Collaborative filtering-based Madura Island tourism recommendation system using RecommenderNet

Kurniawan Eka Permana[1*], *Abdullah Basuki* Rahmat[1], *Dyo Adya* Wicaksana[1], and *Dicky* Ardianto[1]

[1]Department of Informatic, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Indonesia

Abstract. Personalized tourism recommendations in Madura Island are limited, posing challenges for visitors in selecting destinations that match their preferences. This study develops a Collaborative Filtering-based recommendation system using a modified Cosine similarity approach combined with Convolutional Neural Networks (CNN) to improve personalized tourism suggestions for visitors to Madura Island. The investigation validated the system's accuracy in providing tailored recommendations by thoroughly exploring popular attractions and user preferences, highlighting its potential to enhance the overall tourism experience. User ratings for tourist attractions were collected via Google Forms. The collected data were processed, formatted, and modeled using the RecommenderNet architecture, with training and validation to optimize system performance. Data exploration revealed insights into key island attractions. The recommendation system yielded personalized suggestions, exemplified by a sample user's top-rated destinations. The system's performance was evaluated using RMSE, achieving a best score of 0.2579, demonstrating its accuracy in delivering personalized recommendations. The Collaborative Filtering-based recommendation system effectively improved user experiences by providing personalized tourism suggestions. Future work should focus on enhancing algorithmic approaches and expanding data integration to further refine and enrich tourism experiences in Madura Island.

1 Introduction

As the digital world develops, computer technology can be created to meet the needs of people. Human needs can guide the development of computer technology. The advancement of computer technology has the potential to simplify decision-making [1]. In the era of globalization and information innovation, the surge in mass travel between cities is undeniable. Surabaya, as a prominent city, attracts many individuals either as a place to reside or visit. Similarly, attention is drawn to Madura Island, situated to the east of Surabaya. The decision of where to live or visit is complex, considering factors such as safety, cultural background, and public facilities. Recommendation systems play a crucial role in aiding individuals or groups in making informed decisions when selecting a destination city.

Collaborative Filtering (CF) technology has emerged as an effective method in providing relevant and personalized recommendations based on user preferences and past behaviors. While CF has proven effective in various industries, its application in recommending cities, especially in Indonesia, is still limited and warrants further exploration. Indonesian travelers might have specific preferences influenced by cultural norms, such as a preference for halal-friendly destinations, or travel choices guided by local festivities like Nyepi in Bali or the Yogyakarta arts festivals. Moreover, Indonesia's archipelagic nature means that travel preferences can be highly regional, with some users preferring tropical beaches like those in Lombok, while others might seek cooler highlands like those in Bandung. Personalized recommendation systems can account for these cultural and geographical factors, offering tailored suggestions that resonate more deeply with Indonesian users. To address existing gaps and develop targeted solutions, this research focuses on the challenge of the limited number of available recommendations. Existing tourism recommendation systems often consider general factors, neglecting specific elements such as cultural preferences and personal needs. This research aims to improve the accuracy and relevance of recommendations by developing procedures to fill information gaps and utilizing existing data sources effectively.

An application that provides and suggests items when making decisions that the user finds appealing is known as a recommendation system [2-4]. A tourist recommendation system with multiple stages—input, preprocessing, process, and output—has been the subject of research on recommended tourist destinations. The input comprises of tourist attractions, user evaluations, ratings, and meteorological data (weather and climate). Term Frequency-Inverse Document Frequency (TF-IDF) word embedding is used to preprocess the textual data. Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) are used in the feature extraction and

* Corresponding author: kurniawan@trunojoyo.ac.id

recommendation processes. Results presented as an assessments index [5].

In addition, the second level uses LSTM to make a recommendation. The outcomes are presented as top-ranked tourist destinations and location probabilities. Using LSTM, accuracy results show 97.2%. The system additionally makes use of a Bidirectional LSTM and a Gated Recurrent Unit (GRU). Results for accuracy are 94.2% and 96.4%, respectively [6].

In the realm of recommendation systems, they serve as applications providing and suggesting items when users make preferred decisions [4]. These systems anticipate items for movies, music, books, news, and travel, collecting data directly or indirectly from users.

The steps in a recommendation system involve data acquisition, where users can rate items, assign favorites, choose preferences from multiple options, or list liked or disliked items. Data collection can also occur by observing items viewed by users in the system. The collected data undergoes processing using specific algorithms, and the results are returned as recommended items [7].

Recommender systems are considered alternative search engines. Various methods, such as user-based collaborative filtering, content-based filtering, hybrid approaches, and knowledge-based behavioral methods, are employed [8]. In a technical context, one of the most common methods used in developing recommendation systems is Collaborative Filtering (CF). This method utilizes data from multiple sources to find patterns of similarity between users and provide recommendations based on these similarities. In this research, data is sourced from two main channels: user reviews and user tourism history.

The user's tourism history includes information about the tourist attractions they've visited, providing insights into visit patterns and assumptions about users with similar interests. User reviews offer deeper insights into preferences, such as cleanliness, service, and facilities [9].

By combining information from these two sources, the system can build user profiles reflecting their preferences and interests in tourist attractions. The success of Collaborative Filtering-based recommendation systems relies on comprehensive and structured data collection, as the accuracy is highly dependent on the amount and diversity of available data. Additionally, other researchers explore recommendation systems using Question Answer and Knowledge Graph Techniques, as demonstrated in a case study in Vietnam [10]. This technique, part of Natural Language Processing (NLP), involves stages such as question input, one-hot encoding, Seq2seq model, processing Knowledge Graph with deep learning LSTM, and producing answers through Graph Query. The study reports an F1 measure of 83% and precision of 87%.

In another instance, a tourist spot recommendation system application is built, recommending attractions based on various attributes. This system utilizes Collaborative Filtering and a Modified Neural Network using Recommender Net [11]. In summary, this research endeavors to provide targeted and personalized options for individuals or families considering Madura Island as a place to stay or visit. The focus on individual preferences and needs aims to add significant value to users. Furthermore, the research is anticipated to expand the understanding of tourism recommendations in Indonesia and contribute innovative insights to the field. An application for a tourism destination recommendation system was developed for this study. A number of factors or parameters are taken into consideration when making recommendations for tourist destinations, such as tourist destination descriptions, categories, cities, ratings, user location, and user age. Collaborative filtering and a modified neural network with Recommender_net are used in the recommendation system [11-12].

2 Method

Process of collaborative filtering with place and user data. In the model created using a modified neural network, collaboration takes place. Thismodel is a Python module that was adapted from Keras Tensorflow [13]. This research makes a Collaborative filtering recommendation system with a modified Cosine similarity approach with CNN. CNNs put together intricate patterns from the simpler patterns in the training data to attain a high degree of accuracy. For example, CNNs can accurately detect humans in an image using just lines up to an eyebrow, two eyebrows up to a human face, and finally a full human figure [12]. The recommendation system that will be made in this research goes through 7 stages as shown below in the system flow picture in Fig. 1. [7].

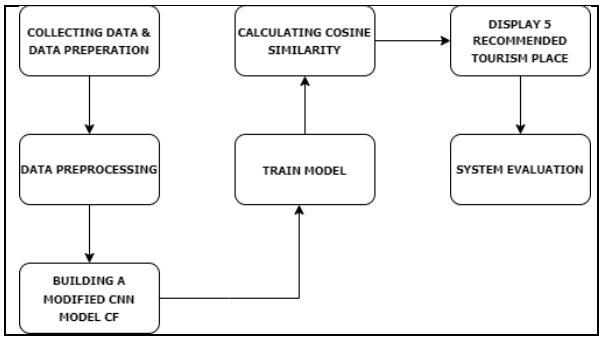

Fig. 1. Research method.

2.1 Data retrieval

In this research, the data collection process involved utilizing a Google Form as the primary tool for gathering information. By employing a form-based approach, the author sought to streamline the process for respondents, making it convenient for individuals both from within Madura Island and from other regions to participate in the study. The use of the Google Form was intended to encourage a wider range of participants to review and rate tourist attractions on Madura Island, whether these attractions were ones they had personally visited or those they wished to recommend. The form was

carefully designed with a series of prompts and commands that guided the respondents in selecting specific tourist attractions. These attractions were then subjected to a review process, where participants were asked to provide ratings based on their experiences or perceptions. This method not only facilitated the collection of diverse opinions but also ensured that the data gathered was relevant and reflective of the respondents' authentic preferences and experiences. The results of data collection is shown in Table 1.

Table 1. Data from rating.

No	User_Id	Place_Id	Place_Rating
0	1	23	4.0
1	1	22	4.5
2	1	10	4.5
3	1	31	4.0
4	1	30	4.0
5	2	30	2.0
6	2	22	4.0
7	2	15	4.5
8	2	3	4.5
9	2	11	2.0
10	3	4	4.0
11	3	15	4.5
12	3	3	3.5
13	3	9	4.0
14	3	5	4.0

The dataset that has not been cleaned is obtained from the results of data collection through the form through a series of processes until it can produce data that is ready for further research. Table 1 shows a sample of 15 data from 3 users who have provided reviews for tourism in Madura, because they give ratings and reviews of 5 tourism, so each user has five ratings.

2.2 Data processing

Data processing is a very vital stage in this research. The following are the steps of the data processing process including:

1. Data Cleaning

Data from the Google Form is checked to ensure there are no incomplete or inappropriate entries. Invalid or empty data is corrected or deleted. Data cleaning is a crucial stage in the process of processing information from Google forms. At this stage, the data collected from the form is carefully checked to ensure that there are no entries that are incomplete or do not match the predefined criteria. If any data is invalid or has gaps, corrective measures or even deletions will be taken. The aim is to ensure the integrity and quality of the data before proceeding to the more in-depth analysis stage. This cleaning process is important so that the data used in the research has high reliability and is suitable for the needs of the analysis to be carried out. We removed any empty or mismatched values from the pertinent data after the data cleaning procedure was finished [14].

2. Dataset Integration

The three data obtained (user, rating, and place) are combined to form one unified dataset containing user information, ratings, and tourist attractions. So we get 3 (three) data in .csv format, namely user, rating, and place [15].

3. Formatting Data

The data was organized in a format suitable for subsequent analysis. This included ensuring format consistency within each column and converting the data to the appropriate format where necessary. Data formatting is the process of changing the features and values of data to make it fit a specific format. Data from users and restaurants is used in later steps and needs to be in numerical format to be readable by machine [16].

2.3 Data modelling

At this stage, modeling is done to build a recommendation system. The Recommender Net based Collaborative Filtering method is used to analyze user preference patterns and provide appropriate recommendations. The research process begins with the first step of loading data related to user preferences for tourist destinations. Then, a preprocessing stage is carried out where users and destinations are encoded as integer indices so that they can be processed further. Next, the preparation of training and validation data is the next step to ensure the model can learn from the data properly. After that, the model is created using an artificial neural network that has been modified with the Collaborative Filtering method as shown in Table 2 before the model is trained we check to take sample data that has been merged and encoded.

Table 2. Sample data after merge.

	User_ Id	Place _Id	Place_ Rating	User	Place
972	54	31	5.0	202	30
312	82	9	4.0	157	8
70	168	2	4.0	61	1
1028	181	34	4.0	65	33
1021	119	34	4.0	168	33

Which allows the model to understand and predict user preferences for destinations. The model training process is based on dividing the data into two parts, namely training data and validation data. Finally, the trained model can be used to provide the best destination recommendations to users who have form id 30, by displaying the top 5 destinations based on their preferences. With this sequence of steps, it is expected that the recommendation system can provide accurate recommendation results and in accordance with the user's wishes.

2.4 Train model

The training procedure will then be carried out using a modified neural network. He_normal is used to initialize weights in the training and validation process, among other parameters [17-18]. The built model is trained using the processed dataset. Next, the model is evaluated to measure its performance. Evaluation includes the use of evaluation metrics such as Root Mean Square Error

(RMSE) to measure the extent to which the model can accurately predict user preferences. The author conducted a model evaluation process to determine and validate the model during the training process by treating the parameters in it using the matplotlib library to display the parameter as shown in Table 3.

Table 3. Experiment parameter.

Model	Epoch	Learning Rate	RMSE
1	10	0.0001	0.3698
2	100	0.0002	0.3950
3	1000	0.0003	0.3319
4	10	0.0004	0.2579
5	100	0.0005	0.2684

The use of a differentiated learning rate according to the training so that the author wants to know the best average RMSE results in each test. A comparatively low mean squared error value from Model 1 indicated a more accurate system [19]. So in this study, different hyperparameter changes are different in order to see the best average RMSE results in which test.

2.5 Evaluation

The training and validation data produces performance in the form of root mean square error, and validation root means square error. A measure of Root Mean Square Error (RMSE) in (1) [20-22].

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

yi = predicted value of y; $\hat{y}$ = actual value of y; N = amount of data.

3 Result and discussion

3.1 Testing environment

The process of making a recommendation system program is carried out using Visual Studio Code software with hardware specifications in the form of a Core i7-10750H @ 2.60GHz processor (6 CPUs), Memory 16384 MB RAM. Solid State Drive 512 Gb. In addition, the program requires several Python libraries such as TensorFlow, keras, numpy, pandas, seaborn and matplotlib.

3.2 Exploratory data analysis

The first testing step uses exploratory data analysis, or EDA, to detect trends, identify outliers (if any), and check assumptions [23]. Our research is actively developing a Collaborative Filtering recommendation system for tourism in Madura. Currently, our focus is on collecting and analyzing data related to popular tourist attractions on the island. Fig. 2 depicts the locations that have received the most reviews. The key attractions include the Suramadu Bridge, Bangkalan Plaza, Makam Syaikhona Kholil, and the Alun-Alun Bangkalan [24]. The primary goal is to enhance user experience by providing personalized recommendations based on their preferences and behaviors. The aim of this testing process's data exploration is to identify patterns, find outliers, and validate the data [7].

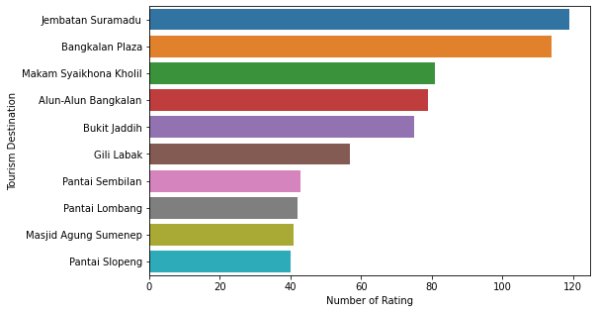

Fig. 2. Most reviewed location.

3.3 Recommendation result

The results of recommendations through testing using the Recommender Net architecture can be seen in table 1, where the results of testing the model are obtained from a sample user of order 30, after giving a rating and review. Then the results are obtained as shown in the Table 4.

Table 4. Tourism recommendation output result.

Tourism	Price (idr)	Rating	Maps	Category
Pantai Sembilan	15000	4	https://maps.app.goo.gl/PFyfugmkYBg2QcG68	Bahari
Gili Labak	150000	4	https://maps.app.goo.gl/Hbg74yP8tC214Ht19	Bahari
Pantai Siring Kemuning	5000	4	https://maps.app.goo.gl/ocjmp1NBkSsDZoVk6	Bahari
Museum Cakraningrat Bangkalan	0	4	https://maps.app.goo.gl/GsfRXTHopHyeg5SJA	Budaya
Masjid Agung Bangkalan	0	4	https://maps.app.goo.gl/BUaTeT4YaDgfUUMt7	Religi

The results differ, as Figure 3 shows the final outcome after training, rather than an average result. Because the system executes a regression task, the recommendation system with collaborative filtering employs RMSE to measure performance [25]. The best RMSE is the 4th model with a value of 0.2579 obtained by using hyper parameters with a learning rate of 0.0004

and epoch value is 10. Although it can be seen in the bar chart image that is close to the smallest value is the 3th training result but it is the last point of the epoch iteration of each model, so the smaller RMSE value will produce a good system.

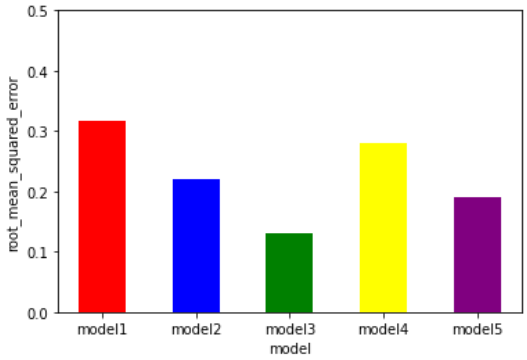

Fig. 3. Bar Diagram of RMSE.

4 Conclusion

In evaluating the Collaborative Filtering recommendation system for tourism in Madura, the data exploration results reveal significant insights into popular tourist attractions on the island. The Recommendation table results show the system output for a sample user, providing an overview of recommended tourist attractions along with attributes such as price, rating, map, and category.

Evaluation of the recommendation system using Recommender Net architecture resulted in the best Root Mean Square Error (RMSE) value of 0.2579 in the fourth model with certain hyperparameter settings, especially the learning rate of 0.0004. Although the RMSE bar chart shows fluctuations during the training iterations, the final RMSE value confirms the effectiveness of the system. This is in line with the original objective listed in the Introduction chapter, which was to create a system that enhances user experience through customized recommendations. These results validate the potential of the system to provide accurate suggestions, fulfilling tourism recommendation preferences.

Furthermore, the findings of this study suggest further exploration and development opportunities. Future studies could explore improving the recommendation algorithm, considering additional factors such as temporal dynamics, user feedback loops, or more advanced machine learning techniques to improve accuracy. This shows that the system still needs improvement. For future model experiments, it is necessary to involve all attribute elements of tourism data such as user age, tour category, and ticket. In addition, alternative recommendation models are needed with other architectures and hyperparameter tuning to find out a better model.

This research is one of the outputs of Penelitian Grup Riset, Penelitian Mandiri, LPPM, University of Trunojoyo Madura.

References

1. A. Faizin and I. Surjandari, *Product recommender system using neural collaborative filtering for marketplace in indonesia*, IOP Conf. Ser. Mater. Sci. Eng., **909**, 1 (2020), doi: 10.1088/1757-899X/909/1/012072
2. P. Yochum, L. Chang, T. Gu, and M. Zhu, *Linked Open Data in Location-Based Recommendation System on Tourism Domain: a Survey*, IEEE Access, **8**, 16409–16439 (2020), doi: 10.1109/aCCESS.2020.2967120
3. 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. Commun.*, **12**, 2, 113–119 (2020), doi: 10.7236/IJIBC.2020.12.2.113
4. R. A. Hamid *et al.*, "How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management," *Computer Science Review*, 39. Elsevier Ireland Ltd, Feb. 01, (2021). doi: 10.1016/j.cosrev.2020.100337
5. H. woo An and N. Moon, *Design of recommendation system for tourist spot using sentiment analysis based on CNN-LSTM,* J. Ambient Intell. Humaniz. Comput., **13**, 3, 1653–1663 (2022), doi: 10.1007/s12652-019-01521-w
6. W. Shafqat and Y. C. Byun, *A context-aware location recommendation system for tourists using hierarchical LSTM model*, Sustain., **12**, 10 (2020), doi: 10.3390/su12104107
7. K. E. Permana, S. Herawati, and W. Setiawan, *Tourism Destination Recommendation System Using Collaborative Filtering and Modified Neural Network*, 60–70 (2023), doi: 10.2991/978-94-6463-174-6_7
8. A. A. Fakhri, Z. K. A. Baizal, and E. B. Setiawan, *Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region*, in Journal of Physics: Conference Series, May 2019, **1192**, 1 (2019) doi: 10.1088/1742-6596/1192/1/012023
9. M. D. Ekstrand, J. T. Riedl, and J. A. Konstan, *Collaborative filtering recommender systems*, Foundations and Trends in Human-Computer Interaction, **4**, 2. 81–173 (2010), doi: 10.1561/1100000009
10. P. Do, T. H. V. Phan, and B. B. Gupta, *Developing a Vietnamese Tourism Question Answering System Using Knowledge Graph and Deep Learning*, ACM Trans. Asian Low-Resource Lang. Inf. Process., **20**, 5, 1–18 (2021), doi: 10.1145/3453651.
11. D. Sukma Pradana and G. P. Hartawan, Perbandingan Algoritma Content-Based Filtering dan Collaborative Filtering dalam Rekomendasi Kegiatan Ekstrakurikuler Siswa
12. O. G. Yalçin, *Applied Neural Networks with TensorFlow 2: API Oriented Deep Learning with*

Python. Springer, (2020). doi: 10.1007/978-1-4842-6513-0

13. R. E. Nakhli, H. Moradi, and M. A. Sadeghi, *Movie Recommender System Based on Percentage of View,* 2019 IEEE 5th Conf. Knowl. Based Eng. Innov. KBEI 2019, 656–660 (2019), doi: 10.1109/KBEI.2019.8734976
14. S. Anunaya, *Data Preprocessing in Data Mining-A Hands On Guide*, Retrieved Novemb., **2** (2022)
15. M. Ilhami and Suharjito, *Film recommendation systems using matrix factorization and collaborative filtering*, 2014 Int. Conf. Inf. Technol. Syst. Innov. *ICITSI 2014 - Proc.*, November, 1–6, (2014), doi: 10.1109/ICITSI.2014.7048228
16. Y. Liu, S. Wang, M. S. Khan, and J. He, *A novel deep hybrid recommender system based on auto-encoder with neural collaborative filtering*, Big Data Min. Anal., 1, 3, 211–221, (2018), doi: 10.26599/BDMA.2018.9020019
17. S. K. Kumar, *On weight initialization in deep neural networks*, 1–9 (2017), http://arxiv.org/abs/1704.08863
18. K. He, X. Zhang, S. Ren, and J. Sun, *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*, Proc. IEEE Int. Conf. Comput. Vis., 2015 International Conference on Computer Vision, ICCV 2015, 1026–1034, (2015), doi: 10.1109/ICCV.2015.123
19. S. Khaki, Z. Khalilzadeh, and L. Wang, *Predicting yield performance of parents in plant breeding: A neural collaborative filtering approach*, PLoS One, **15**, 5, 1–13 (2020), doi: 10.1371/journal.pone.0233382
20. Amity University and Institute of Electrical and Electronics Engineers, Proceedings of the 9th International Conference On Cloud Computing, Data Science and Engineering : Confluence, 10-11 January 2019, Uttar Pradesh, India (2019)
21. R. M. Sallam, M. Hussein, and H. M. Mousa, *An Enhanced Collaborative Filtering-based Approach for Recommender Systems* (2020)
22. Sagedur Rahman, *Extended Collaborative Filtering Recommendation System with Adaptive KNN and SVD*, Int. J. Eng. Manag. Res., **13**, 4, 105–112, (2023), doi: 10.31033/ijemr.13.4.14
23. M. I. T. C. Data, *MIT Critical Data Secondary Analysis of Electronic Health Records*.
24. A. Melese, *Food and Restaurant Recommendation System Using Hybrid Filtering Mechanism*, Mon. J. by TWASP, **4**, 4, 268–281 (2021), https://doi.org/10.5281/zenodo.4712849
25. F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, *Recommendation systems: Principles, methods and evaluation*, Egypt. Informatics J., **16**, 3, 261–273 (2015), doi: 10.1016/j.eij.2015.06.005