

Mapping of Village Population Profile with Schistosomiasis Cases Using Clustering Large Applications

Mohammad Fajri¹, Rais¹, Nurul Fiskia Gamayanti¹, Siti Natazha Dg Mabaji¹, Shalsa Yunita Rahman Jati¹, Rizwan Arisandi²

¹Department of Statistics, Tadulako University, Indonesia

²Bina Nusantara University, Indonesia

Article Info

Article history:

Received : xx

Revised : xx

Accepted : 10-19-2023

Keywords:

Schistosomiasis;

Population Profile;

Cluster Analysis;

Clustering Large Application

ABSTRACT

Schistosomiasis is a tropical disease caused by *Schistosoma mansoni* (intestinal schistosomiasis) and *Schistosoma haematobium* (urogenital schistosomiasis). Schistosomiasis in Indonesia is endemic to Central Sulawesi and is commonly found in the Napu Valley and Bada Valley areas, which are administratively included in Poso District and Sigi District. The purpose of this study is to obtain information on schistosomiasis endemic areas by mapping the population profile of villages with schistosomiasis cases. This mapping is intended to provide an overview of the social and demographic conditions of villages with schistosomiasis cases. The research method used is Clustering Large Applications (CLARA). CLARA is a clustering method that is more robust to unusual data and can be applied to handle large volumes of data. The results of this study are obtained two optimum clusters, each possessing distinct characteristics as determined by schistosomiasis case indicators (population profile). Cluster 1 consists of 102 villages and cluster 2 consists of 68 villages. Cluster 2 has a high dominant value compared to cluster 1 based on the average value of all indicators per cluster. Further research can use other clustering methods or development methods from CLARA, namely CLARANS (Clustering Large Applications Based on Randomized Search). CLARANS has the ability to complete clustering tasks faster than CLARA



Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020

DOI: <https://doi.org/10.30812/varian.v7i2.3423>

Corresponding Author:

Mohammad Fajri,

Department of Statistics, Universitas Tadulako

Email: m.fajri@untad.ac.ad

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



A. INTRODUCTION

Schistosomiasis is an overlooked tropical disease which is currently estimated to affect excess of 140 million individuals. The main species responsible for schistosomiasis in humans are *Schistosoma haematobium* (urogenital schistosomiasis) and *Schistosoma mansoni* (intestinal schistosomiasis) (Deol et al., 2019). Schistosomiasis has been detected in Indonesia, specifically within the Central Sulawesi Province, and control efforts have been underway for over the last three and a half decades. However, transmission still persists to this day (Erlan et al., 2020).

Schistosomiasis in Indonesia is present in the Bada, Napu, and Lindu Highlands. This disease is widespread in several villages in Poso District and Sigi District. Schistosomiasis in Central Sulawesi Province is attributed to the trematode worm species *Schistosoma japonicum*, utilizing the snail, *Oncomelania hupensis lindoensis* as its intermediate host (Nurwidayati et al., 2019). The parasitic worms of the schistosoma emerge from snails (*Oncomelania hupensis lindoensis*) to contaminate freshwater and subsequently infect humans or mammals whose skin comes into contact with the water (Barmawi and Hatta, 2017).

According to the Directorate of Vector-Borne and Zoonotic Disease Prevention and Control (P2PTVZ) of the Ministry of Health of the Republic of Indonesia, the prevalence of schistosomiasis cases in humans in Central Sulawesi Province from 2018 to 2021

was 0,36%, 0,10%, 0,11%, and 0,3% (Direktorat Pencegahan dan Pengendalian Penyakit Tular Vektor dan Zoonotik Kementerian Kesehatan RI, 2020). This fluctuating situation is exacerbated and broadened by population migration, development, economic growth, and trade, particularly with mammalian meat as the host. Despite the overall success of control programs in reducing the number of infections, the presence of infected individuals indicates that reinfection still occurs. This is possible because there are still sources of infection from host animals and humans, enabling contact with infective larvae and the continued spread of the infection. No available treatment can halt reinfection, rendering existing treatments relatively inefficient in breaking the transmission cycle (Gordon et al., 2019).

This should receive more attention, especially from local governments regarding policy distribution to reduce the prevalence of schistosomiasis in Poso District. In order to obtain information about the grouping and mapping of village population profiles in Poso District with schistosomiasis cases, clustering can be performed to observe the similarities in characteristics of each village. This can provide information and insights into the population profiles of villages with schistosomiasis cases, both socially and demographically. Village clustering with Schistosomiasis cases can be done using the Clustering method. There are many clustering methods available, among which Clustering Large Application (CLARA) is a clustering method known for its robustness against outliers and its ability to handle large datasets. Furthermore, CLARA is more efficient than K-Medoids (PAM) in terms of computational time and deviation issues for large datasets (Muslim, 2018).

A previous study was conducted by Marsita (2021). In their research, data from earthquake-prone areas in West Sumatra were utilized using the CLARA and DBSCAN methods. The results obtained indicate the optimum condition for both methods, with a total of 5 clusters based on the silhouette index, with CLARA being the superior method. Another study was conducted by Sakinah et al. (2022) entitled spatial analysis of the spread of schistosomiasis disease using the moran index to support the eradication of schistosomiasis in the province of central Sulawesi based on a web dashboard. According to the results of the previous study, by considering the number of schistosomiasis cases and the proximity of each village, four group distribution patterns were formed. Based on (Towidjojo et al., 2023), the complicating factors in the monitoring and control of schistosomiasis are the social, demographic and behavioral conditions of the community. Therefore, it is necessary to consider these factors in shaping the schistosomiasis distribution pattern of each village but also also the social demographic indicators in each village. The novelty of this study is that it categorizes villages in Poso District not only based on the number of schistosomiasis cases in each village. So, the demographic condition in the villages also considered as criteria of grouping. Based on the aforementioned exposure issues, the purpose of this research is obtaining a mapping of village in Poso district by considering number of schistosomiasis cases and social demographic indicators in each village using CLARA method. This will provide information and descriptions of the social and demographic aspects of villages with schistosomiasis cases.

B. RESEARCH METHOD

The data in this study are the number of schistosomiasis cases in Poso District villages and population profile data, covering social and demographic conditions in the village. This study used data obtained from the Poso District Health Office and the Poso District Central Bureau of Statistics. This study incorporates several variables contained in Table 1.

Table 1. The Used Variables

Name	Variable	Unit
Z	Total Number of Schistosomiasis Cases	Cases
X_1	Height from Sea Level	Meters
X_2	Density of Population	People per meters squared
X_3	Total Population Migration	People
X_4	Total number of health facilities	Units
X_5	Number of health workers	People
X_6	Number of Marketing Facilities	Units
X_7	Number of Transport Facilities	Units

This research used the Clustering Large Applications method with the research flowchart as shown in Figure 1.

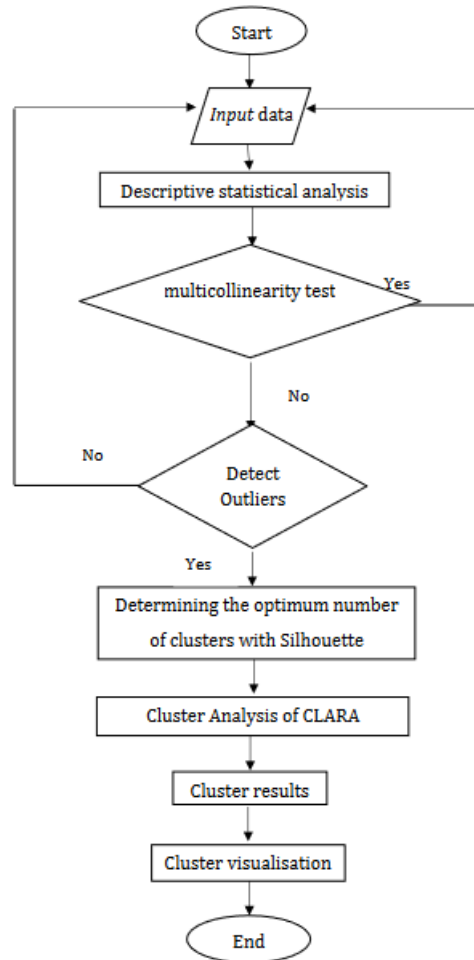


Figure 1. Flowchart

According to the flowchart depicted in Figure 1, the research can be divided into a few stages (Rifa et al., 2019):

1. Collecting and inputting data on the number of schistosomiasis cases in Poso district, Central Sulawesi.
2. Perform a descriptive statistical analysis to establish an overview of the characteristics of the variables used.
3. Standardize data
4. Test whether there is multicollinearity using the Variance Inflation Factor (VIF) test as in Equation (1)

$$VIF_i = \frac{1}{1 - R_i^2} \quad (1)$$

With R_i^2 is the coefficient of determination on variable i . If the VIF value surpasses 10, it can be inferred that multicollinearity is present (Supriyadi et al., 2017).

5. Detect the presence of outliers using the Mahalanobis squared distance measurement as in Equation (2) (Nahdliyah et al., 2019)

$$d_{MP}^2(i) = (x_i - \bar{x})' \sum^{-1} (x_i - \bar{x}) \quad (2)$$

Where,

- $d_{MP}^2(i)$: the squared Mahalanobis distance of the object at the i -th observation
- x_i : data vector of objects in the i -th observation of size $p \times 1$
- \bar{x} : average vector of each variable of size $p \times 1$
- $(x_i - \bar{x})$: matrix of the difference between the data and the average data
- $(x_i - \bar{x})'$: the transpose matrix of the difference between the data and the average data
- \sum : covariance matrix of size $p \times p$, where p is the number of variables
- \sum^{-1} : the inverse of covariance matrix of size $p \times p$, where p is the number of variables

The i -th observation is identified as an outlier if,

$$d_{MP}^2(i) > \chi_{p,1-\alpha}^2 \quad (3)$$

with $\chi_{p,1-\alpha}^2$ being the outlier limit with p number of independent variables and probability $1 - \alpha$ (Utami and Saputro, 2018).

6. Perform the algorithm steps of the Clustering Large Application method (CLARA).

- (a) Determine the value of k , which represents the desired number of clusters to form using the silhouette width as described in Equation (4).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \quad i = 1, 2, \dots, n \quad (4)$$

Where,

$s(i)$: silhouette width value for the i -th data

$a(i)$: average distance between the i -th data and all other data

$b(i)$: average distance for an i -th data point with all observations of observations

According to Faujia et al. (2022), this method measures the validation of the goodness of a data, single cluster or the whole cluster. the higher the silhouette width value, the better the clustering. Silhouette width categories can be seen in Table 2.

Table 2. Silhouette Width Category

Silhouette Width	Interpretation
$0.7 < s < 1$	Strong Cluste
$0.5 < s < 0.7$	Reasonable Cluster
$0.25 < s < 0.5$	Weak or artificial Cluster
$s \leq 1$	No Cluster found

- (b) Generate random subsets of fixed size from a given dataset. The sample size is at least $(40 + 2 * k$ or $n)$
- (c) Execute the Partitioning Around Medoids (PAM) algorithm on each of the subsets to identify the medoid for each subset. The steps of the PAM algorithm are (Fadillah, 2023):

i. Randomly generate k cluster centers (medoids)

ii. Compute the distance between each non-medoid items and the medoid in its cluster, then assign each non-medoid item to the closest medoid using Manhattan distance formula outlined in Equation (5) (Fitrayana and Saputro, 2022).

Then, the total distance is calculated.

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (5)$$

Where,

d_{ij} : measured distance between the i -th object and the j -th object

x_{ij} : data from i -th object on k -th variable

x_{jk} : data from j -th object on k -th variable

p : number of variables

iii. Randomly select a non-medoid object in every cluster to serve as the new medoid candidate.

iv. Calculate the distance between each non-medoid object and the new medoid candidate, then place the non-medoid object in the nearest medoid candidate. The total distance of all non-medoid objects is calculated.

v. Calculate the entire distance difference (S)

$S =$ entire distance on the new medoid candidate - entire distance on the old medoid.

vi. There are two states: if $S < 0$, the new medoid candidate becomes the new medoid; otherwise, if $S > 0$, the iteration stops.

In other words, if the cumulative distance from each object to the newly proposed medoid candidate exceeds the cumulative distance from each object to the current medoid, no medoid swapping occurs. This is because the new medoid candidate isn't more centrally located than the current one. Consequently, the iteration process comes to a halt.

- (d) Calculate the distance between all items in the dataset that are not medoids and those that are medoids.

- (e) The subset with the smallest sum of distances is selected.
- (f) Obtain cluster results from CLARA analysis
- (g) Visualization of clustering result data in the form of radar charts.
- (h) Draw conclusions from the analysis results.

C. RESULT AND DISCUSSION

1. Multicollinearity Test Result

The purpose of multicollinearity test is to determine whether there exists any correlation among the Z, X_1, \dots, X_7 variables. The variables used must fulfill the assumption of no multicollinearity. The assessment of multicollinearity involves examining the Variance Inflation Factors (VIF) for each individual variable. If the VIF value is less than or equal to 10, it is considered indicative of the absence of multicollinearity, and this finding is reported in Table 3.

Table 3. Test Results Variance Inflation Factors (VIF)

Variable	Z	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
VIF	1.04	1.21	1.15	1.14	1.34	1.07	1.09	1.29

Based on Table 3, it is known that the variable Z, X_1, \dots, X_7 has VIF value ≤ 10 , so it can be concluded that the assumption of no multicollinearity between variables is fulfilled.

2. Outlier Test Results

The outlier test was performed using the Mahalanobis Distance criterion at the $P < 0,05$. The Mahalanobis Distance is evaluated using χ^2 at a free degree equal to the number of variable indicators used in the study, which can be calculated using Equation 2.2. In this study, the number of indicators used was 8 so that at the level $p 0,05$ the χ^2 value was 15,50. So that if the highest Mahalanobis distance value is lower than 18,50, it means that there are no outliers.

Based on the results obtained, there are 27 villages detected outliers, namely Dodolo, Alitupu, Wuasa, Watumaeta, Tamadue, Maholo, Winowanga, Mekar Sari, Kalemago, Ratoombu, Toinasa, Pamona, Tentena, Gebang Rejo, Moengko, Kayamanya, East Gabangrejo, West Gabangrejo, Kawua, Madale, Tegal Rejo, Kasintuwu, Sangginora, Kilo, Kalora, Tambarana and Maranda. Therefore, the CLARA method is suitable for use in this study because the CLARA method is able to overcome the existence of outliers in the dataset.

3. Determining The Optimum Cluster

In this study, to cluster villages based on the population profile of villages with schistosomiasis cases in Poso District, optimum clustering was used for a good level of variation within clusters. Based on Figure 2, the number of clusters which has the highest average silhouette width is 2 number of clusters marked with a dotted line. Therefore, it can be concluded that the optimum number of clusters is 2 clusters.

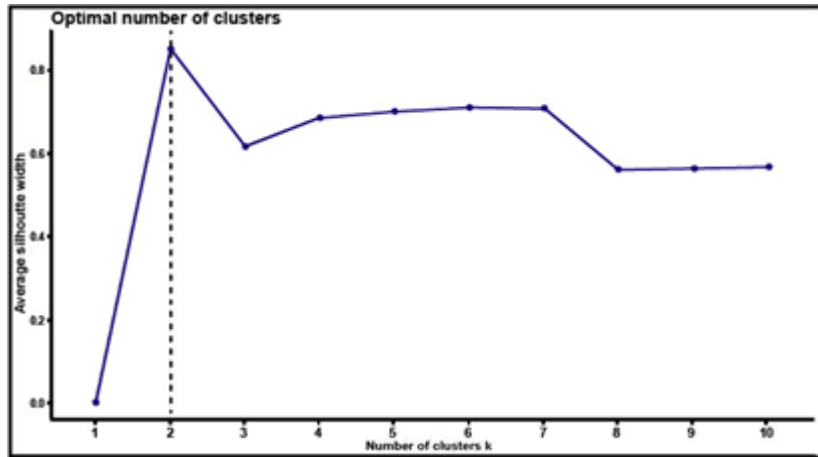


Figure 2. The Silhouette Width Method

4. Cluster Results Using CLARA

In this analysis, grouping objects into 2 clusters is carried out. Clustering using representative objects (medoids) as cluster centers for each cluster can be seen in Table 3.

Table 4. Cluster Centers

Village	Z	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
Amporiwo	0	550	35	0	2	1	0	1
Lanto Jaya	0	0	192	0	3	4	0	0

Based on Table 4, representative objects to become cluster centers are Amporiwo and Lanto Jaya villages, where Amporiwo sub-district is cluster 1 and Lanto Jaya sub-district is cluster 2, so it will follow the cluster. Each village will follow the cluster that has the closest distance to the selected cluster center.

5. Evaluation of Cluster Results

The distance measure used was Manhattan. The data consists of 170 villages that will be grouped based on the 8 variables studied with 27 villages identified as outliers. Evaluation of clustering results can use the average silhouette width value.

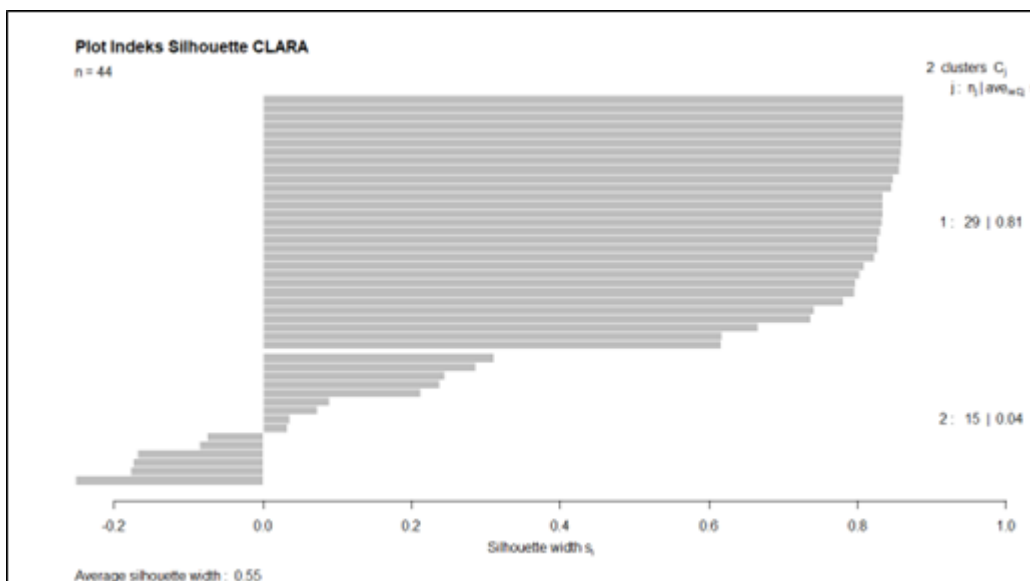


Figure 3. Evaluation of Cluster Results

Based on the simulations conducted as shown in Figure 3, the CLARA analysis with Manhattan distance obtained an average silhouette width value of 0,55. This value can be categorized that the grouping is in the good category. So, CLARA analysis is suitable for the cases.

6. Profiling Cluster Results

Table 5. Cluster Results

No	Village	Z	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	Cluster
1	Dodolo	4.18	0.39	-0.32	-0.43	-0.15	-0.32	-0.44	0.16	1
2	Kaduwaa	0.64	0.34	-0.29	-0.43	-0.15	-0.32	-0.44	0.16	1
3	Alitupu	4.18	0.36	-0.27	-0.43	0.62	0.13	1	0.4	1
4	Wuasa	1.52	0.34	-0.23	-0.43	0.62	4.15	2.43	0.16	1
5	Watumaeta	4.18	0.62	-0.19	-0.43	0.62	0.24	1	0.64	1
6	Sedoa	-0.24	0.23	-0.32	-0.43	0.62	-0.32	-0.44	0.4	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
165	Tambarana	-0.24	-1.06	-0.24	0.63	5.97	0.57	-0.44	1.37	2
166	Bakti Agung	-0.24	-1.06	0.33	-0.06	2.15	0.01	2.43	-0.08	2
169	Membuke	-0.24	-1.06	-0.31	-0.23	0.62	-0.21	-0.44	-0.08	2
170	Maranda	-0.24	-1.06	-0.31	-0.06	1.38	-0.1	-0.44	3.78	2

Based on Table 5, the results of the CLARA analysis formed 2 clusters with members in each cluster are:

- Cluster 1 consists of 102 villages, namely Dodolo Kaduwaa, Alitupu, Wuasa, Watumaeta, Sedoa, Bumi Banyusari, Tamadue, Maholo, Winowanga, Mekar Sari, Kalemago, Watutau, Wanga, Betue, Talabosa, Siliwanga, Hanggira, Lempe, Doda, Bariri, Torire, Rompo, Baliura, Katu, Tuare, Tomehipi, Kageroa, Lengkeka, Kolori, Lelio, Gintu, Runde, Badangkaia, Bakekau, Bulili, Bewa, Pada, Bomba, Taipa, Owini, Meko, Salukaia, Toinasa, Uranosari, Tonusu, Mayakeli, Buyumpondoli, Pamona, Tentena, Sangele, Peura, Dulumai, Leboni, Soe, Wera, Mayoa, Mayasari, Uelene, Pandajaya, Mayajaya, Pandayora, Bangun Jaya, Pendolo, Boe, Bancea, Panjo, Pasir Putih, Barati, Salindu, Singkona, Tolambo, Tindoli, Tokilo, Korobono, Amporiwo, Wayura, Masewe, Pancasila, Olumokunde, Kamba, Petiro, Taripa, Tiu, Poleganyara, Ompo Didiri, Kelei, Matialemba, Kancuu, Magapu, Petirodongi, Tendeadingi, Sawidago, Saojo, Uelincu, Sulewana, Lena, Sangira, Kuku, Panjoka, Sangginora, and Dewua.
- Cluster 2 consists of 68 villages, namely Tampemadoro, Pandiri, Watuawu, Tambaro, Sintuwulemba, Tagolu, Maliwuko, Silanca, Sepe, Bategencu, Toyado, Tongko, Malei, Labuan, Ratoombu, Labuadago, Gebang Rejo, Moengko, Moengko Baru, Kayamanya, Gebangrejo Timur, Gebangrejo Barat, Kayamanya Sentral, Lembomawo, Ranononcu, Kawua, Bukit Bambu, Sayo, Madale, Tegal Rejo, Lawanga Tawongan, Lawanga, Lombugia, Bonesompe, Kasintuwu, Tokorondo, Pinedapa, Tiwaa, Lape, Masani, Towu, Ueralulu, Toini, Masamba, Tabalu, Lanto Jaya, Kasiguncu, Mapane, Saatu, Betania, Bega, Tangkura, Padalembara, Malitu, Pantangolemba, Taunca, Betalemba, Patiwunga, Kilo, Trimulya, Kawende, Kalora, Tambarana, Bakti Agung, Tumora, Tobe, Membuke, and Maranda.

The findings of this study are that two clusters were formed, where each cluster has its own characteristics of village population profile data variables with schistosomiasis cases in Poso District, so it is necessary to identify each cluster formed. The average value of several variables in each cluster to determine the characteristics of each cluster is presented in Table 5.

Table 6. Cluster Averages

Cluster	1	2
Z	0.16	-0.24
X ₁	0.69	-1.03
X ₂	-0.27	0.41
X ₃	-0.29	0.43
X ₄	-0.13	0.2
X ₅	0.01	-0.01
X ₆	-0.05	0.07
X ₇	-0.04	0.06

Based on Table 6, the highest average of a cluster is colored blue while the lowest average is marked in yellow. The interpretations of the characteristics of the two clusters formed are:

1. Cluster 1 has the highest number of schistosomiasis cases, height from sea level and number of health workers with mean values of 0.16 cases, 0.69 meters and 0.01 people. Then the variable population density, the number of population migration, the number of health facilities, the number of marketing facilities and the number of transportation facilities are the lowest with an average value of -0.27 people, -0.29 people, -0.13 units, -0.05 pieces and -0.04 pieces, meaning that the villages in this cluster are areas that have a low number of majority population profile with high schistosomiasis cases.
2. Cluster 2 has the highest indicator characteristics of population density, population migration, number of health facilities, number of marketing facilities and number of transport facilities with average values of 0.41 people, 0.43 people, 0.20 units, 0.07 pieces and 0.06 pieces. Then the variables of schistosomiasis cases, height from sea level and the number of health workers are the lowest with an average value of -0.24 cases, -1.03 meters and -0.01 people. This means that villages in this cluster are areas that have a high number of majority population profile with low schistosomiasis cases.

7. Clustering Visualization Using Radar Chart

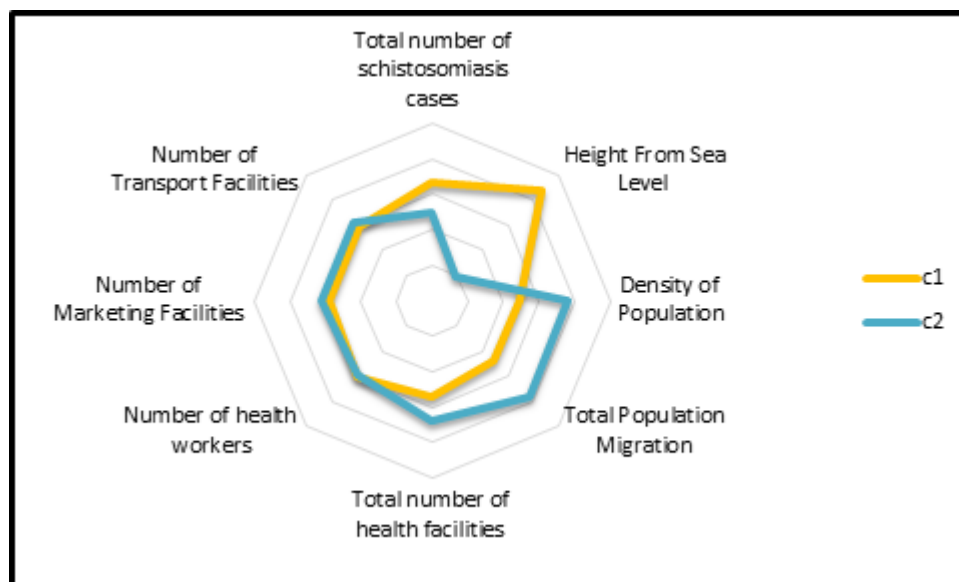


Figure 4. Visualization of Cluster Results Using Radar Chart

The results of using CLARA method in clustering villages with schistosomiasis cases in Poso District is the population profile of the village map and it visualized using a radar chart. This is done to make it easier to know the clustering results obtained. The results can be seen in Figure 4, it can be seen that cluster 2 has a high dominant value compared to cluster 1 based on the cluster average value in Table 6.

In this research, there are two cluster that established to describe villages population profile, either have a schistosomiasis case or not. The grouping has more information instead of the previous research about schistosomiasis conducted by Sakinah et al. (2022), which is only involved the number of schistosomiasis. On this research, by building the group based on schistosomiasis case and the population profile (social and demographic indicators), some pattern can be identified to mapping the characteristic of each village within group.

D. CONCLUSION AND SUGGESTION

In the clustering conducted using the Clustering Large Application (CLARA) algorithm, the results showed that the overall average silhouette width value was 0.55 and fell into the good category, while the silhouette width validity index of the optimum number of clusters formed was 2 clusters with different characteristics of each cluster based on schistosomiasis cases and population profile. Cluster 1 consists of 102 villages and cluster 2 consists of 68 villages. Cluster 2 has a high dominant value compared to cluster 1 based on the average value of all indicators per cluster. Based on the results of this study, hopefully it can be one of the references

for the government in prioritizing which villages in Poso district need more attention in reducing the spread of schistosomiasis cases. Further research can use other clustering methods or development methods from CLARA, namely CLARANS (Clustering Large Applications Based on Randomized Search. CLARANS has the ability to complete clustering tasks faster than CLARA.

ACKNOWLEDGEMENT

The research team would like to extend their gratitude to Research and Society Service Institute (LPPM) of Tadulako University for their financial assistance, which greatly facilitated the progress of this study. Additionally, special thanks are due to Dinas Kesehatan Kabupaten Poso for their valuable assistance in gathering data on schistosomiasis cases.

REFERENCES

- Barmawi, N. and Hatta, I. (2017). Pengetahuan masyarakat lindu terkait schistosomiasis di kabupaten sigi sulawesi tengah. *Jurnal Vektor Penyakit*, 11(2):49–60. <https://doi.org/10.22435/vektor.v11i2.7160.49-60>.
- Deol, A. K., Fleming, F. M., Calvo-Urbano, B., Walker, M., Bucumi, V., Gndou, I., Tukahebwa, E. M., Jemu, S., Mwingira, U. J., Alkohlani, A., et al. (2019). Schistosomiasis assessing progress toward the 2020 and 2025 global goals. *New England Journal of Medicine*, 381(26):2519–2528. <https://doi.org/10.1056/nejmoa1812165>.
- Direktorat Pencegahan dan Pengendalian Penyakit Tular Vektor dan Zoonotik Kementerian Kesehatan RI (2020). Pengendalian coronavirus disease (covid-19). *Kemendes RI*.
- Erlan, A., Widjaja, J., Nur Widayati, A., Maksud, M., Tolistiawaty, I., Mustafa, H., et al. (2020). Implementasi model bada dalam pengendalian schistosomiasis di desa lengkeka, kecamatan lore barat, kabupaten poso, sulawesi tengah implementation of the bada model in schistosomiasis control in lengkeka village, kecamatan lore barat, poso district, central sulawesi. *Buletin Penelitian Kesehatan*, 48(3):199–208. <https://doi.org/10.22435/bpk.v48i3.3246>.
- Fadillah, W. N. (2023). *PENGEMBANGAN SISTEM INFORMASI GEOGRAFIS UNTUK PEMETAAN PENYAKIT DBD DI SUMATERA UTARA MENGGUNAKAN METODE PARTITIONING AROUND MEDOIDS*. PhD thesis, UNIMED. <https://doi.org/10.35580/jmathcos.v6i2.52350>.
- Faujia, R. A., Setianingsih, E. S., and Pratiwi, H. (2022). Analisis kluster k-means dan agglomerative nesting pada indikator stunting balita di indonesia. In *Seminar Nasional Official Statistics*, volume 2022, pages 1249–1258. <https://doi.org/10.34123/semnasoffstat.v2022i1.1511>.
- Fitrayana, P. R. and Saputro, D. R. S. (2022). Algoritme clustering large application (clara) untuk menanganidata outlier. In *Prisma, Prosiding Seminar Nasional Matematika*, volume 5, pages 721–725.
- Gordon, C. A., Kurscheid, J., Williams, G. M., Clements, A. C., Li, Y., Zhou, X.-N., Utzinger, J., McManus, D. P., and Gray, D. J. (2019). Asian schistosomiasis: current status and prospects for control leading to elimination. *Tropical medicine and infectious disease*, 4(1):40. <https://doi.org/10.3390/tropicalmed4010040>.
- Marsita, D. (2021). *CLUSTERING DAERAH RAWAN GEMPA DI SUMATRA BARAT MENGGUNAKAN METODE CLUSTERING LARGE APPLICATION DAN METODE DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE*. PhD thesis, Universitas Muhammadiyah Semarang.
- Muslim, A. B. (2018). Cluster analysis using clara method for data with outlier.
- Nahdliyah, M. A., Widiari, T., and Prahutama, A. (2019). Metode k-medoids clustering dengan validasi silhouette index dan c-index (studi kasus jumlah kriminalitas kabupaten/kota di jawa tengah tahun 2018). *Jurnal Gaussian*, 8(2):161–170. <https://doi.org/10.14710/j.gauss.v8i2.26640>.
- Nurwidayati, A., Frederika, P. P., Sudomo, M., et al. (2019). Fluktuasi schistosomiasis di daerah endemis provinsi sulawesi tengah tahun 2011-2018. *Buletin Penelitian Kesehatan*, 47(3):199–206. <https://doi.org/10.22435/bpk.v47i3.1276>.
- Rifa, I. H., Pratiwi, H., and Respatiwan, R. (2019). Implementasi algoritma clara untuk data gempa bumi di indonesia. In *Seminar & Conference Proceedings of UMT*. <https://doi.org/10.31000/cpu.v0i0.1694>.

- Sakinah, N., Saputra, W., Nurfitri, N., Satriani, S., and Junaidi, J. (2022). Analisis spasial penyebaran penyakit schistosomiasis menggunakan indeks moran untuk mendukung eradikasi schistosomiasis di provinsi sulawesi tengah berbasis web dashboard. *Jambura Journal of Probability and Statistics*, 3(2):120–127. <https://doi.org/10.34312/jjps.v3i2.16580>.
- Supriyadi, E. et al. (2017). Perbandingan metode partial least square (pls) dan principal component regression (pcr) untuk mengatasi multikolinearitas pada model regresi linear berganda. *Unnes Journal of Mathematics*, 6(2):117–128. <https://doi.org/10.15294/ujm.v6i2.11819>.
- Towidjojo, V. D., Nurhafizhah, A. S., and Mardin, S. (2023). Faktor-faktor yang berhubungan dengan perilaku pencegahan schistosomiasis pada masyarakat desa kaduwaa napu kabupaten poso. *Promotif: Jurnal Kesehatan Masyarakat*, 13(1):22–27. <https://doi.org/10.56338/promotif.v13i1.3719>.
- Utami, D. S. and Saputro, D. R. S. (2018). Pengelompokan data yang memuat pencilan dengan kriteria elbow dan koefisien silhouette (algoritme k-medoids).