OPEN ACCESS
Indo-JC Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 7, Issue. 2, August 2022. pp. 45-52
doi:10.34818/indojc.2022.7.2.649

Music Recommender System Using K-Nearest Neighbor and Particle Swarm Optimization

Randika Dwi Maulana Rasyid [#1], Z. K. A. Baizal [#2*]

 # School of Computing, Telkom University
Jl. Telekomunikasi No. 1, Bandung 40257, Indonesia

[1] randikadw@student.telkomuniversity.ac.id
[2] baizal@telkomuniversity.ac.id

Abstract

In this day, users can listen to music anytime digitally and access them through the already available applications. A music recommender system is needed to help users choose music according to their interests and find music to listen to. K-Nearest Neighbor (KNN) is a popular method used in Collaborative Filtering (CF). In many studies, CF with the KNN method has been widely used, but it does not provide good performance. Thus, in this study, we use KNN, which will be optimized using Particle Swarm Optimization (PSO), which can improve the performance of the results obtained against the method used. System testing is done by comparing the performance of the KNN algorithm with the optimization results of KNN-PSO with several variables being observed, including the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values. The results of these recommender will predict the rating value where the KNN method gives MSE 0.07 and RMSE 0.27 while the KNN-PSO method gives MSE 0.04 and RMSE 0.20.

Keywords: Recommender System, *K-Nearest Neighbor*, *Particle Swarm Optimization*

Abstrak

Saat ini pengguna dapat mendengarkan musik kapan saja secara digital dan dapat mengakses melalui aplikasi yang sudah tersedia. Dengan demikian dibutuhkan sistem pemberi rekomendasi musik untuk membantu pengguna dalam memilih musik sesuai minat dan menemukan musik yang akan didengar. *K-Nearest Neighbor (KNN)* adalah algoritma yang populer digunakan dalam *Collaborative Filtering (CF)*. Dalam banyak penelitian, *CF* dengan metode *KNN* sudah banyak digunakan, namun kurang memberikan kinerja yang baik. Dengan demikian, pada penelitian ini kami menggunakan *KNN* yang akan dioptimasi menggunakan *Particle Swarm Optimization (PSO)* yang mana dapat meningkatkan performansi hasil yang didapatkan terhadap metode yang digunakan. Pengujian sistem dilakukan dengan membandingkan performasi antara algoritma KNN dengan hasil optimasi KNN-PSO dengan beberapa variabel yang diamati diantaranya nilai *Mean Squared Error (MSE)* dan *Root Mean Squared Error (RMSE)*. Hasil rekomendasi tersebut akan di prediksi nilai rating dimana metode *KNN* memeberikan nilai *MSE* 0.07 dan *RMSE* 0.27 sementara metode *KNN-PSO* meberikan nilai *MSE* 0.04 dan *RMSE* 0.20.

Kata Kunci: Sistem Pemberi Rekomendasi, *K-Nearest Neighbor*, *Particle Swarm Optimization*

I. Introduction

Technology development in the music field is growing fast, where users can now listen to music digitally and easily access music through available applications. For users, efficiently finding the music of their choice from a large amount of music data can increase user's interest [1]. However, various issues arise related to music

*Corresponding Author
Received on April, 2022. Accepted on July, 2022

services where users need facilities to help find and enjoy the music that users want [2]. Currently, the recommended music is usually popular music being played on the radio and music applications so that users know what song trends are being played. The music application also has an extensive collection of songs from various artists, song titles, and genres. However, some still have difficulty choosing songs according to user tastes. Therefore, a recommender system was built to get the best results in providing music recommender.

The recommender system is a system of recommending to the user to determine the item that the user wants. Several methods can be used to build a recommender system, such as Collaborative Filtering, Content-based Systems, Knowledge-based Recommenders, and Hybrid Recommenders [3]. The system was built using the CF approach as a recommender system where CF will predict user preferences based on past behavior and preferences of similar users [4]. One of the applications in CF is using the KNN algorithm, which will be analyzed in this study.

This study uses the KNN algorithm because it will calculate the similarity value between items and calculate rating predictions. The KNN algorithm looks for k groups in the data where in this study, the music history data that the user has heard will be calculated the similarity between items and will produce music recommendations to the user[5]. The results of these recommendations are predicted by the rating value from the user using the KNN algorithm, and finally, the performance value of the system will be measured.

In previous studies, the use of the KNN algorithm in the recommendation system has often been used and tested. Therefore, this study will use the KNN algorithm, which is re-optimized using the PSO algorithm. PSO is an optimization algorithm used in making decisions in recommending music. The advantage of this algorithm is that it results in speed in solving optimization problems and getting optimal results[6]. From optimization using PSO, it can be seen whether the optimization algorithm can provide optimal recommendation results compared to using only KNN. The last test is to analyze whether PSO can improve the performance of the KNN algorithm. In addition, we analyzed the PSO method to find other ways of providing recommendation results.

II. Literature Review

A. Recommender System

A recommender system is a technique or system for recommending and suggesting a particular product, service, or item. By this recommender system, it can help users in choosing the correct item [3]. The recommender system will help users find interesting and relevant products from the available products [7]. The application is the use of a recommender system that the system predicts existing items for recommendations, such as recommendations for music, books, films, and others that can attract users' attention. It can suggest whether the user will prefer the item based on their history and profile [7].

Generally, the recommender system has four approaches: collaborative filtering, content-based systems, knowledge-based recommenders, and hybrid recommenders [3]. Each method has advantages and disadvantages, but collaborative filtering and content-based systems are the most commonly used methods. In the implementation of this final project, the approach uses CF.

B. Collaborative Filtering (CF)

Collaborative filtering is one technique for recommending an item. The process of CF is to provide recommenders based on user behavior or habits that aim to make users who have the same habit of predicting which items are similar and liked by users. This CF technique is based on past behavior, so this technique is widely known and valuable as a recommender system [8].

Rating is an essential element of the application of the CF method, and the rating is obtained from some users giving an implicit or explicit assessment of an item. An explicit rating is when a user gets a request to give an opinion directly on an item. The implicit rating means the system will automatically obtain user preferences by monitoring user actions. One of the algorithms in the application of the CF method is KNN.

C. K-Nearest Neighbor (KNN)

KNN is a classification algorithm that can classify data values based on similarity or proximity to other values. This value is the point between the data separated into several clusters in making a new sample[9]. KNN can be implemented in the application of the CF system with item-based and user-based methods by calculating similarity values to find the similarity value between users.

In calculating the similarity value or distance between values, similarity has several algorithmic approaches, such as Cosine Similarity and Pearson Correlation[9]. Cosine Similarity can measure the similarity between two values whose value is a vector, then compare the length of the vector in the calculation of Cosine Similarity[10]. Pearson Correlation calculates values such as Cosine Similarity, but the difference is normalized first so that it does not become an inequality in the data [9].

In this study, we use a music dataset as a recommendation system development using the KNN algorithm where the similarity values between items will be searched, and rating predictions will be found. Calculating user similarity is essential in the KNN-based CF algorithm [5].

D. Particle Swarm Optimization (PSO)

PSO is an optimization algorithm method that can determine parameters based on population. PSO simulates the social behavior of animals such as flocks of birds and fish where each flock cooperatively finds food, and each flock member continues to change their foraging pattern according to their own learning experiences and other members of the herd [11]. The PSO method has a technique similar to the genetic algorithm (GA), but the difference is that each particle or solution contains velocity, acceleration, and position. Velocity and acceleration here are responsible for changing the particle's position to explore the space of the possible solutions[12]. In addition, the swarm in the PSO does not limit the entire movement but also continuously searches for optimal solutions in the possible solution space[11].

Several previous studies have used the recommendation system with the KNN algorithm widely. However, the use of the KNN algorithm alone still does not provide exemplary performance in the design of the recommendation system. Therefore, this study makes a music recommendation system using the KNN algorithm, which is optimized using the PSO algorithm. Finally, the performance of the PSO algorithm will be calculated and compared with the KNN algorithm only.

III. Research Method

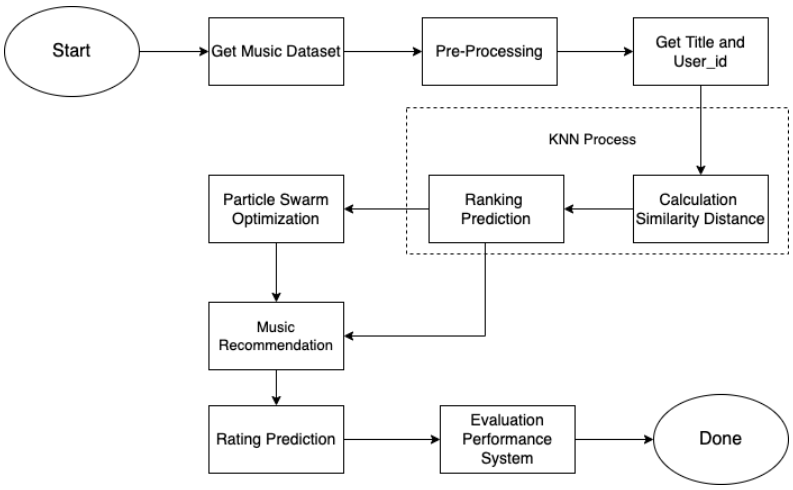

Fig 1. Design System

Based on Fig 1, the system design has several stages in the application of music recommender. The main stages are retrieval of music datasets, calculation of similarity distance, optimization of PSO, and evaluation of system performance. This research used user_id to search for music recommenders and calculate rating predictions.

A. Dataset

Dataset retrieval in constructing this music recommendation system from Million Dataset Song. The dataset is provided free of charge to obtain popular music data, which aims to provide access to music data and can encourage the research needed[13]. We took this dataset because it supports our research in searching music databases. The first music dataset in Table I contains User-id, Song-id, and Listening Count data. The second music dataset in Table II contains Song-id, Title, Release, Artist Name, and Year data. The two datasets will be processed in the music recommender system, which will be combined in the following pre-processing process.

TABLE I. First Music Dataset

User-Id	***Song-Id***	***Listenting Count***
256	3	5
256	4	2

TABLE II. Second Music Dataset

Song-Id	***Title***	***Release***	***Artist Name***	***Year***
12	Red Right Hand	Crying Lightning	Arctic Monkey	2009
256	One More Mystery	The Lost Album	Lewis Taylor	2004

B. Pre-processing

Data pre-processing is essential to getting efficient data quality[14]. In this study, several stages of pre-processing support the recommender process later. The first stage is to combine the two datasets into one data frame. Because both datasets have a song_id column, use the song_id column to equate the two data frames as in Table III.

TABLE III. Merge Dataframe

User_id	***Song_id***	***Listening Count***	***Title***	***Release***	***Artist Name***	***Year***
12	12	6	Red Right Hand	Crying Lightning	Arctic Monkey	2009
256	14	5	One More Mystery	The Lost Album	Lewis Taylor	2004
45	15	3	Fireflies	Ocean Eyes	Owl City	2009

Then the next stage in pre-processing is selecting the data frame because not all users listen to all songs. The selection is by selecting the number of users who have heard at least twenty songs that have been heard. This selection aims to calculate the sparse matrix on the pivot table made in Table IV. In the pivot table, a zero value indicates that the user has never heard the song.

TABLE IV. Pivot Table

	User_id		
Song_id	**1**	***2***	***3***
1	0	1	0
2	1	0	0
3	1	1	0

C. KNN Process

In this study, the implementation of the use of the KNN algorithm is carried out in two stages. Where the first stage is to find the calculation of the similarity value between music from the music history that the user has heard. The value of k is the value of the neighbor determination whose similarity value will be calculated. The value of k is used using the default parameter of the NearestNeighbor Sklearn Library with a value of k = 5. Then the second stage, from the results of the music recommendations given to the user, is given a rating prediction using KNN again, which will be tested by collecting test data and training data.

The KNN calculation is implemented using the similarity calculation formula to find the similarity value between users or items to be tested[9]. The system takes a history of music that has been heard by the user, where the results of the history are used to find the rating value and are tested using test data and training data[15]. The calculation of the KNN value in finding the prediction value of the rating is based on (1).

$$\hat{r}_{ui} = \frac{\sum_{v \in N_i^k(u)} sim(u,v) \cdot r_{vi}}{\sum_{v \in N_i^k(u)} sim(u,v)} \quad (1)$$

D. PSO Optimization Process

PSO optimization is carried out by providing the best parameters, which will later be rerun using the NeirestNeighbor method. The process parameters were obtained from the PSO optimization calculation and the predicted distance response[16]. The setting selected is to produce the closest predictive value between items.

In looking for parameters, PSO looks for the optimal solution by using the KNN algorithm, and the best k value is obtained for determining neighbor values which will calculate the similarity value between music with a value of k = 23 and from that value it is calculated again and gets the results of music recommendations to the user.

E. Performance Evaluation

Measuring the performance of the system that has been built is essential because it is a form of evaluating the performance of the algorithm that has been chosen, for researchers who want to improve machine learning algorithms on different problem domains and compare test results with others to understand performance metrics and choose the most appropriate performance from several possibilities [17].

Measurement of performance in this study using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) because, in this study, the user rating value was calculated with the rating prediction value as the evaluation value between the KNN algorithm and the KNN-PSO method. MSE is the average square of the difference between the observed value and the predicted value, where the variable value here is the predicted rating value from the user. In addition, RMSE calculates a more significant difference to assess errors in ranking predictions [18]. Both MSE and RMSE calculations calculate the predicted value results.

The calculation of the MSE and RMSE values can be seen in formulas (2) & (3), where the value of N is the total number that is rated, y_i is the rating value of the predicted results in the list, and $\hat{y}$ is the rating value of the actual results in the list.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2 \quad (2)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \quad (3)$$

IV. Result And Discussion

We divide the results of the design system that have been built into two groups: the results of recommender using the KNN method and the results of recommendations using the KNN-PSO method.

A. Result of Recommendation with KNN and KNN-PSO

Aa The first stage is to carry out the results of the recommendations for the KNN and KNN-PSO methods, where the first step is to confirm the User_id input contained in the data frame as in Fig 2, which will be given other music recommendations according to the proximity between music that has been heard before.

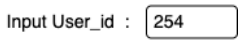

Fig 2. Illustration of Input user_id

It is known that User_id 254 has listened to music 28 times from the user's listener history. From the music history, a recommendation search was made using the KNN method, and the results obtained were 108 music title recommendations, which can be seen in Table V. The results from the KNN-PSO method obtained 431 music title recommendations, which can be seen in Table VI. These results prove that the KNN algorithm with PSO optimization can improve the performance of recommenders and provide more recommendation results to users.

TABLE V. Recommendation Result using KNN

No	Title	Artist Name
1	It's My Life	Bon Jovi
2	The King Of Carrot Flowers Parts Two And Three	Neutral Milk Hotel
3	Valentine	Justice
…	…	…
106	Burn	Nine Inch Nails
107	Shut Your Eyes	Snow Patrol
108	Superstition	The Kills

TABLE VI. Recommendation Result using KNN-PSO

No	Title	Artist Name
1	It's My Party	Lesley Gore
2	The King Of Carrot Flowers Parts Two And Three	Neutral Milk Hotel
3	Under The Gun	The Killers
…	…	…
429	Black Ballon	The Kills
430	Spring Nicht	Tokio Hotel
431	Starry Eyed Surprise	Shifty

The rating prediction stage will be continued from the recommendations shown to the user, where the results will be tested with test data and training data. The last step will be the system performance using the MSE and RMSE parameters.

B. Comparison of MSE and RMSE Value

TABLE VII. Calculation Evaluation

Calculation Evaluation		
Methode	***MSE***	***RMSE***
KNN	0.07	0.27
KNN-PSO	**0.04**	**0.20**

The evaluation of the calculations obtained using the KNN method gives an MSE value of 4.48 and an RMSE of 2.54. The calculation of the KNN-PSO method gives an MSE value of 1.70 and an RMSE of 1.30. The calculation evaluation results show that the use of the optimized KNN method using PSO is suitable for predicting user ratings and can improve the performance of the results obtained.

V. CONCLUSION

Based on the results of testing and analysis of the development of a music recommender system using KNN and KNN-PSO, the KNN-PSO method gives better performance results than the KNN method. The KNN-PSO method can also provide the best parameter setting values and increase the rating prediction value for recommended items.

The results show the prediction value of the rating where the KNN method gives an MSE value of 0.07 and an RMSE value of 0.27. Meanwhile, the KNN-PSO method gives MSE 0.04 and RMSE 0.20 values, so applying the application of the KNN-PSO method can improve rating predictions.

In future research, improve the recommender system's performance with a larger music dataset with different parameters. Then combined using different optimization methods to make a recommender system.

ACKNOWLEDGMENT

The author would like to thank the author's lecturer Mr. Baizal and friends who have supported the author in designing this recommender system.

References

[1] J. H. Su, C. Y. Chin, Y. W. Liao, H. C. Yang, V. S. Tseng, and S. Y. Hsieh, "A Personalized Music Recommender System Using User Contents, Music Contents and Preference Ratings," *Vietnam Journal of Computer Science*, vol. 7, no. 1, pp. 77–92, Feb. 2020, doi: 10.1142/S2196888820500049.

[2] A. Gatzioura, J. Vinagre, A. M. Jorge, and M. Sanchez-Marre, "A Hybrid Recommender System for Improving Automatic Playlist Continuation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 1819–1830, May 2021, doi: 10.1109/TKDE.2019.2952099.

[3] Z. K. A. Baizal,., D. H. Widyantoro, & N. U. Maulidevi, "Computational model for generating interactions in conversational recommender system based on product functional requirements". *Data & Knowledge Engineering*, 128, 101813. 2020, doi: 10.1016/j.datak.2020.101813

[4] Z. A. Baizal,., D. H. Widyantoro, & N. U Maulidevi, "Design of knowledge for conversational recommender system based on product functional requirements". In *2016 international conference on data and software engineering (ICoDSE)* (pp. 1-6). IEEE, October 2016

[5] S. Airen and J. Agrawal, "Movie Recommender System Using K-Nearest Neighbors Variants," *National Academy Science Letters*, vol. 45, no. 1, pp. 75–82, Feb. 2022, doi: 10.1007/s40009-021-01051-0.

[6] R. Katarya and O. P. Verma, "Efficient music recommender system using context graph and particle swarm," *Multimedia Tools and Applications*, vol. 77, no. 2, pp. 2673–2687, Jan. 2018, doi: 10.1007/s11042-017-4447-x.

[7] B. Patel, P. Desai, and U. Panchal, "Methods of Recommender System: A Review," 2017.

[8] Y. Afoudi, M. Lazaar, and M. al Achhab, "Collaborative filtering recommender system," in *Advances in Intelligent Systems and Computing*, 2019, vol. 915, pp. 332–345. doi: 10.1007/978-3-030-11928-7_30.

[9] I. Yoshua, H. Bunyamin, and S. Si, "Pengimplementasian Sistem Rekomendasi Musik Dengan Metode Collaborative Filtering," 2021.

[10] R. Samuel, R. Natan, Fitria, and U. Syafiqoh, "Penerapan Cosine Similarity dan K-Nearest Neighbor (K-NN) pada Klasifikasi dan Pencarian Buku," *Journal of Big Data Analytic and Artificial Intelligence, Vol. 1, No. 1*, 2018.

[11] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: an overview," *Soft Computing*, vol. 22, no. 2, pp. 387–408, Jan. 2018, doi: 10.1007/s00500-016-2474-6.

[12] J. C. Bansal, "Particle swarm optimization," in *Studies in Computational Intelligence*, vol. 779, Springer Verlag, 2019, pp. 11–23. doi: 10.1007/978-3-319-91341-4_2.

[13] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere, "The Million Song Dataset. THE MILLION SONG DATASET," 2011. [Online]. Available: http://code.google.com/p/pyechonest/

[14] W. S. Bhaya, "Review of Data Preprocessing Techniques in Data Mining," 2017.

[15] R. Samuel, R. Natan, Fitria, and U. Syafiqoh, "Penerapan Cosine Similarity dan K-Nearest Neighbor (K-NN) pada Klasifikasi dan Pencarian Buku," 2018.

[16] A. Sateria, I. D. Saputra, and Y. Dharta, "Penggunaan Metode Particle Swarm Optimization (PSO) Pada Optimasi Multirespon Gaya Tekan dan Momen Torsi Penggurdian Material Komposit Glass Fiber Reinforce Polymer (GFRP) Yang Ditumpuk Dengan Material Stainless Steel (SS)," 2018.

[17] G. Canbek, T. T. Temizel, S. Sagiroglu, and N. Baykal, *Binary Classification Performance Measures/Metrics: A Comprehensive Visualized Roadmap to Gain New Insights*. 2017.

[18] T. Silveira, M. Zhang, X. Lin, Y. Liu, and S. Ma, "How good your recommender system is? A survey on evaluations in recommendation," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 5, pp. 813–831, May 2019, doi: 10.1007/s13042-017-0762-9.