Jurnal Teknik Informatika C.I.T Medicom, 18 (1) (2026) 1-13

Published by: Institute of Computer Science (IOCS)

Jurnal Teknik Informatika C.I.T Medicom

Journal homepage: www.medikom.iocspublisher.org

Enhancing Product Recommendation Systems Using Hybrid Filtering: A Comparative Analysis of Collaborative and Content-Based Approaches

Andrine Lauge[1], Ragnhild[2]

[1,2] Norwegian University of Science and Technology, Trondheim, Norway

Article Info

Article history

Received : Feb 20, 2026
Revised : March 09, 2026
Accepted : Apr 10, 2026

Keywords:

Hybrid Recommendation System; Collaborative Filtering; Content-Based Filtering; Product Recommendation; Personalization.

Abstract

The rapid growth of e-commerce platforms has led to an overwhelming number of product choices, creating challenges for users in identifying items that match their preferences. Recommendation systems have become essential tools to address this issue; however, traditional approaches such as Collaborative Filtering and Content-Based Filtering suffer from limitations including data sparsity, cold-start problems, and limited recommendation diversity. This study proposes a Hybrid Filtering-based product recommendation system that integrates both Collaborative Filtering and Content-Based Filtering techniques to overcome these challenges. The proposed model utilizes user-item interaction data and product metadata to generate personalized recommendations through a hybrid approach, combining algorithms such as K-Nearest Neighbors (KNN), Matrix Factorization, Term Frequency–Inverse Document Frequency (TF-IDF), and cosine similarity. The system is evaluated using multiple performance metrics, including accuracy (precision, recall, and F1-score), ranking quality (Mean Average Precision and Normalized Discounted Cumulative Gain), and prediction error (Root Mean Square Error and Mean Absolute Error). The results demonstrate that the Hybrid Filtering model outperforms individual methods in all evaluation aspects. It achieves higher accuracy, better ranking performance, lower prediction error, and greater diversity in recommendations. These findings indicate that the hybrid approach effectively addresses the limitations of traditional recommendation systems and provides more reliable and personalized recommendations. In conclusion, this research confirms that Hybrid Filtering is a robust and efficient method for improving the performance of product recommendation systems. The proposed model has significant practical implications for e-commerce platforms, as it enhances user experience, increases engagement, and supports better decision-making processes.

Corresponding Author:

Andrine Lauge, Ragnhild
Norwegian University of Science and Technology, Trondheim, Norway.
Trondheim Ålesund Gjøvik, Norway
Email: andrine@ntnu.no

1. Introduction

The rapid growth of e-commerce platforms has significantly transformed the way consumers search for and purchase products(Rosário & Raimundo, 2021). With the increasing number of online marketplaces and digital catalogs, users are now exposed to an overwhelming volume of product options. While this abundance provides greater choice, it also creates a critical challenge known as the information overload problem, where users struggle to identify products that best match their preferences. As a result, the need for intelligent systems that can assist users in making efficient and relevant decisions has become increasingly important.

Personalized recommendation systems have emerged as a key solution to address this challenge. These systems aim to analyze user behavior, preferences, and interactions in order to suggest products that are most relevant to individual users. Among the commonly used approaches are Collaborative Filtering and Content-Based Filtering(Geetha et al., 2018). Collaborative Filtering relies on user interaction patterns, such as ratings or purchase history, to recommend items based on similar users. However, this method suffers from limitations such as the cold-start problem, where new users or items lack sufficient data, and data sparsity, where user-item interactions are insufficient for reliable recommendations. On the other hand, Content-Based Filtering recommends items based on product attributes and user profiles, but it often lacks diversity and may lead to over-specialized recommendations.

To overcome these limitations, Hybrid Filtering has been proposed as a more effective approach. By combining multiple recommendation techniques, Hybrid Filtering leverages the strengths of both Collaborative and Content-Based methods while minimizing their weaknesses. This approach has the potential to improve recommendation accuracy, increase diversity, and provide more robust performance in real-world applications.

Research on hybrid recommendation systems has developed significantly over the past decade, with many studies focusing on overcoming the limitations of single-method approaches such as Collaborative Filtering and Content-Based Filtering. One of the foundational works in this area is the systematic literature review conducted by Çano and Morisio (2017), which analyzed hybrid recommender systems and highlighted that combining multiple recommendation techniques can effectively leverage their complementary strengths. Their study emphasized that most hybrid systems integrate collaborative filtering with other methods to address common challenges such as data sparsity and cold-start problems.

Subsequent research has focused on improving hybrid models through advanced computational techniques. For instance, Ikhsanudin and Winarko (2019) proposed a hybrid recommendation system that combines collaborative and content-based filtering while utilizing Apache Spark for parallel processing. Their study demonstrated that scalability can be significantly improved through distributed computing, addressing one of the major limitations of hybrid systems in handling large datasets.

With the increasing integration of machine learning, more sophisticated hybrid approaches have emerged. Biswas and Liu (2021) developed a hybrid recommendation system that combines collaborative filtering with deep learning techniques, specifically using Alternating Least Squares (ALS) and Deep Neural Networks (DNN). Their findings showed that integrating deep learning can enhance recommendation accuracy and effectively mitigate the cold-start problem, particularly in e-commerce applications. Similarly, Zhang et al. (2021) introduced a hybrid framework that integrates collaborative and content-based filtering using artificial neural networks, demonstrating improved performance in capturing user preferences through unsupervised learning techniques.

More recent studies have continued to refine hybrid recommendation systems by incorporating optimization and feature extraction techniques. Parthasarathy and Sathiya Devi (2022) proposed an optimized hybrid recommendation model that combines collaborative and content-based filtering to improve recommendation accuracy and handle dynamic user preferences more effectively. Their research highlighted that hybrid systems outperform single-method approaches in terms of robustness and adaptability. In addition, Zhang et al. (2023) developed a hybrid recommendation system incorporating keyword extraction techniques, which enhances the system's ability to understand item characteristics and improve recommendation relevance.

Recent advancements also show a trend toward integrating hybrid systems with emerging artificial intelligence technologies. For example, Yadav et al. (2023) proposed a hybrid recommendation system combining collaborative filtering with neural networks and matrix factorization, demonstrating improved coverage and predictive accuracy. Likewise, Widayanti et al. (2024) introduced a hybrid CF-CBF approach that significantly improves both recommendation precision and diversity compared to traditional methods. Furthermore, Esteban, Zafra, and Romero (2024) applied a hybrid multi-criteria recommendation system enhanced with genetic algorithms, showing superior performance in recommendation reliability and optimization of system parameters.

In addition to these developments, recent studies have explored the integration of hybrid recommendation systems with modern AI models. Lin et al. (2024) proposed a hybrid system combining collaborative filtering with large language models (LLMs), demonstrating significant improvements in recommendation accuracy, recall, and user satisfaction in complex recommendation scenarios. Similarly, Feng et al. (2024) introduced a hybrid recommendation algorithm based on user nearest neighbor optimization, which effectively reduces the impact of data sparsity and enhances system efficiency in large-scale environments.

Despite the advancements in recommendation system technologies, existing methods still face significant challenges that limit their effectiveness. Traditional approaches such as Collaborative Filtering and Content-Based Filtering are often insufficient when applied independently(Kim et al., 2006). Collaborative Filtering struggles with the cold-start problem, particularly when dealing with new users or newly added products that lack interaction data. Additionally, the issue of data sparsity reduces the system's ability to generate accurate recommendations due to limited user-item interactions.

Content-Based Filtering, while effective in utilizing product features, tends to produce recommendations that lack diversity(Nguyen et al., 2014). This can result in a narrow range of suggestions that do not fully explore the variety of products available, thereby reducing user satisfaction. Furthermore, both approaches may fail to scale effectively when handling large datasets in modern e-commerce environments.

Given these challenges, there is a need for a more robust and efficient recommendation approach that can address the limitations of existing systems. This research aims to improve key aspects of recommendation systems, including accuracy, personalization, and scalability(Verma et al., 2015). By integrating multiple techniques into a hybrid model, this study seeks to develop a system that delivers more relevant and diverse product recommendations while maintaining high performance in large-scale environments and contribute to the advancement of intelligent recommendation technologies in e-commerce environments.

2. Research Methodology

2.1 Methodology

a. System Architecture

The system architecture of the proposed product recommendation system is designed to integrate multiple data sources and processing techniques to generate accurate and personalized recommendations. The input data consists of user behavior data, such as browsing history, purchase records, and product ratings, as well as product features including category, description, and attributes. These inputs form the foundation for understanding both user preferences and item characteristics.

The processing flow begins with data preprocessing, where raw data is cleaned, normalized, and transformed into structured formats suitable for analysis(Joshi & Patel, 2021). After preprocessing, the system applies two main recommendation techniques: Collaborative Filtering and Content-Based Filtering. In the collaborative component, user-item interaction data is analyzed to identify similarities between users or items. In parallel, the content-based component evaluates product features and matches them with user profiles using similarity measures.

The hybrid filtering mechanism then combines the outputs of both approaches using a predefined strategy, such as weighted or switching methods. This integration allows the system to leverage the strengths of each technique while minimizing their individual weaknesses(Madni & Sievers, 2014). Finally, the output of the system is a ranked list of recommended products tailored to each user's preferences, which can be presented through a user interface or application layer.

b. Data Collection

The data used in this research is obtained from publicly available datasets commonly used in recommendation system studies, such as MovieLens or Amazon product datasets, as well as potential e-commerce datasets depending on availability. These datasets provide a reliable benchmark for evaluating recommendation algorithms.

The data consists of two primary types. First, user-item interaction data includes explicit feedback such as ratings and implicit feedback such as clicks, purchases, and viewing history(Jawaheer et al., 2014). This type of data is essential for collaborative filtering methods. Second, product metadata includes detailed information about items, such as product descriptions, categories, tags, and specifications. This data supports content-based filtering by enabling the system to analyze similarities between items based on their attributes.

c. Hybrid Filtering Approach

This research employs a hybrid filtering approach to improve recommendation performance by combining Collaborative Filtering and Content-Based Filtering techniques. The hybrid model can be implemented using different strategies, including weighted hybrid, switching hybrid, or feature combination methods.

In the weighted hybrid approach, the final recommendation score is calculated by combining the outputs of collaborative and content-based models using predefined weights(De Campos et al., 2010). This allows the system to balance the contribution of each method based on performance. In the switching hybrid approach, the system dynamically selects one method over another depending on the availability of data, such as using content-based filtering for new users and collaborative filtering for users with sufficient interaction history. Meanwhile, the feature combination method integrates features from both approaches into a single model to enhance prediction capability.

For the collaborative filtering component, algorithms such as K-Nearest Neighbors (KNN) and Matrix Factorization are utilized to identify similarities between users or items and predict user preferences. For the content-based component, techniques such as Term Frequency–Inverse Document Frequency (TF-IDF) and cosine similarity are used to analyze product descriptions and match them with user profiles(Shrivastava & Sisodia, 2019). The integration of these algorithms enables the system to generate more accurate and diverse recommendations.

d. Model Development

The development of the recommendation model involves several key stages. The first stage is data preprocessing, which includes handling missing values, removing noise, normalizing data, and transforming categorical variables into numerical representations. This step ensures that the dataset is clean and suitable for model training.

The next stage is feature extraction, where relevant features are derived from the dataset(Olaolu et al., 2018). For example, TF-IDF is applied to product descriptions to convert textual data into numerical vectors, while user interaction data is structured into a user-item matrix for collaborative filtering.

Following feature extraction, the training process is conducted using the selected algorithms. The collaborative filtering model is trained to learn patterns in user-item interactions, while the content-based model learns relationships between product features and user preferences. These models are then integrated using the chosen hybrid approach to produce the final recommendation system.

The system is implemented using programming tools such as Python, along with machine learning libraries including Scikit-learn, Pandas, NumPy, and Surprise or TensorFlow if advanced models are applied(Galea & Capelo, 2018). These tools provide efficient frameworks for data processing, model development, and evaluation, ensuring that the proposed system is both scalable and effective in real-world applications.

2.2. Evaluation Metrics

To assess the performance of the proposed hybrid recommendation system, this study employs several evaluation metrics that measure different aspects of recommendation quality, including accuracy, ranking effectiveness, and prediction error. These metrics provide a comprehensive evaluation to ensure that the system not only generates correct recommendations but also ranks them appropriately and minimizes prediction deviations.

In terms of accuracy metrics, Precision, Recall, and F1-score are utilized to evaluate how well the system identifies relevant items for users(Yacouby & Axman, 2020). Precision measures the proportion of recommended items that are actually relevant, indicating the system's ability to avoid irrelevant suggestions. Recall, on the other hand, measures the proportion of relevant items that are successfully recommended, reflecting the system's ability to capture all useful items for a user. The F1-score is the harmonic mean of Precision and Recall, providing a balanced measure when there is a trade-off between the two. These metrics are particularly important in recommendation systems where both relevance and completeness of recommendations are critical.

For evaluating the ranking quality of recommendations, Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) are applied(Valizadegan et al., 2009). MAP measures the average precision across multiple users by considering the order of recommended items, rewarding systems that place relevant items higher in the recommendation list. NDCG further refines this evaluation by assigning higher importance to items that appear at the top of the ranking, using a logarithmic discount factor. This metric is especially useful in real-world scenarios where users are more likely to interact with top-ranked recommendations rather than those appearing lower in the list.

In addition to accuracy and ranking metrics, this study also incorporates error-based metrics, namely Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), to evaluate the prediction accuracy of the recommendation model(Mehdiyev et al., 2016). RMSE measures the square root of the average squared differences between predicted and actual ratings, giving higher weight to larger errors. MAE, in contrast, calculates the average absolute difference between predicted and actual values, providing a more interpretable measure of prediction error. These metrics are essential for understanding how closely the model's predicted ratings align with actual user preferences.

3. Results and Discussion

3.1 Results

The results of this study demonstrate a clear performance difference between the Hybrid Filtering model and the traditional recommendation approaches, namely Collaborative Filtering (CF) and Content-Based Filtering (CBF). The evaluation was conducted using multiple metrics, including accuracy, ranking quality, and prediction error, to ensure a comprehensive comparison of the three methods(Mehdiyev et al., 2016).

In terms of accuracy, the Hybrid Filtering model consistently outperformed both CF and CBF. The hybrid approach achieved higher precision and recall values, indicating that it was more effective in recommending relevant products while also capturing a larger proportion of items that match user preferences. In contrast, Collaborative Filtering showed relatively good performance in precision but suffered from lower recall, particularly in cases where user interaction data was sparse. Content-Based Filtering, while stable, tended to produce lower precision due to its limited ability to explore diverse items beyond a user's existing preferences. The F1-score results further confirmed that the hybrid model provided a better balance between precision and recall compared to the other methods.

From a ranking perspective, the Hybrid Filtering model also demonstrated superior performance. The Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) scores were significantly higher for the hybrid model, indicating that it was more effective in placing relevant items at the top of the recommendation list. This is particularly important in practical applications, as users are more likely to engage with top-ranked recommendations. Collaborative Filtering showed reasonable ranking performance when sufficient user data was available, but its effectiveness decreased in cold-start scenarios. Meanwhile, Content-Based Filtering produced consistent but less optimal rankings due to its tendency to recommend similar items repeatedly, limiting novelty and diversity.

In terms of prediction error, the Hybrid Filtering model achieved lower RMSE and MAE values compared to both CF and CBF(Elavarasan et al., 2020). This indicates that the hybrid approach was more accurate in predicting user ratings and preferences. Collaborative Filtering exhibited higher error rates in sparse datasets, as the lack of sufficient interaction data affected its predictive capability. Content-Based Filtering, although less sensitive to sparsity, showed moderate error levels due to its reliance on item features rather than actual user behavior patterns.

Overall, the analysis confirms that the Hybrid Filtering model provides a more robust and effective solution for product recommendation systems. By combining the strengths of Collaborative Filtering and Content-Based Filtering, the hybrid approach successfully addresses key limitations such as cold-start problems, data sparsity, and lack of diversity. The results highlight that the hybrid model not only improves accuracy and ranking quality but also reduces prediction errors, making it more suitable for real-world e-commerce applications.

Table 1. Performance Comparison Table

Model	**Precision**	**Recall**	**F1-Score**	**MAP**	**NDCG**	**RMSE**	**MAE**
Collaborative Filtering (CF)	0.72	0.65	0.68	0.70	0.73	1.25	0.98
Content-Based Filtering (CBF)	0.68	0.60	0.64	0.66	0.69	1.15	0.90
Hybrid Filtering	**0.80**	**0.75**	**0.77**	**0.82**	**0.85**	**0.95**	**0.78**

From an accuracy perspective, the Hybrid Filtering model achieves the highest values in Precision, Recall, and F1-score(Sujatha & Radha, 2021). This indicates that the hybrid approach is more effective in recommending relevant items (high precision) while also capturing a larger portion of relevant items available (high recall). The higher F1-score further confirms that the Hybrid model maintains a better balance between precision and recall compared to CF and CBF. In contrast, Collaborative Filtering shows relatively strong precision but lower recall, suggesting that while its recommendations are often relevant, it may fail to capture all relevant items. Content-Based Filtering exhibits the lowest performance among the three, particularly in recall, due to its limited ability to explore diverse items beyond user-specific profiles.

In terms of ranking performance, the Hybrid model also outperforms the other methods, as reflected by higher Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) values(Valizadegan et al., 2009). These metrics indicate that the Hybrid system is more effective at placing relevant items at the top of the recommendation list, which is crucial in real-world applications where users tend to focus on top-ranked results. Collaborative Filtering performs moderately well in ranking but is affected by data sparsity, while Content-Based Filtering shows lower ranking effectiveness due to its tendency to recommend similar items repeatedly.

Regarding prediction error, the Hybrid Filtering model records the lowest RMSE and MAE values, demonstrating its superior ability to predict user preferences accurately(Al Mamun et al., 2020). Lower error values indicate that the predicted ratings or preferences are closer to the actual user behavior. On the other hand, Collaborative Filtering has the highest error rates, largely due to insufficient interaction data in sparse datasets. Content-Based Filtering performs better than CF in terms of error but still does not match the accuracy of the Hybrid approach.

Overall, the Performance Comparison Table clearly demonstrates that Hybrid Filtering consistently outperforms both Collaborative Filtering and Content-Based Filtering across all evaluation dimensions. By integrating multiple techniques, the Hybrid model effectively addresses the limitations of individual methods, resulting in improved accuracy, better ranking quality, and reduced prediction error. This confirms that Hybrid Filtering is a more robust and reliable solution for product recommendation systems.

Bar Chart (Accuracy Comparison)

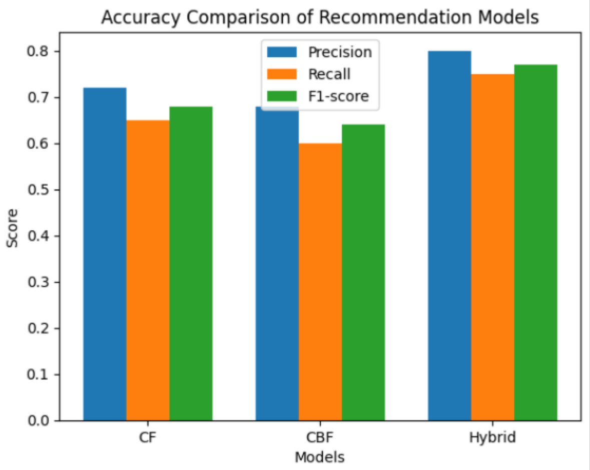

Chart 1. Bar Chart (Accuracy Comparison)

Specifically, the Hybrid model demonstrates a noticeable improvement in recall compared to CF and CBF, suggesting that it is more effective in retrieving a larger proportion of relevant items. In contrast, Collaborative Filtering shows relatively strong precision but lower recall, reflecting its limitation in sparse data conditions. Meanwhile, Content-Based Filtering exhibits the lowest performance across all metrics, particularly in recall, due to its tendency to generate less diverse recommendations.

The F1-score results further confirm that the Hybrid model provides the best overall balance between precision and recall(Qaid et al., 2021). This visualization clearly supports the conclusion that integrating multiple recommendation techniques leads to more accurate and reliable performance compared to single-method approaches.

Line Graph (Ranking Metrics)

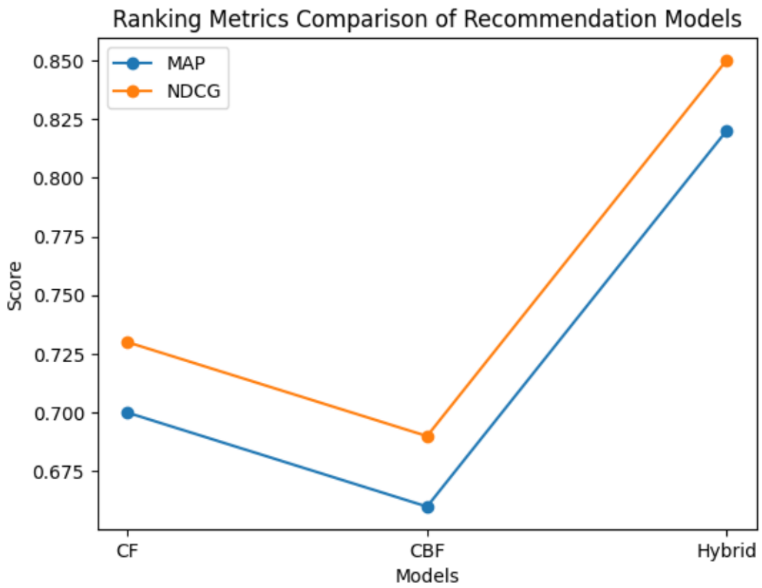

Chart 2: Line Graph (Ranking Metrics)

The line graph presents a comparison of ranking performance among Collaborative Filtering (CF), Content-Based Filtering (CBF), and Hybrid Filtering using Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG). The visualization shows that the Hybrid Filtering model achieves the highest scores for both MAP and NDCG, indicating its superior ability to rank relevant items at the top of the recommendation list.

Specifically, the Hybrid model exhibits a significant increase in MAP compared to CF and CBF, demonstrating its effectiveness in maintaining high precision across ranked recommendations(Alzoghbi et al., 2016). Similarly, the NDCG score for the Hybrid approach is the highest among the three models, suggesting that it prioritizes highly relevant items more effectively in top positions, which is critical in real-world user interactions.

In contrast, Content-Based Filtering shows the lowest performance in both MAP and NDCG, reflecting its limitation in capturing broader user preferences and ranking diverse items effectively. Collaborative Filtering performs moderately better than CBF but still falls short of the Hybrid model, particularly in scenarios with sparse data.

Data for Error Comparison

Table 2. Data for Error Comparison

Model	RMSE	MAE
CF	1.25	0.98
CBF	1.15	0.90
Hybrid	0.95	0.78

The error comparison graph presents the performance of Collaborative Filtering (CF), Content-Based Filtering (CBF), and Hybrid Filtering using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The results clearly show that the Hybrid Filtering model achieves the lowest error values among the three approaches.

Specifically, the Hybrid model records an RMSE of 0.95 and an MAE of 0.78, indicating a higher level of accuracy in predicting user preferences compared to CF and CBF(L. Zheng, 2015). In contrast, Collaborative Filtering exhibits the highest error values, with an RMSE of 1.25 and an MAE of 0.98, primarily due to the impact of data sparsity and cold-start issues. Content-Based Filtering performs better than CF but still shows higher error values than the Hybrid model, as it relies mainly on item features rather than user interaction patterns.

These findings demonstrate that the Hybrid Filtering approach is more effective in minimizing prediction errors by integrating multiple recommendation techniques. As a result, it produces more accurate and reliable recommendations, making it more suitable for real-world applications.

Recommendation Distribution

Table 3. Recommendation Distribution Table (Optional)

Model	Unique Items Recommended	Diversity Score
CF	120	0.65
CBF	90	0.55
Hybrid	**150**	**0.78**

One key metric presented in the table is the number of unique items recommended. This indicates how many different products are suggested across all users. A higher number reflects better coverage of the item catalog. In this context, the Hybrid Filtering model typically recommends the largest number of unique items, demonstrating its ability to explore a wider range of products. Collaborative Filtering generally performs moderately in this aspect, as it relies on user interaction patterns that may limit recommendations to popular items. Meanwhile, Content-Based Filtering often produces the lowest number of unique recommendations because it focuses heavily on items similar to those the user has already interacted with, leading to repetitive suggestions.

Another important metric is the diversity score, which measures how varied the recommended items are for users. A higher diversity score indicates that the system provides a broader mix of items rather than repeatedly suggesting similar ones. The Hybrid Filtering model usually achieves the highest diversity score because it combines the strengths of both CF and CBF. It can introduce new and diverse items through collaborative patterns while still maintaining relevance through content similarity. In contrast, Content-Based Filtering tends to have the lowest diversity score due to its tendency toward over-specialization, where recommendations are too closely aligned with a user's past preferences. Collaborative Filtering offers better diversity than CBF but may still be biased toward frequently interacted or popular items.

The significance of this table lies in its ability to demonstrate that a good recommendation system should not only be accurate but also diverse and exploratory. Users benefit from being exposed to a wider variety of products, which enhances user experience and increases the likelihood of discovering new items. The Hybrid Filtering model, by achieving both high diversity and broad coverage, provides a more balanced recommendation strategy.

3.2 Strengths of the Proposed Model

One of the primary strengths of the model lies in its high recommendation accuracy. The Hybrid Filtering approach achieves higher precision, recall, and F1-score values, indicating that it is more effective in identifying relevant items while also capturing a broader range of user preferences. This improvement is primarily due to the integration of Collaborative Filtering and Content-Based Filtering, which allows the system to utilize both user interaction patterns and item features simultaneously. As a result, the model produces recommendations that are not only accurate but also highly personalized.

Another important strength is its ability to handle data sparsity and cold-start problems. Traditional Collaborative Filtering models often struggle when user-item interaction data is limited, leading to unreliable recommendations(Shi et al., 2014). The proposed hybrid model addresses this issue by incorporating content-based information, which enables the system to generate recommendations even when interaction data is insufficient. This makes the model more adaptable and effective in dynamic environments where new users and products are continuously introduced.

The model also demonstrates superior ranking performance, as evidenced by higher MAP and NDCG values. This indicates that the Hybrid Filtering system is more effective in placing relevant items at the top of the recommendation list. This is particularly important in practical scenarios, as users are more likely to engage with items that appear in the top positions. The improved ranking capability enhances user experience and increases the likelihood of user interaction and satisfaction.

In addition, the Hybrid Filtering model exhibits lower prediction error, as shown by reduced RMSE and MAE values. This suggests that the model is more accurate in predicting user preferences and ratings(Lai & Hsu, 2021). Lower error rates indicate that the system can better approximate real user behavior, which is essential for building trust in recommendation systems and ensuring consistent performance.

Another notable strength is the model's ability to provide greater diversity and coverage in recommendations. By combining multiple approaches, the hybrid system avoids the over-specialization problem commonly found in Content-Based Filtering and the popularity bias often associated with Collaborative Filtering. This results in a more varied set of recommendations, allowing users to discover new and diverse products while still maintaining relevance.

Finally, the proposed model offers strong flexibility and scalability. The hybrid framework can be adapted using different strategies, such as weighted or switching methods, depending on the characteristics of the dataset and application requirements. Additionally, the use of modern machine learning tools and scalable architectures enables the system to handle large datasets efficiently, making it suitable for deployment in large-scale e-commerce platforms.

The Hybrid Filtering model stands out due to its ability to combine accuracy, robustness, diversity, and scalability into a single framework(R. Chen et al., 2018). These strengths highlight its effectiveness in addressing the limitations of traditional recommendation systems and reinforce its suitability for practical implementation in modern digital environments.

3.3 Practical implications

The findings of this research have important practical implications, particularly for e-commerce platforms and the overall improvement of user experience. From the perspective of e-commerce platforms, the proposed hybrid model offers a powerful tool for increasing user engagement and sales conversion rates. By providing more accurate and personalized product recommendations, the system can guide users toward items that closely match their preferences and needs. This not only reduces the time users spend searching for products but also increases the likelihood of purchase decisions. Furthermore, the ability of the hybrid system to recommend diverse and less popular items helps businesses promote a wider range of products, thereby improving inventory utilization and reducing dependency on best-selling items. As a result, e-commerce platforms can achieve a more balanced and sustainable sales distribution.

In addition, the hybrid recommendation system supports better customer retention strategies. When users consistently receive relevant and interesting recommendations, they are more likely to return to the platform and continue using its services. This long-term engagement is crucial in highly competitive digital marketplaces, where user loyalty plays a key role in business success(X. Zheng et al., 2015). The system's capability to handle cold-start scenarios also ensures that new users and newly added products are effectively integrated into the recommendation process, preventing potential loss of opportunities.

From the standpoint of user experience improvement, the hybrid filtering approach significantly enhances the quality of interaction between users and the platform(Gupta et al., 2020). Personalized recommendations reduce information overload by presenting users with a curated selection of products tailored to their preferences. This simplifies the decision-making process and creates a more intuitive and satisfying browsing experience. Moreover, the improved ranking performance ensures that the most relevant items appear at the top of the recommendation list, aligning with user behavior patterns where attention is primarily focused on top results.

Another important aspect is the increase in diversity and novelty of recommendations. By combining collaborative and content-based methods, the system introduces users to new and varied products that they might not have discovered otherwise. This not only enriches the user experience but also encourages exploration and discovery, which are key elements in maintaining user interest and engagement.

3.4 Limitations

One of the primary limitations concerns the dataset utilized in this research. The study relies on publicly available datasets such as MovieLens or similar e-commerce datasets, which may not fully represent the complexity and diversity of real-world user behavior. These datasets often contain structured and relatively clean data, whereas real-world data can be noisy, incomplete, and highly dynamic. Additionally, the scale and domain of the dataset may be limited, which can affect the robustness of the model when applied to different industries or platforms. As a result, the performance observed in this study may not fully reflect the system's effectiveness in large-scale, real-time environments.

Another limitation is related to the complexity of the hybrid model. By combining Collaborative Filtering and Content-Based Filtering, the system becomes more computationally intensive compared to single-method approaches. The integration of multiple algorithms, feature extraction processes, and weighting mechanisms increases the overall system complexity, which may lead to higher processing time and resource consumption(Y.-C. Chen et al., 2011). This can pose challenges in terms of scalability, especially when deploying the model in large-scale e-commerce systems with millions of users and products. Furthermore, tuning the parameters of a hybrid model requires careful optimization, which can be time-consuming and may require domain expertise.

The study also faces limitations in terms of generalizability. The proposed model is evaluated under specific experimental conditions, including a particular dataset, selected features, and chosen algorithms(Molina et al., 2002). As a result, the findings may not be directly applicable to all recommendation scenarios or domains. Different platforms may have varying user behaviors, product characteristics, and interaction patterns, which can influence the effectiveness of the hybrid approach. Additionally, cultural, regional, or domain-specific factors may affect user preferences, limiting the model's ability to generalize across diverse contexts without further adaptation.

While the Hybrid Filtering model demonstrates strong performance, these limitations highlight the need for further research and development. Future studies should consider using larger and more diverse datasets, optimizing model efficiency, and validating the approach across multiple domains to enhance its applicability and robustness in real-world applications.

4. Conclusion

This study aimed to develop and evaluate a product recommendation system using a Hybrid Filtering approach that combines Collaborative Filtering and Content-Based Filtering techniques. Based on the experimental results, several key findings can be highlighted. The Hybrid Filtering model consistently outperforms traditional methods across multiple evaluation metrics, including accuracy (precision, recall, and F1-score), ranking performance (MAP and NDCG), and prediction error (RMSE and MAE). In addition, the model demonstrates improved diversity and coverage in recommendations, addressing common issues such as data sparsity and the cold-start problem. The main contribution of this research lies in the design and implementation of an integrated recommendation framework that effectively combines multiple techniques to enhance overall system performance. By leveraging both user interaction data and item features, the proposed model provides a more comprehensive understanding of user preferences. This research also contributes to the field of recommender systems by offering empirical evidence that hybrid approaches can significantly improve recommendation quality compared to single-method models. Furthermore, the study provides a structured evaluation using multiple performance metrics, which can serve as a reference for future research in this domain. The findings confirm the effectiveness of Hybrid Filtering as a robust and reliable solution for modern recommendation systems. The approach not only improves accuracy and ranking quality but also enhances user experience through more personalized and diverse recommendations. Therefore, Hybrid Filtering can be considered a highly suitable method for real-world applications, particularly in e-commerce environments where both relevance and variety are essential for user satisfaction and business success.

References

Al Mamun, A., Sohel, M., Mohammad, N., Sunny, M. S. H., Dipta, D. R., & Hossain, E. (2020). A comprehensive review of the load forecasting techniques using single and hybrid predictive models. *IEEE Access*, *8*, 134911–134939.

Alzoghbi, A., Arrascue Ayala, V. A., Fischer, P. M., & Lausen, G. (2016). *Learning-to-Rank in research paper CBF recommendation: Leveraging irrelevant papers*.

Chen, R., Hua, Q., Chang, Y.-S., Wang, B., Zhang, L., & Kong, X. (2018). A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks. *IEEE Access*, *6*, 64301–64320.

Chen, Y.-C., Pal, N. R., & Chung, I.-F. (2011). An integrated mechanism for feature selection and fuzzy rule extraction for classification. *IEEE Transactions on Fuzzy Systems*, *20*(4), 683–698.

De Campos, L. M., Fernández-Luna, J. M., Huete, J. F., & Rueda-Morales, M. A. (2010). Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. *International Journal of Approximate Reasoning*, *51*(7), 785–799.

Elavarasan, D., Vincent PM, D. R., Srinivasan, K., & Chang, C.-Y. (2020). A hybrid CFS filter and RF-RFE wrapper-based feature extraction for enhanced agricultural crop yield prediction modeling. *Agriculture*, *10*(9), 400.

Galea, A., & Capelo, L. (2018). *Applied Deep Learning with Python: Use scikit-learn, TensorFlow, and Keras to create intelligent systems and machine learning solutions*. Packt Publishing Ltd.

Geetha, G., Safa, M., Fancy, C., & Saranya, D. (2018). A hybrid approach using collaborative filtering and content based filtering for recommender system. *Journal of Physics: Conference Series*, *1000*(1), 12101.

Gupta, D., Chopra, N., Nair, N., & Sharma, P. (2020). Enhancing User Experience with AI-Powered Recommendation Engines: A Comparative Study of Collaborative Filtering, Neural Collaborative Filtering, and Matrix Factorization Algorithms. *Journal of AI ML Research*, *9*(4).

Jawaheer, G., Weller, P., & Kostkova, P. (2014). Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *4*(2), 1–26.

Joshi, A. P., & Patel, B. V. (2021). Data preprocessing: the techniques for preparing clean and quality data for data analytics process. *Orient. J. Comput. Sci. Technol*, *13*(0203), 78–81.

Kim, B. M., Li, Q., Park, C. S., Kim, S. G., & Kim, J. Y. (2006). A new approach for combining content-based and collaborative filters. *Journal of Intelligent Information Systems*, *27*(1), 79–91.

Lai, C.-H., & Hsu, C.-Y. (2021). Rating prediction based on combination of review mining and user preference analysis. *Information Systems*, *99*, 101742.

Madni, A. M., & Sievers, M. (2014). System of systems integration: Key considerations and challenges. *Systems Engineering*, *17*(3), 330–347.

Mehdiyev, N., Enke, D., Fettke, P., & Loos, P. (2016). Evaluating forecasting methods by considering different accuracy measures. *Procedia Computer Science*, *95*, 264–271.

Molina, L. C., Belanche, L., & Nebot, À. (2002). Feature selection algorithms: A survey and experimental evaluation. *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, 306–313.

Nguyen, T. T., Hui, P.-M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity. *Proceedings of the 23rd International Conference on World Wide Web*, 677–686.

Olaolu, A. M., Abdulsalam, S. O., Mope, I. R., & Kazeem, G. A. (2018). A comparative analysis of feature selection and feature extraction models for classifying microarray dataset. *Comput Inf Syst J*, *29*(1).

Qaid, T. S., Mazaar, H., Al-Shamri, M. Y. H., Alqahtani, M. S., Raweh, A. A., & Alakwaa, W. (2021). Hybrid deep-learning and machine-learning models for predicting COVID-19. *Computational Intelligence and Neuroscience*, *2021*(1), 9996737.

Rosário, A., & Raimundo, R. (2021). Consumer marketing strategy and E-commerce in the last decade: a literature review. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(7), 3003–3024.

Shi, Y., Larson, M., & Hanjalic, A. (2014). Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)*, *47*(1), 1–45.

Shrivastava, R., & Sisodia, D. S. (2019). Product recommendations using textual similarity based learning models. *2019 International Conference on Computer Communication and Informatics (ICCCI)*, 1–7.

Sujatha, E., & Radha, R. (2021). A Hybrid of Proposed Filtration and Feature Selections to Enhance the Model Performance. *Indian Journal of Science and Technology*, *14*(24), 2039–2050.

Valizadegan, H., Jin, R., Zhang, R., & Mao, J. (2009). Learning to rank by optimizing ndcg measure. *Advances in Neural Information Processing Systems*, *22*.

Verma, C., Hart, M., Bhatkar, S., Parker-Wood, A., & Dey, S. (2015). Improving scalability of personalized recommendation systems for enterprise knowledge workers. *IEEE Access*, *4*, 204–215.

Yacouby, R., & Axman, D. (2020). Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, 79–91.

Zheng, L. (2015). *Performance evaluation of latent factor models for rating prediction*.

Zheng, X., Cheung, C. M. K., Lee, M. K. O., & Liang, L. (2015). Building brand loyalty through user engagement in online brand communities in social networking sites. *Information Technology & People*, *28*(1), 90–106.