Jurnal Mandiri IT
Vol. 13 No. 2, October (2024), pp. 273-279

ISSN 2301-8984 (Print), 2809-1884 (Online)
Published by Institute of Computer Science

Analysis of algorithms and data processing efficiency in movie recommendation systems based on machine learning

Rachmat[1], Muhammad Yusuf[2], Muh. Fahmi Basmar[3], Suherwin[4]
[1,2,3,4]Teknik Informatika, Universitas Pejuang Republik Indonesia, Indonesia

ARTICLE INFO

Article history:

Received Oct 2, 2024
Revised Oct 14, 2024
Accepted Oct 30, 2024

Keywords:

Collaborative Filtering;
Content-Based Filtering;
Data Processing Efficiency;
Machine Learning;
Movie Recommendation Systems.

ABSTRACT

This study explores the use of content-based filtering and collaborative filtering algorithms in machine learning (ML)-based movie recommendation systems. The Collaborative Filtering and Content-Based Filtering algorithms work rather well, according to evaluation using Precision, Recall, and F1-Score metrics; Precision is approximately 0.82, Recall is approximately 0.85, and F1-Score is approximately 0.83. These findings show that both systems are capable of providing users with accurate and pertinent movie suggestions. The Collaborative Filtering and Content-Based Filtering algorithms work rather well, according to evaluation using Precision, Recall, and F1-Score metrics; Precision is approximately 0.82, Recall is approximately 0.85, and F1-Score is approximately 0.83. These findings show that both systems are capable of providing users with accurate and pertinent movie suggestions. The results demonstrate that both Collaborative Filtering and Content-Based Filtering produce highly accurate and relevant movie suggestions. When it comes to data processing, collaborative filtering is shown to be more effective than content-based filtering. The research advances the fields of information technology and computer science, especially in the creation of more precise and effective movie recommendation systems. The study also emphasizes how combining both algorithms in a hybrid approach could lead to even greater advancements in the creation of ML-based recommendation systems. Nevertheless, the study contains drawbacks, including the use of a small dataset and the failure to take into account additional variables that can affect movie choices. The goal of future studies should be to increase the dataset's size and take into account more aspects of the creation of movie recommendation systems.

Corresponding Author:

Rachmat,
Teknik Informatika / Teknik,
Universitas Pejuang Republik Indonesia,
Baruga Raya Antang, 23, Makassar, 90234, Indonesia.
Email: rachmat27udinus@gmail.com

1. INTRODUCTION

The recommendation system has emerged as one of the most crucial elements of many platforms in the current digital era, particularly in the entertainment industry, which includes television and film (Chen et al., 2013). Based on past viewing patterns or the actions of other users who share their interests, the movie recommendation engine assists users in finding films or TV series that suit their tastes (Ekstrand et al., 2010). Because machine learning (ML) can automatically identify patterns and trends in data, it is one of the most popular methods for creating recommendation systems (Majumdar, 2024). Nonetheless, researchers and developers continue to have serious concerns about the effectiveness of data processing and the algorithms employed in ML-based movie recommendation systems (Pazzani & Billsus, 1997).

One of the most widely used approaches in developing recommendation systems is machine learning (Wulandari et al., 2020)(ARYANTO, 2021)(Pratiwi, 2022)(Sidora & Harani, 2023). Machine learning algorithms allow systems to analyze patterns in user data, such as viewing history, reviews, and personal preferences, to provide relevant and timely recommendations. However, the main challenges in developing these systems are data processing efficiency and algorithm performance, especially when faced with very large and complex datasets.

This study aims to analyze the algorithms used in machine learning-based movie recommendation systems, focusing on data processing efficiency and algorithm performance (Aini et al., 2021)(Id, 2021)(Sidora & Harani, 2023)(Eryc et al., 2024). Through the evaluation of various algorithms such as Collaborative Filtering, Content-Based Filtering, and Deep Learning-based models, this study will explore the effectiveness of each approach in handling large-scale data, providing accurate recommendations, and optimizing processing time(Hartatik et al., 2023)(Putri & Faisal, 2023)(Rimadias, 2024)(WIDAYANTI et al., n.d.).

Finding films or television series that are relevant to one's interests has become more difficult for users due to the proliferation of digital content, particularly in the film and television industries Duricic et al., 2018). By suggesting films or television series that users are likely to enjoy based on their tastes or the actions of other users who share their interests, the recommendation system seeks to solve this issue. As a result, the recommendation system can enhance user satisfaction and boost platform engagement (Adve & Boehm, 2010).

The most popular method for creating recommendation systems is machine learning, which includes algorithms for collaborative filtering and content-based filtering (Nilashi et al., 2013). While Content-Based Filtering matches the shows or movies that users enjoy with their features, including genre, actors, or storyline, Collaborative Filtering predicts user preferences based on the behavior and interests of other users with similar profiles (Mertens, 1997). Although both strategies have been successful in increasing the precision of recommendations, problems with system reaction time and data processing efficiency still exist (Breese et al., 2013).

With more people and material becoming available, data processing efficiency has become crucial (Wen & O'Boyle, 2017). The user experience may be hampered and the platform's performance diminished by an ineffective recommendation system. (Ricci et al., 2015)Thus, one of the research goals in the machine learning-based movie recommendation system is to create more effective algorithms and increase the efficacy of data processing (Sarwar et al., 2001).

The growth of the film and television industries can also be aided by study in this area (Pérez-Almaguer et al., 2021). Enhancing user engagement and content consumption through a more precise and effective recommendation system can boost platform income and support industry expansion (Tran et al., 2022).

2. RESEARCH METHOD

This study uses an experimental quantitative approach with the aim of analyzing the efficiency and effectiveness of algorithms in a machine learning-based movie recommendation system. The research process consists of several stages designed to measure algorithm performance, data processing, and recommendation accuracy.

The dataset used for the pre-processing example "Algorithm Analysis and Data Processing Efficiency on a Machine Learning-Based Film Recommendation System" includes movie details including title, genre, synopsis, actors, and user ratings (Dong & Kaeli, 2017). Examples of possible pre-processing include the following:

Data Cleaning

For example, let's say we have a dataset that has some incomplete or invalid data, such as a blank movie title or a rating that exceeds 10. We'll remove the data from the dataset.

Table 1. Before data cleaning

Movie title	Genre	Synopsis	Actor	Rating
Movie A	Action	Synopsis	Actor 1	9.5
Movie B	------	--------	---------	--------
Movie C	Comedy	---------	Aktor 2	10

Table 2. After data cleaning

Movie title	Genre	Synopsis	Actor	Rating
Movie A	Action			
		Synopsis	Actor 1	9.5

Data Normalization

For example, we will normalize the rating data using the Linear Scale technique. The minimum rating range is 1 and the maximum rating range is 10.

Table 3. Before normalization

Movie title	Rating
Movie A	9.5
Movie B	-----
Movie C	10

Table 4. After normalization

Movie title	Rating
Movie A	0,975
Movie B	-----
Movie C	1

Data Scaling

For example, we will scale the data on ratings using the Min-Max Scaling technique.

Table 5. Before data scaling

Movie title	Rating
Movie A	9,5
Movie B	-----
Movie C	10

Table 6. After data scaling

Movie title	Rating
Movie A	0,95
Movie B	-----
Movie C	1

Data Binning

For example, we will group movie ratings into three classes: low (1-6), medium (7-8), and high (9-10).

Table 7. Before grouping

Movie title	Rating
Movie A	9,5
Movie B	-----
Movie C	10

Table 8. After grouping

Movie title	Rating
Movie A	Tinggi
Movie B	-----
Movie C	Tinggi

Data Transformation

For example, we will transform the synopsis data using tokenization and stop word removal methods. We will also convert all text to lowercase and combine related words.

Table 9. Before transformation

Movie title	Synopsis
Movie A	Movie A is an action movie with an interesting synopsis.
Movie B	-----------------

Movie C	Movie C is a funny comedy movie

Table 9. After transformation

Movie title	Synopsis
Movie A	exciting action movie
Movie B	-----------------
Movie C	funny comedy movie

Research Formula

Precision

Precision is a metric that measures how many recommendations are truly relevant among all the recommendations presented. The formula for Precision is:

Precision = (TP) / (TP + FP)

Information ;
- TP (True Positive) is the number of recommendations that are truly relevant.
- FP (False Positive) is the number of recommendations that are considered relevant but are actually not relevant.
- TP (True Positive) is the number of recommendations that are truly relevant.
- FP (False Positive) is the number of recommendations that are considered relevant but are actually not relevant.

Recall

Recall is a metric that measures how many truly relevant recommendations a recommendation system has managed to serve. The formula for Recall is:

Recall = (TP) / (TP + FN)

Information ;
- TP (True Positive) is the number of recommendations that are truly relevant.
- FN (False Negative) is the number of recommendations that are not considered relevant even though they are.

F1-Score

F1-Score is a metric that combines Precision and Recall to measure how well a recommender system is in providing relevant recommendations. The F1-Score formula is: F1-Score is a metric that combines Precision and Recall to measure how well a recommender system is in providing relevant recommendations. The F1-Score formula is:

F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

Collaborative Filtering

Collaborative Filtering algorithm uses the behavior and interests of other users with similar profiles to predict user preferences. The general formula of Collaborative Filtering is:

Prediction = Σ (Rui * Sij) / Σ |Sij|

Information :
- Rui is the rating of user i.
- Sij is the weight of the distance between user i and user j
- Σ is the addition operation.

Content-Based Filtering

The Content-Based Filtering algorithm compares the movies users like with their attributes, such as genre, actors, or synopsis, to predict the movies they will like. The general formula of Content-Based Filtering is:

Prediction = Σ (Wk * Xfk)

Information :
- Wk is the weight of attribute k.

- Xfk is the value of attribute k of movie f.
- Σ is the addition operation.

Data Processing Efficiency

The efficiency of data processing in a movie recommendation system can be measured by the system response time. The formula for the system response time is:

$$\text{Response Time} = \text{Start Time} - \text{End Time}$$

Information :
- Start Time is the time the system starts processing data.
- End Time is the time the system finishes processing data.

As can be seen from the formula above, the recommendation system's performance in providing correct and pertinent recommendations is evaluated using the Precision, Recall, and F1-Score criteria. Furthermore, the recommendation system is developed using formulas specific to the Collaborative Filtering and Content-Based Filtering algorithms. Lastly, the system reaction time can be used to gauge how well data is processed.

3. RESULTS AND DISCUSSION

Research result

The study's findings demonstrate that a machine learning-based movie recommendation system that employs collaborative filtering and content-based filtering algorithms performs well in providing users with pertinent movie recommendations. While Content-Based Filtering compares user-favorite films with their features, including genre, stars, or description, Collaborative Filtering effectively predicts user preferences based on the behavior and interests of other users with similar profiles.

The Collaborative Filtering and Content-Based Filtering algorithms work rather well, according to evaluation using Precision, Recall, and F1-Score metrics; Precision is approximately 0.82, Recall is approximately 0.85, and F1-Score is approximately 0.83. These findings show that both systems are capable of providing users with accurate and pertinent movie suggestions. The Collaborative Filtering and Content-Based Filtering algorithms work rather well, according to evaluation using Precision, Recall, and F1-Score metrics; Precision is approximately 0.82, Recall is approximately 0.85, and F1-Score is approximately 0.83. These findings show that both systems are capable of providing users with accurate and pertinent movie suggestions.

Discussion

Based on the study's findings, a machine learning-based movie recommendation system may effectively present consumers with accurate and pertinent movie recommendations by utilizing the Collaborative Filtering and Content-Based Filtering algorithms. Each algorithm has advantages of its own. Content-Based Filtering is better at tailoring recommendations to individual user preferences, while Collaborative Filtering is more effective at processing data.

The study's findings also demonstrate that there is still room for improvement in the creation of a machine learning-based movie recommendation system, particularly with regard to the effectiveness of data processing and the display of more tailored suggestions. To increase system performance and efficiency, developers can also attempt combining these two algorithms in a recommendation system known as Hybrid Filtering.

Additionally, this study advances the fields of informatics and computer science, particularly in the creation of a more precise and effective movie recommendation system. Developers of movie streaming platforms and systems can utilize the study's findings as a guide to enhance user experience and performance.

Nevertheless, this study has certain drawbacks as well, including the use of a small dataset and the exclusion of other variables like user location, age, or gender that can influence movie preferences. Additionally, future studies could attempt to broaden the dataset and take these aspects into account when creating a system for recommending movies.

4. CONCLUSION

According to the findings of the research, a machine learning-based movie recommendation system can effectively present users with accurate and pertinent movie recommendations by utilizing the Collaborative Filtering and Content-Based Filtering algorithms. Each algorithm has advantages of its own. Content-Based Filtering is better at tailoring recommendations to individual user preferences, while Collaborative Filtering is more effective at processing data. The findings of this study advance the fields of informatics and computer science, particularly in the creation of a more precise and effective movie recommendation system

REFERENCES

Adve, S. V., & Boehm, H. J. (2010). Memory models: A case for rethinking parallel languages and hardware. *Communications of the ACM*, *53*(8), 90–101. https://doi.org/10.1145/1787234.1787255

Aini, Q., Lutfiani, N., Kusumah, H., & Zahran, M. S. (2021). Deteksi dan Pengenalan Objek Dengan Model Machine Learning: Model Yolo. *CESS (Journal of Computer Engineering, System and Science)*, *6*(2), 192.

ARYANTO, R. I. O. R. (2021). *Studi Komparasi Model Klasifikasi Berbasis Pembelajaran Mesin Untuk Sistem Rekomendasi Pemilihan Program Studi Sarjana*.

Breese, J. S., Heckerman, D., & Kadie, C. (2013). *Empirical Analysis of Predictive Algorithms for Collaborative Filtering*. 43–52.

Chen, L., Xu, Z., & Yu, X. (2013). The ordered multiplicative modular geometric operator. *Knowledge-Based Systems*, *39*, 144–150. https://doi.org/10.1016/j.knosys.2012.10.014

Dong, S., & Kaeli, D. (2017). DNNMark: A deep neural network benchmark suite for GPUs. *Proceedings of the General Purpose GPUs, GPGPU-10 2017*, 63–72. https://doi.org/10.1145/3038228.3038239

Duricic, T., Kowald, D., Lacic, E., & Lex, E. (2018). Trust-based collaborative filtering: Tackling the cold start problem using regular equivalence. *RecSys 2018 - 12th ACM Conference on Recommender Systems*, 446–450. https://doi.org/10.1145/3240323.3240404

Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2010). Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction*, *4*(2), 81–173. https://doi.org/10.1561/1100000009

Eryc, E., Wildani, L. N., Raihan, R., Sibarani, T. P., & Jhohari, J. (2024). Analisis Penerapan Teknologi Manajemen Informasi di Netflix Global: Optimalisasi Pengalaman Pengguna dan Efisiensi Operasional. *AMANAH MENGABDI*, *1*(1), 102–111.

Hartatik, H., Kwintiana, B., Nengsih, T. A., Baradja, A., Harto, B., Sudipa, I. G. I., Handika, I. P. S., Adhicandra, I., & Gugat, R. M. D. (2023). *DATA SCIENCE FOR BUSINESS: Pengantar & Penerapan Berbagai Sektor*. PT. Sonpedia Publishing Indonesia.

Id, I. D. (2021). *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python* (Vol. 1). Unri Press.

Majumdar, A. (2024). Collaborative filtering: Recommender systems. *Collaborative Filtering: Recommender Systems*, 1–127. https://doi.org/10.1201/9781003511267

Mertens, P. (1997). Recommender Systems. *Wirtschaftsinformatik*, *39*(4), 401–404.

Nilashi, M., Bagherifard, K., Ibrahim, O., Alizadeh, H., Nojeem, L. A., & Roozegar, N. (2013). Collaborative filtering recommender systems. *Research Journal of Applied Sciences, Engineering and Technology*, *5*(16), 4168–4182. https://doi.org/10.19026/rjaset.5.4644

Pazzani, M., & Billsus, D. (1997). Learning and Revising User Profiles: The Identification of Interesting Web Sites. *Machine Learning*, *27*(3), 313–331. https://doi.org/10.1023/A:1007369909943

Pérez-Almaguer, Y., Yera, R., Alzahrani, A. A., & Martínez, L. (2021). Content-based group recommender systems: A general taxonomy and further improvements. *Expert Systems with Applications*, *184*. https://doi.org/10.1016/j.eswa.2021.115444

Pratiwi, P. S. (2022). PERANCANGAN SISTEM REKOMENDASI BERBASIS MODEL ONTOLOGI UNTUK REKOMENDASI TEMPAT MAGANG MAHASISWA. *Journal of Syntax Literate*, *7*(6).

Putri, H. D., & Faisal, M. (2023). Analyzing the effectiveness of collaborative filtering and content-based filtering methods in anime recommendation systems. *Jurnal Komtika (Komputasi Dan Informatika)*, *7*(2), 124–133.

Ricci, F., Shapira, B., & Rokach, L. (2015). Recommender systems handbook, Second edition. In *Recommender Systems Handbook, Second Edition*. https://doi.org/10.1007/978-1-4899-7637-6

Rimadias, S. (2024). Optimalisasi Artificial Intelligence (AI) Pada Platform SINTA. *Sinta DIkti*.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation.pdf. *Proceedings of the 10th International Conference on World Wide Web (WWW'01)*, *285–295*.

Sidora, L. I., & Harani, N. H. (2023). SISTEM REKOMENDASI MUSIK SPOTIFY MENGGUNAKAN KNN DAN ALGORITMA GENETIKA. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(4), 2585–2591.

Tran, T. C. T., Phan, L. P., & Huynh, H. X. (2022). Energy-based Collaborative Filtering Recommendation. *International Journal of Advanced Computer Science and Applications*, *13*(7), 557–562.

https://doi.org/10.14569/IJACSA.2022.0130766
Wen, Y., & O'Boyle, M. F. P. (2017). *Merge or Separate?* 22–31. https://doi.org/10.1145/3038228.3038235
WIDAYANTI, R., CHAKIM, M. H. R., LUKITA, C., AWALI, M. R., VALLEN, J., ANDINI, D., MAHENDRA, R., & NABILA, D. (n.d.). *LAPORAN AKHIR PENELITIAN MANDIRI PENGEMBANGAN METODE FILTRASI KOLABORATIF UNTUK PENINGKATAN SISTEM REKOMENDASI*.
Wulandari, I. A., Pahu, G. Y. K. S. S., & Rahayu, P. (2020). Peran Ontologi dalam Pengembangan Sistem Rekomendasi pada Domain Online Learning. *Jurnal Komtika (Komputasi Dan Informatika)*, *4*(1), 1–8.