Perfume Product Selection Recommender System Using Content-based Filtering Approach

1st Fannisa Eimin Aurora
School of Computing
Telkom Universtiy
Bandung, Indonesia
fanniaurora@student.telkomuniversity.ac.id

2nd Z. K. A. Baizal*
School of Computing
Telkom Universtiy
Bandung, Indonesia
baizal@telkomuniversity.ac.id

2025 International Conference on Data Science and Its Applications (ICoDSA) | 979-8-3315-9854-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICoDSA67155.2025.11157374

***Abstract*—The process of selecting perfumes on e-commerce platforms is often affected by search bias, which challenges consumers in finding products aligned with their preferences. Previous studies using the AHP method relied on respondent satisfaction scores, which skewed towards highly positive responses and resulted in a reported accuracy of 90.07%. This study proposes a recommender system employing Content-based Filtering with the TF-IDF algorithm to enhance accuracy. TF-IDF was chosen for its efficiency in quantifying reviews and user ratings, enabling more relevant recommendations. The system was developed through data collection and filtering, sentiment analysis of reviews, TF-IDF implementation, and re-ranking recommendations based on sentiment and user ratings. The system's performance was evaluated using precision and recall metrics on Top-N recommendations, with Word2Vec employed for comparative analysis. The results of this paper demonstrate that TF-IDF significantly outperforms Word2Vec, achieving a precision of 80% and recall of 95%, compared to Word2Vec's precision of 18% and recall of 89%. Moreover, rating-based models demonstrated superior performance over sentiment-based models, achieving a maximum precision of 82.09% and a recall of 98%. This study highlights TF-IDF's effectiveness in improving recommendation accuracy and its capability to address search bias, providing a more robust and effective solution for perfume recommendations.**

***Index Terms*—content-based filtering, TF-IDF, Word2Vec, recommender systems, perfume**

I. Introduction

Currently, perfumes are among the products with increasing sales on e-commerce platforms. This growth is driven by factors like price, product quality, and promotions, which are key considerations for consumers when making purchase decisions [1] [2]. The variety of perfume products on e-commerce platforms, coupled with these factors, often challenges consumers to find products aligned with their preferences. As a result, users may purchase products that do not match their initial preferences. Therefore, recommender systems can effectively assist consumers in choosing which products to purchase.

The abundance of accessible online information also makes it difficult for consumers to validate their choices [3]. Many consumers visit e-commerce websites primarily to read reviews from other users, which provide insights into the product's quality and performance, before making their purchase decision. This is because reviews from other consumers are highly valued.

Two primary methods are commonly used to address information overload: Collaborative Filtering and Content-based Filtering [4]. Each of these methods has its limitations. Collaborative Filtering utilizes user rating data to provide recommendations, whereas Content-based Filtering recommends items similar to those the user liked in the past [5].

Previous research on perfume recommender systems has employed methods such as Fuzzy Delphi, Neural Network Classification, Analytic Hierarchy Process (AHP), and the Order of Preference Technique with Similarity to the Ideal Solution Method (TOPSIS) [6]. However, these methods have some shortcomings. First, the Fuzzy Delphi method, used to identify influential variables, involves respondent subjectivity, which introduces potential bias. Second, there has been no performance comparison between Artificial Neural Networks and other classification methods, such as Decision Trees. Similarly, AHP and TOPSIS methods, while achieving an accuracy of 90.07%, also have certain limitations for product recommendations.

For AHP, the first issue is its low scalability; as the number of criteria increases, the number of questions users must answer also increases arithmetically. Second, AHP lacks a mechanism for assigning different weights to each criterion, treating all criteria as equally important. Based on previous research, the use of AHP for recommending perfume products is less suitable due to the absence of a weighting mechanism. Furthermore, the accuracy calculated in these studies is based on respondent satisfaction, which tends to lean towards high satisfaction levels but remains limited in its accuracy value.

Therefore, our research proposes the utilization of the TF-IDF algorithm, combined with a reranking process that incorporates sentiment review and rating weight calculations. This study develops a recommender system for selecting perfume products using a Content-based Filtering approach, leveraging the TF-IDF algorithm to optimize the filtering system and assist consumers in choosing perfumes that meet their needs. The system is validated by ranking each product and comparing the results. The Content-based Filtering method is capable of recommending products across different categories to provide more effective recommendations, helping consumers select perfumes that match their preferences. However, the Content-based Filtering approach for recommending perfumes has not been extensively explored.

II. Related Work

In previous studies, recommender systems have been widely developed across various domains, such as perfume recommendations, tourist destinations, and consumer product reviews [7] [8]. We leverage prior research to analyze the relationship between earlier studies and our work to avoid plagiarism. Therefore, this chapter discusses related research to highlight the differences between our study and previous works.

G. Lee et al. [9], in their study, showed that to analyze the chemical composition of products and recommend items with similar ingredients can use Content-based Filtering. This approach also enables users to specify their desired beauty effects rather than providing specific product names, which is particularly useful for users with limited knowledge or those who have not yet discovered products they like. The study showed that Content-based Filtering is effective, achieving an average accuracy of 75% in chemical component analysis.

Research aimed at personalizing skincare needs indicated that Content-based Filtering provides recommendations tailored to users' needs based on items and user preferences. Content-based Filtering actively serves users by providing accurate information, even when the information provided by users is limited [10]. Additionally, another study utilized Content-based Filtering to recommend recipes. In this case, it enabled recipe recommendations based on attributes (ingredients) specified by the user. During the recommendation process, similarity metrics were calculated from the feature vectors in the dataset and the ingredients preferred by the user. The output consisted of the top-N recipes most similar to the user's input, along with their ingredients, for the user to choose from. The algorithm evaluated multiple attributes to determine the closeness of the user-selected ingredients to the items in the dataset and then recommended recipes accordingly [11].

Another study applied Content-based Filtering as the primary approach for recommending journals and conferences [12]. The system used manuscript abstract data as input and utilized TF-IDF metrics for feature selection. As a result, the system could provide relevant journal and conference recommendations within an average time of about 5 seconds, prioritizing the recommendations based on manuscript abstracts, with an accuracy of 61.37%. Moreover, Content-based Filtering (CBF) has been recognized as one of the most successful recommendation techniques, relying on correlations between content. Content-based Filtering uses item information expressed as attributes to calculate similarities between items [13].

In summary, several studies have analyzed that Content-based Systems have a strong track record and successfully provide recommendations with high levels of accuracy in various contexts [14]. Through this research, we delve into the application of Content-based Filtering for recommending perfumes to users by employing the TF-IDF algorithm [15] and Word2Vec as optimization techniques.

III. Methodology

The perfume product recommender system using a Content-based Filtering approach, as illustrated in Fig. 1, outlines a process that begins with data collection and filtering to ensure that only relevant perfume products and reviews are utilized. Next, sentiment analysis is conducted to understand user preferences for the products. The processed data is then used to build a recommendation model using the TF-IDF method, with reranking based on sentiment and ratings to enhance the quality of recommendations. The evaluation is performed using precision and recall metrics to ensure the system's accuracy and relevance. These stages are designed to provide optimal recommendation results that align with user needs.

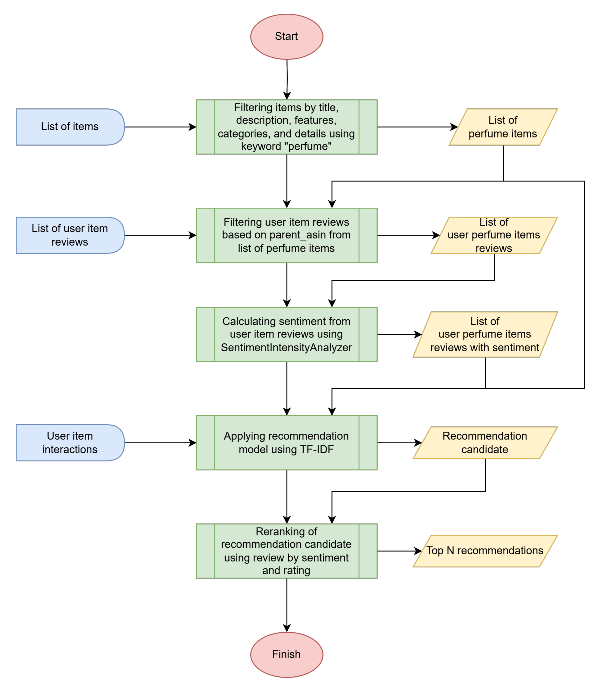

Fig. 1. System design of the perfume Product recommender using content-based filtering approach.

A. Product Filtering by Keywords

This study begins by filtering products based on metadata such as title, description, features, category, and details using keywords such as ”perfume, fragrance, cologne, aroma, eau de toilette, eau de parfum.” This stage aims to identify and select products relevant to the perfume domain. The filtering process involves matching the keywords across various product attributes, resulting in a list of perfume items that will serve as the basis for analysis in subsequent stages.

The dataset used is the Amazon e-commerce product dataset, downloaded from the website amazon-reviews-2023.github.io [16]. This dataset contains information on unique perfume products, which has undergone a meticulous data cleaning process and includes 9 features: brand name, product ID, user ID, review, rating, category, details, description, and features. The dataset was last updated in 2023 and is considered reliable for analyzing consumer preferences.

B. Filtering User Reviews for Perfume Products

Once the list of perfume items is obtained, the next step is to filter user reviews for these products. This filtering is performed using the attribute $parent\_asin$ to link the reviews to the relevant perfume items. This ensures that only reviews directly associated with perfume products are included in the analysis, resulting in a curated list of user reviews for the perfume items.

C. User Review Sentiment Analysis

The third stage involves conducting sentiment analysis on user reviews. For this, tools such as VADER SentimentIntensityAnalyzer are utilized to identify the sentiment within the reviews, categorizing them as positive, negative, or neutral. The sentiment information obtained enriches the user review data, which is later used to support subsequent steps in the recommender system.

D. Implementation of Recommendation Model with TF-IDF

The TF-IDF method is a technique for calculating the weight of each word that is most commonly used in information retrieval. This method is also well-known for its efficiency, simplicity, and accurate results [17]. The technique aims to represent user review texts numerically, enabling the model to match users with relevant products based on their reviews. TF-IDF works by calculating the Term Frequency (TF) and Inverse Document Frequency (IDF) values for each token (word) in every document within the corpus. Term Frequency (TF) represents how often a term appears in a document, while Inverse Document Frequency (IDF) adjusts the term's frequency by considering its occurrence across the entire corpus. This combination helps the method focus on terms that are unique and meaningful within the context of the document. The results from this stage produce an initial list of recommendation candidates generated by the TF-IDF model. This list serves as a foundation for further refinement and ranking in subsequent steps of the recommender system.

E. Reranking Based on Sentiment and Rating

The next stage involves reranking the list of recommendation candidates by considering user review sentiment and ratings. This process is carried out by applying an algorithm that assigns specific weights to these two factors, resulting in a Top-N recommendation list that is more relevant and better aligned with user preferences. This reranking step ensures that the final recommendations not only reflect product similarities but also incorporate user feedback and satisfaction levels, enhancing the overall effectiveness of the recommender system.

IV. Evaluation

The evaluation stage is a step to measure the accuracy of the system in recommending products. This stage assesses how accurate the built system is in providing relevant recommendations. A common method for evaluation is through the use of evaluation metrics [18]. The evaluation metrics applied in this study are precision (1) and recall (2). The classification results are divided into four categories: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These four classification results are utilized to calculate precision and recall, as shown in the following equations [18] [19].

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

To evaluate the effectiveness of the proposed model, several experiments were conducted. First, we used TF-IDF as the filtering algorithm. Second, we employed Word2Vec as a comparison to demonstrate the superiority of the proposed model. The purpose of using both TF-IDF and Word2Vec was to determine whether there is a significant difference between the two methods. To obtain the results of this testing, the evaluation metrics used were precision and recall. These metrics indicate how relevant the recommendations provided to users are. The results of these experiments are presented in Table I.

TABLE I
PERFORMANCE RESULTS OF TF-IDF AND WORD2VEC ON CONTENT-BASED FILTERING

Model	Precision	Recall
TF-IDF	80%	95%
Word2Vec	18%	89%

The results in Table I indicate that the performance of the system built using TF-IDF, based on precision and recall, is superior to Word2Vec. This is evident from the TF-IDF precision value of 80%, meaning the method can accurately identify items truly relevant to users, and the recall value of 95%, demonstrating its ability to find relevant items far exceeding Word2Vec, which only achieves a precision of 18% and a recall of 89%.

For the Word2Vec model, we used the ”gensim.models” module without pretraining, resulting in significantly lower performance compared to TF-IDF. This discrepancy arises because the quality of Word2Vec representations depends heavily on the data, requiring large datasets to achieve optimal results without pretraining. Small or unrepresentative datasets lead to less meaningful embeddings, causing weaker semantic relationships between words. This finding aligns with research by [17], which also reported similar results, indicating that Word2Vec performs poorly with limited data. Word2Vec requires a substantial amount of data to learn word representations effectively and to place similar words in closer proximity. Consequently, Word2Vec struggles to recognize sentiment in datasets with insufficient data [17].

Additionally, there is evidence that this lower performance is influenced by the imbalance in the number of negative, neutral, and positive sentiment instances, with negative sentiment data being a minority. In datasets with limited data, Word2Vec cannot adequately capture the semantic and syntactic information of words. Word2Vec requires large training data to learn word representations. On the other hand, the TF-IDF model achieves good accuracy even with smaller datasets. This is because TF-IDF efficiently identifies the presence or absence of specific keywords within a document. Having established that TF-IDF outperforms Word2Vec, we further compare the performance of rating-based reviews versus sentiment-based reviews. The results of this comparison are presented in Fig. 2.

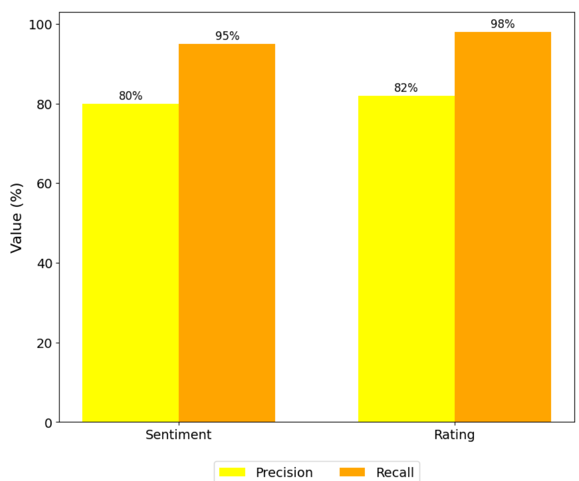

Fig. 2. Comparison of precision and recall between sentiment and rating.

The results in Fig. 2 show that in terms of precision, the value of rating-based reviews is 2% higher. This indicates that rating-based reviews can identify items that are truly relevant to users compared to sentiment-based reviews. Similarly, for recall, rating-based reviews are far superior to sentiment-based reviews, as seen from a difference of 3%, meaning that rating-based reviews can identify 98% of all relevant items. To validate the effectiveness of the TF-IDF method developed based on other user reviews, we evaluated the tests using precision and recall by conducting tests at three levels of Top-N recommendations, namely Top-5, Top-10, and Top-15, for both sentiment-based and rating-based reviews.

In Fig. 3 and Fig. 4, it is shown that the rating-based TF-IDF model outperforms the sentiment-based TF-IDF model in terms of both precision and recall. In the precision graph, the highest value for the rating-based model is achieved at Top-5 (0.8209) and gradually decreases at Top-10 (0.6852) and Top-15 (0.6114). This decline is due to the increasing number of recommendations, which raises the likelihood of including irrelevant items. Meanwhile, the sentiment-based model also shows a declining precision trend from Top-5 (0.8042) to Top-10 (0.6888) and Top-15 (0.5832), but the

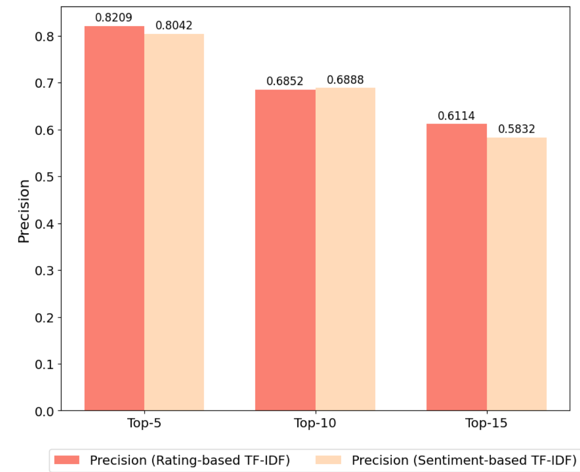

Fig. 3. Precision test results between sentiment and rating.

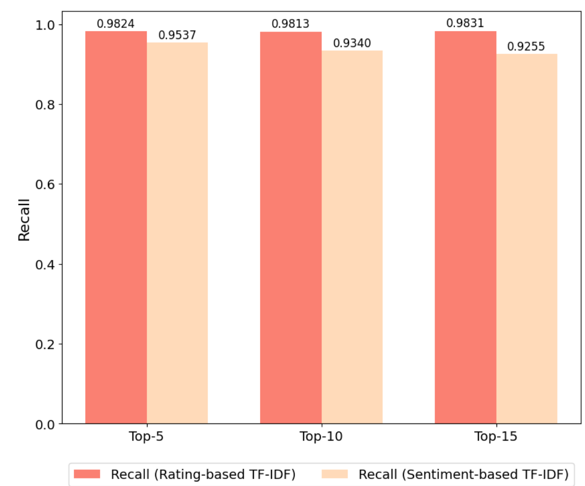

Fig. 4. Recall test results between sentiment and rating.

overall precision of this model is lower than the rating-based model. This indicates that rating data is more effective in generating relevant recommendations.

In the recall graph, the rating-based model exhibits excellent stability with high values across all Top-N: 0.9824 (Top-5), 0.9813 (Top-10), and 0.9831 (Top-15). This indicates that the model can include almost all relevant items in its recommendation list. Conversely, the sentiment-based model also has high recall values but slightly lower than the rating-based model, with scores of 0.9537 (Top-5), 0.9340 (Top-10), and 0.9255 (Top-15). The decline in recall for the sentiment-based model suggests that it is less effective at covering all relevant items, especially in longer recommendation lists.

Overall, the rating-based model is superior in providing relevant recommendations (higher precision) and covering more relevant items (higher recall) compared to the sentiment-based model. This demonstrates that rating data, which provides explicit information about relevance, is more effective than sentiment data. However, the sentiment-based model can serve as a good alternative in scenarios where rating data is unavailable.

V. Conclusion

This study aims to develop a perfume product recommender system using a Content-based Filtering approach optimized with the TF-IDF algorithm. The system is designed to assist consumers in finding perfumes that align with their preferences based on other users' reviews, with performance evaluated using precision and recall metrics. The study's findings show that the TF-IDF algorithm delivers excellent performance with a precision of 80% and a recall of 95%, surpassing the comparison algorithm Word2Vec, which achieved only 18% precision and 89% recall. The superiority of TF-IDF is attributed to its ability to perform well even on small datasets. This method leverages key terms in documents to generate relevant recommendations. In further analysis, the rating-based model demonstrated superior performance compared to the sentiment-based model in both precision and recall. For precision, the rating-based model achieved the highest value of 82.09% at Top-5, gradually decreasing to 61.14% at Top-15. Conversely, the sentiment-based model had lower precision, ranging from 80.42% at Top-5 to 58.32% at Top-15. In terms of recall, the rating-based model showed excellent stability across all Top-N levels (approximately 98%), while the sentiment-based model achieved slightly lower recall, averaging around 93-95%. The superiority of the rating-based model is due to the explicit nature of rating data in providing relevance information, whereas the sentiment-based model relies heavily on text analysis, which is prone to bias and inaccuracies. Nevertheless, the sentiment-based model can serve as a viable alternative when rating data is unavailable.

Overall, this study demonstrates that a Content-based Filtering approach optimized with the TF-IDF algorithm can provide relevant and accurate recommendations for perfume products. This approach helps consumers overcome search biases and simplifies decision-making, thereby enhancing efficiency and user satisfaction in online shopping.

References

[1] K. Kalashi, S. Saed, and B. Teimourpour, "Sentiment-driven community detection in a network of perfume preferences," *arXiv preprint arXiv:2410.19177*, 2024.

[2] J. Kim, K. Oh, and B.-S. Oh, "An nlp-based perfume note estimation based on descriptive sentences," *Applied Sciences*, vol. 14, no. 20, p. 9293, 2024.

[3] B. C. Rodrigues, V. V. Santana, L. P. Queiroz, C. M. Rebello *et al.*, "Harnessing graph neural networks to craft fragrances based on consumer feedback," *Computers & Chemical Engineering*, vol. 185, p. 108674, 2024.

[4] Y. Deldjoo, Z. He, J. McAuley, A. Korikov, S. Sanner, A. Ramisa, R. Vidal, M. Sathiamoorthy, A. Kasirzadeh, and S. Milano, "A review of modern recommender systems using generative models (gen-recsys)," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 6448–6458.

[5] T. A. Manurung and Z. Baizal, "The usage of neural collaborative filtering in enhancing personalized skincare recommendation," in *2024 International Conference on Data Science and Its Applications (ICoDSA)*. IEEE, 2024, pp. 375–379.

[6] F. A. T. Tobing, A. Kusnadi, E. E. Surbakti, W. R. Harsono, A. I. Wijaya, F. Dinarta, M. Irelynn, and R. Paskah, "Men's perfume recommendation system using analytic hierarchy process (ahp) and technique for order of preference by similarity to ideal solution (topsis) method," *International Journal of Science, Technology & Management*, vol. 4, no. 6, pp. 1552–1559, 2023.

[7] Z. A. Baizal, D. Tarwidi, B. Wijaya *et al.*, "Tourism destination recommendation using ontology-based conversational recommender system," *International Journal of Computing and Digital Systems*, vol. 10, 2021.

[8] A. M. Najah and Z. Baizal, "Ontology-based low glycemic index menu recommender system for patients with diabetes," in *2025 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*. IEEE, 2025, pp. 1–6.

[9] G. Lee, X. Jiang, and N. Parde, "A content-based skincare product recommendation system," in *2023 International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2023, pp. 2039–2043.

[10] I. Saifudin and T. Widiyaningtyas, "Systematic literature review on recommender system: Approach, problem, evaluation techniques, datasets," *IEEE Access*, vol. 12, pp. 19 827–19 847, 2024.

[11] S. Bangale, A. Haspe, B. Khemani, and S. Malave, "Recipe recommendation system using content-based filtering," in *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*, 2022.

[12] B. Suvarna and S. Balakrishna, "Enhanced content-based fashion recommendation system through deep ensemble classifier with transfer learning," *Fashion and Textiles*, vol. 11, no. 1, p. 24, 2024.

[13] Q. Liu, N. Chen, T. Sakai, and X.-M. Wu, "Once: Boosting content-based recommendation with both open-and closed-source large language models," in *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, 2024, pp. 452–461.

[14] A. Putra and Z. Baizal, "Laptop recommender system using the hybrid of ontology-based and collaborative filtering," *Sinkron: jurnal dan penelitian teknik informatika*, vol. 8, no. 2, pp. 892–901, 2024.

[15] M. Ridha, M. K. Abdur Rohman, D. Agustin, D. H. Shaputra, Y. Manayla, and A. H. Malikah, "Penerapan machine learning untuk klasifikasi teks depresi pada kesehatan mental dengan svm, tf-idf, dan chi-square," *Journal Software, Hardware and Information Technology*, vol. 5, no. 2, p. 171–182, Jun. 2025. [Online]. Available: https://shift.sin.fst.uin-alauddin.ac.id/index.php/shift/article/view/210

[16] Y. Hou, J. Li, Z. He, A. Yan, X. Chen, and J. McAuley, "Bridging language and items for retrieval and recommendation," *arXiv preprint arXiv:2403.03952*, 2024.

[17] D. E. Cahyani and I. Patasik, "Performance comparison of tf-idf and word2vec models for emotion text classification," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2780–2788, 2021.

[18] Z. A. Baizal and A. T. Wibowo, *Pengantar Sistem Rekomendasi*. Jakarta, Indonesia: Penerbit Erlangga, 2024, 27, 88–89. [Online]. Available: https://openlibrary.telkomuniversity.ac.id/home/catalog/id/218439/slug/pengantar-sistem-rekomendasi.html

[19] M. Ridha, D. Nurjanah, and M. Rakha, "Multilabel classification abusive language and hate speech on indonesian twitter using transformer model: Indobertweet indoroberta," in *2024 International Conference on Intelligent Cybernetics Technology Applications (ICICyTA)*, 2024, pp. 48–54.