Journal of Multimedia Trend and Technology - JMTT
Vol. 3, No. 2, August 2024, ISSN 2964-1330
https://journal.educollabs.org/index.php/jmtt/

Movie Recommendation System Model using Bisecting K-Means Technique and Collaborative Filtering

Hernawati Gohzali[1], Dita Maria Panjaitan[2]

Informatics Engineering Study Program, Universitas Mikroskil, Medan, Indonesia
Email: hernawati.gohzali@mikroskil.ac.id[1], ditamaria@student.mikroskil.ac.id[2]

ARTICLE INFO

History :

Submit on 3 March 2024
Review on 13 May 2024
Accepted on 5 June 2024

Keyword :

Recomendation,
Model,
Film,
Data Mining

ABSTRACT

In the current film industry, the competition is very big. We can see it in online streaming content through the ratings obtained. Film itself is a visual work that is packaged as a product of public entertainment for a specific purpose. However, there are also many films that are considered not to meet the audience's expectations. Even the films presented are sometimes illegal or pirated films. We can also find out whether a film is recommended or not. The problem is that viewers rarely understand how to see recommendations or even provide appropriate film recommendations. This study aims to develop a film recommendation system model using a combination of K-Means bisecting and Collaborative Filtering. The film data used in this study comes from Movie-Lens which consists of 100,000 ratings from 668 users for 10329 film titles in 18 film genres. The training process consists of a cluster process with the K-Means bisecting algorithm and calculating similarity values with collaborative filtering (item-based and user-based). The testing process is carried out to calculate the system error value by calculating the Mean Absolute Error (MAE) value. The results of the study show that recommendations with bisecting K-Means and user-based collaborative filtering get lower MAE values compared to bisecting K-Means and item-based collaborative filtering.

Copyright © 2024 by Author
The copyright of this article belongs entirely to the author

Corresponding Author:

Hernawati Gohzali
Informatics Engineering Study Program, Universitas Mikroskil, Medan, Indonesia
Email: hernawati.gohzali@mikroskil.ac.id

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

INTRODUCTION

Movies have become one of the most popular entertainment media among the public[1][2]. Since 1874 to 2015, as many as 3,361,741 movie titles have been released by the film industry (http://imdb.com)[1]. The large number of movie titles that have been circulated makes it difficult for people to find the movies they want. Movie rating data contained in a website can be processed and used to recommend movies to other users[3]. The consideration is to find movies based on the relationship between one movie and another movie that has been rated by the user to be used as a recommendation to other users[4].

Movie recommendations play a very important role for viewers for the following reasons, (1) Saving Time, With so many movie choices available, viewers often have difficulty choosing a movie that suits their preferences[4][5]. Recommendations help them find movies they might like without having to spend time searching. (2) Personalized Experience, The right recommendations create a more personal and satisfying experience for viewers. By getting suggestions that suit their tastes, viewers are more likely to enjoy the recommended movies[6][7]. (3) Increasing Viewer Satisfaction, when viewers get recommendations for movies they enjoy, they tend to feel more satisfied with the platform or service that provides the recommendations. This increases their loyalty and engagement with the platform[8]. (4) Reducing the Risk of Disappointment, good recommendations can reduce the risk of viewers watching movies that don't suit their tastes, which can lead to disappointment[9]. That way, they can enjoy a positive viewing experience more often.

Therefore, a system is needed that can recommend movies to users. There are quite a few techniques that are currently known to recommend a movie[10]. First, Collaborative Filtering, this technique recommends movies based on similarities in preferences between users[11]. If two users have similar viewing patterns, movies that are liked by one user can be recommended to other users[4]. Then this technique can also analyze the similarities between movies. For example, if a user likes a certain movie, the system will recommend other movies that are similar to the ones they already like[6].

Second, Content-Based Filtering, this technique recommends movies based on the characteristics of the movies that users like, such as genre, director, actor, or synopsis. For example, if a user often watches action movies, the system will recommend more action movies[12].

Third, Deep Learning, this technique uses neural networks models to learn user preferences more deeply and complexly. For example, deep learning models can analyze not only ratings, but also reviews, movie metadata, and user interactions to provide more personalized recommendations[13][1].

Next, there is Social Filtering. Recommendations are given based on the activities and preferences of the user's friends or social networks[10]. For example, if many of the user's friends watch and like a particular movie, the movie can be recommended. There are many more techniques that can be used to determine the best movie[14].

A recommendation system is a mechanism that can provide information or recommendations according to user preferences based on information obtained from the user[9][10]. Therefore, an appropriate recommendation model is needed so that the recommendations given by the system are in accordance with user preferences, and make it easier for users to make decisions in determining the items (films) to be selected[15]. To increase the accuracy of the relationship between users with the same preferences for an item (film), a clustering algorithm is used. One of the recommendation methods used in the recommendation system is Collaborative filtering[14][13].

Collaborative filtering connects each user with the same preference for a movie item based on the rating given by a small number of groups so that each group has something essential in common. Previous studies have tried to combine Collaborative filtering with the K-Means algorithm which produces an efficient recommendation system for processing large amounts of data and high accuracy. Bisecting K-Means is a better algorithm than the K-Means algorithm because it produces uniform clusters and does not produce empty clusters, a good level of accuracy and is more efficient when the number of clusters increases[16]. This study combines Collaborative filtering with Bisecting K-Means to produce a good recommendation system.

METHOD

This research begins with the process of data collection, process analysis, system requirements analysis, system design, implementation and system testing. Movie data is obtained from the Movie-Lens dataset containing 100,000 ratings, 10,329 movies with 668 users. Movies are grouped into 18 genres, namely drama, comedy, short, documentary, talk-show, family, news, romance, animation, music, reality-TV, crime, action, game-Show, adventure, trailer, mystery, fantasy, sci-fi, adult, sport, horror, history, biography, western, war, and film noir. Users can give ratings (scale 1-5) with details of scale 1 being the worst and scale 5 being the best, to the film if they have registered.

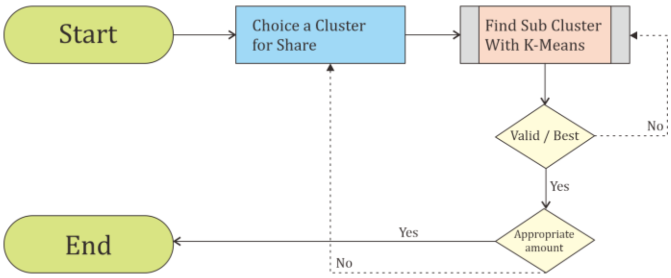

Figure 1, K-Means Bisecting Process Flow

The main idea in collaborative filtering recommendation system is to utilize the opinions of other users to predict items that a user might like or be interested in. The quality of recommendations given using this method is highly dependent on the opinions of other users (neighbours) towards an item. User-Based Collaborative Filtering finds a group of neighbour users who have the same history of preferences as the user who will be targeted for recommendations. After a group of neighbours is formed, the system will combine the neighbours' preferences to generate recommendations to the active user[4].

a. Calculating Similarity.

Pearson Correlation is used to calculate the similarity value between users and items, as in Equation 1.

$$S_{(i,j)} = \frac{\sum u \in U (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum u \in U (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum u \in U (R_{u,j} - \bar{R}_j)^2}} \quad (1)$$

Description:

$S_{(i,j)}$	Similarity value between item i and item j
$u \in_U$	Set of users who rated item i and item j
$R_{u,i}$	Rating of user u on item i
$R_{u,j}$	Rating of user u on item j
R_i	Average rating of item i
R_j	Average rating of item j

b. Calculating Prediction Values.

The second step is to calculate the predicted rating of the items. The way to calculate the predicted value for a new user or item is to use the Weighted Sum equation according to Equation 2.

$$P_{(a,j)} = \frac{\sum_{i \in I}(R_{a,i} * RS_{i,j})}{\sum_{i \in I}|S_{i,j}|} \quad (2)$$

Description:

$P_{(a,j)}$	Prediction of rating of item j by user a
$i \in I$	Set of items i that are similar to item j
$R_{a,i}$	Rating of user a on item i
$s_{i,j}$	Similarity value between items i and j

Item-Based Collaborative Filtering is a recommendation method based on the similarity between the rating of a product and the product purchased. From the level of product similarity, it is then divided by the customer's needs parameter to obtain the product's utility value. The product with the highest utility value is then recommended.

a. Calculating Similarity.

To calculate the similarity value between items and users, the adjusted cosine equation is used according to Equation 3.

$$s_{(i,j)} = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_u)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_u)^2}} \quad (3)$$

Description:

$S_{(i,j)}$	Similarity value between item i and item j
$u \in U$	Set of users who rated item i and item j
$R_{u,i}$	Rating of user u on item i
$R_{u,j}$	Rating of user u on item j
$\bar{R}_u$	Average rating value of user u

b. Calculating Prediction Values.

The second step is to calculate the predicted rating of the items. The way to calculate the predicted value for a new user or item is to use the Weighted Sum equation according to Equation 4.

$$P_{(a,j)} = \frac{\sum_{i \in I}(R_{a,i} * RS_{i,j})}{\sum_{i \in I}|S_{i,j}|} \quad (4)$$

Description:

$P_{(a,j)}$	Prediction of rating of item j by user a
$i \in I$	Set of items i that are similar to item j
$R_{a,i}$	Rating of user a on item i
$s_{i,j}$	Similarity value between items i and j

c. Combination of Bisecting K-Means and Collaborative Filtering.

This study combines Bisecting K-Means and Collaborative Filtering to obtain movie recommendation results. Figure 2 shows the combination of bisecting K-Means and item-based collaborative filtering.

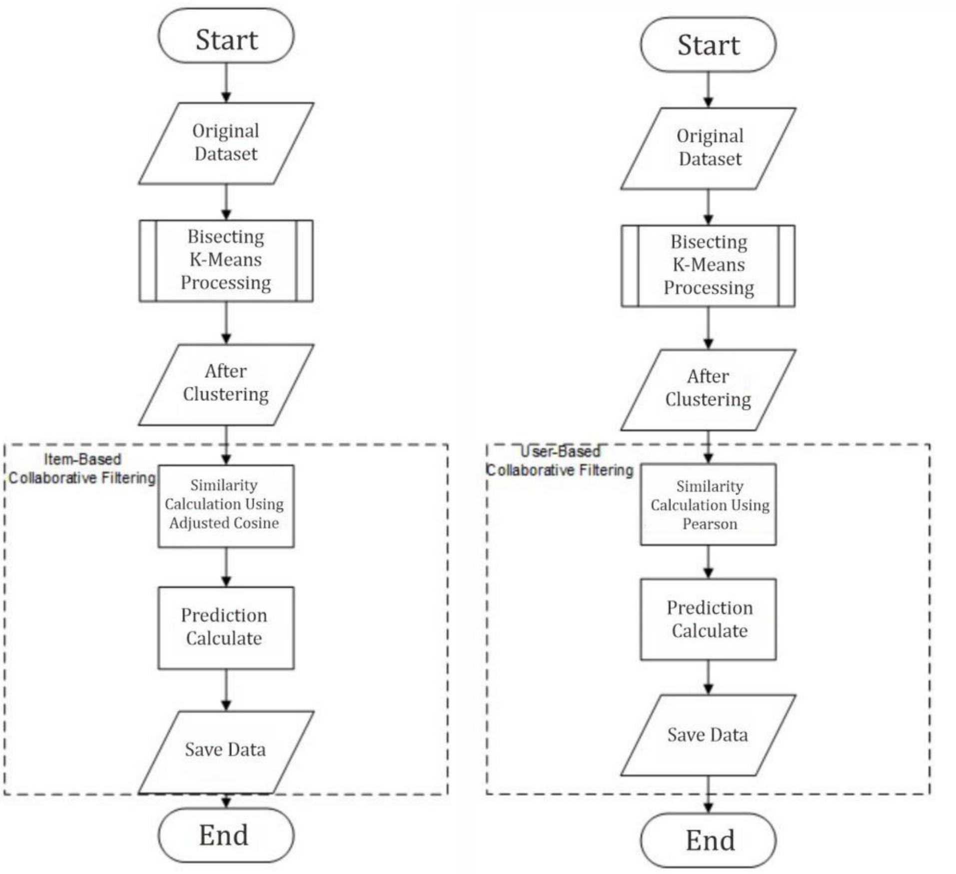

Figure 2. K-Means bisecting flowchart and item-based collaborative filtering

The design consists of designing the website interface of the movie recommendation system using balsamiq and designing the database using Entity Relationship Diagram (ERD). The recommendation system is built using PHP, HTML and CSS programming languages. The database is a relational database, namely MySQL. Then the final stage of testing is carried out with a data set from Movie-Lens. To test the system, the recommendation results are tested based on the rating prediction. The test will calculate the level of accuracy using the Mean Absolute Error (MAE) based on the Neighborhood Size (NS) parameter, then the results are analyzed. The MAE value can be calculated using Equation 5.

$$MAE = \frac{\Sigma_{i=1}^{N} abs(p_i - r_i)}{N} \quad (5)$$

Description:

p_i	Predicted rating value of i-th data
r_i	Actual rating value of i-th data
N	Number of Data

RESULT & DISCUSSION

The first step that was carried out was to try to collect film data originating from websites that had been built previously. The initial view of the website-based movie recommendation system that the user will see is shown in Figure 3.

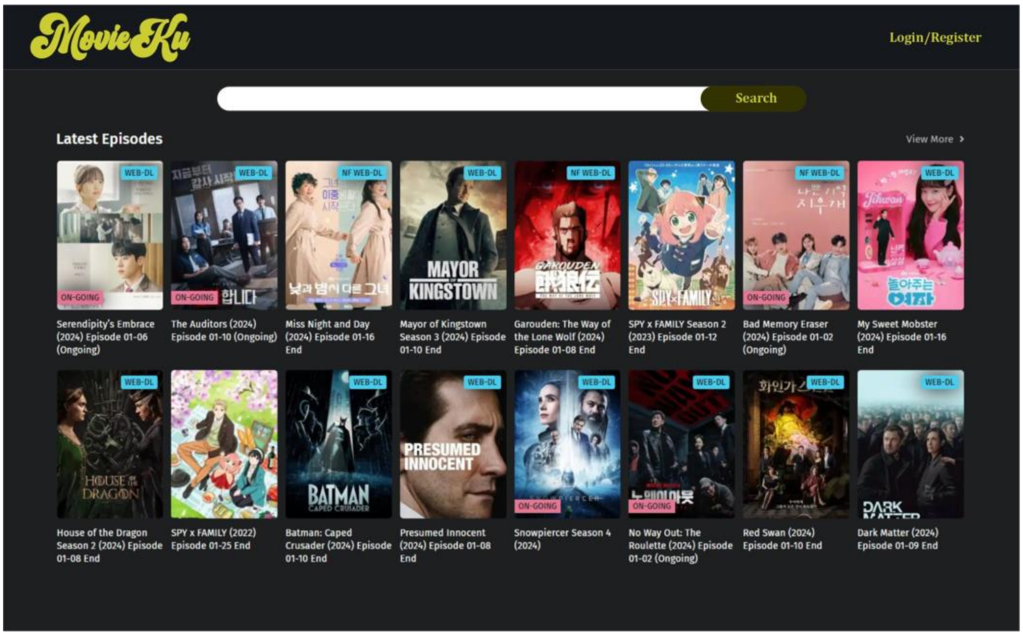

Figure 3. Initial View of Movie Search on Website

On this page, the application logo, login form, registration link, search column that can display several movie titles according to the keywords entered by the user and the genre list menu are displayed. In addition, there are several movies accompanied by images, titles and short descriptions of the film where when selected one will lead to the details of the film.

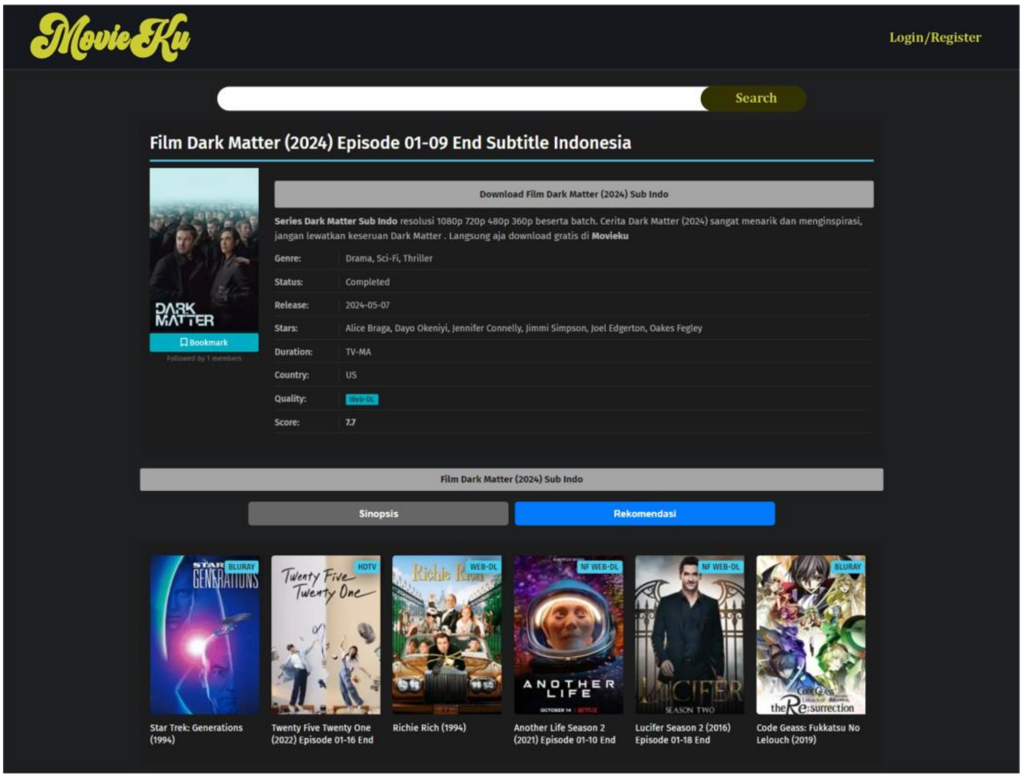

Figure 4. Tampilan rekomendasi film

The movie details display consists of the selected movie image along with the title, genre, description and rating that can be filled in by members. Ratings cannot be filled in by non-members. The right sidebar displays 6 movie recommendations based on genre where these recommendations are displayed based on the Item-based collaborative filtering method and the bottom sidebar displays 6 movie recommendations based on other users' likes where these recommendations are displayed based on the User-based collaborative filtering method. Figure 5 shows the results of the training process with a combination of K-Means bisecting and collaborative filtering for each movie in the dataset.

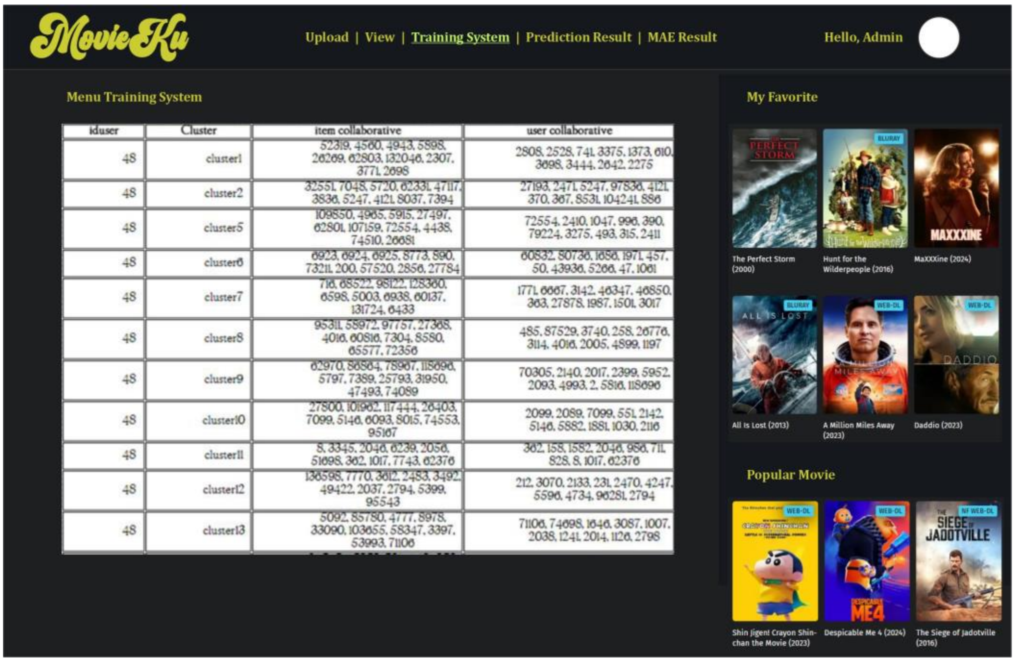

Figure 5, Training Result Website Display

Figure 6 shows an example of the predicted value results from user 48 in cluster 1 for the first 10 films with a combination of bisecting k-means and item-based collaborative filtering.

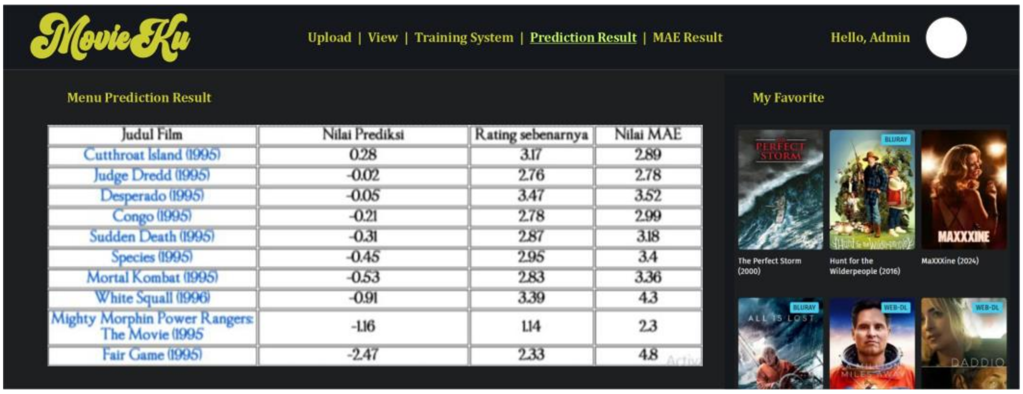

Figure 6. Prediction Result

Table 1 shows the results of the MAE value calculation after the testing process for all films in each cluster against 5 users with the specification of 3 users who have rated and 2 users who have never rated at all. The number of clusters in this test is 18 (according to the number of genres).

Tabel 1. MAE value for each cluster

Cluster	***MAE Item-based***	***MAE User-based***
1	1.18	1.33
2	1.60	1.35
3	1.46	1.48
4	1.58	1.78
5	1.72	1.64
6	1.73	1.83
7	1.26	1.54
8	1.90	1.49
9	1.84	1.82
10	2.07	1.67
11	2.56	2.00
12	1.71	1.40
13	1.47	1.29
14	1.87	1.90
15	1.76	1.54
16	1.60	1.55
17	2.31	2.43
18	1.40	1.38
Means	1.72	1.63

The average MAE value for each cluster on item-based is higher than user-based, meaning the combination of bisecting K-Means and user-based produces better recommendation values. The worst recommendation values occur in clusters 11 and 17 compared to other clusters. This is due to the poor distribution of user rating values, resulting in poor prediction values. Figure 7 shows the uneven distribution of rating values in the dataset for cluster 11, so that the error value in the recommendation system is relatively higher.

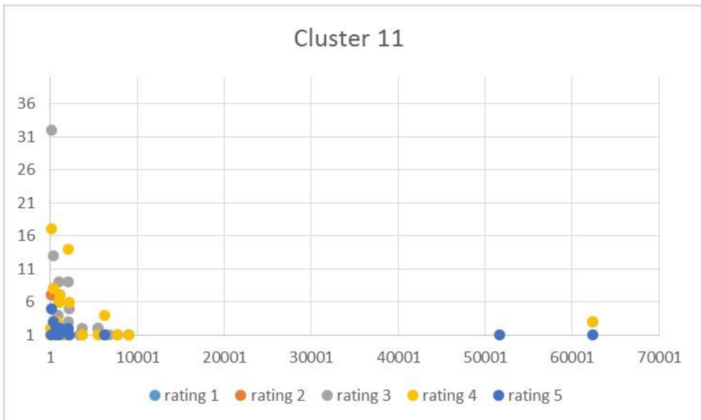

Figure 7. Uneven distribution of rating values in cluster 11

Figure 8 shows the distribution of rating values in the dataset for cluster 1 which is even so that the error value in the recommendation system is relatively lower.

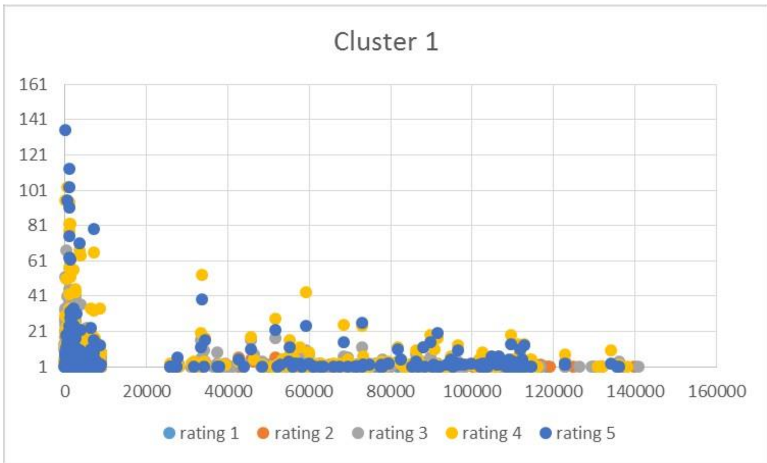

Figure 8. Even distribution of rating values in cluster 1

CONCLUTIONS

This research has produced a website-based movie recommendation system using a combination of K-Means bisecting and Collaborative Filtering algorithms. The website-based recommendation system that has been developed uses information from the Movie-Lens dataset. The error rate in the recommendation system has been calculated using the MAE value. The average MAE value of the combination of K-Means bisecting and user-based CF is 1.63, lower than the average MAE value of the combination of K-Means bisecting and item-based. In addition to the recommendation method, the distribution of rating values in the dataset also greatly affects the MAE value. This is also shown in clusters 11 and 17 with uneven distribution of rating values, which will result in a higher error value in the recommendation system.

REFERENCE

[1] F. Furtado and A. Singh, "Movie recommendation system using machine learning," *Int. J. Res. Ind. Eng.*, vol. 9, no. 1, pp. 84–98, 2020.

[2] A. S. Hapsari, M. Hanif, and others, "Motion graphic animation videos to improve the learning outcomes of elementary school students.," *Eur. J. Educ. Res.*, vol. 8, no. 4, pp. 1245–1255, 2019.

[3] R. H. Singh, S. Maurya, T. Tripathi, T. Narula, and G. Srivastav, "Movie recommendation system using cosine similarity and KNN," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 5, pp. 556–559, 2020.

[4] E. R. Agustian, E. P. Nugroho, and others, "Sistem Rekomendasi Film Menggunakan Metode Collaborative Filtering dan K-Nearest Neighbors," *JATIKOM J. Apl. dan Teor. Ilmu Komput.*, vol. 3, no. 1, pp. 18–21, 2020.

[5] A. Saputra, M. Aziz, A. Rosadi, F. I. Hakim, and P. Rosyani, "Systematic Literature Review: Penerapan Metode Forward Chaining dalam Sistem Rekomendasi Film," *JRIIN J. Ris. Inform. dan Inov.*, vol. 1, no. 2, pp. 320–323, 2023.

[6] A. Zakharia, A. D. Ulhaq, A. B. Suryono, N. C. Nugroho, D. F. Hafith, and N. D. A. Gusmao, "Sistem Rekomendasi Film Indonesia Menggunakan Metode Content-Based Filtering," *Log. J. Ilmu Komput. dan Pendidik.*, vol. 2, no. 4, pp. 671–678, 2024.

[7] H. Tahmasebi, R. Ravanmehr, and R. Mohamadrezaei, "Social movie recommender system based on deep autoencoder network using Twitter data," *Neural Comput. Appl.*, vol. 33, no. 5, pp. 1607–1623, 2021.
[8] S. Salloum and D. Rajamanthri, "Implementation and evaluation of movie recommender systems using collaborative filtering," *J. Adv. Inf. Technol.*, vol. 12, no. 3, 2021.
[9] T. Singh, A. Nayyar, and A. Solanki, "Multilingual opinion mining movie recommendation system using RNN," in *Proceedings of first international conference on computing, communications, and cyber-security (IC4S 2019)*, 2020, pp. 589–605.
[10] T. Widiyaningtyas, I. Hidayah, and T. B. Adji, "User profile correlation-based similarity (UPCSim) algorithm in movie recommendation system," *J. Big Data*, vol. 8, no. 1, p. 52, 2021.
[11] S. Airen and J. Agrawal, "Movie recommender system using parameter tuning of user and movie neighbourhood via co-clustering," *Procedia Comput. Sci.*, vol. 218, pp. 1176–1183, 2023.
[12] K. E. Permana, A. B. Rahmat, D. Ardianto, and D. A. Wicaksono, "Madura island tourism recommendation system based on collaborative filtering using convolutional neural network," in *AIP Conference Proceedings*, 2024.
[13] S. S. Choudhury, S. N. Mohanty, and A. K. Jagadev, "Multimodal trust based recommender system with machine learning approaches for movie recommendation," *Int. J. Inf. Technol.*, vol. 13, pp. 475–482, 2021.
[14] T. Anwar and V. Uma, "Comparative study of recommender system approaches and movie recommendation using collaborative filtering," *Int. J. Syst. Assur. Eng. Manag.*, vol. 12, pp. 426–436, 2021.
[15] Z. Z. Darban and M. H. Valipour, "GHRS: Graph-based hybrid recommendation system with application to movie recommendation," *Expert Syst. Appl.*, vol. 200, p. 116850, 2022.
[16] Q. Zhong, "Design of game-based collaborative learning model," *Open J. Soc. Sci.*, vol. 7, no. 7, pp. 488–496, 2019.