

# Utilization Of Discord Bots In Providing Manhwa Recommendations Using Content-Based Filtering Method

Muhammad Farhan Maulana<sup>1</sup>, Ani Dijah Rahajoe<sup>2</sup>, Made Hanindia Prami Swari<sup>3</sup>

<sup>1</sup>Informatika, Fakultas Ilmu Komputer, UPN "Veteran Jawa Timur"

<sup>2,3</sup>Magister Informatika, Fakultas Ilmu Komputer, UPN "Veteran Jawa Timur"

## ABSTRACT

The high number of manhwa released today makes it difficult for readers to find manhwa that match their preferences, especially when trying to find manhwa that are similar to ones they have read before. The manual search process, either through recommendations from communities or online forums, often results in subjective and inconsistent suggestions. To address this issue, a Discord bot was developed that utilizes the Content-Based Filtering method as an automated solution to manhwa recommendation. This method uses the Cosine Similarity algorithm to measure the similarity between manhwa based on features such as title, genre, synopsis, and author. For comparison, the Euclidean Distance algorithm is used to evaluate the accuracy and performance of the recommendation. From the test results, the Cosine Similarity algorithm showed superior performance in providing recommendations based on the questionnaire results and showed a high level of user satisfaction with the developed Discord bot.

**Keywords:** Manhwa Recommendation, Content-Based Filtering, Cosine Similarity.

## INTRODUCTION

Comics have long been an effective visual communication medium to convey information and entertainment to audiences. Along with the development of technology, comics have now transformed into digital formats that are more easily accessible through various online platforms. One type of comic that is very popular today is manhwa, which is a comic created and published from South Korea (Zagita & Sukandar, 2021). Manhwa has attracted the attention of many readers around the world, mainly because of the variety of genres and stories it offers. With thousands of manhwa titles published, readers have many options to browse and enjoy different stories. However, the high number of manhwa available also brings its own challenges, especially for readers who want to discover new titles that suit their preferences. They often struggle to find manhwa that have similar themes or styles to those they have already read. Therefore, there is a need for a system that can help readers find manhwa that match their interests amidst the vast selection.

Discord, as a popular online communication platform, offers excellent features such as real-time communication, audio and video streaming, screen sharing, and ease of developing bots (Vladoiu & Constantinescu, 2020), which makes it a better choice over similar apps such as Skype, Teamspeak, and Slack. Originally designed for the gamer community, Discord has now evolved into a widely used platform for communication and information exchange (Jeevan Joseph et al., 2022). Discord bots are often used for moderation, music players, and running games within community servers (Mrathe et al., 2022). These advantages make Discord a highly versatile digital communication tool that is popular among users.

The research conducted by (Verma et al., 2021) highlights the development of Discord bots aimed at increasing the interactivity and appeal of Discord servers to users. The bot is designed to facilitate communication, share ideas, provide entertainment, and provide information by utilizing data from Reddit. The Discord bot has the ability to communicate with the server, process data from the database, and provide desired results through an API.

Recommendation is something that is done to provide references to other people. Recommendation systems have been widely used in various applications or websites in providing content suggestions that users might be interested in opening. One method used in recommendation systems is Content-Based Filtering. This method operates on the assumption that if a user likes an item with a certain feature, then the user is likely to like other items that have similar features (Kuo & Cheng, 2022). Examples of implementation of this method include in the recommendation of journal articles, movies, and manhwa, where similarity is calculated using algorithms such as Cosine Similarity to measure the similarity between contents.

The goal of this research is to develop the content-based filtering method as a manhwa recommendation system, with the hope of providing recommendations that match readers' desires and increasing the popularity of lesser-known manhwa so that readers can find new manhwa titles to read.

## **LITERATURE REVIEW**

Previous research conducted by (Kuo & Cheng, 2022) discussed the use of Content-Based Filtering method, Cosine Similarity algorithm, and Rocchio algorithm in a product recommendation system on Online Retail dataset from an online retail company in the UK. This research shows how Cosine Similarity is used to measure the relevance between products

and user profiles, while the Rocchio algorithm is used to adjust user profiles based on transaction history and repeat purchase behavior. The main difference with the current research is the focus on the use of Rocchio's algorithm which utilizes user feedback to dynamically refine profiles, while this research emphasizes more on developing a content-based Cosine Similarity algorithm without requiring user feedback.

The next previous research conducted by (Ula et al., 2021) discussed the development of a web-based song recommendation system using the Content-Based Filtering method with TF-IDF and Cosine Similarity algorithms. This research aims to predict and suggest relevant songs to users based on their preferences, with data pre-processing processes that include case folding, stemming, and stopword removal. System testing was conducted through alpha and beta testing, which showed successful results with the alpha test reaching 100%. The main difference with the current study is the medium used, where the previous study used a website, while this study focuses on bot development in the Discord application. However, both developed the Cosine Similarity algorithm in the Content-Based Filtering method to generate appropriate recommendations.

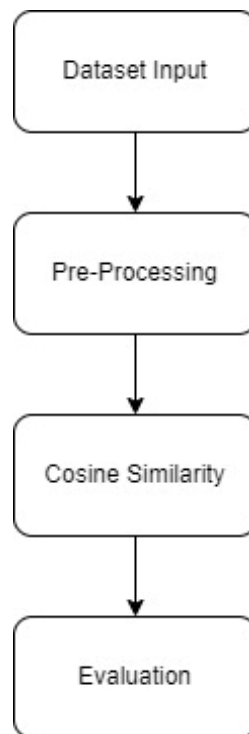
Research by (Isinkaye et al., 2024) used Content-Based Filtering and Deep Learning methods to improve the accuracy of plant disease identification and provide efficient treatment recommendations, using PlantVillage and PlantDoc datasets containing images of healthy and diseased plants. The method relies on analyzing symptoms, environmental conditions, and disease phase to provide appropriate treatment. Although this research focuses on plant disease diagnosis, the principle of using Content-Based Filtering is similar to the research of this thesis, which aims to recommend manhwa based on user preferences by relying on the analysis of content characteristics.

Research by (Mrathe et al., 2022) discusses the application of Machine Learning and Sentiment Analysis techniques on Discord bots to identify emotions in messages and automatically delete inappropriate messages. The system was developed using the Python programming language and run on Linux-based servers such as Raspberry Pi. Tests were conducted on normal messages and negative messages, where detected negative messages will be deleted by the bot as a preventive measure. The main difference with this thesis research is the focus of using Discord bots; Mrathe et al.'s research is focused on message content detection and moderation, while this thesis research aims to provide manhwa recommendations based on user input.

However, there is no research that specifically uses the Content-Based Filtering method to provide manhwa recommendations. This research aims to use the Content-Based Filtering method with the Cosine Similarity algorithm to get manhwa recommendations and is implemented into a bot in the Discord application.

## METHODS

This research uses the Content-Based Filtering method and the Cosine Similarity algorithm to determine the level of similarity of a manhwa. Based on the previously discussed literature review, the author can compile the research methods carried out in this study. The research method is presented in Figure 1.



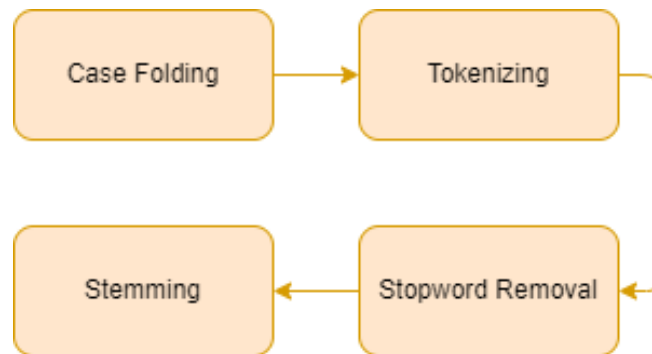
**Figure 1.** Methods

### Data Collection

The dataset used in this research was obtained from Kaggle. The author retrieved data needed in the research such as title, status, genre, synopsis, and author. The amount of data taken amounted to more than 2800 manhwa datasets that were last updated in 2023.

## Pre-Processing

All data from the dataset will be pre-processed before entering the next step. The steps of pre-processing is presented in Figure 2.



**Figure 2.** Pre-Processing Steps

First, case folding is performed to convert all uppercase letters into lowercase letters and remove unnecessary symbols, leaving only alphabetic characters. Next, a tokenizing process is performed to break the text into word-by-word arrays. After that, stopwords removal is used to remove words that are common and have no significant meaning, such as “and”, “in”, or “which”. Finally, the stemming stage is applied to return words to their basic form by removing both prefixes and suffixes, so words such as “reading” and “readed” will be converted to “read”. All these steps aim to optimize the data to make it more relevant and ready to be used in the recommendation algorithm analysis process.

## Cosine Similarity

The Cosine Similarity algorithm is used to measure the degree of similarity between two documents. The algorithm works by calculating the cosine angle between two vectors in a multidimensional space representing the documents. Cosine Similarity values range from -1 to 1, where a value of 1 indicates perfect similarity, 0 indicates no similarity, and -1 indicates opposite similarity (Shkhanukova, 2023). The following is the formula for calculating Cosine Similarity.

$$sim(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

$A \cdot B$  is the dot product between vectors  $A$  and  $B$ . Here's the formula to calculate the multiplication of the vectors.

$$\mathbf{A} \cdot \mathbf{B} = \sum_{i=1}^n A_i B_i = A_1 B_1 + A_2 B_2 + \dots + A_n B_n$$

Whereas  $|\mathbf{A}|$  dan  $|\mathbf{B}|$  describe the vector quantities of the vector and also the vector  $\mathbf{A}$  and also the vector  $\mathbf{B}$ .

$$|\mathbf{A}| = \sqrt{A_1^2 + A_2^2 + \dots + A_n^2}$$

$$|\mathbf{B}| = \sqrt{B_1^2 + B_2^2 + \dots + B_n^2}$$

### Euclidean Distance

The Euclidean Distance algorithm is used as a comparison algorithm to measure the distance between two manhwa documents based on certain features. Euclidean Distance calculates the shortest straight (linear) distance between two points in a multidimensional space representing the documents. The smaller the Euclidean Distance value, the more similar the two documents are. Here is the basic formula to calculate the Euclidean Distance between two vectors  $\mathbf{A}$  and  $\mathbf{B}$  in  $n$ -dimensional space.

$$d(\mathbf{A}, \mathbf{B}) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}$$

$A_i$  dan  $B_i$  are the components of vectors  $\mathbf{A}$  and  $\mathbf{B}$ , respectively. In other words, Euclidean Distance calculates the square root of the sum of the squares of the differences between each pair of elements of the two vectors. In the context of manhwa recommendation, this algorithm helps determine how similar or different two manhwa are based on the attributes analyzed, and the results are compared with the Cosine Similarity algorithm to evaluate which one provides more accurate recommendations.

### Evaluation

The evaluation was conducted to determine the performance of the Discord bot and compare the results provided by Cosine Similarity and Euclidean Distance. The evaluation is done by giving questionnaires to the respondents. There are 29 respondents who are manhwa readers from a community who have given their assessment.

## RESULTS AND DISCUSSION

### Pre-Processing

Before starting pre-processing, the title, genres, synopsis, and authors data are combined into one. The first pre-processing is case folding, using code syntax manually. After that, the data goes through a tokenizing stage to break up the words and convert them into arrays. Then, the data is processed in the stopwords removing stage using the “stopword-json” library which contains stopwords that must be removed. The last stage is stemming, which removes the prefix and suffix of words using the “PorterStemmer” function from the “natural” library. Table 1 is an example of the results of the pre-processing that has been carried out.

**Table 1.** Pre-Processing Result

<pre>[   {     "id": 1,     "title": "Solo Leveling",     "array_data": ["solo", "level", "finish", "action", "adventure", "fantasy", "ten", "year", "ago", "gate", "appear", "connect", "real", "world", "realm", "magic", "monster", "combat", "vile", "beast", "ordinary", "people", "receive", "superhuman", "power", "hunter", "twenty", "year", "sung", "jin", "woo", "hunter", "world", "weakest", "owe", "pathetic", "power", "compare", "measly", "rank", "hunt", "monster", "tirelessly", "low", "rank", "gate", "pay", "mother", "medic", "bill", "misery", "lifestyle", "jin", "woo", "believe", "left", "die", "mission", "terrible", "wrong", "awaken", "hospital", "day", "find", "mystery", "screen", "float", "front", "quest", "log", "demand", "jin", "woo", "complete", "unreal", "intens", "train", "program", "face", "penalty", "initial", "reluctant", "comply", "quest", "rigor", "jin", "woo", "find", "transform", "world", "fearsome", "hunter", "written", "mal", "rewrite", "chugong", "story", "jang", "sung", "rak", "art", "disciple", "art"]   }, ]</pre>
---

### Cosine Similarity Calculation

The Cosine Similarity algorithm is used in this study to provide manhwa recommendations on a Discord bot by measuring the similarity between a manhwa that a user has read and other manhwa in the dataset based on features such as title, synopsis, genre, and author. Utilizing a vector-based approach, the algorithm calculates the cosine value of the angle between two vectors representing the manhwa. In the context of a recommendation bot, after a user enters the title of a manhwa they like, the bot calculates the Cosine Similarity between the user's input manhwa vector and all manhwa vectors in the dataset. The manhwa that has the highest Cosine Similarity value will be recommended to the user, as it shows significant content similarity with the user's preferences. Figure 3 is a visualization of the recommendation results using Cosine Similarity.



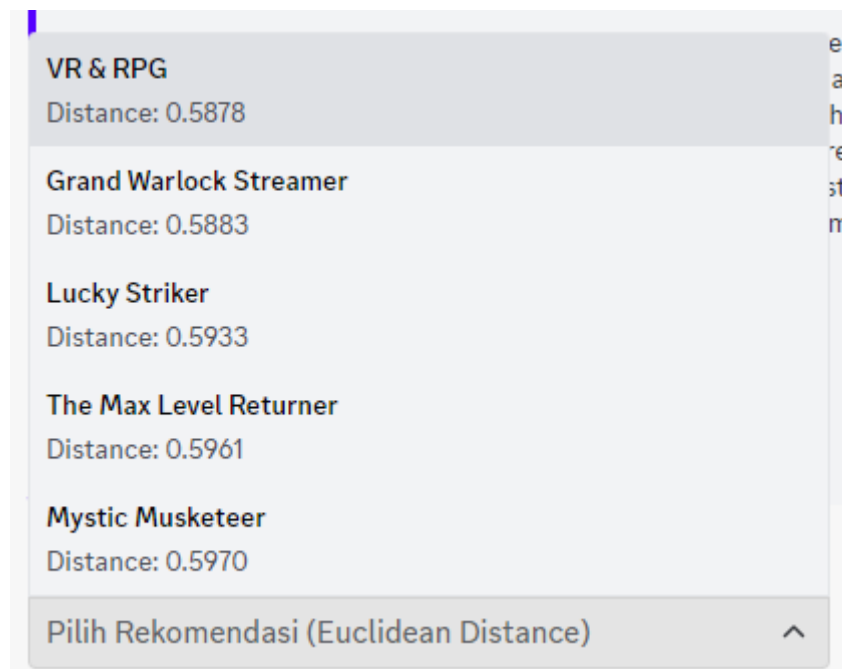
**Figure 3.** Cosine Similarity Recommendation Results

### Euclidean Distance Calculation

The Euclidean Distance algorithm is used as a comparison algorithm to provide manhwa recommendations via Discord bot. This algorithm measures the linear distance between two vectors representing manhwa based on features such as title, synopsis, genre, and author. In the context of recommendation, Euclidean Distance calculates how far or close two manhwa are to each other by summing the squares of the difference of each element of the vector, then taking the square root of the result. The smaller the Euclidean Distance value between two



vectors, the more similar the two manhwa are. Therefore, when a user provides the title of a manhwa they like, the bot uses this algorithm to calculate the distance between that manhwa and other manhwa in the dataset. The manhwa with the shortest distance will be considered the most similar and recommended to the user. The result of Euclidean Distance is then compared with the result of Cosine Similarity to evaluate the effectiveness of each algorithm in producing accurate recommendations that match the user's preferences. Figure 4 is a visualization of the recommendation results using Euclidean Distance.



**Figure 3.** Cosine Similarity Recommendation Results

### Model Evaluation

Table 1 shows the questions that will be asked to the respondents to get an assessment of the performance of the bot system and the two algorithms used. The first question asks for the total number of manhwa that the respondents have read, aiming to understand their level of experience and interest in reading manhwa. The second question asked respondents to name the manhwa titles they had used to get recommendations, which served to determine the extent to which users had interacted with the bot. The third question evaluates the adequacy of the number of recommendations provided by the bot, asking if displaying five recommendations is enough or if the user wants more, which is useful for customizing the bot's output to the user's preferences. The last question asked respondents to rate which algorithm is better between Cosine Similarity and Euclidean Distance, based on their experience, with the aim of

determining the most effective algorithm in providing satisfactory recommendations. The questionnaire as a whole is designed to collect direct feedback from users, which will be used to refine and improve the performance of the manhwa recommendation bot.

**Table 1. Model Evaluation**

No.	Questions
1.	Berapa banyak total manhwa yang sudah pernah anda baca sebelumnya?
2.	Judul manhwa apa saja yang sudah anda coba untuk mendapatkan rekomendasi?
3.	Apakah menurut anda menampilkan 5 jumlah hasil yang direkomendasikan oleh kedua algoritma dalam bot sudah cukup? Jika kurang, maka berapakah yang harus ditampilkan menurut anda?
4.	Setelah menggunakan bot Discord tersebut, manakah di antara kedua algoritma yang digunakan yang lebih baik menurut anda?

The results of the first question in the questionnaire showed that the majority of respondents had read a large number of manhwa, with many of them claiming to have read more than 100 titles. This shows that the users who filled out the questionnaire are active and experienced readers of manhwa, which strengthens the validity of the data as they have a solid basis for providing valuable feedback. From the second question, it was revealed that users often seek recommendations for manhwa that have similar genres or stories to popular titles such as “Solo Leveling” and “Tower of God.” This information shows the pattern of users' interest in certain types of stories, and helps in customizing the recommendation algorithm to focus more on frequently searched genres or themes.

The results of the third question showed that over 90% of respondents felt that displaying five manhwa recommendations was sufficient, although there were about 7% who wanted more options, especially when the recommendations provided did not fully match their preferences. This provides an important insight into the balance between the quality and quantity of recommendations that users want. Finally, from the fourth question, 82% of respondents preferred the Cosine Similarity algorithm over the Euclidean Distance, arguing that the recommendations provided by Cosine Similarity were more accurate and in line with their

preferences. These results indicate that Cosine Similarity is more effective in capturing content similarity based on manhwa features, and is therefore preferred as the primary method for providing recommendations.

## CONCLUSION

The conclusion of this study shows that the use of Discord bot as a tool to provide manhwa recommendations with the Content-Based Filtering method, which utilizes the Cosine Similarity algorithm, is an effective and efficient approach. The bot is able to provide recommendations that match the user's preferences based on the similarity of features between the manhwa that the user has read and other manhwa in the dataset. The Cosine Similarity algorithm proved superior to the Euclidean Distance comparison algorithm in providing more relevant recommendations, as shown by the questionnaire results which stated that 75% of users were satisfied with the recommendations generated by this algorithm.

In addition, this study successfully demonstrated that active manhwa readers, who filled out the questionnaire, have specific needs in finding titles similar to their previous preferences. With the majority of respondents being satisfied with five recommendations, this study confirms the importance of providing recommendations that are not only quality, but also fit the needs of a reasonable quantity. This shows that users value recommendation systems that can provide appropriate and not overwhelming suggestions, which can enhance their reading experience.

Overall, this research makes a significant contribution to the field of recommendation systems, especially in the use of Discord bots for this purpose. The results of this research can serve as a basis for further development in optimizing the Content-Based Filtering method and implementing recommendation algorithms in other digital communication platforms. The developed bot is expected to be continuously improved and adapted to the development of user needs, and can be used as a model for other recommendation applications in the future.

## REFERENCES

- Isinkaye, F. O., Olusanya, M. O., & Singh, P. K. (2024). Deep learning and content-based filtering techniques for improving plant disease identification and treatment recommendations: A comprehensive review. *Heliyon*, 10(9).  
<https://doi.org/10.1016/j.heliyon.2024.e29583>

- Jeevan Joseph, Akshara Anilkumar, Treesa Thomas, & Binny S. (2022). Discord : An all in one messaging application (Case Study). *International Journal of Engineering Technology and Management Sciences*, 6(5), 437–440.  
<https://doi.org/10.46647/ijetms.2022.v06i05.067>
- Kuo, R. J., & Cheng, H. R. (2022). A content-based recommender system with consideration of repeat purchase behavior. *Applied Soft Computing*, 127.  
<https://doi.org/10.1016/j.asoc.2022.109361>
- Mrathe, P. . Dr. A., Ingawle, K., Bagav, S., Gaikwad, S., More, O., & Hiwale, P. (2022). Safe Discord Bot Using ML. *International Journal of Scientific Research in Engineering and Management (IJSREM)*, 06(11), 1–3. <https://doi.org/10.55041/ijsrem16797>
- Shkhanukova, M. (2023). *Cosine distance and cosine similarity* | by Milana Shkhanukova | *Medium*. Retrieved July 19, 2024, <https://medium.com/@milana.shkhanukova15/cosine-distance-and-cosine-similarity-a5da0e4d9ded>
- Ula, N., Setianingsih, C., & Nugrahaeni, R. A. (2021). SISTEM REKOMENDASI LAGU DENGAN METODE CONTENT-BASED FILTERING BERBASIS WEBSITE WEB-BASED SONG RECOMMENDATION SYSTEM USING CONTENT-BASED FILTERING. *E-Proceeding of Engineering*, 8(6), 12193–12199.
- Verma, A., Tyagi, S., & Mathur, G. (2021). A Comprehensive Review on Bot - Discord Bot. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 532–536. <https://doi.org/10.32628/cseit2172100>
- Vladoiu, M., & Constantinescu, Z. (2020). Learning During COVID-19 Pandemic: Online Education Community, Based on Discord. *2020 19th RoEduNet Conference: Networking in Education and Research (RoEduNet), 2020-December*, 1–6.  
<https://doi.org/10.1109/RoEduNet51892.2020.9324863>
- Zagita, N. I., & Sukandar, R. (2021). Pandangan Masyarakat Indonesia Terhadap Budaya Korea Selatan: Studi Kasus Manhwa Noblesse pada Aplikasi Line Webtoon. *COMMENTATE: Journal of Communication Management*, 2(1), 78–93.  
<https://doi.org/10.37535/103002120216>

