Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.3, June 2025, pp. 683~692
e-ISSN: 2548-6861

Web-Based Makeup Recommendation System Using Hybrid Filtering

Putu Mia Setya Utami [1]*, I Nyoman Prayana Trisna [2]*, Wayan Oger Vihikan [3]*
* Teknologi Informasi, Universitas Udayana
miasetyautami@gmail.com[1], prayana.trisna@unud.ac.id[2], oger_vihikan@unud.ac.id[3]

Article Info

Article history:

Received 2025-03-25
Revised 2025-04-08
Accepted 2025-04-14

Keyword:

Collaborative Filtering,
Content-Based Filtering,
Hybrid Filtering,
Recommendation System,
Weighted Hybrid

ABSTRACT

The increasing use of makeup products in the modern era, driven by evolving beauty trends and e-commerce accessibility, presents challenges in selecting products suited to individual skin types and conditions. A recommendation system addresses this issue by enhancing selection efficiency. This study explores the implementation of Content-Based Filtering (CBF) using TF-IDF and Cosine Similarity, Collaborative Filtering (CF) with Singular Value Decomposition (SVD), and a Hybrid Filtering approach integrating both methods through Weighted Hybrid techniques. The system's performance is evaluated across two user scenarios: new users (without prior ratings) and old users (with rating history). The evaluation method includes Precision, Normalized Discounted Cumulative Gain (NDCG), and accumulation of the best scenario based on user opinion. Results show that Hybrid Filtering outperforms CBF and CF, with notable differences between user groups. For new users, 32% prefer Scenario 1, which emphasizes CBF, achieving 80.8% Precision and 89.73% NDCG. For old users, 23% favor Scenario 2, attaining 83.4% Precision and 90.31% NDCG.

I. Introduction

In the modern era, the increasing use of makeup has led to a continuous rise in demand for cosmetic products [1]. Makeup has undeniably become an essential aspect of many women's lives, as it significantly enhances self-confidence [2]. As beauty trends evolve and the variety of available products expands, consumers face challenges in selecting the most suitable product. Therefore, individuals must consider their skin conditions before selecting and applying makeup products to ensure optimal compatibility and effectiveness [3].

Recommendation systems play a crucial role in assisting users in identifying products that best match their characteristics, such as skin type, undertone, and skin tone. By leveraging these systems, users can enhance efficiency in selecting products that align with their needs and preferences. Two commonly employed approaches in recommendation systems are Content-Based Filtering and Collaborative Filtering [4].

Numerous studies have investigated the effectiveness of these approaches. For example, one study implemented TF-IDF-based Content-Based Filtering and Cosine Similarity in a training program recommendation system, achieving an average precision of 88% [5]. Another study applied Collaborative Filtering to a dataset from Goodreads using the Alternating Least Squares (ALS) and Singular Value Decomposition (SVD) algorithms. The results demonstrated that the SVD algorithm outperformed ALS, yielding a lower RMSE of approximately 0.86822 and an MAE of around 0.69032 [6].

While effective, each method has limitations. Content-Based Filtering struggles with the serendipity problem, whereas Collaborative Filtering faces challenges related to the cold-start issue. To address these drawbacks, Hybrid Filtering was developed by integrating both approaches [7]. Several studies have implemented this method. One study combined Collaborative Filtering with the SVD algorithm with Content-Based Filtering using TF-IDF, Cosine Similarity, and Random Forest. The final recommendations were generated using a Weighted Hybrid technique with varying weight distributions, where the optimal performance was achieved with a 0.4:0.6 ratio of Content-Based Filtering to Collaborative Filtering [8]. Another study combined TF-

IDF-based Content-Based Filtering with Convolutional Neural Network (CNN)-based Collaborative Filtering, experimenting with different weight distributions. The findings indicated that a 20% Content-Based Filtering and 80% Collaborative Filtering ratio yielded the highest hybrid score, achieving a user satisfaction rate of 80% [9].

The Weighted Hybrid technique implemented in previous studies has been shown to reduce prediction errors and enhance recommendation accuracy. Building on these findings, this study focuses on building a Hybrid Filtering recommendation system that improves accuracy while addressing the limitations of each method. In addition to emphasizing the Hybrid Filtering approach, this study will also independently analyze the effectiveness of each method and implement the system as a web-based platform to enhance accessibility and usability. Further innovations include the incorporation of a more diverse range of hybrid filtering weight variations, the involvement of real users in the evaluation process to provide a realistic assessment of recommendation effectiveness, and system testing under two scenarios: new users who have not provided any ratings and old users with prior rating histories. Through these advancements, this study seeks to support users in making decisions and enhance their overall satisfaction with the recommended makeup products.

II. METHODS

The development of a makeup recommendation system follows a structured methodology to ensure accuracy and effectiveness in delivering personalized recommendations to users. The overall workflow is depicted in Figure 1. The implementation uses the Python programming language and the Google Colab platform, supported by various libraries such as Pandas, NumPy, Scikit-learn, Surprise, NLTK, Selenium, and Beautiful Soup.

The initial phase involves makeup product data from various local Indonesian brands. The dataset encompasses product categories, including face, lip, cheek, and eye makeup. In addition to product attributes, user review data is incorporated to provide supplementary insights into user experiences with the products. The collected data undergoes a comprehensive preprocessing phase to ensure consistency, remove noise, and standardize formats for further analysis.

Following the preprocessing stage, the refined data is utilized in three primary recommendation approaches: Content-Based Filtering, Collaborative Filtering, and Hybrid Filtering. Hybrid Filtering, implemented in Scenarios 1 to 5, integrates Content-Based and Collaborative Filtering methods using various ratio combinations to enhance recommendation accuracy and mitigate each approach's limitations. Content-Based Filtering, employed in Scenario 6, generates recommendations by analyzing product attributes. Collaborative Filtering, applied in Scenario 7, identifies patterns in user interaction data to provide personalized suggestions.

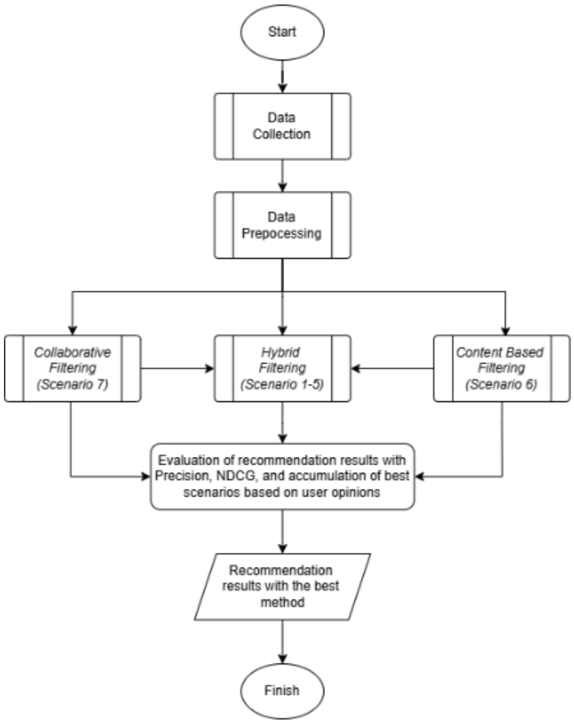

Figure 1. Flowchart of The Research

Each recommendation technique is implemented and evaluated using multiple performance metrics. This evaluation is crucial in determining the most effective approach for improving recommendation quality, ensuring users receive product suggestions tailored to their preferences and skin characteristics. The performance metrics are calculated based on feedback collected directly from users, where users assess the relevance of the recommended products. This user-driven evaluation provides a more realistic understanding of the model's effectiveness.

A. Data Collection

Data collection is a crucial process for obtaining relevant data to support research. In this study, data was from the Female Daily website (https://femaledaily.com/). The data collection process is carried out in several stages. The first stage involved gathering information on makeup brands originating from Indonesia. Subsequently, the makeup product data from the local brands was collected. Next, retrieve detailed information about each product, including product descriptions and user ratings. The collected data was then organized based on the needs of two recommendation approaches. For the Content Based Filtering method, product attributes such as makeup part, makeup type, product name, brand name, shade name, and product description were utilized to capture the characteristics of each item. Meanwhile, the Collaborative Filtering approach leveraged user interactions, specifically using user ID, product ID, and rating (stars), to generate personalized recommendations based on user behavior patterns. Finally, all collected data was stored in a MongoDB database for further analysis.

B. Data Preprocessing

Data preprocessing is a critical stage, aimed at cleaning and preparing the data before it is utilized for analysis or modelling. This step ensures the quality and reliability of the data, thereby enhancing the accuracy and efficiency of the implemented model. The preprocessing workflow is depicted in Figure 2.

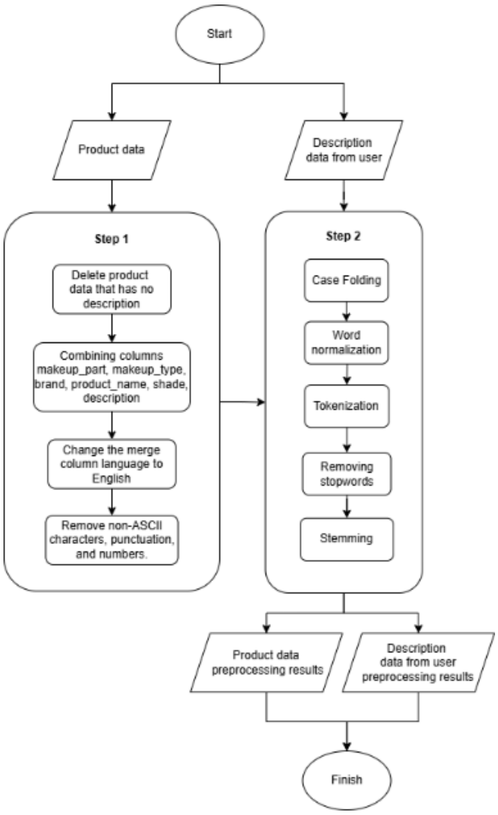

Figure 2. Flowchart of Data Preprocessing

The preprocessing of product descriptions is conducted in two main stages. The first stage involves filtering the data by removing products without descriptions, merging relevant columns, translating product descriptions into English, and eliminating unnecessary non-ASCII characters, punctuation, and numerical values. The second stage comprises a series of text processing techniques, including case folding, word normalization, tokenization, stop word removal, and stemming, to refine the text representation and improve its suitability for further analysis.

C. Content-Based Filtering

Content-Based Filtering is a recommendation system approach that analyses item characteristics and compares them with items previously preferred by the user [10]. In this study, the developed recommendation system employs the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm combined with Cosine Similarity to measure the degree of similarity between items.

To support this approach, the system utilizes product data that has been preprocesses, including attributes such as makeup part, makeup type, product name, brand name, shade name, and product description. In addition, it also incorporates user input in the form of textual descriptions (such as user preferences or product expectations), which have also undergone the same preprocessing steps. By comparing the processed user input with the processed product data, the system can recommend items that are most similar to what the user is looking for.

TF-IDF algorithm is a method used to assess the importance of a word in an entire document collection [11]. The text processing procedure using the TF-IDF method consists of several key stages. First, Term Frequency (TF) is calculated to determine how frequently a term appears within a document. Subsequently, Document Frequency (DF) is computed to identify the number of documents containing the term. Based on the DF value, the Inverse Document Frequency (IDF) is calculated. Finally, the TF and IDF values are multiplied to construct a feature matrix representing the document in vector form. The calculation of TF, DF, IDF, and TF-IDF values is performed according to Equation 1, 2, and 3.

$$tf_{(i,j)} = f_{(i,j)} \tag{1}$$

$$idf_{(i)} = \log(\frac{N}{df_i}) \tag{2}$$

$$W_{(i,j)} = tf_{(i,j)} \; x \; idf_{(i)} \tag{3}$$

Description:
$tf_{(i,j)}$: Frequency of term i appearing in document j.
$f_{(i,j)}$: Count of occurrences of term i within document j.
$idf_{(i)}$: Inverse document frequency
$df_{(i)}$: Number of available documents with the word i in them
N : The total number of documents in the system
$W_{(i,j)}$: Results of weighting of TF-IDF document values

Once the TF-IDF stage is completed, the resulting vectors are utilized to compute item similarity using Cosine Similarity. Cosine Similarity is a method used to measure the level of similarity between two vectors in n-dimensional space [10]. This method measures the similarity between two vectors by assessing the angle between them, with higher values representing a stronger similarity. The product with the highest similarity score is subsequently recommended to the user. The computation of Cosine Similarity follows Equation 4.

$$Cos = \frac{A.B}{|A||B|} = \frac{\sum_{i=1}^{n} A_i \; x \; B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} x \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{4}$$

Description:
A : Vector A B : Vector B
A_i : Term i within document A.
B_i : Term i within document B
$A.B$: Dot product of vectors A and B
$|A|$: Length of vector A
$|B|$: Length of vector B
$|A||B|$: Cross product between $|A|$ and $|B|$

D. Collaborative Filtering

Collaborative Filtering is a technique used to assess and evaluate items based on input or preferences provided by other users [12]. In this study, the Collaborative Filtering method employed is Singular Value Decomposition (SVD). SVD is a matrix factorization technique that represents the relationship between users and items in a latent factor space of dimension f. In this model, each item i is represented by a latent vector q_i, while each user u is represented by a latent vector p_u. The interaction between user u and item i is computed as the dot product of these latent vectors, $q_i^T p_u$, which quantifies the degree of preference or suitability of a user for a particular item [13]. The computation of SVD follows Equation 5 and 6.

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T p_u \quad (5)$$

$$\sum_{r_{ui} \epsilon R_{train}} (r_{ui} - \hat{r}_{ui})^2 + \lambda (b_i^2 + b_u^2 + \|q_i\|^2 + \|p_u\|^2) \quad (6)$$

Description:
$\hat{r}_{ui}$: Predicted rating result
r_{ui} : Actual rating
μ : Global rating
b_i : Bias of item
b_u : Bias of user
q_i : Factor of item
p_u : Factor of user
λ : Regularization

In the process of forming the SVD model, the dataset comprises three key components: user ID, product ID, and rating (stars). The data is partitioned into two subsets such as a training set and a test set. The SVD model is trained using the training data, while its performance is evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) metrics. The SVD model that achieves the best evaluation performance is selected for rating prediction.

After selecting the optimal SVD model, all products in the dataset are stored as the complete set of available products. Based on the user's rating history, the products that have been previously rated by a specific user x are identified and stored as rated products. The set of unrated products is then determined by excluding the rated products from the complete set of available products. The trained SVD model is subsequently utilized to predict the ratings for these unrated products, generating predictions within the rating scale of 1 to 5. To ensure seamless integration within the Hybrid Filtering framework, the predicted ratings are normalized using min-max scaling, thereby standardizing the values for further processing.

Although SVD relies on historical data to generate accurate ratings for users and products, it can still provide predictions for new users or new items by falling back on the global mean rating. This output can be combined with the results from Content-Based Filtering (CBF) to produce more relevant recommendations for new users within the Hybrid Filtering method.

E. Hybrid Filtering

Hybrid Filtering is a method that combines multiple recommendation techniques to improve the effectiveness of the recommendation system [14]. In this study, the Hybrid Filtering method is applied by integrating Content-Based Filtering and Collaborative Filtering. The combination method used Weighted Hybrid Filtering, where each technique is assigned a specific weight, which is then integrated to generate the final recommendation. The implementation of the Hybrid Filtering method is illustrated in Figure 3.

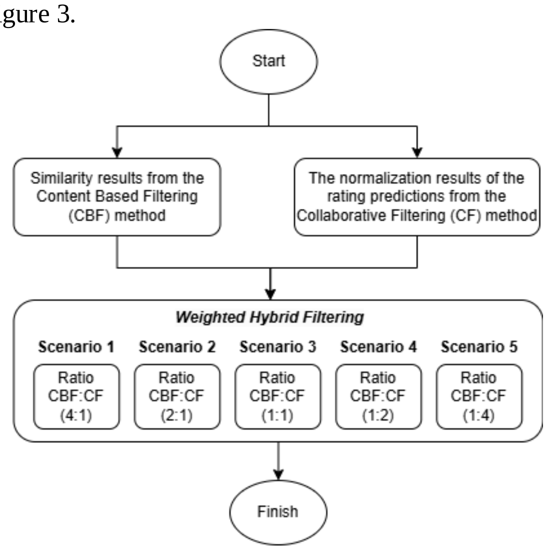

Figure 3. Flowchart of Hybrid Filtering

The process of combining values from Content-Based Filtering and Collaborative Filtering is conducted by testing various static weighting scenarios. This approach allows for the exploration of how different weight distributions affect the overall recommendation performance. The final recommendation score is calculated using a weighted formula that integrates both Content-Based and Collaborative components, as illustrated in Equation 7.

$$R_{hybrid} = R_{CBF} * \left(\frac{W_{CBF}}{W_{total}}\right) + R_{CF} * \left(\frac{W_{CF}}{W_{total}}\right) \quad (7)$$

Description:
R_{hybrid} : Hybrid prediction
R_{CBF} : Rating predictions of CBF
R_{CF} : Rating predictions of CF
W_{CBF} : Ratio of CBF
W_{CF} : Ratio of CF
W_{Total} : Sum of the CBF and CF ratios

F. Evaluation Method

Evaluation methods describe how the performance of a recommendation system is measured based on the applied approach. Several evaluation metrics commonly used include MAE, RMSE, MAPE, Precision, and NDCG.

1) Mean Absolute Error (MAE)

Metric for calculate the error of predicted ratings in a recommendation system [15] . A lower MAE value indicates higher prediction accuracy and better overall performance of the recommendation model. The calculation of MAE can be seen in Equation 8.

$$MAE = \frac{1}{N}\sum_{i=0}|R_i - P_i| \tag{8}$$

Description:
N : Total number of items that have predicted ratings
R_i : Actual rating value of the items used
P_i : Predicted rating value of the predicted items

2) Root Mean Square Error (RMSE)

Metric for calculate the average squared difference between actual ratings and predicted ratings generated by the system [16]. A higher RMSE value indicates greater prediction errors, signifying a substantial level of inaccuracy in the recommendations. The calculation of RMSE can be seen in Equation 9.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}} \tag{9}$$

Description:
n : Number of items rated by the user
y_i : Predicted rating value of item i
x_i : Actual rating value of item i

3) Mean Absolute Percentage Error (MAPE)

Metric for calculate the average percentage error over multiple periods [17]. The computation involves comparing actual data with predicted values. A lower MAPE value indicates higher prediction accuracy. The calculation of MAPE can be seen in Equation 10.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| x\ 100\% \tag{10}$$

Description:
n : Number of data
y_t : Actual value at period t
$\hat{y}_t$: Predicted value at period t

4) Precision

Metric for evaluate the accuracy of a model in predicting data classified as positive [18]. It is calculated by dividing the count of correctly predicted positive instances by the total number of positive predictions generated by the model. The calculation of Precision can be seen in Equation 11.

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

Description:
TP : Number of items with rating values that are successfully predicted and identified as recommendations that match the user's preferences.
FP : Number of items with rating values that are not successfully predicted and identified as recommendations that do not match the user's preferences.

5) Normalized Discounted Cumulative Gain (NDCG)

A method used to assess the performance of a recommendation system based on the relevance of the recommended items [15]. This approach is widely utilized to evaluate the performance of ranking mechanisms and the positioning of relevant items within a recommendation list [19]. The NDCG score ranges from 0.0 to 1.0, where a value of 1.0 signifies an optimal ranking of recommendations. The formulation of NDCG is provided in Equations 12 and 13.

$$DCG = \sum_{i=1}^{ranks}\frac{gains}{log_2(i+1)} \tag{12}$$

$$NDCG = \frac{DCG}{IDCG} \tag{13}$$

Description:
$gains$: Relevance of recommendation results at position i
$ranks$: Number of recommended items
DCG : Relevance for each recommendation result
$IDCG$: Order of the best relevance values based on the actual order

6) Accumulation of the Best Scenario Based on User Opinions

The accumulation of user opinions is a method used to evaluate the overall best scenario based on subjective user preferences. In this process, each user is asked to rank seven recommendation scenarios from the most relevant (rank 1) to the least relevant (rank 7). After all users have submitted their rankings, the results are aggregated to identify which scenarios most frequently occupy each rank position. For example, in the first rank, the scenario that is chosen most often is considered the best-performing, whereas, in the seventh rank, the most frequently chosen scenario is regarded as the least effective according to user perceptions.

TABLE I
EXAMPLE OF USER INPUT RESULTS RELATED TO SCENARIO RANK

User ID	Hybrid Filtering Scenario Rank					CBF Scenario Rank	CF Scenario Rank
	S-1	S-2	S-3	S-4	S-5	S-6	S-7
100	6	5	1	2	3	7	4
12	6	4	1	3	2	7	5
10	6	1	2	3	4	7	5
34	4	3	2	1	5	5	7
55	4	3	1	2	6	5	7

TABLE II
EXAMPLE OF ACCUMULATED RESULTS OF EACH SCENARIO RANK

Rank	Accumulated Number of Rank						
	S-1	**S-2**	**S-3**	**S-4**	**S-5**	**S-6**	**S-7**
1	0	1	**3**	1	0	0	0
2	0	0	2	2	1	0	0
3	0	2	0	2	1	0	0
4	2	1	0	0	1	0	1
5	0	1	0	0	1	2	2
6	3	0	0	0	1	0	0
7	0	0	0	0	0	3	2

Table 1 presents an example of user rankings for the seven scenarios. Subsequently, Table 2 shows the accumulation of ranks assigned to each scenario. Based on the aggregated results, it can be concluded that Scenario 3 (with a 1:1 weight ratio) is the best-performing scenario, as it most frequently occupies the first rank.

III. RESULT DAN DISCUSSION

This section presents the research findings, including the data collection process and the analysis of the recommendation system performance. The results are evaluated to determine the effectiveness of the proposed Content-Based Filtering, Collaborative Filtering, and Hybrid Filtering.

A. *Data Collection*

This research gathered data on makeup products from the Female Daily website using the web scraping technique. This data collection process successfully obtained 8,903 makeup products, encompassing various categories and brands. The extracted information will be utilized in the Content-Based Filtering approach, using key attributes such as makeup part, product name, brand name, shade name, product description.

Furthermore, user review data was collected to enhance the implementation of the Collaborative Filtering method. A total of 149,127 reviews were successfully gathered, providing insights into user interactions and preferences regarding different makeup products. For the development of the Collaborative recommendation model, three primary attributes are considered: user ID, product ID, and rating (1-5).

B. *Data Preprocessing*

The preprocessing of product data was carried out in two stages, resulting in a cleaned dataset of 7,734 products. An example of the processed data is provided in Figure 4.

```
'lip lipstick exclus matt lipstick wardah pink wardah exclus matt lipstick lip sh
ade intens color one swipe long last product matt finish light comfort use produc
t enrich vitamin e antioxid health look lip textur lipstick creami light lip last
hour use appli upper low lip suitabl ingredi'
```

Figure 4. Data Preprocessing Result

Furthermore, the review data underwent a reduction process to ensure alignment with the available products. As a result, the final dataset consists of 132,734 reviews from 39,842 users.

In terms of rating distribution, the data is heavily skewed toward positive ratings: only 1,469 reviews received 1 star, 4,660 received 2 stars, 15,984 received 3 stars, 36,783 received 4 stars, and the majority, 73,837 reviews, were rated 5 stars. This indicates a strong bias toward higher ratings.

Moreover, considering the scale of the user-item matrix, the number of interactions suggests that the dataset is highly sparse, with a sparsity level of approximately 99.5%. This means that the vast majority of possible user-product combinations lack interaction data.

C. *Result For Content-Based, Collaborative, Hybrid Filtering*

1) *Content-Based Filtering*

The similarity between the user's input and the available product descriptions is computed and utilized to generate recommendations. The results of the content-based approach for lipstick recommendations with product ID 6526 are illustrated in Figure 5.

	product_id	product_name	score
0	3690	Watercolor Hydrating Lipstick	0.434364
1	8638	Exclusive Matte Lipstick	0.431550
2	8731	Long Lasting Lipstick	0.424641
7	6011	Lip Cream	0.416786
9	6503	Lipstick	0.409260

Figure 5. Result Content-Based Filtering

2) *Collaborative Filtering*

The optimal SVD model was obtained with n_factors = 100, n_epochs = 30, lr_all = 0.025, and reg_all = 0.05. The model achieved an RMSE of 0.7238, an MAE of 0.5172, and an MAPE of 16.00%. This optimized SVD model is utilized for rating prediction. The predicted lipstick ratings for User ID 12 are illustrated in Figure 6.

	product_id	product_name	score_svd
2107	2389	Lip Moss	1.000000
460	483	Metalip	0.954751
5108	5885	Matte Lipstick	0.954361
1180	1262	Minimattes Collection	0.933914
1443	1560	Poppin Matte The 90s Edition	0.928294

Figure 6. Result Predicted Rating Using SVD

3) *Hybrid* Filtering

Among the five hybrid filtering scenarios evaluated, the results of the combination of CBF and CF at a 1:1 ratio are presented in Figure 7.

	product_id	product_name	final_score
4	3691	Watercolor Hydrating Lipstick	0.511257
7	6011	Lip Cream	0.482176
12	8607	Long Lasting Lipstick	0.474433
37	6632	Gotta Be Matte Lip Cream	0.450298
1	8638	Exclusive Matte Lipstick	0.441368

Figure 7. Result Hybrid Filtering

D. Website Interface

This section discusses the implementation of the recommendation system within the user interface. The implementation encompasses various components, as illustrated in the figure below.

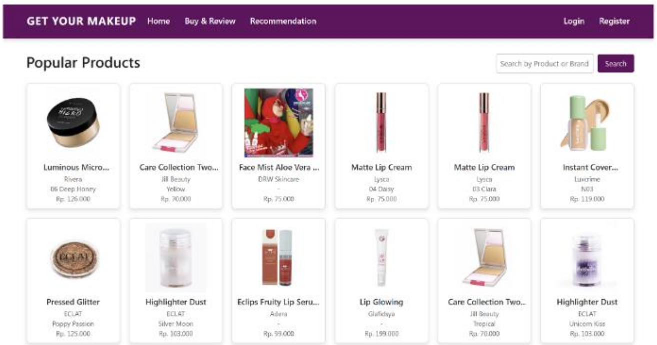

Figure 8. Dashboard Page View

The dashboard page provides details about popular makeup products, as illustrated in Figure 8. Furthermore, users can search for their desired products and can rate products they have previously purchased. This process enables the model to analyze and comprehend the user's preference history.

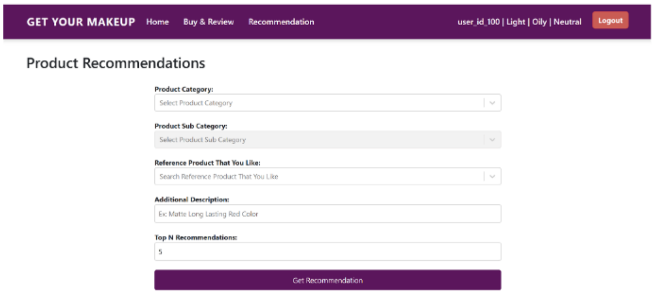

Figure 9. Form Product Recommendation

In Figure 9, users can request recommendations and evaluate the results. Before this, users must input detailed information about the desired and preferred products to enable the system to generate more accurate recommendations. The necessary data includes product categories, subcategories, product references, additional descriptions, and the number of recommendations (Top-N).

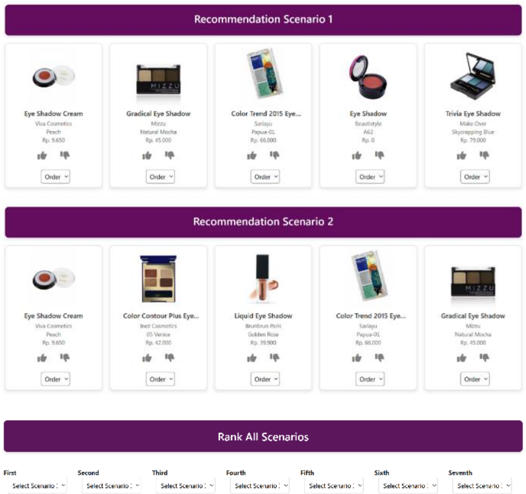

Figure 10. Recommendation Result and Evaluation

After completing the recommendation form, users will be directed to the recommendation results page, as illustrated in Figure 10. On this page, users are required to evaluate seven scenarios: five Hybrid approaches, one Content-Based approach (Scenario 6), and one Collaborative approach (Scenario 7). The results of this evaluation will be analyzed to determine the most effective recommendations for both new and old users.

E. Result and Discussion Evaluation

The recommendation system was tested under two primary scenarios: one for new users and another for old users. Each scenario encompassed seven recommendation methods to assess the system's performance. Throughout the testing process, a total of 200 data were collected, 100 data from new users and 100 from old users. The test results are presented as follows.

TABLE III
ACCUMULATED RESULTS IN NEW USERS

	Ranking Order						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Scenario 1	**32**	13	19	11	9	4	12
Scenario 2	17	**24**	14	13	12	13	7
Scenario 3	8	20	11	**21**	15	14	11
Scenario 4	12	5	13	16	**25**	17	12
Scenario 5	8	13	15	16	12	**23**	13
Scenario 6	16	12	**21**	14	11	14	12
Scenario 7	7	13	7	9	16	15	**33**

The first result presents the optimal scenario based on user preferences and opinions. Users are asked to order/rank the scenarios from best to worst, with Order 1 representing the most favorable scenario and Order 7 indicating the least preferred. The accumulated orders of the best to worst scenarios, as determined by users, are presented in Table 3 and 4.

TABLE IV
ACCUMULATED RESULTS IN OLD USERS

	Ranking Order						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Scenario 1	20	17	**20**	8	12	8	15
Scenario 2	**23**	16	13	18	10	9	11
Scenario 3	17	**21**	19	14	11	9	9
Scenario 4	12	16	10	**26**	15	14	8
Scenario 5	9	8	17	8	20	17	**20**
Scenario 6	11	11	10	14	9	**27**	18
Scenario 7	8	11	11	12	**23**	16	19

Based on the table 1 and 2, the bolded values represent the highest value in each ranking order. For example, in Order 1 for new users, the bolded value is 32, indicating that 32 out of 100 data points selected Scenario 1 as the most favorable for new users. To provide a clearer analysis, these values are visualized in a graph as illustrated in Figure 11. In the graph, purple represents scenarios using the Hybrid Filtering method, pink indicates Scenario 6, which applies Content-Based Filtering, and orange denotes scenarios utilizing the Collaborative Filtering method.

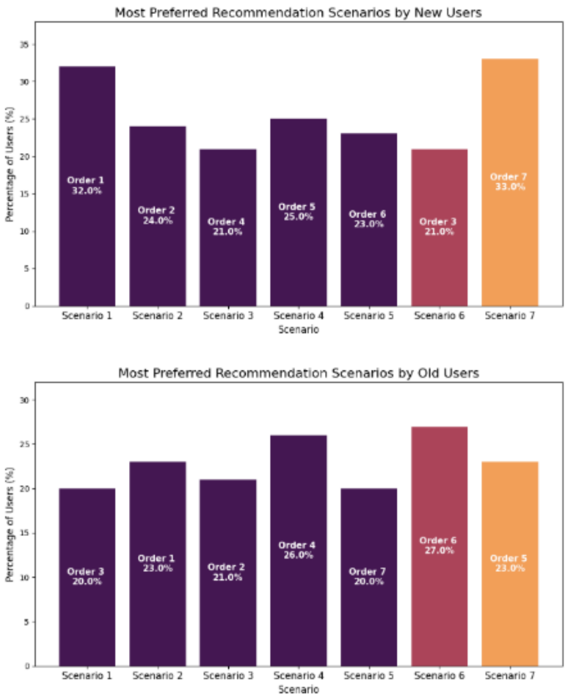

Figure 11. Most Preferred Recommendation Scenarios by New and Old User Opinions

Figure 11 presents the evaluation results of the best-performing recommendation scenarios based on the preferences of both new and old users. Scenario 1 was the most preferred by new users, accounting for 32% of the evaluations, while Scenario 2 was favored by 23% of old users.

For new users, the most effective scenario prioritized a Hybrid Filtering approach with a stronger emphasis on Content-Based Filtering (CBF). This indicates that new users relied more heavily on product descriptions due to the lack of prior interactions. Conversely, the Collaborative Filtering (CF) approach was found to be less effective for this group, as reflected by the lowest ranking of Scenario 7, which relied solely on CF.

In contrast, old users tended to prefer Scenario 2, which combines CBF and CF with a 2:1 ratio. Although CBF remained dominant, the contribution of CF began to show its effectiveness among old users. This is also reflected in the improved ranking of Scenario 7 among old users compared to new users.

The second result pertains to the precision calculation for each scenario. All user input values are processed to determine the average precision. The findings are presented in Figure 12.

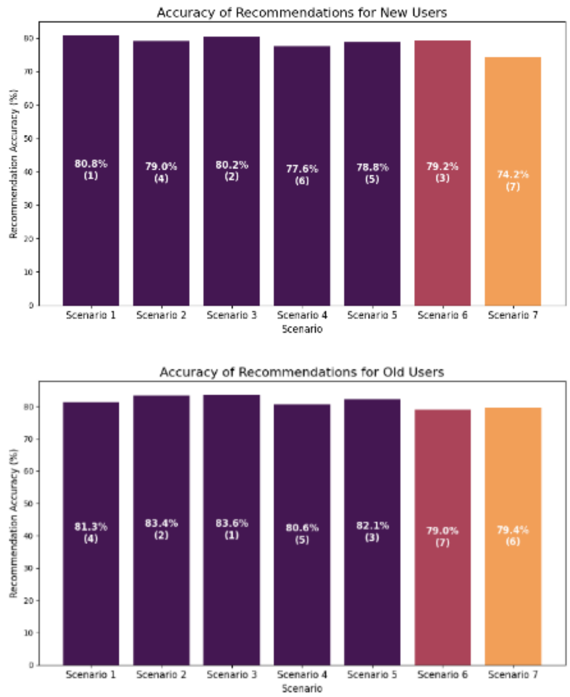

Figure 12. Accuracy of Recommendation for New and Old Users

Based on Figure 12, the evaluation results indicate that Scenario 1 achieved the highest precision among new users, while Scenario 3 ranked highest for old users.

Overall, the graph shows that the differences in average precision across scenarios within each user group are not highly significant. However, there is a noticeable trend of higher precision among old users compared to new users. This suggests that the recommendation system tends to perform more effectively when users have sufficient interaction history. Additionally, the Top 3 scenarios for both user groups all employed Hybrid Filtering approaches, indicating that hybrid methods generally produce more relevant recommendations than single-method approaches.

For new users, precision values varied considerably across scenarios. Scenarios emphasizing Hybrid Filtering with a stronger Content-Based Filtering (CBF) component proved more effective. Scenario 1 recorded the highest precision at 80.8%, while Scenario 7, which relied solely on Collaborative Filtering (CF), had the lowest. This supports the notion that CF is less suitable for new users due to their limited interaction data.

Among old users, the top five scenarios in terms of precision were all based on Hybrid Filtering, while pure CBF and CF approaches ranked lowest. Although CF began to show more positive contributions as user interaction histories grew, it still could not outperform the hybrid approaches. This performance gap can be attributed to the hybrid method's ability to combine CF's strength in modelling user interaction patterns with CBF's ability to utilize product content information, resulting in more accurate and relevant recommendations.

The final calculation involves determining the Normalized Discounted Cumulative Gain (NDCG) for each scenario. The average NDCG value is computed to assess the ranking quality of the recommended products. The results of the NDCG calculations for both new and old users are presented in Figure 13.

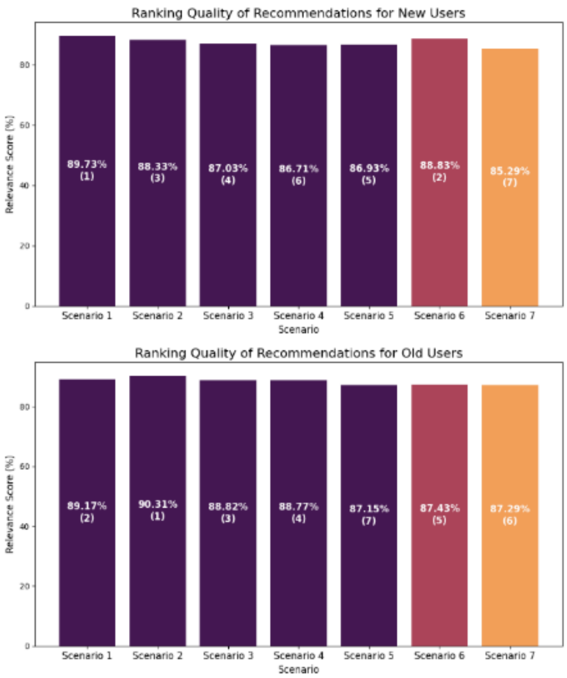

Figure 13. Average NDCG Graph for New and Old Users

As shown in Figure 13, the evaluation results of recommendation ranking quality for each scenario are measured using the average NDCG scores from new and old users. Scenario 1 achieved the highest NDCG among new users, while Scenario 2 ranked highest for old users.

For new users, Scenario 1 recorded the highest NDCG at 89.73%, followed by Scenario 6 at 88.83% and Scenario 2 at 88.33%. These results highlight the effectiveness of Hybrid Filtering approaches, particularly those emphasizing Content-Based Filtering (CBF), in generating relevant recommendation sequences for users with limited interaction history. The strong performance of Scenario 6, which is purely CBF-based, also suggests that descriptive product information can significantly aid in aligning recommendations with user preferences.

Among old users, the overall NDCG values were higher than those of new users, and the differences between scenarios were relatively small. Scenario 2 emerged as the best-performing for this group, indicating that the combination of CBF and CF becomes more effective as users accumulate more interaction data. This supports the idea that hybrid approaches can better capture user behavior over time and deliver more personalized recommendation rankings.

Overall, this study demonstrates that the application of a Hybrid Filtering recommendation system is more effective compared to single-method approaches. These findings are consistent with the study by Pratama et al. in 2023, which also concluded that hybrid methods produce more accurate and relevant recommendations. However, there is a notable difference in the dominant method used. While Pratama et al. emphasized Collaborative Filtering due to the characteristics of their dataset in the tourism sector, where users tend to rely on others' experiences, this study places more emphasis on Content-Based Filtering. The difference is likely due to the nature of the dataset, which involves makeup products where users are more inclined to depend on product information rather than the preferences of others. Consequently, a dominant ratio favoring CBF is found to be more effective for both new and existing users.

The selection of the best-performing recommendation scenario is determined based on user opinion evaluation. This is considered to provide a more representative measure of system performance, as it captures the actual user experience and perceived relevance of the recommendations. Although metrics such as Precision and NDCG offer useful quantitative insights, user feedback is ultimately more reflective of real-world effectiveness. Based on this evaluation, Scenario 1 is identified as the most appropriate for new users, while Scenario 2 is deemed most suitable for existing users.

Although the developed recommendation system performs well, some limitations should be addressed in future work. One of the main issues is the lack of flexibility in adjusting the balance between CBF and CF within the Hybrid approach. The current system applies a fixed combination ratio regardless of the user's level of interaction. This could lead to suboptimal recommendations for users with different behavioral patterns. Therefore, further development should aim to create a more adaptive system that adjusts the weighting of CBF and CF based on user activity. Future evaluations can be conducted by grouping users according to their interaction levels to determine whether an increase in user engagement should lead to a stronger emphasis on CF.

IV. CONCLUSION

Based on the research conducted on the Makeup Recommendation System with a Website-Based Hybrid Filtering Approach, it can be concluded that this recommendation system was developed by implementing three approaches, namely Content-Based Filtering (CBF), Collaborative Filtering (CF), and Hybrid Filtering which combines both. The research data was obtained through a web scraping process from the Female Daily website, covering 7,734 product data and 132,734 rating data that had gone through the preprocessing stage. In its analysis, CBF was applied using the TF-IDF and Cosine Similarity algorithms, while CF used the Singular Value Decomposition (SVD) algorithm. These two approaches were combined in Weighted Hybrid Filtering with five scenarios that were compared with the CBF and CF methods separately, both for new users (without rating history) and old users (with rating history).

The evaluation was carried out using three methods, namely the accumulation of the best scenario based on user opinion, Precision, and Normalized Discounted Cumulative Gain (NDCG). The evaluation dataset consisted of 200 data, which were divided into 100 new user data and 100 old user data. The evaluation results show that the Hybrid Filtering method provides better recommendations than the single method. New users prefer scenario 1 Hybrid Filtering which emphasizes CBF because they do not have a history of interactions. As many as 32 out of 100 evaluation data on new (32%) agree that Scenario 1 is the best choice for new users, with an average Precision of 80.8% and an average NDCG of 89.73%. Meanwhile, old users prefer Scenario 2 with a 2:1 ratio. Although CBF remained dominant, the contribution of CF began to demonstrate its effectiveness for old users with interaction history. As many as 23 out of 100 evaluation data on old users (23%) prefer scenario 2, with an average Precision of 83.4% and NDCG of 90.31%.

This system still has several limitations that can be enhanced in future developments. One notable limitation is the lack of adaptability in adjusting the ratio between CBF and CF based on user interactions. At present, the system employs a fixed ratio without accounting for varying levels of user engagement. Therefore, future improvements should focus on implementing a more dynamic approach that adjusts the ratio accordingly to generate more accurate and personalized recommendations.

REFERENCES

[1] I. Lisnawati and L. Julaeha, ‘Pengaruh Harga, Kualitas Produk Dan Electronic Word Of Mouth Terhadap Keputusan Pembelian Produk Kosmetik Emina’, *Jurnal Ilmiah Multidisiplin*, vol. 3, no. 04, pp. 128–136, 2024.

[2] A. O. Lestari and S. Nawangsari, ‘Pengaruh Citra Merek, Kualitas Produk, Keragaman Produk Dan Beauty Vlogger Terhadap Keputusan Pembelian Produk Kosmetik Make Over (Studi Kasus Pada Konsumen Wanita Di Jabodetabek)’, *Jurnal Akuntansi dan Manajemen Bisnis*, vol. 3, no. 2, pp. 153–166, 2023.

[3] M. Putri, ‘Talk Show: Cara Cerdas Memilih Kosmetik yang Aman’, *Edukasi Dan Pengabdian Masyarakat*, vol. 1, no. 2, pp. 23–28, 2021.

[4] A. Al Amin, A. Sunyoto, and H. Al Fatta, ‘Mereduksi Error Prediksi Pada Sistem Rekomendasi Menggunakan Pendekatan Collaborative Filtering Berbasis Model Matrix Factorization’, *Explore*, vol. 11, no. 2, pp. 8–14, 2021.

[5] M. F. Abdurrafi and D. H. U. Ningsih, ‘Content-based filtering using cosine similarity algorithm for alternative selection on training programs’, *Journal of Soft Computing Exploration*, vol. 4, no. 4, pp. 204–212, 2023.

[6] H. A. Adyatma and Z. K. A. Baizal, ‘Book Recommender System Using Matrix Factorization with Alternating Least Square Method’, *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1286–1292, 2023.

[7] H. H. Arfisko and A. T. Wibowo, ‘Sistem Rekomendasi Film Menggunakan Metode Hybrid Collaborative Filtering Dan Content-based Filtering’, *eProceedings of Engineering*, vol. 9, no. 3, 2022.

[8] D. E. Pratama, D. Nurjanah, and H. Nurrahmi, ‘Tourism recommendation system using weighted hybrid method in Bali Island’, *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp. 1189–1199, 2023.

[9] K. S. Y. Putri, I. M. A. D. Suarjaya, and W. O. Vihikan, ‘Sistem Rekomendasi Skincare Menggunakan Metode Content Based Filtering dan Collaborative Filtering’, *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 4, no. 3, pp. 764–774, 2024.

[10] C. A. Melyani, ‘Hotel recommendation system with content-based filtering approach (case study: Hotel in yogyakarta on nusatrip website)’, *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 15, no. 1, 2022.

[11] Y. A. Wibisono, M. Afdal, M. Mustakim, and R. Novita, ‘Implementasi Algoritma Random Forest Untuk Analisa Sentimen Data Ulasan Aplikasi Pinjaman Online Digoogle Play Store’, *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, pp. 619–626, 2024.

[12] R. K. Hapsari, I. Ibadurrohman, and S. R. Wardhana, ‘Sistem Rekomendasi Film dengan Menggunakan Pendekatan Collaborative Filtering Berdasarkan Class’, in *Prosiding Seminar Implementasi Teknologi Informasi dan Komunikasi*, 2023, pp. 159–164.

[13] F. Ricci, · Lior, R. Bracha, and S. Editors, *Recommender Systems Handbook Second Edition*. 2015.

[14] A. Pramarta and A. Baizal, ‘Hybrid Recommender System Using Singular Value Decomposition and Support Vector Machine in Bali Tourism’, *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 7, no. 2, pp. 408–418, 2022.

[15] I. Jayaperwira, A. T. Wibowo, and D. Nurjanah, ‘Anime rekomendasi menggunakan Collaborative Filtering’, *eProceedings of Engineering*, vol. 10, no. 3, 2023.

[16] R. Kusumaningrum and R. Rismiyati, ‘Film Recommender System Menggunakan Metode Neural Collaborative Filtering’, *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 3, pp. 699–708, 2023.

[17] R. Cipta and T. R. T. P. Nugraha, ‘Evaluasi Prediksi Curah Hujan dengan Algoritma Backpropogation di BMKG Cilacap’, *ZONAsi: Jurnal Sistem Informasi*, vol. 2, no. 2, pp. 97–109, 2020.

[18] M. A. Pradana and A. T. Wibowo, ‘Movie recommendation system using hybrid filtering with word2vec and restricted boltzmann machines’, *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 1, pp. 231–241, 2024.

[19] D. A. Nugroho, C. Lubis, and N. J. Perdana, ‘Film Recommendation System using Neural Collaborative Filtering Method’, *INTECOMS: Journal of Information Technology and Computer Science*, vol. 7, no. 3, pp. 926–937, 2024.