Imputation missing value to overcome sparsity problems

RZ Abdul Aziz[1], Sri Lestari[1], Fitria[1], Febri Arianto[2]

[1]Department of Bachelor of Informatics Engineering, Faculty of Computer Science, Institute of Informatics and Business Darmajaya, Bandar Lampung, Indonesia
[2]Department of Master of Informatics Engineering Study Program, Faculty of Computer Science, Institute of Informatics and Business Darmajaya, Bandar Lampung, Indonesia

Article Info

ABSTRACT

Collaborative filtering (CF) is a method to be used in recommendation systems. CF works by analyzing rating data patterns from previous users to produce recommendations according to their interests. However, it faces a crucial problem, sparsity, a condition where a lot of data is empty, which will affect the quality of the recommendations produced. To state this problem, the purpose of this study is to input methods including mean, min, max, and k-nearest neighbor imputation (KNNI). The steps taken include imputation of empty data, followed by similarity calculations using the cosin similarity method, and evaluation using root mean square error (RMSE). The experimental result shows that the mean method is excellent with an average similarity value of 0.99 and an RMSE value of 0.98.

Corresponding Author:

Sri Lestari
Department of Bachelor of Informatics Engineering, Faculty of Computer Science
Institute of Informatics and Business Darmajaya
ZA. Pagar Alam St., No. 93 Gedong Meneng, Bandar Lampung, Indonesia
Email: srilestari@darmajaya.ac.id

1. INTRODUCTION

Information technology provides suitability both obtaining various information and disseminating information. Information technology can be used to improve services. Optimal service is very important for companies and agencies. Therefore, many organizations implement it to make this happen. One of these media is an online system such as e-commerce. E-commerce will provide product or service recommendations according to the user's interests. The recommendation system is worked by searching for the most relevant information to the user's interests obtained from large amounts data. Furthermore, it is to produce suitable recommendations. The recommendation system is to reduce users' effort and time in searching for information that suits their interests [1]. The most relevant recommendations can be made with content-based filtering [2]–[4] demographic filtering [5], [6], collaborative filtering (CF) [7]–[9], and hybrid filtering [2], [10]–[12]. CF is a successful method and widely used in recommendation systems [13].

This method works by analyzing rating data patterns to make predictions. Moreover, CF method is a simple and efficient method [8]. CF is further divided into two categories, namely memory-based approach and model-based approach [1], [14]. The memory-based approach generally performs well with dense data, whereas with sparse data (sparsity) the approach is less reliable. Sparsity is a condition where data is not fully filled in or is sparse. The users cannot give an expected rating for the item. Apart from that, there are new users who don't like the item or leave the item without giving a rating. If the condition rating data is sparse, it is difficult to determine similarities between users, so the quality of the resulting recommendations is low [15].

Several studies have been carried out to overcome the sparsity problem. Alhijawi *et al.* [16] proposes a statistics-based method by utilizing a user-item rating matrix and an item-feature matrix to build a user interest print (UIP) matrix. UIP is a filled (dense) matrix that stores data on the level of user satisfaction with an item's semantic feature [16]. Furthermore, Yu [17] proposed an item CF algorithm based on attribute similarity. Mohamed *et al.* [15] using cluster-based association rules to solve sparsity problems and increase accuracy. Meanwhile, this study uses imputing missing values to overcome the sparsity problem in CF. Data with a lot of blanks will be filled in with several missing value imputation techniques, namely mean, min, max and k-nearest neighbor (KNN) imputer to replace ratings with empty values (NaN). By imputation missing values, the sparsity problem can be overcome so that the quality of recommendations can be improved.

2.   METHOD

This study was carried out in several stages of activities from literature study, data collection, data preprocessing, followed by imputation and cosine similarity techniques, and evaluation was carried out using root mean square error (RMSE), as in Figure 1.
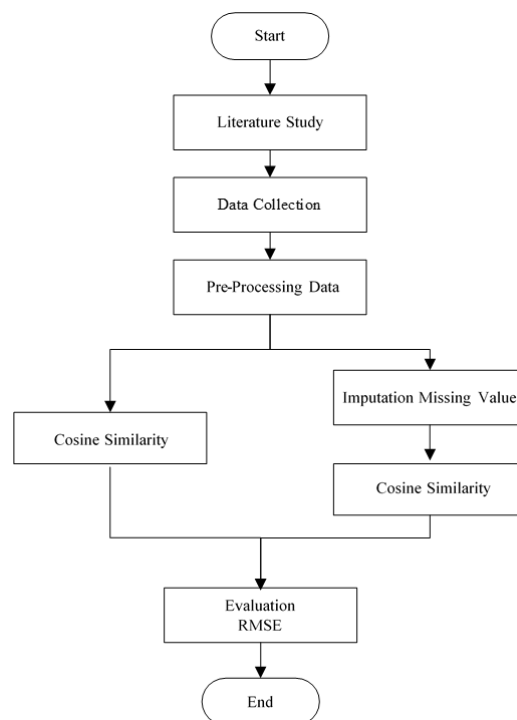

Figure 1. Flowchart of research stages

− Literature study

At this stage, this is done by looking for references from various sources such as books, journal articles, proceeding articles, and the internet related to recommendation systems, CF, various imputation techniques and others.
− Data collection

The used data in this study was obtained from the Kaggle website, a website that provides the data needed by data scientists and the data was obtained in the form of a comma-separated values (CSV) file. The used dataset is the MovieLens dataset. MovieLens is a data collection that is most often used in research, MovieLens stores information related to users and movies [18]. The MovieLens 100 K dataset has 100,000 ratings given by 943 users to 1682 movies on a scale of 1-5, so this data contains a sparsity of 93.7% [19]–[21].
− Pre-processing data

The preprocessing stage involves importing data from the CSV file, and continuing by combining the data between rating data, movie data and user data, and in the form of a pivot table.

− Imputation missing value

Using statistical or machine learning methods to estimate selected observation data to replace empty values. Missing value imputation analyzes patterns is missing as the data output of a classification model [22]. This study uses imputation techniques, namely mean, min, max, and KNN imputer [23]. The KNN imputation algorithm uses observations that have similar values, namely by determining the K value or the number of closest observations that will be used. The imputer KNN equation uses (1) [24].

$$d(x_a, x_b) = \sqrt{\sum_{j=1}^{n}(x_{aj} - x_{bj})} \tag{1}$$

This study uses the Python library to apply imputation techniques. The Python library has an approach to missing values by replacing missing data, with several methods of approximating the "mean" value, the method of approximating the maximum value "max", and the minimum value approach "min" [25].

− Cosine similarity

Cosine similarity is a method that works by calculating the level of similarity between two objects. This study will be used to compare similarities between users. Cosine similarity calculation uses (2).

$$(dj, q) = \frac{\sum_{i=1}^{t}(w_{ij}.w_{iq})}{\sqrt{\sum_{i=1}^{t} w_{ij}^2 \sum_{i=1}^{t} w_{iq}^2}} \tag{2}$$

− Evaluation RMSE

This study uses RMSE to evaluate the performance of CF from implementing imputation and cosine similarity techniques. RMSE is a performance evaluation to calculate the average value which has the squared difference between the actual value and the predicted rating value [26], [27]. RMSE calculates the power value rooted from the result. RMSE is a matrix that has a very high result matrix when a very high error is not desired [28]. RMSE calculation uses (3) [29].

$$RMSE = \sqrt{\sum_{i=1}^{n}(|\rho i - qi|)} \tag{3}$$

3. RESULTS AND DISCUSSION

This study utilizes the imputation method to fill sparsity in the rating data to be used for recommendations. The empty data on the rating value appears because every user does not give a rating value to all available items. Thus, it makes a difficulty for the system to provide recommendations.

− Data preprocessing

In this data preprocessing, it imports data using the Python module, namely Pandas, to read the data set from the CSV file. The data is divided into three files. The first is the rating data containing userId, movieId, and ratting. Then, the second data is the movie contains movieId, movieNames, and genres. The third file is user data, contains userID, gender, age, occupation, and zip. These three data are combined into one and produced the data in Table 1. Furthermore, it converts the movie rating data into a pivot table (Table 2) to get unfilled rating data from each user. From the pivot table data, it displays a lot of empty data or sparsity to carry out imputation in filling the data gaps.

Table 1. Rating movie data

Movie id	Movie_names	Genres	Userid	Rating
1	Toy story (1995)	Animation|children's|comedy	308	4
4	Waiting to exhale (1995)	Comedy|drama	308	5
5	Father of the bride part II (1995)	Comedy	308	4
7	Sabrina (1995)	Comedy|romance	308	4
8	Tom and huck (1995)	Adventure|children's	308	5

Table 2. Pivot table on rating movie data

Movie id	1	2	3	4	5
User id					
1	5	3	4	3	3
2	4	Nan	Nan	Nan	Nan
3	Nan	Nan	Nan	Nan	Nan
4	Nan	Nan	Nan	Nan	Nan
5	4	3	Nan	Nan	Nan

− Imputation

At this stage, it is done by using 4 methods, namely mean, max, min and k-nearest neighbor imputation (KNNI), to fill in the gaps in the pivot table of the movie rating data above. For imputation, it uses the modules available in Python, namely the Pandas module for mean, max and min imputation. Meanwhile, it uses the "sklearn.impute" KNN imputer module for KKN-imputation.

The mean imputation method in the Pandas data module is taken from the average value in each movie column in the userId row. Then, it is added up and filled for each empty column (NaN) and the results of the mean imputation method can be seen in Table 3, but the mean is still producing a decimal value. Meanwhile, the rating required is a rounded value of 1-5 in Table 4.

Table 3. Pivot table on imputation mean

User id	1	2	3	4	5
1	5	3	4	3	3
2	4	3,2061	3,0333	3,5502	3,3023
3	3,8783	3,2061	3,0333	3,5502	3,3023
4	3,8783	3,2061	3,0333	3,5502	3,3023
5	4	3	3,0333	3,5502	3,3023

Table 4. Pivot table on rounding the mean

User id	1	2	3	4	5
1	5	3	4	3	3
2	4	3	3	4	3
3	4	3	3	4	3
4	4	3	3	4	3
5	4	3	3	4	3

The min imputation method also uses the Pandas module. It is taken from the smallest value in each movie column in the userId row and filled in each empty column (NaN). Therefore, the smallest value of the rating is 1, for the min method imputation results can be seen in Table 5.

The imputation max method also uses the Pandas module, the same as the mean and min methods, but for this method the data is taken from the largest value in each movie column in the userId row and filled in each empty column (NaN), the largest value of the rating is 5 for the max method imputation results can be seen in Table 6. The KNN imputer imputation method also uses a module available in Python, it set the K value for this method to 1 to observe the closest value that will be used from an empty data column, the results of the KNN imputer imputation method can be seen in Table 7.

Table 5. Pivot table imputation min on data rating movie

User id	1	2	3	4	5
1	5	3	4	3	3
2	4	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1
5	4	3	1	1	1

Table 6. Pivot table imputation max on data rating movie

User id	1	2	3	4	5
1	5	3	4	3	3
2	4	5	5	5	5
3	5	5	5	5	5
4	5	5	5	5	5
5	4	3	5	5	5

Table 7. Pivot table imputation KNN imputer on rating movie data

User id	1	2	3	4	5
1	5	3	4	3	3
2	4	3	4	4	4
3	4	2	1	3	3
4	5	3	4	5	4
5	4	3	1	4	4

− Cosine similarity

The next step is to carry out cosine similarity calculations in each imputation results table and to determine the closeness of each imputation value and it can be seen in the following table. In Table 8, it is the results of cosine similarity calculations with mean imputation. From the calculation results, it can be seen an average value of 0.99.

Table 8. Mean method of cosine similarity

	0	1	2	3	4
0	0	0.990674	0.989397	0.990442	0.985293
1	0.990674	0	0.996176	0.997156	0.990484
2	0.989397	0.996176	0	0.99609	0.989424
3	0.990442	0.997156	0.99609	0	0.990772
4	0.985293	0.990484	0.989424	0.990772	0

Table 9 states the results of the cosine similarity calculation with min imputation. It can be seen that the average value is 0.88. In Table 10, it is the results of cosine similarity calculations with Max imputation. It can be seen an average value of 0.98.

Table 9. Min method of cosine similarity

	0	1	2	3	4
0	0	0.825976	0.823433	0.823778	0.846395
1	0.825976	0	0.920377	0.925455	0.868059
2	0.823433	0.920377	0	0.955368	0.883788
3	0.823778	0.925455	0.955368	0	0.881011
4	0.846395	0.868059	0.883788	0.881011	0

Table 10. Max method of cosine similarity

	0	1	2	3	4
0	0	0.984894	0.98253	0.98713	0.978187
1	0.984894	0	0.993891	0.997501	0.983568
2	0.98253	0.993891	0	0.995036	0.980453
3	0.98713	0.997501	0.995036	0	0.985677
4	0.978187	0.983568	0.980453	0.985677	0

In Table 11, it is the results of cosine similarity calculations with KNN imputer imputation. From the calculation results, it can be seen an average value of 0.92. Based on the results of the four-imputation data, the average similarity value data is in Table 12. The highest similarity value was shown in the mean imputation method with 0.99. it showed that Mean imputation can improve the quality of recommendations better than the min, max, and KNNI methods.

− The results of RMSE test

After carrying out imputation and calculating the similarity values, RMSE performance testing is carried out with results as in data Table 13.

Table 11. KNNI method of cosine similarity

	0	1	2	3	4
0	0	0.943055	0.923853	0.936458	0.937643
1	0.943055	0	0.93203	0.940095	0.93119
2	0.923853	0.93203	0	0.921892	0.917673
3	0.936458	0.940095	0.921892	0	0.924012
4	0.937643	0.93119	0.917673	0.924012	0

Table 12. Mean of cosine simlarity

Imputation	Mean	Min	Max	KNN-I
Mean of similarity	0.992469	0.889397	0.989933	0.927114

Table 13. The result of RMSE

	Without imputation	Mean	Min	Max
RMSE	1.10871	0.98213	1.15145	1.14454

The sparsity problem is a crucial problem because it will affect the quality of the recommendations provided. One technique to solve this problem is the missing value imputation technique. This research uses mean, min, max, and KKN-I imputation techniques and carries out evaluations by implementing cosin similarity and RMSE.

The test results showed that the excellent value from the mean imputation method with an RMSE value=0.98213. It is the smallest value compared to the others. Moreover, it showed that the resulting prediction value is better. In addition, it is supported by the highest similarity value, namely 0.99. Based on the evaluation results, it can be concluded that the imputation method, especially mean. It can overcome the sparsity problem in CF and it is able to improve the quality of recommendations.

4. CONCLUSION

The quality of recommendations is vital in personal service. Therefore, several problems that frequently occur, such as sparsity, must be resolved immediately. Sparsity is a rare data condition, so it will affect the quality of the recommendations produced. Therefore, this research proposes a missing value imputation method to solve this problem. The missing value imputation methods used in this research include min, max, and KNNI. Apart from that, it uses the cosin similarity and RMSE methods for evaluation. The experimental result shows that the mean imputation method is excellent for the min, max, and KNNI imputation methods with similarity value of 0.99 and the smallest RMSE value with 0.98. This shows that the mean imputation method can solve the sparsity problem and improve the quality of recommendations.

ACKNOWLEDGEMENTS

REFERENCES

[1]   D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *Journal of Big Data*, vol. 9, no. 1, p. 59, Dec. 2022, doi: 10.1186/s40537-022-00592-5.

[2]   Y. Afoudi, M. Lazaar, and M. Al Achhab, "Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network," *Simulation Modelling Practice and Theory*, vol. 113, p. 102375, Dec. 2021, doi: 10.1016/j.simpat.2021.102375.

[3]   M. Rojszczak, "Online content filtering in EU law – A coherent framework or jigsaw puzzle?," *Computer Law and Security Review*, vol. 47, p. 105739, Nov. 2022, doi: 10.1016/j.clsr.2022.105739.

[4]   J. A. Diaz-Garcia, M. D. Ruiz, and M. J. Martin-Bautista, "NOFACE: A new framework for irrelevant content filtering in social media according to credibility and expertise," *Expert Systems with Applications*, vol. 208, p. 118063, Dec. 2022, doi: 10.1016/j.eswa.2022.118063.

[5]   A. Yassine, L. Mohamed, and M. Al Achhab, "Intelligent recommender system based on unsupervised machine learning and demographic attributes," *Simulation Modelling Practice and Theory*, vol. 107, p. 102198, Feb. 2021, doi: 10.1016/j.simpat.2020.102198.

[6]   M. Sharma, B. Pant, and V. Singh, "Demographic profile building for cold start in recommender system: A social media fusion approach," *Materials Today: Proceedings*, vol. 46, pp. 11208–11212, 2021, doi: 10.1016/j.matpr.2021.02.428.

[7]   M. F. Hafidz and S. Lestari, "Solution to Scalability and Sparsity Problems in Collaborative Filtering using K-Means Clustering and Weight Point Rank (WP-Rank)," *Jurnal Rekayasa Sistem dan Teknologi Informasi*, vol. 7, no. 4, pp. 743–750, Aug. 2023, doi: 10.29207/resti.v7i4.4543.

[8]   H. Khojamli and J. Razmara, "Survey of similarity functions on neighborhood-based collaborative filtering," *Expert Systems with Applications*, vol. 185, p. 115482, Dec. 2021, doi: 10.1016/j.eswa.2021.115482.

[9]   W. Yue, Z. Wang, W. Liu, B. Tian, S. Lauria, and X. Liu, "An optimally weighted user- and item-based collaborative filtering approach to predicting baseline data for Friedreich's Ataxia patients," *Neurocomputing*, vol. 419, pp. 287–294, Jan. 2021, doi: 10.1016/j.neucom.2020.08.031.

[10]   K. Kobyshev, N. Voinov, and I. Nikiforov, "Hybrid image recommendation algorithm combining content and collaborative filtering approaches," *Procedia Computer Science*, vol. 193, pp. 200–209, 2021, doi: 10.1016/j.procs.2021.10.020.

[11]   Y. Zhang, Z. Liu, and C. Sang, "Unifying paragraph embeddings and neural collaborative filtering for hybrid recommendation," *Applied Soft Computing*, vol. 106, p. 107345, Jul. 2021, doi: 10.1016/j.asoc.2021.107345.

[12]   M. N and J. Samraj, "A hybrid convolutional neural network with long short-term memory (HCNN-LSTM) model based Edge System Recommendation(ESR) for cloud service providers," *Measurement: Sensors*, vol. 29, p. 100886, Oct. 2023, doi: 10.1016/j.measen.2023.100886.

[13]   H. Yan and Y. Tang, "Collaborative Filtering Based on Gaussian Mixture Model and Improved Jaccard Similarity," *IEEE Access*, vol. 7, pp. 118690–118701, 2019, doi: 10.1109/ACCESS.2019.2936630.

[14]   F. Horasan, A. H. Yurttakal, and S. Gündüz, "A novel model based collaborative filtering recommender system via truncated ULV decomposition," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, p. 101724, Sep. 2023, doi: 10.1016/j.jksuci.2023.101724.

[15]   M. H. Mohamed, M. H. Khafagy, H. Elbeh, and A. M. Abdalla, "Sparsity and cold start recommendation system challenges solved by hybrid feedback," *International Journal of Engineering Research and Technology*, vol. 12, no. 12, pp. 2735–2742, 2019.

[16]   B. Alhijawi, G. Al-Naymat, N. Obeid, and A. Awajan, "Mitigating the Effect of Data Sparsity: A Case Study on Collaborative Filtering Recommender System," in *2019 2nd International Conference on New Trends in Computing Sciences*, IEEE, Oct. 2019, pp. 1–6. doi: 10.1109/ICTCS.2019.8923064.

[17]   P. Yu, "Merging attribute characteristics in collaborative filtering to alleviate data sparsity and cold start," in *Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference*, IEEE, Mar. 2019, pp. 569–573. doi: 10.1109/ITNEC.2019.8729461.

[18]   A. Gonzalez, F. Ortega, D. Perez-Lopez, and S. Alonso, "Bias and Unfairness of Collaborative Filtering Based Recommender Systems in MovieLens Dataset," *IEEE Access*, vol. 10, pp. 68429–68439, 2022, doi: 10.1109/ACCESS.2022.3186719.

[19]   Z. Ding, Z. Qin, Q. X. Wang, and Z. G. Qin, "Random Group Recommendation Model Based on Fuzzy Clustering," *Journal of Electronic Science and Technology*, vol. 18, no. 2, p. 100054, Jun. 2020, doi: 10.1016/j.jnlest.2020.100054.

[20]   G. Jain, T. Mahara, and S. C. Sharma, "Performance Evaluation of Time-based Recommendation System in Collaborative Filtering Technique," *Procedia Computer Science*, vol. 218, pp. 1834–1844, 2022, doi: 10.1016/j.procs.2023.01.161.

[21]   G. Behera and N. Nain, "Collaborative Filtering with Temporal Features for Movie Recommendation System," *Procedia Computer Science*, vol. 218, pp. 1366–1373, 2022, doi: 10.1016/j.procs.2023.01.115.

[22]   K. Phiwhorm, C. Saikaew, C. K. Leung, P. Polpinit, and K. R. Saikaew, "Adaptive multiple imputations of missing values using the class center," *Journal of Big Data*, vol. 9, no. 1, p. 52, Apr. 2022, doi: 10.1186/s40537-022-00608-0.

[23]   M. Kowsher, N. J. Prottasha, A. Tahabilder, K. Habib, M. Abdur-Rakib, and M. Shameem Alam, "Predicting the Appropriate Mode of Childbirth using Machine Learning Algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, pp. 700–708, 2021, doi: 10.14569/IJACSA.2021.0120582.

[24]   A. Fadlil, Herman, and M. Dikky Praseptian, "Single Imputation Using Statistics-Based and K Nearest Neighbor Methods for Numerical Datasets," *Ingenierie des Systemes d'Information*, vol. 28, no. 2, pp. 451–459, Apr. 2023, doi: 10.18280/isi.280221.

[25]   N. Pandey, P. K. Patnaik, and S. Gupta, "Data Pre Processing for Machine Learning Models using Python Libraries," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 4, pp. 1995–1999, Apr. 2020, doi: 10.35940/ijeat.d9057.049420.

[26]   K. Lappalainen, M. Piliougine, and G. Spagnuolo, "Experimental comparison between various fitting approaches based on RMSE minimization for photovoltaic module parametric identification," *Energy Conversion and Management*, vol. 258, p. 115526, Apr. 2022, doi: 10.1016/j.enconman.2022.115526.

[27]   M. J. Mabula, D. Kisanga, and S. Pamba, "Application of machine learning algorithms and Sentinel-2 satellite for improved bathymetry retrieval in Lake Victoria, Tanzania," *Egyptian Journal of Remote Sensing and Space Science*, vol. 26, no. 3, pp. 619–627, Dec. 2023, doi: 10.1016/j.ejrs.2023.07.003.

[28] M. S. Kabul and E. B. Setiawan, "Recommender System with User-Based and Item-Based Collaborative Filtering on Twitter using K-Nearest Neighbors Classification," *Journal of Computer System and Informatics*, vol. 3, no. 4, pp. 478–484, Sep. 2022, doi: 10.47065/josyc.v3i4.2204.

[29] M. W. Liemohn, A. D. Shane, A. R. Azari, A. K. Petersen, B. M. Swiger, and A. Mukhopadhyay, "RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 218, p. 105624, Jul. 2021, doi: 10.1016/j.jastp.2021.105624.

BIOGRAPHIES OF AUTHORS

RZ Abdul Aziz 🆔 🇬 SC ⊘ he is an Associate Professor in the Graduate School of Informatics Engineering at Institute of Informatics and Business Darmajaya, Lampung, Indonesia. He received his Ph.D in Information Science and Technology from OSAKA University Japan. His research activities focus on artificial intelligence and machine learning, global optimation, component-based software engineering, quality software engineering, and statistical data analysis. He can be contacted at email: rz_aziz@darmajaya.ac.id.

Sri Lestari 🆔 🇬 SC ⊘ obtained her Doctorate (Dr from the Electrical Engineering Doctor Program, Universitas Gadjah Mada, Yogyakarta, Indonesia in 2019. She is a lecturer in the Department of Computer Science, Institute of Informatics and Business Darmajaya, Bandar Lampung, Indonesia. Her research interests include artificial intelligence, recommendation system, cf, data mining, decision support systems, and software engineering. Her representative published articles listed as follows: Porat^{Rank} to improve performance recommendation system (Lecture Notes in Electrical Engineering, Springer 2021), decision support system for service quality using SMART and fuzzy ServQual methods (JUITA: Jurnal Informatika, 2021), WP-rank: rank aggregation based CF method in recommender system (International Journal of Engineering and Technology (UAE), 2018), performance comparison of rank aggregation using borda and copeland in recommender system (International Workshop on Big Data and Information Security (IWBIS 2018)), NRF: normalized rating frequency for CF (The 2018 International Conference on Applied Information Technology and Innovation (ICAITI 2018)), design and analysis model application system teaching online media. (Proceedings of the International Conference on Information Technology and Business (ICITB), 2016). She can be contacted at email: srilestari@darmajaya.ac.id.

Fitria 🆔 🇬 SC ⊘ is a lecturer at the Institute of Informatics and Business Darmajaya, Department of Informatics Engineering, Lampung Indonesia. Obtained a Master's degree in Computers from Gajah Mada University, Yogyakarta with a Computer Science Study Program. His research activities focused on artificial intelligence, data mining algorithms, and image processing. She can be contacted at email: fitria@darmajaya.ac.id.

Febri Arianto 🆔 🇬 SC ⊘ currently continuing his studies in the Master of Informatics Engineering study program with a concentration in Data Science at IIB Darmajaya. Apart from that, he also work as IT DevOps at an online news company. He can be contacted at email: febriarianto464@gmail.com.