

МОБОД-2022

Рекомендательные сервисы в продакшнене

Николай Анохин

8 сентября 2022 г.

Информация о курсе

Входной опрос



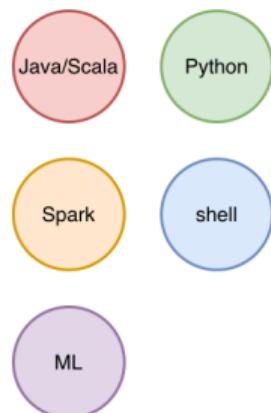
<http://shorturl.at/dosU1>

Николай Анохин

Академический опыт



Навыки



Индустриальный опыт

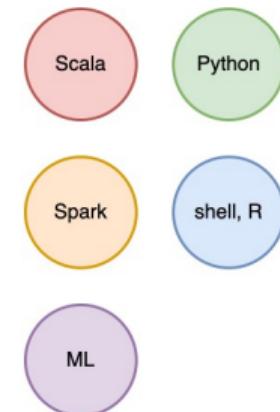


Дарья Никанорова

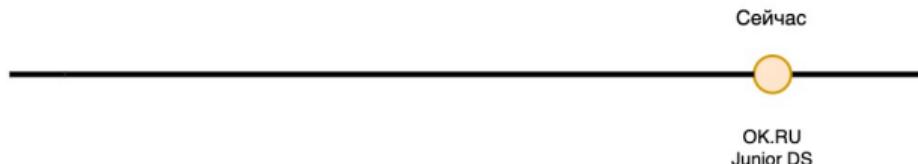
Академический опыт



Навыки



Индустриальный опыт



Telegram

Чат курса: <https://t.me/+qri5NHAS4Pc3YTky>

- 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 будущим 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:

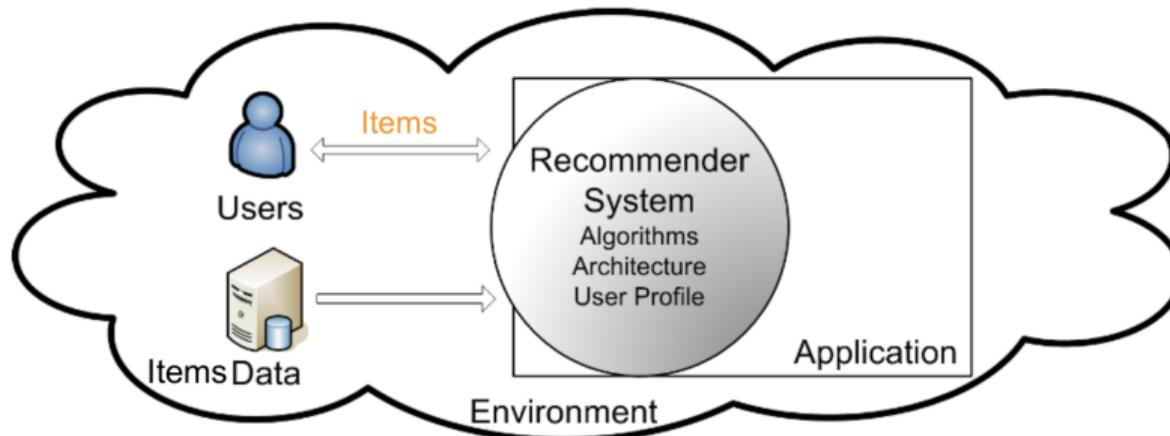
20:24 Петя: - - - - -

20:25 Саша: + + +

20:25 Ольга: + + + + +

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 image is a screenshot of the Netflix homepage. At the top, there is a navigation bar with links for "Home", "TV Shows", "Movies", "New & Popular", and "My List". On the right side of the header are icons for search ("Search"), "Kids", notifications ("3 notifications"), and account settings. The main title "THE TINDER SWINDLER" is displayed prominently in large green letters. Below the title is a "Play" button and a "More Info" button. To the right of the title, there is a volume icon and a rating of "16+". The background of the page features a dark, atmospheric image of a flooded area with a person standing in the water. Below the main title, there is a section titled "TV Shows" with thumbnails for several popular series: "THE WOMAN IN THE HOUSE ACROSS THE STREET FROM THE GIRL IN THE WINDOW" (TOP 10), "OZARK" (TOP 10), "THE SINNER" (NEW EPISODES), "STAY CLOSE", "After Life" (NEW EPISODES), and "DOWNTON ABBEY". At the bottom, there is a section titled "Trending Now" with thumbnails for "Sweet Magnolias" (NEW EPISODES), "THE TINDER SWINDLER" (TOP 10), "THE POWER OF THE DOG", "THE CROWN", "BRIDGERTON", and "MONEY HEIST" (TOP 10).



The image is a screenshot of the Netflix website. At the top, there is a navigation bar with links for "Home", "TV Shows", "Movies", "New & Popular", and "My List". On the far right of the header are icons for search ("Kids", a bell with a red notification dot, and a smiley face). The main content area features a large banner for the movie "THE TINDER SWINDLER". The banner includes a play button, a "More Info" button, a volume icon, and a rating of "16+". Below the banner, there is a section titled "TV Shows" with 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". Further down, there is a section titled "Trending Now" with thumbnails for "Sweet Magnolias", "THE TINDER SWINDLER", "THE POWER OF THE DOG", "THE CROWN", "BRIDGERTON", and "MONEY HEIST". Each thumbnail includes a "TOP 10" badge and a "NEW EPISODES" button.



NETFLIX

Home TV Shows Movies New & Popular My List

THE TINDER SWINDLER 1

Play More Info

16+

TV Shows

TOP 10 N THE WOMAN IN THE HOUSE ACROSS THE STREET FROM THE GIRL IN THE WINDOW NEW EPISODES 2 OZARK TOP 10 N THE SINNER NEW EPISODES N HERLÁN CORRÉA'S STAY CLOSE NEW EPISODES N After Life NEW EPISODES DOWNTON ABBEY N

Trending Now

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

Navigation icons: back, forward, search, etc.



The image is a screenshot of the Netflix website. At the top, there is a navigation bar with links for "Home", "TV Shows", "Movies", "New & Popular", and "My List". On the far right of the top bar are icons for search ("Search"), "Kids", notifications ("3 notifications"), and account settings. The main content area features a large banner for the movie "THE TINDER SWINDLER". The banner includes a play button, a "More Info" button with a circular badge containing the number "1", a volume icon, and a rating of "16+". Below the banner, there is a section titled "TV Shows" with thumbnails for several series: "THE WOMAN IN THE HOUSE ACROSS THE STREET FROM THE GIRL IN THE WINDOW" (TOP 10), "OZARK" (TOP 10), "THE SINNER" (NEW EPISODES), "STAY CLOSE", "After Life" (NEW EPISODES), and "DOWNTON ABBEY". At the bottom, there is a section titled "Trending Now" featuring thumbnails for "Sweet Magnolias" (NEW EPISODES), "THE TINDER SWINDLER" (TOP 10), "THE POWER OF THE DOG", "THE CROWN", "BRIDGERTON", and "MONEY HEIST".



The screenshot shows the Netflix homepage. At the top, there is a navigation bar with links to Home, TV Shows, Movies, New & Popular, and My List. To the right of the navigation are search, Kids, notifications (with 3 notifications), and account settings icons. The main title 'THE TINDER SWINDLER' is displayed prominently, with a large orange circle containing the number '1' indicating it's the top recommendation. Below the title are two buttons: 'Play' and 'More Info'. To the right of the title, there is a rating of '16+' and a volume icon. The background of the main section is a dark, atmospheric image of a bridge over water at night. Below the main title, there is a section titled 'TV Shows' featuring several other shows: 'THE WOMAN ACROSS THE STREET' (ranked 4, with a red 'TOP 10' badge), 'OZARK' (ranked 2, with a red 'TOP 10' badge), 'THE SINNER' (ranked 1, with a red 'TOP 10' badge), 'STAY CLOSE' (ranked 3, with a red 'TOP 10' badge), 'After Life', and 'DOWNTON ABBEY'. Each show has a 'NEW EPISODES' button. Below this, there is a section titled 'Trending Now' featuring 'Sweet Magnolias' (ranked 4, with a red 'TOP 10' badge), 'THE TINDER SWINDLER' (ranked 3, with a red 'TOP 10' badge), 'THE POWER OF THE DOG', 'THE CROWN', 'BRIDGERTON', and 'MONEY HEIST'. Each show in this section also has a 'NEW EPISODES' button. At the bottom of the page, there is a navigation bar with icons for back, forward, and search.



The screenshot shows the Netflix homepage. At the top, there is a navigation bar with links to Home, TV Shows, Movies, New & Popular, and My List. To the right of the navigation are search, Kids, notifications (with a count of 3), and account settings icons. A large orange circle with the number '5' is positioned in the top right corner.

The main feature is the TV show 'THE TINDER SWINDLER'. Below it are two buttons: 'Play' and 'More Info'. The show's image features a bridge over water at night. To the right of the show's image are volume and 16+ rating controls.

Below the main show, there is a section titled 'TV Shows' with several recommended titles:

- 1. THE WOMAN ACROSS THE STREET FROM THE GIRL IN THE WINDOW (marked with a yellow circle containing '4')
- 2. OZARK (marked with a yellow circle containing '2')
- 3. THE SINNER (marked with a yellow circle containing '3')
- 4. STAY CLOSE (marked with a yellow circle containing '1')
- 5. AFTER LIFE (marked with a yellow circle containing '5')
- 6. DOWNTON ABBEY

Below this section is a 'Trending Now' section with the following titles:

- 1. SWEET MAGNOLIAS (marked with a yellow circle containing '4')
- 2. THE TINDER SWINDLER (marked with a yellow circle containing '3')
- 3. THE POWER OF THE DOG
- 4. THE CROWN
- 5. BRIDGERTON
- 6. MONEY HEIST

At the bottom of the page, there is a navigation bar with icons for back, forward, and search.



The screenshot shows the Netflix homepage. At the top, there is a navigation bar with links to Home, TV Shows, Movies, New & Popular, and My List. On the right side of the header are icons for search, Kids, notifications (with a count of 3), and account settings. Below the header, the main title 'THE TINDER SWINDLER' is displayed with a large orange circle containing the number '1' to its right. Below the title are two buttons: 'Play' and 'More Info'. To the right of the title, there is a rating of '16+' and a volume icon. The background of the main section is a dark, atmospheric image of a bridge over water at night. Below the main title, there is a section titled 'TV Shows' with several thumbnail images for other shows. One thumbnail for 'THE WOMAN ACROSS THE STREET' has a yellow circle with the number '4' to its left. Another thumbnail for 'OZARK' has a yellow circle with the number '2' to its left. A third thumbnail for 'THE SINNER' has a yellow circle with the number '3' to its left. The thumbnails for 'STAY CLOSE', 'After Life', and 'DOWNTON ABBEY' do not have numbers. In the 'Trending Now' section below, there are thumbnails for 'Sweet Magnolias', 'THE TINDER SWINDLER', 'THE POWER OF THE DOG', 'THE CROWN', 'BRIDGERTON', and 'MONEY HEIST'. The thumbnail for 'THE TINDER SWINDLER' in this section also has a yellow circle with the number '3' to its left. The bottom of the page features a navigation bar with icons for back, forward, and search.



RS в индустрии и науке

Компании

Amazon.com, YouTube, Netflix, Yahoo, Tripadvisor, Spotify, Booking.com

Конференции

RecSys, KDD, WSDM, SIGIR

Книги

Recommender Systems Handbook [RRSK10], Mining of Massive datasets [LRU14]

Зачем RS бизнесу

- Увеличить продажи

Зачем RS бизнесу

- Увеличить продажи
- Продвигать более разнообразные айтемы

Зачем RS бизнесу

- Увеличить продажи
- Продвигать более разнообразные айтемы
- Улучшить пользовательский опыт

Зачем RS бизнесу

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

Зачем RS бизнесу

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

Зачем RS пользователям

- Найти лучший айтем

Зачем RS пользователям

- Найти лучший айтем
- Найти **все** подходящие айтемы

Зачем RS пользователям

- Найти лучший айтем
- Найти **все** подходящие айтемы
- Найти последовательность или набор айтемов

Зачем RS пользователям

- Найти лучший айтем
- Найти **все** подходящие айтемы
- Найти последовательность или набор айтемов
- Залипнуть

Зачем RS пользователям

- Найти лучший айтем
- Найти **все** подходящие айтемы
- Найти последовательность или набор айтемов
- Залипнуть
- Помочь другим сделать выбор

Зачем RS инженерам



- Делать высоконагруженный отказоустойчивый сервис :D
- Анализировать большие данные
- Окунуться в волшебный мир матана машинного обучения
- Объективно измерять результат своей работы
- Узнать много нового о людях
- Все это за зарплату

RS vs другие задачи ML [SBE⁺21]

- RS ориентированы на продакшен

RS vs другие задачи ML [SBE⁺21]

- RS ориентированы на продакшен
- Наблюдаемые данные очень разреженные

RS vs другие задачи ML [SBE⁺21]

- RS ориентированы на продакшен
- Наблюдаемые данные очень разреженные
- Отсутствующие данные – missing not-at-random

RS vs другие задачи ML [SBE⁺21]

- RS ориентированы на продакшен
- Наблюдаемые данные очень разреженные
- Отсутствующие данные – missing not-at-random
- Отсутствующие данные – либо ненаблюдаемые позитивные, либо негативные

RS vs другие задачи ML [SBE⁺21]

- RS ориентированы на продакшен
- Наблюдаемые данные очень разреженные
- Отсутствующие данные – missing not-at-random
- Отсутствующие данные – либо ненаблюдаемые позитивные, либо негативные
- Рекомендательные сервисы живут в петле обратной связи

Пройдя этот курс, вы...

- Разработаете свой рекомендательный сервис (почти) с нуля
- Научитесь использовать Spark для нужд RS
- 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
- Подготовитесь к собеседованию по RS на junior позицию

Что дальше

Программа модуля

Неделя	Тема	Квиз	Семинар	Домашка
1	Рекомендательные сервисы в продакшене	✓	✓	
2	Метрики и базовые подходы	✓	✓	
3	Классические алгоритмы 1	✓	✓	
4	Классические алгоритмы 2	✓	✓	✓
5	Нейросетевые рекомендеры	✓	✓	
6	Нерешенные проблемы и новые направления	✓	✓	
7	Рекомендации и Reinforcement Learning	✓	✓	

Оценка $\in [0, 1]$

- Идеальное выполнение домашки = 0.6 (нижняя граница “тройки”)
- Идеальное выполнение квизов и домашки = 0.9 (нижняя граница “пятерки”)
- Выберем трех самых активных слушателей и накинем им по 0.1
- Дополнительные баллы можно получить на зачете



Литература I

-  Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman, *Mining of massive datasets*, 2nd ed., Cambridge University Press, USA, 2014.
-  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.

