

# Рекомендательные сервисы в продакшене

Николай Анохин

24 сентября 2021 г.

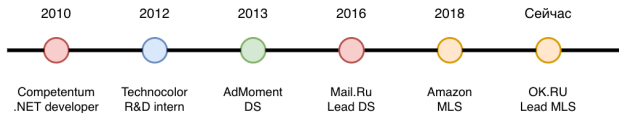
## Обзор модуля

## Обо мне

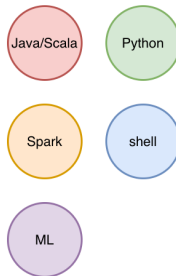
### Академический опыт



### Индустриальный опыт



### Навыки

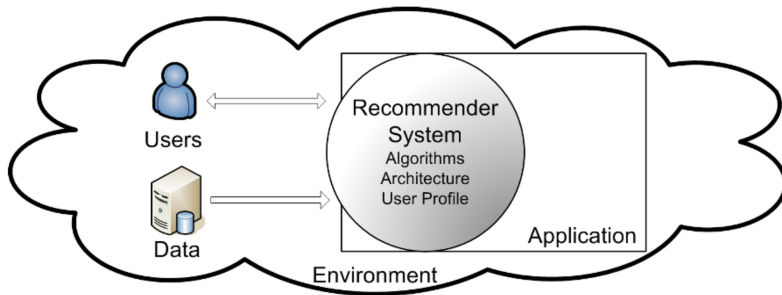


# Программа

Дата	Тема	Семинар	Домашка
2021-09-30	Рекомендательные сервисы в продакшене	✓	
2021-10-07	Метрики и базовые подходы	✓	
2021-09-14	Классические алгоритмы	✓	✓
2021-09-21	Нейросетевые рекомендеры	✓	
2021-09-28	Нерешенные проблемы и новые направления	✓	

## Зачем нужны рекомендательные сервисы

**Recommender Systems (RS)** are software tools and techniques providing suggestions for items to be of use to a user [RRSK10].



## Зачем RS бизнесу

- Увеличить продажи
- Продвигать более разнообразные товары
- Улучшить пользовательский опыт
- Добиться большей лояльности
- Лучше понимать пользователей

## Зачем RS пользователям

- Найти лучший товар
- Найти **все** подходящие товары
- Найти последовательность или набор товаров
- Залипнуть
- Найти рекоммендер, которому можно доверять
- Реализовать творческие потребности
- Помочь другим сделать выбор

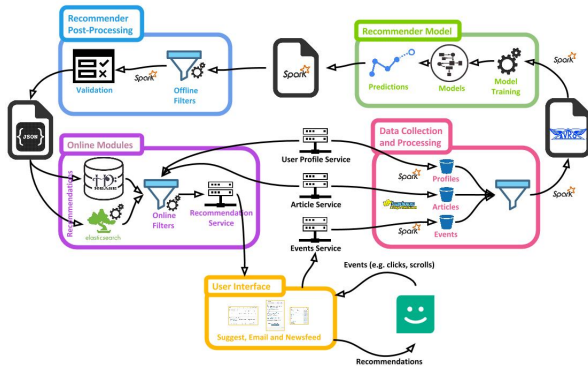


## Зачем RS инженерам

- Делать высоконагруженный отказоустойчивый сервис
- Анализировать большие данные
- Окунуться в волшебный мир магии машинного обучения
- Объективно измерять результат своей работы
- Все это за зарплату

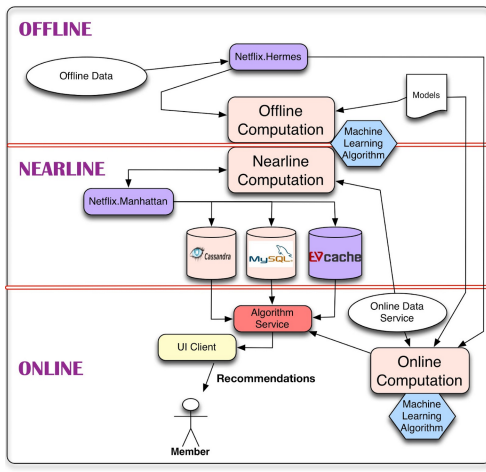
## Архитектуры рекомендательных сервисов

# Обзор типичных компонентов RS / Mendeley (2016) [JH16]



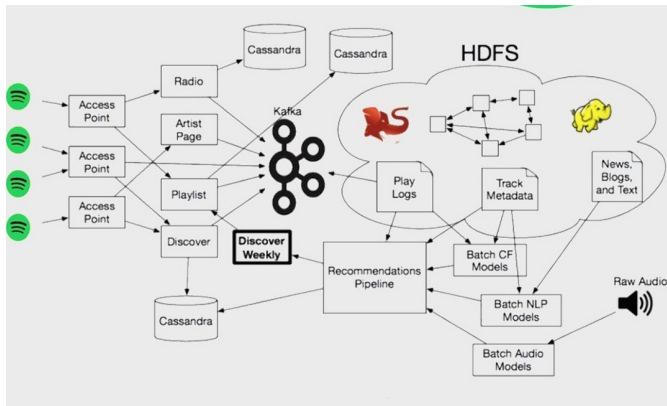
Машинное обучение – небольшая часть рекомендательного сервиса

## Подходы к обработке данных / Netflix (2013) [NN13]



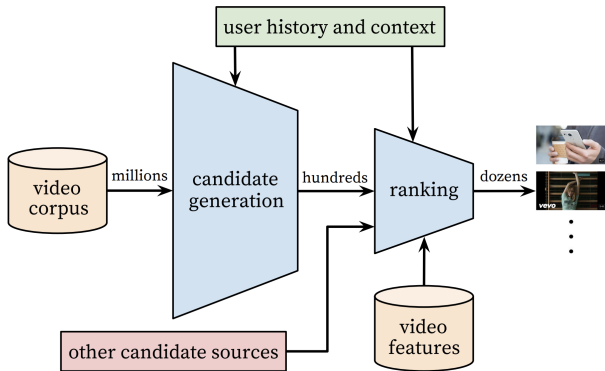
Чем ближе вычисления к real-time, тем больше ограничений и компромиссов

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



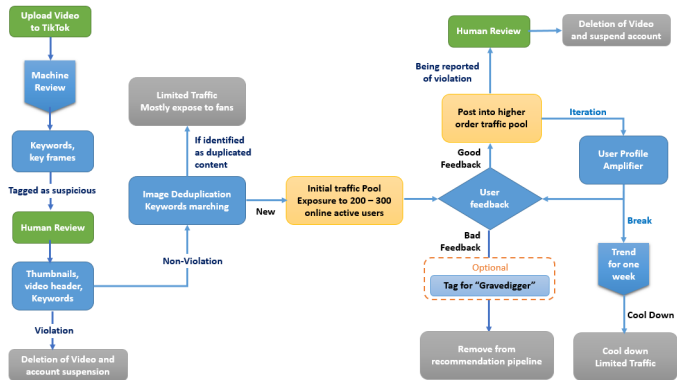
Холодные айтемы  
и пользователи  
будут всегда: ду-  
маем, что с ними  
делать

## Двухступенчатая архитектура / Youtube (2016) [CAS16]



Айтемов так много, что учесть **полный контекст** для не может даже Google

## Работа с контентом / TikTok (2020) [Wan20]



Потребности людей нельзя упаковать в удобную метрику: вокруг МЛ нужен пре- и пост-процессинг

## Метрики и эксперименты



## Разрабатываем инструменты для анализа метрик

## Задача

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

## Фундаментальная Проблема Causal Inference

Для конкретного объекта невозможно вычислить causal effect напрямую, потому что нельзя пронаблюдать 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 АТЕ 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 АТЕ

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

- Метрики распределены по-разному: нужно подбирать подходящие тесты
- Используются методы снижения дисперсии оценок (cuped, diff-in-diff)
- Собираются тысячи метрик: часто для интерпретации нужны специалисты

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







## Итоги

Рекомендательные сервисы улучшают жизнь бизнесу и пользователям. Жизнь инженеров они делают очень интересной.

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.

A/B 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.
-  Xavier Amatriain Netflix and Justin Basilico Netflix, *System architectures for personalization and recommendation*, Mar 2013.

