JURNAL STATISTIKA DAN APLIKASINYA

e-ISSN 2620 – 8369
**VOLUME 8 (ISSUE 1) JUNE 2024: 87 - 98**
https://doi.org/10.21009/JSA.08108

# CLUSTERING OF DENGUE HEMORRHAGIC FEVER TYPOLOGIES IN MUNICIPALITIES OF CENTRAL JAVA

**Fareka Erdien[1], Bagus Sumargo [2*], Nisriina Nazhiifah[3], Siti Julpia Kirana[4], Dania Siregar[5], Mulyono[6]**

*[1,2,3,4,5]Department of Statistics, Faculty of Mathematics and Natural Sciences, Jakarta State University*
*Rawamangun Muka St., East Jakarta 13320, Indonesia*
*[6]Department of Computer Science, Faculty of Mathematics and Natural Sciences, Jakarta State University*
*Rawamangun Muka St., East Jakarta 13320, Indonesia*
*Corresponding author's e-mail: [*]**bagussumargo@unj.ac.id***

*ABSTRACT*

*Article History:*

*Received: 18 May 2024*
*Revised: 7 June 2024*
*Accepted: 13 June 2024*
*Published: 30 June 2024*
*Available online.*

*Keywords:*

*DHF; Clustering; Environmental Health.*

Dengue hemorrhagic fever is a tropical climate disease that has become a global health issue. Cases of dengue hemorrhagic fever (DHF) generally occur in areas with high temperatures. Central Java Province, Indonesia, is one of the regions that has high temperatures, making it vulnerable to dengue cases. The study aimed at grouping DHF-endemic areas in Central Java to assist the government in determining policies to control or prevent DHF. Cluster analysis will be carried out to determine specifically the spread of clusters based on the factors causing DHF. The cluster analysis method used in this study is the Hierarchical Agglomerative Cluster method, namely Single Linkage, Complete Linkage, Average Linkage, Ward's Method, and Centroid Method. However, the value of Average Linkage cophenetic correlation was the largest compared to other agglomerative methods. It allows for effectively identifying endemic areas and assists in targeted policymaking for dengue control and prevention.

The results showed that there were 3 clusters formed. Out of 3 clusters, 2 clusters are clean water type, and 1 cluster is sanitation type. All existing cluster types are indicators of environmental health. Therefore, environmental health is closely related to the presence of dengue fever in a community environment. In this case, it is necessary to pay attention to the community for environmental health to overcome or prevent the occurrence of DHF, where clean water and sanitation are both indicators of environmental health. Therefore, environmental health is closely related to the presence of dengue fever in a community environment.



*How to cite this article:*
F. Erdien, B. Sumargo, N. Nazhiifah, S.J. Kirana, D. Siregar, Mulyono, "CLUSTERING OF DENGUE HEMORRHAGIC FEVER TYPOLOGIES IN MUNICIPALITIES OF CENTRAL JAVA", Jurnal Statistika dan Aplikasinya, vol. 8, iss. 1, pp. 87 – 98, June 2024


Journal homepage: *https://journal.unj.ac.id/unj/index.php/statistika*
Journal e-mail: *jsa@unj.ac.id*
Research Article · Open Access

## 1. INTRODUCTION

Dengue hemorrhagic fever is a tropical climate disease that has become a global health issue [1]. Not only that, but dengue fever has also attacked Indonesia for a long time. Indonesia has the 2nd highest dengue cases out of 30 countries [2]. The number of municipalities infected with DHF in Indonesia in 2018 has increased, from 434 (84.44%) in 2017 to 440 (85.60%) in 2018. The mortality rate (CFR) due to DHF is more than 1% which is categorized as high. CFR in 2018 decreased compared to the previous year from 0.72 in 2017 to 0.71. In 2017 there were 30 provinces with morbidity rates of less than 49 per 100,000 population. Meanwhile, in 2018 the province has a morbidity rate of less than 49 per 100,000 population decreased to 26 provinces [3].

One of the provinces in Indonesia that is experiencing serious problems related to dengue is Central Java. Public health conditions in Central Java province in 2018 in the past month showed that 34.46% of people experienced health complaints. The incidence rate (IR) of DHF in Central Java Province decreased in 2018 by 10.2 per 100,000 population, compared to 2017 which was 21.68 per 100,000 population. In addition, the DHF case fatality rate (CFR) in Central Java decreased from the CFR in 2017 which was 1.24 percent to 1.05 percent in 2018 [4]. The number of cases is declining but remains a health problem that needs a solution to overcome.

Dengue hemorrhagic fever is a disease caused by the dengue virus. Dengue viruses belong to members of the family Flaviviridae and the genus Flavivirus. The four known dengue viruses are DENV-1, DENV-2, DENV-3, and DENV-4 [5]. High fever, muscle aches, joint pain, a drop in pulse pressure, and severe shock are all common symptoms of dengue hemorrhagic fever [6]. Temperature, humidity, and rain are among the climatic factors that influence the beginning of the emergence of mosquito breeding. The heavy rain causes inundation, which increases the breeding of mosquitoes. Low humidity affects the lifespan of mosquitoes, where the humidity level of 60% is the lowest limit that allows mosquitoes to multiply. Meanwhile, at a temperature of 26°C - 32°C is the average optimal temperature for mosquito growth [7].

According to the Semarang Climatology Station, in 2017, the average air temperature in Central Java was 26.8°C to 28.9°C. The relative humidity ranges from 69.0 to 86.0 percent on average. Heavy rainfall was also recorded in October at 484.00 mm [8]. The increase in dengue hemorrhagic fever virus is also due to many factors, including poor waste management and water distribution, urbanization, globalization, and an increasing amount of plastic waste [5].

In a previous study on dengue hemorrhagic fever in Central Java province, it was concluded that the factors that contributed to the increase in dengue cases in Central Java province were the number of health centers per 1,000 population, the number of Pondok Bersalin Desa (polindes) per 1,000 population, population density, the percentage of residents who have access to sustainable drinking water, and the percentage of clean water. The quality of free bacteria, fungi, and chemicals and the number of protected spring facilities [9]. Additionally, other studies have found a significant relationship between rainfall and DHF cases [10].

In previous studies, dengue hemorrhagic fever modeling was carried out in central Java province using Geographically Weighted Negative Binomial Regression (GWNBR). What can be concluded from the study is that the number of protected spring facilities is an influential factor in the increase in dengue cases in all regions of Central Java. In addition, several factors contribute to the high number of dengue cases in Central Java, such as Perilaku Hidup Bersih dan Sehat (PHBS) households and healthy houses [11].

Based on previous studies, the K-Means Clustering Method is the best method for grouping areas based on the level of DHF vulnerability in Central Java using 4 clusters [14]. In the grouping of districts/cities based on DHF indicators by Mustafidah, four clusters were formed using the Complete Linkage and Average Linkage methods [15]. In this study, cluster analysis will be carried out to determine specifically the spread of clusters in Central Java based on the factors causing DHF. The cluster analysis method used in this study is the Hierarchical Agglomerative Cluster method, namely Single Linkage, Complete Linkage, Average Linkage, Ward's Method, and Centroid Method.

## 2. METHODS

### Material and Data

The data obtained from the Central Java Provincial Health Office and BPS Jawa Tengah Province in 2018. The response variable to be used is the number of dengue cases in 35 municipalities in Central Java in 2018. The following is a variable used in this study [11].

**Table 1. Research Variables**

| Variables | Description |
|---|---|
| X1 | Percentage of Households with Access to Adequate Drinking Water |
| X2 | The Percentage of Healthy Houses |
| X3 | The Number of Village Maternity Hut |
| X4 | Number of Protected Spring Facilities |
| X5 | Percentage of Households with Access to Proper Sanitation |

### Research Method

There are several steps to model the number of DHF cases using the Cluster Average Linkage Method. Cluster analysis is used to group the observed objects into clusters based on the measurement of the observed variables, so that objects in the cluster are the same/similar and between clusters are not similar. The fundamental problem in this analysis is the unsatisfactory definition of what constitutes a cluster. The benefits of clustering include: (1) data exploration, (2) data reduction, (3) layering or separation of objects, and (4) prediction (forecasting). The clustering stages can be presented in the form of a tree diagram (trend diagram or dendogram) which allows tracking the observed clustered objects more easily and informatively.

The cluster analysis method used is the Agglomerative Hierarchy method. Agglomerative method, where the clustering starts from each object as a cluster, then continues with merging (closer clusters are combined) and so on so that finally one cluster is obtained which consists of all objects. The next problem is the use of similarity measures between groups. Several measures of similarity: (1) Single linkage, (2) Complete linkage, (3) Average linkage, (4) Centroid, and (5) Ward's error sum of squares method. The following is an explanation of some of the linkage in the hierarchical clustering method [12]:

1. Single Linkage
   The distance of two clusters is measured by the closest distance between an object in one cluster and an object in another cluster [12].
   $$dmj = min\,(dkj, dlj) \tag{2.1}$$

   where:
   $dmj$ : distance between clustersmand clusters $j$.
   $dkj$ : distance between the nearest neighbors of $k$ and $j$,
   $dlj$ : distance between the nearest neighbors $l$ and $j$.

2. Complete Linkage
   The distance of two clusters is measured by the furthest distance between an object in one cluster and an object in the other cluster [12].
   $$dmj = max\,(dkj, dlj) \tag{2.2}$$

3. Average Linkage

The distance between two clusters is measured by the average distance between an object in one cluster and an object in the other cluster [12].

$$dmj = \frac{(Nk\ dkj + Nl\ dlj)}{Nm} \tag{2.3}$$

where:
$Nk, Nl$ and $Nm$ = number of objects in the $k$, $l$, and $m$ clusters

4. Centroids
The distance between two clusters was measured as the Euclidian distance between the two clusters' centroids. The centroid method is the mean value of observations on variables in a set of cluster variables. The advantage is that outliers have little effect when compared to other methods, which are calculated by the formulation [13].

$$dmj = \frac{(Nk\ dkj + Nl\ dlj)}{Nm} - \frac{Nk\ Nl\ dkl}{Nm^2} \tag{2.4}$$

5. Ward's error sum of squares
The distance between the two clusters is the sum of the squares of the deviations from the centroid method. The goal of the ward method is to minimize the in-cluster sum of squares.

$$dmj = \frac{(Nj + Nk)\ dkj + (Nj + Nl)dlj - Nj\ dkl\ )}{Nj + Nm} \tag{2.5}$$

This method is different from other methods because it uses an analysis of variance approach to calculate the distance between clusters or this method minimizes the number of squares. It can only be counted if the cluster has more than one object element. A cluster that has only one object is zero. Ward's method is calculated based on the following equation:

$$SSE\ = \sum N_i = 1(X_i\ -\ \bar{X})'\ (X_i\ -\ \bar{X}) \tag{2.6}$$

where:
$SSE$ : Sum of Square Error
$X_i$ : The column vector whose entry is the object value with $i\ =\ 1,2,3,\dots,N$
$X$ : The column vector whose entries are the average of the object values in the Cluster
$N$ : The number of objects

In cluster analysis, the assumption that must be met is the absence of multicollinearity between variables. This assumption can be tested by looking at the correlation between variables with Pearson correlation and the VIF value of each variable. Pearson correlation can be written as follows.

$$r_{xy} = \frac{n\sum\ x_iy_i - (\sum\ x_i)(\sum\ y_i)}{\sqrt{\left(n\sum\ x_i^2 - (\sum\ x_i)^2\right)\left(n\sum\ y_i^2 - (\sum\ y_i)^2\right)}} \tag{2.7}$$

where:
$r_{xy}$ : correlation between *x* and *y*
$n$ : number of samples

The multicollinearity test uses the VIF (Variance Inflation Factors) value. The VIF for the k-th variable can be expressed as.

$$VIF(k) = \frac{1}{1 - \rho^2(k)} \tag{2.8}$$

where $\rho^2(k)$ is the correlation coefficient between $x_k$ and the linear regression estimate of $x_k$ obtained from the remaining x variables.

There are several steps to model the number of dengue cases using the Cluster Average Linkage Method. The following are the steps of the analysis carried out: (1). Look for descriptive statistics of each data; (2). The multicollinearity test uses the VIF (Variance Inflation Factors) value. If VIF > 10, then there are cases of multicollinearity in the predictor variables and must be handled; (3). Hierarchical cluster analysis to classify data; (4). Modeling the Cluster Average Linkage Method, the clustering process can be seen through a dendogram; (5). Scree Plot Clustering using Elbow Method, to see the optimal number of clusters; (6). Determine the characteristics of each cluster, identification of characteristics in this study was carried out by descriptive analysis by finding the average value of each variable in each cluster.

## 3. RESULTS

In this discussion, a cluster analysis will be carried out on 35 municipalities in Central Java based on the factors causing cases of Dengue Hemorrhagic Fever. The cluster analysis method used is the Agglomerative Hierarchy method. Before conducting the analysis, the descriptive statistics of each variable in the data will be seen first.

**Table 2. Statistics Descriptive of Variables**

| Statistics | X1 | X2 | X3 | X4 | X5 | X6 |
|---|---|---|---|---|---|---|
| Min | 61.91 | 59.69 | 44.70 | 23.00 | 0.00 | 9.24 |
| Q1 | 74.31 | 71.60 | 67.24 | 86.00 | 77.50 | 71.31 |
| Median | 78.75 | 78.30 | 75.56 | 147.00 | 1078.00 | 77.08 |
| Means | 79.36 | 79.22 | 74.10 | 156.00 | 8439.30 | 74.54 |
| Q3 | 83.97 | 88.06 | 83.61 | 224.00 | 9542.00 | 86.83 |
| Max | 99.50 | 97.25 | 100.00 | 502.00 | 83519.00 | 97.47 |

The assumption that must be fulfilled in cluster analysis is that there is no multicollinearity between variables. This assumption can be tested by looking at the correlation between variables and the VIF value of each variable. All the Pearson correlation values between variables are not more than 0.7. For example, the correlation between variables X1 and X2 is 0.3552; variables X3 and X5 have a Pearson correlation value of -0.2095; and so, on so that no correlation value is found more than 0.7. Therefore, it is indicated that there is no multicollinearity between variables. Further, it can be seen in the VIF value of each variable, all variables have a VIF value of less than 10, so it can be concluded that there is no multicollinearity in the data.

Prior to conducting cluster analysis, normalization was performed on the data. In this study, normalization was carried out using the Z-Score Normalization method which had an average of 0. After normalizing the data, a hierarchical cluster analysis was carried out to classify the data. Hierarchical cluster analysis is a method that does not require an assumption made in the number of groups or group structure. Cluster analysis itself is a method of grouping with the number of groups to be created is not yet known. The grouping is done in a multilevel way starting with each separate object, then the objects with the closest distance are made into one cluster and so on until a single cluster is formed. This process is called agglomeration or concentration.

In this study, the clustering process is carried out using the Average Linkage method, which is to calculate the average distance from all objects in one cluster to all objects in other clusters. The clustering process can be illustrated by the following dendogram.

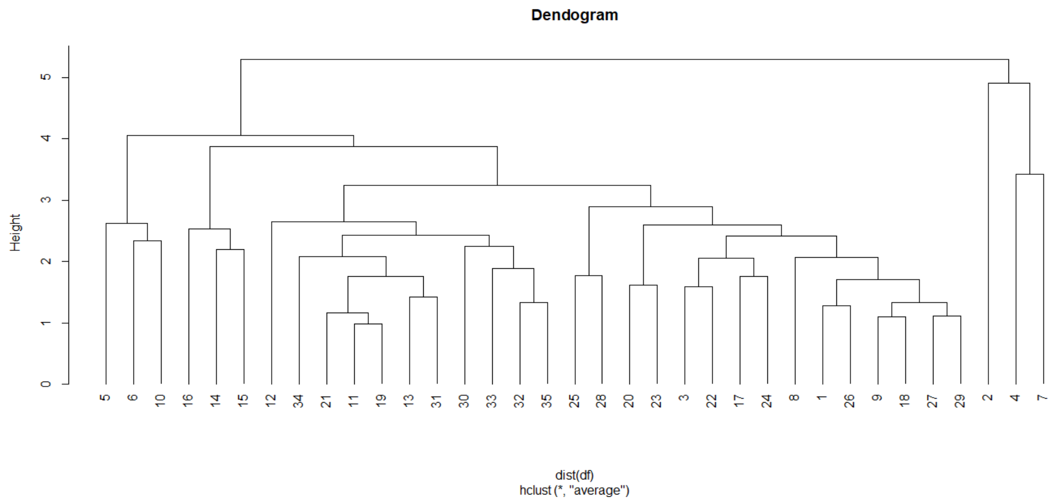


**Figure 1. Dendogram Clustering Results using the Average Linkage method**

**Table 3. Cophenetic Correlation Values**

| Single Linkage | Complete linkage | Average Linkage | Ward's Method | Centroid Method |
|---|---|---|---|---|
| 0.9350 | 0.9726 | 0.9781 | 0.6608 | 0.9535 |

The results of the analysis using the Average Linkage yielded a cophenetic correlation value of 0.978 where this value was the largest compared to other agglomerative methods as shown in Table 6. Determination of the best number of clusters was carried out using the Elbow method. At a certain point there will be a drastic decrease with an indentation called the elbow criterion where this value will be the best number of clusters. If seen in Figure 2, the elbow criterion is when the number of clusters is 5. Then the number of selected clusters is 5 clusters.

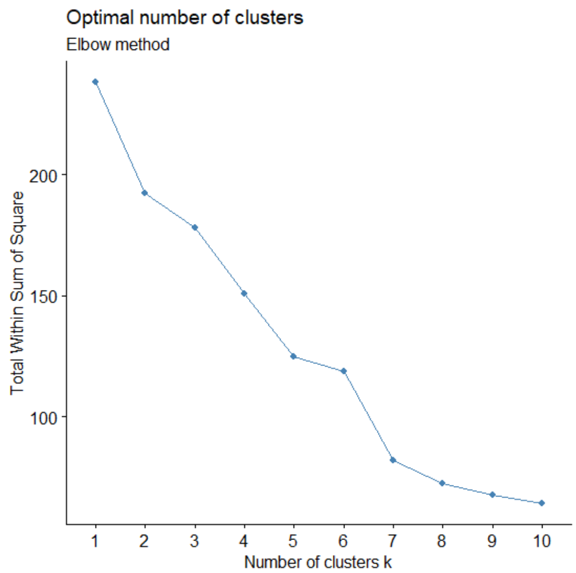


**Figure 2. Scree Plot Clustering Results using the Elbow Method**

The results of grouping Municipality in Central Java into 5 clusters are as follows. With $k = 5$, clusters are easier to interpret and understand by stakeholders, aiding in policy making. A smaller number of clusters is easier to manage in public health programs.

**Table 4. Results of Clustering Municipality in Central Java by Factors Causing Dengue Hemorrhagic Fever**

| Municipality | Clusters | Municipality | Clusters |
|---|---|---|---|
| Cilacap | 1 | Holy | 1 |
| Banyumas | 2 | Jepara | 1 |
| Purbalingga | 1 | Demak | 1 |
| Banjarnegara | 3 | Semarang | 1 |
| Kebumen | 4 | Temanggung | 1 |
| Purworejo | 4 | Kendal | 1 |
| Wonosobo | 3 | Stem | 1 |
| Magelang | 1 | Pekalongan | 1 |
| Boyolali | 1 | Poor | 1 |
| Klaten | 4 | Tegal | 1 |
| Sukoharjo | 1 | Brebes | 1 |
| Wonogiri | 1 | Magelang | 1 |
| Karanganyar | 1 | Surakarta | 1 |
| Sragen | 5 | Salatiga | 1 |
| Grobogan | 5 | Semarang | 1 |
| Blora | 5 | Pekalongan | 1 |
| Rembang | 1 | Tegal | 1 |
| Starch | 1 | | |

After forming 5 clusters, the next step is to determine the characteristics of each cluster. Identification of characteristics in this study was carried out by descriptive analysis by finding the average value of each variable in each cluster. So that the results of these descriptive statistics can provide an overview of the characteristics of each cluster.

**Table 5. Average Variable in Each Cluster**

| | Clusters | | | | |
|---|---|---|---|---|---|
| **Variables** | **1** | **2** | **3** | **4** | **5** |
| X1 | 81.21 | 75.08 | 75.23 | 67.68 | 79.16 |
| X2 | 79.38 | 70.94 | 72.48 | 82.57 | 81.79 |
| X3 | 75.27 | 75.28 | 49.35 | 82.84 | 71.36 |
| X4 | 113.62 | 245.00 | 246.50 | 386.00 | 212.67 |
| X5 | 4600.85 | 83519.00 | 30708.00 | 9816.00 | 456.33 |
| X6 | 77.86 | 66.15 | 15.05 | 83.88 | 78.89 |

The clusters formed using the Average Linkage method for municipalities in Central Java based on the factors causing Dengue Hemorrhagic Fever were 5 clusters. Most of the municipalities in

Central Java are in cluster 1, which is 26 municipalities. Cluster 2 consists of 1 municipality, cluster 3 consists of 2 regencies/cities, cluster 4 consists of 3 municipalities, and cluster 5 consists of 3 municipalities. The following is a breakdown of each cluster.

1. Cluster 1 members are Cilacap, Purbalingga, Magelang, Boyolali, Sukoharjo, Wonogiri, Karanganyar, Rembang, Pati, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Magelang, Surakarta City, Salatiga, Semarang, Pekalongan, and Tegal. Cluster 1 has a high level of paying attention to the factors that cause Dengue Hemorrhagic Fever. The characteristic of this cluster is that it has a very high percentage of households that have access to safe drinking water but the number of polindes is very low.
2. Cluster 2 members are Banyumas. Cluster 2 has a moderate level of paying attention to the factors that cause Dengue Hemorrhagic Fever. The characteristics of this cluster are that it has a very high number of protected springs and a high percentage of healthy houses.
3. Cluster 3 members are Banjarnegara and Wonosobo. Cluster 3 has a very low level of paying attention to the factors that cause DHF. The characteristics of this cluster are that it has a very low percentage of healthy homes and households that have access to proper sanitation with a high number of polindes and protected spring facilities.
4. Cluster 4 members are Kebumen, Purworejo, and Klaten. Cluster 4 has a very high level of paying attention to the factors that cause DHF. The characteristics of this cluster are that it has a very high score on the percentage of PHBS households, the percentage of healthy homes, the number of polindes and households that have access to proper sanitation.
5. Cluster 5 members are Sragen, Grobogan, and Blora. Cluster 5 has a low level of paying attention to the factors that cause DHF. The characteristics of this cluster are that it has a very low number of protected spring facilities, but high enough for access to proper drinking water, households with PHBS, and access to proper sanitation.

## 4. DISCUSSIONS

Based on previous research, the K-Means Clustering Method is the best method for grouping areas based on the level of DHF vulnerability in Central Java using 4 clusters [14]. Mustafidah concluded that 4 clusters were formed resulting from the Complete Linkage and Average Linkage methods in grouping districts/cities based on DBD indicators [15]. Based on the results of the analysis, 5 clusters were formed. The results of this analysis show that the best method used is the Average Linkage Method with the formation of 5 clusters. However, in this discussion, 3 clusters will be formed. The 3 clusters from the cluster that has the highest average on certain variables, namely clusters 1, 2, and 4. Cluster 1 is the cluster with the highest average X1 variable. Cluster 2 is the cluster with the highest average of X5 variable. Cluster 4 is the cluster with the highest average X2, X3, X4, and X6 variables. The remaining 2 clusters, namely clusters 3 and 5, will be combined with the previous 3 clusters based on the same average of the highest variables. Cluster 3 has the second highest average X5 variable so that this cluster is combined with cluster 2. While cluster 5 has the second highest average X1 variable so that this cluster is combined with cluster 1. Three new clusters are formed.

**Table 6. Averages for The Three New Clusters**

| | **Clusters** | | |
|---|---|---|---|
| | **A** | **B** | **C** |
| X1 | 80.19 | 75.15 | 67.68 |
| X2 | 80.58 | 71.71 | 82.57 |
| X3 | 73.32 | 62.32 | 82.84 |
| X4 | 163.14 | 245.75 | 386.00 |
| X5 | 2528.59 | 57113.50 | 9816.00 |
| X6 | 78.37 | 40.60 | 83.88 |

The following is a breakdown of each cluster.

1. Cluster A is a cluster that has the best percentage of households that have access to safe drinking water with an average of 80.19%. Cluster A members are Cilacap, Purbalingga, Magelang, Boyolali, Sukoharjo, Wonogiri, Karanganyar, Rembang, Pati, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Magelang, Surakarta, Salatiga, Semarang, Pekalongan, Tegal, Sragen, Grobogan, and Blora.
2. Cluster B is the cluster that has the largest number of protected water facilities, with an average of 57113.50 ~ 57114. Cluster B members are Banyumas, Banjarnegara and Wonosobo.
3. Cluster C is the one that has the best 4 factors, namely the percentage of clean and healthy households with an average of 82.57%, the percentage of healthy homes with an average of 82.84%, the number of Polindes (Village Maternity Boarding Schools) with an average of 386, and the percentage of households that have access to proper sanitation with an average of 83.88%. Cluster C members are Kebumen, Purworejo, and Klaten.

**Table 7. Number of DHF Cases, Number of Population in 2018, Percentage of DHF Cases, and Cluster in Each Central Java Region**

| Number of DHF Cases | Number of Population in 2018 | Percentage of DHF Cases | Clusters | Central Java Region |
|---|---|---|---|---|
| 136 | 1719504 | 7,909E-05 | A | Cilacap |
| 57 | 1679124 | 3.395E-05 | B | Banyumas |
| 190 | 925193 | 0.0002054 | A | Purbalingga |
| 80 | 918219 | 8.713E-05 | B | Banjarnegara |
| 41 | 1195092 | 3.431E-05 | C | Kebumen |
| 21 | 716477 | 2,931E-05 | C | Purworejo |
| 41 | 787384 | 5.207E-05 | B | Wonosobo |
| 82 | 1279625 | 6.408E-05 | A | Magelang |
| 133 | 979799 | 0.0001357 | A | Boyolali |
| 20 | 1171411 | 1.707E-05 | C | Klaten |
| 35 | 885205 | 3,954E-05 | A | Sukoharjo |
| 24 | 957106 | 2.508E-05 | A | Wonogiri |
| 80 | 879078 | 9,1E-05 | A | Karanganyar |
| 316 | 887889 | 0.0003559 | A | Sragen |
| 460 | 1371610 | 0.0003354 | A | Grobogan |
| 388 | 862110 | 0.0004501 | A | Blora |
| 75 | 633584 | 0.0001184 | A | Rembang |
| 82 | 1253299 | 6.543E-05 | A | Starch |
| 58 | 861430 | 6.733E-05 | A | Holy |

| | | | | |
|---|---|---|---|---|
| 216 | 1240600 | 0.0001741 | A | Jepara |
| 39 | 1151796 | 3.386E-05 | A | Demak |
| 173 | 1040629 | 0.0001662 | A | Semarang |
| 104 | 765594 | 0.0001358 | A | Temanggung |
| 32 | 964106 | 3.319E-05 | A | Kendal |
| 68 | 762377 | 8.919E-05 | A | Stem |
| 182 | 891892 | 0.0002041 | A | Pekalongan |
| 50 | 1299724 | 3,847E-05 | A | Malang |
| 70 | 1437225 | 4.87E-05 | A | Tegal |
| 30 | 1802829 | 1.664E-05 | A | Brebes |
| 50 | 121872 | 0.0004103 | A | Magelang |
| 22 | 517887 | 4.248E-05 | A | Surakarta |
| 32 | 191571 | 0.000167 | A | Salatiga |
| 97 | 1786114 | 5.431E-05 | A | Semarang |
| 30 | 304477 | 9.853E-05 | A | Pekalongan |
| 5 | 249003 | 2.008E-05 | A | Tegal |

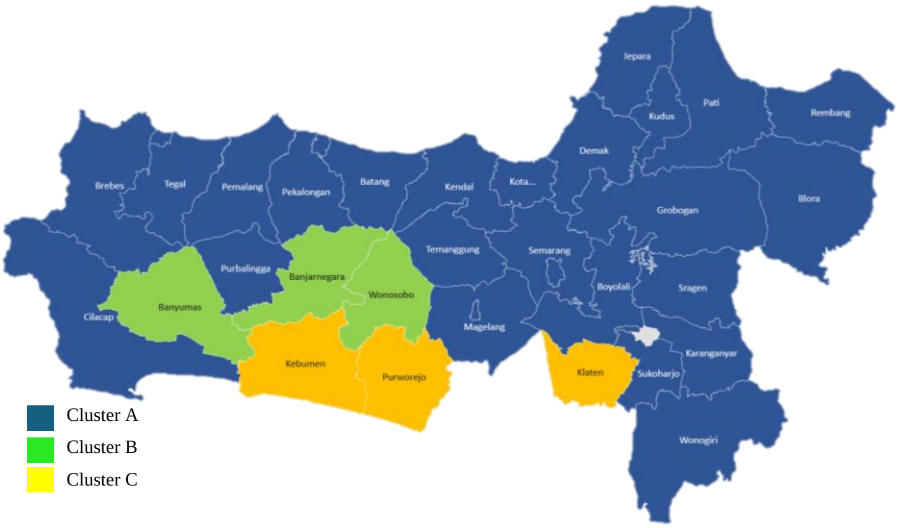


**Figure 3. Distribution of the Clusters in Central Java Province**

Three clusters have been formed. The distribution of clusters in Central Java province can be seen in Figure 3. Based on the figure, most municipalities in Central Java are in cluster A. Where cluster A is a cluster that has the best percentage of households that have access to safe drinking water. Thus, this means a low number of dengue cases in 2018 in Central Java province because the majority of municipalities in Central Java have the best percentage of households that have access to safe drinking water. After obtaining three new clusters, the average number of dengue cases in each cluster was searched based on the data on the number of dengue cases in each area of Central Java in Table 10. In addition,

**Table 8. Averages Number of DHF Cases, Average Population in 2018, and Percentage of DHF Cases in Each Cluster**

| Clusters | Average Number of DHF Cases | Average Population in 2018 | Percentages |
|---|---|---|---|
| A | 112.3793103 | 966314.7586 | 0.0001163 |
| B | 5933333333 | 1128242.333 | 5.259E-05 |
| C | 27.33333333 | 1027660 | 2.66E-05 |

Based on the results of the analysis with, it can be seen in Table 8 that the average percentage of DHF cases in cluster A with 29 municipalities has a percentage of 0.0001163%, Cluster B with a total of 3 municipalities has an average percentage of 5.259 E-05% of DHF cases, and Cluster C with a total of 3 municipalities with an average percentage of 2.66E-05% of DHF cases. Based on the results of the analysis, the average proportion of dengue cases in each cluster is very low.

## 5. CONCLUSION

There are 3 different clusters that were formed. Cluster A with 29 municipalities has typology percentage of households that have access to safe drinking water. Cluster B with a total of 3 municipalities has the largest number of protected water facilities. Cluster C with a total of 3 municipalities has the best 4 factors, namely the percentage of clean and healthy living behavior of households. At the same time, cluster A has the highest average percentage of DHF cases, while cluster B has the lowest average percentage of DHF cases. The average percentage of DHF cases in each cluster is very low, meaning that high temperatures do not cause too many DHF cases. Out of 3 clusters, 2 clusters are clean water type, and 1 cluster is sanitation type. All existing cluster types are indicators of environmental health. therefore, environmental health is closely related to the presence of dengue fever in a community environment. So, in this case, it is necessary to pay attention to the community for environmental health to overcome or prevent the occurrence of DHF.

## 6. ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to all individuals and organizations who contributed to this research. We also extend our appreciation to the health officials and environmental agencies who provided valuable insights and assistance throughout the study. Their local knowledge and expertise were instrumental in the successful completion of this research. The authors acknowledge the financial support from various institutions that made this research possible. Additionally, we are grateful for the technical support and resources provided by universities. Their contributions have been invaluable in shaping the final outcomes of this study.

## 7. REFERENCES

[1] W.H. Wang, A. N. Urbina, M. R. Chang, W. Assavalapsakul, P.-L. Lu, Y.-H. Chen, S.-F. Wang, "Dengue hemorrhagic fever – A systemic literature review of current perspectives on pathogenesis, prevention and control," *Journal of Microbiology, Immunology and Infection*, vol. 53, no. 6, pp. 963-978, 2020.

[2] World Health Organization (WHO), *Global Strategy for Dengue Prevention and Control 2012-2020*, World Health Organization Press, France, 2012.

[3] Ministry of Health RI, *Indonesian Health Profile*, Indonesia, 2019.

[4] Central Java Provincial Health Office in Indonesia, *Central Java Health Profile 2018*, Central Java, 2019.

[5] M. S. Mustafa, V. Rasotgi, S. Jain, V. Gupta, "Discovery of fifth serotype of dengue virus (DENV-5): A new public health dilemma in dengue control," *Medical Journal Armed Forces India*, vol. xx, no. xx, pp. 87-70, 2015.

[6] A. Candra, "Dengue Hemorrhagic Fever Epidemiology, Pathogenesis, and Its Transmission Risk Factors," *Aspirator Journal of Vector Borne Diseases Studies*, vol. 2, no. 2, pp. 110-119, 2013.

[7] J. Ariati and D. A. Musadad, "Incidence of Dengue Haemorrhagic Fever (DHF) and Climate factors in Batam City of Riau Islands Province," *Journal of Health Ecology*, vol. 11, no. 4, pp. 279-286, 2012.

[8] The Central Statistics Agency (BPS) Central Java Province, *Central Java Province in Number*, Central Java, 2018.

[9] I. F. Fatati, H. Wijayanto, A. M. Sholeh, "Spatial Regression Analysis and Distribution Patterns in Dengue Hemorrhagic Fever (DBD) Cases in Central Java Province," *Media Statistika*, vol. 10, no. 2, pp. 95-105, 2017.

[10] C. L. Rompis and O. J. Sumampouw, "Does Rainfall Affect the Incidence of Dengue Hemorrhagic Fever?" *Indonesian Journal of Public Health and Community Medicine*, vol. 1, no. 1, 2020.

[11] B. Sumargo, S. Julpia, S. Rohmah, "Dengue Hemorrhagic Fever Modeling using Geographically Weighted Negative Binomial Regression," in Proc. of International Conference on Statistics, vol. 2588, 2021.

[12] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed., London: Prentice Hall, 2002.

[13] S. Singgih, *SPSS Multivariate Statistics Exercise Book*, Jakarta: PT Elex Komputindo, 2002.

[14] N. Imani, T. Shafira, L. Kurnia, E. Rusnita, and E. Widodo, "Analisis Perbandingan Metode Single Linkage dan K-Means Clustering (Studi Kasus: Kasus Demam Berdarah Dengue (DBD) di Provinsi Jawa Tengah Tahun 2015)," *Semnasteknomedia Online*, vol. 6, no. 1, pp. 2-4, 2018.

[15] M. E. Mustafidah and M. D. Purnama, "Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kasus DBD Menggunakan Complete Linkage dan Average Linkage," *MATHunesa: Jurnal Ilmiah Matematika*, vol. 12, no. 2, pp. 337-343, 2024.

[16] S. Brown, R. Tauler, and B. Walczak, Eds., *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*. Elsevier, 2020.

[17] S. Hulu and J. Sinaga, *Analisis Korelasi: Pearson, Spearman, dan Kendall*. Yogyakarta: Deepublish, 2019.