INTENSIF, Vol.10 No.1 2026
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

Enhancing Tourist Experiences in North Toraja through K-Means Clustering-Based Recommendation System

Received:
9 June 2025
Accepted:
19 January 2026
Published:
17 February 2026

[1*]Srivan Palelleng, [2] Eko Suripto Pasinggi, [3]Juprianus Rusman
[1,2,3] Department of Information Engineering, Universitas Kristen Indonesia Toraja, Makale, South Sulawesi, Indonesia
E-mail: [1]srivan_palelleng@ukitoraja.ac.id, [2]ekopasinggi@ukitoraja.ac.id, [3]rusman.jr@ukitoraja.ac.id

*Corresponding Author

Abstract— Background: North Toraja in South Sulawesi, Indonesia, is a culturally rich region with high tourism potential due to its unique traditions. The government has invested in infrastructure to boost tourism and regional income (PAD), but has insufficiently used information systems for promotion. An innovative system that can assist tourists in navigating the diverse attractions in North Toraja based on their interests needs to be developed. **Objective**: This research aims to develop a recommendation system for tourist attractions in North Toraja using K-means Clustering and the Similar Characteristics Method. **Methods**: We used Orange Data Mining to perform K-means clustering, and then used similarity-based methods to determine the closeness of characteristics among attractions. The system analyzes based on the fields of cultural, geographical, facility, and landscape features, resulting in four distinct clusters. The clusters were defined as three tourist attractions in cluster C1, eleven in C2, four in C3, and fourteen in C4. We also developed a system interface that allows travelers to input preferences, view personalized recommendations, and access detailed information. The system's novelty lies in its specific application of K-Means Clustering to leverage these local attributes for granular categorization for effective promotion of North Toraja's diversity. **Conclusion**: Our approach effectively groups attractions with similar characteristics, enhancing exploration based on user interests. The high altitude and similar geographical features of North Toraja result in attractions that share natural characteristics, making this system an advancement in technology-driven tourism solutions.

Keywords—Recommendation System; North Toraja; K-Means Clustering; Similar Characteristics; Tourism

Corresponding Author:

Author Srivan Palelleng,
Department of Information Engineering,
Institution Universitas Kristen Indonesia Toraja,
Email srivan_palelleng@ukitoraja.ac.id
Orchid ID: https://orcid.org/0000-0003-2793-7189

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

I. INTRODUCTION

Toraja is one of the regions in Indonesia that has tourism potential [1]. Tana Toraja and North Toraja regions are in the South Sulawesi province. These two regions are some of Indonesia's most popular tourist destinations, especially because of their unique cultures and traditions. Toraja is famous for its strong customs, traditional house architecture called Tongkonan, and magnificent and ritualistic death ceremonies [2]. There are several reasons why Toraja is a popular tourist destination: 1) Toraja is known for its dead culture, which is unique and different compared to other places. The funeral ceremony in Toraja, known as Rambu Solo, is a grand event that often lasts several days and has various ritual stages. 2) The traditional Torajan house called Tongkonan has a roof that curves upwards like a boat. This architecture has a profound cultural and spiritual meaning for the Torajan people. 3) Toraja has stunning natural scenery, including mountains, terraced rice fields, and vast green valleys. 4) In Toraja, various exciting tourist attractions, such as Londa and Lemo, are graves in cliffs and caves. 5) Diversity of Rituals and Festivals: In addition to death ceremonies, Toraja also has various other festivals and rituals celebrated lively, such as Ma'Nene, or the ritual of cleaning ancestors' bodies. Ke'te Kesu, Kalimbuang, Lolai, Londa, and Singki are the most frequently visited tourist attractions. The ones rarely visited are tourist attractions in the countryside or villages due to the narrow roads or the fact that those places are less exposed.

The tourism industry is expected to improve to support regional development and increase Regional Original Income (PAD) in Toraja. According to [3], Several studies on tourism in Indonesia asserted that tourist visits were positively associated with PAD in cities such as Kudus, Gianyar, Pariaman, Manado, Palembang, and North Toraja. Furthermore, hotels/other accommodations and restaurants significantly affect Regional Original Income in North Toraja Regency [4]. Increased tourism can lead to job creation, boost the local economy, and provide employment opportunities for residents [5]. Additionally, it can promote cultural exchange and preserve traditional customs and practices by generating interest and appreciation among visitors [6].

Until now, the government continues to focus on infrastructure development to support tourism activities in the North Toraja Regency. This development includes the construction of an airport, roads, and other facilities [7]. The ease of accessing information related to Tana Toraja Regency's tourism profile is still a priority in developing Tana Toraja's tourism potential through information technology [8]. Furthermore, many information technology facilities still need to be developed. However, this research fills the gap by implementing a recommendation system for tourist attractions. It can help the tourist information system by developing technology in North

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

Toraja Regency. The system should enhance tourists' experiences by providing personalized suggestions based on their preferences and interests. Also, it is expected to increase the visibility of lesser-known attractions, helping attract more people to visit across the region. The recommendation system aims to help visitors discover attractions that align with their interests, whether they seek cultural experiences, natural landscapes, adventure activities, or culinary delights. One potential challenge in implementing the system is accurately capturing and interpreting users' diverse preferences. Therefore, we are working with the North Toraja Culture and Tourism Office to design and build this technology from the government's perspective. Here, we only make the algorithm and do not create the interface.

In this paper, we developed a recommendation system for tourist attractions in the North Toraja, South Sulawesi, Indonesia region based on their characteristics. Their characteristics have similarities in terms of their culture, geography, facilities, and natural landscapes [9]. The recommendation system analyzes the attributes of attractions and groups them into clusters. By doing so, it can suggest destinations that align with visitor preferences, offering personalized recommendations. Therefore, we combine K-means clustering and similar characteristics methods to improve reliability and achieve better results.

K-means clustering is a widely used unsupervised machine learning algorithm for partitioning a dataset into K distinct, non-overlapping subsets or clusters [10], [11]. Specifically, research used K-means for a recommendation system [12], [13]. Each cluster is represented by its centroid, which is the mean of the points in the cluster. The algorithm aims to minimize the within-cluster sum of squares, ensuring that points in the same cluster are as similar as possible. In contrast, points in different clusters are as dissimilar as possible. K-means is popular for its simplicity and efficiency, making it suitable for large datasets. However, it requires the number of clusters, K, to be specified in advance and is sensitive to the initial placement of centroids. In recommendation systems, K-means clustering can be employed to group users or items based on their characteristics or behaviors [14]. The system can provide more personalized recommendations by clustering similar users, as users within the same cluster likely have identical preferences. Similarly, clustering items can help identify groups of related products, allowing the system to recommend items that are frequently associated with each other [15]. This approach enhances the accuracy and relevance of recommendations, improving user satisfaction. Additionally, K-means can assist in identifying outliers or niche user groups, enabling the system to cater to diverse user needs more effectively.

North Toraja Regency has already been the subject of extensive research in agriculture [16], animal husbandry [17], plantations [18], and tourism [19]. Studies have shown that the Toraja regency is highly suitable for coffee plantations due to its fertile soil and favorable climate [20].

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

Research in animal husbandry has highlighted the potential for increasing livestock productivity [21]. Additionally, tourism studies have emphasized the area's cultural richness, with traditional ceremonies and unique architecture attracting visitors from around the world [22]. Developing eco-friendly accommodations and guided cultural tours could be practical in enhancing tourism in the region [23]. The infrastructure research, such as better roads and communication networks, would also make the area more accessible to tourists [24]. Integrating technology, such as virtual reality experiences, could enhance visitor engagement and education [25]. The digital tools can provide immersive insights into the region's history and traditions, attracting tech-savvy travelers. Furthermore, collaborating with government and local communities to develop authentic experiences can ensure tourism growth benefits residents and preserves cultural heritage.

Many studies have been conducted in the field of recommendation systems for tourism [26], [27], [28], focusing on developing algorithms that recommend suitable destinations based on travelers' interests and preferences. Tourism recommendation systems can be implemented using content-based, collaborative filtering, knowledge-based, and demographic approaches, each leveraging user preferences, similarities among users, or profile attributes to generate personalized recommendations [29]. Additionally, conversational recommendation systems utilizing ontology have been developed to support interactive recommendations and explanations [30]. Several studies also propose hybrid recommendation architectures to enhance the effectiveness of tourism recommendation systems [31]. Despite these advancements, further research is still required to design tourism recommendation systems that are more adaptive to local tourism characteristics and capable of effectively addressing diverse user preferences in real-world applications. Furthermore, this study addresses this need by presenting a novel approach that differentiates itself from existing works through its unique combination of K-means clustering and a custom relevancy calculation algorithm tailored to the specific local tourism characteristics of North Toraja, thereby clearly highlighting its novelty.

The purpose of this research is making tourism recommendation systems by combining our algorithm and k-means clustering to suggest the most suitable places for tourists. This algorithm can help refine personalized recommendations over time. We obtained data from the North Toraja Culture and Tourism Office; the data relates to the relevancy of a tourist attraction within a specific category. According to the data, 31 tourist attractions are included in this study and categorized into 16 categories. Using this method, we can determine the distance between tourist attractions by utilizing K-Means clustering and similarity of characteristics.

The contribution of this paper is that we developed a recommendation system combining our relevancy calculation algorithm and K-means clustering and similar characteristics to get the

INTENSIF, Vol.10 No.1 2026
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

closeness value of each tourist attraction. By identifying the closeness value of each attraction, our system was able to help users find the most suitable destination for their travel needs.

This article is presented as follows: Related work reviews the Toraja regency and the recommendation system in tourism. The method section states the framework and methodology, dataset, algorithm, and clustering. The result contains our findings. In the Discussion section, we discuss the advantages and disadvantages of our approach. We discuss our future research in the following section and conclude the research in the conclusion section.

II. RESEARCH METHOD

2.1. Dataset

The North Toraja Culture and Tourism Office provided data on the relevancy of a tourist attraction within a specific category. The category relevance score is from 0 to 5, 0 for irrelevance, and 5 for most relevance. These sixteen categories include paid, stone graves, traditional villages, mountains, rivers, traditional houses, trekking, rafting, agriculture, museums, culinary, swimming, cycling, traditional ceremonies, fishing, and crowded areas. Table 1 shows all the data of 31 tourist attractions and 16 categories. Furthermore, the relevance score helps tourists to identify which categories most appeal to them, also allowing for better resource allocation and promoting strategies. We enable researchers, marketers, and policymakers to freely access and utilize this open-source data. Anyone interested in the data can use it to conduct further analysis or develop new insights, provided they cite this paper as the source.

2.2. Our Algorithm

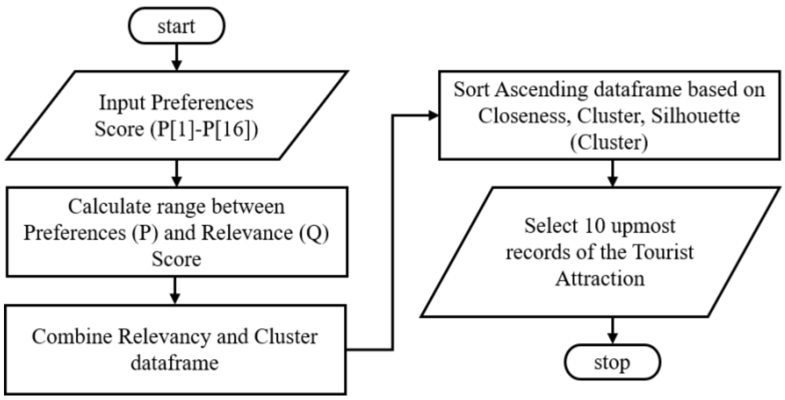


Fig 1. User preferences relevancy calculation flowchart

As shown in Figure 1, it starts with the user entering their desired preference categories in the first step. They can choose any number between 1 and 5 for 16 categories. However, they can also fill only some of the categories. The system will mark the unfilled categories as 0. Then, the calculated range score function, as seen in algorithm 1, is executed to determine the level of

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

proximity between each tourist spot and the user's preferences. In our algorithm, the lower the closeness value, the more suitable the tourist spot is to the user's preference. After that, the relevance dataframe is combined with the cluster dataframe to get a dataframe containing the number, tourist spot, closeness, cluster, and Silhouette (Cluster). After that, the dataset will be sorted in ascending order starting with closeness, cluster, and Silhouette (Cluster). The result is that the top 10 tourist attractions are recommendations.

Input: Preferences
Output: Relevancy dataframe

```
def calculate_range_score(p):
    F_cr = []
    T_an = []
    Num_Ar = []
    for i in range(len(df_ta)):
        Cal_result = 0
        Ta_current = df_ta.iloc[i]
        T_an.append(Ta_current['TouristAttraction'])
        for j in range(len(F_a)):
            if ([P[j]! = 0 and Ta_current[F_a[j]] ! = 0):
                Cra = abs(P[j] − Ta_current[F_a[j]])
                Cal_result = Cal_result + Cra
            else:
                Cal_result = Cal_result + 5
        F_cr.append(Cal_result)
        Num_Ar.append(i + 1)
    df_result = pd.Dataframe(index = Num_Ar)
    df_result['place'] = T_an
    df_result['closeness'] = F_cr
    return df_result
```

Algorithm 1. Calculate range score function

Algorithm 1 is a similar characteristic method, where P is the user's preference, where the user specifies what level of preference, they have concerned the tourist attractions they wish to visit. An array variable called F_{cr} stores the sum of preferences and relevance. T_{an} is an array to store the names of tourist attractions. There is an array variable called Num_{Ar}, which keeps track of the index of the df_{result}. df_{ta} is a tourist attraction dataframe that is complete with the relevance of each category Q. $Ta_{current}$ is a variable that store one row in the tourist attraction dataframe. F_a is an array of categories Q. Variable i is an iteration for df_{ta}, while J is an iteration for F_a. Cra is the result of subtracting abs from P and $Ta_{current}[F_a[j]]$. If $[P[j]! = 0\ and\ Ta_{current}[F_a[j]]\ ! =\ 0$ then, Cra, which is the result of subtracting $\text{abs}(P[j] - Ta_{current}[F_a[j]])$, then Cra will be added to the Cal_{result} variable; besides, the new Cal_{result} is the result of Cal_{result}, which was added to 5. After that, the result of Cal_{result} is entered into the F_{cr} array. Likewise, the Num_{Ar} variable will add a new row with the sum of iteration $i + 1$. After

INTENSIF, Vol.10 No.1 2026
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

that, T_{an} and F_{cr} are entered into the df_{result} dataframe, which is the output of the calculate_range_score function.

Table 1. Relevancy between tourist attractions and categories

No	Tourist Attraction	Distance (m)	Paid (Q[1])	Stone Grave (Q[2])	Traditional Village (Q[3])	Mountain (Q[4])	River (Q[5])	Tradional House (Q[5])	Trekking (Q[7])	Rafting (Q[8])	Agriculture (Q[9])	Museum (Q[10])	Culinary (Q[11])	Swimming (Q[12])	Cycling (Q[13])	Traditional Ceremony (Q[14])	Fishing (Q[15])	Crowded (Q[16])
1	Ke'te Kesu'	5600	5	5	5	2	0	5	3	0	0	0	3	0	5	5	0	5
2	Museum Pongtiku	100	4	0	0	0	0	0	0	0	0	5	0	0	5	0	0	0
3	Buntu Pune	2600	4	4	4	4	0	5	3	0	0	0	0	0	3	4	0	0
4	Pala' Tokkek	7000	4	4	4	4	0	0	3	0	0	0	0	0	0	4	0	0
5	Gumuk Pasir	26000	3	0	0	5	0	0	4	0	0	0	0	0	0	0	0	0
6	Londa	6800	5	5	5	4	0	0	5	0	0	0	4	0	4	5	0	0
7	Museum Landorundun	100	3	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0
8	D. W. Lemb. Nonongan	7100	2	3	4	5	4	4	3	0	5	0	4	0	4	4	4	0
9	Tirotiku	14700	0	0	0	5	0	0	3	0	0	0	0	0	5	0	0	0
10	Lempe	15300	5	0	0	5	0	3	0	0	0	0	0	0	4	4	0	5
11	To' Tombi	12400	5	0	0	5	0	3	0	0	0	0	0	0	5	0	0	5
12	Bukit Nato	13200	5	0	0	5	0	0	3	0	0	0	0	0	4	0	0	4
13	Pong Torra	17900	5	0	0	4	0	0	4	0	0	0	0	0	4	0	0	4
14	Sarambu Sikore	31000	4	0	0	4	5	0	5	0	0	0	0	4	0	0	0	3
15	Buntu Sopai II	14200	4	0	0	5	0	0	5	0	5	0	0	0	0	4	0	3
16	Salib Gunung Singki'	1600	3	2	0	3	0	0	4	0	0	0	0	0	0	0	0	0
17	Kolam Alam Limbong	3200	3	0	0	3	0	0	4	0	0	0	0	5	0	3	0	3
18	Tambolang	1300	3	3	3	3	0	0	4	0	0	0	0	0	0	3	0	0
19	Pana'	80000	2	4	0	3	0	0	3	0	0	0	0	0	0	3	0	0
20	D. W. Sesean Suloara'	11000	2	3	4	3	4	4	3	0	5	0	4	0	5	3	3	0
21	To' Kumila'	15000	3	3	3	3	3	3	4	0	5	4	4	0	5	3	3	0
22	Lo'ko' Mata	13600	5	5	4	4	0	0	0	0	0	0	0	0	4	0	0	4
23	Lombok Parinding	7600	4	3	3	3	0	4	3	0	0	0	0	0	0	3	0	0
24	Kalimbuang Bori'	7600	5	5	5	4	0	5	4	0	4	0	0	0	5	4	0	5
25	Palawa'	10500	5	0	5	0	0	5	0	0	0	5	0	0	5	5	0	5
26	Museum Ne' Gandeng	8100	5	0	5	0	4	5	0	0	4	5	0	0	5	5	0	5
27	To' Barana'	12900	5	0	5	0	5	5	5	5	4	0	0	5	5	5	5	5
28	Galugu Dua	12500	5	0	4	0	0	4	0	0	0	0	0	0	0	4	0	0
29	Marimbunna	5300	0	5	4	3	0	3	4	0	0	0	0	0	0	3	0	0
30	Sarambu Manurun	47300	0	0	0	4	5	0	4	0	0	0	0	5	0	0	0	0
31	Marante	13200	0	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0

INTENSIF, Vol.10 No.1 2026
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

2.3. Clustering

Figure 2 shows the flowchart of our K-means process. Clustering is performed using Orange Data Mining version 3.37. We chose to use this software because we have limited data to analyze. 0n the raw file in Table 1, we perform preprocessing using a normalized feature ranging from 0 to 1. Based on the centroid silhouette score, cluster 4 gets the highest value among the other clusters, which means dividing it into 4 is the best option. The Silhouette Plot image is displayed in the result section. The result of this process is a dataframe with Cluster and Silhouette (Cluster).

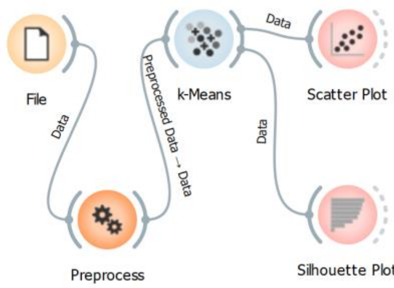


Fig 2. K-Means Clustering Process in Orange Data Mining

2.3.1. K-means

The k-means clustering method is a vector quantization technique originating in signal processing. It partitions n observations into k clusters based on their nearest mean, serving as a prototype for clusters. Each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid). This partitioning of the data space into Voronoi cells results in a partitioning of the data space. In K-means clustering, within-cluster variances are minimized (squared Euclidean distances), but not regular Euclidean distances, which is the more challenging problem of the Weber problem: the mean optimizes squared errors, whereas only the geometric median minimizes Euclidean distances. By using k-medians and k-medoids, it is possible, for example, to find better Euclidean solutions to a problem.

In k-means clustering, given a set of observations $(x_1, x_2, x_3, \cdots x_n)$ where each observation is a real vector of d dimensions, each observation is divided into k sets $[s_1, s_1, \cdots s_k]$ in order to minimize the within-cluster sum of squares (WCSS) (i.e., variance). In terms of its formal objective, it consists of finding the following information:

$$\underset{s}{\operatorname{argmax}} \sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2 = \underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} |S_i| VarS_i \quad (1)$$

In the above equation, μ_i refers to the mean (also known as the centroid) of points in S_i, i.e.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

$$\mu_i = \frac{1}{|S_i|}\sum_{x\in S_i} x, \tag{2}$$

It is important to note that $|S_i|$ is the size of S_i, while symbol $||\cdot||$ is the norm of L^2. There is a pairwise squared deviation of points in the same cluster, which is the same thing as minimizing that deviation:

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \frac{1}{|S_i|} \sum_{x,y\in S_i} ||x-y||^2 \tag{3}$$

Based on the identity, the equivalence can be deduced:

$$|S_i| \sum_{x\in S_i} ||x-\mu_i||^2 = \frac{1}{2}\sum_{x,y\in S_i} ||x-y||^2 \tag{4}$$

There is a constant variance in the data, so in order to maximize the sum of squared deviations between points in different clusters (between-cluster sum of squares, BCSS), the variance at each point is maximized. The law of total variance in probability theory is also related to this deterministic relationship in terms of the law of deterministic variance.

2.3.2. Silhouette

Using silhouettes, you can interpret and validate data consistency within clusters. Using this technique, one can obtain a concise graphic representation of how well objects have been classified. The silhouette value measures how closely an object aligns with its cluster (cohesion) compared with other clusters (separation). A silhouette ranges from -1 to +1. Cluster matching is high for objects within their cluster but poor for objects neighboring them.

Consider that the data have been clustered into k clusters using any technique, such as k-means or k-medoids. Consider that the data have been clustered into k clusters using any technique, such as k-means or k-medoids. In the case of data point and c_I (that is, point i in the cluster c_I), let

$$a(i) = \frac{1}{|c_I| - 1} \sum_{j\in C_I, i\neq j} d(i,j) \tag{5}$$

We divide $d(i,j)$ by $|CI| = 1$ because we do not include the distance $d(i,i)$ in the sum of all data points within the same cluster. $|C|$ is the number of points belonging to cluster $|C|$, and $d(i,j)$ is the distance between cluster CI data points i and j. As a measure of how well Ii is assigned to its cluster, $a(i)$ can be viewed as a smaller value indicates a better assignment.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

The mean dissimilarity of point i to a cluster C_j is defined as the mean distance between i and all points in C_j (where $C_j \neq C_I$). As a result, we can now determine the average for each data point $i \in C_i$

$$b(i) = \min_{j \neq I} \frac{1}{|C_j|} \sum_{j \in C_j} d(i, j) \tag{6}$$

In the formula, the mean distance between i and all points in any other cluster (i.e., in any cluster where i is not a member) is set to be the smallest (hence the m operator). Because it is the next-best cluster fit for point i, the cluster with the smallest mean dissimilarity is called the "neighboring cluster." Now that we have defined one data point i, we may define a silhouette (value) for it.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, if\ |C_I| > 1 \tag{7}$$

And

$$s(i) = 0, if\ |C_I| = 1 \tag{8}$$

We set $S(i) = 0$ when $a(i)$ does not match the definition of $a(i)$ for clusters with size = 1. Despite being arbitrary, this choice is neutral in the sense that it falls somewhere between -1 and 1.

Overall, this research methodology is designed to systematically develop a personalized tourism recommendation system for North Toraja. Starting with a comprehensive dataset from the North Toraja Culture and Tourism Office detailing 31 attractions across 16 categories, our approach integrates two key components: a custom relevance algorithm and K-means clustering. The relevance algorithm calculates proximity scores between user preferences and attraction attributes, while K-means clustering, implemented through Orange Data Mining with $k = 4$ optimal clusters determined by Silhouette analysis, groups attractions based on similar characteristics. By combining the results of these two processes—personalized proximity scores and cluster membership—the system then sorts and presents the top 10 recommendations, ensuring that the suggestions provided are not only relevant to individual interests but also consider the overall similarity of attractions.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

III. RESULT AND DISCUSSION

The findings of this research are that an innovative personalized tourist recommendation system for North Toraja was successfully developed by integrating K-means clustering, which grouped attractions into four main clusters, and similarity assessment to provide tailored suggestions based on user preferences, thereby enhancing the tourism experience. Firstly, we generate clusters and silhouettes (clusters) based on the clustering process using Orange Data Mining. We have divided the clusters into four categories: Cluster C1, C2, C3, and C4. A silhouette value is a metric used to evaluate how well data points are matched with clusters using a clustering algorithm.

Table 2. Tourist Attraction with cluster and silhouette (cluster)

Tourist Attraction	Cluster	Silhouette (Cluster)
Museum Pongtiku	C1	-0.105342
Museum Ne' Gandeng	C1	0.240128
Palawa'	C1	0.290057
Pana'	C2	0.00308987
Lo'ko' Mata	C2	0.0675115
Galugu Dua	C2	0.0966316
Ke'te Kesu'	C2	0.165516
Kalimbuang Bori'	C2	0.113666
Tambolang	C2	0.171975
Londa	C2	0.199061
Marimbunna	C2	0.262659
Lombok Parinding	C2	0.270321
Pala' Tokkek	C2	0.276358
Buntu Pune	C2	0.33464
To' Barana'	C3	0.0848511
To' Kumila'	C3	0.306647
D. W. Lemb. Nonongan	C3	0.372792
D. W. Sesean Suloara'	C3	0.394258
Museum Landorundun	C4	0.0633574
Lempe	C4	0.0699925
Marante	C4	0.0868312
Buntu Sopai II	C4	0.0958837
Salib Gunung Singki'	C4	0.139596
To' Tombi	C4	0.171007
Kolam Alam Limbong	C4	0.18621
Tirotiku	C4	0.211173
Sarambu Manurun	C4	0.21977
Gumuk Pasir	C4	0.245202
Sarambu Sikore	C4	0.245908
Pong Torra	C4	0.265845
Bukit Nato	C4	0.277361

Table 2 shows Silhouette metrics in relation to its cluster. Cluster C1 has three tourist attractions, cluster C2 has eleven, cluster C3 has four, and cluster C4 has fourteen. There is something unique here; for example, even though all three museums (Pongtiku, Ne'gandeng, and

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

Lantorundun) are museums, Lanorundun is in a different cluster from the other two. Based on Table 1, this difference is because Pongtiku and Negandeng have similar relevance scores in museums ($Q10$) and are much higher than Landorundun. Furthermore, Ne'gandeng has the highest silhouette score compared to the two museums because it was grouped into more categories: traditional villages, rivers, houses, agriculture, cycling, traditional ceremonies, and crowded.

Here, we calculate the correlation between category crowded and other categories using Pearson correlation. A correlation value of $r = 1$ indicates a perfect positive correlation, while a correlation value of $r = -1$ indicates a perfect negative correlation. Those other categories are stone graves, traditional villages, mountains, rivers, traditional houses, trekking, rafting, agriculture, museums, culinary, swimming, cycling, traditional ceremonies, and fishing. We assume that when a tourist attraction is crowded, tourists are most likely to visit those attractions, as shown in Figure 3.

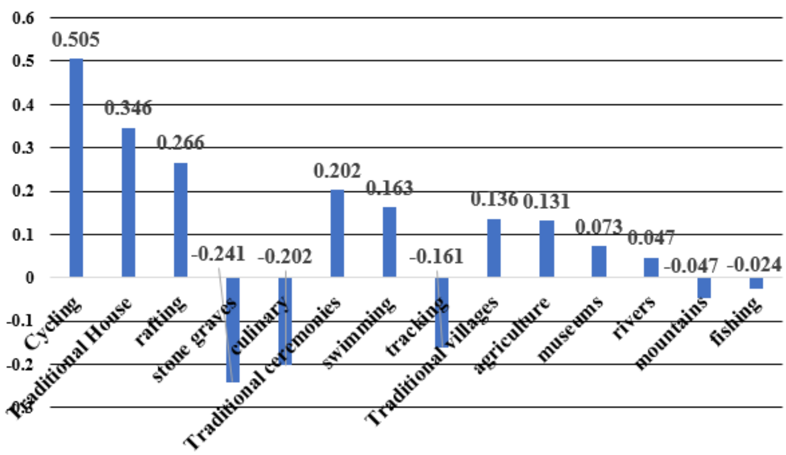


Fig 3. Correlation between crowded category and other categories

A high correlation between the crowded category and other categories would indicate that those attractions are popular among tourists. Figure 3 shows that cycling gains the highest r value in Toraja. The easy accessibility by bicycle often implies that the places are located near cities. As a result, we conclude that the more accessible the place, the more popular it is. Second, the traditional house category received more visitors, as the Tongkonan is still extremely popular with tourists. The most uncrowded place in Toraja is the stone grave, possibly due to its remote location. However, places with both stone graves and traditional houses, such as Ke'te Ke'su, Buntu Pune, and Kalimbuang Bori', are very crowded and popular with tourists.

Figure 4 illustrates the Silhouette Plot generated by Orange Data Mining using Euclidean distance. The figure shows visually that each cluster has a bar representing each tourist attraction's silhouette score. Cluster C4 has the most tourist attractions, while Cluster C1 has the fewest.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

Currently, we will only analyze cluster C3, which contains four tourist attractions, including D. W. Sesean Suloara', D. W. Lemb. Nonongan, To' Kumila', and To Barana. Since both D. W. Sesean Suloara' and D. W. Lemb. Nonongan are tourist villages, it is easy to see that they both have close scores. In Table 1, we find similarities in the categories of paid, stone graves, traditional villages, rivers, traditional house tracking, agriculture, and culinary. Besides D. W. Sesean Suloara' and D. W. Lemb. Nonongan, To Barana is also a tourist village whose score is higher than those of D. W. Sesean Suloara' and D. W. Lemb. Nonongan. However, To' Barana does not have stone graves and culinary, which means it has a silhouette score lower than D. W. Sesean Suloara' and D. W. Lemb. Nonongan.

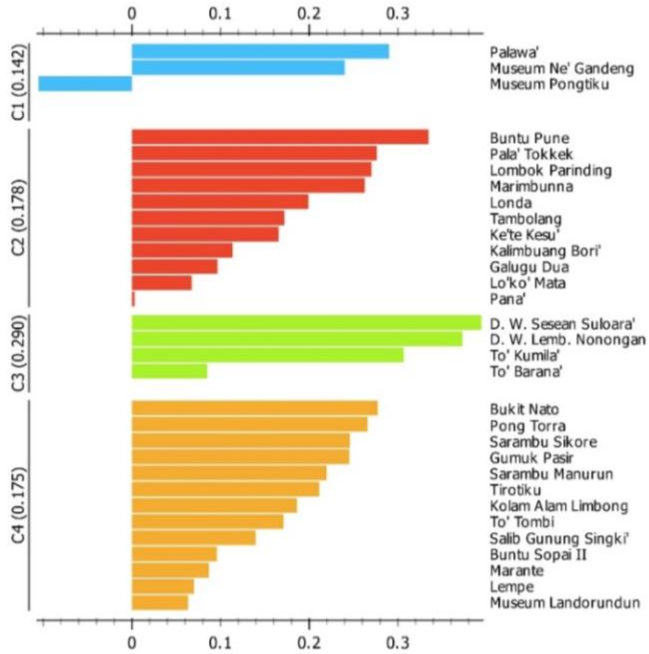
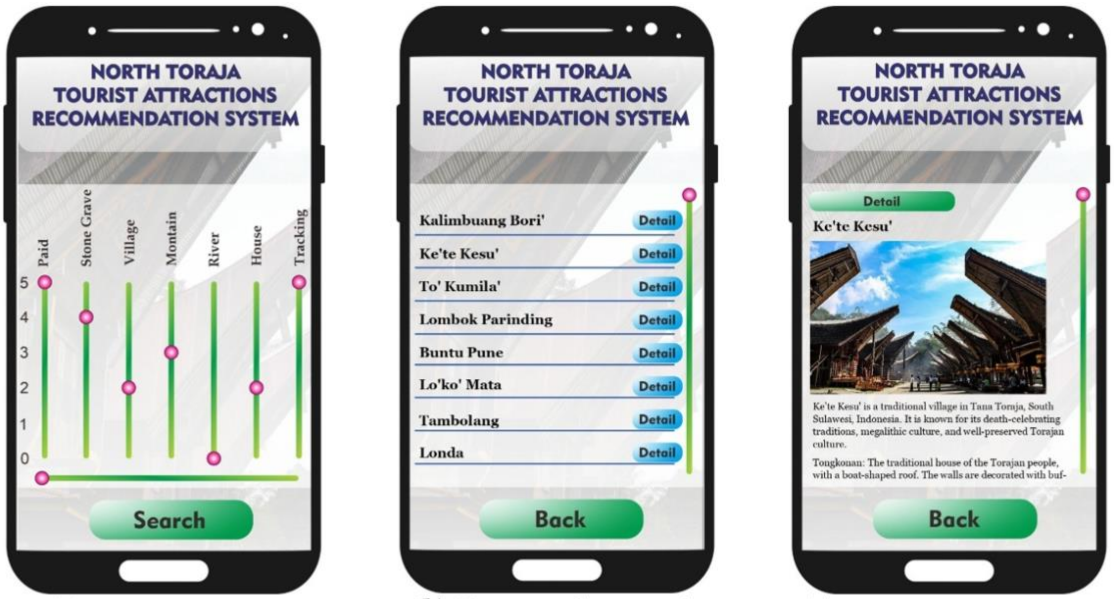


Fig 4. Silhouette Plot using Euclidean distance

As illustrated in Fi1gure 5, it comprises three distinct stages represented by different interfaces: (a) preferences slider, (b) recommendation result, and (c) detail of tourist attraction. The system is designed to enhance the user experience by allowing travellers to personalize their searches based on individual preferences. It also allows travellers to receive tailored recommendations and access detailed information about each recommended location. Figure 5(a) shows the first step, where users can adjust their preferences using sliders. These sliders enable them to specify their desired attributes, such as paid, stone grave, village, mountain, river, house, and other customizable parameters. The sliders give users enhanced control over their search criteria, ensuring the recommendation system is aligned with their expectations. Once the desired preferences have been entered, the search button can be pressed to initiate the recommendation process. As a result of the user preferences, the system generates relevant recommendations based

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

on our algorithm. Figure 5(b) depicts the second step; it presents the recommendation results in a list. This interface displays multiple tourist attractions that are relevant to user preferences. On the right side of each listed tourist attraction, a “Detail” button is available. This button provides additional information about specific tourist attractions. As seen in Figure 5(c), it gives a detailed view of a selected tourist attraction. When a user clicks on the “Detail” button from the recommendation list, the system transitions to this interface, displaying comprehensive information about the chosen destination. The results of this research are in line with or supported by previous studies demonstrating the effectiveness of K-means clustering in recommendation systems, where it is employed to group items based on their characteristics to provide more personalized and relevant suggestions [15].



(a) Preferences slider (b) Recommendation result (c) Detail of tourist attraction

Fig 5. Mockup tourist attraction recommendation system in mobile

Table 3. Tourist attraction recommendations based on one specific tourist attraction

No	Tourist Attraction	Recommended Tourist Attraction
1	Ke'te Kesu'	Kalimbuang Bori', Londa, D. W. Sesean Suloara', To' Barana', Buntu Pune, To' Kumila', D. W. Lemb. Nonongan, Museum Ne' Gandeng, Palawa', Lombok Parinding.
2	Sarambu Sikore	To' Barana', Kolam Alam Limbong, Buntu Sopai II, Sarambu Manurun, Pong Torra, Kalimbuang Bori', To' Kumila', Bukit Nato, Londa, D. W. Lemb. Nonongan
3	Museum Ne' Gandeng	To' Barana', Palawa', Kalimbuang Bori', Ke'te Kesu', To' Kumila', D. W. Lemb. Nonongan, D. W. Sesean Suloara', Lempe, Buntu Pune, Londa.
4	Marante	Tambolang, Lombok Parinding, To' Kumila', D. W. Sesean Suloara', Pana', Salib Gunung Singki', Marimbunna, Pala' Tokkek, Buntu Pune, D. W. Lemb. Nonongan.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

Table 3 shows the closeness characteristics of one tourist attraction to other tourist attractions. As we can see, Ke'te Kesu' is a traditional Torajan village known for its well-preserved Tongkonan houses, rice barns, and burial sites. It is like Kalimbuang Bori', a Megalithic stone burial site, an essential part of Torajan burial customs. Also, Ke'te Kesu' has characteristics identical to Londa, a famous cave cemetery closely related to Toraja's burial traditions. Sarambu Sikore is a waterfall located in the highlands of Toraja, offering stunning natural scenery. This waterfall is close to Kolam Alam Limbong, a natural pool with refreshing water, ideal for relaxation. While there are no waterfalls in To' Barana', this site is part of the nearby Toraja landscape, making them common stops for visitors. Museum Ne' Gandeng showcases the history and cultural heritage of Toraja.

The findings of this research are that an innovative personalized tourist recommendation system for North Toraja was successfully developed by integrating K-means clustering. In this case (as shown in Table 2), for example, it shows the recommended sites that have strong cultural and historical connections. Museum Ne' Gandeng has a close characteristic with To' Barana,' Palawa,' and Kalimbuang Bori' because these places are traditional villages with cultural significance. The last is Marante, known for its cliffside burial sites and traditional Torajan houses. The place is like Tambolang & Lombok Parenting: both sites contain unique burial traditions and Tongkonan houses. Furthermore, like Pala' Tokkek & Buntu Pune, historical and cultural sites preserve Toraja traditions.

As we can see, tourist attractions in North Toraja are very diverse, ranging from stone graves, rivers, traditional houses, and even traditional ceremonies. However, there are some similarities between them from one place to another. This is because North Toraja Regency, all the places located at a high altitude, can still be said to have the same surface area. Therefore, using our approach, K-means clustering and similar characteristics can produce a useful recommendation system. This system can effectively group attractions with similar characteristics, making it easier for tourists to explore the region and enhancing their overall experience by suggesting sites that match their interests. One potential challenge in implementing the system is obtaining accurate and up-to-date data about each attraction's features and visitor preferences. Additionally, the system needs to account for seasonal variations and cultural sensitivities that affect the relevance of recommendations. Ensuring the recommendations remain personalized and diverse can also be complex, requiring continuous refinement of the clustering algorithm.

IV. CONCLUSION

Several criteria for determining similarities between tourist attractions include cultural, geographical, facilities, and other natural landscapes. Furthermore, we combine similar

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

characteristics methods and K-means to develop a recommendation system for tourist attractions in the North Toraja regency using K-means clustering and similar characteristics methods. The North Toraja Culture and Tourism Office provided data on the relevancy of a tourist attraction within a specific category. Orange Data Mining is used to cluster the data, which shows Silhouette metrics and its cluster. We designed a system to enhance the user experience by allowing travellers to personalize their searches based on individual preferences. The system enables travellers to receive tailored recommendations and access detailed information about each recommended location—the closeness characteristics of one tourist attraction to other tourist attractions. For example, Ke'te Kesu' is a traditional Torajan village known for its well-preserved Tongkonan houses, rice barns, and burial sites. It is like Kalimbuang Bori', a Megalithic stone burial site, an essential part of Torajan burial customs. Many tourist attractions have almost identical characteristics because North Toraja Regency, located at a high altitude, can still be said to have the same geographical area. Therefore, using our approach, K-means clustering and similar characteristics can produce a helpful recommendation system. This system can effectively group attractions with similar characteristics, making it easier for tourists to explore the region and enhancing their overall experience by suggesting sites that match their interests. The following aspects can be explored in future research: 1) Obtain real-time feedback from tourists, 2) Integrate official tourism data to make recommendations based on seasonal availability, and 3) Create a user-centric recommendation system that adapts to individual tourist preferences.

While this recommendation system successfully clusters attractions based on similar characteristics and provides personalized recommendations, it has several limitations. One potential challenge in implementing this system is obtaining accurate and up-to-date data on each attraction's features and visitor preferences. Furthermore, the system needs to account for seasonal variations and cultural sensitivities that impact the relevance of recommendations. Ensuring personalized and diverse recommendations can also be complex, requiring ongoing refinement of the clustering algorithm. Future research could explore the following areas: 1) obtaining real-time feedback from tourists, 2) integrating official tourism data to generate recommendations based on seasonal availability, and 3) creating a user-centric recommendation system that adapts to individual tourist preferences.

Author Contributions: *Srivan Palelleng: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Supervision. Eko Suripto Pasinggi: Investigation, Data Curation, Software, Writing - Original Draft. Juprianus Rusman: Investigation, Data Curation. All authors have read and agreed to the published version of the manuscript.*

Funding: No funding was provided by any institution for the conduct of this study.

Acknowledgments: We would like to thank the North Toraja Culture and Tourism Office for

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

their valuable contribution to this project. Their cooperation and support have been instrumental in collecting accurate and comprehensive information. In addition, we are grateful for their dedication to preserving and promoting the rich cultural heritage of North Toraja.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: The data provided by the North Toraja Culture and Tourism Office is open source, allowing researchers, marketers, and policymakers to access and use it freely. All data we used is presented in this paper. Anyone interested in the data can use it to conduct further analysis or develop new insights, provided they cite this paper as the source.

Informed Consent: There were no human subjects.

Animal Subjects: There were no animal subjects.

ORCID:
Srivan Palelleng: https://orcid.org/ 0000-0003-2793-7189
Eko Suripto Pasinggi: https://orcid.org/0000-0001-6654-4834
Juprianus Rusman: https://orcid.org/0000-0002-6522-2152

REFERENCES

[1] M. Hasyim, "Foreign Tourists' Perceptions of Toraja as a Cultural Site in South Sulawesi, Indonesia," *African Journal of Hospitality, Tourism and Leisure*, vol. 8, no. 3, pp. 1–13, 2019.

[2] M. L. Unm and O. Sampebua, "TONGKONAN KE'TE' KESU' AS A TRADITIONAL ARCHITECTURAL TOURIST ATTRACTION IN TANA TORAJA," *Jour. of build. mater. Sci.*, vol. 3, no. 1, Mar. 2021, doi: 10.30564/jbms.v3i1.2929.

[3] R. U. Bayanuddin, Nursini, and S. Suhab, "Pengaruh Sektor Parawisata Pada Kapasitas Fiskal Daerah: Studi Kasus Toraja Utara," *Development Policy and Management Review (DPMR)*, vol. 1, no. 2, pp. 129–148, 2021.

[4] A. Pongtuluran and A. A. Sanusi Fattah, "Analysis of Hotel and Restaurant Tax Efficiency, Effectiveness, and Contribution to Regional Original Revenue (PAD) of North Toraja Regency", Accessed: Feb. 14, 2025. [Online]. Available: https://www.academia.edu/download/75574977/G09114759.pdf

[5] J. Yang, R. Yang, M.-H. Chen, C.-H. (Joan) Su, Y. Zhi, and J. Xi, "Effects of rural revitalization on rural tourism," *Journal of Hospitality and Tourism Management*, vol. 47, pp. 35–45, Jun. 2021, doi: 10.1016/j.jhtm.2021.02.008.

[6] Z. Chen, "Visualizing experiencescape – from the art of intangible cultural heritage," *Current Issues in Tourism*, vol. 25, no. 4, pp. 559–578, Feb. 2022, doi: 10.1080/13683500.2021.1892040.

[7] L. S. Tanduklangi, M. Y. Jinca, and L. Caroles, "Infrastructure Development Strategies In Supporting District Tourism Tana Toraja," *Migration Letters*, vol. 21, no. 2, pp. 566–583, 2024, doi: 10.59670/ml.v21i2.6205.

[8] Reza Fathurrahman, Dwi Putri Nadila, Nurfahmi Hidayat, H. Ahmad Khairul Shiddiq, and Ahmad Khairul Shiddiq, "Web-Based Tourism Recommendation System for South Sulawesi: Design, Implementation, and Evaluation," *j. of digital technol. and comput. sci.*, vol. 3, no. 1, pp. 1–15, Oct. 2025, doi: 10.61220/digitech.v3i1.20255.

[9] K. Y. Sinay, S. Mariati, H. Brahmantyo, and M. Rahmanita, “Analisis Daya Dukung Pada Wisata Buntu Burake, Kabupaten Tana Toraja, Provinsi Sulawesi Selatan,” *JIP*, vol. 27, no. 3, p. 277, Nov. 2022, doi: 10.30647/jip.v27i3.1623.

[10] R. E. Caraka *et al.*, “Cluster Around Latent Variable for Vulnerability Towards Natural Hazards, Non-Natural Hazards, Social Hazards in West Papua,” *IEEE Access*, vol. 9, pp. 1972–1986, 2021, doi: 10.1109/ACCESS.2020.3038883.

[11] T. Toharudin *et al.*, “Investigating Adolescent Vulnerability in Indonesia: A Socio-Remote Sensing Big Data Analytics Study Using Night Light Data,” *IEEE Access*, vol. 12, pp. 14800–14818, 2024, doi: 10.1109/ACCESS.2024.3355963.

[12] Research Scholar, Computer Engineering, Dr. Subhash University, Junagadh, Gujarat, India, B. M. Bangoria, S. S. Panchal, S. R. Panchal, J. M. Maheta, and S. R. Dhabaliya, “Multidimensional Dynamic Destination Recommender Search System Employing Clustering: A Machine Learning Approach,” *IJST*, vol. 17, no. 40, pp. 4187–4197, Oct. 2024, doi: 10.17485/IJST/v17i40.2266.

[13] S. Glaret Shirley, K. Subrahmanyam, D. Susrija, and P. Akhila, “K-Means Algorithm and Clustering Technique for A Recommender System,” *ijaste*, vol. 1, no. 1, pp. 302–312, Feb. 2023, doi: 10.24912/ijaste.v1.i1.302-312.

[14] K. Kim and H. Ahn, “A recommender system using GA K-means clustering in an online shopping market,” *Expert Systems with Applications*, vol. 34, no. 2, pp. 1200–1209, Feb. 2008, doi: 10.1016/j.eswa.2006.12.025.

[15] B. Mulyawan, M. Viny Christanti, and R. Wenas, “Recommendation Product Based on Customer Categorization with K-Means Clustering Method,” *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 508, p. 012123, May 2019, doi: 10.1088/1757-899X/508/1/012123.

[16] A. R. T. Rassing, A. Gunawan, and P. I. Pratiwi, “Agricultural Landscape Model Based on the Culture of the Indigenous People of Tana Toraja, South Sulawesi,” *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1384, no. 1, p. 012022, Aug. 2024, doi: 10.1088/1755-1315/1384/1/012022.

[17] R. Dewi, R. Tandu, H. G. Lunkenheimer, M. Nyho, R. Pasoloran, and R. La’biran, “Tedong (Buffalo): Symbol of Nobility, Humanity, and Entertaiment in Funeral Ceremony in The Indigenous Torajan, Indonesia,” *Intl. J. Rel.*, vol. 5, no. 8, pp. 179–190, May 2024, doi: 10.61707/j1qzmj60.

[18] Y. B. S. Panggabean, M. Arsyad, Mahyuddin, and Nasaruddin, “Sustainability agricultural supply chain in improving the welfare of North Toraja Arabica coffee farmers,” *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1107, no. 1, p. 012065, Dec. 2022, doi: 10.1088/1755-1315/1107/1/012065.

[19] D. R. K. Kausar and M. P. Gunawan, “Managing heritage tourism in Toraja: strengthening local values and improving tourists’ experiences,” *Journal of Heritage Tourism*, vol. 13, no. 6, pp. 550–561, Nov. 2018, doi: 10.1080/1743873X.2017.1411356.

[20] Sunanto *et al.*, “The Impact of Sustainable Technology Adoption on Coffee Farming in Tana Toraja, Indonesia: A Call for Comprehensive Support,” *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1364, no. 1, p. 012016, Jun. 2024, doi: 10.1088/1755-1315/1364/1/012016.

[21] V. S. Lestari, D. P. Rahardja, and S. N. Sirajuddin, “Application of biosecurity in pig farms,” presented at the PROCEEDINGS OF THE 13TH TSME INTERNATIONAL CONFERENCE ON MECHANICAL ENGINEERING 2023, Chiang Mai, Thailand, 2024, p. 040013. doi: 10.1063/5.0223850.

[22] P. Mana, R. P. R. Allo, N. Paranoan, and M. Parapak, “Analysis of Buntu Burake Tourism Object Management and Its Impact on Community Welfare In Tana Toraja,” in *Proceedings of the Conference on SDGs Transformation Through the Creative Economy: Encouraging Innovation and Sustainability (TCEEIS 2023)*, vol. 271, I. Abrori, I. Supheni, M. Mudhofar, W. Murniati, N. Irawan, A. Ikhsan Lubis, W. Qiang, Moh. Hudi Setyobakti, Muh. Barid Nizarudin Wajdi, Tarjo, and E. Sujana, Eds., in Advances in

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v10i1.25597

Economics, Business and Management Research, vol. 271. , Dordrecht: Atlantis Press International BV, 2023, pp. 32–37. doi: 10.2991/978-94-6463-346-7_7.

[23] T. Natsir, B. Rauf, F. Syafar, A. W. Haedar, and F. Najamuddin, "Visualization Mapping of the Socio-Technical Architecture based on Tongkonan Traditional House," *JOIV : Int. J. Inform. Visualization*, vol. 7, no. 3, pp. 1007–1015, Sep. 2023, doi: 10.30630/joiv.7.3.1788.

[24] L. K. Wibisono, M. E. Limbongan, and D. Ramba, "Building Tana Toraja Towards a Smart City," *Eduvest*, vol. 4, no. 2, pp. 480–487, Feb. 2024, doi: 10.59188/eduvest.v4i2.1039.

[25] D. Iskandar, S. Hanafi, F. Nadiyah, G. Ramadhan, and P. Larasati, "Proposed Virtual Reality Implementation in Toraja Tourism to Improve Accessibility and Foreign Interests," in *Proceedings of the 5th International Conference on Administration Science, ICAS 2023, 29 November 2023, Bandung, West Java, Indonesia*, Bandung, Indonesia: EAI, 2024. doi: 10.4108/eai.29-11-2023.2347792.

[26] 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, p. 114324, Apr. 2021, doi: 10.1016/j.eswa.2020.114324.

[27] L. Esmaeili, S. Mardani, S. A. H. Golpayegani, and Z. Z. Madar, "A novel tourism recommender system in the context of social commerce," *Expert Systems with Applications*, vol. 149, p. 113301, Jul. 2020, doi: 10.1016/j.eswa.2020.113301.

[28] X. Nan, Kayo Kanato, and X. Wang, "Design and Implementation of a Personalized Tourism Recommendation System Based on the Data Mining and Collaborative Filtering Algorithm," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–14, Aug. 2022, doi: 10.1155/2022/1424097.

[29] S. Renjith, A. Sreekumar, and M. Jathavedan, "An extensive study on the evolution of context-aware personalized travel recommender systems," *Information Processing & Management*, vol. 57, no. 1, p. 102078, Jan. 2020, doi: 10.1016/j.ipm.2019.102078.

[30] Z. K. A. Baizal, D. Tarwidi, and B. Wijay, "Tourism Destination Recommendation Using Ontology-based Conversational Recommender System," *IJCDS*, vol. 10, no. 1, pp. 829–838, Nov. 2021, doi: 10.12785/ijcds/100176.

[31] K. A. Fararni, F. Nafis, B. Aghoutane, A. Yahyaouy, J. Riffi, and A. Sabri, "Hybrid recommender system for tourism based on big data and AI: A conceptual framework," *Big Data Min. Anal.*, vol. 4, no. 1, pp. 47–55, Mar. 2021, doi: 10.26599/BDMA.2020.9020015.