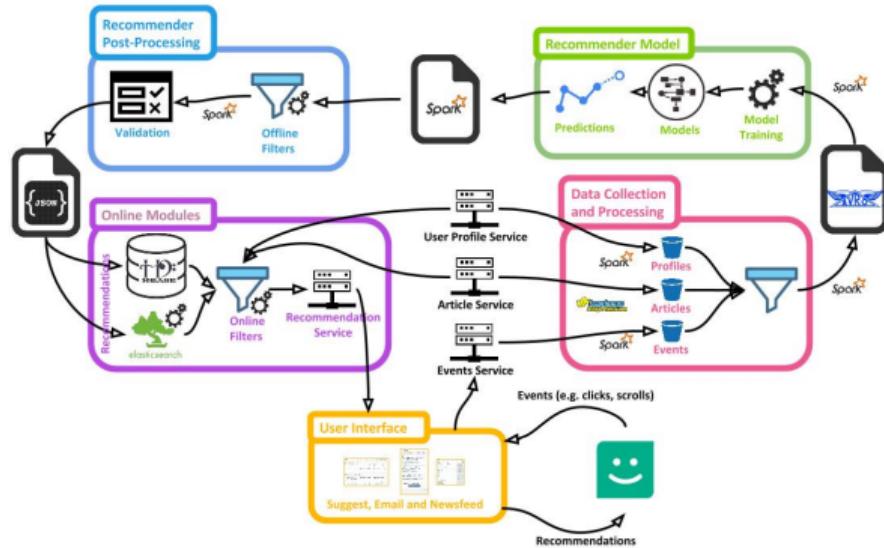
13 февраля 2022 г.

Входной опрос: <https://tinyurl.com/yu4ytbfd>



# Архитектуры рекомендательных сервисов

# Обзор типичных компонентов RS / Mendeley (2016) [JIH16]

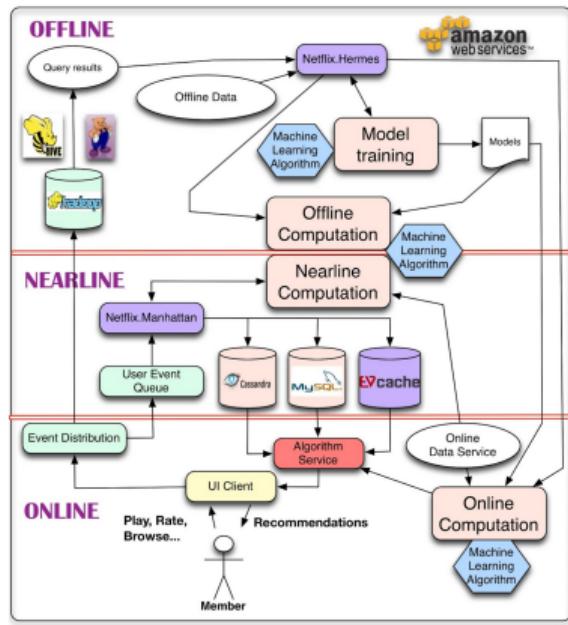


1

Машинное обучение – небольшая часть рекомендательного сервиса. Другие компоненты часто требуют не меньше усилий.

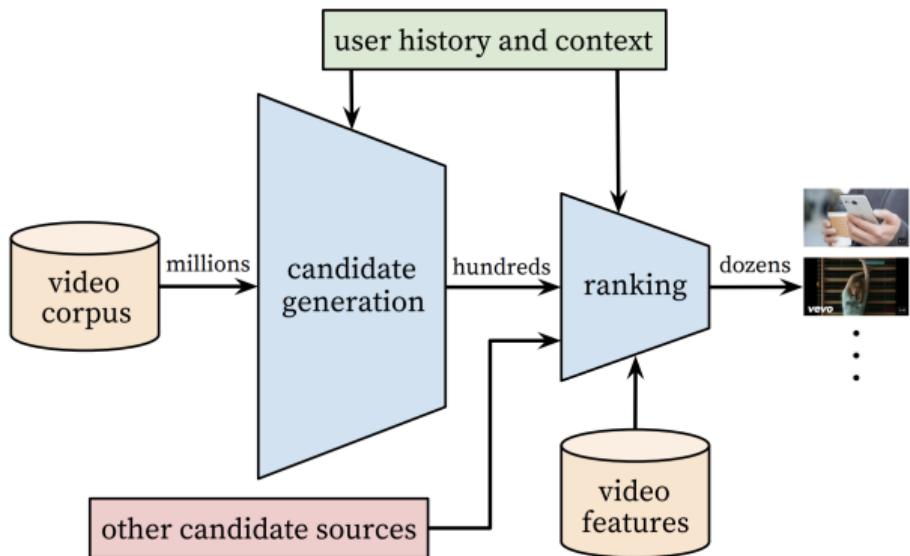
<sup>1</sup>Ссылки: Hadoop + Spark + HBase + Elasticsearch

# Обработка данных под высокой нагрузкой / Netflix (2013) [NN13]



Двигаясь от offline к real-time, мы можем быстрее реагировать на изменения контекста. При этом возникают ограничения на сложность алгоритмов.

## Рекомендации айтемов из больших каталогов / Youtube (2016) [CAS16]



Айтемов так много, что учесть **полный контекст** не может даже Google. Для быстрого отбора кандидатов применяются грубые фильтры.

## Загадка

Что общего между

- населением городов
- количеством друзей у пользователей в социальной сети
- размерами лесных массивов
- количеством прослушиваний песен в Spotify

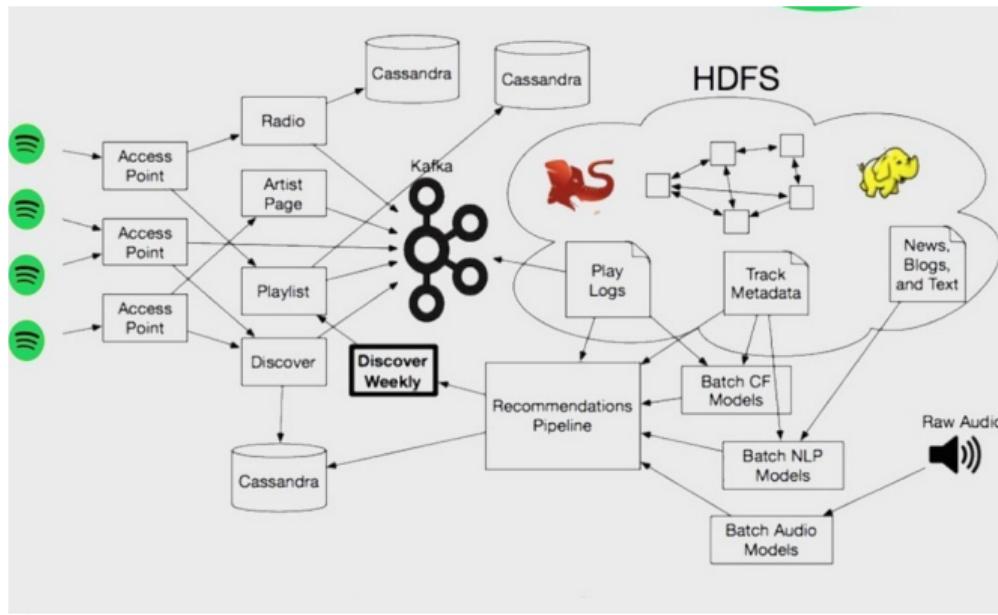
## Power law

$$p(x) = \frac{C}{x^\alpha}, \quad x > x_{min}$$



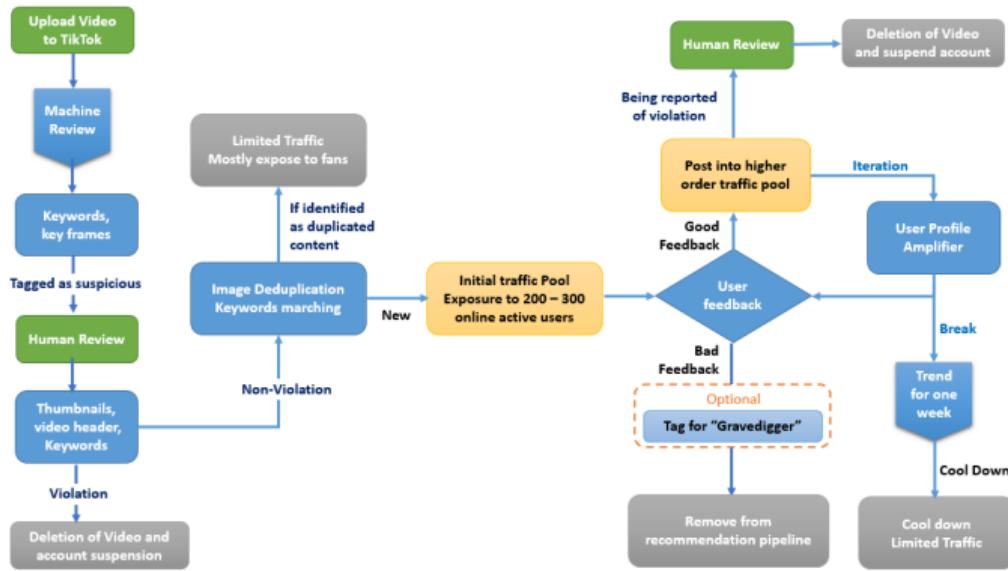
Правило 80/20

# Холодный старт и длинный хвост / Spotify (2016) [Spo16]



Холодные айтемы и пользователи будут всегда. Использование контента - один из вариантов решения проблемы

# Несоответствие таргетов моделей и business-value / TikTok (2020) [Wan20]



Потребности людей нельзя упаковать в удобную метрику. Кроме машинного обучения в рекомендательных сервисах приходится использовать пре- и пост-процессинг, чтобы гарантировать business-value.

## Как в действительности выглядит архитектура RS



## Какие сложности учитывает архитектура RS

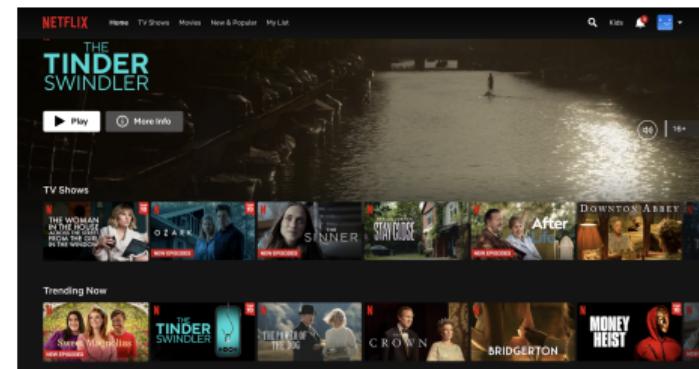
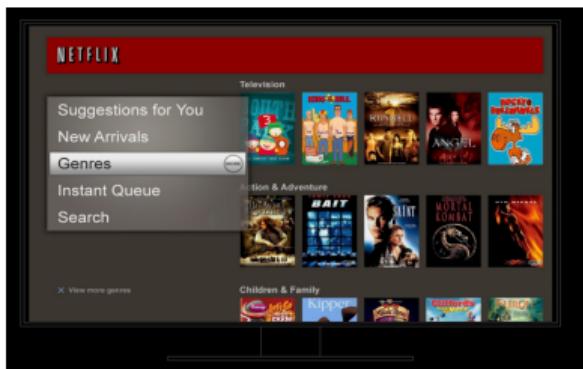
- Высокая нагрузка рекомендательных сервисов
- Большие каталоги айтемов
- Холодный старт пользователей и айтемов
- Несоответствие business-value и метрик оптимизации

## Какие технические средства могут понадобиться

- Отказоустойчивые продакшен-сервисы (HBase, Cassandra, Elasticsearch)
- Передача данных (kafka)
- Хранение данных (Hadoop HDFS)
- Batch обработка данных (Spark)
- Потоковая обработка данных (Kafka, Spark Streaming)

## Метрики и эксперименты

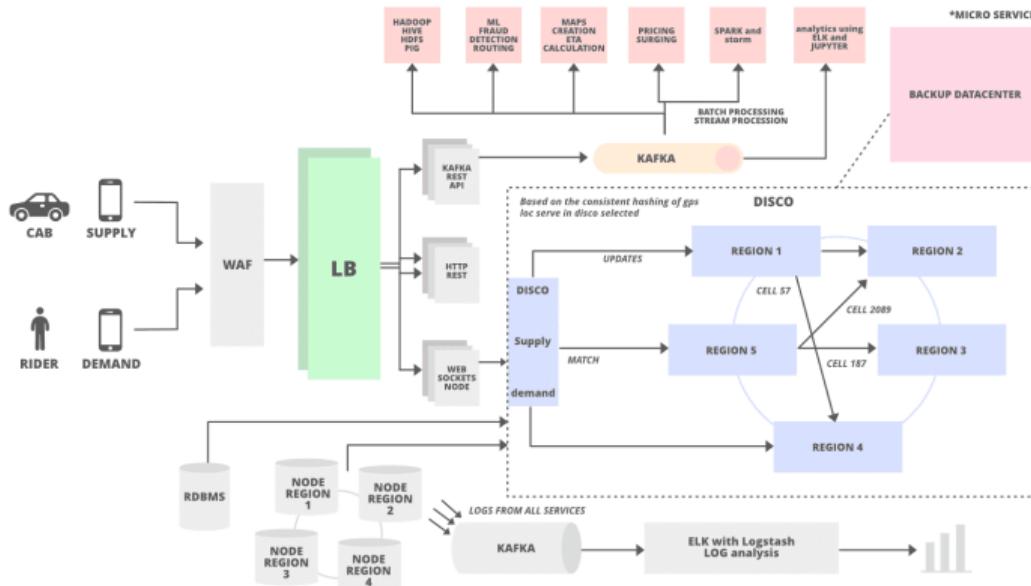
# Netflix 2010-2021 [NET21]



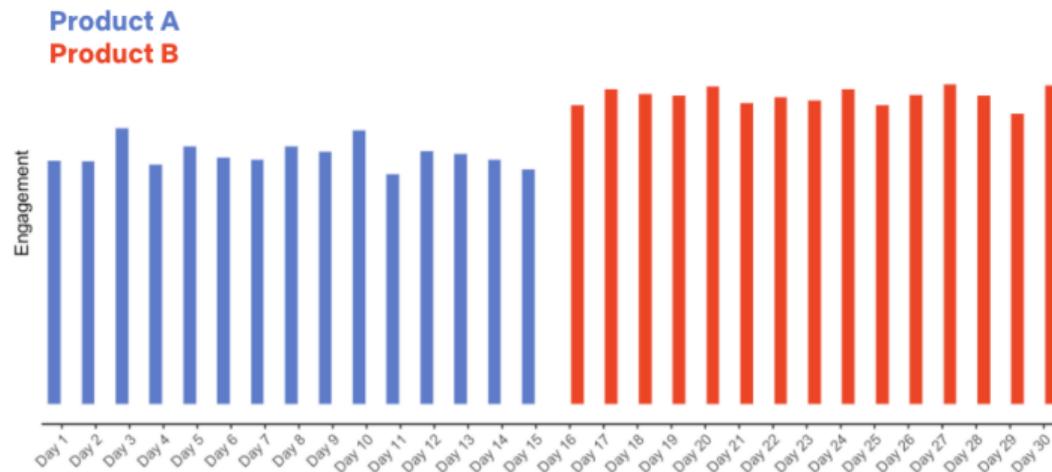
Хотим принимать решения на основе данных →

Начинаем собирать метрики →

Разрабатываем инструменты для принятия решений



## Наивный подход к измерению эффекта



## Задача

Какой причинно-следственный эффект на распределение целевой метрики окажет выбранное воздействие  $T$ ?

## Фундаментальная Проблема Causal Inference

Для конкретного пользователя невозможно вычислить causal effect напрямую, потому что нельзя проанаблюдать значение целевой переменной при более чем одном значении  $T^a$

---

<sup>a</sup>Без дополнительных предположений эту проблему не решить [GH07]



## Фреймворк Potential Outcomes

Воздействие на  $i$  пользователя:

$$T_i = \begin{cases} 0, & \text{если показываем control} \\ 1, & \text{если показываем treatment} \end{cases}$$

Соответствующие потенциальные исходы:

$$y_i^0 \text{ и } y_i^1$$

Требуется оценить:

Average Treatment Effect

$$ATE = E [y_i^1 - y_i^0]$$



## Randomized Controlled Experiment

### Схема эксперимента

Все доступные пользователи независимо друг от друга случайным образом распределяются в control либо treatment с одинаковой вероятностью

### Предположение 1:

Можно оценить значение некоторой характеристики для всей популяции, имея выборку из этой популяции.

### Предположение 2: Stable Unit Treatment Value Assumption

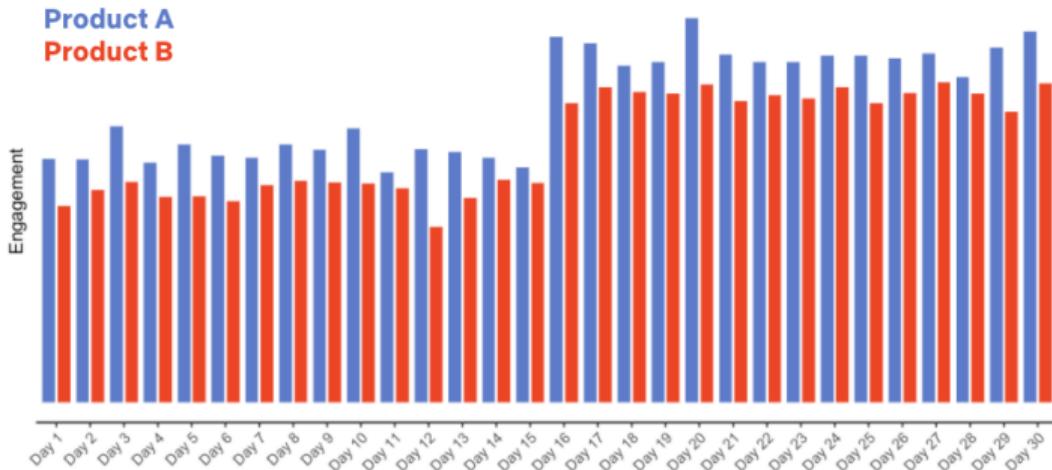
Потенциальные исходы для каждого пользователя зависят только от свойств этого пользователя, но не свойств и исходов других пользователей.



## Оцениваем ATE в RCE

$$ATE = E[y_i^1 - y_i^0] = E[y_i^1] - E[y_i^0] \sim \text{avg}_{i \in T}(y_i^1) - \text{avg}_{i \in C}(y_i^0) = \bar{y}_1 - \bar{y}_0$$

- нужно оценить две характеристики –  $E[y_i^0]$  и  $E[y_i^1]$ , поэтому используем выборки  $C$  и  $T$
- проще всего сделать оценку, если выборка несмещенная
- чем больше данных, тем точнее оценка



## Доверительный интервал на ATE

Доверительный интервал  $(L, U)$  с уровнем доверия  $\alpha$ :

$$P(L < \theta < U) = 1 - \alpha$$

Формула Уэлча:

$$\bar{y}_1 - \bar{y}_0 \pm t_{\alpha/2,r} \sqrt{\frac{s_1^2}{n_1} + \frac{s_0^2}{n_0}}, \quad r = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_0^2}{n_0}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_0^4}{n_0^2(n_0-1)}}$$

Где:

- $n_1$  и  $n_0$  – количество пользователей в treatment и control
- $s_1^2$  и  $s_0^2$  – оценки дисперсии метрики в treatment и control
- $t_{\alpha/2,r}$  – табличное значение для  $r$  степеней свободы



## На практике

- Перед запуском
  - Выбираем ключевую метрику, несколько сопутствующих метрик и контролируем, что не “уронили” важные
  - Выбираем длительность эксперимента, оценивая мощность теста :D
- При анализе
  - Метрики распределены по-разному: нужно подбирать подходящие тесты
  - Используются методы снижения дисперсии оценок (cuped, diff-in-diff)

Если вы попали в компанию, в которой есть культура принятия решений на основе данных – сохраняйте ее всеми силами. Если нет – пропагандируйте.

Архитектуры рекомендательных сервисов  
oooooooooooo

Метрики и эксперименты  
oooooooooooo

Итоги  
●○○

# Итоги

В основе рекомендательных сервисов лежит машинное обучение. При проектировании нужно учитывать множество дополнительных факторов, например требования к скорости обработки данных, эффект длинного хвоста и возможность холодного старта.

А/В эксперимент – надежный способ оценки эффекта от изменений в сервисе.

## Литература I

-  Paul Covington, Jay Adams, and Emre Sargin, *Deep neural networks for youtube recommendations*, Proceedings of the 10th ACM Conference on Recommender Systems (New York, NY, USA), RecSys '16, Association for Computing Machinery, 2016, p. 191–198.
-  Andrew Gelman and Jennifer Hill, *Data analysis using regression and multilevel/hierarchical models*, vol. Analytical methods for social research, Cambridge University Press, New York, 2007.
-  Kris Jack, Ed Ingold, and Maya Hristakeva, *Mendeley suggest architecture*, Oct 2016.
-  *Decision making at netflix (series)*, Sep 2021.
-  Xavier Amatriain Netflix and Justin Basilico Netflix, *System architectures for personalization and recommendation*, Mar 2013.

## Литература II

-  GALVANIZE Spotify, *Ever wonder how spotify discover weekly works? data science*, Aug 2016.
-  Catherine Wang, *Why tiktok made its user so obsessive? the ai algorithm that got you hooked.*, Jun 2020.