

Рекомендательные сервисы в продакшене

Николай Анохин

19 мая 2022 г.



Входной опрос

<https://forms.gle/vBpRB7UAHxtzCA1X8>

Информация о курсе

Николай Анохин

Академический опыт



Навыки



Индустриальный опыт

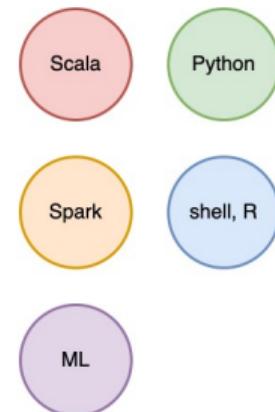


Дарья Никанорова

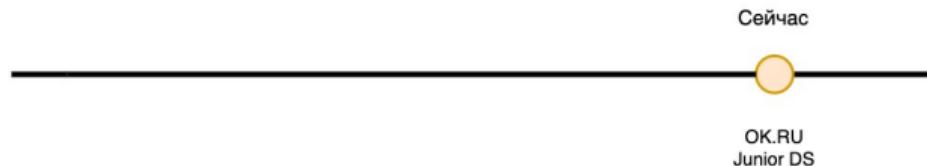
Академический опыт



Навыки



Индустриальный опыт



Материалы курса

Канал в discord: #2_rec_sys

Блог: <https://data.vk.company/blog/view/415/>

Репозиторий на GH: <https://github.com/anokhin/recsys-made-2022>

- Вопросы вне занятия можно задать в личке или в чате курса (лучше)
- Тегайте нас, чтобы мы не пропустили ваш комментарий в общем потоке сообщений
- Если ответа не последовало в течении 24 часов, то мы, вероятно, не увидели ваше сообщение. Не стесняйтесь его продублировать

Как задать вопрос

- Голосом
- В специально выделенное для этого время
- Перед тем как спросить будет хорошим тоном поставить несколько знаков вопроса

20:23 Саша: ????

20:23 Преподаватель: Ждём вопроса от Саши

20:24 Саша: Какая метрика хорошо работает в задаче рекомендаций?

Если что-то пошло не так

- Пропал голос
- Исчезло изображение
- Плохо слышно
- Любые проблемы другого характера

Сразу пишем в чат много минусов и не ждем других участников. Если вы увидели, что в чате кто-то написал много минусов, а у вас всё хорошо, то поставьте несколько плюсов:

20:24 Петя: - - - - -

20:25 Саша: + + +

20:25 Ольга: + + + + +

Программа модуля

Дата	Тема	Квиз	Семинар	Домашка
2022-05-19	Рекомендательные сервисы в продакшнене	✓	✓	
2022-05-26	Метрики и базовые подходы	✓	✓	
2022-06-02	Классические алгоритмы	✓	✓	✓
2022-06-09	Нейросетевые рекомендеры	✓	✓	
2022-09-16	Нерешенные проблемы и новые направления	✓	✓	

Оценка $\in [0, 1]$

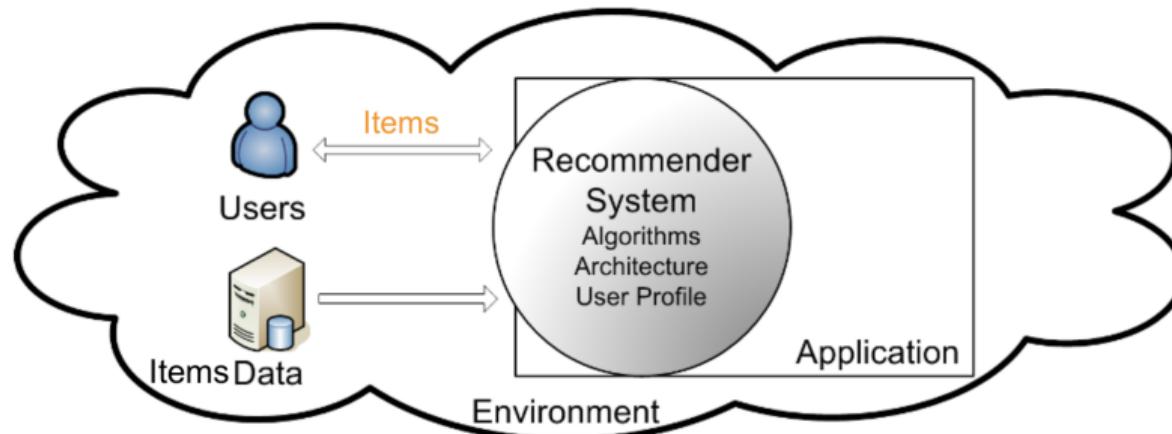
- Квизы = 0.4, домашка = 0.6
- Успех = оценка ≥ 0.6
- Дополнительные баллы можно получить на зачете
- Выберем трех самых активных слушателей и накинем им по 0.1

Пройдя этот курс, вы...

- Разработаете свой рекомендательный сервис (почти) с нуля
- Сможете выбирать правильные инструменты под конкретную задачу
- Узнаете о проблемах, возникающих в боевых RS, и научитесь их решать
- Будете в курсе SOTA моделей и задач RS, над которыми работают ученые

Зачем нужны рекомендательные сервисы

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



The screenshot shows the Netflix homepage. At the top, there is a large banner for the TV show "THE TINDER SWINDLER". Below the banner, there are two buttons: "Play" and "More Info". To the right of the banner, there is a search icon, a "Kids" icon, a notification icon with a red "3", a smiley face icon, and a dropdown menu. Below the banner, there is a section titled "TV Shows" featuring several other shows: "THE WOMAN IN THE HOUSE ACROSS THE STREET FROM THE GIRL IN THE WINDOW", "OZARK", "THE SINNER", "STAY CLOSE", "After Life", and "DOWNTON ABBEY". Each show has a thumbnail image and a "TOP 10" badge. Below this, there is a section titled "Trending Now" featuring "Sweet Magnolias", "THE TINDER SWINDLER", "THE POWER OF THE DOG", "THE CROWN", "BRIDGERTON", and "MONEY HEIST". Each show has a thumbnail image and a "TOP 10" badge.

The screenshot shows the Netflix homepage. At the top, there is a banner for the TV show "THE TINDER SWINDLER" with a season 1 indicator. Below the banner are two buttons: "Play" and "More Info". To the right of the banner are icons for search, Kids, notifications (with 3 notifications), and account settings. The main content area features a large image of a bridge over water. Below this, there are sections for "TV Shows" and "Trending Now", each displaying several movie and TV show thumbnails with their titles and release information.

TV Shows

- N THE WOMAN IN THE HOUSE ACROSS THE STREET FROM THE GIRL IN THE WINDOW TOP 10 NEW EPISODES
- N OZARK TOP 10
- N THE SINNER TOP 10 NEW EPISODES
- N HERMAN CORNELL'S STAY CLOSE TOP 10
- N After Life TOP 10 NEW EPISODES
- DOWNTON ABBEY

Trending Now

- N Sweet Magnolias TOP 10 NEW EPISODES
- N THE TINDER SWINDLER TOP 10
- N THE POWER OF THE DOG TOP 10
- N THE CROWN TOP 10
- N BRIDGERTON TOP 10
- N MONEY HEIST TOP 10 NEW EPISODES

The screenshot shows the Netflix homepage. At the top, there is a banner for the TV show "THE TINDER SWINDLER" with a season 1 indicator. Below the banner, there are two main sections: "TV Shows" and "Trending Now". The "TV Shows" section features thumbnails for "THE WOMAN IN THE HOUSE ACROSS THE STREET FROM THE GIRL IN THE WINDOW", "OZARK", "THE SINNER", "STAY CLOSE", "After Life", and "DOWNTON ABBEY". The "Trending Now" section features thumbnails for "Sweet Magnolias", "THE TINDER SWINDLER", "THE POWER OF THE DOG", "THE CROWN", "BRIDGERTON", and "MONEY HEIST". Each thumbnail includes a "NEW EPISODES" button. The overall design is dark-themed.

The screenshot shows the Netflix homepage with a large banner for the TV show "THE TINDER SWINDLER" (Season 1). Below the banner, there are sections for "TV Shows" and "Trending Now".

TV Shows:

- THE WOMAN IN THE HOUSE ACROSS THE STREET FROM THE GIRL IN THE WINDOW (Top 10)
- OZARK (Top 10) (2)
- THE SINNER (New Episodes)
- HERMAN CORNELL'S STAY CLOSE (New Episodes)
- After Life (New Episodes)
- DOWNTON ABBEY (New Episodes)

Trending Now:

- SWEET MAGNOLIAS (New Episodes)
- THE TINDER SWINDLER (Top 10) (3)
- THE POWER OF THE DOG
- THE CROWN
- BRIDGERTON
- MONEY HEIST (Top 10) (New Episodes)

The screenshot shows the Netflix homepage with a large banner for the TV show "THE TINDER SWINDLER". Below the banner, there are sections for "TV Shows" and "Trending Now".

TV Shows:

- 1. THE TINDER SWINDLER
- 2. Ozark
- 3. The Sinner
- 4. The Woman Across the Street
- 5. Stay Close
- 6. After Life
- 7. Downton Abbey

Trending Now:

- 4. Sweet Magnolias
- 3. The Tinder Swindler
- 2. The Power of the Dog
- 1. The Crown
- 5. Bridgerton
- 6. Money Heist

The screenshot shows the Netflix homepage with several UI elements highlighted by orange circles:

- Top Show Recommendation:** "THE TINDER SWINDLER" (TV Show) with a "Play" button and a "More Info" button. A yellow circle with the number 1 is positioned above the title.
- User Profile:** A circular icon in the top right corner containing the number 5, representing the user's account.
- TV Shows Section:** A grid of TV show thumbnails.
 - **Row 1:** "THE WOMAN ACROSS THE STREET" (4), "OZARK" (2), "THE SINNER", "STAY CLOSE", "After Life", "DOWNTON ABBEY".
 - **Row 2:** "Sweet Magnolias" (4), "THE TINDER SWINDLER" (3), "THE POWER OF THE DOG", "THE CROWN", "BRIDGERTON", "MONEY HEIST".
 A yellow circle with the number 4 is over "THE WOMAN ACROSS THE STREET". A yellow circle with the number 2 is over "OZARK". A yellow circle with the number 3 is over "THE TINDER SWINDLER".- Trending Now Section:** A grid of TV show thumbnails.
 - **Row 1:** "THE WOMAN ACROSS THE STREET" (4), "OZARK" (2), "THE SINNER", "STAY CLOSE", "After Life", "DOWNTON ABBEY".
 - **Row 2:** "Sweet Magnolias" (4), "THE TINDER SWINDLER" (3), "THE POWER OF THE DOG", "THE CROWN", "BRIDGERTON", "MONEY HEIST".
 A yellow circle with the number 4 is over "Sweet Magnolias". A yellow circle with the number 3 is over "THE TINDER SWINDLER".

Other visible UI elements include a search bar, a "Kids" icon, a notification bell with a count of 3, a smiley face icon, and a volume icon. The age rating "16+" is also present.

The screenshot shows the Netflix homepage with several UI elements highlighted by orange circles:

- Top Recommendation:** "THE TINDER SWINDLER" (TV Show) with a "1" badge.
- Buttons:** "Play" and "More Info".
- User Icons:** A profile icon with "6" and a notification icon with "5".
- Filters:** "TV Shows" and "16+".
- TV Shows Section:** "THE WOMAN ACROSS THE STREET" (4), "OZARK" (2), "THE SINNER", "STAY CLOSE", "After Life", and "DOWNTON ABBEY".
- Trending Now Section:** "Sweet Magnolias" (4), "THE TINDER SWINDLER" (3), "THE POWER OF THE DOG", "THE CROWN", "BRIDGERTON", and "MONEY HEIST".

RS vs другие задачи ML [SBE⁺21]

- RS ориентированы на продакшен
- Наблюдаемые данные очень разреженные
- Наблюдаем разное количество данных в зависимости от популярности
- Отсутствующие данные – либо ненаблюдаемые позитивные, либо негативные
- Отсутствующие данные – missing not-at-random
- Рекомендательные сервисы живут в петле обратной связи

Зачем RS бизнесу

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

Зачем RS пользователям

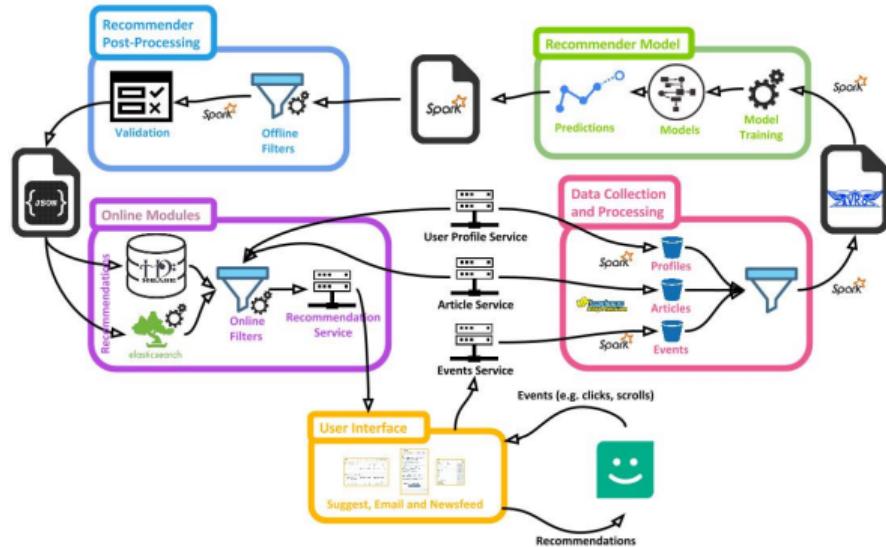
- Найти лучший товар
- Найти **все** подходящие товары
- Найти последовательность или набор товаров
- Залипнуть
- Найти рекомендер, которому можно доверять
- Помочь другим сделать выбор

Зачем RS инженерам

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

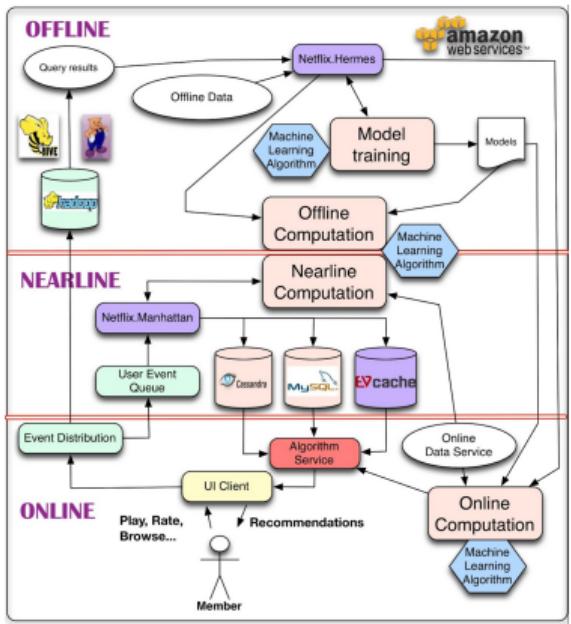
Архитектуры рекомендательных сервисов

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



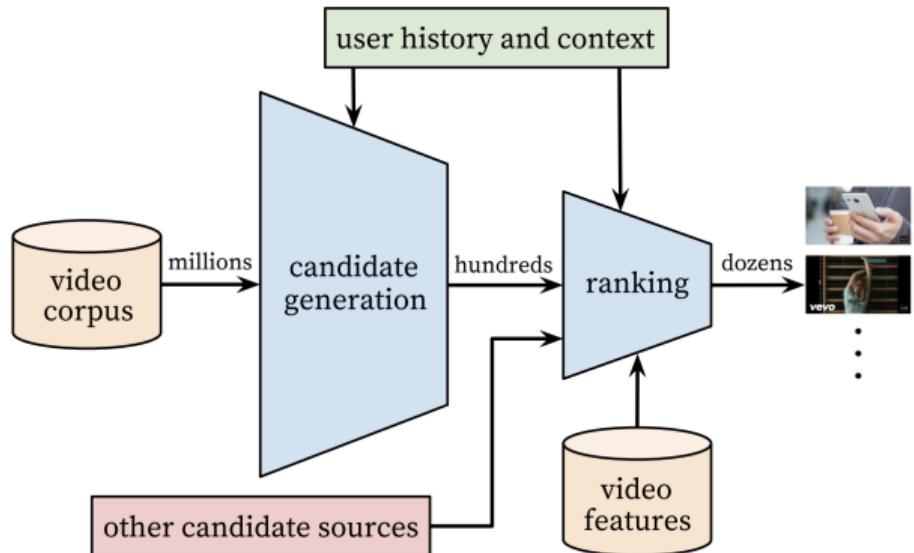
Машинное обучение – небольшая часть рекомендательного сервиса. Другие компоненты часто требуют не меньше усилий.

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

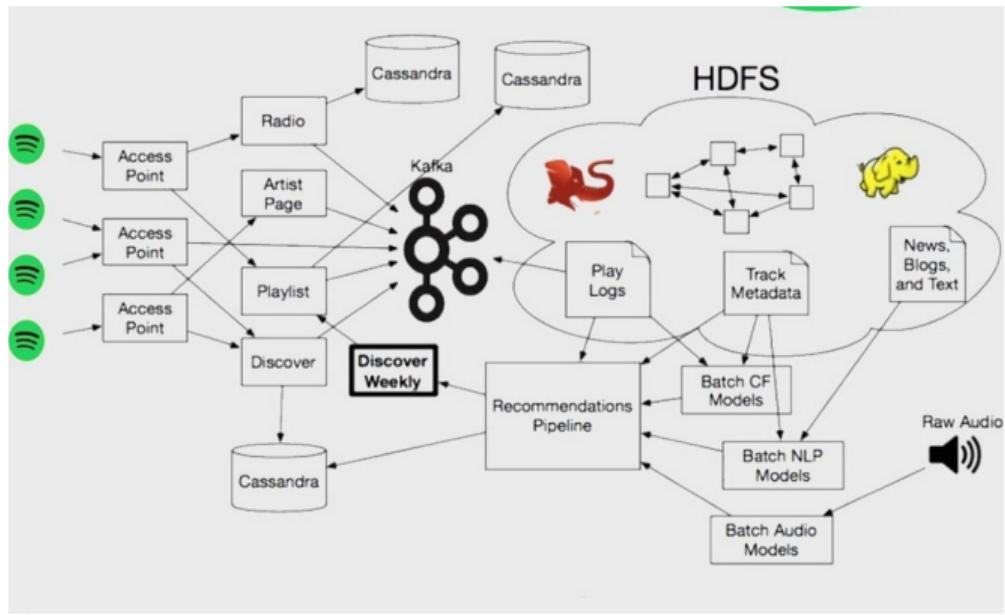
Power law

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



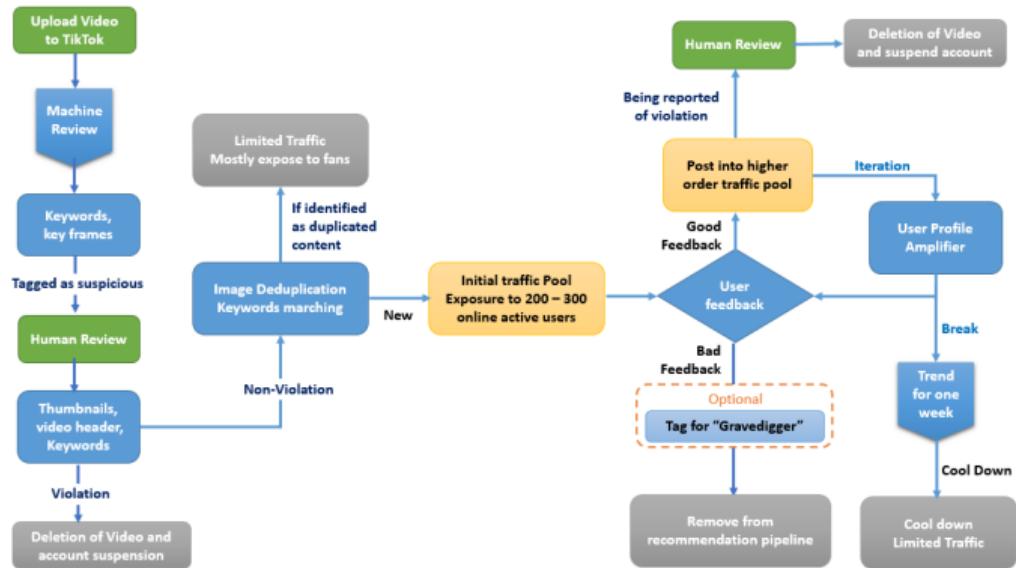
Правило 80/20

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

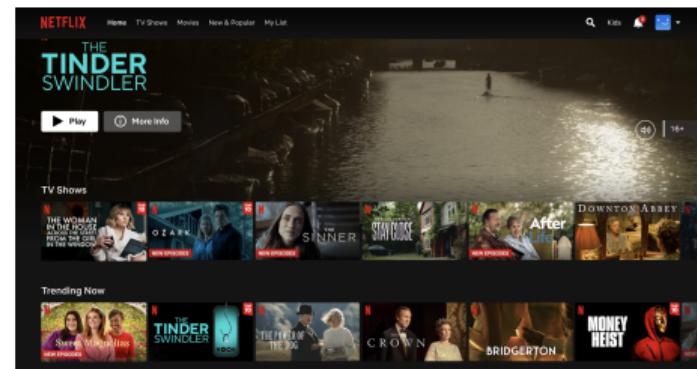
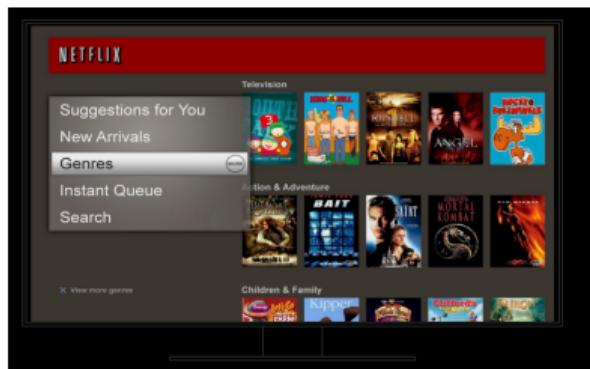
- Высокая нагрузка рекомендательных сервисов
- Большие каталоги айтемов
- Холодный старт пользователей и айтемов
- Несоответствие business-value и метрик оптимизации

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

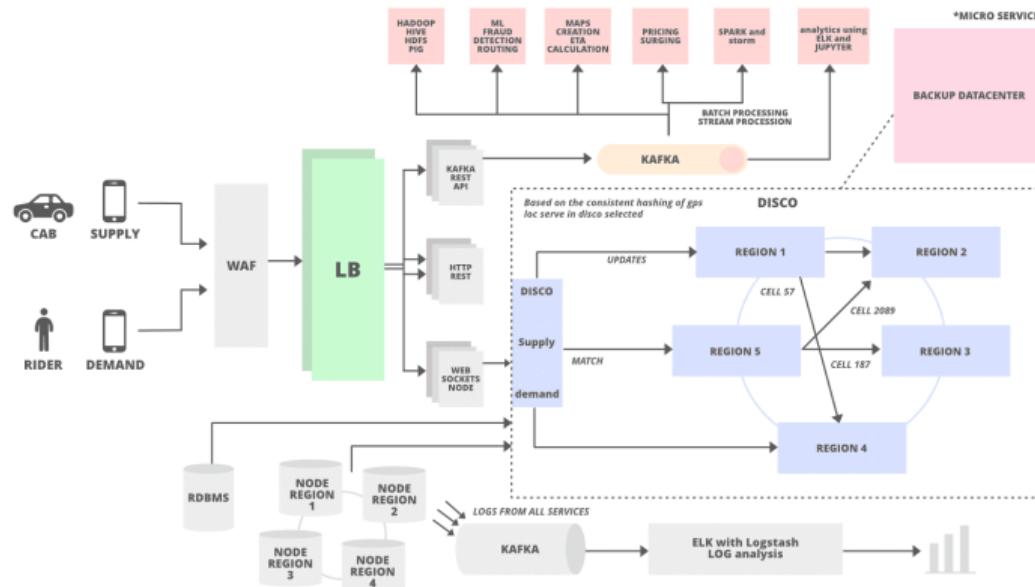
Netflix 2010-2021 [NET21]



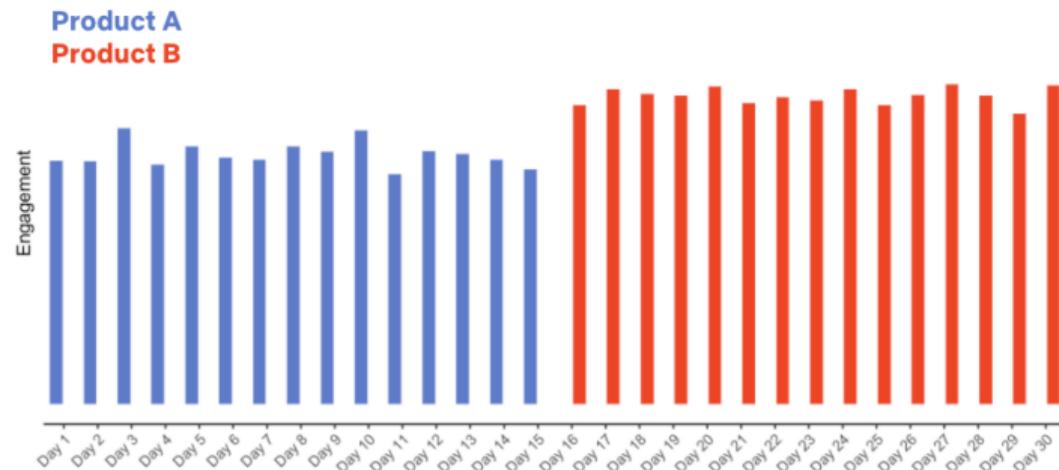
Хотим принимать решения на основе данных →

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

^aБез 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.

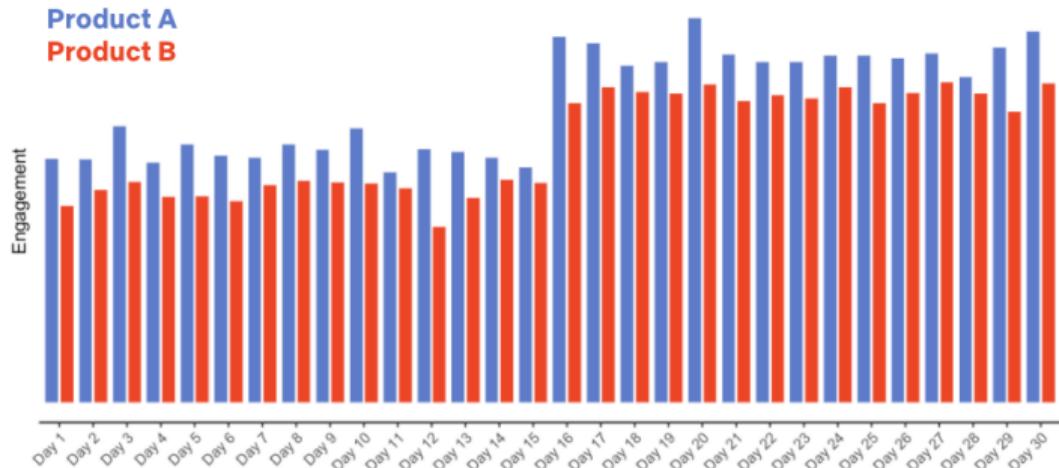
Предположение 2: Stable Unit Treatment Value Assumption

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 АТЕ

Доверительный интервал (L, U) с уровнем доверия α :

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

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.

Сложности RCE в индустриальных рекомендерах [GKT⁺¹⁹]

- Как выбрать Самую Главную Метрику (OEC)?
- Как оценить долгосрочный эффект?
- Разный эффект на разных сегментах пользователей
- Отсутствие культуры экспериментирования
- Масштабирование платформы для экспериментов
- Сетевые эффекты
- Наложение эффектов от экспериментов

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.
-  Somit Gupta, Ronny Kohavi, Diane Tang, Ya Xu, Reid Andersen, Eytan Bakshy, Niall Cardin, Sumita Chandran, Nanyu Chen, Dominic Coey, Mike Curtis, Alex Deng, Weitao Duan, Peter Forbes, Brian Frasca, Tommy Guy, Guido W. Imbens, Guillaume Saint Jacques, Pranav Kantawala, Ilya Katsev, Moshe Katzwer, Mikael Konutgan, Elena Kunakova, Minyong Lee, MJ Lee, Joseph Liu, James McQueen, Amir Najmi, Brent Smith, Vivek Trehan, Lukas Vermeer, Toby Walker, Jeffrey

Литература II

Wong, and Igor Yashkov, *Top challenges from the first practical online controlled experiments summit*, SIGKDD Explor. Newsl. **21** (2019), no. 1, 20–35.

-  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.
-  Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, *Recommender systems handbook*, 1st ed., Springer-Verlag, Berlin, Heidelberg, 2010.
-  Harald Steck, Linas Baltrunas, Ehtsham Elahi, Dawen Liang, Yves Raimond, and Justin Basilico, *Deep learning for recommender systems: A netflix case study*, AI Magazine **42** (2021), no. 3, 7–18.

Литература III

-  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.