PRODUCT RECOMMENDER SYSTEM USING DATA MINING

SEMINAR REPORT

Submitted by

ASHISH GAUTAM-RA2011003020102

ANIMESH SINGHAL-RA2011003020038

DIVYANSHI GUPTA-RA2011003020082

Under the guidance of

Dr. S. URMELA

Mr. M. SADHASIVAM

In partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

RAMAPURAM CAMPUS, CHENNAI-600089

MAY 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that the Seminar-I report titled "**PRODUCT RECOMMENDER SYSTEM USING DATA MINING**" is the bonafide work of ASHISH GAUTAM-RA2011003020102, ANIMESH SINGHAL-RA2011003020038, DIVYANSHI GUPTA-RA2011003020082 submitted for the course 18CSP106L Seminar – II. This report is a record of successful completion of the specified course evaluated based on literature reviews and the supervisor. No part of the Seminar Report has been submitted for any degree, diploma, title, or recognition before.

SUPERVISOR SIGNATURE	**HOD SIGNATURE**
Dr. S Urmela, Ph.D,.	Dr.K.Raja
Assistant Professor	Head of Department
Dept. of Computer Science & Engineering	Dept. of Computer Science & Engineering
SRM Ramapuram	SRM Ramapuram

Submitted for the Viva Voice Examination held on ………………… at SRM Institute of Science and Technology, Ramapuram Campus, Chennai-600089.

EXAMINER 1 **EXAMINER 2**

ABSTRACT

A product recommender system using collaborative filtering and popularity-based model aims to suggest items to users based on their past interactions with the system. Collaborative filtering utilizes the past behaviour of users to recommend items that users with similar behaviour have interacted with. The system uses a dataset of past interactions between users and products, such as purchase history or browsing history, to identify patterns and relationships between products. These relationships are then used to generate recommendations for new products that are likely to be of interest to a particular user. Collaborative filtering algorithms can be implemented using a variety of techniques, such as matrix factorization or nearest neighbour methods, and can be applied to a wide range of application domains, including e-commerce, media, and entertainment. Popularity-based model, on the other hand, suggests items that are popular among all users. The system would gather data on customer interactions with the platform, such as purchase history, browsing behaviour, and ratings, and use this data to identify the most popular items. These items would then be recommended to other users, increasing the likelihood of them being purchased. By combining both models, the system can recommend items that are both popular and similar to the user's past behaviour. This can lead to more personalized and accurate recommendations. The system can be implemented by first collecting data on user interactions with the system, and then using machine learning algorithms to train the model. The model can then be used to suggest items to new users based on their past interactions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

A product recommender system is a type of artificial intelligence (AI) system that analyzes user behavior data to provide personalized recommendations for products or services. These systems are widely used in e-commerce, media, and other industries to improve customer engagement, increase sales, and enhance customer satisfaction.

Product recommender systems work by analyzing data such as user preferences, past purchases, and browsing history to identify patterns and generate personalized recommendations for products or services that are likely to be of interest to the user. These recommendations can be delivered through various channels such as email, web, mobile app, or social media.

The development of machine learning algorithms and big data analytics has greatly improved the accuracy and efficiency of product recommender systems, enabling businesses to provide highly personalized recommendations to their customers. As a result, product recommender systems have become an essential tool for businesses looking to improve customer engagement, increase sales conversion rates, and enhance customer loyalty.

PROBLEM STATEMENT

A problem statement for product recommender system could be to design and develop a personalized product recommender system that can suggest relevant products to users based on their preferences and past behavior, with the aim of improving customer engagement, retention, and sales conversion rates for an e-commerce platform. The system should be scalable, efficient, and able to handle a large number of users and products. It should also be able to adapt to changing user preferences and trends over time, while ensuring privacy and security of user data.

There are several challenges associated with developing and implementing a product recommender system. A product recommender system requires a large amount of high-quality data to generate accurate recommendations. However, the data can be incomplete, noisy, or biased, making it difficult to develop accurate machine learning algorithms.

When a new user joins the platform or a new product is introduced, there is no user data available to generate personalized recommendations. This makes it challenging to provide relevant recommendations to new users or for new products.

While the primary goal of product recommender systems is to improve customer engagement and sales, there is a risk that users may perceive recommendations as intrusive or irrelevant. It is important to balance the benefits of personalized recommendations with user privacy and satisfaction.

Addressing these challenges requires a combination of technical expertise, data management, and user experience design, making product recommender systems a complex and multidisciplinary field.

SCOPE AND OBJECTIVE

The scope and objective of the scope of a product recommender system involves a range of technical and business aspects that require collaboration between data scientists, software engineers, user experience designers, and business stakeholders.

A product recommender system can have many objectives, such as:

1. Improve customer engagement: By suggesting relevant products to customers based on their preferences, a product recommender system can increase the chances of customers finding what they are looking for, and thereby improving their overall shopping experience.
2. Increase sales conversion rates: By presenting customers with personalized recommendations, a product recommender system can help to increase the likelihood of customers making a purchase, thereby boosting sales and revenue for the e-commerce platform.
3. Enhance customer retention: By providing personalized recommendations and offers, a product recommender system can help to improve customer loyalty and retention, reducing the likelihood of customers switching to competitors.
4. Optimize inventory management: By analyzing customer preferences and purchase patterns, a product recommender system can help to optimize inventory management, reducing the likelihood of overstocking or understocking products.
5. Improve customer satisfaction: By providing customers with personalized recommendations, a product recommender system can help to improve customer satisfaction, as customers are more likely to find products that meet their needs and preferences.

EXISTING SYSTEM

There are various existing systems and tools for product recommender system, such as:

1. Spotify: Spotify's recommendation system uses machine learning to analyze user listening behavior and generate personalized playlists and recommendations based on their preferences.
2. YouTube: YouTube's recommendation system uses a combination of collaborative filtering and deep learning algorithms to analyze user behavior and generate personalized video recommendations.
3. Netflix: Netflix's recommendation system is widely regarded as one of the most advanced in the industry. It uses machine learning algorithms to analyze user behavior data and provide personalized recommendations for movies and TV shows.

These systems are examples of how different companies use recommendation systems to personalize the user experience, increase engagement and sales, and improve customer retention.

There are several existing systems of product recommender systems, including:

The Current system continues to recommend the items which you have already purchased and can be categorized as a one-time

purchase.

The Current system suggests on the basis of ratings which can be sometimes deceptive as people tend to give either too high or too low ratings.

LITERATURE SURVEY

Study	Objective	Methodology	Dataset	Results
Walek and Fajmon (2022)	Recommending Suitable Products in E-shop Using Explanations	Conducted experiments using High or very high similarity to purchased products and recommended products with explanations	Product reviews from Amazon	Found that the best results were obtained using a combination of unigrams and bigrams with a support vector machine (SVM) classifier
Gharei, Dadkhah and Daryoush (2021)	To develop a content-based clothing recommender system using deep neutral network	Used deep convolution neural network to detect categories, product, gender and feature extraction from an image	Fashion product images from Kaggle site	Found that it doesnot recommend products that are already in user purchase history and accuracy is about 73.7%
Razaimehr and Dadkhah (2021)	Injection Shilling Attack tool for Recommender System	Created a tool to add a different attack types on datasets	Movielens (100K and 1M) from Kaggle site	When a fake user attack the system so top 10 recommendation is affected and we shoud avoid that
Farha Islam, Md Shohel Arman*, Nusrat Jahan, Musabbir Hasan Sammak, Nusrat Tasnim, Imran Mahmud (2022)	Model and Popularity Based Recommendation System: A Collaborative Filtering Approach	Used deep learning methodology using the CF approach, CNN, and cosine similarity for the product recommendation system.	Amazon E-commerce Dataset is taken from Kaggle	Generated a utility matrix, transpose it, and then use Truncated Singular-Value Decomposition, also known as Truncated Singular-Value Decomposition
Bharathipriya C, Aswini D, Francis Jency X, Kirubakaran R, Swathi B (2021)	Product Recommender System Using Collaborative Filtering Technique	User clustering method is used and improved cosine similarity is used to pair up the similar users.	Five Different types of Amazon Datasets are used for our implementation	Evaluation metrics shows us that we have produced a high quality recommendation.
You Lyu (2021)	Recommender Systems in e-Commerce	Content-based recommendation is mixed within the framework of recommendation	Different data sources taken from different E-Commerce Sites	The user model uses the recommendation algorithm to recommend the results to the user by collecting

Study	Objective	Methodology	Dataset	Results
				the users' preferences and combining it with the recommended object model.
Anand Kumar Pandey, B. Ankayarkanni (2020)	Recommending E-Commerce Products on Cold Start and Long Tail Using Transaction Data	Recommending the products to the customers of some kind based on the previous purchases byusing clustering and classification.	Amazon Dataset taken from Kaggle is used in the project	Cold start and longtail problem handled to make the sales better for the retailers and achieve max profit.
Sangwhan Cha, Moonhyung Lee (2020)	Personalized Hybrid Recommender System using Average Visit Intervals	Supervised, unsupervised, semi-supervised, and reinforced learning methods are used in the implementation.	25M dataset was selected to be tested on to check whether the system can process very large amounts of data	The mean RMSE and MAE values of 0.7653 and 0.6632 can be seen as quite high compared to other algorithms
Parul, Kavita Khanna (2021)	Literature Survey: Recommender Systems	Two prominent fuzzy logic applications – fuzzy inference system and fuzzy MCDM method are utilized to support the decision of service choice.	Standard data-set.	The performance of TARS is evaluated using two datasets of different sparsity levels viz. Jester dataset and MovieLens dataset (available online) and compared with traditional Collaborative Filtering based approach for generating recommendations.
E.W.T. Ngai, Li Xiu, D.C.K. Chau (2008)	Application of data mining techniques in customer relationship management: A literature review and classification	Business intelligence and knowledge discovery are the most common academic discipline for data mining research in CRM so, they are being used in this research work.	academic database of literature between the period of 2000–2006 covering 24 journals	It results in providing the reasonable insights and shows the incidence of research on this subject. Research on the application of data mining in CRM will increase significantly in the future based on past publication rates and the increasing interest in the area.

Study	Objective	Methodology	Dataset	Results
Hao Wang (2021)	ZeroMat: Solving Cold-start Problem of Recommender System with No Input Data	In this paper, they are proposing a new method called ZeroMat to resolve this problem. Their method is inspired by Probabilistic Matrix Factorization.	MovieLens dataset has a rating scale ranging from 1 to 5	Their method is an ideal tool for the cold-start problem. Matrix factorization techniques are far from being good enough. Even compared with the classic matrix factorization, our method is competitive in both MAE and fairness metric.

ARCHITECTURE

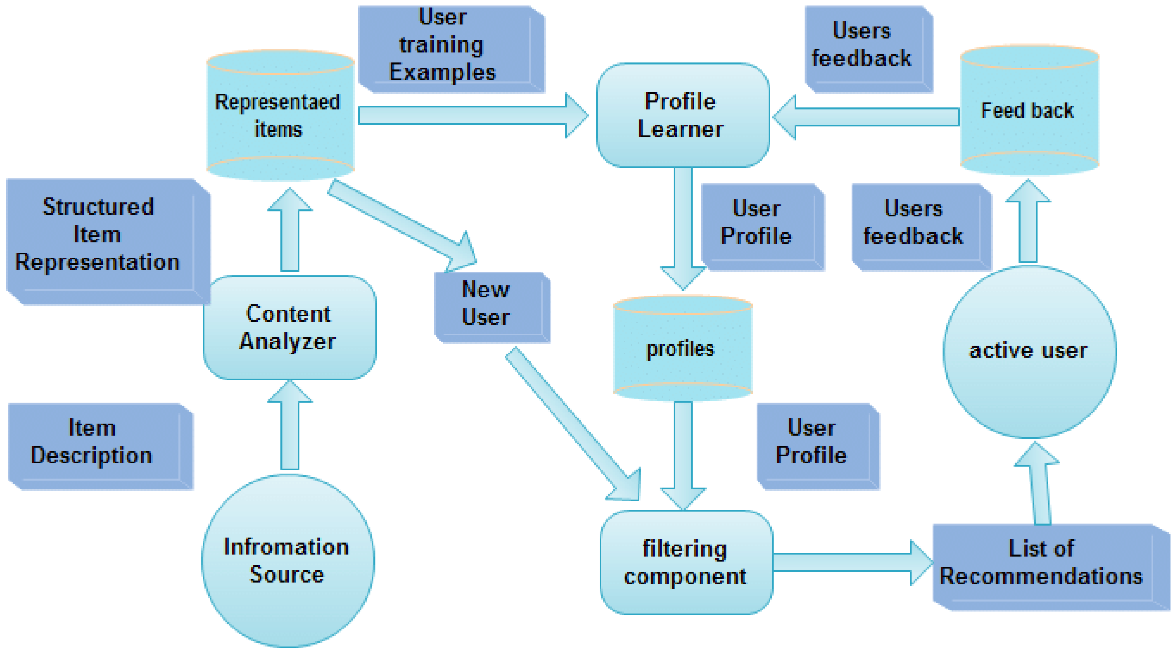

Fig. 1: Content Based Filtering

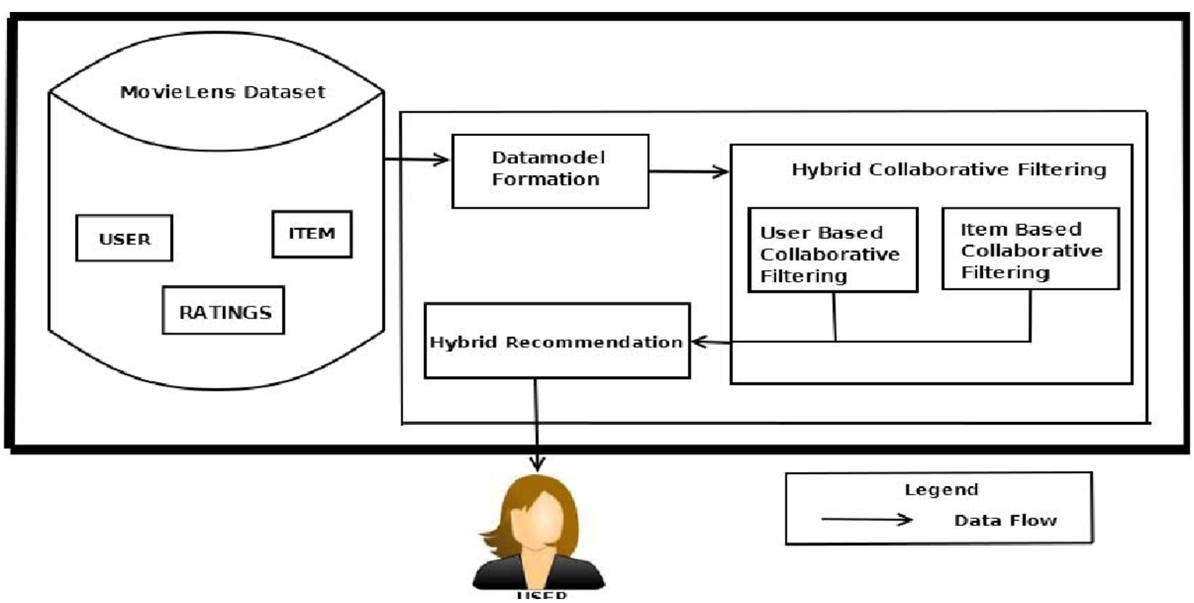

Fig. 2: Collaborative Filtering

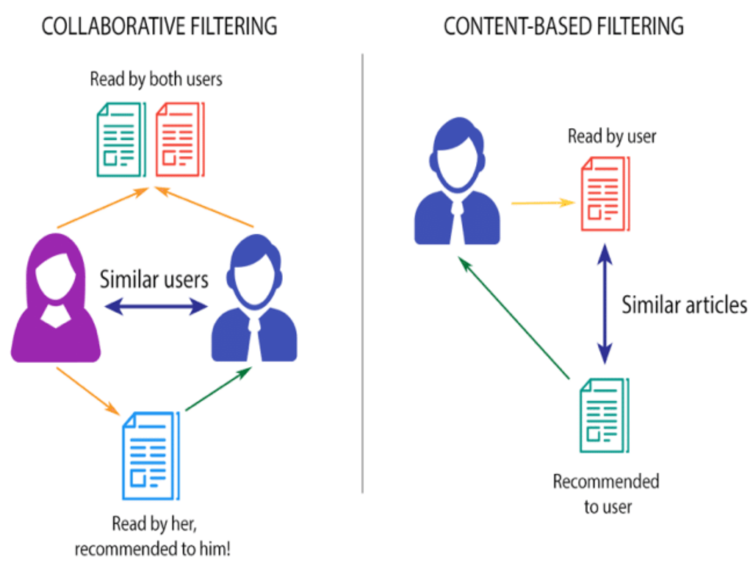

Fig. 3: Collaborative Vs. Content Based Filtering

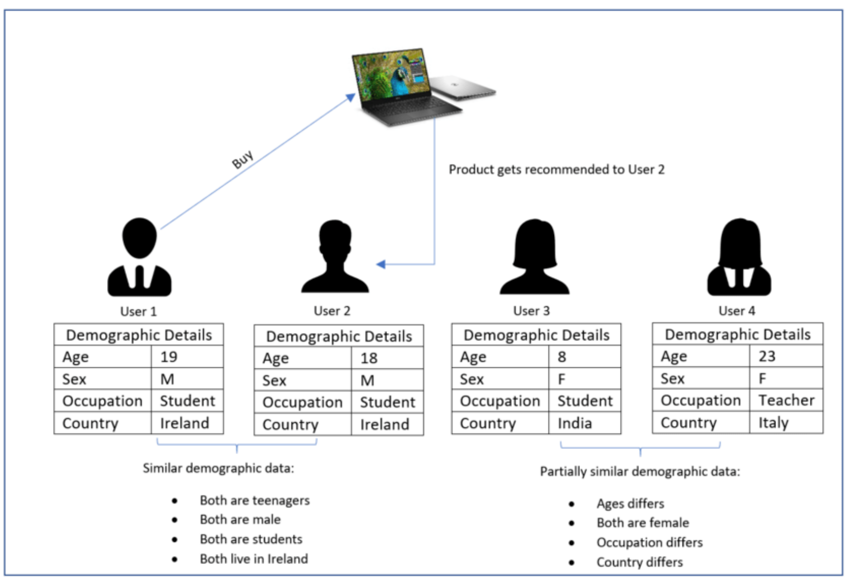

Fig. 4: Demographic Filtering

PROPOSED WORK

Recommendation Systems are a type of information filtering systems as they improve the quality of search results and provides items that are more relevant to the search item or are realted to the search history of the user. Recommendation system finds similarity between the product a user clicks on and the other products and then recommends if their is some sustainable similarity. Youtube recommends videos and even decides which video to play next on autoplay. Facebook and Instagram recommend friends and pages to folow. Instagram also recommends reels that you might be interested in. Netflix recommends movies based on user taste and genre selection, it even recommends thumbnails of movies based on user data. Amazon, BigBasket and other E-Commerce websites recommend ads and products based on the items we buy or search for.

There are basically three types of recommender systems:-

1. **Demographic Filtering-** They offer generalized recommendations to every user, based on movie popularity and/or genre. The System recommends the same movies to users with similar demographic features. Since each user is different , this approach is considered to be too simple. The basic idea behind this system is that movies that are more popular and critically acclaimed will have a higher probability of being liked by the average audience.
2. **Content Based Filtering-** They suggest similar items based on a particular item. This system uses item metadata, such as genre, director, description, actors, etc. for movies, to make these recommendations. The general idea behind these recommender systems is that if a person liked a particular item, he or she will also like an item that is similar to it.
3. **Collaborative Filtering-** This system matches persons with similar interests and provides recommendations based on this matching. Collaborative filters do not require item metadata like its content-based counterparts.
4. **Collaborative Vs Content Based Filtering-** Collaborative filtering and content-based filtering are two commonly used approaches in recommender systems for making personalized recommendations. Collaborative filtering involves analyzing the past behaviors of users to identify patterns and make recommendations based on similar

behaviors of other users. On the other hand, content-based filtering involves analyzing the attributes of items (e.g., product features, genre, etc.) and recommending similar items to users based on their past preferences.

The steps involved in building a product recommender system using average rating are as follows:

1. Gather data: Collect the ratings given by users for different products.
2. Calculate the average rating: Calculate the average rating for each product based on the ratings given by users.
3. Sort products by rating: Sort the products by their average rating, from highest to lowest.
4. Recommend top-rated products: Recommend the top-rated products to new users based on their preferences.
5. Implement the system: Implement the recommender system in a user-friendly interface that allows users to easily browse and select products.

IMPLEMENTATION

In [1]:

```
#Basic Libraries
import numpy as np
import pandas as pd

#Visualization Libraries
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

#Text Handling Libraries
import re
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.metrics.pairwise import linear_kernel, cosine_similarity
```

In [2]:

```
df = pd.read_csv('../input/bigbasket-entire-product-list-28k-datapoints/BigBaske
t Products.csv',index_col='index')
```

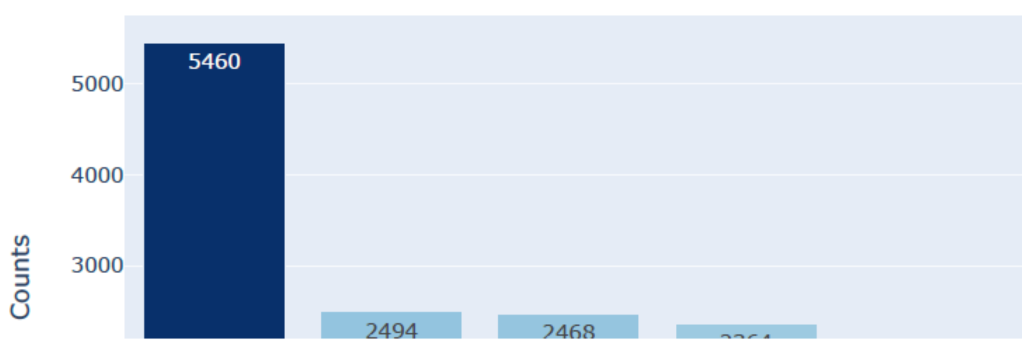

Fig. 5(a): Count of Items in Each Category

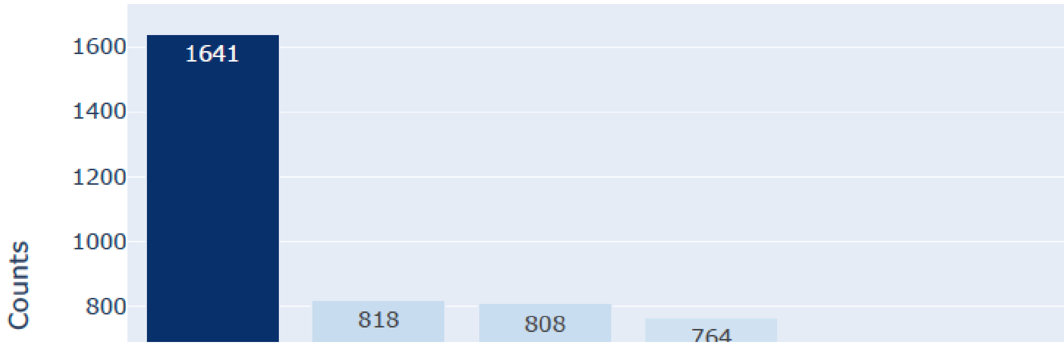

Fig. 5(b): Top 10 Bought Sub_Categories

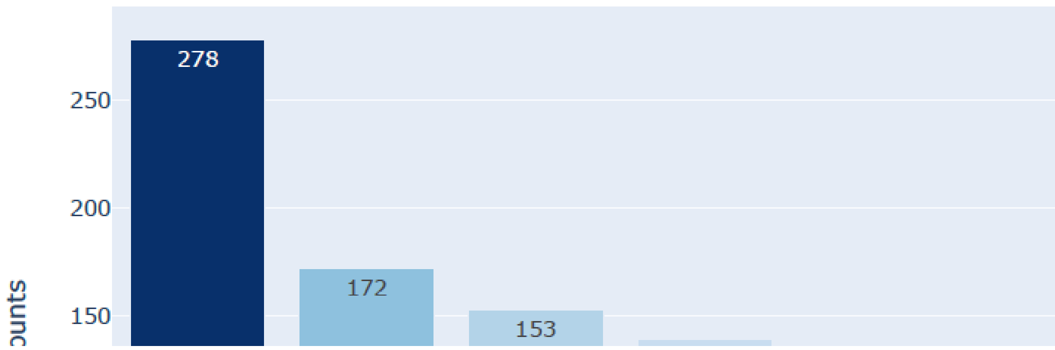

Fig. 5(c):Top 10 Brand Items based on Item Counts

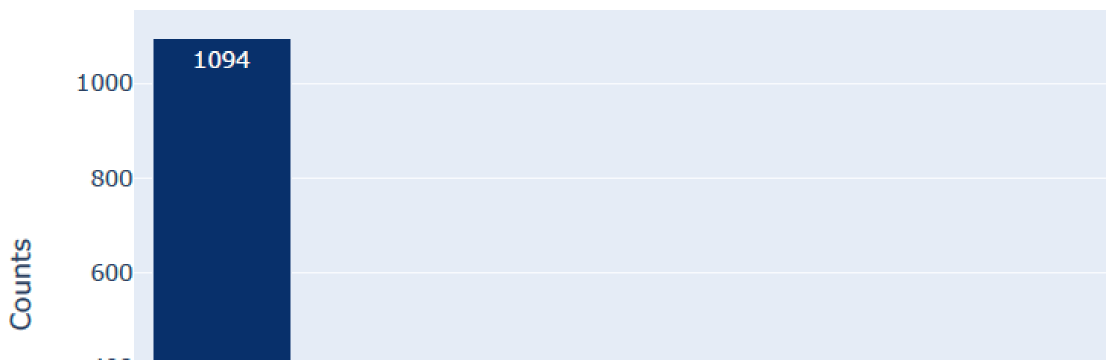

Fig. 5(d): Top 10 Types of Products based on Item Counts

In [31]:

```
get_recommendations_1('Cadbury Perk - Chocolate Bar')
```

Out[31]:

```
index
27049                          Pickle - Mixed
6601                    Pickle - Kaduku Mango
17934                 Pickle - Mix Vegetable
27105                          Pickle - Prawn
3962                    Pickle - Tender Mango
16875                Olive Oil - Carrot Pickle
3444                       Pickle - Cut Mango
17237        Andhra Special Red Chilli Pickle
27234     Pickle - Lime (South Indian Style)
4955                     Pickle - Gooseberry
Name: product, dtype: object
```

Final Output:

Out[50]:

	Old Recommendor	New Recommendor
0	Rectangular Plastic Container - With Lid, Mult...	Glass Water Bottle - Aquaria Organic Purple
1	Jar - With Lid, Yellow	Glass Water Bottle With Round Base - Transpare...
2	Round & Flat Storage Container - With lid, Green	H2O Unbreakable Water Bottle - Pink
3	Premium Rectangular Plastic Container With Lid...	Water Bottle H2O Purple
4	Premium Round Plastic Container With Lid - Yellow	H2O Unbreakable Water Bottle - Green
5	Premium Rectangular Plastic Container With Lid...	Regel Tritan Plastic Sports Water Bottle - Black
6	Premium Round & Flat Storage Container With Li...	Apsara 1 Water Bottle - Assorted Colour
7	Premium Round Plastic Container With Lid - Blue	Glass Water Bottle With Round Base - Yellow, B...
8	Premium Round Plastic Container With Lid - Mul...	Trendy Stainless Steel Bottle With Steel Cap -...
9	Premium Round Plastic Container With Lid - Pink	Penta Plastic Pet Water Bottle - Violet, Wide ...

FUTURE SCOPE

The future scope of product recommender systems is vast and includes the following potential advancements:

1. Contextual recommendation: Product recommender systems can integrate contextual data such as location, time, and weather to provide more relevant and personalized recommendations. For example, a system could recommend products based on the user's location, weather, and time of day.
2. Hybrid recommendation techniques: Future systems could combine various recommendation techniques, such as content-based filtering, collaborative filtering, and knowledge-based filtering, to provide more accurate and diverse recommendations.
3. Explainable AI: Explainable AI is an emerging trend that aims to provide transparency and accountability in machine learning algorithms. Future product recommender systems could incorporate explainable AI techniques to explain how recommendations are generated and to build trust with users.
4. Augmented Reality: With the advancement of augmented reality (AR) technology, future product recommender systems could integrate AR features that allow users to view products in a real-world context. This can improve user engagement and provide a more immersive shopping experience.
5. Personalized pricing: Future product recommender systems could incorporate personalized pricing strategies that offer different prices to different users based on their preferences and behavior. This can increase sales and improve customer loyalty.
6. Social recommendations: Social recommendations can leverage users' social networks to generate recommendations for products that are popular among their friends and followers. This can improve user engagement and trust in the system.

Overall, the future scope of product recommender systems is promising and will continue to evolve with advancements in AI, data analytics, and user experience design. As the technology becomes more sophisticated, product recommender systems will play an increasingly important role in enhancing customer engagement, increasing sales, and improving customer satisfaction.

While product recommender systems offer many benefits, there are also some potential disadvantages, including:

Limited diversity: Recommender systems can sometimes suggest the same type of products repeatedly, which can limit the user's exposure to a diverse range of products. This can result in missed opportunities to discover new products and can lead to a lack of variety in the user's purchases.

Overreliance on past behavior: Recommender systems are based on past user behavior, which can result in recommendations that are too narrow or similar to past purchases. This can lead to missed opportunities to recommend products that may be outside of the user's typical preferences.

Limited data availability: Recommender systems require a significant amount of data to generate accurate recommendations, which can be a challenge for smaller e-commerce companies or for companies with limited user data. In such cases, the recommendations generated by the system may be less accurate and may not meet the user's needs.

Lack of transparency: Recommender systems often use complex algorithms and data analysis techniques, which can make it difficult for users to understand how recommendations are generated. This can lead to a lack of trust in the system and can result in users not using the recommendations.

Privacy concerns: Recommender systems require access to user data, which can raise privacy concerns for some users. Users may be uncomfortable with their personal data being used to generate recommendations, which can lead to a lack of adoption of the system.

Overall, while product recommender systems offer many benefits, it is important for companies to be aware of these potential disadvantages and to work to address them through transparency, diversity in recommendations, and sensitivity to privacy concerns.

CONCLUSION

In conclusion, A recommendation system is an information service system that connects users and projects: on the one hand, it helps users discover potential projects of interest; on the other hand, it helps project providers to deliver projects to users who are interested in it.

We have designed and implemented the system using content based filtering, collaborative filtering and demographic filtering. The dataset considered has the ratings given by the other users to a specific product and depending on the similarity between the rated product we try to recommend the products to our current user. The future work of the project includes improving the efficiency of the system. And it should also be able to give appropriate recommendations to the users who don't have any previous purchase history or to the new users. In future we can try to use recurrent neural networks and deep learning. With the help of deep learning techniques, we can overcome some of the drawbacks of the matrix factorization technique. Deep learning uses recurrent neural networks to accommodate time in the recommender system which is not possible in the matrix factorization method. We can also work on providing sub-optimal recommendations to the user and record the reaction of the user and it can be used in the future by the system.

REFERENCES

[1] Bobadilla, JesúS, Fernando Ortega, Antonio Hernando, and JesúS Bernal. "A collaborative filtering approach to mitigate the new user cold start problem." Knowledge-based systems 26 (2012): 225-238.

[2] Zhao, Wayne Xin, Sui Li, Yulan He, Edward Y. Chang, Ji-Rong Wen, and Xiaoming Li. "Connecting social media to e-commerce: Cold-start product recommendation using microblogging information." IEEE Transactions on Knowledge and Data Engineering 28, no. 5 (2015): 1147-1159.

[3] Sundaresan, Neel. "Recommender systems at the long tail." In Proceedings of the fifth ACM conference on Recommender systems, pp. 1-6. 2011.

[4] J. Li, K. Lu, Z. Huang, and H. T. Shen, "Two birds one stone: On bothcold-start and long-tail recommendation," in ACM MM, ser. MM '17.ACM, 2017, pp. 898–906.

[5] Li, Jingjing, Ke Lu, Zi Huang, and Heng Tao Shen. "On both Cold-Start and Long-Tail Recommendation with Social Data." IEEE Transactions on Knowledge and Data Engineering (2019).

[6] Houlsby, Neil, José Miguel Hernández-Lobato, and Zoubin Ghahramani. "Cold-start active learning with robust ordinal matrix factorization." In International Conference on Machine Learning, pp. 766-774. 2014.

[7] Asha, P., Sri Neeharika, K., Sindhura, T, "Metoo Movement Analysis Through the Lens of Social Media", International Journal of Recent Technology and Engineering,8(3),2019.

[8] Li, Yung-Ming, Chun-Te Wu, and Cheng-Yang Lai. "A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship." Decision Support Systems 55, no. 3 (2013): 740-752.

[9] Gigimol, S., and Sincy John. "A Survey on Different Types of Recommendation Systems." Engineering and Science 1, no. 4 (2016): 111-113.

[10] Christina, Ruth, Greeshma Liz Shajan, and B. Ankayarkanni. "CART-A Statistical Model for Predicting QoE using Machine Learning in Smartphones." In IOP Conference Series: Materials Science and Engineering, vol. 590, no. 1, p. 012001. IOP Publishing, 2019.

[11] Herrera, Victor M., Taghi M. Khoshgoftaar, Flavio Villanustre, and Borko Furht. "Random forest implementation and optimization for Big Data analytics on LexisNexis's high performance computing cluster platform." Journal of Big Data 6, no. 1 (2019): 68.

[12] Valcarce, Daniel, Javier Parapar, and Álvaro Barreiro. "Item-based relevance modelling of recommendations for getting rid of long tail products." Knowledge-Based Systems 103 (2016): 41-51.