

Measuring Tourist Experience in Semarang City through an Advanced Recommendation System

Rudi Sutomo¹(✉), Daffa Kaisha Pratama¹

¹ Information System Department, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia

rudi.sutomo@umn.ac.id, daffa.kaisha@student.umn.ac.id

Article Info

Article history:

Received Oktober 26, 2023

Revised November 2, 2023

Accepted November 24, 2023

Keywords:

Content-Based Filtering
Collaborative Filtering
System Recommendation
Tourism

ABSTRACT

In the tourism sector which includes recreation and holiday activities, the Indonesian tourism sector has a very important role because of its impact on the country's foreign exchange reserves. Indonesia, with its diverse attractions ranging from nature, culture, religion, family activities, shopping and gastronomy, presents many choices for tourists. The city of Semarang is actively increasing its tourism offerings, but the sheer number of choices can overwhelm tourists. This research presents an advanced recommendation system based on collaborative filtering and content-based filtering techniques. By leveraging historical travel data, including attraction visits, ratings, and frequently visited categories, the system provides tailored suggestions. Content-based filtering prioritizes tourist attractions such as Chinatown Semarang, Kampoeng Djadhoel Semarang, Kapal Mosque Semarang, Tugu Muda Semarang, and Tinjomoyo Forest Tourism Semarang based on ratings. Collaborative filtering resulted in recommendations such as Gua Maria Kerep Ambarawa (rating: 4.8), La Kana Chapel (rating: 4.5), Palagan Ambarawa Monument (rating: 4.4), Eling Bening Tourism (rating: 4.3), and Kampoeng Kopi Banaran (rating: 4.3). In a world where choices are many and time is limited, this advanced recommendation system simplifies travel decisions, elevating ordinary trips into extraordinary adventures. This heralds the future of tourism, where technology aligns with exploration to uncover Semarang's hidden treasures.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Rudi Sutomo

Information System, Universitas Multimedia Nusantara, Tangerang, Indonesia

Email: rudi.sutomo@umn.ac.id

1. INTRODUCTION

Data from the UN-WTO reveals promising global trends in tourism. In 1950, the United States welcomed 25 million international tourists, a number that surged to 1.087 billion worldwide by 2014. If this growth trajectory, with an average annual increase of 6.5%, is maintained, projections indicate that by 2020, global tourist arrivals will reach 1.6 billion, and by 2030, the figure is expected to soar to 1.8 billion [1].

Central Java is one of the provinces in Java Island, located on the crossing route between West Java and East Java. Many tourists often skip Central Java because it is only a crossing area. Semarang City has various tourist objects such as nature tourism, cultural tourism, religious tourism, family tourism, shopping, and culinary tourism. Semarang City's economic activities, specifically those connected with tourism, are anticipated to expand as the number of visitors increases, helping to improve regional and local income. The range of tourist attractions, including cultural, religious, historical, gastronomic, and natural, is the backbone of efforts to grow tourism in the Semarang. Complete infrastructure, including airports, train stations, terminals, and toll roadways, enhance this by enabling simple access for tourists to Semarang City[2][3].

The number of tourists in Semarang City has decreased from year to year. In 2018, the number of tourists was 66,107 people. The number of tourists decreased by 33.42% compared to 2017, which amounted to 99,282 people. The table above shows that in 2017, the number of tourist visits was 99,282 people, experiencing a decrease of 2.43% from the number of visits of the previous year in 2016 amounted to 101,756 people[3].

Based on the problems above, to help improve the development of the tourism sector, especially the city of Semarang, it is necessary to analyze recommendations based on existing data and process them so that visualization in graphic form and ranking order of recommendations based on the Content-Based Filtering and Collaborative Filtering methods in Semarang City[4].

2. LITERATURE REVIEW

This study uses two different approaches to analyze recommendations: the first is content-based filtering, which relies on product or service attributes and assigns them to specific categories. Collaborative filtering, on the other hand, compares a user's choices and behavior with the intentions of other users who have the same user preferences or ratings. This literature offers valuable insights into the exploration of tourist destination preferences in the context of recommendation systems. Several studies have investigated the effectiveness of these two approaches, and highlighted their strengths and weaknesses. For example, the other research conducted research on content-based recommendation systems, emphasizing their ability to provide personalized suggestions by analyzing destination attributes and aligning them with user preferences [5].

However, a prominent gap in the literature is the application of this approach specifically to the tourism context of Semarang City. Although there is a lot of information about recommendation systems in various fields, research regarding their application in the Semarang City tourism sector is still limited. This research aims to address this gap by offering a detailed analysis of content-based and collaborative filtering techniques in the context of the tourist destination of Semarang City, thereby contributing to the advancement of knowledge in this field [6].

2.1 Content-Based Filtering

To identify which items are most relevant to each user, content-based filtering contrasts representations of the content of objects (documents) with models of readers' interests. Finding the optimum expression for both the items (item profile) and the users (user profile) becomes an issue due to file maps mapping of the actual user's interests to a simplified model space that comes close to representing the user's genuine interests. To facilitate matching between the profiles, a user's profile, and an item's profile should utilize the exact representation technique using the same representation method representation (for instance, model by keywords). A rating score is the result of the matching process, and it represents how well an item's profile matches the user's user[7].

2.2 Collaborative Filtering

The degree of similarity between users is a factor in collaborative techniques. Finding a group or collection of user X whose preferences, likes, and dislikes are similar to user A's of user A is the first step in this strategy. The area around A is X. User A is then given recommendations for the new goods in X that are popular with most users. The effectiveness of a collaborative algorithm depends on how precisely it can locate the surrounding area of the target user. As user data must be shared, traditional collaborative filtering-based systems need help with the cold-start and privacy issues. However, to produce a recommendation using collaborative filtering techniques, item feature knowledge is unnecessary to create an offer using collaborative filtering techniques, item feature knowledge is unnecessary[8].

The user chooses the type of collaborative filtering of collaborative filtering selected, the utility of product use (Item-Based), and the degree of similarity of user encouragement (User-Based). The Collaborative Filtering approach, which is used in the tourism sector, and d recommendations for the locations of tourist sites [9].

2.3 Dataset

A dataset is a group of data organized in a specific manner. Datasets are often displayed as tables with rows and columns. Typically, each row and column represents a different variable. A method is needed to make sure that the dataset that has been gathered has a proper category label before data can be used to create a dataset for the classification system. There are significantly few labeled datasets, and finding them might be complicated [10].

It was finding the machine learning method that would deliver high accuracy for a critical data type critical. Since we recognize that different machine learning algorithms identify text data differently, this comparative analysis will examine the effectiveness of various machine learning algorithms before determining which algorithm is better for which type of data. Because of this, it's essential to know which method to use with a certain kind of dataset [11].

The data source was obtained from Kaggle with the name Indonesian Tourism Destination data and the data owner is Mr. A Prabowo. The updated data is 21 July 2021 and in the data that has been obtained there are only a few choices of tourist cities and the data has not yet been processed so this is something that needs

to be studied using the recommender system method in order to present results or models that can later help increase tourism in Indonesia [[12].

3. METHOD

Users could receive recommended items or services using content-based or collaborative filtering, two separate strategies. Content-based filtering is used based on the features or traits of a good or service, the characteristics or qualities of a good or service; content-based content-based filtering is used. For instance, in movies, content-based filtering can suggest related films based on the genre, leading actor, director, or movie. Content-based content-based filtering can present related films based on the genre, the leading actor, the director, or the plot [4].

Meanwhile, collaborative filtering uses user preferences and behavior to find similarities with other users with similar intentions. For example, if users A and B have the same choices for a particular movie, then the movie that user B likes is likely also to be selected by user A.

Thus, the primary difference between content-based and collaborative filtering is that the former places more importance on the attributes of goods or services. At the same time, the latter approach uses information about user behavior [13].

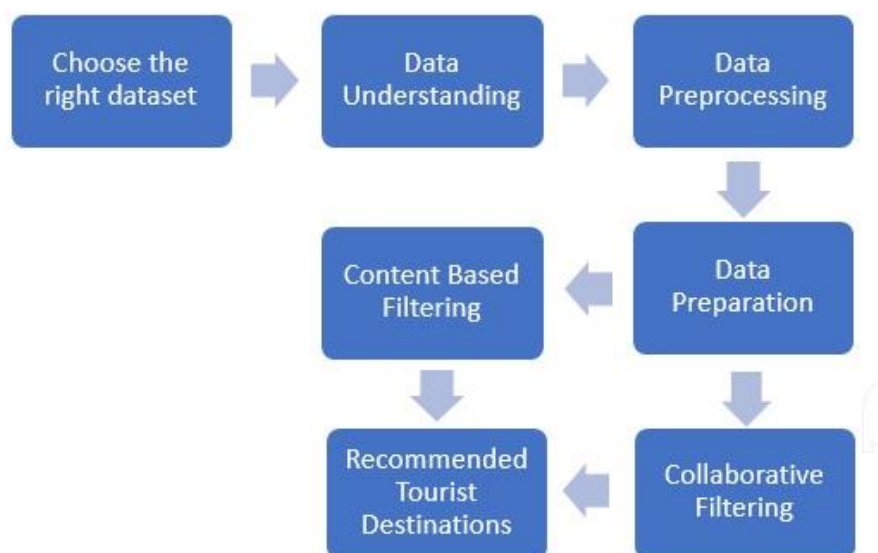


Figure 1. Proposed Method

3.1 Data Understanding

In using the dataset, several cities have data categories; for this study, the focus is more on the city of Semarang. The data used, namely `package_tourism`, is used to retrieve city data and tourist destinations. Then use the `tourism_rating` data to measure visitor satisfaction ratings. So that all data can be assessed, `tourism_with_id` data is used, and finally, to find out visitor data, user data is used [14].

3.2 Data Processing

Determining only specific data that will be used for the recommendation system, `tourism_with_id` is determined to be a place variable and `tourism_rating` to be a rating variable and used to be a user variable. Five features. Some features have been removed because they are not used [15].

3.3 Data Preparation

The data is checked to determine whether there is a null in the core data. Then much data is contained in the core data; information is obtained that the current amount of data is 57 in Semarang. After that, a new preparation variable is created, which contains data from the core data, and duplicate values are removed in the `Place_Id` feature. Then incorporate these new variables into tourist destinations. Data preparation aims to eliminate the same data that can cause bias and accommodate features that will only be used for recommendations [16].

3.4 Model Content-Based Filtering

The recommendation method, "content-based filtering," suggests to users products that are similar based on the characteristics of a product or service. This approach is predicated on the concept that customers are more likely to be satisfied with products that share the same qualities or characteristics as products they have in the past appreciated [17].

A cosine similarity function is the method used to determine how similar two item characteristics are. A similarity rating of 1 between two profile components indicates that two tourist attractions are identical. When both profile elements have a similarity value of 0, it is stated that the two establishments are distinct from one another. The higher the value, the two tourist destinations under consideration are considered comparable in their cosine similarity function results from tourist destinations under consideration are considered more equal in their cosine similarity function results from the higher the value, and vice versa [18].

An approach based on content will be used for this research. Using user profiles and item profiles that are similar, the system will choose and rank items. Because each item's content can be deduced from its representation, this approach allows users to understand why certain things are deemed necessary to them. The fact that this approach is keyword-focused and has downsides is just one of them. Based on different attribute types connected with structured objects from text, this approach cannot capture more complicated relationships at a deeper semantic level. The principle of proximity stipulates that the distance between the two descriptions of the object and the representation of the user is the basis for comparing the two their similarity [19].

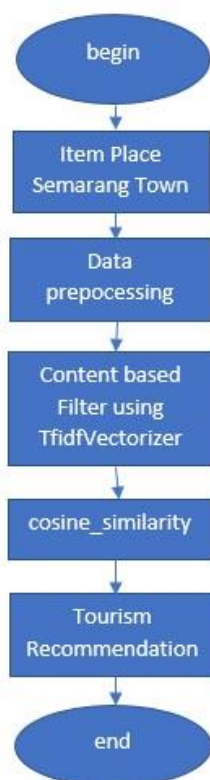


Figure 2. Content-Based Filtering Method

3.4 Model Collaborative Learning

Based on the concept of recommendation marketing, collaborative filtering assumes that a person's decision-making process heavily weighs the advice of friends and relatives. When dealing with people online, nearest neighbors—users who share the same preferences as the current user or have similar purchasing behaviors—replace family members and friends [19]. A collection of users and a set of things placed serve as the two separate forms of background data that collaborative filtering (see Figure 3). Users' ratings of products and other users are the primary means through which users engage with one another and items. These ratings are used to forecast the rating a user would give an article in the future during suggestion sessions. Suppose we assume that the place interacts with a collaborative filtering recommendation system. In that case, the recommendation system's first step is to locate the closest neighbors (users with rating behaviors similar to User) and extrapolate the rating of User from the ratings of the comparable users [20][21].



Figure 3. Collaborative Filtering Method

3.5 Recommended Tourist Destinations

This study will present recommendations from data processing results with two approaches to the recommendation system method so that the recommendation results are more accurate and on target. Several libraries support suggestions for content-based filtering using Python programming, so the results are shown in the following figure 4, below this.

	place_name	place_category
0	Semarang Chinatown	Budaya
1	Kampoeng Djadhoel Semarang	Budaya
2	Masjid Kapal Semarang	Tempat Ibadah
3	Tugu Muda Semarang	Budaya
4	Hutan Wisata Tinjomoyo Semarang	Cagar Alam

Figure 4. Content-Based Filtering Recommendation

Recommendations for collaborative filtering using Python programming and supported by several libraries The results are shown in the following figure.5. below this.

```
List of recommendations for: User 209
=====

-----
Places with the highest tourist ratings from users
-----
Masjid Kapal Semarang : Tempat Ibadah

Water Blaster Bukit Candi Golf : Taman Hiburan

Hutan Pinus Kayon : Cagar Alam

Tugu Muda Semarang : Budaya

Air Terjun Semarang : Cagar Alam

-----
```

Figure 5. Collaborative Filtering Recommendation Rating by Users

Based on Travel Rating, it uses the collaborative filtering recommendation method using Python programming. It is supported by several libraries, one of which is the tensor flow tensor, so the results are shown in the following figure. below this

```
-----
Top 5 place recommendations in Semarang City:
-----
1 . Waduk Jatibarang
   Cagar Alam , Entrance Ticket Price 2500 , Travel Rating 4.3

2 . Monumen Palagan Ambarawa
   Budaya , Entrance Ticket Price 7500 , Travel Rating 4.4

3 . Gua Maria Kerep Ambarawa
   Cagar Alam , Entrance Ticket Price 2000 , Travel Rating 4.8

4 . La Kana Chapel
   Taman Hiburan , Entrance Ticket Price 35000 , Travel Rating 4.5

5 . Kampong Kopi Banaran
   Taman Hiburan , Entrance Ticket Price 200000 , Travel Rating 4.3

=====
```

Figure 6. Collaborative Filtering Recommendation Rating by Travel Ratings

Thus the results of the two system recommendation approaches from the two complementary results seen from the user side and the rating of tourist visit locations so the data looks accurate and can be a consideration for tourists visiting the city of Semarang.

4. RESULT AND DISCUSSION

Experiments were carried out on a dataset from Kaggle, which describes several tourist destinations in Indonesia, especially Semarang.

4.1 Result in Content-Based-Filtering

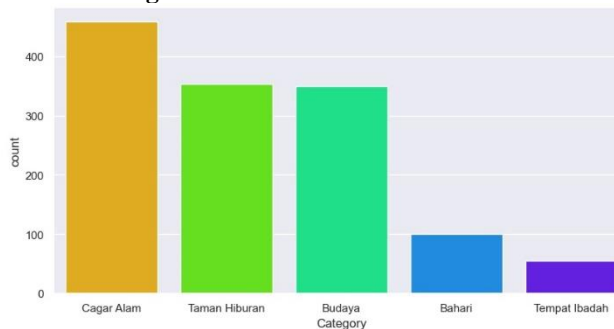


Figure 7. Content-Based Filtering Recommendation Count by Category

It can be seen in the visualization above; each rating has an almost similar percentage. The following is the percentage and number of specific samples from the visualization above, sorted from the most to the lowest:

- a. Rating 2: 303 sample, 23.0%
- b. Rating 4 : 270 sample, 20.5%
- c. Rating 5 : 262 sample, 19.9%
- d. Rating 3 : 259 sample, 19.7%
- e. Rating 1 : 223 sample, 16.9%

It can be seen that 35% of the total data is in the Nature Reserve Category and then the Amusement Park; the following is the percentage and number of specific samples from Category features sorted from most significant to lowest:

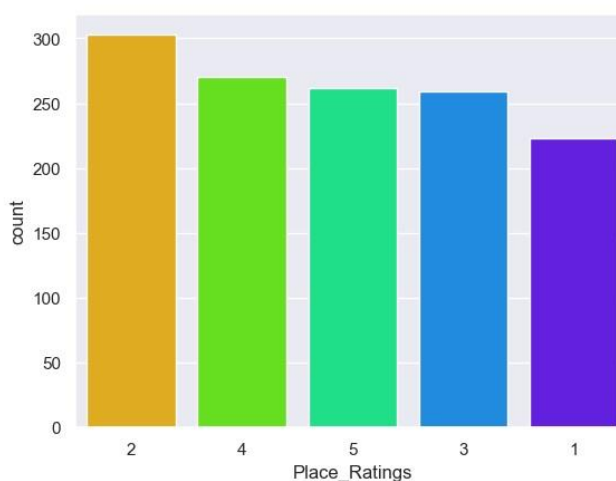


Figure 8. Collaborative Filtering Recommendation Count by Place Ratings

- a. Places of Worship: 385 samples, 3.8%
- b. Nature Reserve: 459 samples, 34.9%
- c. Amusement Parks: 353 samples, 26.8%
- d. Culture : 350 samples, 26.6%
- e. Marine : 100 samples, 7.6%
- f. Places of Worship : 55 samples, 4.2%

This recommendation system displays recommendations based on tourist attractions visited by tourists or previous tourists so that they show recommendations similar to previous tours, as seen in Figures 7 and 8 above.

4.2 Result in Collaborative Learning

This recommendation system displays recommendations based on ratings where collaborative filtering aims to recommend tourist destinations even though tourists or tourists have never been to a tourist destination before. And the result is five recommendations for tourist attractions with the first order to the end, namely: Maria Kerep Ambarawa Cave with a rating of 4.8, La Kana Chapel with a rating of 4.5, Palagan Ambarawa Monument with a rating of 4.4, Eling Bening Tourism with a rating of 4.3, Kampoeng Kopi Banaran with a rating of 4.3.

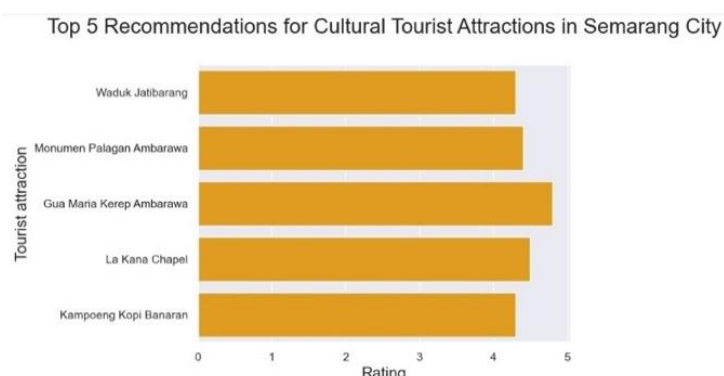


Figure 9. Collaborative Filtering Recommendation Tourist Destination by Rating

5. CONCLUSION

In conclusion, the results obtained from the experimental phase of the recommendation system in 2023 underscore its significant contribution to enhancing tourist experiences in Semarang City. The system effectively provides tailored suggestions based on ratings, utilizing both content-based filtering and collaborative filtering methodologies. For content-based filtering, the top-rated destinations include Semarang Chinatown, Kampong Djadhoel Semarang, Masjid Kapal Semarang, Tugu Muda Semarang, and Hutan Wisata Tinjomoyo Semarang. On the other hand, collaborative filtering prioritizes Maria Kerep Ambarawa Cave with a remarkable rating of 4.8, followed by La Kana Chapel (4.5), Palagan Ambarawa Monument (4.4), Eling Bening Tourism (4.3), and Kampong Kopi Banaran (4.3). These findings demonstrate the system's efficacy in offering valuable recommendations, ultimately enhancing the tourist experience and contributing to the tourism sector's growth and development in Semarang City.

ACKNOWLEDGEMENTS

Multimedia Nusantara University provided financial support for this research. We thank you for your support and assistance during this time.

REFERENCES

- [1] M. T. Astuti, "Branding Strategy for Tourism Destination in Semarang City," *Pros. ICSMR*, pp. 1–20, 2020, [Online]. Available: <http://conference.loupiasconference.org/index.php/ICSMR/article/view/1>
- [2] M. Damayanti, H. Wahyono, M. Rahdriawan, W. P. Tyas, P. C. Sani, and J. Riptek, "Penerapan Smart Tourism Di Kota Semarang," *J. Riptek*, vol. 14, no. 2, pp. 128–133, 2020, [Online]. Available: <http://ripteck.semarangkota.go.id>
- [3] R. Ginanjar, S. Dian, and W. Prajanti, "Economics Development Analysis Journal Development Strategies for Tourism Destinations in Semarang Old Town Article Info," *Econ. Dev. Anal. J.*, vol. 10, no. 1, pp. 105–122, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/edaj>
- [4] M. nazim uddin, J. Shrestha, and G. S. Jo, *Enhanced Content-Based Filtering Using Diverse Collaborative Prediction for Movie Recommendation*. 2009. doi: 10.1109/ACIIDS.2009.77.
- [5] J. C. John and C. Limited, "Intelligent Travel Recommendation Systems for Transforming Nigeria ' s Intelligent Travel Recommendation Systems for Transforming Nigeria ' s Tourism," no. September, 2023.
- [6] R. Glauber and A. Loula, *Collaborative Filtering vs. Content-Based Filtering: differences and similarities*. 2019.
- [7] V. Maidel, P. Shoval, B. Shapira, and M. Taieb-Maimon, "Evaluation of an ontology-content based filtering method for a personalized newspaper," *RecSys '08 Proc. 2008 ACM Conf. Recomm. Syst.*, pp. 91–98, 2008, doi: 10.1145/1454008.1454024.
- [8] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00592-5.
- [9] M. Hikmatyar and Ruuhwan, "Book Recommendation System Development Using User-Based Collaborative Filtering," *J. Phys. Conf. Ser.*, vol. 1477, no. 3, 2020, doi: 10.1088/1742-6596/1477/3/032024.
- [10] E. T. Matsubara, M. C. Monard, and R. C. Prati, "Exploring unclassified texts using multiview semisupervised learning," *Emerg. Technol. Text Min. Tech. Appl.*, pp. 139–161, 2007, doi: 10.4018/978-1-59904-373-9.ch007.
- [11] S. U. Hassan, J. Ahamed, and K. Ahmad, "Analytics of machine learning-based algorithms for text classification," *Sustain. Oper. Comput.*, vol. 3, no. March, pp. 238–248, 2022, doi: 10.1016/j.susoc.2022.03.001.
- [12] A. Prabowo, "Indonesia Tourism Destination," 2021. <https://www.kaggle.com/datasets/aprabowo/indonesia-tourism-destination> (accessed Mar. 21, 2022).
- [13] C. Rana and J. Shokeen, "A review on the dynamics of social recommender systems A review on the dynamics of social recommender systems Jyoti Shokeen * and Chhavi Rana," no. January, 2018, doi: 10.1504/IJWET.2018.10016164.
- [14] Yuniawati Ekaningrum, Fredianaika Istanti, and Evada El Ummah Khoiro, "The Influence Of Service Quality On Customer Satisfaction In Surabaya Tourist Destinations During The Covid-19 Pandemic," *Int. J. Humanit. Educ. Soc. Sci.*, vol. 1, no. 3, pp. 161–168, 2021, doi: 10.55227/ijhess.v1i3.65.
- [15] P. Thiengburanathum, "An intelligent destination recommendation system for tourists," *PQDT - UK Irel.*, no. March, 2018.
- [16] D. Chicco, L. Oneto, and E. Tavazzi, "Eleven quick tips for data cleaning and feature engineering," *PLOS Comput. Biol.*, vol. 18, no. 12, p. e1010718, Dec. 2022.
- [17] L. He, N. N. Liu, and Q. Yang, "Active dual collaborative filtering with both item and attribute feedback," *Proc. Natl. Conf. Artif. Intell.*, vol. 2, pp. 1186–1191, 2011, doi: 10.1609/aaai.v25i1.8085.
- [18] R. H. Mondri and A. Wijayanto, "Recommendation System With Content-Based Filtering Method for Culinary Tourism in

- Mangan Application,” *ITSMART J. Ilm. Teknol. dan Inf.*, vol. 8, no. 2, pp. 65–72, 2019.
- [19] Z. Lu, Z. Dou, J. Lian, X. Xie, and Q. Yang, “Content-Based Collaborative Filtering for News Topic Recommendation,” pp. 217–223.
- [20] A. Felfernig, M. Jeran, G. Ninaus, F. Reinfrank, and M. Stettinger, “Basic Approaches in Recommendation Systems”.
- [21] Y. I. Lubis, D. J. Napitupulu, and A. S. Dharma, “Implementation of Hybrid Filtering (Collaborative and Content-based) Methods for the Tourism Recommendation System,” *12th Conf. Inf. Technol. Electr. Eng.*, pp. 6–8, 2020.