Available online at http://scik.org
J. Math. Comput. Sci. 2022, 12:62
https://doi.org/10.28919/jmcs/7086
ISSN: 1927-5307

A COMBINATION OF ALGORITHM AGGLOMERATIVE HIERARCHICAL CLUSTER (AHC) AND K-MEANS FOR CLUSTERING TOURISM IN MADURA-INDONESIA

EKA MALA SARI ROCHMAN*, ACH. KHOZAIMI, IKA OKTAVIA SUZANTI, HUSNI,
ROHMATUL JANNAH, BAIN KHUSNUL KHOTIMAH, AERI RACHMAD

Department of Informatics Engineering, Faculty of Engineering, University of Trunojoyo Madura, Indonesia

Abstract: The development approach through the tourism sector is one of the programs launched by the government since 2016. However, the development approach is not carried out in all areas because the number of accommodation and public facilities is minimal and uneven, one of which is in Madura. With so many tourist objects in Madura, it is necessary to distribute the development of public facilities and analyze tourism that has a non-strategic distance to public facilities to help increase tourist visits. This study builds a system for clustering tourist attractions in each district in Madura based on the distance to public facilities which include hotels, gas stations, restaurants, and mosques which are important criteria and considerations for tourists in visiting a tourist location. The method used in this research is a combination of the AHC method with K-Means. The test results of the AHC, K-Means method, and the combination of AHC and K-Means methods using the Silhouette Coefficient method indicate that the AHC and K-Means combination method is the best method with a Silhouette Coefficient value of 0.8055 for k=2 and is classified as a strong structure, for the K method. -Means produces the highest Silhouette Coefficient value of 0.638. While the AHC

*Corresponding author
E-mail address: ekamala.sari@yahoo.com
Received December 15, 2021

method produces the highest Silhouette Coefficient value of 0.707.

Keywords: clustering tourist attractions; AHC; K-means; tourist location.

2010 AMS Subject Classification: 93A30.

1. INTRODUCTION

Tourism is one of the sources of state and regional revenue, therefore the development of a well-managed tourism sector will be able to attract domestic and foreign tourists to come and spend their money in tourist activities [1]. According to Yoeti, efforts to attract tourists to visit tourist destinations must have several tourism components which include tourist transportation, accommodations, bars and restaurants, tourist objects, and tourist attractions [2]. Meanwhile, according to Marpaung, tourism is a temporary movement carried out by humans to get out of routine work so that they need facilities to meet their needs [3].

Although the Indonesian government emphasizes tourism development as one of its development sector priorities, this approach has not been applied consistently across the region. In Madura, there is very little tourism activity (East Java Provincial Tourism Office 2007). The number of accommodations in Bangkalan and Sampang remained the same between 2005 and 2007. A slight increase was found in Pamekasan where the number of accommodations increased from 10 to 11 and in Sumenep the accommodation business increased from 5 to 7 [4]. According to data from the Central Java Statistics Agency in 2016-2017, it shows that the number of accommodations owned by each Regency on Madura Island is in the tenth lowest rank, even Sampang and Bangkalan Regencies have the least accommodation compared to other Regencies throughout East Java [5].

Accommodation is a means to provide lodging services that can be complemented by food and drink services and other services. The gas station variable was chosen because the gas station has an important role to refuel the means of transportation used to reach tourist destinations, both private transportation, and transportation provided by a travel agency. In addition, according to Yoeti, public infrastructure consists of a network of roads, bridges, transportation, and other infrastructure consisting of the provision of clean water, electricity, telecommunications facilities, and gas stations [2]. The next variable used in this study is the mosque. In addition to the general facilities and infrastructure described above, places of worship are important supporting infrastructure for tourists. The mosque was chosen because Indonesia is a country that has the largest Muslim population in the world. The mosque is one of the Islamic attributes in tourism [6]. Grouping of tourist objects in Madura will be processed using the Clustering technique and is one way of processing data in Data mining which is a process using statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from various databases. Clustering is an activity (task) that aims to group data that has similarities between one data and another into clusters or groups so that data in one cluster has a maximum similarity (similarity) and data between clusters has a minimum similarity [7]. Currently, there are many methods used for clustering such as LVQ (Learning Vector Quantization), SOM (Self Organizing Map), Fuzzy C-Means, K-Means, and so on. The method that has been used to classify tourist objects is the K-means method and the Fuzzy C-Means method. The K-Means method is the most popular method used because it has several advantages, including this algorithm is simple and easy to implement, besides the K-Means algorithm can group data in large enough quantities with relatively fast and efficient computation time [8]. Fuzzy C-Means algorithm is more used for datasets with many (varied) attributes, while K-Means Cluster is more used for datasets with few attributes [9][10][11].

Meanwhile, in 2015 a study was conducted using the K-Means method on the nutritional status of toddlers, this study states that the K-Means algorithm only has an accuracy value of 34% [9]. One of the shortcomings of the K-Means method is that there is no definite provision in determining the best initial center of the cluster while determining the initial center of a different cluster will result in different memberships [12][13][14].

Based on previous studies that have been carried out using the Clustering technique, the contribution of this research is to group tourist attractions in Madura by combining the AHC and K-Means Cluster methods with the parameter (attribute) of distance to public facilities (mosques, hotels, restaurants and gas stations). The combination of the AHC method with the K-Means Cluster is intended so that the results of the cluster formed are better. The AHC method will be used to determine the initial center of the cluster with the Single Linkage approach which looks for the distance of two clusters according to the shortest distance between two members in one cluster, then the grouping process will be carried out using the K-Means Cluster method.

2. PRELIMINARIES

Data mining is a method used in the large-scale data processing. Therefore, data mining has a very important role in several areas of life including industry, finance, weather, science, and technology [15]. In data mining, some methods can be used such as classification, clustering, regression, variable selection, and market basket analysis [7]. According to Larose, there are six functions in data mining, namely, a description function, an estimation function, a prediction function, a classification function, a cluster function, and an association function. According to Berry and Browne, the six data mining functions can be divided into [16]:

1. Minor functions or additional functions, which include description, estimation, and prediction.
2. Major function or main function which includes classification, grouping, and association.

A. Cluster

Clustering is a way of processing data in data mining that aims to group or divide data into several parts [17]. Grouping analysis or clustering is the process of dividing data in a set into several groups whose data similarity in one group is greater than the similarity of the data with data in other groups [15]. Clustering is a type of classification on a finite set of objects. Clustering consists of several similar objects grouped. The relationships between objects are represented in the matrix between the rows and columns according to the objects. Objects are specified as patterns or points in dimensional space, the probability of the distance between pairs of points is calculated using the Euclidean Distance technique [18].

The potential for clustering can be used to determine the structure in the data which can be further used in a wide variety of applications such as classification, image processing, and pattern processing [10]. In the cluster analysis process, the method used to divide the data into subsets of data based on the similarity or similarity that has been determined previously. So, cluster analysis, in general, can be said that [19]:

1. The data contained in one cluster has a high degree of similarity.
2. The data contained in a different cluster has a low level of similarity.

B. Metode Agglomerative Hierarchical Cluster (AHC)

The AHC method is a bottom-up hierarchical clustering method that combines n clusters into a single cluster. The bottom-up algorithm is good at identifying small groups. This method begins by placing each data object as a separate cluster and then combining these clusters into a larger cluster until all objects are united in a single cluster [9]. To calculate the distance between clusters in the AHC algorithm can be done using the Single Linkage method. The single Linkage Algorithm is to find the distance of two clusters according to the shortest distance between two members in two clusters [13]. Measurement of the distance of two clusters in Single Linkage using the minimum distance formula (minimum proximity) in equation (1).

$$d(UV)W = min(d_{uw}, d_{vw}) \tag{1}$$

d_{uw} is the distance between the nearest neighbors of the cluster U and W d_{vw} is the distance between the nearest neighbors of cluster V and W.

AHC with a single linkage method is used to determine the initial centroid (the central point of the cluster) used in the K-Means Cluster method in classifying tourist objects. In general, the centroid of the K-Means Cluster method is determined randomly so that the solution is local optimum [13].

The steps of the AHC method are as follows [9]:

1. Starting with N clusters, each cluster contains a single entity and an N × N symmetric matrix of distance or similarity $D = d_{ik}$.
2. Find the distance matrix for the closest pair of clusters. Suppose that the distance between the most similar clusters U and V is denoted by d_{uv}.
3. Merge clusters U and V. Label the new cluster formed with (UV), then update the entries in the distance matrix by:

a. Delete the rows and columns corresponding to clusters U and V.

b. Adds a row and column that gives the distances between the cluster (UV) and the remaining clusters.

4. Repeat steps 2 and 3 N-1 times. All objects are in a single cluster after the last algorithm. Then note the identity of the merged cluster and the levels (distance/similarity) at which the join is placed.

After the iteration ends, the average calculation of the amount of data in the last iteration is carried out. The average is set as the centroid for the K-Means Cluster method

C. K-Means Cluster

K-Means Cluster is a non-hierarchical data grouping method that partitions data into two or more forms. This method partitions the data so that data with the same characteristics are put into one group while different data are grouped into another group [12]. K-Means Cluster is a top-down algorithm that is good at identifying large groups. K-Means Cluster can group large amounts of data with relatively fast and efficient computation time. The result of clustering with the K-Means Cluster is very dependent on the initial center of the cluster. The results of grouping with the K-Means Cluster are good if the initial center of the cluster is determined correctly [8]. The steps of the K-Means Cluster method are as follows [20]:

1. Determine k as the desired number of clusters and the desired distance matrix.
2. Select k data set x as the centroid. In this study, the centroid value was taken from the results of the calculation of the AHC method.
3. Allocate all data from the nearest centroid to the distance matrix that has been determined by using the Euclidean Distance formula in equation 2 below:

$$d(x,y) = |\,|x-y|\,| = \sqrt{\sum_{i=1}^{n}(xi-yi)^2} \quad (2)$$

xi is the object x to -i, while yi is object y to -i, and n is the number of objects

4. Recalculate the new centroid based on the data that follows each cluster.

5. Repeat steps (3) and (4) until no data moves.

Record the results of the cluster after no data transfer occurs.

D. Silhouette Coefficient

Testing on the grouping of tourist objects in Madura aims to determine the level of performance of the method used. Silhouette Coefficient is an evaluation method to test the accuracy of a cluster that has been formed from the clustering process. This method is a combination of the separation and cohesion methods [21], the calculation steps are:

1. Calculate the average distance of the object with all other objects in the same cluster

$$a(i) = \frac{1}{[A]-1} \sum j \in_{A,j \neq i} \quad d(i,j) \tag{3}$$

a(i) is the average difference of object (i) to all other objects in A, d(i,j) is the distance between data i and data j, while A itself is a cluster

2. Calculate the average distance of the object with all other objects in other clusters, then take the minimum value with the equation: (4) where:

$$d(i,C) = \frac{1}{|c|} \sum j \in C \quad d(i,j) \tag{4}$$

d(i,C) = Average difference of object (i) to all other objects in C

C = cluster other than cluster A or cluster C is not the same as cluster A

3. After calculating d(i,C) for all C, then the smallest value is taken with the equation:

b(i) = minc≠adi,C (5)

Cluster B that reaches the minimum (that is, d(i,B)) is called a neighbor of object(i). This is the second-best cluster for object(i).

4. Calculate the value of the Silhouette Coefficient with the equation:

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \tag{6}$$

Based on table 1, the Silhouette Coefficient value is divided into 4 ranges, the first in the range of more than 0.7 to 1 is classified as a strong structure, which means that the structure for each cluster membership is correct and the resulting cluster is the best, the value of a(i) or the distance between data in one cluster is small or close to 0 and the value of b(i) or the distance between data is large so that the Silhouette Coefficient value is close to 1 [21]. The medium structure has a value range of more than 0.5 to 0.7, meaning that the results of placing data in each cluster are standard, the value of a(i) is moderate and the value of b(i) is large. The weak structure has a value range of more than 0.25 to 0.5, meaning that the resulting cluster structure is weak and requires additional methods, the a(i) value is close to 1 and the b(i) value is almost the same as the a(i) value. Meanwhile, unstructured cluster membership has a value range of less than 0.25 which indicates that the resulting cluster does not have an unclear structure, the value of a(i) is greater than the value of b(i)

Table 1. Silhouette Values

NO	Value Silhouette Coefficient	Structure
1	0.7<SC<=1	Strong Structure
2	0.5<SC<=0.7	Medium Structure
3	0.25<SC<=0.5	Weak Structure
4	SC<=0.25	No Structure

The results of the calculation of the Silhouette Coefficient value vary with a range of -1 to 1. The clustering value can be said to be good if it is positive, namely (ai < bi) and ai is close to 0. With this, the maximum Silhouette Coefficient value will be 1 when ai = 0. If s(i) = 1 indicates that cluster i has been in the right cluster. However, if the value of s(i) is 0 then object i is between two clusters, so the object can be said to have an unclear structure [21].

3. MAIN RESULTS

A. Data Collection

The research was conducted on tourist objects located on the island of Madura. Tourist object data was obtained from the Department of Tourism, Culture, and Sports. Data on the distance of tourist attractions to public facilities (mosques, hotels, restaurants, and gas stations) were obtained from google maps. The method used is a combination of the AHC Single Linkage model with the K-Means Cluster. The Cluster feature parameter used is the distance of the tourist attraction to hotels, restaurants, gas stations, and mosques. Table 1 shows the number of public facilities in each district.

Table 2. Number of public facilities in each district

No.	Regency	Number of hotels	Number of restaurants	Number of gas stations	Number of mosques
1	Bangkalan	6	12	9	17
2	Sampang	7	11	8	14
3	Pamekasan	13	13	11	14
4	Sumenep	17	10	11	17
Total		43	46	49	62

B. AHC and K-MEANS Combination

The output generated in the grouping of tourist objects in Madura with the combination of the AHC method with K-Means Cluster is a group of tourist objects in Madura which consists of several groups according to the value of k. Based on Figure 1, the process of grouping tourist objects begins with the input of the tourist attraction files along with the distance to public facilities. Then the inputted data is processed using the AHC method. After grouping with the AHC method, the data will be grouped using the K-Means method and the resulting grouping of tourist objects. Then the results of cluster membership were tested using the Silhouette Coefficient method.

To clarify the flow of the system as a whole, the following describes the sequence of running the system from the beginning until the Silhouette Coefficient testing process is carried out.

1. Input the tourist attraction file along with the distance to public facilities into the tourist attraction grouping application.
2. Choose the value of k according to the cluster to be formed.
3. Data Normalization Process.
4. The grouping process uses the AHC method. The flowchart of the grouping process using the AHC method can be seen in Figure 2.

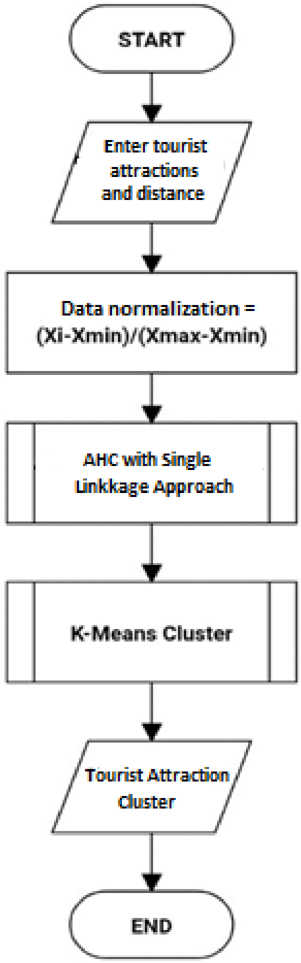


Figure 1. System Flow

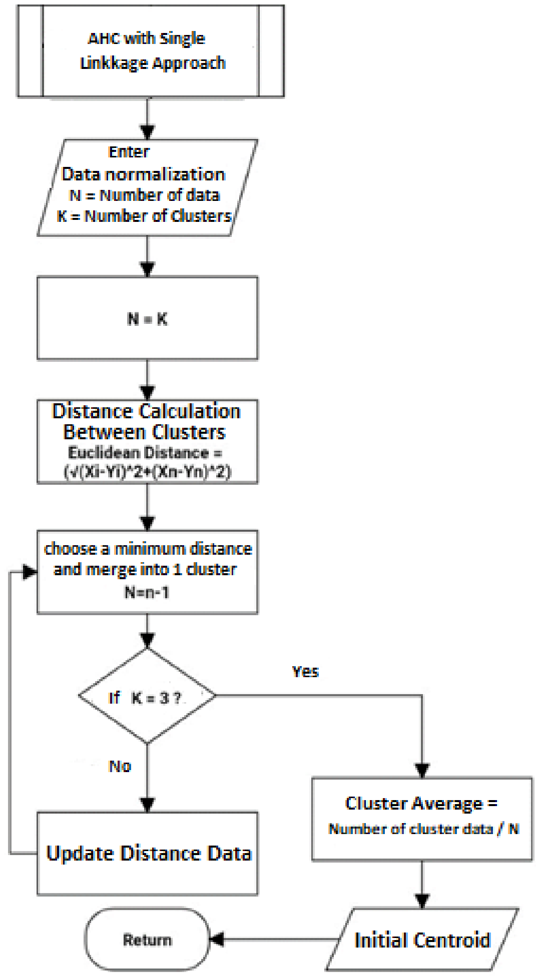


Figure 2. Flowchart of AHC Method

Based on Figure 2, the first step in the calculation process for the AHC method with the Single Linkage approach is that each data is initialized as a cluster. In this section, each data is used as a cluster. Then calculate the distance between clusters (data) with the Euclidean Distance formula. After calculating the distance of each cluster, the data that has the minimum distance are combined and made into one cluster. If the iteration reaches the last data or several k then the process is continued with the analysis of the cluster results according to the number of clusters that have been formed. Furthermore, from several clusters that have been formed, the average value of the data in each cluster is calculated and used as the centroid value. The centroid value is taken and used as the centroid in the K-Means Cluster method to group tourist objects

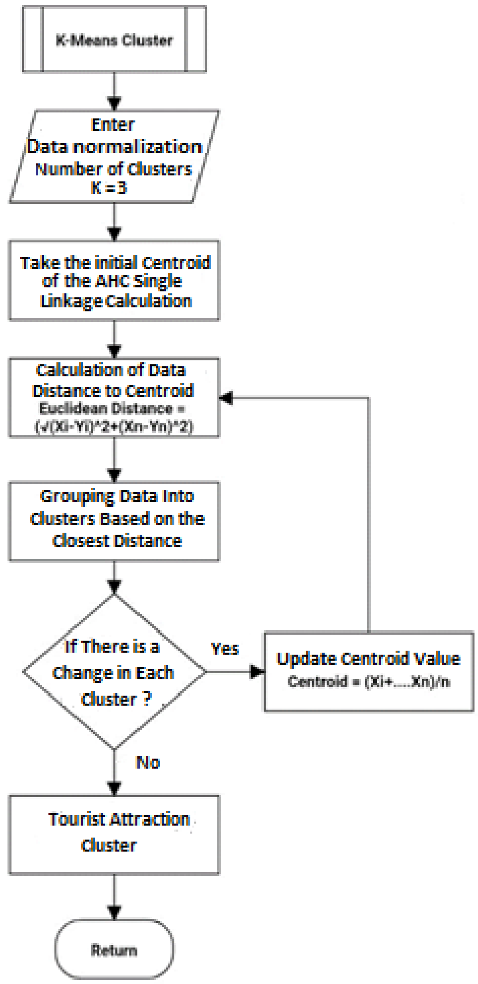


Figure 3. K-Means Flowchart

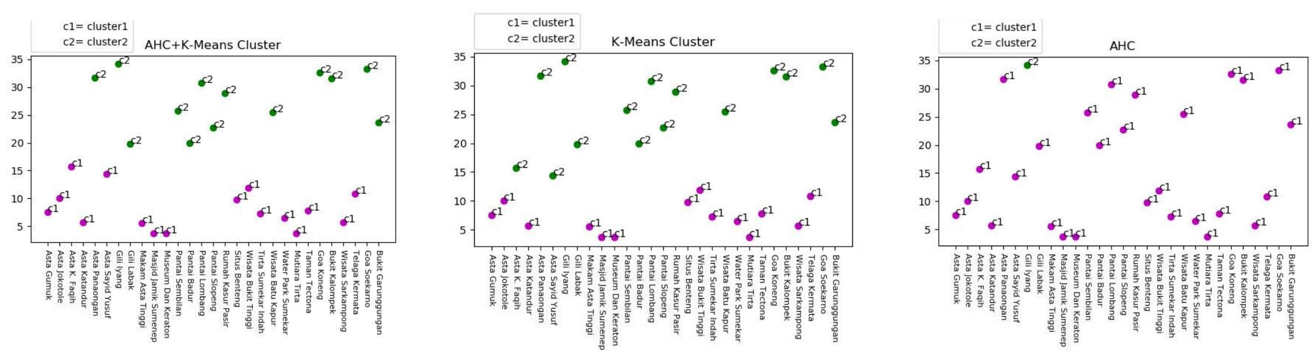
5. The next process is the calculation of the K-Means Cluster method using the centroid value generated from the AHC method. The flowchart of the K-Means Cluster method can be seen in Figure 3.

   Based on Figure 3, the first step of the K-Means Cluster method is the initialization of the centroid generated by the Single Linkage AHC method as the initial centroid of the K-Means Cluster method. Determination of k clusters according to the number of centroids that have been formed in the AHC method. Furthermore, the calculation of the distance of each object to each centroid is carried out using the Euclidean Distance formula. Then each data is combined according to the closest distance to the centroid. Then the calculation of the new centroid is carried out according to the data of the new members that have been grouped and the distance calculation is carried out again according to the distance to the new centroid. This calculation continues until the membership data does not change the cluster.

6. From the grouping of the AHC and K-Means Cluster methods, groups of tourist objects are generated according to the input k value.
7. The next process is testing the results of cluster membership with the Silhouette Coefficient method.
8. The Silhouette Coefficient value is displayed.
9. The process is complete.

C. Grouping of tourist objects based on the distance of public facilities in Sumenep Regency

The first test scenario was carried out on each data for each district in Madura. Tourist objects are grouped based on the distance to public facilities in the form of hotels, gas stations, restaurants, and mosques by combining all public facilities in each district into grouping criteria. This test was carried out using the values of k=2 and k=3. Following are the results of grouping tourist attractions in Sumenep with a value of k=2, which can be seen in Figure 4 (a)(b) and (c)



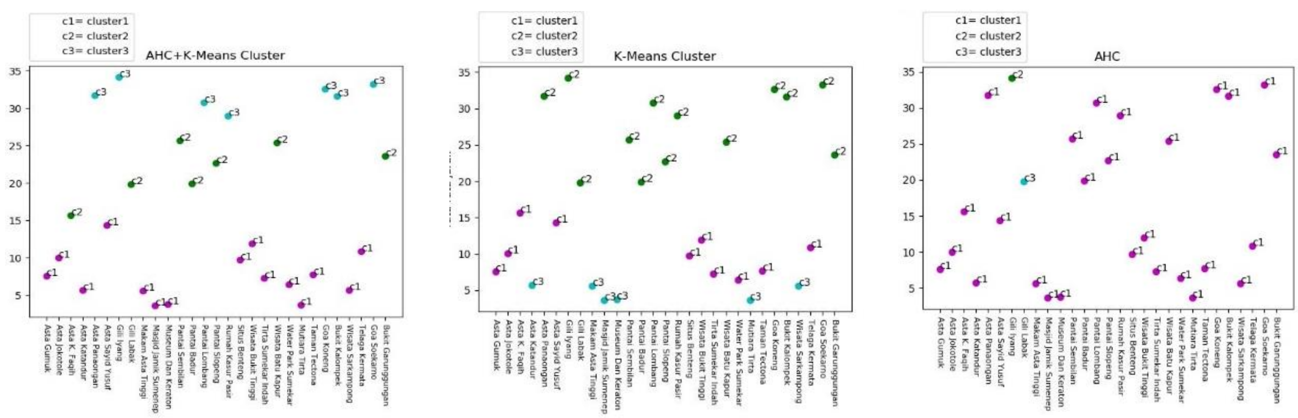
(a) (b) (c)

Figure 4. Grouping of attractions in Sumenep with a value of k=2

Figure (a) is the result of grouping using the AHC and K-Means method, Figure (b) is the result of grouping using the K-Means Cluster method, and Figure (c) is the result of grouping using the AHC method. The difference in grouping results with the combination of the AHC + K-Means method with the K-Means Cluster is the Gili Labak, Badur Beach, and Slopeng Beach attractions, while the AHC method only produces 1 tourist attraction as a member of Cluster 2, namely Gili

Iyang. The AHC + K-Means method produces a Silhouette Coefficient value of 0.6228, the K-Means Cluster method produces a Silhouette Coefficient value of 0.5698 while the AHC method produces a Silhouette Coefficient value of 0.271.

Therefore, the AHC + K-Means method is better than the K-Means Cluster and AHC methods. The results of grouping tourist objects in Sumenep Regency based on the distance to public facilities in Sumenep Regency with a value of k = 3 can be seen in Figures 5 (a), (b) and (c). Figure (a) is the result of grouping using the AHC + K-Means method, Figure (b) is the result of grouping using the K-Means Cluster method, and Figure (c) is the result of grouping using the AHC method.



(a) (b) (c)

Figure 5. Grouping of attractions in Sumenep with a value of k=3

Based on the grouping process that has been carried out the AHC + K-Means method produces a Silhouette Coefficient value of 0.499, while the K-Means Cluster method produces a Silhouette Coefficient of 0.323, while the AHC method produces a Silhouette Coefficient of 0.090. The a(i) value in the AHC+K-Means Cluster method is denser or smaller than the a(i) value in the K-Means. Cluster method so that the Silhouette Coefficient value for the AHC+K-Means Cluster method is higher than the K-Means Cluster method. As for the AHC method, the results of the Cluster membership that are formed are not structured, this indicates that there is a mismatch in the placement of Cluster members. The AHC+K-Means Cluster method is better than the K-Means Cluster method or the AHC method for grouping tourist objects in Sumenep Regency with a value of k=3.

D. Testing the value of k on the grouping of tourist objects based on public facilities

Test scenario 6 was carried out to determine the effect of the k value on the Silhouette Coefficient value generated in the grouping process with a combination of AHC and K-Means methods. This test is carried out on data in each Bangkalan Regency with initial k = 2 and final k = 10. The first test was carried out on data on tourist objects in Sumenep Regency.

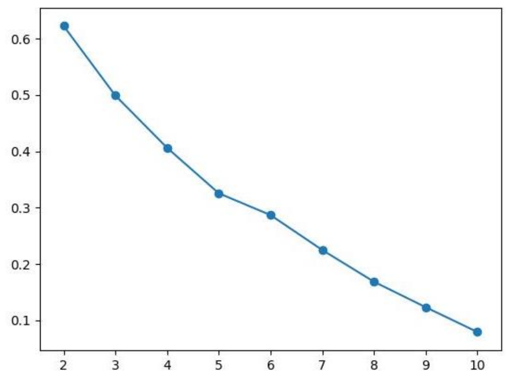


Figure 6. Testing tourist attraction data in Sumenep Regency

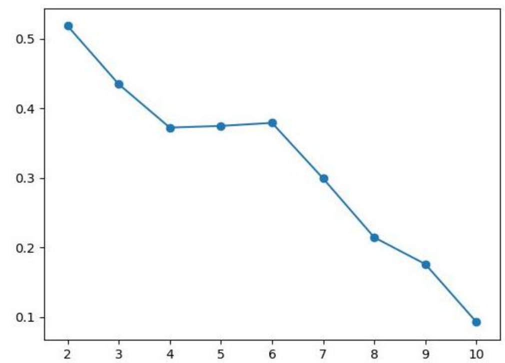


Figure 7. Testing tourist attraction data in Pamekasan Regency

Based on Figure 6, the highest Silhouette Coefficient value is obtained when the value of k=2 is 0.6228, while the lowest Silhouette Coefficient value is at k=10 with a value of 0.0793. Subsequent testing was carried out on tourist attraction data in Pamekasan Regency. The test results can be seen in Figure 7. Based on Figure 7, the highest Silhouette Coefficient value is obtained when the value of k=2 is 0.5182 and the lowest Silhouette Coefficient value is below 0.0926 at k=10. The next test was carried out on the data of tourist objects in Sampang Regency. The test results can be seen in Figure 8.

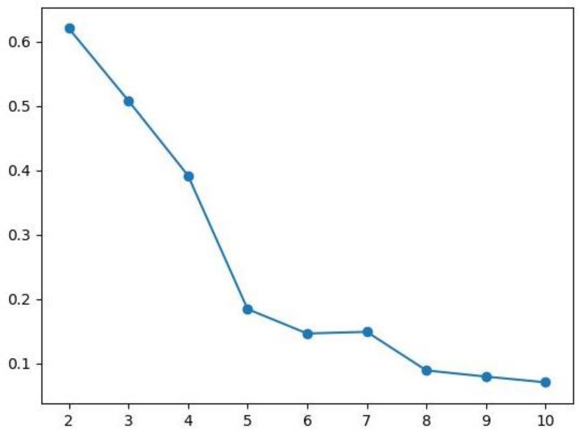


Figure 8. Testing tourist attraction data in Sampang Regency

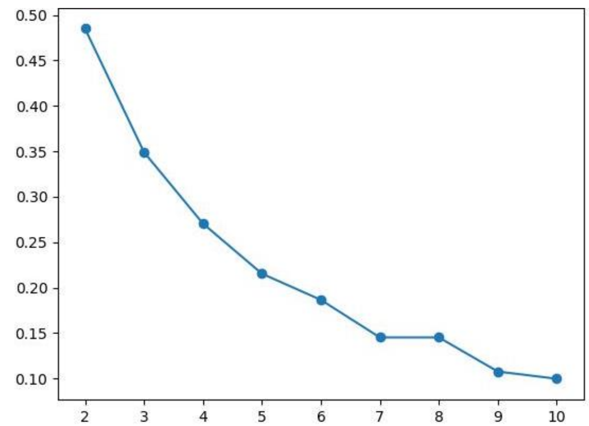


Figure 9. Testing tourist attraction data in Bangkalan Regency

Based on Figure 9, the highest Silhouette Coefficient value is obtained when the value of k=2 is 0.6207 and the lowest Silhouette Coefficient value is below 0.070 when k=10. Subsequent testing was carried out on tourist attraction data in Bangkalan Regency. The test results can be seen in Figure 4.136. The highest Silhouette Coefficient value is obtained when the value of k=2 is more than 0.4851 and the lowest Silhouette Coefficient value is 0.0997 when k=10. Based on the tests carried out on the value of k, it can be concluded that the value of k is inversely proportional to the value of the Silhouette Coefficient. The smaller the value of k, the greater the value of the Silhouette Coefficient generated by the AHC + K-Means method.

E. Performance Testing Results of AHC + K-Means and K-Means methods using Silhouette Coefficient

Based on the trial scenario that has been carried out on the grouping of tourist attractions in several districts on Madura Island based on public facilities and some features, the test results are presented in tabular form to determine the best performance formed by the AHC + K-Means method and its comparison with the K-Means method. and the AHC method. Table 3 is the result of testing the performance of the AHC + K-Means, K-Means, and AHC methods.

Table 3. Performance test results

District/ criteria	AHC + K-Means Method		K-Means Method		AHC Method	
	***k*=2**	***k*=3**	***k*=2**	***k*=3**	***k*=2**	***k*=3**
Sumenep/ Public Facilities	0.622	0.499	0.578	0.282	0.251	-0.090
Pamekasan/ Public Facilities	0.518	0.434	0.336	0.427	0.137	0.460
Sampang/ Public Facilities	0.620	0.508	0.620	0.384	0.118	0.534
Bangkalan/ Public Facilities	0.485	0.349	0.414	0.105	0.238	0.281
Sumenep/ Hotel	0.683	0.519	0.683	0.427	-0.031	0.455
Sumenep/ Restaurant	0.672	0.595	0.652	0.576	0.672	0.388
Sumenep/ Gas Station	0.545	0.423	0.494	0.408	0.200	0.416
Sumenep/ Mosque	0.675	0.575	0.658	0.427	0.248	-0.101
Pamekasan/ Hotel	0.707	0.626	0.478	0.626	0.707	0.489
Pamekasan/ Restaurant	0.674	0.589	0.492	0.542	0.674	0.424
Pamekasan/ Gas Station	0.654	0.407	0.654	0.309	0.654	0.520
Pamekasan/ Mosque	0.556	0.436	0.544	0.355	0.628	0.436
Sampang/ Hotel	0.805	0.665	0.449	0.638	0.620	0.534
Sampang/ Restaurant	0.674	0.453	0.246	0.459	0.620	0.624
Sampang/ Gas Station	0.703	0.503	0.665	0.514	0.703	0.596
Sampang/ Mosque	0.594	0.594	0.594	0.594	0.594	0.427
Bangkalan/ Hotel	0.571	0.644	0.557	0.533	0.559	0.483
Bangkalan/ Restaurant	0.588	0.456	0.588	0.429	0.588	0.372
Bangkalan/ Gas Station	0.495	0.302	0.433	0.233	0.379	0.009
Bangkalan/ Mosque	0.407	0.297	0.368	0.146	0.199	0.226
Average Value	0.6128	0.4940	0.5254	0.4211	0.4389	0.3745

Based on Table 3. The AHC + K-Means method has better performance than the K-Means method and the AHC method. The average result of the Silhouette Coefficient value in the AHC + K-Means method is 0.6128 for the value of k=2 and 0.4940 for the value of k=3, while the K-Means method is 0.5254 for k=2 and 0.4211 for k=3. Meanwhile, the AHC method produces an average of 0.4389 for k=2 and 0.3745 for k=3. The highest performance was obtained by the AHC + K-Means method in grouping tourist attractions in Sampang Regency based on the distance to the hotel with a Silhouette Coefficient value of 0.80559 and classified as a strong structure.

CONCLUSION

From the results of several research trial scenarios that have been carried out, conclusions can be drawn, namely:

1. Based on the Silhouette Coefficient value, the AHC and K-Means combination method is better than the AHC or K-Means Cluster method
2. The highest Silhouette Coefficient value is obtained when the value of k=2. Therefore, the value of k is inversely proportional to the value of the Silhouette Coefficient.
3. The highest Silhouette Coefficient value for the AHC method is 0.707 in the grouping of tourist objects in Pamekasan based on the distance to the hotel.
4. The highest Silhouette Coefficient value for the K-Means Cluster method is 0.683 in the grouping of tourist objects in Sumenep based on the distance to the hotel.
5. The highest Silhouette Coefficient value for the AHC and K-Means combination method is 0.805 in the grouping of tourist objects in Sampang based on the distance to the hotel.

ACKNOWLEDGMENT

The authors would like to thank University of Trunojoyo Madura, for the opportunity to make research with contract number 3070/UN46.4.1/PT.01.03./2021. And also, thanks to the Government tourism office in Madura (Bangkalan, Sampang, Pamekasan and Sumenep) for their assistance in providing data in this research.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] E. M. S. Rochman, I. Pratama, A. Rachmad, Implementation of fuzzy mamdani for recommended tourist locations in Madura-Indonesia. J. Phys.: Conf. Ser. 1477 (2020), 022033.

[2] D. R. Anamisa, E. M. S. Rochman, A. Rachmad, Analysis of the result of the ant colony system adaptive on

tourism object. J. Multidiscip. Eng. Sci. Technol. 6 (2019), 9446-9448.

[3] D. R. Anamisa, A. Rachmad, E. M. S. Rochman, Ant colony system based ant adaptive for search of the fastest route of tourism object Jember, East Java. J. Phys.: Conf. Ser. 1477 (2020), 052053.

[4] E. M. S. Rochman, I. Pratama, A. Rachmad, Application based of tourist attraction selection with fuzzy Tahani. Jurnal Pekommas, 5 (2020), 195-202.

[5] A. Rachmad, E. M. S. Rochman, D. Kuswanto, I. Santosa, R. K. Hapsari, T. Indriyani, E. Purwanti, Comparison of the traveling salesman problem analysis using neural network method. In: International Conference on Science and Technology (ICST 2018), Atlantis Press (2019), pp.1057-1061.

[6] Anonimous, Central Bureau of Statistics of East Java, Number of hotel accommodations by Regency/City in East Java Province, BPS Provinsi Jawa Timur (Statistics Jawa Timur), 2019.

[7] Irda, The influence of Islamic attributes on Islamic tourism Satisfaction of foreign tourists visiting the city of Padang, Menara Ekonomi, 4 (2019), 20-29.

[8] C. Meng, Y. Lv, L. You, Y. Yue, Intrusion detection method based on improved K-means algorithm, J. Phys.: Conf. Ser. 1302 (2019), 032011.

[9] Jamal and D. Yanto, RFM analysis and K-Means algorithm for customer loyalty clustering, Energy, (9) (2019), pp. 1-8.

[10] I. Alpiana, L. Anifah, Application of the KnA method (Combination of K-Means and Agglomerative Hierarchical Clustering) with a single linkage approach to determine nutritional status in toddlers, Indonesian J. Eng. and Technol. 1 (2) (2019).

[11] L. Maulida, The application of data mining in classifying tourist visits into leading tourist objects in the province. DKI Jakarta with k-means, Jurnal Informatika Sunan Kalijaga, 2 (3) (2018), 167-174.

[12] E. M. S. Rochman, A. Rachmad, Clustering tourist destinations based on number of visitors using the K-Mean method. In: 1st International Multidisciplinary Conference on Education, Technology, and Engineering 2019, Atlantis Press (2020), pp. 312-314.

[13] R. Wulansari, D. Hartama, Data mining: k-means algorithm on grouping foreign tourists to Indonesia by province, Seminar Nasional Sains dan Tekhnologi Informasi, 1 (1) (2018), 322-366.

[14] B. Al Kindhi, T. Arief Sardjono, M. Hery Purnomo, Hybrid k-means, fuzzy c-means and hierarchical clustering

for DNA hepatitis C, Expert Syst. Appl. 121 (2019), 373-381.

[15] A. Aprilia, W. Mistarika Rahmawati, M. Hakimah, Determination of the category of nutritional status of children under five using a combination of agglomerative clustering methods and k-means, Prosiding Seminar Nasional Sains dan Teknologi Terapan, 1 (1) (2019), 595-600.

[16] A. Kamalia, Study of tourism potential and tourist perceptions of tourism development in the monkey forest of nepa, batioh village, banyuates sampang, Bioscience Tropic, 3 (3) (2018), 53-60.

[17] A. P. Windarto, Implementation of data mining on rice imports by major country of origin using algorithm using k-means Clustering methode, Int. J. Artif. Intell. Res. 1 (2) (2017), 26-33.

[18] M. Nishom, Comparison of the accuracy of euclidean distance, minkowski distance and manhattan distance on the K-Means Clustering algorithm based on Chi-Square, Jurnal Pengembangan IT, 4 (1) (2019), 20-24.

[19] U. R. Raval, T. Jani, Implementing & improvisation of k-means Clustering algorithm, Int. J. Computer Sci. Mobile Comput. 5 (5) (2016), 191-203.

[20] E. Ariska, Implementasi agglomerative hierarchical Clustering pada data produksi dan data penjualan perusahaan, Repositori Institusi USU, 2018.

[21] A. K. Wardani, K-Means algorithm implementation for clustering of patients disease in kajen clinic of pekalongan, Jurnal Transformatika, 14 (1) (2016), 30-37.