Ranking Analysis of Tourist Destinations in Jakarta Using Collaborative Filtering Method

Anggi[*1], Sudaryono[2], Nur Azizah[3]
[1,2,3] Master of Informatics Engineering, University of Raharja, Indonesia
E-mail: [*1]**anggi@raharja.info**, [2]sudaryono@raharja.info, [3]nur.azizah@raharja.info

Abstract

The high workload coupled with advances in technology means that employees sit all day in front of computers. This can result in stress and mental fatigue. Traveling can relieve soreness from sitting too much and make the heart happy. Indonesia is a country rich in tourist attractions. Jakarta as the capital of Indonesia itself has at least 1,016 recreation and entertainment locations. A recommendation system can be a solution to make it easier to choose tourist destinations that suit user interests and preferences. This research was created to provide users with a list of tourist destination choices based on collaborative filtering and provide method accuracy values with RMSE. The results of this research are that the system is proven to be able to provide tourist destination references for users and shows an RMSE value of 0.34. This value still shows that there is an error in the method's predictions. There is still a lot of development that can be done so that this system has more features and wider uses.

Keywords *— Decision Support System, Collaborative Filtering, Tourist Destinations*

1. INTRODUCTION

High workloads, busy routines, responsibilities at work, and coupled with advances in technology, make employees sit all day in front of a computer or laptop, which makes their eyes tired and lacks movement. This can result in stress, mental fatigue and burnout.

Traveling can provide an opportunity to re-stimulate the mind to make it fresh again, relieve aches from sitting too much, reduce feelings of boredom, and make the heart happy. Tourist travel today is not a luxury activity. Indonesia is a country rich in natural beauty, culture, history and ethnic diversity. This means that Indonesia has many interesting and varied tourist attractions, offering extraordinary experiences at affordable prices.

Jakarta as the capital of Indonesia has at least 1,016 recreation and entertainment locations[1]. Such a large number certainly offers a wide choice of tourist destinations based on price, location, type of tourism, etc. A recommendation system can be a solution to make it easier to choose tourist destinations that suit user interests and preferences, so that tourist trips become more efficient, satisfying and enjoyable.

This research is about building a tourist destination recommendation system. The process of creating the system is based on several attributes or parameters such as tourist destination location, tourist destination description, category, city, rating, user, user location, user age, and tourist ticket prices. This recommendation system uses a collaborative filtering method and a Modified Neural Network using Recommender net.[2,3].

2. RESEARCH METHOD

A. Recommendation System

A recommendation system is an application that provides and recommends items that users like[4]-[6]. Implementing recommendations in a system usually anticipates items such as recommendations for films, music, etc. This system works by collecting data directly or indirectly from users.

The steps taken in the recommendation system include data acquisition from users which can be done by:

- Ask users to rate an item.
- Ask users to designate at least one of their favorite items.
- Give users many options and ask them to choose the one that suits them best.
- Ask users to list the items they like or dislike most.

Data collection is not directly related to the user but is done by observing the items viewed by the user on the system. The collected data is then processed using a certain algorithm. The results are then returned as recommended items with parameters from the user.

Recommendation systems are also alternative search engines for items that users are looking for. When developing a recommendation system, there are several ways to solve the problem, such as user-based collaborative filtering, content-based filtering, and hybrid. However, some researchers added knowledge-based behavioral methods[7,8].

B. Collaborative Filtering

In general, collaborative filtering (CF) is a nearest neighbor algorithm (NN) that is used both in its original form and in machine learning (ML), especially in supervised learning, to predict user preferences in recommendation system. Here, neural graph collaborative filtering (NGCF) aims to solve the critical problem of mapping from existing features : “collaborative signals hide in user-item interactions.”[9].

Collaborative Filtering is a method that predicts the usefulness of items seen by previous users[10]. This method is a process of assessing items using other people's reviews[11].

Collaborative Filtering is a method that utilizes the ratings or preferences given by groups of users to identify commonalities and patterns in those preferences. This method is used to generate recommendations based on similar preferences between users[12].

This research uses Collaborative User Based Filtering where the formula used is Pearson Correlation. This formula is used in user-based Collaborative Filtering to measure the similarity between two users based on their preferences for items. The Pearson correlation formula can be stated as follows:

$$\text{Pearson Correlation} = \frac{\sum((\text{preferensi}_i - \text{rata-rata preferensi}_i) \times (\text{preferensi}_j - \text{rata-rata preferensi}_j))}{\sqrt{\sum(\text{preferensi}_i - \text{rata-rata preferensi}_i)^2} \times \sqrt{\sum(\text{preferensi}_j - \text{rata-rata preferensi}_j)^2}}$$

In the formula, preferences$i$and preferences$j$is the preference of the two users being compared, while the average preference$i$and average preferences$j$is the average preference of each user[13].

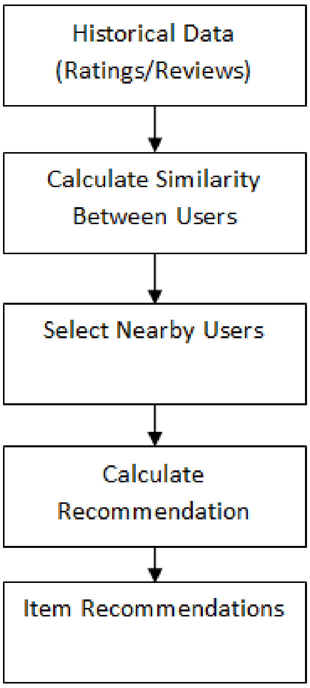

Figure 1 . Collaborative Filtering flowchart[13]

C. RMSE

Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction error). Residuals are a measure of how far a data point is from the regression line; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit. Root mean square error is commonly used in climatology, forecasting, and regression analysis to verify experimental results.

Root mean square error (RMSE) is the square root of the average square error resulting from calculations[14]. The lower the RMSE results, the better the prediction results.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(A_t - F_t)^2}{n}}$$

Where :
At = Actual data value
Ft = Value of forecasting results
N = amount of data
$\sum$ = Summation (Add all values)

3. RESEARCH RESULTS AND DISCUSSION

A. Tools and Materials Used

This recommendation system is run using Google Collab on a laptop whose specifications are AMD Ryzen 7 3750H (8CPUs) ~2.3Ghz, 8GB RAM, AMD Radeon Vega 10 Graphics 4GB memory, 128 GB SSD, 1TB HDD. Apart from that, Python libraries such as TensorFlow, Keras, image Data Generator, OS, NumPy, pandas, seaborn, and matplotlib are also used.

B. Exploratory Data Analysis

Exploratory Data Analysis is data exploration used for the initial testing process which aims to identify patterns, find outliers (if any), and validate assumptions [15]. Exploration data consists of tourist attractions with the highest ratings by users in each region. Figures 2 and 3 are the results of tourism data exploration. Figure 2 shows tourist attractions that have been rated by users. The tourist attraction that is given the most ratings by users is Pacenongan Culinary Tourism.

Next, Figure 3 shows the tourism categories in Jakarta in the form of culture, amusement parks and others. Apart from that, data exploration shows the number of users and the city of origin who rated the tourist attraction.

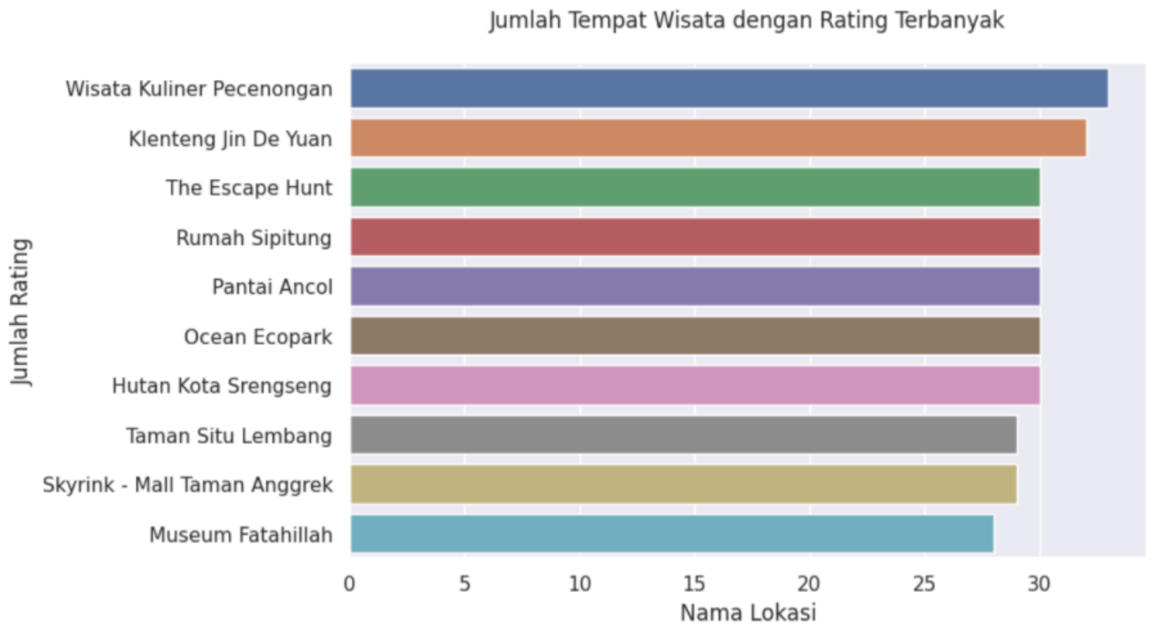

Figure 2: The Relationship Between Tourist Locations and the Number of Visitors

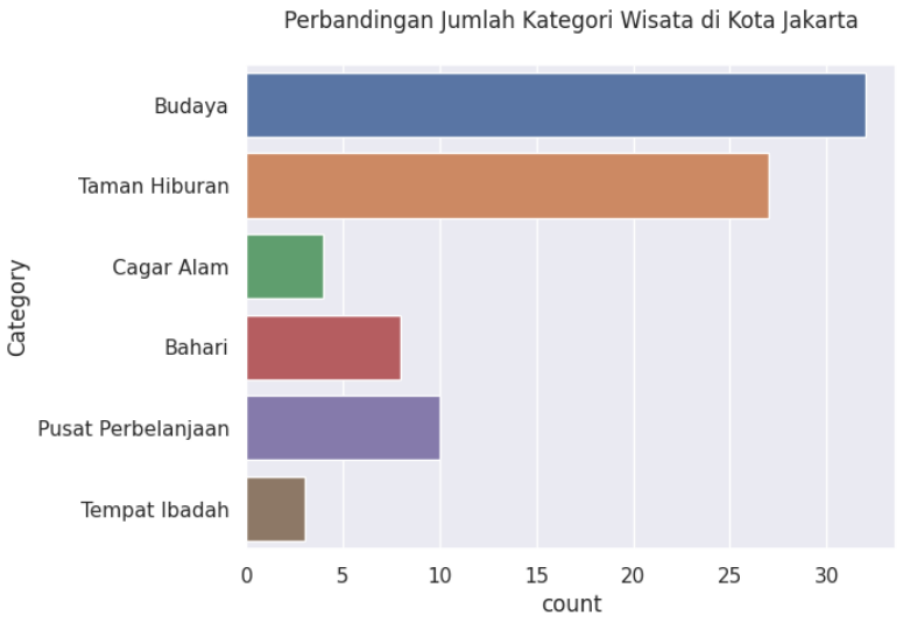

Figure 3: Comparison of the Number of Locations with Tourism Visitors

C. Recommendation Results

The results of this research consist of two things: tourism recommendation output and RMSE performance. Figure 4 shows recommendations for ten tourist destinations in the city of Jakarta.

The performance of this recommendation system with Collaborative Filtering uses RMSE because this system performs regression tasks[16,17]. RMSE shows the metric model produced when creating a tourism recommendation system using the research steps described in the research methods section. Figures 5, 6, and 7 show RMSE during train and validation data with different comparisons. Complete RMSE results can be seen in Table 1. The test results show that RMSE is above 0.3.

The recommendation system that has been created is a recommendation system that involves three data, namely use, place and rating data. The system architecture is made simple by only involving four embedding layers. First, by modifying the layers and component parameters (learning rate, epoch, optimizer) contained in the model. Apart from that, it is also necessary to involve other recommended attributes, in this case namely tourist categories and tickets.

```
2/2 [==============================] - 0s 5ms/step
Daftar rekomendasi untuk: User 106
============================================

-----------------------------------------------------------
Tempat dengan rating wisata paling tinggi dari user
-----------------------------------------------------------
Grand Indonesia Mall : Pusat Perbelanjaan
Bumi Perkemahan Cibubur : Taman Hiburan
Pulau Semak Daun : Bahari
Taman Agrowisata Cilangkap : Taman Hiburan

-----------------------------------------------------------
Top 10 place recommendation
-----------------------------------------------------------
1 . Kampung Cina
     Budaya , Harga Tiket Masuk  15000 , Rating Wisata  4.5

2 . Klenteng Jin De Yuan
     Tempat Ibadah , Harga Tiket Masuk  0 , Rating Wisata  4.5

3 . Pantai Ancol
     Bahari , Harga Tiket Masuk  25000 , Rating Wisata  4.4

4 . Museum Tekstil
     Budaya , Harga Tiket Masuk  5000 , Rating Wisata  4.5

5 . Margasatwa Muara Angke
     Cagar Alam , Harga Tiket Masuk  25000 , Rating Wisata  4.2

6 . Pasar Petak Sembilan
     Pusat Perbelanjaan , Harga Tiket Masuk  0 , Rating Wisata  4.4

7 . Freedom Library
     Budaya , Harga Tiket Masuk  0 , Rating Wisata  5.0

8 . Hutan Kota Srengseng
     Taman Hiburan , Harga Tiket Masuk  1000 , Rating Wisata  4.3

9 . Taman Spathodea
     Taman Hiburan , Harga Tiket Masuk  0 , Rating Wisata  4.6

10 . Plaza Indonesia
     Pusat Perbelanjaan , Harga Tiket Masuk  0 , Rating Wisata  4.7
```

Figure 4 – Ranking Results of 10 Recommendations

Figure 5 – 70:30 RMSE Diagram

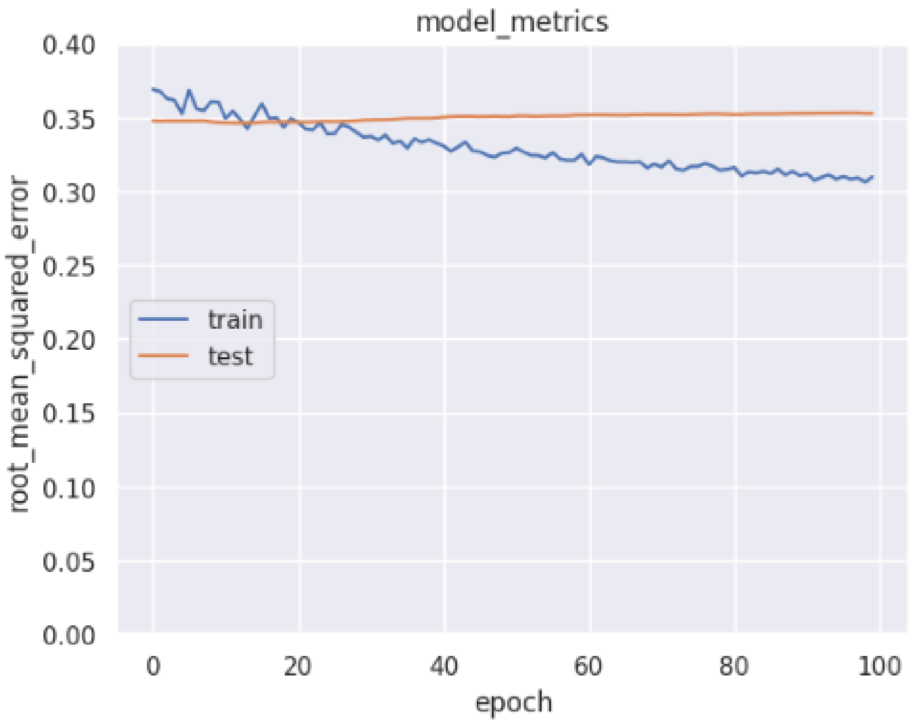

Figure 6 – 80:20 RMSE Diagram

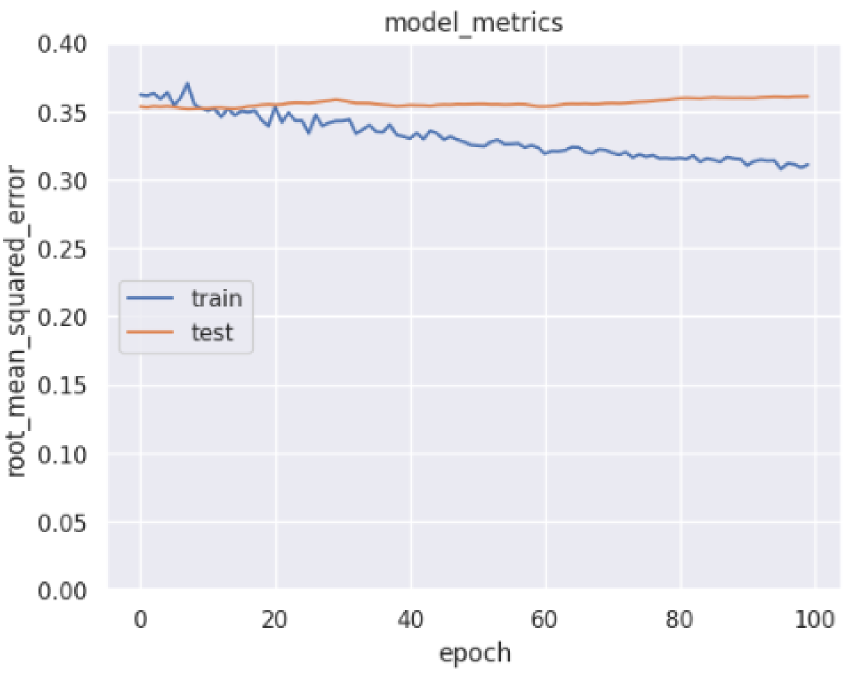

Figure 7 – 90:10 RMSE Diagram

Data Comparison	**70:30**	**80:20**	**90:10**
RMSE value	0.389	0.338	0.349

Table 1 – Comparison of RMSE Values

4. CONCLUSION

In this article, a recommendation system that uses Collaborative Filtering with Recommender_Net has been implemented. Collaborative Filtering is carried out between "users" and "destinations" to produce the best recommendations based on ratings. Then, the training and testing process is carried out using Recommender_Net which consists of four embedding layers. The parameters used in Recommender_Net are sigmoid activation, learning rate 0.0004, Adam optimizer, and epoch 100. The best percentage of training data and training and testing data is 80:20. The results of the train data and test data show that the RMSE value is 0.34. This value is lower than 1.0, which is the RMSE value generally considered the limit for acceptable accuracy. However, the RMSE value of 0.34 is also not a perfect value. This value still shows that there is an error in the method's predictions.

There is still a lot of development that can be done so that this system has more features and wider uses. This prediction error in RMSE can be reduced by increasing the accuracy of the method, for example by using larger training data or by using a more complex algorithm.

Apart from that, recommendations for alternative models with other architectures and hyperparameter tuning are needed to find out a better model.

5. REFERENCES

[1] DKI Jakarta Province in Figures 2021(pdf). DKI Jakarta Provincial Central Statistics Agency: 74. Accessed April 11, 2021.

[2] DS Pradana and GP Hartawan, "Comparison of Content-Based Filtering and Collaborative Filtering Algorithms in Recommendations for Student Extracurricular Activities," Progressive, vol. 18, no. 2, pp. 151–160, 2014.

[3] OG Yalçın, Applied Neural Networks with TensorFlow 2. 2021.

[4] P. Yochum, L. Chang, T. Gu, and M. Zhu, “Linked Open Data in Location-Based Recommendation Systems on Tourism Domain: a Survey,” IEEE Access, vol. 8, pp. 16409–16439, 2020, doi: https://doi.org/10.1109/aCCESS.2020.2967120.

[5] RA Hamid et al., “How smart is e-tourism? A systematic review of smart tourism recommendation systems applying data management,” Comput. Sci. Rev., vol. 39, p. 100337, 2021, doi: https://doi.org/10.1016/j.cosrev.2020.100337.

[6] C.-S. Jeong, K.-H. Ryu, J.-Y. Lee, and K.-D. Jung, “Deep Learning-based Tourism Recommendation System using Social Network Analysis,” Int.J.Internet, Broadcast. ommun., vol. 12, no. 2, pp. 113–119, 2020, [Online]. Available:http://dx.doi.org/https://doi.org/10.7236/IJIBC.2020.12.2.113.

[7] A. Fakhri, ZKA Baizal, and EB Setiawan, "Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region," J. Phys. Conf. Ser., vol. 1192, no. 1, 2019, doi:https://doi.org/10.1088/1742-6596/1192/1/012023. 70 KE Permana et al.

[8] N. Nassar, A. Jafar, and Y. Rahhal, “A novel deep multi-criteria collaborative filtering model for recommendation systems,” Knowledge-BasedSyst.,vol.187,p.104811, 2020, doi: https://doi .org/10.1016/j.knosys.2019.06.019.

[9] Xiang Wang, Le Wu, Xiangnan He, Kun Zhang, Meng Wang. 2022. A Survey on Accuracy-Oriented Neural Recommendation: From Collaborative Filtering to Information-Rich Recommendation. Journal of Information Technology.

[10] Pratama, YA, David Wijaya, Paulus, Arwin Halim. 2013. DIGITAL CAKERY WITH COLLABORATIVE FILTERING ALGORITHM. Medan : ISSN. 1412-0100 VOL 14, NO 1.

[11] Elisabeth Sondakh, Gabriela.2020. Design of a Recommendation System for Selecting a Coffee Drink Menu Using a Collaborative Filtering Algorithm. Tangerang: Bachelor Thesis, Multimedia Nusantara University. Konstan, JA, & Riedl, J. 2012. Recommender systems: from algorithms to user experience. United States : Springer.

[12] Konstan, JA, & Riedl, J. 2012. Recommender systems: from algorithms to user experience. United States : Springer.

[13] Ricci, Francesco, Lior Rokach, Bracha Shapira, and Paul B. Kantor. 2011. Recommender Systems Handbook. United States : Springer.

[14] Prasetyo, VR, Lazuardi, H., Mulyono, AA, & Lauw, C. (2021). Application of the RapidMiner Application to Predict the Rupiah Exchange Rate against the US Dollar Using the Linear Regression Method. National Journal of Information Technology and Systems, 7(1), 8–17.

[15] M. Komorowski, DC Marshall, JD Salciccioli, and Y. Crutain, “Secondary Analysis of Electronic Health Records,” Second. Anal. Electrons. Heal. Rec., no. September, pp. 1–427, 2016, doi:https://doi.org/10.1007/978-3-319-43742-2.

[16] FO Isinkaye, YO Folajimi, and BA Ojokoh, “Recommendation systems: Principles, methods and devaluation,”Egypt.InformaticsJ.,vol.16,no.3,pp.261–273,2015,doi:https://doi.org/10.1016/j.eij.2015.06.005.

[17] M. Saadati, S. Shihab, and MS Rahman, “Movie recommender systems: Implementation and performance evaluation,” arXiv, 2019.