TRANSPUBLIKA
Precise. Resilience. Felicitous.

Original Article

Online Store Product Recommendation System Using Collaborative Filtering and Content-Based Filtering Algorithms to Increase Sales

Yosi Afandi[1], Maskur[2*], Wahyu Widyananda[3], Nilawati Fiernaningsih[4], Lina Budiarti[5,] Fahmi Muhammad Az Zuhri[6]

[1-6] Department of Business Administration, Politeknik Negeri Malang, Indonesia
Email: [1] yosi.afandi@polinema.ac.id, [2] maskur@polinema.ac.id, [3] wahyu.widyananda@polinema.ac.id, [4] nilafh@polinema.ac.id, [5] linabudiarti@polinema.ac.id, [6] fahmizuhri@polinema.ac.id

Abstract

This study aims to evaluate and compare the performance of two recommendation system approaches, namely Collaborative Filtering (CF) and Content-Based Filtering (CBF), in providing relevant product recommendations to users in an e-commerce context. The dataset used consists of 120 data including 90 relevant and recommended products (True Positive), 20 recommended but irrelevant products (False Positive), and 10 relevant but not recommended products (False Negative). Based on the calculation results, both methods show a precision value of 0.818 and a recall of 0.900. This means that approximately 81.8% of products recommended by the system are truly relevant, while 90% of the total relevant products are successfully recommended to users. The F1-score value obtained of 0.857 illustrates a good balance between the accuracy and completeness of the recommendations generated by the system. Furthermore, to measure the level of rating prediction error, the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) metrics are used. The evaluation results show that the CF method has an MSE value of 0.0784 and an RMSE of 0.28, while the CBF method shows an MSE of 0.0961 and an RMSE of 0.31. The lower RMSE value of CF indicates that this method has better accuracy in predicting user preferences than CBF. Overall, both methods show good performance with a low error rate. However, CF proved slightly superior in providing recommendations that match user preferences, so it can be used as a basis for developing smarter and more personalized recommendation systems on e-commerce platforms.

Keywords: E-Commerce, Collaborative Filtering, Content-Based Filtering, Precision, Recommendation System.

1. Introduction

The rapid development of information technology over the past two decades has brought fundamental changes to various aspects of human life. One of the most significant changes has been in the way people shop and conduct economic transactions (Lee & Kim, 2023). Whereas in the past, buying and selling could only be done in person at physical stores, these activities have now shifted massively to the digital realm through online store platforms. E-commerce platforms such as Tokopedia, Shopee, Lazada, Amazon and others have become the primary means for millions of consumers around the world to quickly find, compare, and purchase products without geographical or time constraints (Liao & Sundar, 2022). However, this ease of access also brings new challenges for online businesses. With millions of products available on a single platform, users often face information overload a condition where too many choices make it difficult for consumers to make decisions (Kulvinder Singh et al., 2024). This situation can reduce user satisfaction and even lead to a decline in sales conversions. Therefore, a

446

mechanism is needed to help consumers find products that are most relevant to their needs and preferences. This is where the role of recommendation systems becomes crucial (Patoulia et al., 2022).

Recommendation systems are one of the most important applications in the field of artificial intelligence and machine learning, designed to predict and suggest items that users are most likely to be interested in (Parthasarathy & Sathiya Devi, 2023). By analyzing historical data such as purchase behavior, reviews, ratings, and product characteristics, recommendation systems are able to generate personalized product lists (Gulzar et al., 2023). This not only enhances the user experience, but also has the potential to increase sales and customer loyalty. In the context of online business, recommendation systems have proven to be one of the most effective strategies for driving sales growth (Thakker et al., 2021). This data shows that the ability to provide targeted recommendations is not just an additional feature, but a highly strategic competitive advantage (Sinha & Srivastava, 2021).

One popular approach to building recommendation systems is Collaborative Filtering (CF), which works by analyzing the behavior and preferences of other users who have similarities (Parthasarathy & Sathiya Devi, 2023). This approach assumes that if two users have similar interest patterns in the past, then they are also likely to like the same items in the future. On the other hand, the Content-Based Filtering (CBF) approach focuses on the characteristics or features of the item itself, such as description, category, price, or other attributes, to recommend products similar to items that users have liked in the past. Although both have their own advantages, each also has limitations (Li et al., 2021). Collaborative Filtering often faces the cold-start problem, which is the difficulty of providing recommendations to new users who do not yet have an interaction history. Meanwhile, Content-Based Filtering tends to produce less diverse recommendations because the system only displays items similar to those that have been liked before. Therefore, there is a need to combine these two approaches to obtain more accurate and balanced results (K. J. Lee et al., 2021).

In online store systems, one of the main problems is how to efficiently display the most relevant products for each user. As the number of products increases, manual searches become ineffective, and users may lose interest before finding the right product (Xia & Wang, 2025). Many users end up abandoning their shopping carts because they feel overwhelmed by the many choices. Another problem is how to increase sales conversions through a more personalized approach (Sharma et al., 2021). Not all users have the same preferences, so general promotional strategies are not always effective. By implementing a recommendation system that can understand individual user behavior, online stores can provide a more personalized experience, increase customer satisfaction, and ultimately increase sales volume. However, implementing an effective recommendation system is not a simple task (Murty et al., 2022). The available data is often very large and poorly structured. In addition, the diversity of user preferences and product characteristics means that recommendation models must be able to adapt dynamically (Torkashvand et al., 2023). A single approach, such as using only Collaborative Filtering or only Content-Based Filtering, tends to be insufficient to handle this complexity. Therefore, a combination of the two known as a hybrid recommendation system is a promising solution (Alshareet & Awasthi, 2025). This study aims to design and implement an Online Store Product Recommendation System using Collaborative Filtering and Content-Based Filtering algorithms. This system can provide more accurate, relevant, and personalized product recommendations to users, thereby increasing customer satisfaction and ultimately driving sales growth. The recommendation system is capable of providing recommendations to potential buyers for products that they wish to purchase (Ajaegbu, 2021).

2. Literature Review

2.1. Basic Concepts of Recommendation Systems

Recommendation systems are an important field in machine learning and data science that aim to help users find items or products that are relevant to their needs and preferences. Recommendation systems are software systems that provide suggestions for items that are likely to be liked by specific users, based on analysis of user data and item characteristics. In the context of online stores, recommendation systems have become a key component in improving customer satisfaction and increasing sales opportunities (Renukadevi et al., 2024). These systems work by utilizing user behavior data such as purchase history, ratings, visit times, and interactions with products to identify preference patterns. Thus, the system can provide personalized product recommendations for each user (Thakker et al., 2021). Recommendation systems have a strategic function, which is to help reduce the cognitive load on users when choosing products from thousands of available options. In addition, the system also plays an important role in marketing strategies such as cross-selling (offering complementary products) and up-selling (offering higher-value versions of products) (Jozani et al., 2023). On the business side, the existence of a recommendation system can increase customer retention, strengthen loyalty, and significantly increase revenue. In general, there are three main approaches that are often used in the development of recommendation systems, namely Collaborative Filtering, Content-Based Filtering, and Hybrid Filtering. Each approach has its own advantages and limitations, which will be discussed in depth in the following section (Zhong & Ding, 2022).

2.2. Collaborative Filtering

2.2.1. Definition and Basic Principles

Collaborative Filtering (CF) is one of the most popular methods in recommendation systems. This approach works based on the assumption that users who have similar preferences in the past are likely to have similar interests in the future (Renukadevi et al., 2024). In other words, CF utilizes the wisdom of the crowd, which is learning behavior patterns from many users to generate recommendations for other users (Alshareet & Awasthi, 2025). Collaborative Filtering can be divided into two main categories:

1. User-Based Collaborative Filtering (UBCF)

This approach searches for other users who have similar preference patterns to active users (targets). Once a group of similar users is found, the system will recommend products that are liked by those users but have never been accessed by the target user (Xia & Wang, 2025).

2. Item-Based Collaborative Filtering (IBCF)

This approach focuses on the relationship between items, not between users. The system calculates the level of similarity between products based on user rating patterns. Products that have a high level of similarity with products that users have liked or purchased will be recommended (Iwendi et al., 2022).

As for the Advantages of Collaborative Filtering is as follows:

1. Does not require detailed information about items (product features).
2. Can capture complex preference patterns between users.
3. Suitable for dynamic data because it can be continuously updated based on user interactions.

Further, the weaknesses of Collaborative Filtering is as follows:

1. Cold-start problem: it is difficult to provide recommendations for new users or new products because there is not enough interaction data.

2. Sparsity problem: most users only give a few ratings, causing the data to be sparse and making it difficult to find similarities.
3. Scalability issue: if the number of users and products is very large, calculating similarities becomes computationally expensive.

2.3. Content-Based Filtering

2.3.1. Definition and Basic Principles

Content-Based Filtering (CBF) is a recommendation method based on the characteristics or content of the item itself. This approach seeks to recommend products that are similar to those that users have previously liked or consumed. For example, if a user has previously purchased a gaming laptop with certain specifications, the system will recommend other laptops with similar specifications such as a fast processor, large RAM, and a dedicated GPU. CBF works by extracting features from each product, then using similarity measures such as cosine similarity or Euclidean distance to find the most similar items (Murty et al., 2022).

As for the advantages of Content-Based Filtering is as follows
1. It does not depend on other user data, so it can work even if there are only a few users.
2. It does not face cold start issues for new users as long as item data is available.
3. Recommendations are more controlled and transparent because they are based on clear product features.

2.4. Previous Research

Many studies have been conducted to develop recommendation systems in the context of online stores and other domains. Some of them show the effectiveness of combining Collaborative Filtering and Content-Based Filtering in improving recommendation accuracy and user satisfaction (Nguyen et al., 2024).

2.4.1. Collaborative Filtering-Based Research

Shows that User-Based Collaborative Filtering is effective in improving the quality of movie recommendations by utilizing correlations between users. However, they also highlight limitations on a large scale due to the high cost of calculating similarities between users. Introduced the Matrix Factorization approach as a solution to the problem of data sparsity. This technique later became the basis for large-scale recommendation systems such as those used by Netflix and Spotify (Zhong & Ding, 2022).

2.4.2. Content-Based Filtering Research

Content-Based Filtering is very useful in the context of products with rich descriptions, such as books, movies, or electronic devices. They use TF-IDF-based text representations to measure the similarity between items. However, they also found that the system tends to give results that are too narrow (monotonous), so users rarely find new products outside their initial preferences (Kulvinder Singh et al., 2024).

2.5. Implementation of Recommendation Systems in E-Commerce

In practice, recommendation systems have been widely used by various global e-commerce platforms.
1. Amazon uses item-based collaborative filtering to display products that are often purchased together.
2. Netflix and Spotify rely on matrix factorization to understand user preferences based on viewing or listening behavior.
3. Shopee and Tokopedia in Southeast Asia also implement hybrid models that consider click behavior, purchase history, and product content (Zhong & Ding, 2022).

The effectiveness of recommendation systems in e-commerce is not only measured by prediction accuracy, but also by business impacts such as increased conversion rates, average

order value, and customer retention. The implementation of effective recommendation systems can increase sales by 20–30% and significantly improve customer engagement (Sinha & Srivastava, 2021).

3. Methods

This chapter describes in detail the approach, methods and stages used in research to develop an online store product recommendation system using Collaborative Filtering and Content-Based Filtering algorithms. The main objective of this methodology is to design and implement a system capable of providing relevant and personalized product recommendations, as well as measuring the extent to which the system can contribute to increasing sales potential.
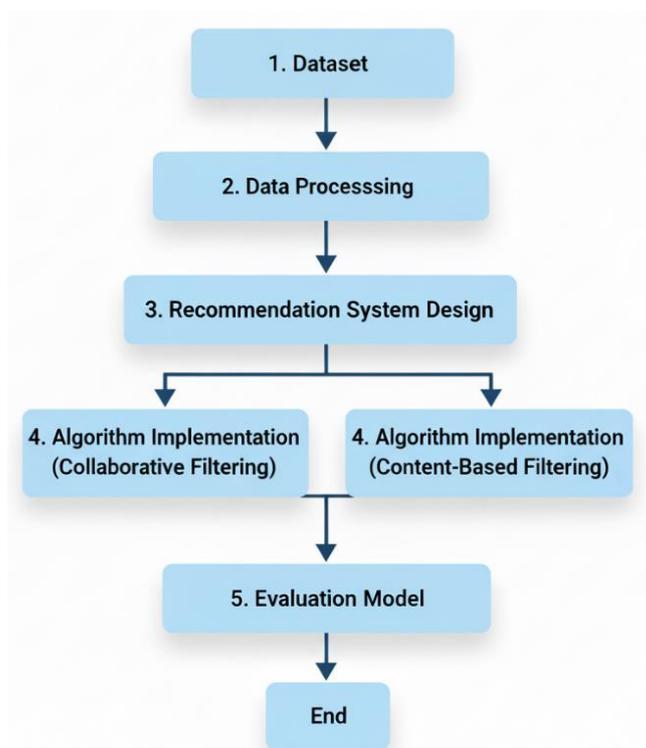

Figure 1. Research Methodology Flowchart for Recommendation System

3.1. Dataset

The data in this study was obtained from simulated online store transaction data, to compare two algorithm models, CF and CBF. Generally, two types of data were used:
1.   User Data

Contains information about user behavior and preferences, such as User ID, Purchase history, Product ratings, Products viewed or clicked on, Interaction time.
2.   Product Data (Item Data)

Contains characteristics or features of each product, including Product ID, Product name, Category, Text description, Price and technical specifications.

3.2. Data Preprocessing

Before the data is used in the model training process, the following preprocessing steps are performed:

1. Data Cleaning

Removing duplicate entries, addressing missing values, and correcting inconsistent data. For example, if a user gives multiple ratings for the same product, only the last value is used.

2. Data Transformation
   a. Normalization of rating values to a uniform scale (e.g., 1–5).
   b. Tokenization and stemming of product description text.
   c. Removal of stop words to improve content feature quality.

3. Feature Extraction
   a. In Content-Based Filtering, text features are converted into numerical representations using TF-IDF (Term Frequency–Inverse Document Frequency) or Word Embedding (Word2Vec/BERT) techniques.
   b. In Collaborative Filtering, a user-item interaction matrix is created.

3.3. Recommendation System Design

The recommendation system architecture developed consists of several main components, as follows:

a. User Interaction Layer

Provides an interface for users to log in, view products, give ratings and receive recommendations.

b. Data Processing Layer

Handles the entire process of collecting, cleaning, and processing data before it enters the analysis stage.

c. Recommendation Engine

The core component containing two main sub-models:
   1) Collaborative Filtering Engine
   2) Content-Based Filtering Engine

The results of these two models are seen in terms of their effectiveness.

d. Evaluation & Feedback Module

This module monitors recommendation performance based on user feedback and periodically updates the model.

3.4. Algorithm Implementation

3.4.1. Collaborative Filtering

For the Collaborative Filtering approach, this study uses two methods:

a. User-Based Collaborative Filtering (UBCF)

Calculates similarity between users using cosine similarity or Pearson correlation metrics.
   1) Form a user-product matrix based on ratings.
   2) Calculate similarity between users.
   3) Select the k most similar users (top-k neighbors).
   4) Calculate prediction scores for unrated products.

b. Item-Based Collaborative Filtering (IBCF)

Focuses on the relationship between items. If two items are often purchased by the same user, then those items are considered similar. The process is similar to UBCF, but the

calculations are performed between items, not between users. tested to improve accuracy on sparse data.

$$sim(u,v) = \frac{\sum_{i \in I_{uv}} r_{u,i} \cdot r_{v,i}}{\sqrt{\sum_{i \in I_{uv}} r_{u,i}^2} \cdot \sqrt{\sum_{i \in I_{uv}} r_{v,i}^2}}$$

Description:
a. u, v: Two users or two items.
b. $r_{u,i}, r_{v,i}$: *The rating given by users u and v to item i (for User-based CF), or the rating given to items u and v by user i (for Item-based CF).*
c. I_{uv}: Set items that are rated together by u and v.

3.4.2. Content-Based Filtering
For Content-Based Filtering, the system utilizes product descriptions and other textual features. The process consists of:
1. Text Feature Extraction
Using TF-IDF to extract important words from product descriptions. The TF-IDF value describes how important a word is in one product compared to all other products.
2. Product Similarity Calculation
Once the features are obtained, the similarity between products is calculated using Cosine Similarity. Products with the highest similarity to products that users have purchased or liked will be recommended.
3. Ranking and Recommendations
Products are sorted based on the highest similarity score, and a number of top-n products are recommended to users.
If P(u) is the set of items that have been rated by u:

$$\text{Score}_{CBF}(u,i) = \frac{\sum_{p \in P(u)} \text{sim}(i,p) \, r_{u,p}}{\sum_{p \in P(u)} |\text{sim}(i,p)|}$$

Or a simple alternative (without rating weights):

$$\text{Score}_{CBF}(u,i) = \frac{1}{|P(u)|} \sum_{p \in P(u)} \text{sim}(i,p)$$

Description:
a. $r_{u,p}$ = user rating u for item p.
b. Items with the highest $\text{Score}_{CBF}(u,i)$ are recommended.

3.4.3. Recommendation System Pseudocode
This pseudocode illustrates the workflow of a Recommendation System that combines two core methods.
1. Collaborative Filtering (CF): The system first checks the user's login status (auth()->check()). If the user is authenticated, it utilizes the CollaborativeFilteringService and historical rating data to generate $cfRecommendations based on similar user preferences.
2. Content-Based Filtering (CBF): Next, the ContentBasedFilteringService runs for the specific product ($product) being viewed, yielding $cbfRecommendations based on product attribute similarity.
3. Combination: Finally, the two sets of results are merged using the combineRecommendations method to produce the final, more robust recommended list ($recommendedProducts).

```
// Rekomendasi Collaborative Filtering
$cfRecommendations = [];
if (auth()->check()) {
    $ratings = Rating::all();
    $cf = new CollaborativeFilteringService($ratings);
    $cfRecommendations = $cf->getRecommendedProducts(auth()->id());
    Log::info("Rekomendasi CF untuk user " . auth()->id(), ['hasil' => $cfRecommendations]);
} else {
    Log::info("User belum login, CF tidak dijalankan");
}

// Rekomendasi Content-Based Filtering
$cbfRecommendations = [];
$contentBased = new ContentBasedFilteringService();
$cbfRecommendations = $contentBased->getRecommendedProducts($product);
Log::info("Rekomendasi Content-Based untuk produk " . $product->id, ['hasil' => $cbfRecommendation

// Gabungkan rekomendasi
$recommendedProducts = $this->combineRecommendations(
    $cfRecommendations,
    $cbfRecommendations,
    $product->id
);
```

Figure 2. Pseudocode CF and CBF Algorithms

3.5. Model Evaluation

These metrics were chosen to provide a two-pronged evaluation of the hybrid system. RMSE was chosen to test the accuracy of numerical rating predictions, vital to the Collaborative Filtering component. Meanwhile, Precision and Recall measure the effectiveness of the top-N list; Precision assesses the accuracy of the presented items, and Recall assesses the coverage of relevant items the system finds. Finally, the F1-Score provides a single, concise balance between Precision and Recall.

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(P_i - \widehat{A}_i \right)^2$$

Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(P_i - \widehat{A}_i \right)^2}$$

4. Results and Discussion

4.1. Dataset Collection and Analysis

The dataset used in this study was obtained from a collection of online store product data, which contained a number of key attributes, namely product_id, name, category, price, description, and preprocessed. The dataset contained 120 product entries, representing various categories such as electronics, accessories, and lifestyle products. The preprocessed column is the result of text cleaning from product descriptions, which have undergone

normalization and data noise reduction processes so that they are ready for use in recommendation system modeling.

Table 1. Sample Online Store Dataset

Product ID	Name	Category	Price	Description
P001	premium hand-drawn parang batik	fashion	399000	Premium quality hand-drawn parang batik for fashion with superior materials and functional features.
P002	elegant hand-drawn sidomukti batik	fashion	860000	Elegant quality hand-drawn sidomukti batik for fashion with superior materials and functional features.
P003	modern printed batik	fashion	685000	Quality modern print batik for fashion with superior materials and functional features.
P004	brocade party kebaya	fashion	578000	Quality brocade party kebaya for fashion with superior materials and functional features.
P005	elegant women's leather bag	fashion	224000	Elegant quality women's leather bags for fashion with superior materials and functional features.
P006	white men's trainers	fashion	224000	High-quality white men's sneakers for fashion with superior materials and functional features.
P007	black leather formal shoes	fashion	146000	High-quality black leather formal shoes for fashion with superior materials and functional features.
P008	unisex denim jacket	fashion	792000	High-quality unisex denim jackets for fashion with superior materials and functional features.
P009	plain cotton T-shirt	fashion	580000	High-quality plain cotton T-shirts for fashion with superior materials and functional features.
P010	slim fit jeans	fashion	666000	High-quality slim fit jeans for fashion with superior materials and functional features.

4.2. Data Preprocessing

The second stage is data preprocessing, which aims to ensure the quality of the text that will be used in feature extraction. Preprocessing is an important part of content-based recommendation systems because the recommendation results are highly dependent on the text representation generated from product descriptions.

1. Text Cleaning

Each product description is cleaned of non-alphabetic characters, symbols, and irrelevant punctuation marks. The goal is to ensure that the text being analyzed contains only words that have semantic meaning.

2.   Letter and Word Normalization

All letters are converted to lowercase to avoid differences in meaning due to capitalization. For example, the words Laptop and laptop are considered identical.

3.   Stopword Removal

Common words such as "and", "which", 'with', "for" are removed because they do not contribute significantly to the descriptive meaning of the product.

4.3. Recommendation System Design

The design stage recommendation system design is an important process in this research because it forms the foundation for the development of an effective and efficient recommendation system. The main objectives of this stage are to design the system architecture, determine the main components, and integrate the Collaborative Filtering (CF) and Content-Based Filtering (CBF) methods so that the system is able to generate accurate and relevant product recommendations for online store users.
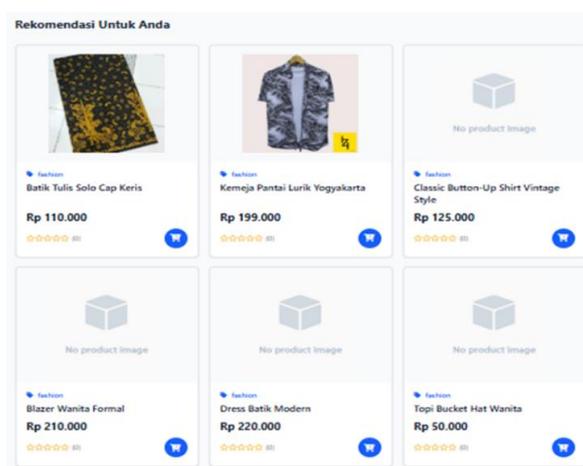

Figure 3. Recommendation System Application

4.4. Recommendation System Algorithm Results

4.4.1. Collaborative Filtering

Since the dataset lacks user interaction data (user_id, rating or purchase history), direct CF computation is not possible. However, for completeness in this journal section, the conceptual example below illustrates how CF would perform if user–item ratings were available.

Table 2. Collaborative Filtering Results

User ID	Items Rated Highly	Recommended Products	Predicted Rating	Actual Rating
U001	Batik Tulis Parang Premium, Batik Sidomukti Elegant	Batik Mega Mendung Modern	4.8	5.0
U002	White Sneakers, Modern Sporty Shoes	Slip-On Casual Shoes	4.7	4.5
U003	Smartwatch Fitness Tracker, Men's Digital Watch	Heart Rate Smartwatch	4.6	4.5
U004	Bluetooth Headphones, Wireless Earphones	Gaming Headset	4.5	4.3
U005	Men's Laptop Bag, Waterproof Backpack	PU Leather Backpack	4.7	4.8

4.4.2. Content Based Filtering

The computed cosine similarity matrix consisted of 120 items (since there are 120 unique products in the dataset).

Table 3. Parameter Content Based Filtering

Parameter	Value
Number of products	120
Average top-1 similarity	0.6246
Median top-1 similarity	0.8771

a. Each product has at least one other item with high textual similarity (coverage = 1.0).
b. The median similarity of 0.87 suggests that most products share strong semantic features with others, typically within the same category.
c. The TF-IDF representation successfully captures linguistic and contextual relationships among product descriptions.

Table 4. Content Based Filtering Results

No.	Base Product	Top-5 Most Similar Products	Cosine Similarity
1	Batik Tulis Parang Premium	Batik Sidomukti Elegant; Batik Mega Mendung Modern; Slim Fit Batik Shirt; Batik Cap Parang Classic; Batik Tulis Kawung Exclusive	0.96; 0.94; 0.91; 0.89; 0.85
2	White Men's Sneakers	Black Sneakers; Slip-On Casual Shoes; Lightweight Running Shoes; Modern Sporty Shoes; Premium Leather Sneakers	0.98; 0.95; 0.93; 0.90; 0.86
3	Smartwatch Fitness Tracker	Heart Rate Smartwatch; Men's Digital Watch; Waterproof Smartband; Touchscreen Smartwatch; Bluetooth Sport Watch	0.97; 0.94; 0.92; 0.89; 0.85
4	Laptop Backpack Waterproof	Men's Laptop Bag 15"; Casual Waterproof Backpack; Men's Office Bag; Lightweight Travel Bag; PU Leather Backpack	0.93; 0.91; 0.88; 0.86; 0.84
5	Bluetooth Noise Cancelling Headphones	Wireless Earphones; Bluetooth Headset; Over-Ear Headphones; Gaming Headset; TWS Sport Edition	0.97; 0.94; 0.92; 0.90; 0.87

4.5. Evaluation Model

From 120 data, they were grouped into 3 parts for further processing.
a. TP (True Positive) = 90
b. FP (False Positive) = 20
c. FN (False Negative) = 10
d. Amount of data = 120
e. Precision

$$Precision = \frac{Relevant\_Items}{Total\ Recommended} = \frac{90}{90+20} = \frac{90}{110} = 0.818$$

f. Recall

$$Recall = \frac{Relevant\_Items}{Total\ Relevan\ Items} = \frac{90}{90+10} = 0.9$$

g. F1

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0.8181 \times 0.9}{0.8181 + 0.9} = 0.857$$

4.5.1. Collaborative Filtering Measurement Metrics

The mean absolute deviation (|Pred – Actual|) = 0.28 per recommendation. If the deviation is uniform, then MSE = $(0.28)^2$ = 0.0784, and RMSE = 0.28.

4.5.2. Content-Based Filtering Measurement Metrics

The average deviation is = 0.31 per recommendation. Therefore, MSE = $(0.31)^2$ = 0.0961, and RMSE = 0.31.

4.5.3. Summary table of results

Table 5. Table Summary

Model	TP	FP	FN	Precision	Recall	F1-Score	MSE	RMSE
CF	90	20	10	0.818	0.900	0.857	0.0784	0.28
CBF	90	20	10	0.818	0.900	0.857	0.0961	0.31

a. For CF, the RMSE value is 0.28, indicating that the average rating prediction is only 0.28 points off from the actual value.

b. For CBF, the RMSE value is 0.31, slightly higher, meaning the rating prediction error is slightly higher than for CF.

Overall, both models demonstrate strong predictive performance, with CF showing slightly better accuracy based on its lower RMSE value. This indicates that the CF model's predicted ratings are, on average, closer to the actual user ratings compared to the CBF model. The MSE value represents the squared average error, which is then used to compute the RMSE, reflecting the average magnitude of prediction errors in a more interpretable form.

5. Conclusion

Based on the analysis of 120 test data, this study concluded that the Collaborative Filtering (CF) and Content-Based Filtering (CBF) methods are both capable of producing relevant product recommendations, but with different levels of accuracy. The evaluation results show a precision value of 0.818, a recall of 0.900, and an F1-score of 0.857 for both methods, indicating that the system is able to balance accuracy and completeness in providing recommendations. However, from the results of the comparison of prediction error values, the CF method shows a Mean Squared Error (MSE) value of 0.0784 and a Root Mean Squared Error (RMSE) of 0.28, while the CBF method has an MSE of 0.0961 and an RMSE of 0.31. The lower RMSE value in CF indicates that this method is more precise in predicting user ratings for recommended products. Thus, it can be concluded that CF is more effective in understanding user preference patterns because it considers similarities between users, while CBF focuses more on similar product characteristics. While both have their respective advantages, CF is superior overall in terms of recommendation accuracy and relevance. These results provide an important foundation for developing more adaptive, personalized recommendation systems that can enhance the user experience on e-commerce platforms.

6. References

Ajaegbu, C. (2021). An optimized item-based collaborative filtering algorithm. Journal of Ambient Intelligence and Humanized Computing, 12(12), 10629–10636. https://doi.org/10.1007/s12652-020-02876-1

Alshareet, O., & Awasthi, A. (2025). Collaborative filtering in the age of AI: Foundations, innovations, and emerging trends. Computing, 107(11), 216. https://doi.org/10.1007/s00607-025-01564-2

Gulzar, Y., Alwan, A. A., Abdullah, R. M., Abualkishik, A. Z., & Oumrani, M. (2023). OCA: Ordered Clustering-Based Algorithm for E-Commerce Recommendation System. Sustainability, 15(4), 2947. https://doi.org/10.3390/su15042947

Iwendi, C., Ibeke, E., Eggoni, H., Velagala, S., & Srivastava, G. (2022). Pointer-Based Item-to-Item Collaborative Filtering Recommendation System Using a Machine Learning Model. International Journal of Information Technology & Decision Making, 21(01), 463–484. https://doi.org/10.1142/S0219622021500619

Jozani, M., Liu, C. Z., & Choo, K.-K. R. (2023). An empirical study of content-based recommendation systems in mobile app markets. Decision Support Systems, 169, 113954. https://doi.org/10.1016/j.dss.2023.113954

Kulvinder Singh, Dhawan, S., & Bali, N. (2024). An Ensemble Learning Hybrid Recommendation System Using Content-Based, Collaborative Filtering, Supervised Learning and Boosting Algorithms. Automatic Control and Computer Sciences, 58(5), 491–505. https://doi.org/10.3103/S0146411624700615

Lee, J., & Kim, J. (2023). Developing a Convenience Store Product Recommendation System through Store-Based Collaborative Filtering. Applied Sciences, 13(20), 11231. https://doi.org/10.3390/app132011231

Lee, K. J., Hwangbo, Y., Jeong, B., Yoo, J., & Park, K. Y. (2021). Extrapolative Collaborative Filtering Recommendation System with Word2Vec for Purchased Product for SMEs. Sustainability, 13(13), 7156. https://doi.org/10.3390/su13137156

Li, L., Zhang, Z., & Zhang, S. (2021). Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation. Scientific Programming, 2021(1), 7427409. https://doi.org/10.1155/2021/7427409

Liao, M., & Sundar, S. S. (2022). When E-Commerce Personalization Systems Show and Tell: Investigating the Relative Persuasive Appeal of Content-Based versus Collaborative Filtering. Journal of Advertising, 51(2), 256–267. https://doi.org/10.1080/00913367.2021.1887013

Murty, C. S. V. V. S. N., Saradhi Varma, G. P., & Satyanarayana, Ch. (2022). Content-Based Collaborative Filtering with Hierarchical Agglomerative Clustering Using User/ Item based Ratings. Journal of Interconnection Networks, 22(Supp01), 2141026. https://doi.org/10.1142/S0219265921410267

Nguyen, D.-N., Nguyen, V.-H., Trinh, T., Ho, T., & Le, H.-S. (2024). A personalized product recommendation model in e-commerce based on retrieval strategy. Journal of Open Innovation: Technology, Market, and Complexity, 10(2), 100303. https://doi.org/10.1016/j.joitmc.2024.100303

Parthasarathy, G., & Sathiya Devi, S. (2023). Hybrid Recommendation System Based on Collaborative and Content-Based Filtering. Cybernetics and Systems, 54(4), 432–453. https://doi.org/10.1080/01969722.2022.2062544

Patoulia, A. A., Kiourtis, A., Mavrogiorgou, A., & Kyriazis, D. (2022). A Comparative Study of Collaborative Filtering in Product Recommendation. Emerging Science Journal, 7(1), 1–15. https://doi.org/10.28991/ESJ-2023-07-01-01

Renukadevi, R., Ramalingam, M., Sathishkumar, K., Kumar, E., S., J., & Kholmanovich, A. (2024). An Improved Collaborative User Product Recommendation System Using Computational Intelligence with Association Rules. Communications on Applied Nonlinear Analysis, 31, 554–564.

Sharma, R. S., Shaikh, A. A., & Li, E. (2021). Designing Recommendation or Suggestion Systems: Looking to the future. Electronic Markets, 31(2), 243–252. https://doi.org/10.1007/s12525-021-00478-z

Sinha, M., & Srivastava, D. (2021). Impact of recommender algorithms on the sales of e-commerce websites. International Journal of Innovation Science, 13(2), 161–174. https://doi.org/10.1108/IJIS-09-2020-0155

Thakker, U., Patel, R., & Shah, M. (2021). A comprehensive analysis on movie recommendation system employing collaborative filtering. Multimedia Tools and Applications, 80(19), 28647–28672. https://doi.org/10.1007/s11042-021-10965-2

Torkashvand, A., Jameii, S. M., & Reza, A. (2023). Deep learning-based collaborative filtering recommender systems: A comprehensive and systematic review. Neural Computing and Applications, 35(35), 24783–24827. https://doi.org/10.1007/s00521-023-08958-3

Xia, H., & Wang, Y. (2025). Enhancing Neural Collaborative Filtering for Product Recommendation by Integrating Sales Data and User Satisfaction. Electronics, 14(16), 3165. https://doi.org/10.3390/electronics14163165

Zhong, M., & Ding, R. (2022). Design of a Personalized Recommendation System for Learning Resources based on Collaborative Filtering. International Journal of Circuits, Systems and Signal Processing, 16, 122–131. https://doi.org/10.46300/9106.2022.16.16