**Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования**

**«Московский государственный технический университет имени Н.Э. Баумана**

**(национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)**

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

# Отчет по лабораторной работе №2 по курсу

**”Проектирование Рекомендательных Систем”**

**Тема** Сравнение алгоритмов коллаборативной фильтрации

**Студент** Варламова Е. А.

**Группа** ИУ7-33М

**Оценка (баллы)**

**Преподаватели** Быстрицкая А.Ю.

Москва — 2024 г.

**Оглавление**

[Введение](#_bookmark0) [3](#_bookmark0)

1. [Аналитический раздел](#_bookmark1) [4](#_bookmark1)
   1. [Коллаборативная фильтрация](#_bookmark2) [4](#_bookmark2)
   2. [Коллаборативная фильтрация по пользователю](#_bookmark3) [4](#_bookmark3)
   3. [Коллаборативная фильтрация по объекту](#_bookmark4) [4](#_bookmark4)
2. [Конструкторский раздел](#_bookmark5) [5](#_bookmark5)
   1. [MovieLens Rating](#_bookmark6) [5](#_bookmark6)
3. [Технологический раздел](#_bookmark7) [6](#_bookmark7)
   1. [Средства реализации](#_bookmark8) [6](#_bookmark8)
   2. [Библиотеки](#_bookmark9) [6](#_bookmark9)
4. [Исследовательский раздел](#_bookmark10) [7](#_bookmark10)
   1. [Условия исследований](#_bookmark11) [7](#_bookmark11)
   2. [Зависимости времени выполнения коллаборативной фильтрации от парамет-](#_bookmark12)

[ра количества используемых похожих объектов для каждой из мер близости](#_bookmark12) . [7](#_bookmark12)

* 1. [Зависимости значений ROC-AUC от параметра количества используемых по-](#_bookmark15) [хожих объектов](#_bookmark15) [9](#_bookmark15)

[ЗАКЛЮЧЕНИЕ](#_bookmark17) [12](#_bookmark17)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ](#_bookmark17) [13](#_bookmark17)

# ВВЕДЕНИЕ

Цель работы – сравнение алгоритмов коллаборативной фильтрации по пользователю и по объекту.

Для достижения поставленной цели потребуется:

* привести описание алгоритмов;
* привести описание используемых для исследования данных;
* провести сравнение алгоритмов по времени работы и точности пред- сказания.

# Аналитический раздел

## Коллаборативная фильтрация

Коллаборативная фильтрация – это метод рекомендации, при котором анализируется только реакция пользователей на объекты – оценки, которые выставляют пользователи. [[1](#_bookmark18)]

Оценки могут быть как явными, так и неявными (например, длитель- ность нахождения пользователя на странице товара). Целью фильтрации яв- ляется предсказание оценки пользователем по оценкам других. Чем больше имеется оценок, тем точнее получатся рекомендации. [[1](#_bookmark18)]

## Коллаборативная фильтрация по пользователю

В данном методе предполагается, что пользователи, которые в прошлом имели похожие предпочтения, будут иметь похожие предпочтения и в буду- щем. Для построения рекомендаций с использованием этого метода выпол- няются следующие шаги:

1. Создание матрицы, где строки представляют пользователей, а столбцы объекты; значения в ячейках матрицы отражают оценки или действия пользователей по отношению к объектам;
2. Вычисление сходства между пользователями (обычно используются ко- синусное сходство или корреляция Пирсона);
3. Для конкретного пользователя генерируются рекомендации на основе сходства с другими пользователями через агрегацию оценок или дей- ствий похожих пользователей.

## Коллаборативная фильтрация по объекту

В данном методе предполагается, что пользователи будут интересо- ваться теми объектами, которые похожи на уже положительно оцененные объекты. Для построения рекомендаций с использованием метода выполня- ются следующие шаги:

1. Создание матрицы, где строки представляют объекты, а столбцы поль- зователей; значения в ячейках матрицы определяют оценки или дей- ствия пользователей по отношению к объектам;
2. Вычисление сходства между объектами;
3. Генерация рекомендаций с опорой на сходство между объектами, кото- рые пользователь уже оценил.

# Конструкторский раздел

## MovieLens Rating

В качестве источника данных был взят датасет, располагающийся в свободном доступе на веб-сайте MovieLens [[2](#_bookmark19)]. Набор данных включает в себя множество записей, поля которых описывают идентификатор пользова- теля, объект, оценку (от 1 до 5) и время появления рецензии. Предобработки проводить не потребовалось.

# Технологический раздел

[[3](#_bookmark20)].

## Средства реализации

В качестве используемого был выбран язык программирования Python

Данный выбор обусловлен следующими факторами:

* Большое количество исчерпывающей документации;
* Широкий выбор доступных библиотек для разработки;
* Простота синтаксиса языка и высокая скорость разработки.

При написании программного продукта использовалась среда разра-

ботки Visual Studio Code. Данный выбор обусловлен тем, что данная среда распространяется по свободной лицензии, поставляется для конечного поль- зователя с открытым исходным кодом, а также имеет большое число расши- рений, ускоряющих разработку.

## Библиотеки

При анализе и обработке датасета, а также для решения поставленных задач использовались библиотеки:

* + pandas;
  + numpy;
  + matplotlib;
  + sklearn [[4](#_bookmark21)];
  + LibRecommender [[5](#_bookmark22)].

Данные библиотеки позволили полностью покрыть спектр потребно- стей при выполнении работы.

# Исследовательский раздел

## Условия исследований

Исследование проводилось на персональном вычислительной машине со следующими характеристиками:

* + процессор Intel Core i5,
  + операционная система MacOs Big Sur,
  + 8 Гб оперативной памяти.

Временные затраты определялись с использованием библиотеки time.

Значение ROC-AUC определялось с использованием библиотеки Lib- Recommender.

В данном исследовании значение параметра, отвевающего за количе- ство используемых похожих объектов, менялось от 1 до 30. Также для каж- дого алгоритма и значения параметра использовались косинусная мера, кри- терий Пирсона и мера Жаккара.

## Зависимости времени выполнения коллаборативной фильтра- ции от параметра количества используемых похожих объек- тов для каждой из мер близости

На рисунке [4.1](#_bookmark13) представлен график зависимости времени выполнения коллаборативной фильтрации по пользователю и по объекту от параметра количества используемых похожих объектов для каждой из мер близости.

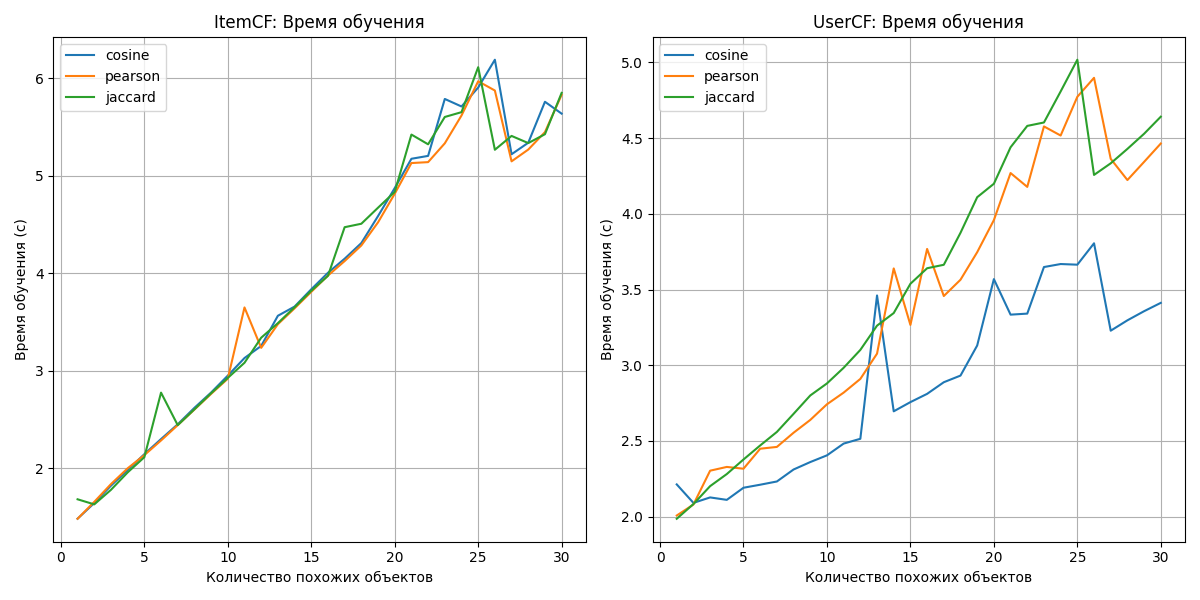


Рис. 4.1: График зависимости времени выполнения коллаборативной филь- трации по пользователю и по объекту от параметра количества используемых похожих объектов для каждой из мер близости.

## Зависимости значений ROC-AUC от параметра количества ис- пользуемых похожих объектов

На рисунке [4.3](#_bookmark16) представлен график зависимости ROC-AUC от параметра количества используемых похожих объектов для каждой из мер близости.

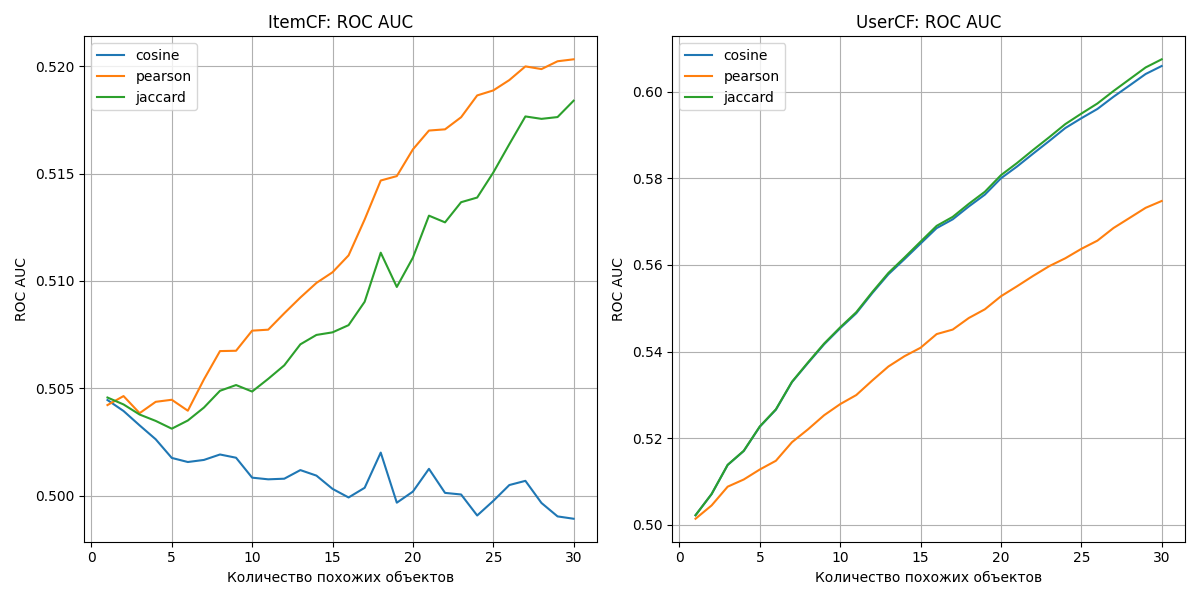


Рис. 4.2: График зависимости ROC-AUC от параметра количества используемых похожих объектов для каждой из мер близости.

# ЗАКЛЮЧЕНИЕ

В ходе выполнения работы было проведено сравнение алгоритмов кол- лаборативной фильтрации по пользователю и по объекту.

Были решены следующие задачи:

* приведено описание алгоритмов;
* приведено описание используемых для исследования данных;
* проведено сравнение алгоритмов по времени работы и точности пред- сказания.

В результате проведенных исследований заметно, что при фильтрации по пользователю, лучше всего по времени себя показала косинусная мера на значениях параметра больше 5. В случае фильтрации по объекту, сложно вы- делить какую-то из мер близости.

Также стало понятно, что на рассматриваемом датасете с точки зрения ROC-AUC большие значения количества похожих объектов ведут к улучшению качества рекомендаций.

# Список литературы

1. Владимировна Смоленчук Татьяна. Метод коллаборативной фильтрации для рекомендательных сервисов // Вестник науки и образования. 2019.

№ 22-1.

1. MovieLens [Электронный ресурс]. Режим доступа: [https:](https://grouplens.org/datasets/movielens/)

[//grouplens.org/datasets/movielens/](https://grouplens.org/datasets/movielens/) (дата обращения 16.09.2023).

1. Python official page [Электронный ресурс]. Режим доступа: [https://](https://www.python.org/) [www.python.org/](https://www.python.org/) (дата обращения 10.05.2023).
2. Scikit-learn official page [Электронный ресурс]. Режим доступа: [https:](https://scikit-learn.org/stable/)

[//scikit-learn.org/stable/](https://scikit-learn.org/stable/) (дата обращения 10.05.2023).

1. LibRecommender: official PyPl project page [Электронный ресурс]. Режим доступа: <https://pypi.org/project/LibRecommender/> (дата обращения 16.09.2023).