



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

Факультет компьютерных наук

# Годовой проект: Построение рекомендательной системы игр на Steam

**Куратор:**

Милоградский Александр

**Выполнили:**

Ермилов Артемий

Неволина Арина

Кубаева Ассоль

Поспелов Артем

Москва, 2024



# Краткое описание проекта

Построение рекомендательной системы игр на Steam

**Задача:** построение рекомендательной системы игр на Steam

**Используем данные трех типов:**

- **Датасет информации об играх:**
  - Steam Games Dataset (Kaggle)
  - $\approx 76\,000$  игр
- **Датасет отзывов пользователей:**
  - Комбинация нескольких датасетов отзывов:
    - Steam Game Reviews 2021 (Kaggle)
    - Steam Store Apps Reviews (Kaggle)
    - Собственный датасет отзывов пользователей
      - Время парсинга - 4 man-weeks
  - $\approx 108$  миллионов отзывов
- **Датасет библиотек пользователей:**
  - Собственный датасет информации о библиотеках пользователей и их плейтайму
  - Парсинг с помощью Steam API, время парсинга  $\approx 4$  man-months
  - $\approx 492$  миллиона наблюдений пользователь-игра-время





# Что было сделано: EDA

Построение рекомендательной системы игр на Steam

- В выборке подавляющее большинство (90%) **непопулярных** игр, демонстрирующих **отличные от остальных игр тренды**. Необходимо **исключить их** из выборки
- В выборке подавляющее большинство (75%) пользователей имеет **крайне мало отзывов** ( $\leq 11$ ). Необходимо учесть это при работе с взаимодействиями
- Около **43%** всех наблюдений игра-пользователь-время имеют **время в игре, равное 0**. То есть игра была куплена, но в нее не играли
- Среднее **количество наблюдений** на пользователя в датасете **отзывов** ( $\approx 7$ ) **намного меньше**, чем в датасете **библиотек** ( $\approx 170$ )
- Во всех выборках **есть выбросы**, которые **необходимо обработать**







# Что было сделано: бейзлайн

Построение рекомендательной системы игр на Steam

Были построены 4 рекомендательных системы:

- 1) **User-based** коллаборативная фильтрация
- 2) **Item-based** коллаборативная фильтрация
- 3) Рекомендация популярных игр
- 4) Рекомендация новинок

Тип взаимодействия - **отзывы**

Функции для рек. систем были написаны вручную, так как готовые имплементации (implicit, surprise) не позволяли делать рекомендации для новых пользователей





# Что было сделано: деление на train и test

Построение рекомендательной системы игр на Steam

Существующие в **RecSys** подходы к разделению данных:

- **Leave One Out**
- **Random Holdout**
- **Time Based Split**

**Выбранный нами подход: Random Holdout**

- Размер тестовой выборки - **30%**
- Сплит в рамках **каждого пользователя**
- Тестовая выборка состоит **только из положительных отзывов**







# Что было сделано: метрики

Построение рекомендательной системы игр на Steam

## Ключевые метрики:

- 1) Mean Recall @ K
  - 2) Mean Precision @ K
  - 3) Mean Average Precision @ K
  - 4) Mean nDCG @ K
- Выбранные значения K: 1, 5, 10, 25, 50, 100
  - Усреднение по пользователям



$$precision@k = \frac{\text{number of recommended relevant items among top } k}{\text{number of recommended items } k}$$

$$recall@k = \frac{\text{number of recommended relevant items among top } k}{\text{number of all relevant items in the system}}$$

$$AP@k = \frac{1}{r} \sum_{i=1}^k precision@i \cdot R_i$$

$r$  = number of relevant items

$$R_i = \begin{cases} 1, & \text{if document } i \text{ is relevant} \\ 0, & \text{if document } i \text{ is not relevant} \end{cases}$$

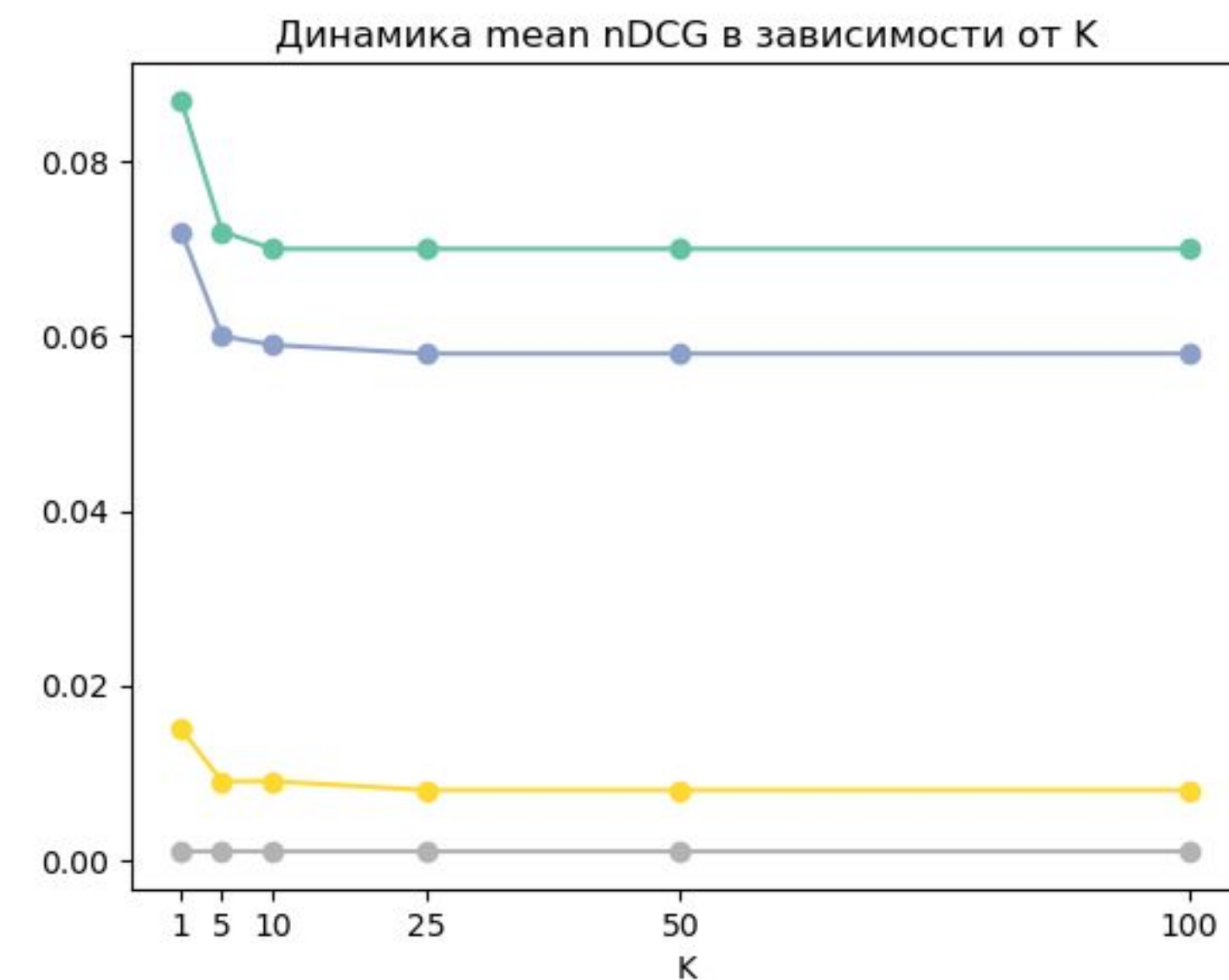
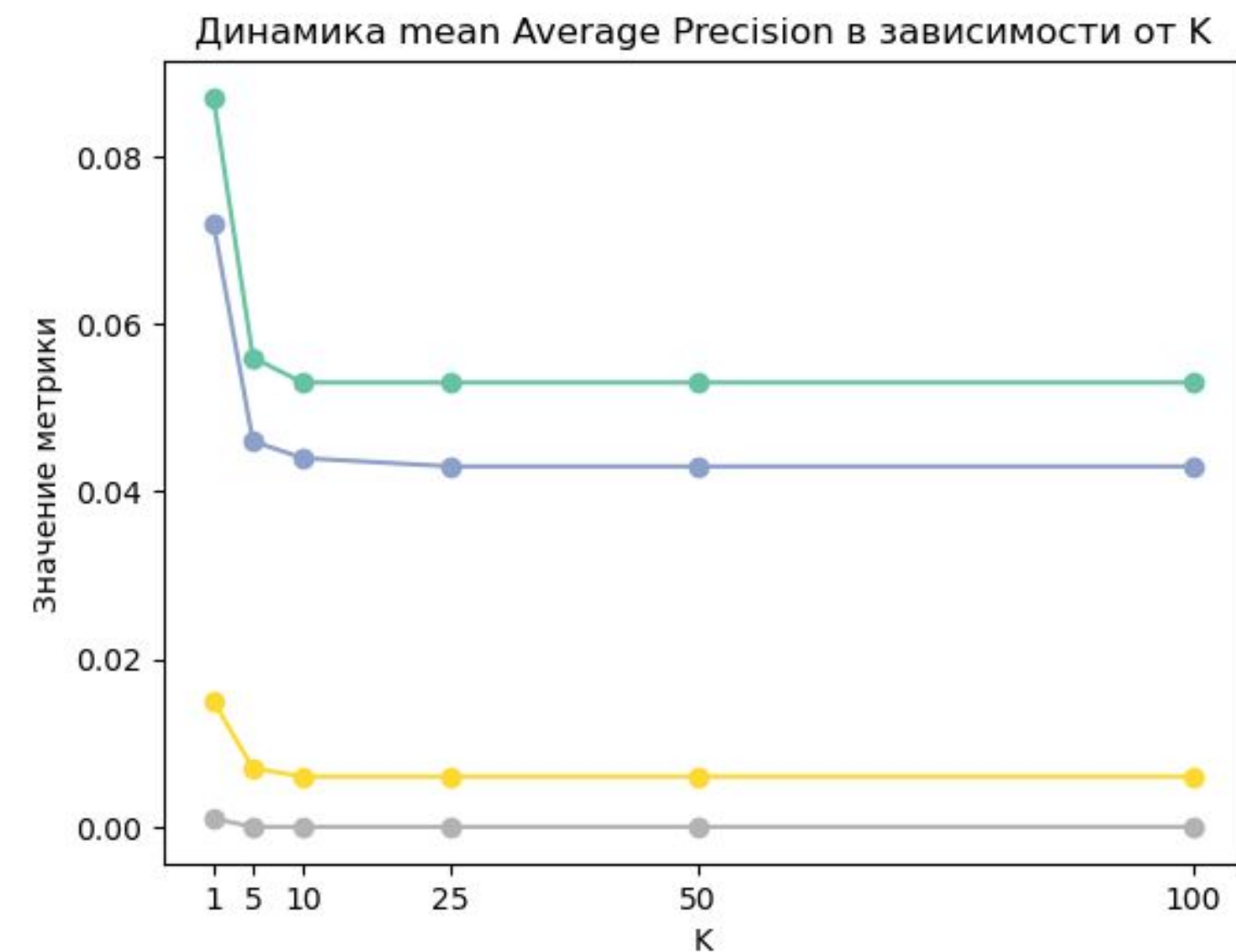
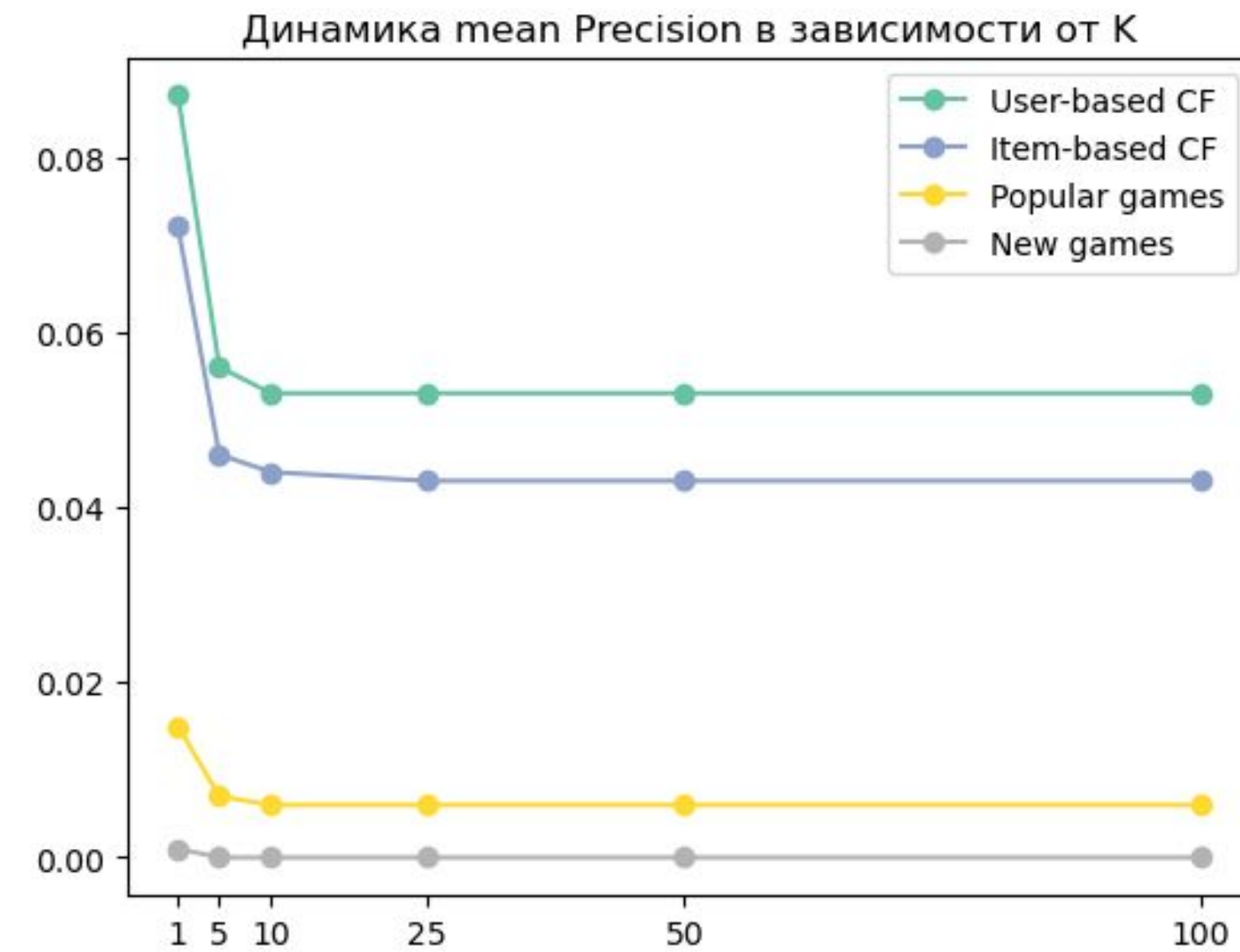
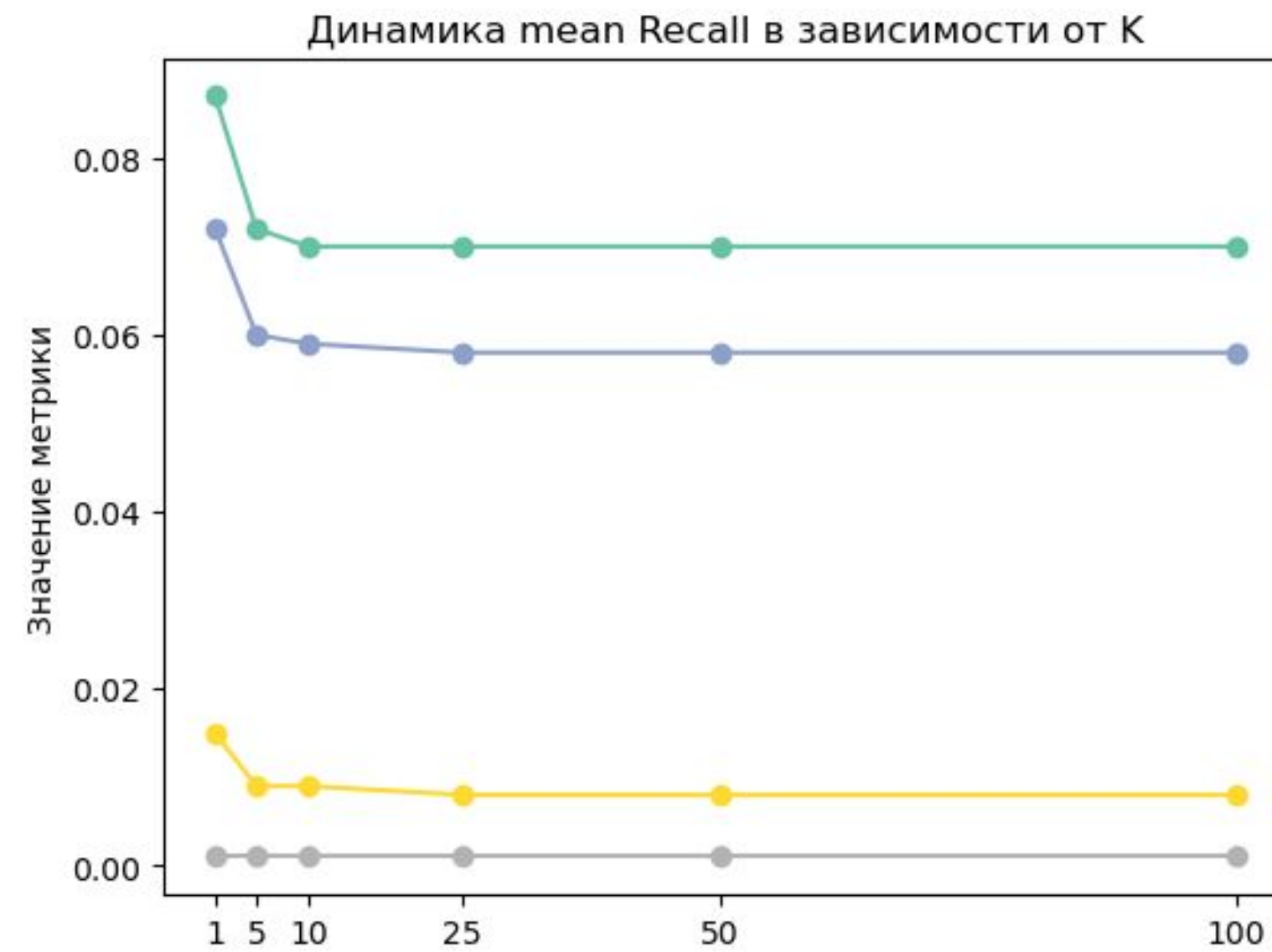
$$MAP@k = \frac{1}{|Q|} \sum_{q \in Q} AP_q@k$$

$$NDCG@k = \frac{DCG@K}{IDCG@K} = \frac{\sum_{i=1}^{k \text{ (actual order)}} \frac{Gains}{\log_2(i+1)}}{\sum_{i=1}^{k \text{ (ideal order)}} \frac{Gains}{\log_2(i+1)}}$$



# Что было сделано: метрики

Построение рекомендательной системы игр на Steam





# Плюсы и минусы бейзлайна

Построение рекомендательной системы игр на Steam

Плюсы:	Минусы:
Коллаборативная фильтрация - простой, но <b>мощный алгоритм</b>	<b>Отзывы менее показательны</b> в качестве взаимодействия с игрой, чем <b>наигранное время</b>
Коллаборативная фильтрация хорошо работает даже при <b>малом количестве</b> любимых игр	Коллаборативная фильтрация <b>не может</b> рекомендовать что-то <b>совсем отличное</b> от того, во что пользователь играл, но что <b>могло бы ему понравиться</b>
Рекомендация популярных игр - <b>отличная стартовая точка</b> для новых пользователей, входящих в гейминг на ПК	Неперсонализированные рекомендации <b>не справляются</b> с рекомендацией интересных игр для пользователя
Рекомендация новых игр - способ открыть для себя <b>набирающие популярность</b> новинки	Рекомендация популярных игр выдает <b>старые, известные практически всем</b> игры





# Что было сделано: telegram бот

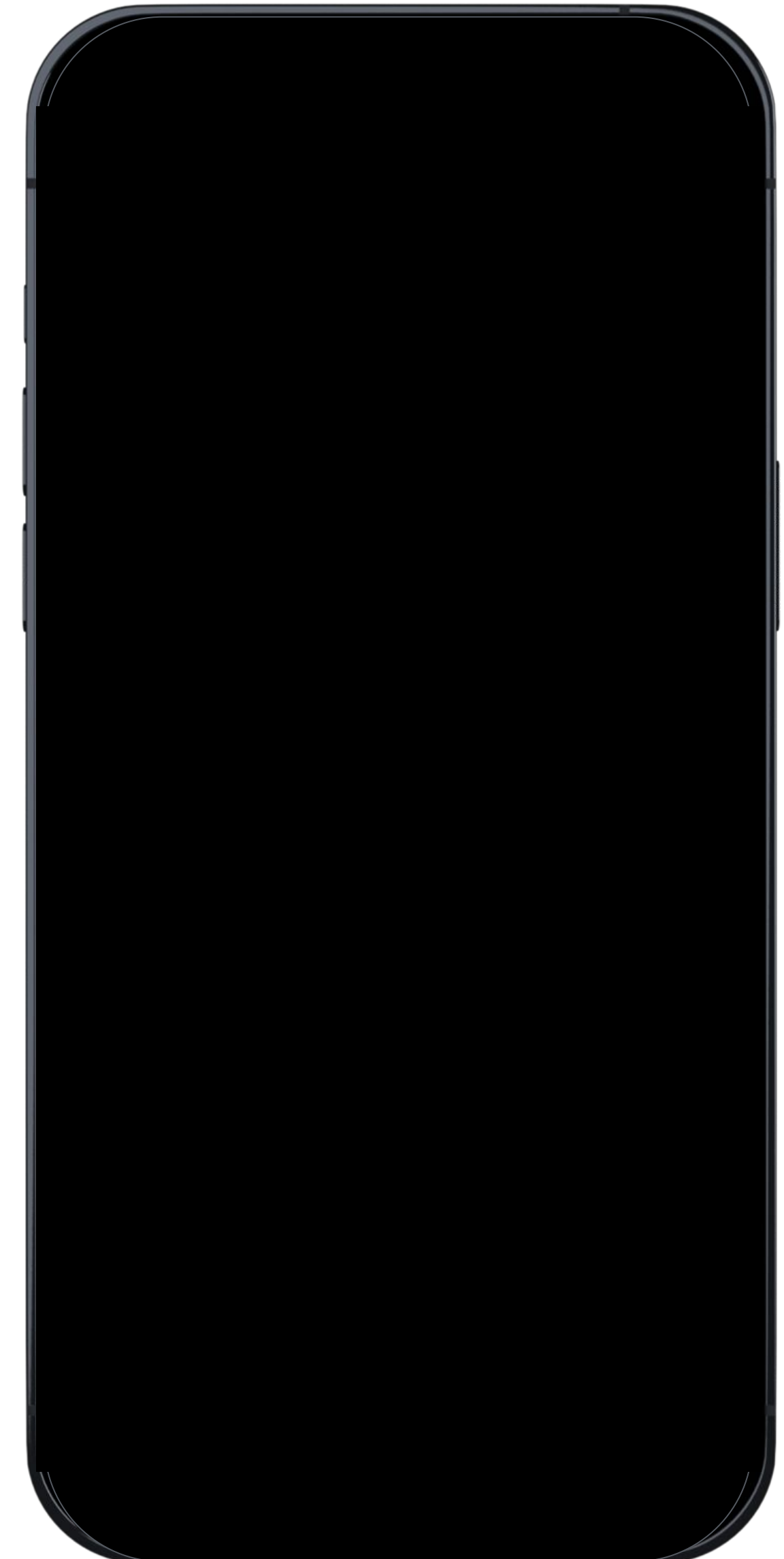
Построение рекомендательной системы игр на Steam

Имя: **Games For You Bot**

Ник: **@games4youbot**

## Функции бота:

- **добавить игру** в список для рекомендаций
- **удалить игру** из списка для рекомендаций
- **получить текущий список** любимых игр
- **очистка списка** любимых игр
  
- **найти информацию** об игре
  
- **установить значение** количества рекомендуемых игр (по умолчанию 10)
- **получить список** рекомендованных игр
- **порекомендовать** игры, **похожие на конкретную игру**
  
- **оставить отзыв** о работе бота
- **вывести подсказку** о функциях бота





# Что сделано к чекпоинту 5

Построение рекомендательной системы игр на Steam

## Бот:

- было проведено тестирование бота пользователями, выявлены направления по исправлению, часть из них уже исправлена (например,
  1. некоторые хорошие игры были отфильтрованы нами как мусорные
  2. не совсем было ясно, как пользоваться ботом
  3. в рекомендациях часто встречаются разные версии одной и той же игры и др.)
- добавлено использование базы данных в боте
- оптимизирована кодовая база



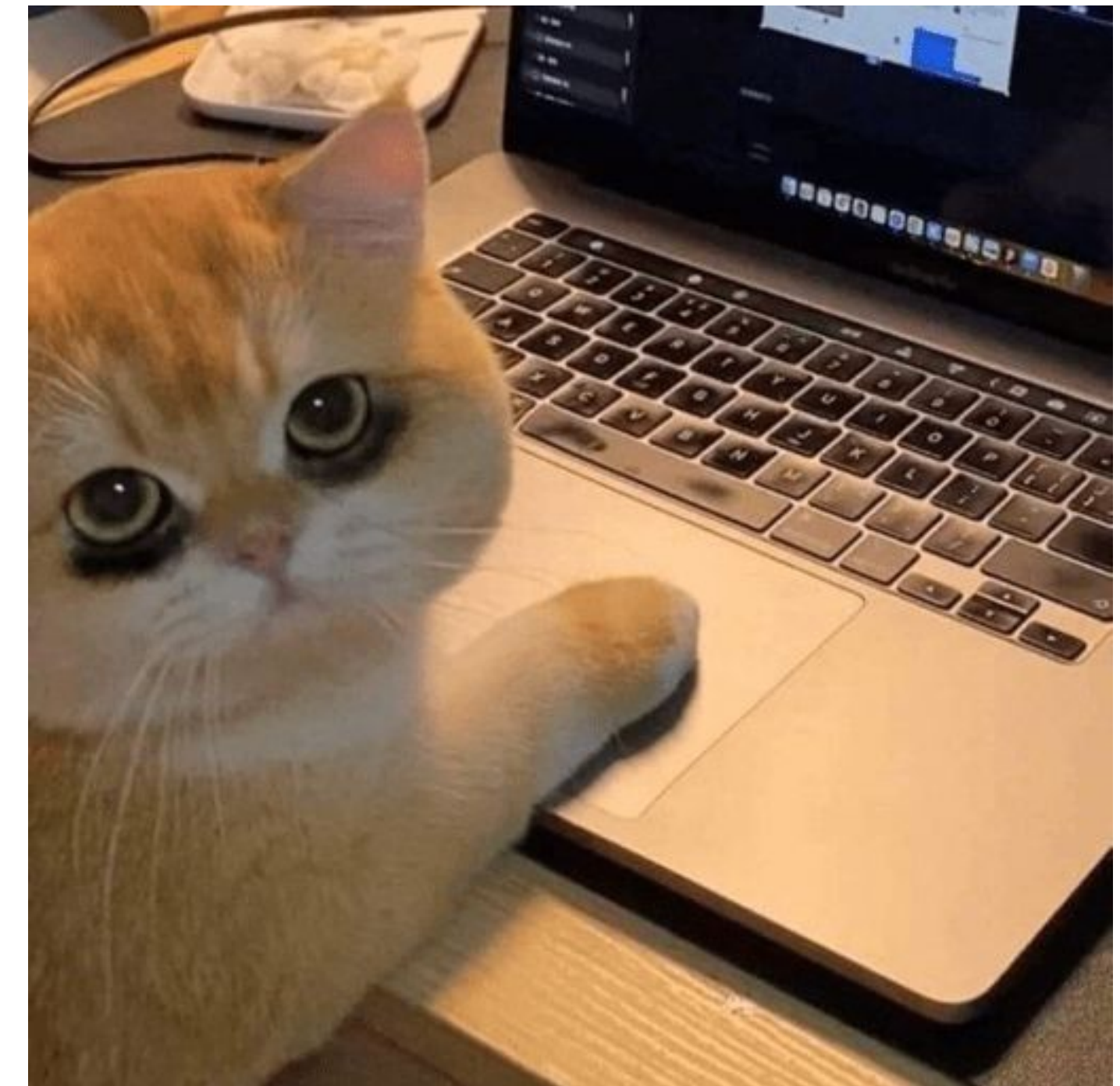


# Что сделано к чекпоинту 5

Построение рекомендательной системы игр на Steam

## Алгоритмы рекомендаций:

- **исправлена фильтрация треш-игр** (фильтрация по количеству отзывов об игре, ранее после фильтрации оставалось  $\approx 8K$  игр, сейчас  $\approx 23K$ )
- **выполнен переход на плейтайм** (выбросы были удалены, значения плейтайм были центрированы по игре, пользователю, выборке в целом)
- **персонализировали рекомендацию новых и популярных игр** (оставлены игры с теми же разработчиками, категориями, тегами, затем уже выбраны самые популярные/новые из них)
- **улучшена организация кодовой базы:** все функции вынесены в отдельные ру файлы







# Что дальше?

Построение рекомендательной системы игр на Steam

- Думаем над тем, как использовать алгоритмы с плейтаймом в боте (основная сложность: получение информации о плейтайме от пользователей)
- Хотим написать функцию, позволяющую регулировать выдачу разных версий одной и той же игры в рекомендациях
- Хотим протестировать другие алгоритмы кроме коллаборативной фильтрации
- Хотим использовать двухступенчатую модель для рекомендаций (первая ступень: уже написанная нами коллаборативная фильтрация, вторая ступень: более сложная модель, например, нейросети)





# Полезные ссылки

Построение рекомендательной системы игр на Steam

## Гитхаб репозиторий:

<https://github.com/avermilov/Steam2Rec>

## Телеграмм бот:

<https://t.me/games4youbot>

## Репозиторий бота:

[https://github.com/ar120/recommendation\\_game\\_bot2](https://github.com/ar120/recommendation_game_bot2)

## Контакты:

Артемий Ермилов (@artermiloff)

Арина Неволина (@nevolinaa)

Ассоль Кубаева (@assolkubaeva)

Артем Поспелов (@posptem)

куратор - Александр Милоградский (@nemexur)











НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ