

МОБОД-2023

Рекомендательные сервисы в продакшнене

Николай Анохин

14 сентября 2023 г.



Информация о курсе

●oooooo

Зачем нужны рекомендательные сервисы
oooooooooooooooooooo

Что дальше
oooooo

Информация о курсе



Входной опрос



<https://shorturl.at/jvxzY>

Николай Анохин

Академический опыт



Навыки



Индустриальный опыт

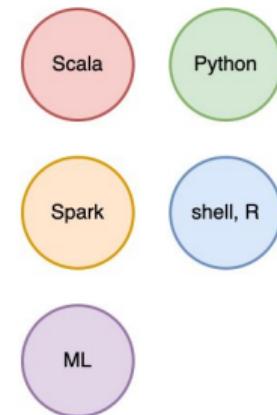


Дарья Никанорова

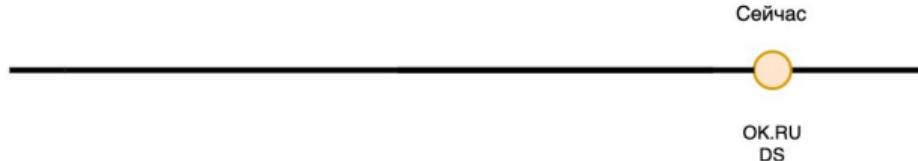
Академический опыт



Навыки



Индустриальный опыт



Telegram

Чат курса: <https://t.me/+b66BQ4PCWro5M2Ey>

- 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: + + + + +



Информация о курсе
oooooooo

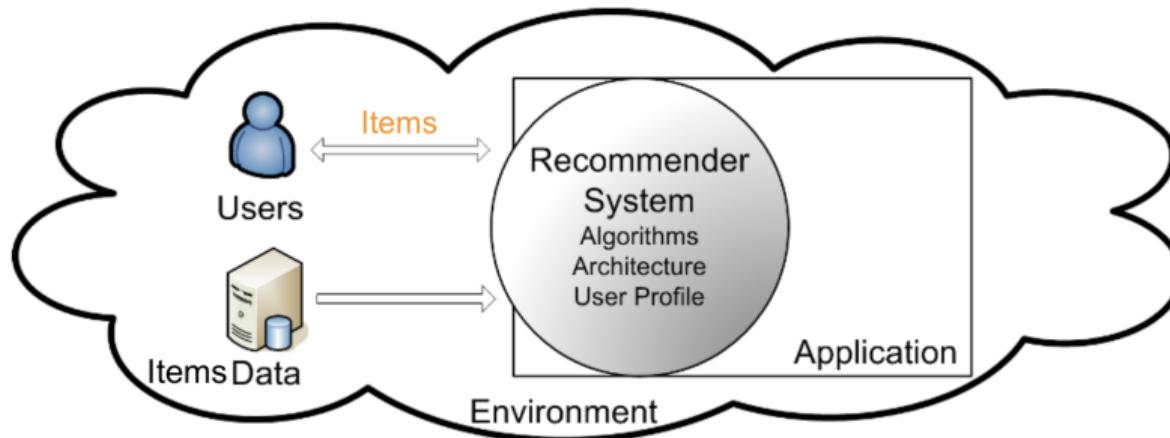
Зачем нужны рекомендательные сервисы
●oooooooooooooooooooo

Что дальше
oooooooo

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 TINDER SWINDLER

Play

More Info

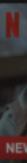


16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

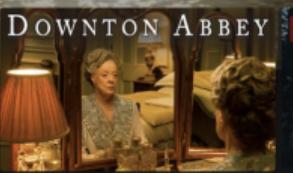
Play

More Info



16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

Play

More Info



16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

Play

More Info



16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

Play

More Info



16+

TV Shows



Trending Now



5

THE TINDER SWINDLER

1

Play

More Info



16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

Play

More Info



16+

TV Shows



Trending Now



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]



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



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



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

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



ML Инженер



Информация о курсе
oooooooo

Зачем нужны рекомендательные сервисы
oooooooooooooooooooo

Что дальше
●ooooo

Что дальше



Что нужно знать до начала курса

Базовая математика

- Операции над векторами и матрицами.
- Дифференцирование и нахождение минимумов функций.
- Базовые алгоритмы и структуры данных.

Основы машинного обучения

- Постановка задачи ML, классические алгоритмы.
- Базовое представление о том как работают современные нейронные сети.

Технические скиллы

- Знание языка python.
- Знакомство с архитектурой высоконагруженных сервисов (БД + Big Data).
- Умение писать простые запросы на PySpark.
- Знакомство с Docker, умение запустить контейнер.
- Умение пользоваться git.



Программа модуля

Репозиторий: <https://gitlab.atp-fivt.org/courses-public/mobod/public>

Неделя	Тема	Квиз	Семинар	Домашка
0	Вводное занятие (вы сейчас на нём!)		✓	
1	Рекомендательные сервисы в продакшене	✓	✓	
2	Метрики и базовые подходы	✓	✓	
3	Классические алгоритмы 1	✓	✓	
4	Классические алгоритмы 2	✓	✓	✓
5	Нейросетевые рекомендеры 1: отбор кандидатов	✓	✓	
6	Нейросетевые рекомендеры 2: ранжирование	✓	✓	✓
7	GNN в рекомендациях	✓	✓	
8	Нерешенные проблемы и новые направления	✓	✓	
9	Рекомендации и Reinforcement Learning	✓	✓	



Оценка $\in [0, 1]$

- Идеальное выполнение домашек = 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.

