Hierarchical Clustering-Based Geospatial Analysis for a Personalized Tourism Destination Recommender System

Imam Marzuki
Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
7022222004@student.its.ac.id

Djarot Hindarto
Department of Informatics, Faculty of Information and Communication Technology, Universitas Nasional, Jakarta, Indonesia
djarot.hindarto@civitas.unas.ac.id

Afdhol Dzikri
Department of Informatics Engineering, Politeknik Negeri Batam, Batam, Indonesia
afdhol@polibatam.ac.id

Fardani Annisa Damastuti
Department of Multimedia Creative Technology, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia
fardani@pens.ac.id

Yunifa Miftachul Arif
Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia
yunif4@ti.uin-malang.ac.id

Reza Fuad Rachmadi
Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia | Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
fuad@its.ac.id

Mochamad Hariadi
Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia | Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
mochar@ee.its.ac.id (corresponding author)

Received: 20 March 2025 | Revised: 24 April 2025 and 25 May 2025 | Accepted: 31 May 2025

 DOI: https://doi.org/10.48084/etasr.11055

ABSTRACT

This study applies hierarchical clustering with cosine distance to analyze and visualize tourist travel patterns across various provinces in Indonesia. The methodology includes grouping tourism travel data using hierarchical clustering, assessing cluster quality with the silhouette score, and visualizing the results through dendrograms and geospatial maps. The clustering results reveal distinct travel patterns across regions, which can form more targeted tourism recommendations. The evaluation shows that hierarchical clustering achieved the highest silhouette score of 0.843, demonstrating superior performance compared to

the other methods. These findings contribute to the field of tourism management and decision-making by offering data-driven insights for more personalized and effective travel planning.

Keywords-tourism recommender system; hierarchical clustering; cosine distance; geospatial analysis; silhouette score

I. INTRODUCTION

The rapid growth of the tourism industry has increased the demand for more personalized and adaptive travel recommendation systems [1]. Traditional recommendation systems typically rely on user reviews and ratings but often overlook geospatial factors, such as the geographical distribution of tourist attractions and the spatial relationships between destinations. Several studies have applied clustering methods in tourism recommendation systems; however, they still have limitations in effectively utilizing spatial relationships between destinations [2-4]. Therefore, a geospatial clustering-based approach is necessary to enhance the accuracy and relevance of tourism recommendations. One of the most widely used clustering methods in hierarchical data analysis is hierarchical clustering, an unsupervised learning technique that organizes data into a hierarchical structure [5]. Unlike partition-based clustering methods, such as K-means, hierarchical clustering does not require a predetermined number of clusters, making it more flexible for handling dynamic and continuously growing tourism datasets. This approach enables the travel recommendation system to generate itineraries that not only align with user preferences but also consider the geographical relationships between destinations. This study proposes a framework for personalized tourism recommendations based on geospatial clustering using the hierarchical clustering method. This approach aims to enhance the relevance of recommendations by considering the spatial relationships between tourist destinations, resulting in more efficient itineraries. The proposed method integrates Geographic Information System (GIS) techniques with machine learning to improve both the accuracy of the recommendation system and the overall user experience.

II. RELATED WORK

Various studies have developed a Tourist Destination Recommendation System (TDRS) using hybrid approaches that integrate Collaborative Filtering (CF), Content-based Filtering (CB), and Demographic Filtering (DF). Some systems combine CF and CB with the K-Nearest Neighbors (KNN) algorithm, as well as DF with decision tree [6]. There is also a Malayalam-language system that utilizes K-means and hierarchical clustering [7], as well as an approach that adapts to local preferences, such as halal tourism [8]. All of these approaches emphasize the importance of user profiling in travel recommendations. Many studies use clustering methods to process tourism data. Clustering is often used to group destinations and tourists [9-14]. Some works compare K-means with Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to achieve better clustering results [11]. Other studies have employed hierarchical clustering for rural tourism zoning [15] and attraction grouping [16]. Researchers evaluate these methods using silhouette coefficient and other cluster metrics. One study employs a denoising autoencoder with K-means for personalized recommendations [17]. Another study uses ensemble learning for better classification accuracy [18]. Several systems now integrate geospatial data with recommendation models. Researchers use spatial clustering to match user preferences with location features [19]. Some systems group tourist spots based on distance to hotels and public facilities [16]. One study uses K-means to classify tourist origin countries [14]. Other studies have combined blockchain with Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) for spatial risk management [20]. In another study, a visual model maps image preferences to travel spots [21]. These approaches have been demonstrated to enhance the relevance of spatial recommendations.

User-generated content plays an important role in modern TDRS. Researchers systematically collect reviews from social media and travel websites [22, 23]. They use topic modeling and sentiment analysis to extract preferences. In addition, techniques such as Latent Dirichlet Allocation (LDA) and Generative Pre-trained Transformer (GPT) help process large unstructured texts. Contextual factors, such as time and weather, have been demonstrated to improve recommendation accuracy [22]. Furthermore, some models employ web scraping as a means of automatic data collection [18]. These systems are particularly effective at adapting to real-time user behavior. Some relevant works originate from fields outside the tourism domain. Movie recommendation systems offer clustering and personalization methods [17, 24]. Webpage and document recommenders use advanced machine learning algorithms [9, 18]. Ontology modeling supports smart tourism data management [25]. Other studies have focused on travel route optimization and time-series analytics [10, 26]. In [27], a systematic review was conducted to map current trends in tourism recommender systems. These studies provide a foundation for the development of intelligent tourism technologies.

III. RESEARCH METHOD

An overview of the proposed system architecture is detailed in Figure 1. This architecture visualizes the system's main workflow, from data collection to recommendation generation. As illustrated in Figure 1, the architecture of the TDRS consists of a series of interconnected components. The process begins with user data input, followed by data normalization to standardize feature scales. The algorithm then calculates data similarity using cosine distance. hierarchical clustering groups the data based on similarity levels, and the system generates tourist destination recommendations for users.

A. Data Collection

The primary data for this study was sourced from Indonesia's Central Statistics Agency, Badan Pusat Statistik (BPS), detailing domestic tourist trips across provinces in 2024. This official dataset ensures data validity and supports the development of accurate recommendation models. The dataset is presented in Table I.

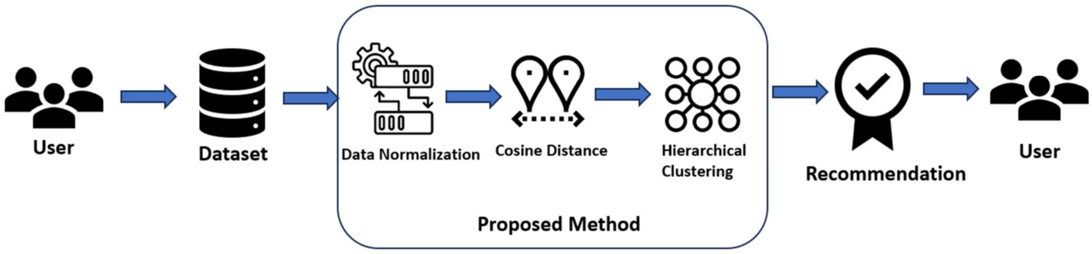

Fig. 1. Proposed system architecture.

TABLE I. TOURIST TRAVEL DATASET BY DESTINATION PROVINCE IN 2024

No	Province	Longitude	Latitude	January	February	…	October	November
1	Aceh	95.3169	5.5625	1,009,785	937,728	…	1,033,267	1,010,852
2	North Sumatera	98.6669	2.6167	4,055,407	3,392,029	…	3,400,972	3,358,557
3	West Sumatera	100.3634	-0.3054	1,767,621	1,758,610	…	1,503,097	1,500,273
4	Riau	101.4447	0.5333	1,546,102	1,449,021	…	1,443,299	1,370,852
5	Jambi	103.6119	-1.6106	665,652	610,850	…	742,582	734,310
…	…	…	…	…	…	…	…	…
34	Southwest Papua	131.2978	-1.3361	87,002	75,511	…	73,886	75,754
35	Papua	138.0804	-4.2699	129,823	102,856	…	84,662	85,704
36	South Papua	140.7221	-7.0352	22,881	17,468	…	10,769	10,193
37	Central Papua	135.7483	-3.9636	47,850	39,547	…	35,605	38,393
38	Papua Mountains	138.757	-3.7916	32,767	35,410	…	31,668	34,146

a. source: https://imammarzuki.github.io/dataset/publikasi3/.

The research dataset in Table I contains the number of tourist trips per month, including the name of each province and its geographical coordinates. Tourist popularity varies by province, providing valuable insights into seasonal trends and travel patterns. The geographic data supports spatial analysis of trip distribution, making this dataset essential for a data-driven tourism recommendation system [28].

B. *Data Normalization*

This study applies a modified min-max normalization method [29], with its formula shown in (1).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Min-max normalization scales feature values (X) to a 0–1 range using the minimum (X_{min}) and maximum (X_{max}) values in the dataset. The result, X_{norm}, ensures uniformity across features and improves consistency in data analysis.

C. *Cosine Distance*

Cosine distance is a measurement method used to determine the degree of difference between two vectors in a multidimensional space [30]. It is derived from the complement of cosine similarity and is calculated using (2).

$$Cosine\ Distance = 1 - Cosine\ Similarity \quad (2)$$

Cosine similarity is formulated in (3).

$$Cosine\ Similarity = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}} \quad (3)$$

where A and B are data vectors, each consisting of multiple elements. A_i represents an element in vector A, whereas B_i represents an element in vector B. Each element has a specific value.

D. *Hierarchical Clustering*

In this study, the hierarchical clustering algorithm begins with a dataset $D = \{x_1, x_2, \ldots, x_n\}$, where each data point is initially treated as an individual cluster. The algorithm calculates cosine distances between points and applies a weighted linkage criterion, considering cluster sizes. In each iteration, the algorithm identifies and merges the two closest clusters based on the weighted linkage. After each merge, the algorithm updates the distances among the remaining clusters. This process repeats until the desired number of clusters is reached or all points form a single cluster. The final output is a dendrogram that visualizes the hierarchical merging structure.

IV. RESULTS AND EVALUATION

A. *Experimental Results*

Experiments conducted on the dataset using hierarchical clustering produced the dendrogram presented in Figure 2. This dendrogram shows the grouping of 38 Indonesian provinces based on cosine distance, producing five main clusters. These clusters group provinces with similar features or characteristics, as seen in the different colored connecting lines and the distance of incorporation between provinces.

The experimental results show the ranking of provinces based on hierarchical clustering. The system identifies the province with the highest score in each cluster, reflecting the dominance of tourist travel in that region. This province serves as the basis for tourism recommendations. The provincial ranking data are presented in Table II, and the system uses these results for further analysis.

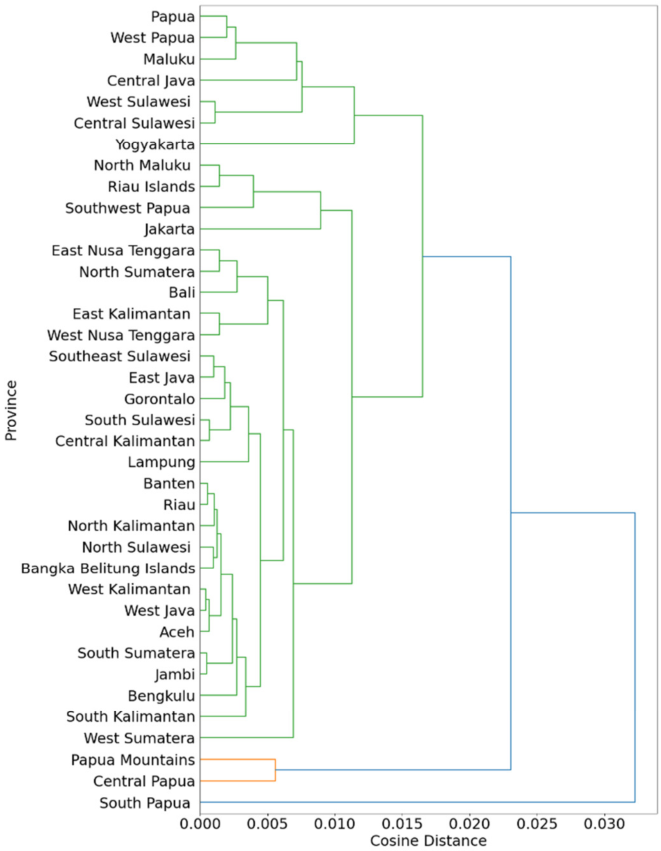

Fig. 2. Dendrogram of clustered Indonesian provinces using cosine distance.

TABLE II. PROVINCE RANKINGS

Rank	Province	Cluster	Average Trips
1	Yogyakarta	1	3,117,321
2	East Java	2	18,037,881
3	West Java	2	13,641,103
4	Jakarta	2	7,122,665
5	Banten	2	3,941,927
…	…	…	…
34	Papua	3	93,911
35	West Papua	3	31,108
36	Central Papua	4	33,172
37	Papua Mountains	4	27,800
38	South Papua	5	13,643

The results of the grouping are visualized through an interactive geomap, which illustrates the geographical distribution of each provincial cluster. The system reveals spatial relationships between provinces, helping to understand regional tourist distribution patterns. Geomaps provide geographical context for interpreting ranking results. Figure 3 shows the geomap visualization results.

B. Evaluation

The quality of the clustering was evaluated using the silhouette score, which measures how well the data points fit within their assigned clusters compared to those in other clusters [31]. The silhouette score is calculated using (4).

$$S(i) = \frac{b(i)-a(i)}{max(a(i),b(i))} \quad (4)$$

The silhouette value of $S(i)$ ranges from -1 to 1, where $a(i)$ indicates the average distance to fellow cluster members and $b(i)$ indicates the average distance to other nearby clusters. A high score indicates that the data are well-clustered, whereas a low score indicates a position at the cluster boundary or the possibility of a clustering error.

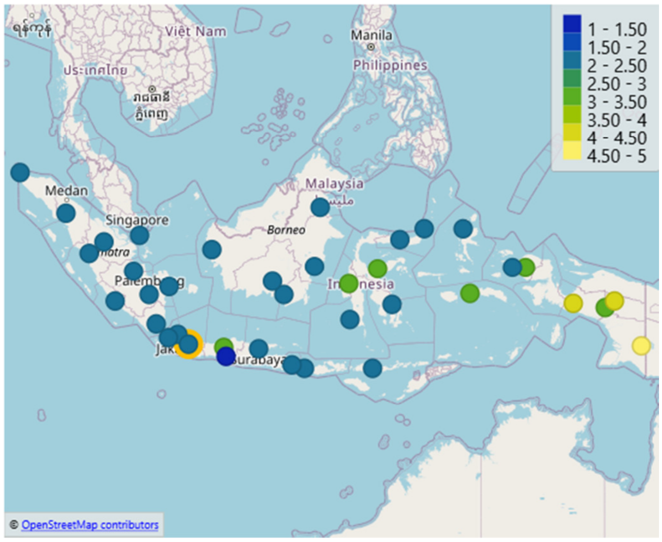

Fig. 3. Geomap of clustered Indonesian provinces.

Figure 4 illustrates the silhouette plot, which shows the distribution of silhouette scores and assesses the quality of clustering. The score ranges from -1 to 1: a score close to 1 indicates the cluster is clearly formed, a score close to 0 indicates the data are at the cluster boundary, and a negative score indicates incorrect grouping. Each color represents a single cluster that has a different size and distribution of scores. The red dotted line shows the average value of the silhouette score of 0.843. The value indicates that the clustering algorithm works very well overall. However, some clusters, such as clusters 3 and 5, show low scores, indicating poor cluster separation.

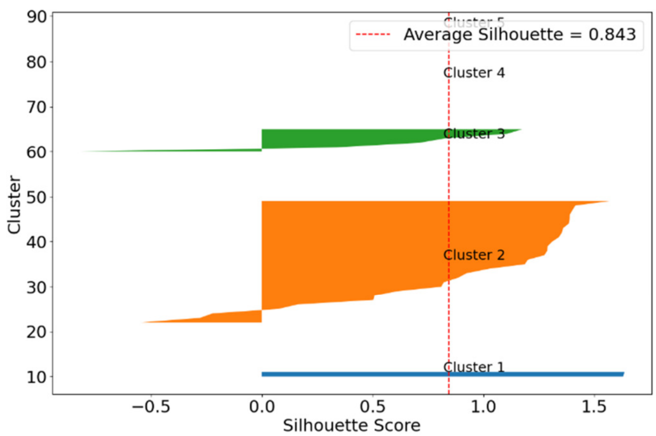

Fig. 4. Silhouette plot of clustering results.

C. Comparison with Other Methods

The performance of hierarchical clustering was compared with that of two other clustering methods, K-means and DBSCAN, using the silhouette score as the evaluation metric. The results revealed differences in each method's ability to group data based on existing structures. Table III presents a comprehensive summary of the evaluation results for all three methods. Hierarchical clustering achieved the highest silhouette score (0.843), followed by K-means (0.746) and DBSCAN (0.734). These results indicate that hierarchical clustering produces the best clustering performance, whereas K-means and DBSCAN are less optimal.

D. Comparison with Existing Studies

The present study is positioned in relation to other studies in the context of clustering-based TDRS, as shown in Table IV. Table IV compares several cluster-based TDRS approaches using the silhouette score for evaluation. The first study used LDA and GPT-2 to analyze destination reviews from social media. This study produced a moderate score of 0.57. The second study applied the K-means algorithm to data on foreign tourists in Indonesia. This method resulted in a score of 0.80 and identified two main clusters for promotional segmentation. The fourth study, which is the author's contribution, used data on domestic travel between regions with hierarchical clustering and cosine distance and recorded the highest score, 0.843.

TABLE III. COMPARISON OF HIERARCHICAL CLUSTERING WITH K-MEANS AND DBSCAN METHODS

No	Clustering method	Silhouette score
1	Hierarchical clustering	0.843
2	K-means	0.746
3	DBSCAN	0.734

TABLE IV. COMPARISON OF CLUSTERING-BASED TOURISM RECOMMENDER SYSTEMS BASED ON SILHOUETTE SCORE EVALUATION

No	Study	Data type	Clustering algorithm	Silhouette evaluation	Silhouette score	Remarks
1	[23]	Unstructured destination reviews (social media)	LDA + LLM (GPT2) + clustering	Silhouette score for Coorg	0.57	Text-based insights with LLM and topic modeling; moderate clustering performance
2	[14]	Foreign tourist origin data to Indonesia	K-means	Cluster number evaluation	0.80 (K=2)	Two dominant clusters: high vs low visitors; useful for promotional strategy segmentation
3	[16]	Tourist destination data in Madura & distance to public facilities	AHC + K-means combination	Silhouette coefficient	0.8055	AHC + K-means combination outperforms AHC (0.707) and K-means (0.638); suitable for areas with uneven access to public infrastructure
4	[Ours]	Domestic travel data across Indonesian provinces	Hierarchical clustering + cosine distance	Method comparison	0.843	Dendrogram and geospatial maps reveal distinct travel patterns

V. CONCLUSION AND FUTURE WORK

The present study employed the hierarchical clustering method to analyze tourist travel patterns in Indonesia. The proposed method successfully grouped the travel patterns of each province, as illustrated through dendrograms and geospatial visualizations. The evaluation results demonstrate the efficacy of the proposed method, as indicated by the high silhouette score of 0.843, which is higher than the scores obtained by other methods. These findings provide a robust foundation for the development of a data-driven tourism recommendation system.

Future research could expand the scope by incorporating additional variables, such as socio-economic and cultural factors, to enhance the understanding of traveler preferences. Other clustering methods, such as spectral clustering, could also be explored to compare performance with hierarchical clustering. The evaluation of cluster quality could be further improved by integrating additional metrics, such as the Dunn index or the Calinski-Harabasz index, to provide a more comprehensive analysis. Furthermore, integrating spatial data could help identify more complex travel patterns, whereas temporal analysis could reveal seasonal trends in traveler movements. This approach is expected to provide deeper insights for data-driven tourism development strategies.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support from the Indonesian Education Scholarship (BPI), the Education Fund Management Institute (LPDP), and the Center for Higher Education Financing and Assessment (PPAPT) under the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia.

REFERENCES

[1] L. Gamidullaeva, A. Finogeev, M. Kataev, and L. Bulysheva, "A Design Concept for a Tourism Recommender System for Regional Development," *Algorithms*, vol. 16, no. 1, Jan. 2023, Art. no. 58, https://doi.org/10.3390/a16010058.

[2] R. A. Nugroho, A. M. Polina, and Y. D. Mahendra, "Tourism Site Recommender System Using Item-Based Collaborative Filtering Approach," *International Journal of Applied Sciences and Smart Technologies*, vol. 2, no. 2, pp. 119–126, Dec. 2020.

[3] Z. Fan and L. Lu, "Tourism Data Clustering Analysis Based on Random Forest Algorithm," in *2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers*, Dalian, China, 2022, pp. 1211–1214, https://doi.org/10.1109/IPEC54454.2022.9777525.

[4] A. S. K. Xin, H. Y. Ting, and A. F. Atanda, "Trends In Tourism Recommendation Systems: A Review," *Journal of Computing Research and Innovation*, vol. 9, no. 2, pp. 85–107, Sep. 2024, https://doi.org/10.24191/jcrinn.v9i2.438.

[5] D. Wu, "Research on Rural Tourism Feature Classification Method Based on Hierarchical Cluster Analysis," in *2024 Second International Conference on Data Science and Information System*, Hassan, India, 2024, pp. 1–4, https://doi.org/10.1109/ICDSIS61070.2024.10594260.

[6] M. E. B. H. Kbaier, H. Masri, and S. Krichen, "A Personalized Hybrid Tourism Recommender System," in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications*, Hammamet, Tunisia, 2017, pp. 244–250, https://doi.org/10.1109/AICCSA.2017.12.

[7] V. K. Muneer and M. K. P. Basheer, "A Comparative Study of Collaborative Filtering and Content-Based Approaches for Improving the Accuracy of Travel Recommender Systems for Malayalam Language," *International Journal of Advanced Networking and Applications*, vol. 14, no. 6, pp. 5717–5721, 2023, https://doi.org/10.35444/IJANA.2023.14608.

[8] Y. M. Arif, H. Nurhayati, S. M. S. Nugroho, and M. Hariadi, "Destinations Ratings Based Multi-Criteria Recommender System for Indonesian Halal Tourism Game," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 1, pp. 282–294, Feb. 2022, https://doi.org/10.22266/ijies2022.0228.26.

[9] I. Pauletic, L. N. Prskalo, and M. B. Bakaric, "An Overview of Clustering Models with an Application to Document Clustering," in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics*, Opatija, Croatia, 2019, pp. 1659–1664, https://doi.org/10.23919/MIPRO.2019.8756868.

[10] N. Lou, "Analysis of the Intelligent Tourism Route Planning Scheme Based on the Cluster Analysis Algorithm," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, 2022, Art. no. 3310676, https://doi.org/10.1155/2022/3310676.

[11] A. Jauhari, D. R. Anamisa, F. A. Mufarroha, and I. O. Suzanti, "Grouping Madura Tourism Objects with Comparison of Clustering Methods," in *2022 IEEE 8th Information Technology International Seminar*, Surabaya, Indonesia, 2022, pp. 119–123, https://doi.org/10.1109/ITIS57155.2022.10009968.

[12] A. Muhammad and D. Saputra, "Machine Learning in the Development of Integrated Tourist Destinations," in *2021 International Conference on Computer Science and Engineering*, Padang, Indonesia, 2021, pp. 1–6, https://doi.org/10.1109/IC2SE52832.2021.9791976.

[13] D. Shrestha, T. Wenan, D. Shrestha, N. Rajkarnikar, and S.-R. Jeong, "Personalized Tourist Recommender System: A Data-Driven and Machine-Learning Approach," *Computation*, vol. 12, no. 3, Mar. 2024, Art. no. 59, https://doi.org/10.3390/computation12030059.

[14] H. Hasanah, N. A. Sudibyo, and R. M. Galih, "Data Mining Using K-Means Clustering Algorithm for Grouping Countries of Origin of Foreign Tourist," in *Basic and Applied Science Conference*, Jakarta, Indonesia, 2021, pp. 88–94, https://doi.org/10.11594/nstp.2021.1112.

[15] W. Wang, S. Cheng, Z. Chen, and K. Liu, "Classification of Rural Tourism Features Based on Hierarchical Clustering Analysis Knowledge Recognition Algorithm," *Mathematical Problems in Engineering*, vol. 2022, no. 1, Jul. 2022, Art. no. 2956020, https://doi.org/10.1155/2022/2956020.

[16] E. M. S. Rochman *et al.*, "A combination of algorithm agglomerative hierarchical cluster (AHC) and K-means for clustering tourism in Madura-Indonesia," *J. Math. Comput. Sci.*, vol. 12, Jan. 2022, Art. no. 62, https://doi.org/10.28919/10.28919/jmcs/7086.

[17] A. A. Al-Asadi and M. N. Jasim, "Deep Learning-Based Rate Prediction Model for Recommender System Using Clustering Techniques," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 2, pp. 477–485, Feb. 2023, https://doi.org/10.22266/ijies2023.0430.38.

[18] Chaithra, L. G. Malleshappa, and J. Sreenivasaiah, "Classification of Web Pages Using the Machine Learning Algorithms with Web Page Recommendations," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 4, pp. 637–650, Aug. 2022, https://doi.org/10.22266/ijies2022.0831.57.

[19] C. Li, J. Cui, and Y. Li, "Research on System Development of Spatial Clustering in Tourism Recommendation," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Apr. 2024, Art. no. 20240840, https://doi.org/10.2478/amns-2024-0840.

[20] D. Hindarto, M. Hariadi, and R. F. Rachmadi, "A Blockchain-based Landslide Mitigation Recommendation System for Decision-Making," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 20452–20460, Feb. 2025, https://doi.org/10.48084/etasr.9806.

[21] Y. Zhang and S. Liu, "A Picture-Based Approach to Tourism Recommendation System," *The Frontiers of Society, Science and Technology*, vol. 5, no. 5, pp. 124–130, Apr. 2023, https://doi.org/10.25236/FSST.2023.050520.

[22] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, "Tourism recommendation system based on semantic clustering and sentiment analysis," *Expert Systems with Applications*, vol. 167, Apr. 2021, Art. no. 114324, https://doi.org/10.1016/j.eswa.2020.114324.

[23] M. N. Sai, N. L. Royal, N. Chakravarthi, N. S. Sai, D. Gupta, and M. Venugopalan, "Automatic Topic Identification in Destination Reviews Using Hybrid Topic Modelling Techniques," in *2024 15th International Conference on Computing Communication and Networking Technologies*, Kamand, India, 2024, pp. 1–8, https://doi.org/10.1109/ICCCNT61001.2024.10726104.

[24] G. S. Ananth, K. Raghuveer, R. Dayananda, and R. Kashyap, "A Novel and Hybrid Approach of an Indian Demographic Movie Recommender System," *Open Access Library Journal*, vol. 7, no. 7, Jul. 2020, Art. no. e6483, https://doi.org/10.4236/oalib.1106483.

[25] H. Liang, "Intelligent Tourism Personalized Recommendation Based on Multi-Fusion of Clustering Algorithms," *Advances in Multimedia*, vol. 2021, no. 1, Dec. 2021, Art. no. 4517973, https://doi.org/10.1155/2021/4517973.

[26] M. Ali, A. Alqahtani, M. W. Jones, and X. Xie, "Clustering and Classification for Time Series Data in Visual Analytics: A Survey," *IEEE Access*, vol. 7, pp. 181314–181338, 2019, https://doi.org/10.1109/ACCESS.2019.2958551.

[27] A. Solano-Barliza *et al.*, "Recommender systems applied to the tourism industry: a literature review," *Cogent Business & Management*, vol. 11, no. 1, Dec. 2024, Art. no. 2367088, https://doi.org/10.1080/23311975.2024.2367088.

[28] I. Marzuki, "Number of Domestic Tourist Trips by Destination Province (Trip) in 2024." Github. [Online]. Available: https://imammarzuki.github.io/dataset/publikasi3/.

[29] H. Henderi, T. Wahyuningsih, and E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *International Journal of Informatics and Information Systems*, vol. 4, no. 1, pp. 13–20, Mar. 2021, https://doi.org/10.47738/ijiis.v4i1.73.

[30] M. Rahim, S. S. Abosuliman, R. Alroobaea, K. Shah, and T. Abdeljawad, "Cosine similarity and distance measures for p,q–quasirung orthopair fuzzy sets: Applications in investment decision-making," *Heliyon*, vol. 10, no. 11, Jun. 2024, Art. no. e32107, https://doi.org/10.1016/j.heliyon.2024.e32107.

[31] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics*, Sydney, Australia, 2020, pp. 747–748, https://doi.org/10.1109/DSAA49011.2020.00096.