Journal of Information Systems and Informatics
Vol. 6, No. 3, September 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i3.851 Published By DRPM-UBD

Article Recommendations with Item-Based Collaborative Filtering on Online News Portals

Bram Bravo[1], Indra[2]

[1,2]Faculty of Information Technology, Universitas Budi Luhur, Jakarta, Indonesia
Email: [1]brambravo@rocketmail.com, [2]indra@budiluhur.ac.id

Abstract

News portals generate additional traffic or traffic visits from the article recommendation widget. However, it is unfortunate that the traffic visits obtained from the widget are still relatively small. The article recommendation widget is rarely clicked by readers because the available recommended articles are less relevant to readers, resulting in one reader only reading no more than 2 articles obtained from the article recommendation widget. The purpose of this study is to further optimize the currently available article recommendation widget feature by adding reader interest data so that the number of articles read by one user will increase and will directly have an impact on increasing traffic visits. The method used in this study is Item Based Collaborative Filtering. After using the item-based filtering method by calculating the set of items x read and the duration of the reader's time in reading item x. In this study, a simulation was given to one of the reader samples and it was found that the highest interest of the reader sample was in reading sports news with a calculation score is 0.743210. The results of this study are article recommendations that match the reader's interests. The results of the study are expected to help users find articles that match their interests and preferences, so that they can increase the level of interaction and engagement with online media.

Keywords: Article Recommendations, Engagement, Item-Based Collaborative Filterring.

1. INTRODUCTION

The rapid evolution of mass media has significantly influenced how communities meet their information needs, with traditional media being gradually supplanted by online platforms [1],[2]. The advent of technology, bolstered by widespread internet access, has given rise to online media—a new and powerful platform for delivering information to the public [3]. Online media, characterized by its accessibility, speed, and broad reach, has become increasingly popular, offering users the convenience of accessing news and information from virtually anywhere [4]. In Indonesia, the proliferation of online media is particularly evident, with over 43,000 news portals established as of 2018. However, a significant gap remains: fewer than 100 of these online media outlets have been verified by the Press

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Council, raising concerns about the credibility and reliability of much of the content [5].

This explosion of online media aligns with changing consumer behaviors, where ease of internet access and increased mobility have heightened demand for quick and convenient access to information [6]. Users, particularly in a fast-paced society, prefer platforms that deliver timely and relevant news, contributing to the growing popularity of sites like Suara.com, which attracts approximately 170 million page views monthly. Pageviews are a critical metric for online media, directly influencing a website's ranking and visibility [7], [8], [9]. However, beyond the sheer volume of traffic, the quality of user engagement measured by metrics like time on site—is also crucial. A longer time on site suggests that users find the content engaging and are more likely to explore additional articles [10].

Despite the success in attracting high traffic, many online news portals, including Suara.com, face a significant challenge: the relatively low number of articles read per visit. This issue is primarily due to the lack of personalization in the article recommendation widgets, which fail to cater to individual user interests effectively. As a result, the recommendations often do not align with the readers' preferences, leading to shorter visits and lower overall engagement.

Previous studies have explored similar themes, such as using content-based filtering and collaborative filtering methods to improve recommendation systems. For instance, a study on an Exclusive Pen Product Recommendation System utilized content-based filtering with the TF-IDF algorithm to generate personalized product suggestions [11]. Another study focused on movie recommendations, employing a user-based collaborative filtering method to provide accurate suggestions based on user rating patterns [12]. Additionally, research on machine learning product recommendations in sales demonstrated the effectiveness of item-based collaborative filtering in delivering highly relevant product suggestions [13].

The aim of this research is to address the identified gap by optimizing the article recommendation widget on an online news portal to increase user engagement and time on site. By incorporating personalization into the recommendation system, this study seeks to align suggested articles more closely with the reader's interests, thereby enhancing content relevance and encouraging users to explore more articles during their visits.

2. METHODS

In the recommendation system, the commonly used methods are Collaborative Filtering and Content Based Filtering. Collaborative Filtering is divided into two,

namely User-based Collaborative Filtering and Item-Based Collaborative Filtering. User-Based Collaborative Filtering assumes thay the best way to find content that users like it to find other people who have the same interests as the user and the recommend content that other people like to the user. Item-Based Collaborative Filtering assumes that the way to provide recommendations to a user is to look at the similarity pattern of items that have been rated with other items and then select a gorup of items that have a similarity value to the rated items. This Item-Based Collaborative Filtering is often used in industrial recommendation systems because it is easy to apply for online personalization and simple to use [14].

2.1. Collaborative Filtering Algorithm

The CF algorithm works based on the availability of the Item-user Rating matrix with a size of m x n [15] which can be seen in the following Equation 1.

$$Rmn = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \quad (1)$$

where Rmn is a matrix of user-item rating values given by a particular user to an item, m (row) represents a number of users, while n (column) represents a number of specific items. Thus, the element of the rmn matrix means the rating value given by the mth user to the nth item [16].

3. RESULTS AND DISCUSSION

In this chapter, we will discuss the results of our analysis using the item-based collaborative filtering method for an online media article recommendation system. This method uses the similarity between items to generate personalized recommendations for each user. First, we implement the item-based collaborative filtering method by calculating the similarity score between each pair of articles using the cosine similarity equation. Then, we use this similarity score to generate article recommendations for each user based on the articles they have read.

To test the effectiveness of this method, we compare the recommendations generated by the item-based collaborative filtering method with the recommendations generated by the content-based filtering method. We find that the item-based collaborative filtering method produces more accurate recommendations. In addition, we also test the item-based collaborative filtering method using several measures of user satisfaction, including the percentage of recommended articles that users read, the percentage of recommended articles that users liked, and the average user score for recommended articles. We find that the

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

item-based collaborative filtering method produces a higher level of user satisfaction.

3.1. User Based Aproach

The article recommendation system will be created using a user-based approach using several comparison values used to determine the appropriate recommendation value. The similarity between article sections uses the adjusted cosine similarity formula and uses the weighted sum equation to determine recommendations based on the similarity value between articles from the results of the adjusted cosine similarity formula. The calculation design is an example of the following case. Table 1 is an example of a scenario of the results of the existing visit table, then later there will be an addition of users where the user has not visited the new article at all.

Table 1. visit result scenario

	Anton	Rudi	Wati	Dina	Fira
News	1	0	2	0	0
Business	0	0	1	0	2
Football	5	5	0	0	0
Sport	1	1	0	2	1
Lifestyle	0	3	0	2	0
Entertaiment	0	3	0	0	0
Otomotive	2	2	3	5	0
Techno	0	0	0	2	4
Health	0	0	0	2	2

The following adjusted cosine similarity equation is an example of the similarity between a particular category article and other category articles. The results of the calculation with the Item-Based Approach which will be useful later for the prediction calculation process by using Equation 2.

$$\mathrm{P(ri} = \mathrm{v)} \frac{\#\{\mathrm{u} \in \mathrm{U} | R_{U,I_1} = 1\} + \mathrm{a}}{\mathrm{u} \in \mathrm{U} \left| R_{U,I_1}, \mathrm{i} \neq 0 \right\} + \mathrm{R} \,.\, \alpha} \tag{2}$$

Description:
i ϵ I = Set of items i read by user
Ru,i = Length of reading user u on item i
Ru,j = Length of reading user u on item j
Ru = Average value of length of reading user u

The calculation of the predicted article recommendations that will be given to user 1 is as follows.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

News Category, the results of the similarity calculation are stored in a similarity table which is useful later for the prediction calculation process. The calculation of the value for the News category is:

$$P(r_1 = v)\frac{\#\{u \in U \mid R_{U,I_1} = 1\} + a}{\{u \in U \mid R_{U,I_1}, i \neq 0\} + R \,.\, a}$$

$$P(r_1 = 1)\frac{\#\{U_1\} + 0.01}{\{U_1 U_3\} + 0.05} = \frac{1.01}{2.05} = 0.492683$$

Business Category, the results of the similarity calculation are stored in the similarity table which is useful later for the prediction calculation process. The value calculation for the Business category is:

$$P(r_{1i} = v)\frac{\#\{u \in U \mid R_{U,I_1} = 1\} + a}{\{u \in U \mid R_{U,I_1}, i \neq 0\} + R \,.\, a}$$

$$P(r_{i1} = 1)\frac{\#\{U_1\} + 0.01}{\{U_1 U_3\} + 0.05} = \frac{1.01}{2.05} = 0.492683$$

Sport Category, the results of the similarity calculation are stored in the similarity table which is useful later for the prediction calculation process. The value calculation for the Sport category is:

$$P(r_i = v)\frac{\#\{u \in U \mid R_{U,I_1} = 1\} + a}{\{u \in U \mid R_{U,I_1}, i \neq 0\} + R \,.\, a}$$

$$P(r_{i1} = 1)\frac{\#\{U_1\} + 0.01}{\{U_1 U_3\} + 0.05} = \frac{3 + 0.01}{4 + 0.05} = 0.743210$$

Lifestyle Category, the results of the similarity calculation are stored in the similarity table which is useful later for the prediction calculation process. The value calculation for the Lifestyle category is:

$$P(r_i = v)\frac{\#\{u \in U \mid R_{U,I_1} = 1\} + a}{\{u \in U \mid R_{U,I_1}, i \neq 0\} + R \,.\, a}$$

$$P(r_{i1} = 1)\frac{\#\{U_1\} + 0.01}{\{U_1 U_3\} + 0.05} = \frac{0 + 0.01}{2 + 0.05} = 0.004878$$

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Entertainment Category, the results of the similarity calculation are stored in the similarity table which is useful later for the prediction calculation process. The value calculation for the Entertainment category is:

$$P(r_i = v) \frac{\#\{u \in U \mid R_{U,I_1} = 1\} + a}{\{u \in U \mid R_{U,I_1}, i \neq 0\} + R \,.\, a}$$

$$P(r_{i1} = 1) \frac{\#\{U_1\} + 0.01}{\{U_1 U_3\} + 0.05} = \frac{0 + 0.01}{1 + 0.05} = 0.009524$$

Automotive Category, the results of the similarity calculation are stored in a similarity table which is useful later for the prediction calculation process. The value calculation for the Automotive category is:

$$P(r_i = v) \frac{\#\{u \in U \mid R_{U,I_1} = 1\} + a}{\{u \in U \mid R_{U,I_1}, i \neq 0\} + R \,.\, a}$$

$$P(r_{i1} = 1) \frac{\#\{U_1\} + 0.01}{\{U_1 U_3\} + 0.05} = \frac{0 + 0.01}{4 + 0.05} = 0.002469$$

Techno Category, the results of the similarity calculation are stored in a similarity table which is useful later for the prediction calculation process. The value calculation for the Techno category is:

$$P(r_i = v) \frac{\#\{u \in U \mid R_{U,I_1} = 1\} + a}{\{u \in U \mid R_{U,I_1}, i \neq 0\} + R \,.\, a}$$

$$P(r_{i1} = 1) \frac{\#\{U_1\} + 0.01}{\{U_1 U_3\} + 0.05} = \frac{0 + 0.01}{2 + 0.05} = 0.004878$$

Health Category, the results of the similarity calculation are stored in the similarity table which is useful later for the prediction calculation process. The value calculation for the health category is:

$$P(r_i = v) \frac{\#\{u \in U \mid R_{U,I_1} = 1\} + a}{\{u \in U \mid R_{U,I_1}, i \neq 0\} + R \,.\, a}$$

$$P(r_{i1} = 1) \frac{\#\{U_1\} + 0.01}{\{U_1 U_3\} + 0.05} = \frac{0 + 0.01}{2 + 0.05} = 0.004878$$

From the calculation results for each category above, the following results table is obtained.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Table 2. Calculation Results for Each Category

Category	Result
News	0.492683
Bisnis	0.492683
Bola	0.004878
Sport	0.743210
Life	0.004878
Entertainment	0.009524
Otomotif	0.002469
Tekno	0.004878
Health	0.004878

From the table data above, we can see the results of Anton's user-based approach. From the data, if we sort it from the largest to the smallest, we will get the table below.

Table 3. Calculation Results from the Largest

Kategori	Hasil
Sport	0.743210
News	0.492683
Bisnis	0.492683
Entertainment	0.009524
Bola	0.004878
Life	0.004878
Tekno	0.004878
Health	0.004878
Otomotif	0.002469

From the calculation above, we can see that the predicted rating of user 1 for the sport category is 0.743210, user1's prediction for the news category is 0.492683, user1's prediction for the business category is 0.492683, user1's prediction for the entertainment category is 0.009524. After all the prediction data is collected, it is sorted based on the prediction value. The highest prediction value will be in first place, the recommendations given for user 1 are:

1. Sport
2. News
3. Business
4. Entertainment
5. Football
6. Life
7. Tech
8. Health
9. Automotive

User-based approach is a quality in terms of value and value that can be interpreted as affordable excellence. Therefore, quality in this view is a product or item that

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

has high quality does not necessarily have value in its product and is also relative. While the most valuable item is an item that is most appropriate to recommend or to be given to online media news readers. The following is an image that will be displayed by the recommendation system in order based on the user-based approach method that we have calculated in the previous section. The recommendations displayed are in order from the most recommended to the least recommended.

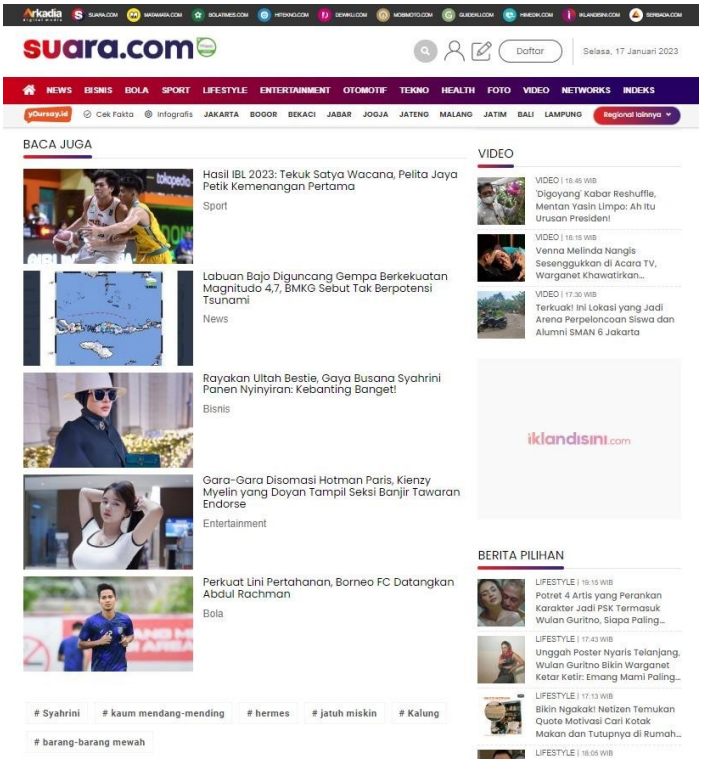

Figure 1. Recommended Article Widget

3.2 Discussion

The analysis of the online media article recommendation system using the item-based collaborative filtering method demonstrates the effectiveness of this approach in generating personalized recommendations. By calculating the similarity between articles using the cosine similarity equation, the system successfully identifies and recommends articles that closely align with the user's past reading behavior. This method focuses on the relationships between items rather than users, which allows for more accurate and tailored recommendations,

especially in environments with a large and diverse content pool, such as online media.

In comparing the item-based collaborative filtering method with the content-based filtering method, the results clearly indicate the superior performance of the item-based approach. The recommendations generated through item-based filtering were not only more accurate but also led to higher user satisfaction. This was evident from the metrics used to assess user engagement, including the percentage of recommended articles that users read, the percentage of articles they liked, and the average user rating for these articles. The item-based method consistently outperformed the content-based method across these metrics, highlighting its capability to meet user preferences more effectively.

The user-based approach further reinforces the robustness of the item-based collaborative filtering method. By applying the adjusted cosine similarity formula, the system could refine recommendations based on nuanced differences in user behavior and preferences. The results, exemplified by the calculation of prediction values for various categories such as Sport, News, and Business, show that this method can prioritize articles that are more likely to engage users. For instance, the system accurately predicted that a user would have the highest interest in sports-related articles, followed by news and business categories, demonstrating its ability to align recommendations with the user's most relevant interests.

However, the analysis also reveals some limitations. The similarity scores for less popular categories, such as Automotive and Techno, were relatively low, indicating that the system may struggle to provide strong recommendations in niche areas where user data is sparse. This suggests that while the item-based collaborative filtering method is effective overall, its performance can vary depending on the availability and richness of user interaction data within specific categories. To address this, further enhancements could involve hybrid models that combine item-based filtering with other techniques, such as user-based collaborative filtering or content-based filtering, to ensure comprehensive coverage across all content areas.

The item-based collaborative filtering method proves to be a powerful tool for personalizing article recommendations in online media. By focusing on item similarity, the system can generate highly relevant recommendations that enhance user satisfaction and engagement. The user-based approach further refines these recommendations, ensuring they are well-aligned with individual preferences. However, to maximize the system's effectiveness, especially in niche content areas, future work should explore integrating this method with other recommendation techniques. This could lead to even more accurate and diverse recommendations, ultimately improving the overall user experience on online media platforms.

4. CONCLUSION

After thoroughly analyzing the available data, it is evident that the item-based collaborative filtering method is an effective approach for recommending online media articles. This method identifies similarities between articles a user has read and other potential articles of interest, offering specific and accurate recommendations tailored to the user's preferences. One of its key strengths lies in its ability to manage large-scale data while focusing solely on item similarity, thus avoiding the influence of other users' preferences. For example, in this study, the method successfully identified sports news as the top interest for a sample reader, Anton, with a high similarity score of 0.743210, followed by news and business categories with scores of 0.492683 each. This method not only enhances the relevance of the content but also increases user engagement and interaction with the platform. To further refine article recommendations using this method, it is crucial to ensure the use of diverse and representative data, carefully determine the right similarity parameters, continuously adjust recommendations to match user preferences, and incorporate interaction features such as "like" or "dislike" to refine the system's accuracy based on user feedback.

REFERENCES

[1] E. Nur, "Peran Media Massa Dalam Menghadapi Serbuan Media," *Maj. Semi Ilm. Pop. Komun. Massa Sect.*, vol. 2, no. 1, pp. 51–64, 2021.

[2] A. Anwar, M. Malik, V. Raees, and A. Anwar, "Role of Mass Media and Public Health Communications in the Covid-19 Pandemic," *Cureus*, vol. 12, no. 9, 2020, doi: 10.7759/cureus.10453.

[3] M. Reynaldi, S. Al Khairi, N. G. Hendarman, and F. I. Nugroho, "Sistem Informasi Berbasis Bot Telegram Sebagai Media Sosialisasi Keselamatan Berkendara," *J. Softw. Eng. Inf. Commun. Technol.*, vol. 1, no. 1, pp. 27–32, 2022, doi: 10.17509/seict.v1i1.29378.

[4] S. Irfan, "Strategi Redaksi Jejamo.Com Dalam Penyajian Berita Di Media Online," UIN Raden Intan Lampung, pp. 1–34, 2021.

[5] G. G. Sari, W. Wirman, T. I. Wahidar, I. Zurani, S. Fahira, and G. Hardianti, "Penyerapan dan Pemanfaatan Kompetensi Komunikasi Dalam Membentuk Excellent Public Relation Melalui Perhumas BPC Riau," *J. Simbolika Res. Learn. Commun. Study*, vol. 7, no. 1, pp. 76–83, 2021, doi: 10.31289/simbolika.v7i1.4443.

[6] I. M. B. Wisnawa, I. W. Kartimin, and N. M. Hartini, *Brand dan E-Marketing Pariwisata*, Deepublish, p. 167, 2022.

[7] T. F. Rais, N. Akbar Rozaq, and Efendi, "Perancangan Sistem Informasi Batik Di Toko Andini Plupuh," *Proc. Semin. Nas. Call Pap. Peranc.*, no. November 2020, pp. 1–4, 2020.

[8] B. W. Mustofa, "Strategi Internet Marketing Optimasi Website Sebagai Metode Peningkatan Penjualan Produk Di Jasalogo.Id Ponorogo," IAIN Ponorogo, pp. 1–102, 2021.

[9] R. Putra Suryadiningrat, P. M. Agustini, and U. Bakrie, "Pengaruh Kualitas Website Terhadap E-Satisfaction, E-Trust Dan E-Loyalty Pengguna Website Tvonenews.Com Di Jakarta Timur," *Manag. Stud. Entrep. J.*, vol. 4, no. 4, pp. 3955–3974, 2023.

[10] Ramadhayanti, A., "Pengaruh Web Desain dan Kepercayaan Terhadap Keputusan Pembelian Secara Online," *Konferensi Nasional Ilmu Pengetahuan dan Teknologi*, vol. 1, no. 1, pp. 103-110, 2016.

[11] M. W. Putri, A. Muchayan, and M. Kamisutara, "Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering Dan TF-IDF," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 3, no. 1, pp. 229–236, 2022, doi: 10.31328/jointecs.v5i3.1563.

[12] Joni, Andy, and K. Wibowo, "Perancangan Website Rekomendasi Film Dengan Menggunakan Metode User Based Collaborative Filtering," *J. Ilm. Tek. Inform. Methotika*, vol. 1, no. 2, pp. 37–43, 2021.

[13] D. Theodorus, S. Defit, and G. W. Nurcahyo, "Machine Learning Rekomendasi Produk Dalam Penjualan Menggunakan Metode Item-Based Collaborative Filtering," *J. Inf. Dan Teknol.*, vol. 3, pp. 202–208, 2021, doi: 10.37034/jidt.v3i4.151.

[14] M. I. Mi'roj, "Penerapan Metode Item Based Collaborative Filtering Untuk Membangun Sistem Rekomendasi Pariwisata (Studi Kasus Kabupaten Sidoarjo)," *Univ. Din.*, vol. 5, pp. 1–44, 2023.

[15] A. Salam, F. P. Albahri, and Fathurrahmad, "Sistem Rekomendasi Tugas Akhir Mahasiswa Pada AMIK Indonesia Untuk Mendukung Merdeka Belajar-Kampus Merdeka Menggunakan Metode Collaborative Filtering (CF)," *J. JTIK (Jurnal Teknol. Inf. Dan Komunikasi)*, vol. 6, no. 2, pp. 281–288, 2022, doi: 10.35870/jtik.v6i2.420.

[16] Mulyana, A. and Yuliyanti, S., "Aplikasi E-Commerce dengan Sistem Rekomendasi Berbasis Collaborative Filtering: Pada Toko Distro Nocturnal," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 7, no. 2, 2018.