engineering proceedings

MDPI

Proceeding Paper

A Recommendation System for E-Commerce Products Using Collaborative Filtering Approaches †

Neelamadhab Padhy *, Sridev Suman , T Sanam Priyadarshini and Subhalaxmi Mallick

School of Engineering and Technology, Department of Computer Science and Engineering, GIET University, Gunupur 765022, Odisha, India; 21cseds010.sridevsuman@giet.edu (S.S.); 21cseds041.tsanampriyadarshini@giet.edu (T.S.P.); 21cseds043.subhalaxmimallick@giet.edu (S.M.)
* Correspondence: dr.neelamadhab@giet.edu

† Presented at the 3rd International Electronic Conference on Processes—Green and Sustainable Process Engineering and Process Systems Engineering (ECP 2024), 29–31 May 2024; Available online: https://sciforum.net/event/ECP2024.

Abstract: The objective of this article is to recommend products using association rule mining from an E-commerce site. This helps us to recommend products through utilizing the filtering concept. In this article, we use the Apriori and FP-Growth algorithms. Our model not only suggests products but also gives tips on how to make strong suggestion systems that can deal with a lot of data and give quick responses. Our objective is to predict ratings so that the users could be recommended and buy products. There are 1,048,100 records in the dataset. This dataset consists of four features, and these are are follows: {user-id, productid, Ratings, and timing}. Here, we consider the rating as our dependent attribute, and others factors are independent features. In this article, we use collaborative filtering algorithms (SVD, SVD+, and ALS) and also item-based filtering techniques (KNNBasic) to recommend products. Apart from these, sssociation rule mining, hybridization of Apriori, and FP-Growth are used. K-means clustering is used to identify anomalies as well as to create a dashboard, using Power BI for data visualization. Apart from these, we have also developed a hybridization algorithm using Apriori and FP-Growth. Among all the recommendation algorithms, SVD outperforms in recommending the product, and the average RMSE and MAE values are 1.31, and 1.04, respectively.

Keywords: E-commerce; recommendation system; Apriori algorithm; FP-Growth; association rule mining; collaborative filtering (SVD, SVD+, ALS); item based filtering (KNNBasic)

Citation: Padhy, N.; Suman, S.; Priyadarshini, T.S.; Mallick, S. A Recommendation System for E-Commerce Products Using Collaborative Filtering Approaches. *Eng. Proc.* **2024**, *67*, 50. https://doi.org/10.3390/engproc2024067050

Academic Editor: Juan Francisco García Martín

Published: 24 September 2024

1. Introduction

E-commerce platforms have transformed how people shop, providing vast arrays of products and services. However, the abundance of choices can overwhelm customers, leading to decision fatigue. To address this, recommendation systems play a crucial role in filtering and suggesting products that align with user preferences, thus enhancing the shopping experience and boosting sales. This paper delves into the mechanisms of recommendation systems, focusing on algorithms such as Apriori and FP-Growth, which leverage association rule mining to discover meaningful patterns in user behavior. In e-commerce, recommender systems vary from manually operated to automated ones. Schafer, J. B., et al. [1] show the importance of person-to-person correlation in recommendations. For example, data sharing among non-competing sites can improve recommendations, and Amazon, CDNOW, eBay, and Reel have diverse systems with inputs like purchase data, scale ratings, text comments, and editor's choices. In the future, personalized recommendations may increase via collaborative filtering for better recommendations. Recommender policies include aggregating data for new users, sharing data across sites, and removing blatant negative reviews. Recommender systems play a pivotal part in increasing the efficacy of internet shopping sites by providing individualized product suggestions to users. Content-based and demographic-based techniques are commonly used. Hussien, F. T. A., et al. [2] discussed that systems can convert their client behavioral data into a single measurement by which they can differentiate between products that are worth spending on and those that are not while using various methods for calculating similarity. However, demographic-based methods are related to particular groups of individuals who have similar profiles. Collaborative filtering is widely used in this context where customer preferences and historical data determine personalized recommendations which lead to improved consumer experience and an increase in sales, as seen by Zhou, M., et al. [3]. Micro-behaviors in e-commerce recommendations are used to understand user interactions and how they affect product suggestions. The RIB (Recommender with Interpretable Behavior) model has been introduced, which models the consequences of micro-behaviors on recommendations through sequences of user actions in a very effective way. This framework has been experimentally tested with real e-commerce data and shows significant enhancements in recommendation performance compared to traditional approaches. Efficient association rule mining for recommender systems stresses the importance of confidence and support metrics in measuring the correlation between item sets, as seen by Lin, W. [4]. This research also explains a recommendation strategy that is based on user associations through like and dislike rules that aim to improve the performance of recommendation systems. The study highlights the significance of modern approaches which can help to improve the efficiency of recommendation systems. It looks into different variants of ASARM (Adaptive Support Association Rule Mining) which modify the minimum support to generate a specific number of rules that give better recommendations. It examines the precision–recall performance distribution across score thresholds as well as the effects of various users' liking probabilities on recommendation performance.

The system uses multi-dimensional association rule mining to come up with product recommendations customized as per customer profiles and transaction data. Clustering techniques are used to enhance efficiency by minimizing the time complexity of recommendation generation, as seen by Parikh, V., et al. [5] Real-time recommender systems pose computational challenges. This study showed why the adoption of recommendation systems is significant for improving customer satisfaction, boosting sales, and creating customer loyalty in the e-commerce sector.

Kumar, B., et al. [6] discussed the Markov chain model which uses probabilities to predict hidden states based on a transition matrix. In e-commerce, usability analysis and usability issues are found through metrics and user behavior analysis. Machine learning and association rule mining play an important role in this model. Data analytics in e-commerce is very important for inventory management, fraud detection, and customer personalization. Historical and statistical data are analyzed to gain an advantage over competitors. Machine learning models like logistic regression are good for predicting binary outputs as they use a non-linear sigmoid function, as seen by Dogan, O., et al. [7] Fuzzy Association Rule Mining (FARM) is an algorithm that focuses on sales amounts and improves the traditional association rule mining (ARM) approach, which only focuses on sales and not on the amounts. It makes rules businesses can use to improve their sales and understanding of their customers. FARM rules aid product picks for the customers of e-commerce platforms. FARM rules are more helpful than other rules that the traditional ARM method produces. FARM rules can give vendors smart ways to connect with their customers. Key decision-makers can use them to boost sales strategies. Company managers can obtain facts to suggest items customers may want. This increases buyers' interest and improves the sales results of the e-commerce platform. In summary, FARM rules work better to engage customers than ARM rules and improve sales performance. A study by Chen, A. H. L., et al. [8] focused on making a precise system for recommendations that uses the buying habits of customers and product-selling customer behaviors. It focuses on how to group customers using the RFM model, and RFM stands for Recency, Frequency, and Monetary value. The model uses data which are based on periods of time as well as on customer engagement factors. The analysis divides the customers into loyal customers and potential customer groups, and it sorts the products as best-sellers, profitable items, and VIP items. Clustering algorithms, ANOVA (Analysis of Variance), and ANOM (Analysis of Means) analyze and optimize the clustering parameters. Sreelakshmi, A., et al. [9] discussed the Apriori and FP-Growth algorithms to increase sales in a supermarket.

2. Material and Methods

The implementation of recommendation systems involves various algorithms and techniques like Apriori, FP-Growth, K-NN, and collaborative and content-based filtering approaches to recommend products. These techniques are called a collaborative filtering approach to recommend a product.

The below proposed model in Figure 1 is used for recommendation systems in the E-commerce system. There are three phases to the system.

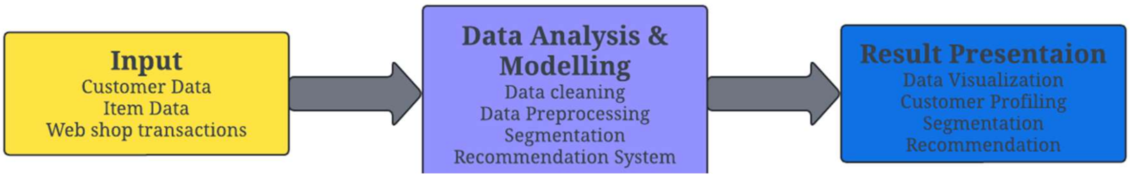

Figure 1. Proposed model for product recommendation system.

Phase-1: In this phase, we describe how to collect and preprocess the customer data, item data, etc. In the 1st phase, we input the features and create the DataFrame, and then it is ran through the data preprocessing system. In this phase, the data are cleaned by using the data normalization technique.

Phase-2: In the next phase, we use different algorithms to process the data and predict the product. There are 1,048,100 records in the dataset. This dataset consists of 4 features and these are as follows: {userid, productid, ratings, and timing}. Here, we consider the rating as our dependent attribute and others factors are independent features. By taking these features, we have developed the recommendation system. Our objective is to predict the rating so that the user can buy the product. This phase discusses the process of collecting and preprocessing the customer's data. Through this step, the data are cleaned for further use for decision-making purposes.

Phase-3: This phase entails customer segmentation and recommendation. In this step, we use the Apriori and FP-Growth algorithms and their hybridization unsupervised learning algorithm.

2.1. Apriori Algorithm

The Apriori algorithm is a classic method in association rule mining that identifies frequent item sets and derives association rules. It operates on the principle that if an item set is frequent, all its subsets must also be frequent. The algorithm proceeds in two steps: candidate generation and pruning. In the candidate generation step, the algorithm generates potential item sets of increasing lengths, while in the pruning step, it removes item sets that do not meet the minimum support threshold.

2.2. FP-Growth Algorithm

The FP-Growth method is a better option than Apriori. Instead of making possible item groups, FP-Growth creates a condensed data structure known as the FP-Tree. This keeps the item set connection information. Then, the method breaks down the FP-Tree again and again to obtain often-used item sets. This reduces the work on the computer more as compared to Apriori.

Table 1 discusses the computational cost of the individual algorithms and apart from these, how much time is taken for hybridizations (which is the combination of both algorithms).

Table 1. Computational cost between Apriori, FP-Growth and hybridization algorithm.

Algorithm Used	Execution Time in Seconds
Apriori	0.0761
FP-Growth	0.0662
Hybrid Algorithm	0.0043

3. Practical Implementation

This algorithm is particularly useful in scenarios where the dataset size is manageable, and the emphasis is on clear and understandable rules. In e-commerce, the algorithm can help identify patterns such as commonly co-purchased items, thereby aiding in the development of product bundles and cross-selling strategies.

3.1. K-Nearest Neighbours

The outliers depicted in Figure 2 as the output of the KNN algorithm can be considered as anomalies or outliers that have crossed the threshold distance for the recommendation model, and these outliers can be flagged as anomalies.

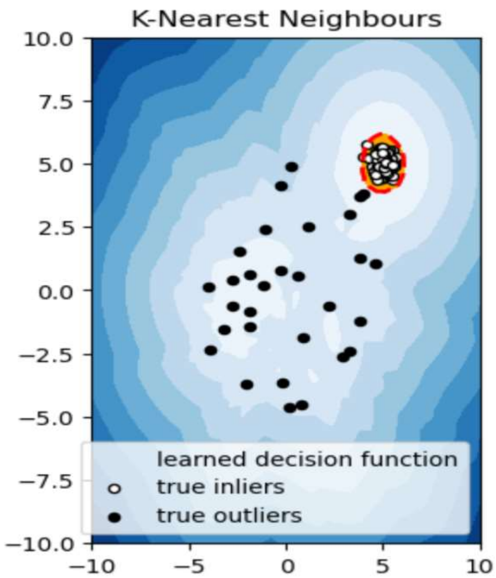

Figure 2. KNN classifier for anomaly detection.

3.2. The Preparation of the Recommender Model (Apriori + FP-Growth)

Training the model:

The recommendation system was trained on a dataset comprising user interactions with products, primarily focusing on user ratings. The FP-Growth and Apriori algorithms were employed to mine frequent item sets where items corresponded to products and ratings. The algorithms efficiently handled the sparsity of the rating data, identifying patterns of co-rated products and enabling the creation of association rules. The hybridization algorithm was developed for the product recommendation system and the accuracy obtained was 0.81%

Table 2 shows the performance analysis and Table 3 discusses the top 10 recommended products.

Table 2. Classification report.

Precision–Recall	f1-Score Support
3 0.80 0.82 0.85 1	
4 0.82 0.78 0.79 1	
accuracy 0.80 2	
macro avg 0.81 0.80 0.81 2	
weighted avg 0.81 0.80 0.802	

Table 3. Top 10 recommendationed product using the hybridization algorithm of the user AKM1MP6P0OYPR.

Rank	Product ID	Rating
1	B000053HC5	5
2	B000053HH5	5
3	B000144I2Q	5
4	B00022VZ1O	5
5	B0002E52S4	5
6	B00077AA5Q	5
7	B00077KMXG	5
8	B0007X7W6K	5
9	B000CMS0XU	5
10	B000FQ2JLW	5

3.3. Isolation Forest

Isolation forest is an unsupervised machine learning algorithm commonly used for anomaly detection. It can be implemented to detect anomalies in the user's behavior and anomalies in product sales. It uses binary trees to detect anomalies. It detects anomalies by identifying the outliers in the dataset, as shown in Figure 3 below.

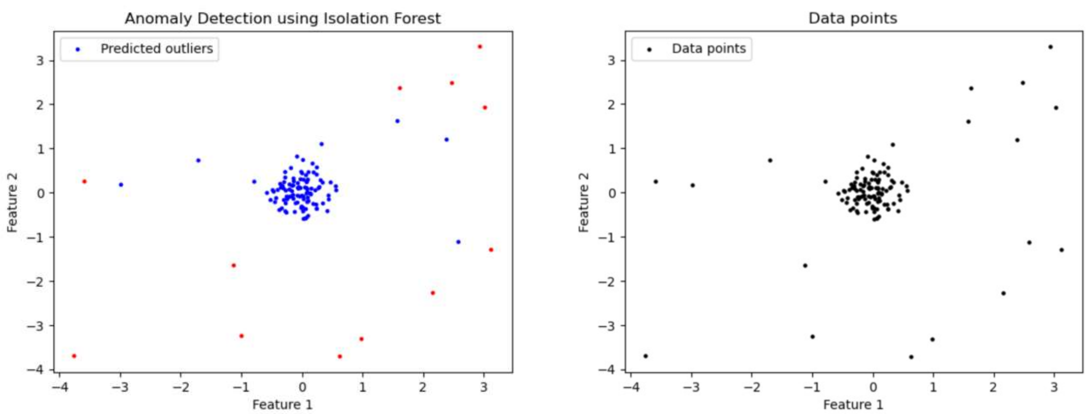

Figure 3. Dashboard's plot.

The data points are plotted in the first graph; the isolation forest formed in the middle is the forest of similar data based on the ratings and the points marked far away from the forest are the outliers or the anomalies present in the ratings dataset.

The violin plot in Figure 4 shows the normal product ratings data in the representation, and the anomalies present in the product ratings data are represented beside it. These representations are based on two widely used classifiers, namely the Naïve Bayes algorithm and the K-Nearest Neighbors (KNN) algorithm, which are for the classification of the dataset and to figure out the anomalies from the vast dataset of customer feedback about the various products.

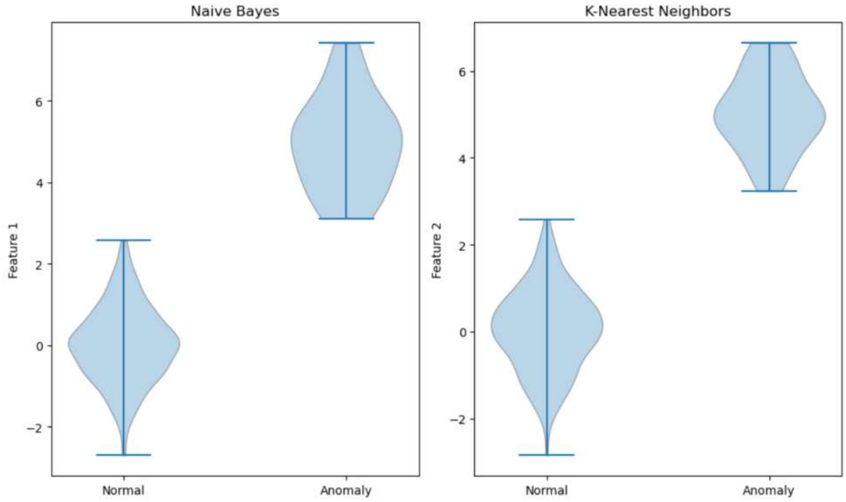

Figure 4. Normal product ratings data.

3.4. Cluster Analysis

Cluster analysis is a method of data mining in which the data are analyzed and smaller groups are created with similar items. It is used to identify relationships and patterns hidden in the data. It is an unsupervised machine learning-based algorithm that can function with unlabeled data. It divides the data into multiple groups, also known as clusters.

Users can be grouped into clusters for similar interests and preferences. These clusters can be used to personalize recommendations for a particular user, which enhances the experience of customers. Figure 5 below demonstrates similar interests and preferences expressed as clusters.

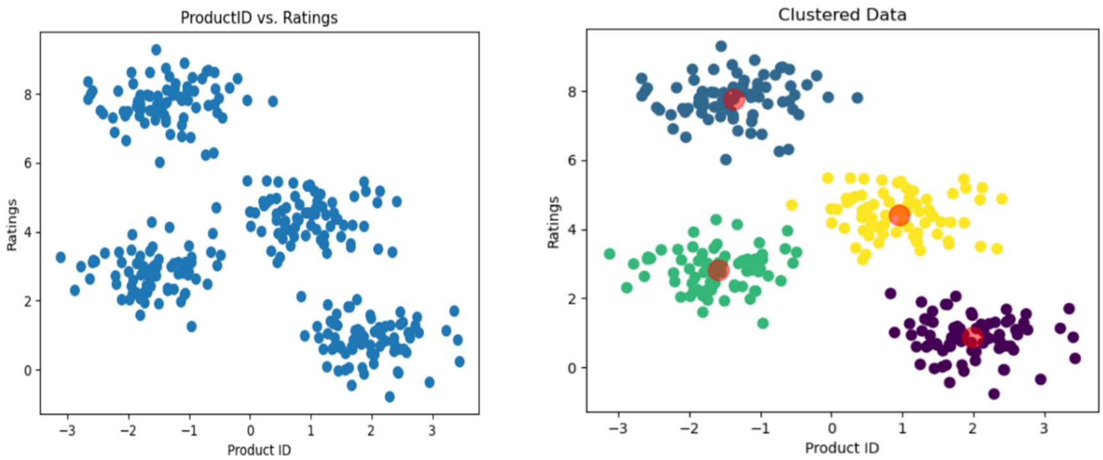

Figure 5. Similar interests and preferences.

3.5. Collaborative Filtering Evaluation

SVD: This is one of the matrix factorization techniques that is used for collaborating and filtering purposes. The matrix is usually divided into three other matrices and these are U means user features, Σ(sigma) means singular values, and VT (item features). These three components are useful for further decomposition of the dimensionality of the data as well as identifying the patterns and relationships between the users and items. It is primarily used for prediction-improving purposes.

SVD++ (SVD with Implicit Feedback): This method is primarily used for implicit feedback, such as clicks, views, and purchases. It is an extended version of SVD. It helps improve accuracy and can also understand user behavior when explicit ratings are missing. When needed for both implicit as well as explicit actions, we use SVD++ for accurate recommendation purposes.

ALS (Alternating Least Squares): This is another factorization method. It is also similar to SVD but here it is primarily used for solving optimization problems. One use is for user factors, and another is for item factors. It is suited for parallel computation as well as dealing with huge amounts of datasets. It is also used to fill in missing data in the user–item matrix.

Item-based Collaborative Filtering (KNNBasic): This is another type of collaborative filtering approach and we have implemented it through the K-NN classifier. It focuses more on items and calculates items based on ratings and items that are recommended. This is also called similarity-based recommendation.

RQ: Is it possible to quantify the performance of different collaborative filtering approaches used to compare in terms of RMSE and MAE values for the given e-commerce dataset?

The solution to RQ:

Yes, it is possible to compare the performance metrics of different collaborative filtering algorithms. From the above graph, it is concluded that SVD performs well in comparison to other SVD++, and ALS in terms of RMSE and MAE metrics. ALS has the highest number of errors, indicating that it is not suitable for recommendation purposes. Figure 6 demonstrates the comparison of collaborative filtering approaches in terms of the performance metrics RMSE and MAE.

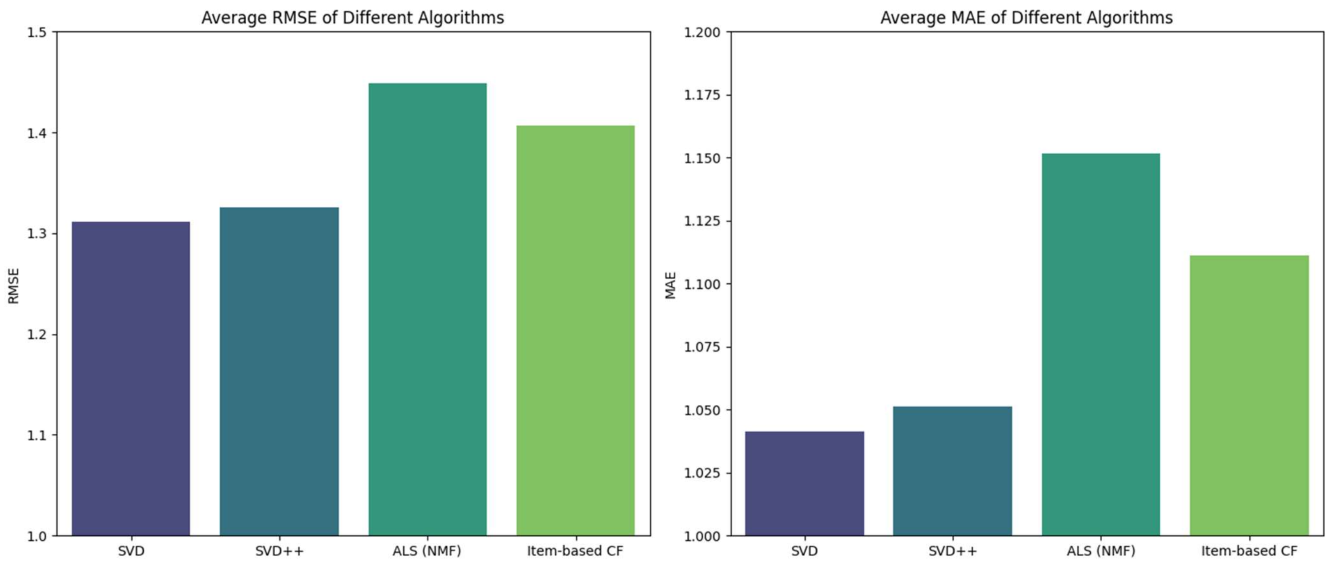

Figure 6. Performance comparison of different collaborative filtering approaches.

SVD (Singular Value Decomposition)

This is one kind of recommendation algorithm that helps to choose the right product, and it works using the matrix factorization method.

3.5.1. SVD Algorithm (Singular Value Decomposition)

Table 4 shows the performance evaluation of SVD algorithms.

Table 4. Evaluating RMSE and MAE metrics of algorithm SVD on 3 split(s).

	Fold 1	Fold 2	Fold 3	Mean	Std
RMSE (testset)	1.3116	1.3101	1.3130	1.3116	0.0012
MAE (testset)	1.0414	1.0404	1.0422	1.0414	0.0007
Fit time	24.08	24.19	26.55	24.94	1.14
Test time	4.50	3.74	4.69	4.31	0.41
Average RMSE			1.3115663911238664		
Average MAE			1.041355101469066		

3.5.2. SVD++ (SVD with Implicit Feedback)

Table 5 shows the performance evaluation of SVD++ algorithms.

Table 5. Evaluating RMSE and MAE metrics of algorithm SVDpp on 3 split(s).

	Fold 1	Fold 2	Fold 3	Mean	Std
RMSE (testset)	1.3238	1.3245	1.3275	1.3253	0.0016
MAE (testset)	1.0505	1.0513	1.0525	1.0514	0.0008
Fit time	10.44	10.65	9.25	10.11	0.61
Test time	2.81	2.41	3.07	2.76	0.27
Average RMSE (SVD++)			1.3252702367404818		
Average MAE (SVD++)			1.0513992653534794		

3.5.3. ALS (Alternating Least Squares)

Table 6 shows the performance evaluation of ALS algorithms.

Table 6. Evaluating RMSE and MAE metrics of algorithm NMF on 3 split(s).

	Fold 1	Fold 2	Fold 3	Mean	Std
RMSE (testset)	1.4475	1.4518	1.4464	1.4485	0.0023
MAE (testset)	1.1512	1.1547	1.1494	1.1518	0.0022
Fit time	38.48	39.01	41.98	39.82	1.54
Test time	1.31	1.91	2.41	1.87	0.45
Average RMSE (ALS)			1.4485265623308452		
Average MAE (ALS)			1.1517890526633938		

3.5.4. Item-Based Collaborative Filtering Recommended Algorithm

Item-based collaborative filtering (Item-based KNN (KNNBasic))

Table 7 shows the performance evaluation of the item-based collaborative filtering recommended algorithm.

1. Apriori, FP-Growth, and Hybridization (Apriori, FP-Growth)
2. SVD (Singular Value Decomposition)
3. SVD++ (SVD with Implicit Feedback)
4. ALS (Alternating Least Squares)
5. Item-Based Collaborative Filtering (KNNBasic)
6. Cluster Analysis

Table 7. Evaluating RMSE and MAE metrics of algorithm KNNBasic on 3 split(s).

	Fold 1	Fold 2	Fold 3	Mean	Std
RMSE (testset)	1.4081	1.4084	1.4049	1.4071	0.0016
MAE (testset)	1.1126	1.1110	1.1110	1.1115	0.0007
Fit time	25.82	22.09	22.21	23.38	1.73
Test time	1.92	2.27	1.74	1.98	0.22
Average RMSE (Item-based CF)			1.4071198922254247		
Average MAE (Item-based CF)			1.1115450913336984		

The above algorithms were used in our research work. Apriori + FP-Growth were used for generating the\association among the products. Then, cluster analysis was used for the segmentation of the products into particular groups. Item-based collaborative filtering (KNNBasic) was used for initial recommendations. The rest of the algorithms were used for personalized recommendation of the products as well as for scalability purposes.

4. Data Visualization of Product Recommendation Using Power BI Techniques

In the below-mentioned figure, we created one dashboard through which we visualized the data properly.

Dashboard for Data Visualization

The dashboard's plot in Figure 7 shows the sum of ProductID and USERID, which can be interpreted as products bought by particular users by their USERID, which is also used for the legend of the plot. The other plot in Figure 8 shows the sum of ProductID and ratings which can be interpreted as products and their ratings given by the users, where the ratings are ranged between one to five, and where ratings are used as the legend for the plot. The other plot shows the count of USERID and the ratings which can be interpreted as the ratings given by the users in the range of one to five.

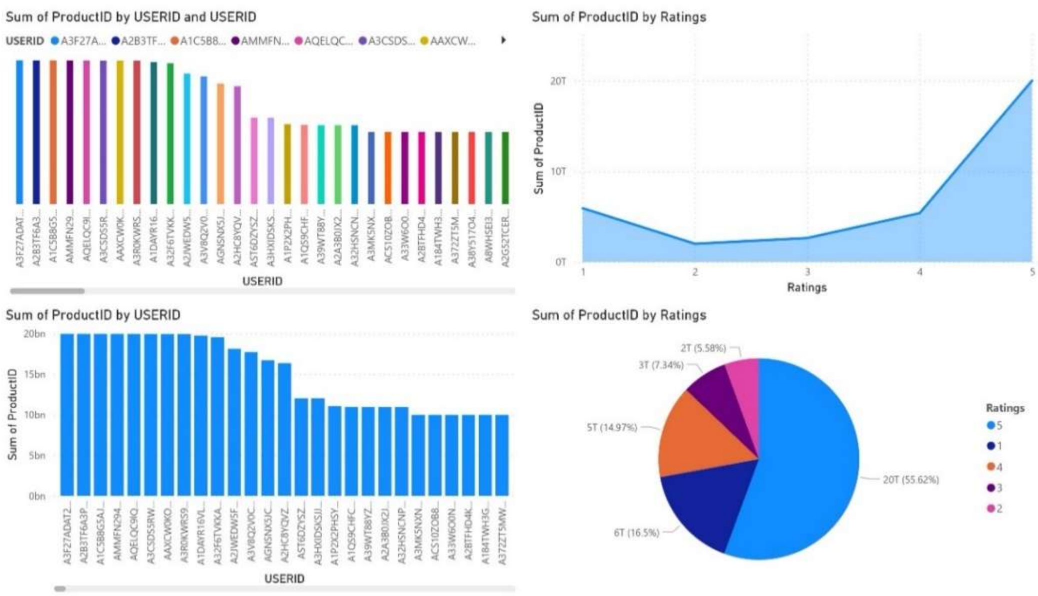

Figure 7. Power BI dashboard for data visualization.

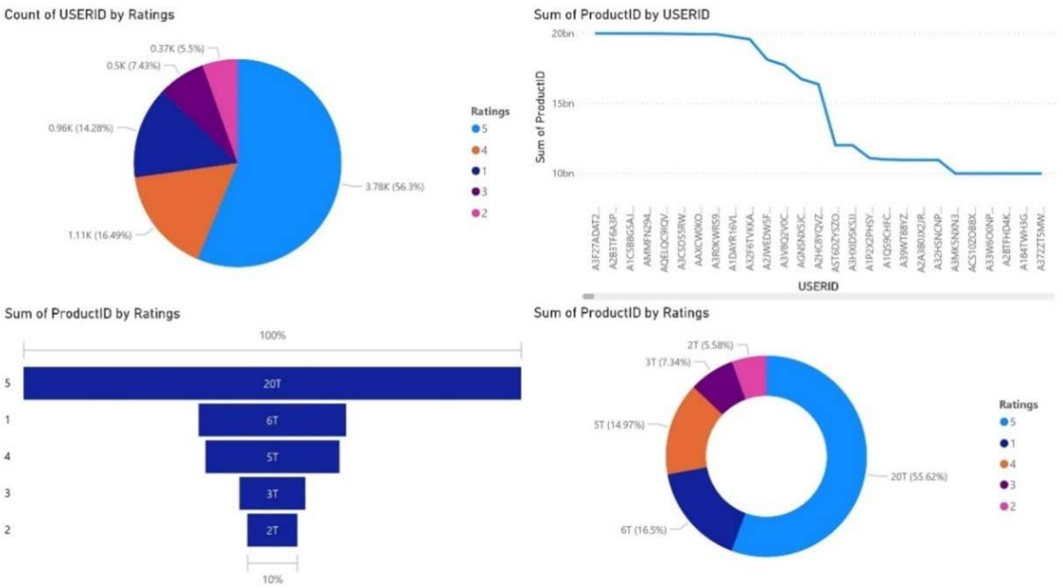

Figure 8. ProductID and rating visualization.

Benefits and limitations of the algorithm

1. **SVD (Singular Value Decomposition):**

Benefits: We have used this algorithm for dimensionality reduction purposes. **Limitations**: When the dataset size is large, it is computationally expensive.

2. **SVD++ (SVD with Implicit Feedback):**

Benefits: When implicit feedback is required, especially for clicks, views, and purchases, it provides better accuracy rather than SVD++. Limitations: It requires more computational resources than SVD. As we used this method for implicit feedback that required additional parameters the complexity of the model increases.

3. **ALS (Alternating Least Squares):**

Benefits: In comparison to the above two approaches, this method is highly scalable because it can handle huge amounts of data. This algorithm's nature is parallel computational. It effectively handles missing values from the dataset. **Limitations**: There might be an overfitting risk if hyperparameters are not properly handled.

4. **Item-Based Collaborative Filtering (KNNBasic):**

Benefits: It is one of the simple recommendation models that is used for item-by-item purposes. It effectively handles the cold start problem as well as sparse data. **Limitations**: The recommendation quality depends on the quality as well as the quantity of the item present in the dataset.

We used the collaborative filtering approach on the e-commerce dataset. Based on this, we predicted the user-item interactions. RMSE as well as MAE metrics are estimated. To improve the accuracy of product recommendations:

1. We have implemented algorithms like SVD, SVD++, ALS, and an item-based collaborative filtering approach for accurate recommendation.
2. Apriori and FP-Growth algorithms were used for generating the most frequent items as well as finding their association rule, recommending a strong association value that suggests that the items could be purchased for together.

By utilizing the above algorithms, we achieved the following objectives:

1. The matrix factorization technique plays an important role in product recommendation and association rule mining (Apriori, FP-Growth).
2. For enhancing personal recommendations, we used the SV++ algorithm.

3. For improving scalability, we used the ALS algorithm. Apart from these, we also checked the hybridization of Apriori and FP-Growth, which generates strong association rules.
4. For diversity recommendations, we used cluster analysis.

5. Conclusions

Recommendation systems are integral to the success of e-commerce platforms, providing personalized shopping experiences that drive user engagement and sales. The Apriori and FP-Growth algorithms are powerful tools in association rule mining, each with their strengths and limitations. The collaborative filtering approach is used for product recommendations like SVD, SVD++, ALS, and the item-based filtering approach (KNNBasic). Out of all the filtering approaches, only SVD provides prominent results. SVD algorithms such as RMSE and MAE produce fewer results in comparison to other algorithms. Apart from these, a hybridization algorithm was developed to recommend products, and its obtained accuracy was 81%. Future research should focus on hybrid models that combine the efficiency of FP-Growth with the interpretability of Apriori and should explore other data mining techniques to further enhance recommendation accuracy.

Author Contributions: Conceptualization, N.P. and S.S.; methodology, T.S.P.; software, S.M.; validation, N.P., S.S. and S.M.; formal analysis, S.S.; investigation, S.M.; resources, N.P.; data curation, S.S.; writing—original draft preparation, S.S.; writing—review and editing, N.P.; visualization, S.M.; supervision, S.S.; project administration, S.M.; funding acquisition, N.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Supporting data for the study are available at the Kaggle repository.

Acknowledgments: We would like to thank the School of Engineering and Technology, GIET University, for providing the resources and support necessary for this research, and also, we would like to express our gratitude to our supervisor, Neelamadhab Padhy, for his invaluable guidance and support throughout this research. We also thank our peers and families for their encouragement and assistance.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Schafer, J.B.; Konstan, J.; Riedl, J. Recommender systems in e-commerce. In Proceedings of the 1st ACM Conference on Electronic Commerce, Denver, CO, USA, 3–5 November 1999; pp. 158–166.
2. Hussien, F.T.A.; Rahma, A.M.S.; Wahab, H.B.A. Recommendation systems for e-commerce systems an overview. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021; Volume 1897, p. 012024.
3. Zhou, M.; Ding, Z.; Tang, J.; Yin, D. Micro behaviors: A new perspective in e-commerce recommender systems. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Los Angeles, CA, USA, 5–9 February 2018; pp. 727–735.
4. Lin, W.; Alvarez, S.A.; Ruiz, C. Efficient adaptive-support association rule mining for recommender systems. *Data Min. Knowl. Discov.* **2002**, *6*, 83–105. [CrossRef]
5. Parikh, V.; Shah, P. E-Commerce recommendation system using Association rule mining and clustering. *Int. J. Innov. Adv. Comput. Sci.* **2015**, *91*, 944–952.
6. Kumar, B.; Roy, S.; Sinha, A.; Iwendi, C.; Strážovská, L'. E-commerce website usability analysis using the association rule mining and machine learning algorithm. *Mathematics* **2022**, *11*, 25. [CrossRef]
7. Dogan, O.; Kem, F.C.; Oztaysi, B. Fuzzy association rule mining approach to identify e-commerce product association considering sales amount. *Complex Intell. Syst.* **2022**, *8*, 1551–1560. [CrossRef]

8. Chen, A.H.L.; Gunawan, S. Enhancing Retail Transactions: A Data-Driven Recommendation Using Modified RFM Analysis and Association Rules Mining. *Appl. Sci.* **2023**, *13*, 10057. [CrossRef]
9. Sreelakshmi, A.; Padhy, N.; Senapaty, M.K. An optimized approach towards increasing the sale rate in a Grocery Mart by using Association Rule Mining Approaches. In Proceedings of the 2024 International Conference on Emerging Systems and Intelligent Computing (ESIC), Bhubaneswar, India, 9–10 February 2024; IEEE: New York, NY, USA, 2024; pp. 538–543.