

МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

**РЕАЛИЗАЦИЯ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ ДЛЯ  
ИНТЕРНЕТ-МАГАЗИНА**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студентки 4 курса 411 группы  
направления 02.03.02 — Фундаментальная информатика и информационные  
технологии  
факультета КНиИТ  
Ситниковой Русланы Эдуардовны

Научный руководитель  
ст. преподаватель

\_\_\_\_\_

М. И. Сафрончик

Заведующий кафедрой  
к. ф.-м. н., доцент

\_\_\_\_\_

С. В. Миронов

Саратов 2023

## ВВЕДЕНИЕ

Объем информации в современном мире достаточно высок, так что пользователи цифровых систем не могут себе позволить охватить всю предлагаемую информацию целиком и вынуждены ограничиваться некоторыми выборками. С другой стороны, владельцы данных систем должны привлекать и удерживать пользователей, удовлетворяя их интересам и конкурируя при этом с другими системами.

Вручную проанализировать каждого пользователя крупной системы невозможно, так как их число может достигать тысяч, десятков тысяч или даже сотен. Решением данной проблемы стало создание рекомендательных алгоритмов, позволяющих на основе данных о пользователях или объектах системы найти предложение, которое может заинтересовать конкретного пользователя. Существуют различные виды рекомендательных систем, у каждого из них есть свои преимущества и недостатки.

Целью рекомендательной системы является предоставление пользователю независимой оценки объекта, который может быть полезен для него, или составление предположительного мнения пользователя обо всех объектах системы, чтобы помочь в принятии решения. Обобщенная цель рекомендательной системы — это вычисление предполагаемой оценки пользователя для объекта, который ему еще неизвестен.

Чаще всего алгоритмы рекомендательных систем обобщают и анализируют знания о пользователях, например, поведение пользователя, или об объектах системы. В настоящее время практически каждый крупный сервис использует рекомендательные системы для работы с пользователями, так как это способствует повышению конкурентоспособности.

Целью выпускной квалификационной работы является создание товарной рекомендательной системы для интернет-магазина одежды.

Для этого были поставлены следующие задачи:

- поиск необходимого набора данных, который сочетает в себе данные о товарах и покупках пользователей;
- предварительная обработка данных;
- реализация пользовательского интерфейса в виде веб-сайта интернет-магазина;
- реализация рекомендательной системы для предложения пользователям

товаров.

**Структура и объем работы.** Для решения поставленных задач выполнена выпускная квалификационная работа, включающая в себя введение, 2 основные главы, заключение, список использованных источников из 20 наименований и 2 приложения. Работа изложена на 62 страницах, содержит 30 рисунков и 2 таблицы.

Первая глава имеет название «Теоретические аспекты создания рекомендательных систем» и содержит основную информацию о рекомендательных системах: определения, виды, методы оценки.

Вторая глава имеет название «Реализация товарной рекомендательной системы для интернет-магазина», данная глава содержит подробное описание процесса выполнения работы.

Выпускная квалификационная работа заканчивается заключением, списком использованных источников, а также приложениями с кодом А-Б.

## **Основное содержание работы**

**Рекомендательные системы.** Рекомендательная система — это система, позволяющая формировать предложения в соответствии с интересами и предпочтениями пользователей. Рекомендательные системы не являются статичными и постоянно подстраиваются под пользователя, а также перестраиваются в зависимости от получаемых данных.

Базовыми сущностями рекомендательных систем являются пользователь и объект, зависимый от предметной области применения системы. Объектом может быть предмет одежды, иной товар, книга или новость. Задачей рекомендательной системы является определение объекта, не используемого пользователем в течение какого-либо промежутка времени или не используемого вовсе, но полезного для него в текущем контексте.

Основные понятия, связанные с рекомендательными системами.

- Предметом рекомендации является объект, который нужно предложить пользователю. В данной работе предметом рекомендации является товар из каталога магазина.
- Цель рекомендации. Для интернет-магазина целью использования рекомендательной системы является увеличение продаж и удержание аудитории.
- Контекст рекомендации — действия пользователя в определенный момент времени. Контекстом рекомендации в данной работе является добавление пользователем товара в корзину.
- Источник рекомендации. Источником рекомендации является сам интернет-магазин.

Рекомендации также могут различаться по степени персонализации:

- неперсональные рекомендации для всех пользователей одинаковы;
- персональные рекомендации основаны на информации о клиенте, для этого может быть использована история покупок пользователя.

**Виды рекомендательных систем и основные определения.** Существует несколько типов рекомендательных систем:

- Неперсонализованные. В такой системе все рекомендации являются неперсональными, пользователям предлагаются исключительно популярные товары.
- Коллаборативная фильтрация. Существует два вида фильтрации: в первом

случае происходит отбор пользователей, похожих на того, для которого нужно найти рекомендацию, после чего предлагаются товары, купленные найденными пользователями, во втором случае — поиск товаров, похожих на купленные пользователем.

- На основе содержания — рекомендации основаны на сходстве товаров с теми, которые пользователь приобретал ранее. Сходство определяется по признакам товара, например, категория или наличие общих тегов.
- Гибридные системы — системы, сочетающие в себе несколько типов. Гибридные системы могут помочь решить проблему «холодного старта», когда данных не хватает — сначала с помощью, например, системы рекомендаций на основе содержания нужно отобрать похожие товары, а затем применить к полученным данным коллаборативную фильтрацию. Данные системы являются сложными в построении. Обычно такой тип систем основан на коллаборативном подходе, анализе контента и знаниях о предметной области. В гибридных системах помимо вышеуказанных методов может быть использован искусственный интеллект.
- Методы, основанные на знаниях. Данные методы построения рекомендательных систем могут быть использованы в тех случаях, когда недостаточно информации о предыдущих покупках пользователя, например, при покупке жилья или транспорта.

Для коллаборативной фильтрации и рекомендательных систем на основе содержания нужно найти похожие друг на друга товары. Для расчета схожести можно воспользоваться косинусным расстоянием, представленным формулой 1.

$$\cos\theta = \frac{\bar{a}\bar{b}}{\|\bar{a}\|\|\bar{b}\|} = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}}, \quad (1)$$

где  $\bar{a}$  и  $\bar{b}$  — два объекта, между которыми нужно вычислить расстояние.

Для поиска похожих друг на друга пользователей можно воспользоваться алгоритмом  $k$ -ближайших соседей. Для этого нужно сначала с помощью одной из мер расстояния, например, приведенной выше косинусной меры, вычислить расстояние от данного пользователя до других, затем отобрать из них  $k$  наиболее близких, далее получить предполагаемую оценку для товаров, которые можно предложить в качестве рекомендации, и отобрать товары с наибольшей оценкой.

Для вычисления предполагаемой оценки можно воспользоваться формулой 2.

$$\bar{r}_{aj} = \frac{1}{\sum_{i \in N(a)} s_{ai}} \sum_{i \in N(a)} s_{ai} r_{ij}, \quad (2)$$

где  $\bar{r}_{aj}$  — предполагаемая оценка  $j$ -го товара пользователем  $a$ ,  $s_{ai}$  — коэффициент схожести  $a$  на  $i$ ,  $r_{ij}$  — оценка  $j$ -го товара  $i$ -м пользователем.

Для рекомендательных систем можно построить так называемую «матрицу предпочтений». Количество строк матрицы предпочтений равно количеству пользователей в системе, а количество столбцов — количеству объектов. Матрица заполняется оценками пользователей, которые могут быть числовыми, бинарными или унарными.

**Контекст.** В рекомендательных системах также существует понятие контекста — условий, при которых пользователь выбирает объект или дает ему оценку. Контекст имеет значительную роль для поиска подходящих для рекомендации объектов. Существует несколько видов контекста:

- физический контекст: время, местоположение, общие данные о пользователе;
- социальный контекст: роль окружающих людей;
- контекст устройства: характеристики устройства, которое пользователь использует для получения информации;
- модальный контекст: данные о целях пользователя, его настроении и тому подобное.

Контекст можно использовать в дополнение к алгоритмам коллаборативной фильтрации для отбора входных или выходных значений.

**Оценка рекомендаций.** Чтобы оценить качество рекомендаций, нужно сравнить предполагаемую оценку товара, предсказанную системой, а также фактическую оценку данного товара пользователем. Для этого можно воспользоваться следующими метриками.

- Средняя абсолютная ошибка (MAE). Это модуль разницы между фактической ошибкой и предсказанной. Вычисление средней абсолютной ошибки приведено на формуле 3. Чем меньше вычисленное значение метрики, тем меньше разница между предсказанной и фактической оценкой.
- Средняя квадратичная ошибка (RMSE). Это квадратный корень из суммы квадратов разниц фактической и предсказанной ошибки, пример вычис-

ления приведен на формуле 4. Так же, как и для MAE, чем меньше вычисленное значение метрики, тем меньше разница между предсказанной и фактической оценкой.

$$MAE = \frac{\sum_{i \in n} |y_i - p_i|}{n}, \quad (3)$$

где  $n$  — количество оценок,  $y_i$  — фактическая оценка товара,  $p_i$  — предсказанная оценка товара.

$$RMSE = \sqrt{\frac{\sum_{i \in n} (y_i - p_i)^2}{n}}, \quad (4)$$

где  $n$  — количество оценок,  $y_i$  — фактическая оценка товара,  $p_i$  — предсказанная оценка товара.

**Клиент-серверная архитектура.** Клиент-серверная архитектура является часто используемой в настоящее время. Данная архитектура позволяет экономить ресурсы клиента, а также позволяет ему использовать различные устройства для получения информации и доступа к ресурсам сервера, например, компьютер или мобильный телефон. Серверная часть может быть представлена облачными вычислениями.

Клиенты в данной архитектуре всегда посылают запрос на сервер, а сервер отправляет клиенту ответ на полученный запрос. Преимуществами клиент-серверной модели являются:

- упрощение передачи информации между клиентом и сервером;
- возможность разделения процессов по различным физическим машинам;
- уменьшение повторения информации путем хранения ее на серверном устройстве, а не на устройстве клиента.

**Реляционные базы данных.** Информацию, используемую в веб-приложениях нужно хранить в наиболее удобном и доступном виде, чтобы избежать излишних затрат ресурсов на ее получение, обработку и передачу. Одним из способов хранения и использования информации являются реляционные базы данных.

База данных — это унифицированная совокупность данных, совокупно используемая широким кругом пользователей, которая динамически изменяется в процессе функционирования. База данных называется реляционной, если форматом представления данных является таблица.

**Реализация товарной рекомендательной системы для интернет-магазина.** В ходе работы создан интернет-магазин с функцией рекомендации товаров на основе добавленных в корзину предметов. Интерфейс магазина представлен в виде веб-сайта.

В работе используется два датасета. Один из них изначально содержит информацию о товарах: название, цену, производителя и так далее. Вторым датасетом содержит информацию о заказах покупателей: идентификатор товара, дату, имя пользователя, оценку товара.

Для создания интернет-магазина используется язык программирования Python версии 3.11.3, для управления Python-модулями также использован инструмент Poetry.

**Работа с датасетами.** В работе были использованы два датасета, один из которых содержит информацию о покупках пользователей, второй — информацию о товарах. Была произведена стандартная обработка датасетов, а также составлена матрица предпочтений пользователей, используемая для вычисления коэффициента близости между пользователями. Вторым датасетом был преобразован путем связывания идентификаторов товаров с датасетом, содержащим информацию о товарах.

Используется следующая информация о товарах:

- идентификатор продукта;
- оригинальное название;
- цена;
- валюта;
- количество проданных товаров;
- рейтинг товара;
- количество оставленных оценок;
- цвет товара;
- размер товара;
- страна производства;
- название производителя.

А также следующая информация о заказах:

- идентификатор продукта;
- идентификатор пользователя;
- рейтинг товара, поставленный пользователем;



- дата совершения заказа.

**Создание интернет-магазина.** Для создания интернет-магазина был создан Poetry-проект, использующий фреймворк Django. К проекту подключена база данных SQLite, в которой содержится информация о товарах, а также матрица предпочтений пользователей.

Интернет-магазин состоит из следующих страниц:

- главная страница;
- страница авторизации;
- страница каталога товаров с возможностью фильтрации и сортировки объектов;
- страница определенного товара, на которой авторизованный пользователь может видеть последнюю оценку данного товара;
- корзина с добавленными и рекомендованными товарами.

**Получение рекомендаций для зарегистрированного пользователя.** Механизм расчета рекомендаций, можно представить следующей последовательностью шагов.

1. Поиск  $k$ -ближайших соседей к заданному пользователю. В данной работе  $k$  равно 10.
2. Получение оценок товаров, купленных отобранными пользователями.
3. Калибровка оценок в зависимости от расстояния до пользователя, которому нужно выдать рекомендацию.
4. Получение предполагаемых оценок для товаров, купленных ближайшими соседями.
5. Выбор товаров с максимальной предполагаемой оценкой в качестве рекомендации.
6. Если данный товар уже был куплен пользователем — выбор следующего товара до тех пор, пока не будут выбраны товары, которые пользователь еще не покупал.

Для поиска ближайших соседей используется косинусное расстояние.

**Получение рекомендаций для анонимного пользователя.** В случае анонимного пользователя происходит следующая последовательность действий.

1. Происходит поиск всех пользователей, купивших товар, для которого нужно найти подходящую рекомендацию.
2. Для каждого пользователя подсчитывается количество покупок каждого

из купленных им товаров.

3. Полученный словарь с подсчетом количества покупок сортируется по убыванию.
4. Среди полученных товаров отбирается тройка самых популярных, то есть купленных наибольшее количество раз.
5. Если по каким-либо причинам три товара отобрать не удалось, недостающее количество закрывается самыми популярными товарами.
6. Набирается три списка пользователей, купивших соответствующие выбранные товары.
7. Происходит поиск трех соответствующих списков всех товаров, купленных пользователями из списков, полученных в предыдущем пункте.
8. Происходит поиск пересечения списков товаров, полученных в предыдущем пункте.
9. Если пересечение найдено, то товары из пересечения возвращаются как рекомендации.
10. Если пересечение не найдено, происходит поиск пользователя, купившего товар из корзины несколько раз, а затем выбирается максимум три товара среди его последних покупок, отсортированных по убыванию рейтинга. Если три товара не набралось, недостающих остаток заполняется самыми популярными товарами среди всех товаров магазина.
11. Если же в предыдущем пункте не удалось найти нужное пересечение, то в качестве рекомендации выдается три самых популярных товара среди всех товаров магазина.

**Оценка рекомендательной системы.** Для оценки реализованной рекомендательной системы была отобрана случайная выборка, составленная из части пользователей системы, для которых была рассчитана рекомендация одного товара — первого, среди предложенных в наборе. В большинстве случаев данный товар является уже купленным, то есть имеет оценку, данную пользователем. Таким образом, можно вычислить MAE и RMSE метрики для системы.

На случайной выборке из 100 пользователей система имеет следующие показатели:

- значение метрики MAE равно 0,499;
- значение метрики RMSE равно 0,634.

При этом в 22 случаях произошло завышение предполагаемой оценки, в

61 — занижение и в 17 — равенство оценок.

Таким образом, система при предсказании оценки пользователя имеет погрешность примерно в половину балла рейтинга в меньшую сторону, чаще занижая предполагаемую оценку.

## ЗАКЛЮЧЕНИЕ

В ходе работы были проанализированы некоторые виды рекомендательных систем, а также способы их оценки. Была разработана рекомендательная товарная система для интернет-магазина одежды, сочетающая в себе алгоритмы коллаборативной фильтрации, а также методы составления рекомендаций на основе содержания, для чего была использована информация об объектах системы.

Данная система может быть улучшена путем оптимизации использования таких ресурсов, как память. Другим возможным вариантом является распределение системы между несколькими вычислительными ресурсами для ускорения вычислений.

В ходе решения задач поиска и обработки необходимого набора данных были использованы такие методы обработки данных, как поиск пропусков и выбросов, кодирование номинальных признаков, работа с реляционными структурами данных.

В ходе решения задачи реализации пользовательского интерфейса были получены навыки разработки Django-проекта.

В ходе решения задачи реализации рекомендательной системы были проанализированы некоторые виды рекомендательных алгоритмов, а также способы их оценки. Разработанная система при выставлении оценки ошибается примерно на половину балла рейтинга в сторону занижения предполагаемой оценки.

Таким образом, поставленные задачи решены в полном объеме и достигнута поставленная цель — была разработана рекомендательная товарная система интернет-магазина одежды.