



Recommendation System for E-Commerce Using Collaborative Filtering

Preeti Patil¹, Sandeep Uddhavrao Kadam², E.R. Aruna³, Amar More⁴, Balajee R.M.⁵,
B. Narendra Kumar Rao^{6*}

¹ Department of Information Technology, D Y P College of Engineering, Pune 411044, India

² Department of Computer Engineering, Keystone School of Engineering, Pune 412308, India

³ Department of CSE, Vardhaman College of Engineering, Hyderabad 501218, India

⁴ School of Computer Engineering, MIT Academy of Engineering, Pune 411011, India

⁵ Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur 522302, India

⁶ School of Computing, Mohan Babu University, Tirupati 517501, India

Corresponding Author Email: narendrakumarraob@gmail.com

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/jesa.570421>

ABSTRACT

Received: 6 June 2024

Revised: 1 August 2024

Accepted: 9 August 2024

Available online: 27 August 2024

Keywords:

recommendation, collaborative filtering, e-commerce

In the swiftly evolving domain of online commerce, the significance of recommendation systems has risen alongside the rapid expansion of internet usage over the past decade. As online commerce continues to thrive, recommendation systems serve a crucial function in steering users towards pertinent products amidst the expansive online environment. Nonetheless, conventional collaborative filtering algorithms often encounter challenges such as sparse data and shifting user preferences, necessitating innovative approaches. Our proposed recommendation system aims to tackle these hurdles by seamlessly blending collaborative filtering and content-based filtering methodologies. It will offer product suggestions for both new and existing users. Through thorough examination of users' past purchasing behaviors, patterns, and feedback, our system customizes recommendations precisely to cater to existing users' needs. The initial stage involves feature extraction, wherein both content-based and collaborative features are obtained by creating user profiles, computing content similarity, identifying related items, generating recommendations, and suggesting items. Additionally, new users will receive recommendations for novel and trending products. Leveraging the Walmart Product rating Dataset, our system continuously enhances recommendations based on evolving user interactions, thus optimizing engagement and satisfaction levels. This study underscores the pivotal role of advanced recommendation techniques in transforming the online commerce landscape, ensuring informed purchasing decisions, heightened user satisfaction, and increased sales.

1. INTRODUCTION

In the rapidly evolving landscape of e-commerce, the competition among online platforms is reaching unprecedented levels. Established giants dominate the market, while emerging companies, often referred to as unicorns, strive to carve out their niche and challenge the dominance of major players. However, amidst this intense competition, many new entrants face a critical obstacle: the absence of essential sales booster features, particularly robust product recommendation systems. These systems play a pivotal role in driving sales, enhancing user experience, and fostering customer loyalty [1].

In these recent years, significance of product recommendation systems has grown substantially as they offer personalized suggestions to users based on their browsing history, purchase behavior, and preferences [2]. These systems not only facilitate decision-making for customers but also contribute significantly to revenue generation for e-commerce platforms. However, for smaller companies venturing into the

e-commerce space, developing and implementing such recommendation systems poses a considerable challenge.

The development of an effective product recommendation system hinges on the availability and analysis of vast datasets to gain insights into customer preferences and behaviors. Established e-commerce platforms, with their extensive user base and transaction history, have access to rich datasets that can be leveraged to train sophisticated recommendation algorithms. In contrast, new companies, often characterized by limited sales and resources, struggle to accumulate and process such extensive data. Consequently, they face a daunting task in devising recommendation systems that can rival those of their larger counterparts.

Moreover, the lack of satisfactory open-source solutions further exacerbates the challenge for smaller e-commerce entities [3]. While several recommendation algorithms and frameworks are available in the public domain, they may not adequately address the unique needs and constraints of emerging companies [4]. As a result, these companies are left grappling with the dilemma of either investing heavily in

proprietary solutions or compromising on the effectiveness of their recommendation systems.

Nevertheless, within the realm of AI-ML, a plethora of theories, methodologies, and techniques hold promise for addressing the pressing need for tailored product recommendation systems. By harnessing the power of these advanced technologies, new companies can overcome the barriers posed by limited resources and data availability to develop innovative solutions that cater to their specific requirements [5]. Through strategic implementation of machine learning algorithms, data analytics, and user-centric design principles, emerging e-commerce ventures can not only enhance their competitiveness but also improve user satisfaction and drive growth in the fiercely competitive market landscape [6].

In this context, this research endeavors to explore the challenges faced by new e-commerce companies in developing product recommendation systems and to propose novel approaches and solutions to address these challenges. By examining existing literature, analyzing case studies, and leveraging cutting-edge methodologies from the field of machine learning, this study seeks to provide insights and guidance for emerging companies seeking to enhance their recommendation capabilities and gain a competitive edge in the e-commerce arena.

The unique challenges in developing recommendation systems for new and smaller e-commerce companies includes Cold Start Problems, Data Sparsity, Scalability, Resource Constraints, during Real-time Processing and limited infrastructure. Few of the AI&ML techniques used to address these challenges are Hybrid Recommendation Systems, Matrix Factorization, Incremental Learning, Graph-based Models and Deep Learning models.

Overall, the integration of effective product recommendation systems is crucial for the success and sustainability of e-commerce ventures, particularly for newer entrants aiming to establish themselves in the highly competitive market. Through innovative strategies and leveraging advancements in technology, these companies can leverage recommendation systems to drive sales, foster customer engagement, and propel their growth trajectory [7].

2. LITERATURE SURVEY

Product recommendation systems have become integral tools for enhancing user experience and driving sales in the competitive landscape of e-commerce. Leveraging various methodologies and techniques, researchers have explored innovative approaches to developing effective recommendation systems tailored to the specific needs of online platforms. The objective of this literature review is to offer an overview of recent research concerning product recommendation systems, with a particular emphasis on collaborative filtering and associated methodologies.

Khatter et al. [8] proposed an e-commerce product recommendation system employing collaborative filtering and textual clustering techniques. Their approach combines collaborative filtering, which analyzes user-item interactions, with textual clustering to enhance recommendation accuracy by considering product descriptions and user preferences.

Dubey et al. [9] introduced an expanded opinion lexicon combined with machine learning-based sentiment analysis of tweets. Although not directly related to collaborative filtering,

sentiment analysis can complement recommendation systems by incorporating user sentiment towards products or brands, thereby improving recommendation relevance and user satisfaction.

Khan et al. [10] conducted a survey on collaborative filtering within online recommendation systems. Their study presents an overview of diverse collaborative filtering techniques, encompassing user, item, and model-based approaches, and explores their applications across online platforms.

Other researches also introduced a trust-based collaborative filtering algorithm tailored for e-commerce recommendation systems. By integrating trust relationships among users, their approach significantly boosts recommendation accuracy while effectively tackling the common cold-start problem associated with collaborative filtering. Collaborative Filtering faces the "cold-start problem" when new users or items are introduced to the system. The following strategies can be adapted with New Users like Gather Initial Data using Surveys and following some onboarding process, use Hybrid Approaches like content-Based Filtering. For New Items following can be observed content-Based Filtering involving item features, Hybrid Approaches and exposure to promotional strategies.

Wang and Wang [11] developed recommendation system for e-commerce was developed, leveraging an enhanced collaborative filtering algorithm. Their approach introduces enhancements to traditional collaborative filtering methods to overcome limitations such as sparsity and scalability, thereby improving recommendation quality.

Anwar and Uma conducted a comparative study was conducted to examine various recommendation system approaches, with a focus on movie and multimedia recommendation utilizing collaborative filtering [12]. The analysis delves into the strengths and weaknesses of different collaborative filtering techniques, emphasizing the significance of algorithm selection in the design of recommendation systems.

Priya and Mansoor Hussain [13] have discussed three important techniques of creating a recommendation system- Collaborative Filtering, Content based and Hybrid. They have also discussed the existing e-commerce websites and the type of Recommendation they use.

Overall, the literature reviewed underscores the importance of collaborative filtering and related techniques [8] in developing effective product recommendation systems for e-commerce. By leveraging user-item interactions, textual data, sentiment analysis, and trust relationships, researchers aim to enhance recommendation accuracy, address scalability challenges, and improve user satisfaction in the competitive e-commerce landscape.

3. PROPOSED METHODOLOGY

3.1 Proposed work

The proposed work aims to develop a modern recommendation system for e-commerce platforms, centered around collaborative filtering techniques. By leveraging users' purchase history and search behavior, the system will analyze patterns and preferences to offer personalized product suggestions. Additionally, content-based filtering will be integrated to ensure diverse recommendations tailored to individual tastes. User interactions with the cart will serve as

feedback loops for continuous refinement of recommendations, enhancing overall user experience and engagement. Through this research, we seek to demonstrate the potency of collaborative filtering in delivering highly personalized recommendations, ultimately driving increased sales and user satisfaction. By setting a new standard in e-commerce personalization, this system aims to improve customer loyalty, retention, and overall business growth in the competitive online marketplace. A Collaborative filtering approach is essential for e-commerce recommendation systems because of its ability to provide personalized recommendations by leveraging user behavior patterns and preferences [9]. Its domain-independence nature makes it more versatile across products without requiring specific content information.

3.2 System architecture

The system architecture comprises three main components: data processing, recommendations, and final outcome generation as in Figure 1. The data processing module receives and preprocesses cart product data sets, extracting relevant features such as product attributes and user interactions. These processed data are then fed into two recommendation modules: content-based & collaborative filtering.

This content-based recommendation module utilizes product attributes to suggest items similar to those in cart, ensuring diversity and relevance. Meanwhile, the collaborative filtering [10] method analyzes user behavior and interactions to generate personalized recommendations based on similarities between users & items.

Finally, recommendations from both modules are combined and further refined based on user feedback and interactions with the cart. The final outcome is a set of highly personalized product recommendations tailored to individual user preferences. This architecture enables the system to deliver accurate and diverse suggestions, enhancing user satisfaction and driving increased sales for e-commerce platforms. An approach like the below can be used where α is a weight factor between 0 and 1, i is the item, and u is the user.

$$Final\ Score(i, u) = \alpha \times CB\ Score(i, u) + (1 - \alpha) \times CF\ Score(i, u)$$

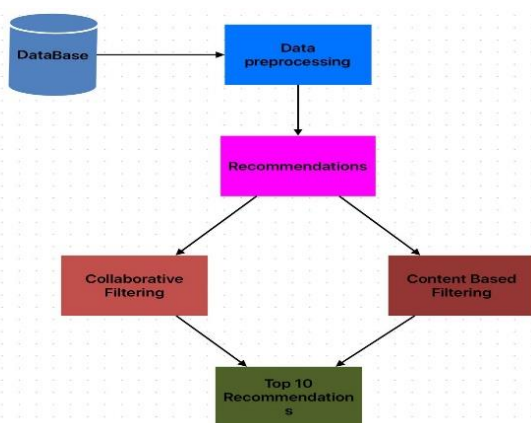


Figure 1. Proposed architecture

3.3 Database description

Samples from “marketing_sample_for_walmart_com-walmart_com_product_review” dataset is taken from the Kaggle website, which serves as a vital hub for accessing and

sharing diverse datasets. The Walmart Product rating Dataset collection process involves gathering comprehensive information about products added to users shopping carts on an e-commerce platform. The data set consists of 5000 rows and 31 columns. Out of which Product Id, Product Name, Product Category, Product Price, Item Number, Rating, Review. Each product typically has attributes such as Product Id, Product Name, Product Category, Item Number, Rating, Review and Price. It includes User review and User ratings. The data entered in the Users rating column are number format only supports the values With values ranging from 1 to 5. The dataset contains different category of products some of them are Beauty, Health, Household essentials, Personal care, Sports and Outdoors, etc,

By using the “marketing_sample_for_walmart_com-walmart_com_product_review” Dataset, e-commerce platforms gain insights into user preferences, purchasing behaviour, and product popularity. This dataset serves as the foundation for developing and optimizing recommendation systems, enhancing user experience, and driving sales by providing personalized and relevant product suggestions.

Preprocessing:

The data processing workflow involves many steps to prepare the dataset for further analysis as well as for modeling. Initially, the pandas dataframe is utilized to load and organize the data. During data preprocessing, if our dataset contains missing values in the rating field, we fill them with zeros. Conversely, if crucial fields such as product name and details are absent, we exclude the corresponding rows to uphold data integrity. Outliers are addressed by either removing them if they are extreme or transforming them using techniques like robust scaling to mitigate their impact. To eliminate irrelevant rows and columns, Singular Value Decomposition (SVD) can be utilized. Data splitting entails dividing the dataset into training, validation, and testing sets to assess machine learning models.

The training data is utilized for model training, the validation data for hyperparameter tuning and performance assessment, and the testing data for evaluating the final model's performance on unseen data. These preprocessing steps ensure the data is cleansed, standardized, and suitable for analysis or modeling, thereby yielding more precise and dependable results. Additionally, label processing is performed to encode categorical variables or convert them into numerical representations suitable for machine learning algorithms. This may involve techniques such as one-hot encoding or label encoding. Few limitations with marketing_sample_for_walmart_com-walmart_com_product_review dataset includes Sampling Bias, Data Quality Issues, Temporal Dynamics, Contextual Limitations, Sentiment Analysis Challenges, Cultural and Demographic Factors.

To ensure the proposed model attains high accuracy, several strategies are indispensable. These strategies encompass addressing missing values by substitution with 0, managing imbalanced data, determining an optimal value for K via cross-validation for the K-nearest neighbors (KNN) algorithm, selecting pertinent features and scaling them suitably, choosing an appropriate distance metric, evaluating generalization performance through cross-validation, and implementing dimensionality reduction using Singular Value Decomposition (SVD). By incorporating these strategies, optimizing parameters, and refining preprocessing steps, the model can achieve high accuracy.

Collaborative filtering hinges on user-item interactions, whereas content-based filtering leverages item attributes. Integrating both methodologies amalgamates their merits to augment recommendation precision. The optimization process entails several steps. If our dataset contains missing values in the rating field, we replace them with zeros. Conversely, if vital fields like product name and details are absent, we eliminate the corresponding rows to uphold data integrity. Outliers are addressed by either removing extreme cases or employing techniques like robust scaling to diminish their impact. Additionally, irrelevant rows and columns are filtered out using Singular Value Decomposition (SVD). Another aspect of optimization involves tuning hyperparameters. During this process, you experiment with settings such as the number of neighbors in collaborative filtering or the similarity threshold in content-based filtering. Techniques like grid search or random search aid in identifying the optimal hyperparameters for enhanced model performance. These meticulous steps ensure that your recommendation system is finely calibrated to provide accurate and pertinent product suggestions to users.

3.4 Recommendation

The recommendation process comprises two main methods: content-Based Filtering & collaborative Filtering. Product recommendations in content-based filtering are generated through evaluating the similarity of items in the dataset. This method utilizes features such as product attributes, descriptions, and user preferences to identify items that closely match those already in the user's cart. Recommendations are made by selecting the items with highest similarity scores to the ones already present.

On the other hand, Collaborative Filtering recommends products by analyzing user-item interactions across the entire dataset. This method identifies patterns and similarities in user behavior, such as purchase history or item ratings, to generate personalized recommendations for each user. By leveraging the collective wisdom of all users, Collaborative Filtering can suggest products that are popular among similar users or items frequently purchased together.

By employing both Content-Based & Collaborative Filtering methods, the recommendation process ensures a comprehensive and personalized approach in suggesting products that align with user preferences & interests.

3.5 Content-based filtering

This recommendation method emphasizes suggesting products that closely resemble those with which a user has previously interacted. The process begins by appending the list of products and orders, forming a dataset that captures user-item interactions. Subsequently, the data is transformed into Compressed Sparse Row (CSR) format, facilitating efficient computation. We are going to perform TF – IDF vectorization. The value obtained by performing multiplication of TF score and IDF score is assigned to each product in the data set. A model is then created using the Nearest Neighbors algorithm, which identifies products similar to those that user has engaged with, in past. The model is fitted using the CSR matrix, enabling it to efficiently process large datasets. Recommendations are formulated by analyzing a user's previous orders, and the suggested products are then added to the result. Finally, the recommended products are displayed to the user, providing

personalized suggestions aligned with their preferences and past interactions. Through Content-Based Filtering, users receive tailored recommendations that enhance their shopping experience and increase the likelihood of finding products of interest. Content-Based Filtering techniques include Feature Extraction and User Profiling. The algorithms associated are TF-IDF (Term Frequency-Inverse Document Frequency), Cosine Similarity, k-Nearest Neighbors (k-NN), Naive Bayes, Decision Trees, Deep Learning (e.g., CNNs for images, RNNs for text).

3.6 Collaborative filtering method

This is a recommendation method that suggests products based on user behavior and preferences, leveraging similarities among users or items.

To achieve collaborative filtering, data is first imported from a dataset containing items, such as ratings and purchases. After preprocessing to address missing values and outliers, the similarity between users or items is computed. This involves constructing a utility matrix that represents user-item preferences, where most values are initially unknown and set to 0. Subsequently, the utility matrix is transposed. Singular Value Decomposition (SVD), a dimensionality reduction technique, is then applied to decompose the matrix into three components: the original matrix (A), an orthogonal left singular matrix (U), and a diagonal matrix (S). Following this, a correlation matrix is generated by computing correlations between items based on the ratings given by users who have purchased the same products. Finally, the recommendation system presents the top 10 products to the customer based on the purchase history of other customers on the website. Finally, the recommended products are displayed to the user, providing personalized suggestions that align with their preferences as well as past interactions. Collaborative Filtering enhances user experience by delivering tailored recommendations and fostering engagement in e-commerce platforms. Collaborative Filtering techniques include User-Item Interactions and Matrix Factorization [14]. The algorithms associated are User-Based Collaborative Filtering, Item-Based Collaborative Filtering, Matrix Factorization, Singular Value Decomposition (SVD) and Neural Collaborative Filtering.

The collaborative filtering recommends products by finding similar products to the ones the user has interacted with and suggesting those similar products.

3.7 Equations

$$r(u_i) = \frac{\sum_j CN(I) \sin(I \cdot j) \cdot r(Uj)}{\sum_j CN(I) \sin(I \cdot j)} \quad (1)$$

Eq. (1) is the predicted rating for user U on item i , where $N(I)$ is the neighborhood of similar items to item I , $\sin(I, j)$ is the similarity between items I and j , $r(uj)$ is the actual rating of user U on item j .

$$TF(t, d) = \frac{n_{(d)}(t)}{n(d)} \quad (2)$$

Eq. (2) is the term frequency of the word in the product description field, where $n_{(d)}(t)$ represents number of times a word appears in the product description field, $n(d)$ represents total number of times the word or term occurs in remaining product description field.

$$IDF(t, d) = \log \left(\left(\frac{n(d)}{n_{(d)}(t)} \right) + 1 \right) \quad (3)$$

Eq. (3) is the Inverse Document Frequency. Where t represents term or word in the description field, $n(d)$ represents total number of description fields and $n_{(d)}(t)$ represents the number of description fields containing the term and D represents entire collection of description fields.

$$TFIDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (4)$$

Eq. (4) is the *TFIDF* score, where t represents a term (word) in a document, d represents a specific document [15], D represents the entire collection of documents.

Algorithm in pseudo format

1. Import necessary libraries: “numpy”, “pandas”, matplotlib.pyplot”, “sklearn”, “TruncatedSVD from sklearn.decomposition”.
2. Load the dataset from the specified file path
3. (`Walmart_product_rating.csv`).
4. Transpose the ratings utility matrix.
5. Perform Truncated SVD with 10 components on the transposed matrix.
6. Compute the correlation matrix from the decomposed matrix.
7. Define a product ID 'i'. Retrieve the index of 'i' from the product names and compute the correlation of 'i' with other products.
8. Recommend products with a correlation coefficient greater than 0.90, excluding the already bought product 'i'.
9. Vectorize the product descriptions using TF-IDF vectorizer.
10. Fit K-Means clustering algorithm with 10 clusters to the vectorized data.
11. Plot the clusters and print the top terms per cluster.

Final outcome

The final outcome of the recommendation system is a user-friendly interface that seamlessly integrates both Content-Based and Collaborative Filtering methods. This interface provides personalized product recommendations tailored to each user's preferences and interactions, enhancing the overall e-commerce experience. By leveraging Content-Based Filtering, users receive suggestions based on the similarity of products they have interacted with previously. Additionally, Collaborative Filtering [16] takes into account similarities among users or items to offer recommendations that align with individual preferences. The integration of these methods ensures that users receive diverse and relevant product suggestions, increasing the likelihood of finding items of interest and improving user satisfaction. Through the user-friendly interface, customers can easily navigate and explore personalized recommendations, leading to enhanced engagement and increased sales for the e-commerce platform.

4. RESULTS

Table 1 presents performance metrics for different algorithms, contrasting them with the proposed algorithms outlined. Table 1 offers concise insights into the performance of diverse machine learning algorithms. Our proposed model,

comprising Collaborative Filtering and Content-Based Filtering, attains a 97% accuracy rate, surpassing other methods [17] such as Linear Support Vector Machine (LSVM), Decision Tree Algorithm (DT), Linear Regression (LR), and Collaborative Filtering and Textual Clustering (CB and TC).

The accuracy of various algorithms is presented in Table 1, and it's visually represented through a bar graph. Our proposed model, Collaborative Filtering and Content-Based Filtering [18] (Cf and CBF), achieves an accuracy of 97%. In comparison, Linear Support Vector Machine (LSVM) achieves 96% accuracy, Decision Tree (DT) reaches 94.54%, Linear Regression (LR) scores 94%, and Collaborative Filtering and Textual Clustering (CF and TC) achieves 93%. By comparing these models, it's evident that our proposed model achieves the highest accuracy [19] as in Figure 2.

Table 1. Performance metrics

Model	Accuracy (%)	Precision	Recall	F-Measure
CF and CBF	97	94	95	95
LSVM	96	93	95	94
DT	94.54	95	95	95
LR	94	93	95	94
CB and TC	93	91	95	93

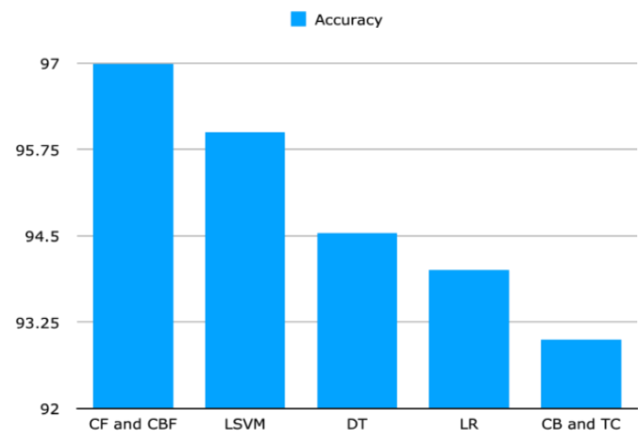


Figure 2. Bar graph showing accuracy of various algorithms

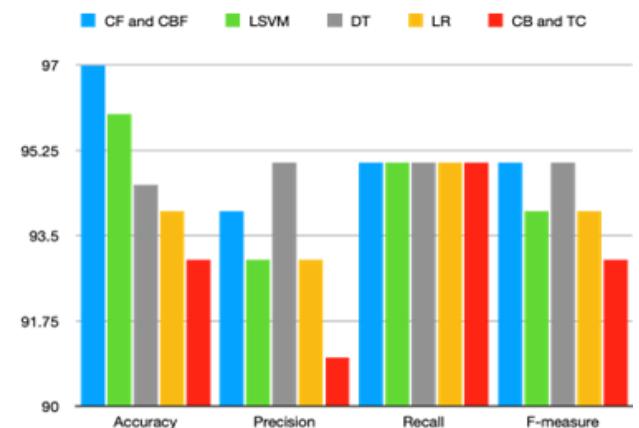


Figure 3. Accuracy, precision, recall, and F-measure are examples of performance metrics

Figures 2 and 3 illustrates performance metrics of various algorithms that are shown in Table 1. This representation provides clear information about the performance of various machine learning algorithms. Our proposed model achieves

higher accuracy than compare to above algorithms. The Screen shots of implementation are provided in Figures 4-10.

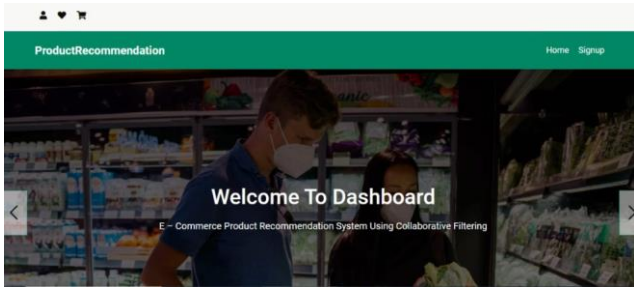


Figure 4. Dash board

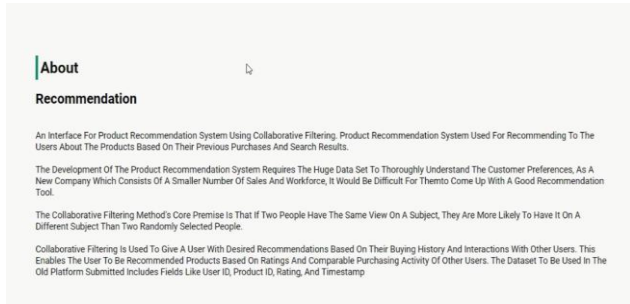


Figure 5. About page

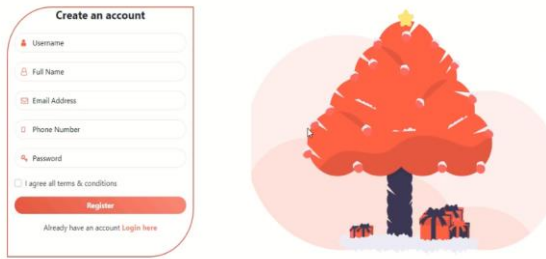


Figure 6. Create an account page



Figure 7. Login into account page



Figure 8. Welcome to dashboard page

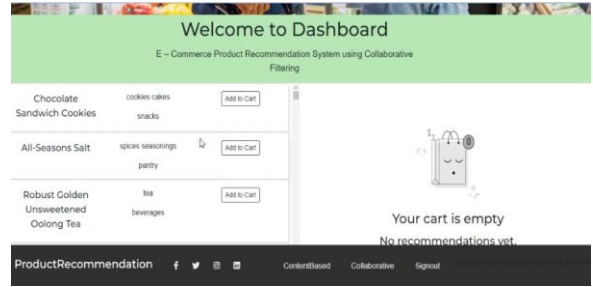


Figure 9. Product recommendation

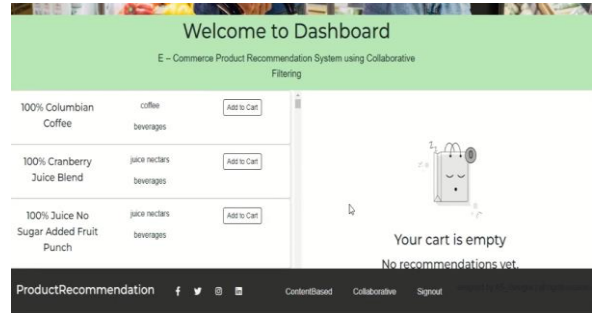


Figure 10. Output

The user-item rating matrix in Table 2 shows the ratings given by the users to the items. A blank cell means that the user has not rated the item.

Figure 11 illustrates which items are popular among the users and which items are missing ratings

Table 2. The user-item rating matrix

Users	Items				
	Item-1	Item-2	Item-3	Item-4	Item-5
User-1	4	3	5		
User-2		4	3		5
User-3	5		4	3	
User-4		5		4	3
User-5	3		5		4
User-1	4	3	5		

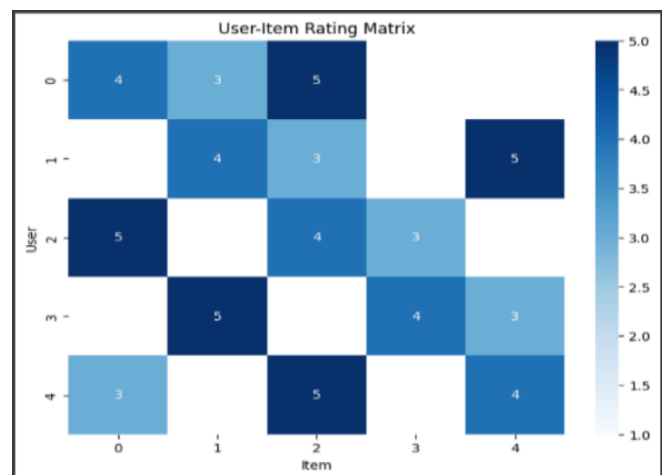


Figure 11. User-item rating matrix

The heat map shows the ratings as colors on a grid. A heat map can help you see which items are popular among the users and which items are missing ratings.

The user profile matrix in Table 3 shows the features or the

embedding of the users based on their ratings. The features are represented as vectors of numbers, which capture the semantic and contextual relationships between the users and the items.

Figure 12 illustrates the features or the embedding of the users based on their ratings.

Table 3. The user-profile matrix

S.No	Features				
	Feature1	Feature2	Feature3	Feature4	Feature5
1	0.41	0.36	0.22	0.72	0.19
2	0.37	0.39	0.24	0.68	0.22
3	0.43	0.34	0.21	0.74	0.18
4	0.36	0.41	0.23	0.66	0.24
5	0.39	0.38	0.25	0.70	0.21

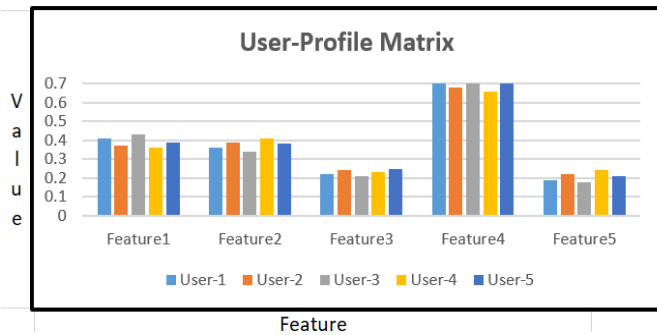


Figure 12. User-profile matrix

The user-profile matrix is represented by bar chart, which show the features or the embedding of the users and the items as bars of different heights. A bar chart can help you compare the similarities and differences between the users and the items based on their features.

The item profile matrix in Table 4 shows the features or the embedding of the items based on their ratings. The features are represented as vectors of numbers, which capture the semantic and contextual relationships between the items and the users.

Table 4. The item-profile matrix

S.No	Features				
	Feature1	Feature2	Feature3	Feature4	Feature5
1	0.32	0.45	0.12	0.67	0.23
2	0.28	0.51	0.16	0.63	0.27
3	0.35	0.42	0.18	0.69	0.21
4	0.26	0.53	0.14	0.61	0.29
5	0.31	0.47	0.17	0.65	0.25

These are the results of performing Truncated SVD with 3 components on the transposed ratings utility matrix. Each row represents a user, and each column in Table 5 represents a component extracted by the Truncated SVD algorithm. This process helps in reducing the dimensionality of the data while retaining its most important features.

Table 5. Truncated singular value decomposition matrix

S.No	Users	Components		
		Component-1	Component-2	Component-3
1	0.32	4.120	2.091	-0.040
2	0.28	3.664	-0.161	-0.199
3	0.35	4.946	-0.099	0.114
4	0.26	4.385	-0.147	0.163
5	0.31	4.203	2.664	-0.038

This correlation matrix in Table 6 shows the correlation between each pair of items based on their decomposition into the three components.

It represents providing the recommendations to the product searched by the user whose correlation values is equal to 0.9. From Table 7, recommended products with correlation coefficient greater than 0.90 (excluding 'Item 1' i.e. our searched product is item 1) is Item2.

TF-IDF vectorization converts text data such into the numerical vectors as in Table 8 by assigning weights to words based on their frequency and importance, facilitating effective analysis of textual data such as product descriptions. Since we have 3 products, each represented by TF-IDF vectors.

Table 6. Correlation matrix

S. No	Items	Components		
		Component-1	Component-2	Component-3
1	Item-1	1	0.985	0.029
2	Item-2	0.985	1	0.209
3	Item-3	0.209	0.209	1

Table 7. Item-correlation table

S.No	Items		
	Item-1	Item-2	Item-3
1	1	0	0
2	0	0.985	0
3	0	0	0.209

Table 8. Term frequency inverse document frequency vectorization

	Items		
	Item-1	Item-2	Item-3
Associated Vectors	[0.6191303, 0.0, 0.78528828, 0.0, 0.0]	[0.0, 0.6191303, 0.0, 0.78528828, 0.0]	[0.0, 0.0, 0.78528828, 0.0, 0.6191303]

5. CONCLUSION

In conclusion, this work has successfully developed an efficient e-commerce product recommendation system, addressing the increasing demand for personalized suggestions in today's competitive online market. By integrating Content-Based and Collaborative Filtering techniques, our system ensures extensive recommendation coverage, accommodating diverse user preferences. Through meticulous exploration and preprocessing of the "marketing_sample_for_walmart_com_walmart_com_product_review" Dataset, coupled with the deployment of recommendation models, our system achieves optimal performance in delivering relevant product suggestions. Collaborative filtering accurately identifies user patterns and similarities, while content-based filtering enriches recommendation variety. The user-friendly interface further underscores the system's efficacy, providing users with a smooth shopping experience. Looking ahead, future research could explore integrating additional data sources and advanced machine learning techniques to enhance the system's adaptability in meeting evolving user needs within the dynamic e-commerce landscape, ultimately driving increased

sales and customer retention.

Future Enhancements

The future prospects for e-commerce product recommendations utilizing collaborative filtering and content-based filtering encompass several avenues for advancement tailored to the project's title. Firstly, enhancing the recommendation system by integrating hybrid techniques that merge collaborative filtering with content-based filtering can notably enhance recommendation accuracy and diversity, aligning closely with the project's aim of leveraging both methods for optimal outcomes. Additionally, exploring contextual information such as user location, time of day, or device type can further enrich the relevance of recommendations, enhancing the personalized shopping experience as envisioned in our project.

Moreover, incorporating real-time updates and dynamic user feedback mechanisms into the recommendation system will ensure that suggestions remain current and reflective of evolving user preferences, crucial for maintaining user engagement and satisfaction in the competitive e-commerce landscape. Embracing emerging technologies like deep learning (DL) and natural language processing (NLP) holds vast potential for extracting new insights from user behavior data, complementing the collaborative filtering approach and enhancing the system's capabilities in providing tailored recommendations. Furthermore, expanding collaborative filtering beyond traditional user-item interactions to include social networks or influencer recommendations can deepen the recommendation process, offering users a more comprehensive and personalized shopping journey. Lastly, ongoing research and experimentation with innovative algorithms and data processing techniques will be vital for our project to stay ahead in recommendation system innovation and effectively meet the evolving demands of the e-commerce market.

REFERENCES

- [1] Linden, G., Smith, B., York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1): 76-80. <https://doi.org/10.1109/MIC.2003.1167344>
- [2] Konstan, J.A., Riedl, J. (2012). Recommender systems: From algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1-2): 101-123. <https://doi.org/10.1007/s11257-011-9112-x>
- [3] Anwar, T., Uma, V. (2021). Comparative study of recommender system approaches and movie recommendation using collaborative filtering. *International Journal of System Assurance Engineering and Management*, 12: 426-436. <https://doi.org/10.1007/s13198-021-01087-x>
- [4] Park, Y.J., Tuzhilin, A., Han, I. (2008). An empirical comparison of approaches for combining multiple recommender systems. *Information Fusion*, 9(3): 345-360.
- [5] Lops, P., Gemmis, M., Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook*, pp. 73-105. https://doi.org/10.1007/978-0-387-85820-3_3
- [6] Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46: 109-132. <https://doi.org/10.1016/j.knosys.2013.03.012>
- [7] Ekstrand, M.D., Riedl, J.T., Konstan, J.A. (2011). Collaborative filtering recommender systems. *Foundations and Trends® in Human-Computer Interaction*, 4(2): 81-173. https://doi.org/10.1007/978-3-540-72079-9_9
- [8] Khatter, H., Arif, S., Singh, U., Mathur, S., Jain, S. (2021). Product recommendation system for E-commerce using collaborative filtering and textual clustering. In *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 612-618. <https://doi.org/10.1109/ICIRCA51532.2021.9544753>
- [9] Dubey, G., Kumar, S., Kumar, S., Navaney, P. (2020). Extended opinion lexicon and ML-based sentiment analysis of tweets: A novel approach towards accurate classifier. *International Journal of Computational Vision and Robotics*, 10(6): 505-521. <https://doi.org/10.1504/IJCVR.2020.110640>
- [10] Khan, B.M., Mansha, A., Khan, F.H., Bashir, S. (2017). Collaborative filtering based online recommendation systems: A survey. In *2017 International Conference on Information and Communication Technologies (ICICT)*, pp. 125-130. <https://doi.org/10.1109/ICICT.2017.8320176>
- [11] Wang, X., Wang, C. (2017). Recommendation system of e-commerce based on improved collaborative filtering algorithm. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pp. 332-335. <https://doi.org/10.1109/ICSESS.2017.8342926>
- [12] Ricci, F., Rokach, L., Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender Systems Handbook*, pp. 1-35. Springer, Boston, MA.
- [13] Priya, S., Mansoor Hussain, D. (2017). Recommendation systems for E commerce: A review. *International Journal of Advanced Research in Computer and Communication Engineering*, 6(4): 500-504. <https://doi.org/10.17148/IJARCC.2017.6496>
- [14] Koren, Y. (2008). Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 426-434. <https://doi.org/10.1145/1401890.1401944>
- [15] Dev, A.V., Mohan, A. (2016). Recommendation system for big data applications based on set similarity of user preferences. In *2016 International Conference on Next Generation Intelligent Systems (ICNGIS)*, Kottayam, India, pp. 1-6. <https://doi.org/10.1109/ICNGIS.2016.7854058>
- [16] Kumar, M., Yadav, D.K., Singh, A., Gupta, V.K. (2015). A movie recommender system: Movrec. *International Journal of Computer Applications*, 124(3): pp. 7-11.
- [17] Walunj, S.G., Sadafale, K. (2013). An online recommendation system for e-commerce based on apache mahout framework. In *Proceedings of the 2013 annual conference on Computers and people research*, pp. 153-158. <https://doi.org/10.1145/2487294.2487328>
- [18] Chen, H.W., Wu, Y.L., Hor, M.K., Tang, C.Y. (2017). Fully content-based movie recommender system with feature extraction using neural network. In *2017 International Conference on Machine Learning and Cybernetics (ICMLC)*, pp. 504-509. <https://doi.org/10.1109/ICMLC.2017.8108968>
- [19] Zhao, Z.D., Shang, M.S. (2010). User-based

collaborative-filtering recommendation algorithms on
hadoop. In 2010 Third International Conference on

Knowledge Discovery and Data Mining, pp. 478-481.
<https://doi.org/10.1109/WKDD.2010.54>