МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «МИСИС»**

Институт компьютерных наук

Кафедра инженерной кибернетики

Направление подготовки: «09.04.03 Прикладная информатика»

Квалификация: **магистр**

Группа: **МПИ-22-1**

**Курсовая работа**

на тему:

**«Рекомендательная система «Интернет – магазин одежды»**

учебная дисциплина

«ЭКСПЕРТНЫЕ И РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ»

**Выполнили:**

**Студенты: Несмеянова Юлия Васильевна**

**Проверил:** д.т.н., профессор кафедры инженерной кибернетики Соловьев Александр Владимирович

**Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Дата: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Москва 2023**

**Введение**

Рекомендательные системы — это алгоритмы и программы, которые анализируют предпочтения пользователей различных систем и пытаются предсказать, что может им понравиться и на этом основании предоставить рекомендацию. Рекомендательные системы шикороко импользуются в различных сервисах для рекомендаций товаров, услуг, контента и т. д.

В эпоху цифровизации и особенно в пост-ковидный период все больше сфер человеческой жизни переходит в онлайн-формат. В большом количестве информации, значение рекомендательных систем особенно важно как для владельцев сервисов, так и для пользователей. В онлайн - продажах такие системы позволяют сократить время поиска и найти необходимые товары для пользователей, а также увеличить прибыль для продавца.

Цель данной курсовой работы - рассмотреть подходы к рекомендательным системам и реализовать свою модель для онлайн-машазина одежды.

Задачи:

- изучить теоретические основы построения рекомендательных систем;

- рассмотреть особенности предметной области;

- рассмотреть существующие реализации рекомендательных систем в онлан-продажах;

- реализовать свою рекомендательную систему с использованием языка python

Результат работы - рекомендательна система с использованием языка python и его библиотек для формирования рекомендаций для интернет-магазина одежды.

**2. Методы и подход к разработке рекомендательной системы**

**2.1 Специфика предметной области**

Любая рекомендальная система имеет собственные особенности в засимости от области для которой она была построенна. При построение рекомендательной системы для магазина одежды необходимо учитывать две основные особенности:

1) Специфику рекомендательных систем в области онлайн-ритейла

2) Особенности предметной области “одежда”

Под понятием “онлайн - ритейл” в первую очередь понимают процесс продажи товаров через интернет[2]. Такая форма торговли имеет большое количество особенностей и преимуществ. В частности, она позволяет охватить очень большое количество потенциальных клиентов в независимости от расстояния от их территориального расположения до физического расположения склада товаров. Крупные международные площадки предоставляют продавцам возможность торговать фактически по всему миру. При таком количестве потенциальных клиентов, просто невозможно использовать для всех стандарные рекомендации товаров. Использование рекомендательных систем может значительно влиять на итоговые показатели продаж. По данным консалтинговой компании McKinsey 35% выручки компании Amazon составляет выручка от продаж товаров, рекомендованных алгоритмами[3]. Поэтому для рекомендательных систем в интернет-продажах важно рекомендовать не только товар, которые потенциально может понравиться пользователю, но и который он может купить.

Предметная область одежда так же имеет особенности, которые отличают ее от, например, простых товаров повседневного спроса. Основные моменты, на которые стоит обращать внимание в данном случае:

1) Пол и возвраст покупателя - являются главными факторами для выбора одежды. Данные параметры чаще всего предлагается заполнить в профиле пользователя, однако, если они не указаны их нужно предсказать. Маленьким детям одежду чаще всего покупают мамы, поэтому также необходимо выделять женщин с детьми.

2) Размеры - ключевой фактор в онлайн продажах одежды. По некоторым данным, возврататы товаров в этой области могут достигать до 70% от общего объема продаж и главной причиной возврата покупатели чаще всего указывают неверный или неподошедший размер[4]. Использование рекомендательных систем может помочь снизить этот процент.

3) Ценовая категория - также имеет большое значение. В индустрии моды ценовые различия на схожие позиции могут достигать существенных значений. Брендовая одежда, в особенности сумки и обувь могут считаться отдельной категорией, так как они ближе к предметам роскоши или коллекционирования. Рекомендовать их на общем основании по схожисти стиля или размера не самая эффективная стратегия.

4) Лояльность бренду - для некоторых пользователей играет определенную роль. Даже в рамках одной ценовой категории есть покупатели, которые имеют любимые или предпочитаемые марки.

5) Модные тенденции - также имеют значения для некоторых категорий покупателей. Таким пользователям эффективнее предлагать самые популярные и покупаемые товары за определенный период.

6) Сезонность - очевидный факт, что купальники лучше продаются в июле, чем в декабре.

**2.2 Подходы к созданию системы**

Для построение примера собственной рекомендательной системы был выбран фрагмент датасета Amazon ReviewData

Amazon ReviewData – многомиллионный набор обзоров, рейтингов и метаданных продуктов (описание, категория, цена, бренд, характеристики, фото), а также данные о просмотре ссылок. Общее количество обзоров в этом наборе данных более 233 тысяч миллионов[5].

В представленном датасете имеется несколько вариантов представления информации:

1) Полные датасеты со всей информацией;

2) Небольшие подмножества для экспериментов;

3) Метаданные для каждой категории;

4) Отфильтрованные подмножества, в которых все пользователи и элементы имеют не менее 5 отзывов.

Для работы были взяти небольшие подмножетва с отзывами и характеристиками:

1) Датасет с 17 ,2 тыс. уникальных отзывов;

2) Датасет с характеристиками 1,1 тыс. уникальных товаров

Исходный датасет, на рисунке 1, с отзывами содержит большое количество столбцов с информацией о покупателе, товаре, оценке, времени отзыва, тексте отзыва и т.д.



Рисунок - 1 Исходный датасет Amazon ReviewData

Однако для построения рекомендательной системы нам необходимы только оценки, поэтому датасет был преобразован в матрицу оценок следующего вида (рис 2). Где id покупателя указано в первом столбце и id товаров указано по колонкам, на пересечение проставленны оценки.

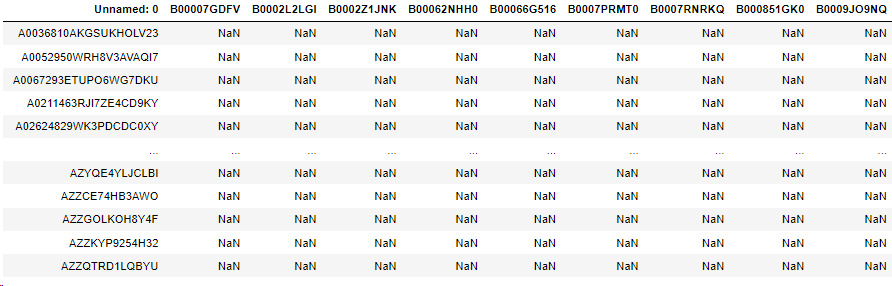


Рисунок 2 - Преобразование датасета с отзывами

Так как количество товаров и количество пользователей сильно отличается, то данная матрица сильно разрежена (максимум товаров, оценненым одним пользователем - 6). Однако, сжатое хранение не потребовалось и рекомендательная система хорошо работает на данной матрице.

Второй предствленный датасет содержит информацию о товарах. Исходный вид датасета продемонстрирован на рисунке 3 где json-файл с различными характеристиками.

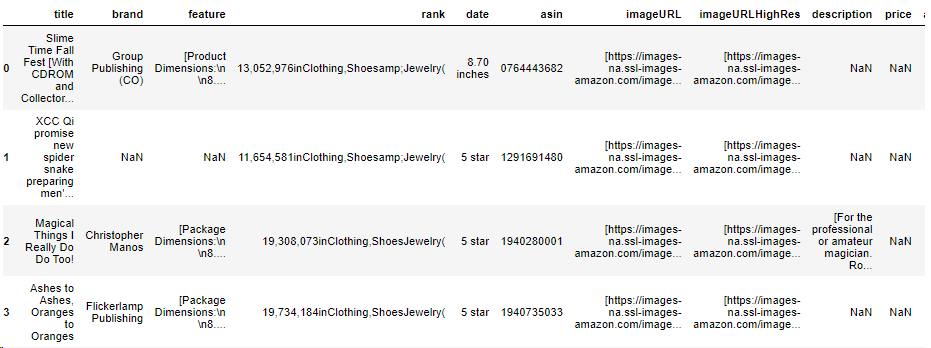


Рисунок 3 - Исходный датасет с описаниями товаров

Для использования в рекомендательной системе второй датасет был приведен к следующему формату как показано на рисунке 4.

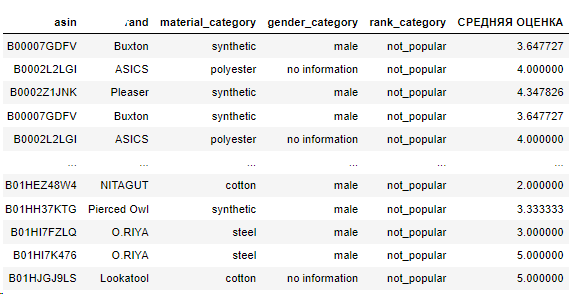


Рисунок 4 - Преобразование датасета с характеристиками товаров

В качестве характеристик остались наиболее важные параметры для данной предметной области, которые были обозначены в разделе 2.1. - бренд, материалы, половая пренадлежность, популярность и отдельна выведена средняя оценка.

Имеющиеся два набора данных позваоляют подойти к рекомендательной системе с нескольких подходов:

1) Коллаборативная фильтрация

2) Основанные на контенте

3) Гибридные

Для реализации был выбран гибридный подход, так как он позволяют совместить преимущества систем основанных на пользователях и на товарах.

**3. Описание разработки**

Разработка велась путем написания функций на python. На первом этапе на основании подготовленной матрицы оценок рассчитывались близкие друг другу покупатели и гипотетические оценки для предполагаемых товаров.

В качестве меры расстояния был выбран квадрат Евклидова расстояния, так как данный способ подсчета близости хорошо себя показывает, если нужно придать большие веса более отдаленным друг от друга объектам [6]. Что в нашем случае верно, так как большая часть пользователей покупает разные товары. Формула имеет следующий вид:

Таким образом, на основании схожести между пользователями реализована первая функция - def recommend(start\_matrix, list\_ids, items). На вход данная функция принимает исходную матрицу, а так же списки id пользователей и товаров в том же порядке. На основании выбранной метрики расстояния вычисляются схожие пользователи и для конкретно указанного покупателя и товара вычисляется предполагаемая оценка как показано на рисунке 5.

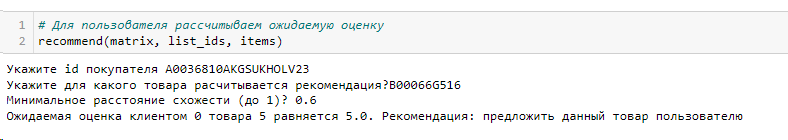


Рисунок 5 - Результат работы функции recommend ()

Однако такой подход не очень удобен, поскольку необходимо задавать как пользователя, так и товар.

Для того что бы формировать общие рекомендации для пользователя реализована вторая функция. Функция основанная на характеристиках товара ищет схожие товары по имеющимся характеристикам. Характеристики товаров также представляют собой матрицу вида:

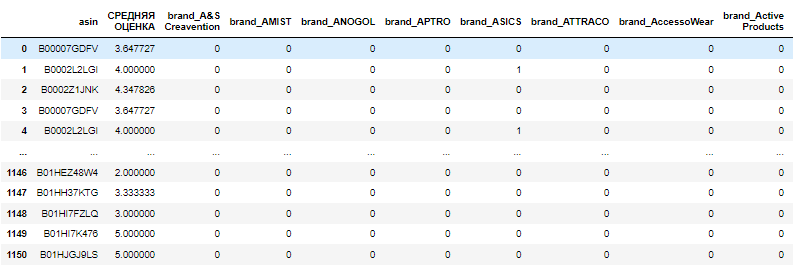


Рисунок 6 - Матрица характеристик товаров

Для конкретного пользователя по уже использованной матрице оценок вычисляется самый любимый товар. Затем для этого товара выбираются пять наиболее похожих. Среди наиболее похожих товаров также отфильтровываются товары с наиболее высокой оценкой – они и составляют рекомендацию.

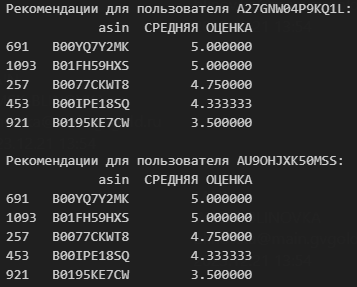


Рисунок 7 – Пример вывода рекомендаций для каждого пользователя

Таким образом как показано на рисунке 7 для каждого пользователя формируется пять схожих товаров со средней оценкой рекомендации.

**4. Руководство пользования и администрирования.**

Чтобы обеспечить высокую производительность, асинхронность обработки запросов и поддержку веб-сокетов было выбрана комбинация FastAPI и Uvicorn. Такой подход позволяет развернуть систему рекомендаций на веб-сервисы, мобильные и десктопные приложения. Запросы могут поступать сразу от нескольких пользователей. Также стоит отметить что, в файле с проектом присутствует файл Jupyter «РЕКОМЕД.ipynd» в котором полностью расписана разработка рекомендационной системы без API.

Перед запуском нужно обязательно настроить пути к данным в файле «main.py» в директории «api\_service». В параметре «PATH» нужно указать путь до папки «data» которая находится в корне проекта. Переменная «N» отвечает за количество выводимых рекомендаций для одного пользователя.

Для того чтобы запустить систему нужно скачать люблю версию python 3 и выполнить установку библиотек, перечень которых представлен в файле requirements.txt. Для установки нужно в корневой папке проекта из терминала прописать команду «python -m pip install -r requirements.txt».

Затем, находясь в этой же директории, выполнить команду «uvicorn main:app --host localhost». Далее нужно открыть браузер и перейти по ссылке «http://localhost:8000».

Для пользователей и разработчиков предлагаются следующие запросы:

1) «http://localhost:8000/» - Возвращает информацию о доступных маршрутах;

2) «http://localhost:8000/users» - Возвращает список всех пользователей;

3) «http://localhost:8000/recommendations/user\_id» - Получить рекомендации для пользователя user\_id;

4) «http://localhost:8000/items» - Возвращает список всех уникальных товаров;

5) «http://localhost:8000/recommendations/user\_id/item\_id» - возвращает ожидаемую оценку пользователем " user\_id " товара " item\_id ".

**5. Заключение**

Заключение к курсовой работе представляет собой сводку основных результатов, полученных в ходе исследования и разработки рекомендательной системы для одежды. В первой главе были изучены основные аспекты рекомендательных систем, описана предметная область и проанализирован выбранный датасет Amazon ReviewData, что позволило получить представление о структуре данных и содержании информации.

Следующий раздел курсовой работы сфокусировался на реализации рекомендационной системы с применением Евклидового пространства. Этот метод дал возможность выявить сходство между товарами на основе их признаков и использовать данную информацию для формирования персонализированных рекомендаций, что является одним из ключевых моментов в разработке системы.

В заключительном разделе была представлена инструкция по запуску и эксплуатации API рекомендационной системы. Это обеспечивает удобство использования системы и делает ее доступной для конечных пользователей, позволяя им получать рекомендации по выбору одежды на основе реализованных алгоритмов.

Таким образом, выполнение данной курсовой работы позволило ознакомиться с основными принципами и методами построения рекомендательных систем для сферы моды и разработать функциональную систему, способную предоставлять персонализированные рекомендации пользователям. Дальнейшее усовершенствование системы может включать в себя использование других методов рекомендаций и расширение функциональности для обеспечения более точных и разнообразных рекомендаций для пользователей.

**Список использованных источников**

1. Kim Falk / Ким Фальк - Practical Recommender Systems / Рекомендательные системы на практике[2020, PDF, RUS]

2. Интернет-ресурс - <https://romi.center/ru/learning/glossary/online-retail/> дата обращения - 02.12.2023

3. Интернет-ресурс - <https://vc.ru/marketing/533278-rekomendatelnye-sistemy-na-641-milliard-dollarov> дата обращения - 02.12.2023

4. Интернет-ресурс - <https://new-retail.ru/business/e_commerce/kak_internet_magaziny_reshayut_zadachu_podbora_vernogo_razmera_odezhdy_i_obuvi_obzor_luchshikh_reshe8152/> дата обращения - 02.12.2023

5. Интернет-ресурс - https://nijianmo.github.io/amazon/index.html дата обращения - 02.12.2023

6. Интернет-ресурс - https://otus.ru/journal/evklidovo-rasstoyanie-v-analize-klasternogo-tipa/ дата обращения - 02.12.2023