

Рекомендательные системы ШАД

Весна 2025

Лекция 1:
Введение

Команда курса

Лекторы:

Ткаченко Даня , Лавка

Зайцева Алена , Лавка

Хрыльченко Кирилл , Портал

Семинаристы:

Нигматуллин Рома , constructor.com, ex. Лавка

Красильников Леша , Портал, ex. Лавка

Матвеев Артем , Портал

Байкалов Владимир , Портал

Даня Ткаченко

- окончил **мехмат** МГУ, матлог
- пришел в Яндекс **в 2018ом** улучшать качество поиска по видео
- **с 2019ого** года занимался рексистемами в разных местах **портала**
- **с 2020** года делал рекомендации в **Дзене**
- **в 2022** году пришел делать рекомендации в **Лавку**, и **в 2023** постепенно начал руководить там всем **ML**

Алена Зайцева

- окончила **Бауманку (ИУ7, программная инженерия)**
- поработала в стартапе, где писали "у нас есть attention дома"
- пришла **в Яндекс в 2023-м** году, поработала над поиском, ллм-ками и рекомендациями
- **руководжу командой рекомендаций** и драйвлю использование AI в разных частях **Лавки**

Кирилл Хрыльченко

- занимается машинкой с 2017-го года
- окончил **ВМК МГУ** (бакалавриат матстата, магистратуру ммп)
- пришел в **Яндекс** в 2020-м году делать **нейросетевые технологии** для рекомендаций; с 2022-го года руковожу командой, занимающейся **R&D рексистем** для всей экосистемы Яндекса
- помогал **улучшать рекомендации** Музыки, Маркета, Кинопоиска, Алисы, Лавки, а также Поиск и Рекламу

Что будет в курсе

Лекция и семинар **1**

Лекция и семинар **2**

Старт **контеста**

Лекция и семинар **3**

Выдача **ДЗ-1**

Лекция и семинар **4**

Лекция и семинар **5**

Лекция и семинар **6**

Выдача **ДЗ-2**

Лекция и семинар **7**

Лекция и семинар **8**

Введение

Оффлайн и онлайн метрики

Кандидатогенерация, модели

Про ранжирование

Нейросетевое ранжирование

Нейросетевая кандидатогенерация

Case Study

Тренды в RecSys

Оценки

Домашка

Первая из **6** баллов

Вторая из **10** баллов

Бонусные баллы

Топ-3/Топ-10 в конкурсе

Баллы от семинаристов
за проактивность

Контест

3 бейзлайна

2 балла за первый

4 балла за второй

6 баллов за третий

отл **24**, хор **20**, уд **14**

Дисклеймер

Про что этот курс и не про что этот курс

- **Курс посвящен не только рекомендациям как математической задаче**
- **В том числе пробуем дать представление (не очень длинное 🙄) о том, как может выглядеть рек-система in your friendly neighborhood рантайме**
- **Цель курса: чтобы вы смогли пройти к нам собес на джуна**
- **Мы = абстрактная продуктовая или исследовательская команда по рекам**

Да кто такие эти ваши RecSys-темы

A **recommender system (RecSys)**, or a **recommendation system** (sometimes replacing *system* with terms such as *platform*, *engine*, or *algorithm*), is a subclass of **information filtering system** that provides suggestions for items that are most pertinent to a particular user.^{[1][2][3]} Recommender systems are particularly useful when an individual needs to choose an item from a potentially overwhelming number of items that a service may offer.^{[1][4]}

И где они обитают

**...обычно где-то, где кто-то должен выбрать что-то,
но он не очень хочет этого делать.**

Зачем

Бизнес:

- Ключевые метрики (e.g. **GMV, CPC** , etc)
- **Timespent** для подписочных сервисов
- Доказанный эффект на **longterm** -метрики (**retention, NPS** , etc)
- Дешево в проде, все еще можно говорить, что **AI** 😊

Зачем

Пользователи:

- Обычно **не** сильно **любят** выбирать
- **Редко** знают, что хотят
- Хотим знать **за них**

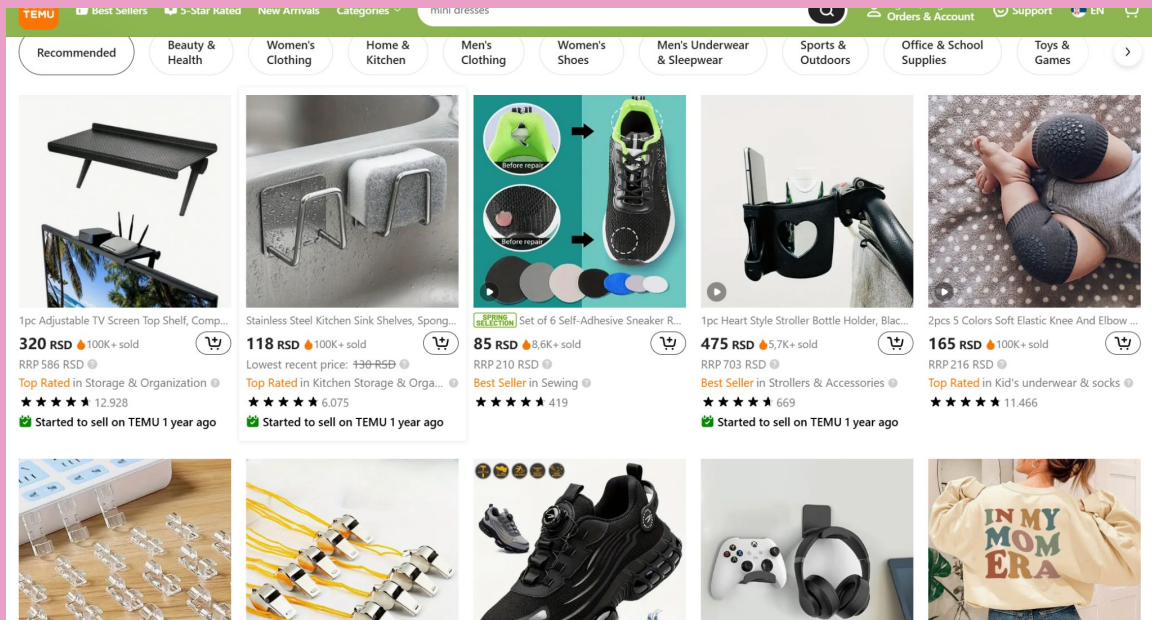
Зачем

Мы:

- Можно заниматься **не только** трансформерами
(но это не точно, подробнее дальше в курсе)
- Всегда **на стыке** продукта и ML
- Нам платят
(если все хорошо)



И где они обитают



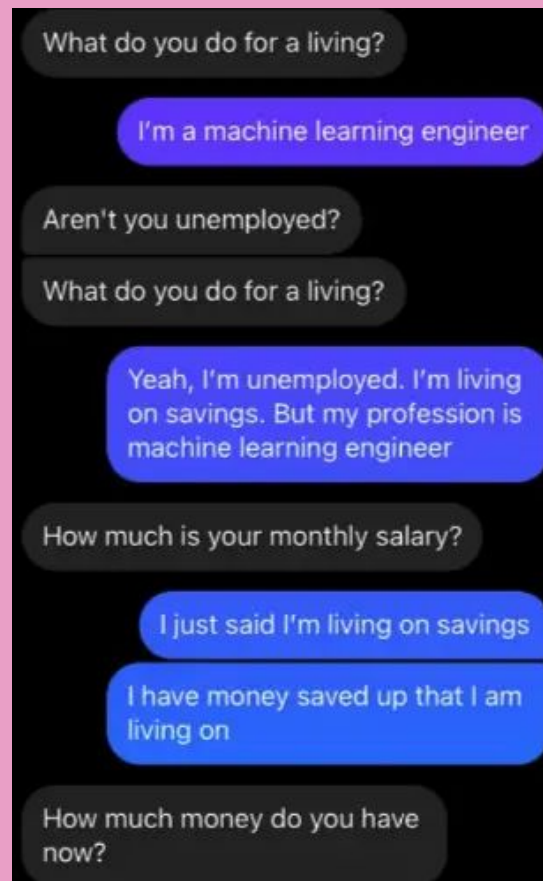
- Бизнес может больше продавать
- Пользователю проще выбирать из миллионов товаров

И где они обитают



- **Бизнесу нужны**
 - а) подписки**
 - б) рекламная выручка**
- **Пользователям нужны матчи**

**При помощи
рекомендательной
системы можно
оптимизировать
функцию от этого [1]**



И где они обитают



- Бизнесу нужен **жесткий** таймспент ради рекламы и иногда счастье авторов
- Пользователю нужно меньше смотреть тикток и **начать жить** [2]

И где они обитают



Google Ads

- Бизнесу **надо денег** с аукционов второй цены
- Для честного аукциона нужна **хорошая вероятность клика** [3]
- Ну вы поняли, это рексистема
- Пользователям **ничего** из этого **не надо**
- Смешную картинку не нашел

Коротенькая история Датасет MovieLens



- **MovieLens – изначально сайт с системой рекомендаций фильмов, запущенный GroupLens в 1997**
- **Все еще работает (qr-code в углу)**
- **Делали, чтобы собирать данные для рисерча**
- **Был go-to-dataset-ом много лет, замеры легко найти в статьях через одну, все еще широко используется [4]**
- **Перестал обновляться в районе 2019 из-за privacy concern-ов**

Коротенькая история Netflix Prize

- **Контеcт, проводившийся Netflix-ом в 2006–2009гг.**
- **Датасет и бейзлайн к нему предоставлялся Netflix (их алгоритм рекомендаций на тот момент)**
- **100,480,507 оценок, которые поставили 480,189 пользователей 17,770 фильмам, с датами**
- **1,000,000\$ самому крутому, кто побьет их алгоритм на 10% на RMSE**
- **Победители выиграли по времени (разница 20 минут)**

Коротенькая история Последствия Netflix Prize

- **Коллаборативная фильтрация с помощью матричной факторизации is king (SVD/ALS)**
- **Но не без ансамблирования и экстенсивной предобработки**
- **Привлек большое внимание к рекомендательным системам, взрыв ресерча**
- **Довольно рано стала понятна ценность правильной метрики (и что RMSE неправильная)**
- **Hinton et al. Предложили первые нейронки (RBM)**
- **Подробно тут [5]**

Коротенькая история Появление нейронок

- **2013, DSSM-like рекомендации: двухбашенные MLP с одной башней для пользователя и одной для айтема [6]**
- **2016, Wide&Deep: комбинация из линейной (memorization) и нейронной (generalization) модели [7]**
- **2017, DCN: эмуляцию взаимодействий второго порядка при помощи Cross-слоев [8]**
- **2018, SASRec, переформулировали задачу и впили attention (в прод естественно такое не завезти) [9]**
- **2024, HSTU: первый генеративный подход, который сработал в проде [10]**

**Так как
порекомендовать
что-нибудь?**

Модельный кейс:

- Рекомендации новостей из классической статьи Yahoo про многоруких бандитов в рекомендациях [11]
- В проде уже что-то работает
- Нас просят сделать рекомендации прямо сейчас, и нет времени писать фронтенд-логи (дальше в курсе будет про важность этого)
- У нас есть только позитивные взаимодействия

Попробуем собрать что-то совсем простое.

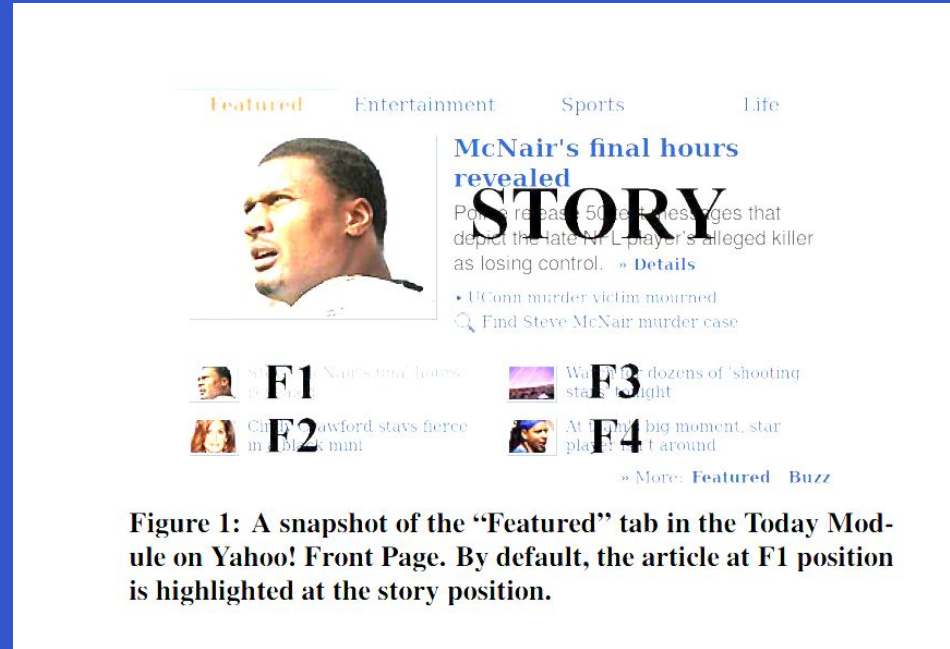


Figure 1: A snapshot of the “Featured” tab in the Today Module on Yahoo! Front Page. By default, the article at F1 position is highlighted at the story position.



Самый полезный слайд во всем курсе

(но это неточно)

Как решить вашу ML-задачу

- Что **хочет** бизнес?
- Что мы можем предсказать, что плюс-минус этому **соответствует** ?
- Какая **модель** может это предсказать?
- Какие **признаки** ?
- Как строится **датасет** ?
- Как провалидировать результаты **в оффлайне**
- Как провалидировать результаты **в онлайн**



Что хочет бизнес?

бизнес

предикт

модель

датасет

валидация



Что хочет бизнес?

**Допустим, бизнес хочет как
можно больше прочитанных
новостей.**



бизнес

предикт

модель

датасет

валидация



Что мы можем
предсказать, что
плюс-минус этому
соответствует?



бизнес

предикт

модель

датасет

валидация



**Что мы можем предсказать, что плюс-минус
этому соответствует?**

**Допустим, вероятность клика. Если у нас
есть эта вероятность, мы просто можем
отсортировать по ней.**

бизнес

предикт

модель

датасет

валидация



**Какая модель
может это
предсказать?**



бизнес

предикт

модель

датасет

валидация



Какая модель может это предсказать?

Something something CrossEntropy.

В качестве модели возьмем градиентный бустинг для простоты, скорости инференса и патриотических соображений.

бизнес

предикт

модель

датасет

валидация



Какие признаки у
модели?



бизнес

предикт

модель

датасет

валидация



Какие признаки у модели?

Пользователь, Мир вокруг, Новость, Их взаимодействие

бизнес

предикт

модель

датасет

валидация



Как строится датасет?

бизнес

предикт

модель

датасет

валидация



Как строится датасет?

Разделение, позитивы, негативы.

Разделение – горячий топик. На практике в индустрии **всегда** разделяют по времени. Больше об этом дальше в курсе.

бизнес

предикт

модель

датасет

валидация



Как провалидировать результаты

бизнес

предикт

модель

датасет

валидация



Как провалидировать результаты

В оффлайне: ваши любимые способы для бинарной классификации.

В рантайме: шмальнуть АБ на пользователей. Нужно придумать метрики, чтобы знать, что все хорошо. И что все плохо :smirk:

Но вообще такое изи заработает. Source: trust me bro.

бизнес

предикт

модель

датасет

валидация



Спасибо!

Даня Ткаченко,
Служба ML-сервисов Лавки,
Белград 2025

- [1] Machine Learning Recommendation Algorithms in Online Dating**
- [2] The Mystery of DNNs: Debunked and Explained | TikTok**
- [3] Как работает аукцион в Директе**
- [4] MovieLens | GroupLens**
- [5] Recommender Systems in Industry: A Netflix Case study**
- [6] Deep Neural Networks for YouTube Recommendations**
- [7] [1606.07792] Wide & Deep Learning for Recommender Systems**
- [8] [1708.05123] Deep & Cross Network for Ad Click Predictions**
- [9] [1808.09781] Self-Attentive Sequential Recommendation**
- [10] [2402.17152] Actions Speak Louder than Words: Trillion-Parameter Sequential Transducers for Generative Recommendations**
- [11] [1003.0146] A Contextual-Bandit Approach to Personalized News Article Recommendation**