Restaurant Recommendation System Based on User Ratings with Collaborative Filtering

Achmad Arif Munaji\textsuperscript{1}, Andi Wahju Rahardjo Emanuel\textsuperscript{2}

\textsuperscript{1,2}Magister Informatika, Universitas Atma Jaya Yogyakarta, 44 Babarsari Street, Yogyakarta, Indonesia

E-mail: \textsuperscript{1}arief.munaji65@gmail.com, \textsuperscript{2}andi.emanuel@uajy.ac.id

Abstract. Recommendation systems are widely used as a reference in marketing products or businesses. The number of choices given sometimes makes a person confused in making choices. The recommendation system provides a solution in overcoming this. Collaborative filtering is a fairly popular algorithm used in recommendation systems that are based on references and information obtained from users. User rating is a variable that is widely used as a reference in establishing a recommendation system. User rating can influence other users in making their choice of the same product. The purpose of this study is to conduct an experiment in the application of a collaborative filtering algorithm and use the Pearson correlation function as a method used to see the level of user similarity in making a restaurant recommendation system. The methods used in this research are: (1) datasets preparation, (2) datasets filtering, (3) collaborative filtering, (4) pearson correlation, (5) recommendations, and (6) evaluation. Collaborative filtering combined with the Pearson correlation produces recommendations that are suitable for users based on characteristics or the same level of interest between users, making it easier for users to make choices. This is also indicated by a small error rate based on the RMSE results.

1. Introduction

In living everyday life, humans are always faced with a condition where they must make a choice. The many choices offered often make us confused in making choices. To overcome this, we often ask for suggestions or recommendations from other people. In the modern era, many problems like this have been overcome by implementing a recommendation system. The use of recommendation systems will often be encountered in marketing systems, such as online applications and online trading. This application is carried out with the aim of seeing potential customers and providing recommendations for a product related to the common interests of users [1]. Providing recommendations for certain products or goods will have a certain impact, such as increasing the level of popularity because the product or goods will often be seen by users and increasing the chances of the product or item being sold [2].

Collaborative filtering is an algorithm used to create a recommendation system. This algorithm is based on the following opinion: someone has the same views and interests, their preferences and interests are strong, and can make choices based on reference to their previous preferences [3]. Collaborative filtering uses certain aspects as a basis for making recommendations, such as user behavior in the past, items purchased by users, and ratings of items that have been purchased [4].

The purpose of this study is to conduct experiments in the application of collaborative filtering algorithms and use the Pearson correlation function as a method used to see the level of user similarity.
in making restaurant recommendation systems based on ratings given by visitors or users in the past. The stages of research carried out include: (1) datasets preparation, (2) datasets filtering, (3) collaborative filtering, (4) pearson correlation, (5) recommendations, and (6) evaluation.

2. Literature review
A recommendation system is a system designed for users with the aim of providing recommendations for a particular item. There are many areas where recommendation systems can be applied, such as social media, e-learning, web news articles, e-commerce, and many more [5]. Collaborative filtering is an algorithm used in making recommendations based on user personalization [6]. The use of collaborative filtering is very helpful in making recommendation systems on large datasets [7] [8]. In its implementation, the calculation of user similarity is carried out and by using this concept as a reference that users who have the same characteristics will tend to assess the same items as well. In collaborative filtering, the use of grouping techniques is done by grouping the same users into several groups [9]. After the similarity between items is obtained, collaborative filtering will calculate the predicted value of the missing data [10].

In measuring the value of similarity, there are many methods that can be used, such as Pearson Correlation, Spearman Rank, Discounted Similarity, and others. Previous research has tried to see how popular and good these methods are in measuring the similarity value. Based on the results of research that has been done, the Weighted Pearson Correlation produces a good accuracy value in prediction. In addition, the results of these studies also suggest that data differences also affect the choice of methods to be used [11]. The completeness of the data used is an important issue that must be considered for many researchers. Data sparsity is a problem related to data that is often found. A new collaborative filtering method is proposed to overcome this problem by adopting a clustering algorithm [12]. Improvements to the similarity equation used in calculating user portraits, item characteristics, and user activity data were proposed in previous studies with the aim of increasing the accuracy of the recommendation results [13]. Although Pearson Correlation is very well used to calculate similarity, previous studies have suggested that, in certain cases, the Pearson correlation shows inaccurate results [14].

3. Method
In this section, the author will explain the stages used in research. The stages of the methodology can be seen in Figure 1.

![Figure 1. Research methods.](image-url)
As shown in Figure 1, the research methodology used in this study consists of six stages. These stages include: datasets collection, datasets filtering, collaborative filtering, pearson correlation, recommendation, and evaluation. The following is an explanation regarding the stages of the research methodology:

3.1. Datasets preparation
At this stage the authors collect a dataset that will be used as research training data. The dataset used is a dataset obtained from one of the dataset provider websites kaggle.com [15]. There are two datasets used in this study, namely the restaurant dataset and the rating dataset.

3.2. Datasets filtering
At this stage, the author will filter the content contained in each dataset. This is done because not all of the content on the dataset will be used during the recommendation making process and lighten when the process takes place. The following is a brief overview of the dataset content used in table form:

| placeID | name                          | address                                      |
|---------|-------------------------------|----------------------------------------------|
| 134999  | Kiku cuernavaca               | Revolucion                                   |
| 132825  | Puesto de tacos               | Esquina santos degollado y leon guzman       |
| 135106  | El rincn de san Francisco     | Universidad 169                              |
| 132667  | Little pizza emilio portes    | calle emilio portes gil                      |
| 132613  | Carnitas mata                 | lic. Emilio portes gil                       |

Table 1 displays the contents of the restaurant dataset that will be used and is the result of the selection that has been done. The five data rows represent the first five rows contained in the dataset. In total, the restaurant dataset contains 130 lines and is a restaurant in Mexico. The following is a brief explanation of the contents of the dataset: the placeID column contains the restaurant location id, the name column contains the name of the restaurant, and the address column contains the address the restaurant is located.

| userID | placeID | rating |
|--------|---------|--------|
| U1077  | 135085  | 2      |
| U1077  | 135038  | 2      |
| U1077  | 132825  | 2      |
| U1077  | 135060  | 1      |
| U1068  | 135104  | 1      |

Table 2 displays the ranking dataset that will be used in this study. Similar to the dataset in Table 1, the Table 2 dataset is also a selected table. Table 2 represents the first five rows in the ranking dataset. The ranking dataset contains 1161 rows and three columns. The userID column contains the user id, the placeID column contains the restaurant location id, and the rating column contains the rating given by the user.

3.3. Collaborative filtering
At this stage, the authors apply the algorithm in making a recommendation system. The recommendation system will be created using collaborative filtering based on filtering users. This technique uses other users as a guide in recommending an item to user input. Search for users who have the same reference or opinion as input data, then provide recommendations for the appropriate
item to the input. Following are the steps for collaborative filtering in making a restaurant recommendation system based on user rating references:

- Select users with restaurants the user has visited.
- Based on the ratings given for each restaurant, find the top X neighbors.
- Get a recording of the restaurant users have visited for each neighbor.
- Calculate restaurant similarity scores using the Pearson Correlation Function.
- Provides restaurant recommendations based on the highest score.

Table 3. Experimental data.

| No | x   | y   | \((x_i - \bar{x})(y_i - \bar{y})\) | \((x_i - \bar{x})^2\) | \((y_i - \bar{y})^2\) |
|----|-----|-----|-----------------------------------|-----------------|-----------------|
| 1  | 3   | 5   | -0.224489796                      | 0.020408163     | 2.469387755     |
| 2  | 4   | 2   | -1.224489796                      | 0.734693878     | 2.040816327     |
| 3  | 5   | 3   | -0.795918367                      | 3.448979592     | 0.183673469     |
| 4  | 1   | 4   | -1.224489796                      | 4.591836735     | 0.326530612     |
| 5  | 3   | 5   | -0.224489796                      | 0.020408163     | 2.469387755     |
| 6  | 4   | 4   | 0.489795918                       | 0.734693878     | 0.326530612     |
| 7  | 2   | 1   | 2.775510204                       | 1.306122449     | 5.897959184     |
| sum|     |     | -0.428571429                      | 10.85714286     | 13.71428571     |

\[ r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \]

3.4. Pearson correlation
At this stage, the author also uses the Pearson Correlation Function as a method to compare all similar users and find which users are most similar. In its application, the Pearson Correlation will later calculate the linear relationship between the two variables:

For example, there are two variables x and y, where each of the two variables represents a rating given by two users and has a value of n. \(\bar{x}\) as the mean of the x and \(\bar{y}\) as the mean of the y. The following is a simple example of applying Pearson (x, y) using experimental data in table 3:

\[ r = \frac{-0.428571429}{3.295017884 \times 3.703280399} = -0.03512 \]

3.5. Recommendation
At this stage the author will provide the best restaurant recommendations with the highest score based on the results of the recommendation system.

3.6. Evaluation
At this stage an evaluation will be carried out using the Root-Mean-Square Error (RMSE). RMSE is an evaluation technique that is often used in collaborative filtering [16]. RMSE is formulated as follows:

\[ RMSE = \sqrt{\frac{1}{N} \sum_{u,i}(P_{ui} - r_{ui})^2} \]

In the above formula, \(P_{ui}\) represents the rating prediction, \(r_{ui}\) is the observation of the actual rating or rating, and N is the rating that users have given the restaurant. The rating prediction is a rating from the results of the recommendations produced, while the observation rating is a restaurant's current
rating before the recommendation is made. The smaller the resulting RMSE value, the more accurate the predictions will be [17].

4. Result and discussion

4.1. Implementation process and recommendations

Implementation process and recommendations is as follows:

- Create a new user input containing 10 lines of restaurant name, restaurant address, and rating. This data can be seen in Table 4. This new data is data that will be used as testing data in the recommendation system creation experiment.

| Name                  | address                               | rating |
|-----------------------|---------------------------------------|--------|
| Little pizza Emilio Portes Gil | Calle Emilio Portes Gil               | 5      |
| Rockabilly            | Agustin De Iturbide                   | 1      |
| Restaurante 75        | Villa de Pozos 4497 Villa de Pozos    | 2      |
| Abondance Restaurante Bar | Industrias 908 Valle Dorado          | 3      |
| Sirlone               | Carr. Mexico                          | 4      |
| Unicols Pizza         | Plaza Del Carmen                      | 4      |
| Restaurant El Muladar de Calzada | De Guadalupe 460 San Miguelito      | 3      |
| La Posada del Virrey  | Av. V. Carranza                       | 3      |
| Restaurant and Bar and Clothesline Carlos N Charlies | Domingo 10 711 El Empleado          | 2      |
| Giovannis             | 40 Norte Civac 1RA. Seccion           | 5      |

- Adds the placeID from the restaurant dataset to the newly created user input data. Extract the restaurant place id from the restaurant dataset and then add it to the user enter as shown in Table 5.

| placeID | name                  | address                               | rating |
|---------|-----------------------|---------------------------------------|--------|
| 132667  | Little pizza Emilio Portes Gil | Calle Emilio Portes Gil               | 5      |
| 132937  | Rockabilly            | Agustin De Iturbide                   | 1      |
| 135070  | Restaurante 75        | Villa de Pozos 4497 Villa de Pozos    | 2      |
| 135069  | Abondance Restaurante Bar | Industrias 908 Valle Dorado          | 3      |
| 132854  | Sirlone               | Carr. Mexico                          | 4      |
| 132856  | Unicols Pizza         | Plaza Del Carmen                      | 4      |
| 135033  | Restaurant El Muladar de Calzada | De Guadalupe 460 San Miguelito      | 3      |
| 132862  | La Posada del Virrey  | Av. V. Carranza                       | 3      |
| 134983  | Restaurant and Bar and Clothesline Carlos N Charlies | Domingo 10 711 El Empleado          | 2      |
| 135013  | Giovannis             | 40 Norte Civac 1RA. Seccion           | 5      |

- With the id location of the restaurant on user input, filtering of users who have visited the same restaurant can be done. After that, the grouping is done based on the user id and ends with the creation of groups so that the user who has the most input has a greater priority. Figure 2 shows an example of a group display that has been created based on the userID.
See the similarity of users in the dataset and enter the user. The next step will be to compare all users with the users who have been determined and choose the most similar using the Pearson Correlation Coefficient. The Pearson Correlation will be calculated between the user and group sections that were previously created.

Table 6. Example of similarity results.

| similarityIndex | userID   |
|-----------------|----------|
| -0.866025       | U1038    |
| -1.000000       | U1003    |
| 1.000000        | U1008    |
| -1.000000       | U1014    |
| 0.000000        | U1016    |

Table 6 shows an example of the calculation results to see the similarity of users. In this table the similarity value of each user is shown in the SimilarityIndex column.

Users with top x that have similarities to user input. In this step, the top 50 rows of user data will be retrieved that have the same level of user input.

Table 7. Example of the top user row.

| similarityIndex | userID   |
|-----------------|----------|
| 1.0             | U1008    |
| 1.0             | U1037    |
| 0.0             | U1045    |
| 0.0             | U1101    |
| 0.0             | U1059    |

Table 7 shows an example of the top 5 rows that have similarities with entries. After this process is complete, we can start making restaurant recommendations to enter users.

Selected user ratings for each restaurant. In this step, we will calculate the weighted average of the restaurant ratings that have been given using the Pearson Correlation as a weight. Weights obtained from multiplication of restaurant ratings with the results of previous Pearson similarities that have a similarity index.

Table 8. Example of rank and weight multiplication results

| similarityIndex | userID   | placeID | rating | weightedrating |
|-----------------|----------|---------|--------|----------------|
| 1.0             | U1008    | 132866  | 1      | 1              |
| 1.0             | U1008    | 132858  | 1      | 1              |
| 1.0             | U1008    | 132854  | 2      | 2              |
| 1.0             | U1008    | 135108  | 1      | 1              |
| 1.0             | U1008    | 132851  | 1      | 1              |
After ranking the ranking with weights is complete as shown in Table 8, the next step is to proceed to collect new rankings and divide the rank by the number of weights to produce a recommendation score.

- Recommendation results. The ranking is done based on the recommendation score obtained to see which reports have the best recommendation score and are recommended by the algorithm. Table 9 shows the 10 best restaurants based on algorithm results.

4.2. Evaluation

After the recommendation making process is complete, the next step is to conduct an evaluation. RMSE is used at this stage. The following is the implementation of evaluation using RMSE:

\[
RMSE = \sqrt{\frac{1}{N} \sum (P_{u,i} - r_{u,i})^2} = 1.359589956010579
\]

The RMSE results above are the result of cumulative errors that are rooted. Then we can find the average error of each data by dividing the RMSE results by the amount of data that has been normalized or data that has the same interest with user input, namely 462, then the average error per data is 0.003. These results indicate a small error rate in the data or a good level of accuracy.

| placeID | name                          |
|---------|-------------------------------|
| 132854  | Sirlone                       |
| 132851  | KFC                           |
| 132847  | Don Burguers                  |
| 135054  | Restaurante y Pescaderia Tampico |
| 132869  | Dominos Pizza                 |
| 135108  | Potzocalli                    |
| 132858  | Hamburguesas Valle Dorado     |
| 132846  | El Lechon Potosino            |
| 132830  | Rincon Huasteco               |
| 132866  | Chaires                       |

5. Conclusion

The use of collaborative filtering based on user references has been widely applied in making recommendation systems [18]. In this study, collaborative filtering is used based on filtering users. User ratings for certain restaurants are used as a reference for other users to create a recommendation system. The calculation of the similarity between one user and another user is done using the Pearson Correlation. After obtaining the similarity value for each user, the user's rank is multiplied by the Pearson similarity value to obtain the weight. Equality and weight values are added based on their respective indices. Then to get the best restaurant recommendations, the division is carried out between the number of similarity scores and the weight of each index and ends with a sorting process based on the highest score of each index. The last step that needs to be done is to check the error rate against the resulting prediction rating. Based on the evaluation using RMSE, it was found that this experiment has good accuracy.

References

[1] Wei J, He J, Chen K, Zhou Y and Tang Z 2017 Collaborative filtering and deep learning based recommendation system for cold start items Expert Syst. Appl. vol 69 pp 1339–51

[2] Kawasaki M and Hasuika T 2018 A recommendation system by collaborative filtering including information and characteristics on users and items 2017 IEEE Symp. Ser. Comput. Intell. SSCI 2017 - Proc. vol 2018-January pp 1–8
[3] Jiang L, Cheng Y, Yang L, Li J, Yan H and Wang X 2019 A trust-based collaborative filtering algorithm for E-commerce recommendation system J. Ambient Intell. Humaniz. Comput. vol 10 no 8 pp 3023–34

[4] Mathew P, Kuriakose B and Hegde V 2016 Book recommendation system through content based and collaborative filtering method Proc. 2016 Int. Conf. Data Min. Adv. Comput. SAPIENCE 2016 pp 47–52

[5] Elssaedi M M, Salih O M and Ahmeed A 2019 Optimizing selection using Pearson correlation coefficient: a case study in movie industry AIP Conf. Proc. vol 2146 issue August

[6] Liu X 2019 Improved collaborative filtering algorithm based on multi-dimensional fusion similarity Proc. - 2019 IEEE Int. Conf. Smart Internet Things, SmartIoT 2019 pp 440–3

[7] Samundeewary K and Krishnamurthy V 2018 Comparative study of recommender systems built using various methods of collaborative filtering algorithm ICCIDS 2017 - Int. Conf. Comput. Intellett. Data Sci. Proc. vol 2018-January pp 1–6

[8] Mu Y, Xiao N, Tang R, Luo L and Yin X 2019 An Efficient Similarity Measure for Collaborative Filtering Procedia Comput. Sci. vol 147 pp 416–21

[9] Koohi H and Kiani K 2016 User based collaborative filtering using fuzzy C-means Meas. J. Int. Meas. Confed. vol 91 issue May pp 134–9

[10] Moe W W, Aye N and Htwe A 2017 Performance comparison of collaborative filtering prediction methods on recommendation system Am. Sci. Res. J. Eng. Technol. Sci. vol 28 no 1 pp 2313–4410

[11] Al Hassanieh L, Jaoudeh C A, Abdo J B and Demerjian J 2018 Similarity measures for collaborative filtering recommender systems 2018 IEEE Middle East North Africa Commun. Conf. MENACOMM 2018 pp 1–5

[12] Ma L, Wang X, Chen A, Gao R, Tang Y and Xiao L 2017 A new collaborative filtering algorithm based on data smoothing Proc. - 2017 2nd Int. Conf. Multimed. Image Process. ICMIP 2017 pp 305–8

[13] Ju Y, Bailin L and Zhixiang Z 2017 A collaborative filtering recommendation algorithm with improved similarity calculation 2nd Int. Conf. Sens. Netw. Comput. Eng. vol 79 pp 158–61

[14] Liu Y, Feng J and Lu J 2017 Collaborative filtering algorithm based on rating distance Proc. 11th Int. Conf. Ubiquitous Inf. Manag. Commun. IMCOM 2017

[15] Kaggle 2017 Restaurant Data with Consumer Ratings | Kaggle Retrieved on 15 July 2020 from https://www.kaggle.com/uciml/restaurant-data-with-consumer-ratings?select=geoplaces2.csv

[16] Zhang J, Wang Y, Yuan Z and Jin Q 2020 Personalized real-time movie recommendation system: practical prototype and evaluation Tsinghua Sci. Technol. vol 25 no 2 pp 180–91

[17] Jiang L, Cheng Y, Yang L, Li J, Yan H and Wang X 2019 A trust-based collaborative filtering algorithm for E-commerce recommendation system J. Ambient Intell. Humaniz. Comput. vol 10 pp 3023–34

[18] Chen VX and Tang T Y 2019 Incorporating singular value decomposition in user-based collaborative filtering technique for a movie recommendation system: A comparative study ACM Int. Conf. Proc. Ser. pp 12–5