



## A Clustering Approach for Mapping Dengue Contingency Plan

Farida Amila Husna<sup>1</sup>, Diana Purwitasari<sup>2\*</sup>, Bayu Adjie Sidharta<sup>3</sup>, Drigo Alexander Sihombing<sup>4</sup>,  
Amiq Fahmi<sup>5</sup>, Mauridhi Hery Purnomo<sup>6</sup>

<sup>1,2,3,4</sup>Department of Informatics Engineering, Faculty of Intelligent Electrical and Informatics Technology,  
Institut Teknologi Sepuluh Nopember, Indonesia

<sup>5,6</sup>Department of Electrical Engineering, Faculty of Intelligent Electrical and Informatics Technology,  
Institut Teknologi Sepuluh Nopember, Indonesia

<sup>5</sup>Department of Information System, Faculty of Computer Science,  
Universitas Dian Nuswantoro, Indonesia

<sup>6</sup>Department of Computer Engineering, Faculty of Intelligent Electrical and Informatics Technology,  
Institut Teknologi Sepuluh Nopember, Indonesia

### Abstract.

**Purpose:** The dengue epidemic has an increasing number of sufferers and spreading areas along with increased mobility and population density. Therefore, it is necessary to control and prevent Dengue Hemorrhagic Fever (DHF) by mapping a DHF contingency plan. However, mapping a dengue contingency plan is not easy because clinical and managerial issues, vector control, preventive measures, and surveillance must be considered. This work introduces a cluster-based dengue contingency planning method by grouping patient cases according to their environment and demographics, then mapping out a plan and selecting the appropriate plan for each area.

**Methods:** We used clustering with silhouette scoring to select features, the best cluster formation, the best clustering method, and cluster severity. Cluster severity is carried out by levelling the attributes of the average value to low, medium, high, and extreme, which are related to the plans each region sets for village type and season type.

**Result:** In five years of data (2016-2020) ±15K cases from Semarang City, Indonesia, feature selection results show that environmental and demography group features have the biggest silhouette score. With these features, it is found that K-Means has a high silhouette score compared to DBSCAN and agglomerative with three optimum numbers of clusters. K-Means also successfully mapped the cluster severity and assigned the cluster to a suitable contingency policy.

**Novelty:** Most of the research on DHF cases is about predicting DHF cases and measuring the risk of DHF occurrence. There are not many studies that discuss the policy recommendations for dengue control.

**Keywords:** Dengue, Clustering, K-Means, Contingency Plan

Received June 2022 / Revised July 2022 / Accepted October 2022

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



### INTRODUCTION

Automatic analysis of microscopic images for disease detection and diagnosis is becoming very important. Dengue Hemorrhagic Fever (DHF) is still one of the leading health problems that threaten people in Indonesia. As a tropical area, Indonesia is quite vulnerable to this disease, so the number of sufferers and the area of the spread of DHF are increasing along with increased mobility and population density. Indonesia is where DHF always occurs (endemic areas), whose cases tend to increase yearly and potentially cause outbreaks in each region. Efforts to eradicate dengue fever based on the Policy of the Ministry of Health Number 581/MENKES/SK/VII/1992 [1], include prevention by conducting Mosquito Nest Eradication (PSN), discovery, assistance, and reporting, epidemiological investigations, and disease observation, prevention as necessary, other countermeasures, and health education. However, until 2019 the morbidity (morbidity) of DHF was still high compared to the previous increase of 110.56% to 138,127 cases. The number of deaths (mortality) also increased by 96.8% to 919 in the same year [2]. From this, it

---

\*Corresponding author.

Email addresses: [faridaamila@gmail.com](mailto:faridaamila@gmail.com) (Husna), [diana@if.its.ac.id](mailto:diana@if.its.ac.id) (Purwitasari),  
[bayuaaas@gmail.com](mailto:bayuaaas@gmail.com) (Sidharta), [drigosihomebinga@gmail.com](mailto:drigosihomebinga@gmail.com) (Sihombing),  
[amiq.fahmi@dsn.dinus.ac.id](mailto:amiq.fahmi@dsn.dinus.ac.id) (Fahmi), [hery@ee.its.ac.id](mailto:hery@ee.its.ac.id) (Purnomo)

DOI: [10.15294/sji.v9i2.36885](https://doi.org/10.15294/sji.v9i2.36885)

is known that the existing policies have not been able to reduce the morbidity and mortality due to dengue fever.

The dengue epidemic draws attention to effective management programs to reduce morbidity and mortality, making dengue no longer a public health concern. There are no vaccines to prevent dengue infection or specific treatment medicines. According to Runge-Ranzinger's previous study, most countries did not have contingency control and comprehensive and detailed focused control for dengue outbreaks.

Governments tend to rely on intensified vector control in response to attacks, with minimal intervention in the integrated management aspects of clinical care, epidemiology, laboratories, and surveillance [3]. Since dengue fever mosquitoes have spread throughout the country, they cannot rely only on intensified vector control as their outbreak response, with minimal focus on integrated management of clinical care, epidemiological, laboratory, and vector surveillance risk communication. This is because the spread of viruses occurs from complex interactions between viruses, hosts, and vectors. These interactions are influenced by environmental and socio-economic conditions, such as temperature, rainfall, humidity, population density, sanitation, housing, education, and access to water [4], [5]. A contingency plan is needed because it links these elements together and describes additional timing and response actions to be taken when an outbreak is imminent or has started to prevent outbreaks of Dengue and control similar epidemics.

Contingency control policies can be more easily understood by visualizing them with a map. The contingency map represents the demographic and environmental relationships inherent in dengue cases involving a contingency plan. Contingency maps can graphically depict a more transparent plan for problem areas. Contingency plan maps help disseminate information in the policy-making process, predict the impact of disasters, and carry out disaster management activities [6]. Besides, mapping a dengue contingency plan is not easy since it involves clinical and managerial issues, vector control, alerting, and surveillance must be considered.

The previous study used the K-means method to map the level of DHF vulnerability in the Kediri district. Spatio-temporal visualization shows the distribution of DHF-vulnerable areas from 2016 to 2019 based on multiple factors affected by DHF: deceased patients, population, precipitation, and public facilities. Using this data can affect not only how the case data is processed to obtain the vulnerability level, but also rainfall and public facilities data. Using the K-Means clustering algorithm, the 344 villages are classified into low, medium, and high vulnerability levels. Web-based spatiotemporal visualization helps to compare levels of DHF sensitivity year by year. Based on an evaluation of an average silhouette coefficient of 0.57, it shows good clustering results [7]. The results of good silhouette scores have led to using the K-Means algorithm and silhouette score as a clustering technique and evaluation metric when creating maps. However, this study did not compare it to other methods for obtaining better clusters

Other previous studies used hierarchical clustering techniques to classify dengue cases and mortality data reported in different states of India. Clustering uses Agglomerative Clustering technique type Ward. The results are plotted on a map of India using a shapefile in RStudio. This clustering technique is inspirational when trying to create a map. The 2018 data is predicted by a logarithmic transformation using a linear regression model. The K-nearest neighbour algorithm is used to predict 2018 cluster data. The results show that the incidence or intensity of dengue fever is significantly reduced in many states [8]. However, this study is not linked to existing policy and will only serve as an alert medium for predicting severity.

Another previous research about clustering in dengue is the application of DBSCAN in disease surveillance. The algorithm depends heavily on the density-based concept of clustering. It is designed to identify clusters of different shapes and patterns. This study did an experimental evaluation to know the effectiveness and efficiency of DBSCAN using case information of reported Dengue Fever (DF) incidences for the study period 2011-2013, which was maintained in the health department of the Municipal Corporation of Delhi. The present study is successful in determining the hotspot of DF. Silhouette Coefficient has been calculated to determine the accuracy of cluster detection. This inspired the use of DBSCAN in dengue cases. These findings have important public health implications for controlling and preventing DF incidences [9]. However, this study only focuses on hotspot detection and does not include detecting the causative factor in DHF.

The problem in this work originates from dengue cases in Indonesia which have not significantly decreased even though many programs have been implemented. This contingency plan is expected to help decision-makers determine plans for each dengue-endemic area. The programs that have been implemented prioritize preventive measures by spraying mosquitoes. It takes a system that can map the dengue contingency plan by considering all aspects thoroughly. From the literature study above, it is known that several methods can be used to create maps using clustering. One of them is K-Means, Agglomerative, and DBSCAN. So, in this study, a comparison of methods using demographic, environmental, and physical data.

On the other hand, World Health Organization introduced Dengue Contingency Planning to create a framework for developing a national emergency response plan, with regional coordination that recognizes the program's components at the micro-level. The plans include a surveillance system, alert algorithm, managerial capacity, vector control, and clinical services [10].

Surveillance System (P1). As epidemiological information, passive routine reporting of dengue cases monitors dengue's spatial and temporal distribution in various symptoms, determines the intervention risk and priority areas, and acts as a trigger for outbreak alerts. The surveillance system plan's necessary actions are improving laboratory support with standardized and quality-controlled test procedures to ensure the use of simplified and standardized case classification.

Alert Algorithm (P2). Timely analysis of local surveillance data provides the basis for outbreak alerts as epidemiological data thresholds trigger them. If the data is inadequate to establish a reference or baseline level, monitoring the data for early detection of outbreaks is difficult. The vital actions in the alert algorithm plan are defining signal thresholds based on syndrome surveillance and considering a step-by-step warning scheme.

Managerial Capacity (P3). This plan improves partner engagement, capacity building, and infrastructure development by providing functional connectivity to ensure a structured and coordinated response regarding financial management, risk communication, outbreak investigation, and outbreak definition. Its success will occur in the activities of daily living, the pre-epidemic initial plan, and the expansion of routine interventions.

Vector Control (P4). Prevention of dengue fever and containment of outbreaks still depends on vector control. The unique characteristics of this method depend on social or geographical circumstances. The vital actions in the vector control plan are strengthening community involvement, focusing on timely interventions that take multiple approaches, and regularly monitoring pesticide resistance.

Clinical Services (P5). The annual need for increased clinical-service personnel corresponds to the week of increasing vector density. Clinical service plans must define requirements for additional personnel, equipment, reagents, treatment units and whether the plans should include clinical re-education courses. A clinical services plan takes these actions by promoting hospital emergency planning, ensuring regular and timely training, checking mortality, and timely notification of clinicians.

## **METHODS**

In this section, we present the main contribution of this paper. The proposed method is described in Figure 1. The first stage of this study is collecting dengue cases, demographic, and climate data. The second pre-processing stage includes removing attributes with many null values, reformatting broken values, encoding categorical attributes numerically, and normalizing. The third stage is categorizing features based on categories that influence the spread of DHF, namely demography, physical, and environmental [11], [12]. The fourth stage selects the best clustering method among Agglomerative, K-Means, and DBSCAN. The fifth stage is to perform clustering with K-Means. The sixth stage is levelling severity by clustering the mean value attributes of each cluster into four levels (low, medium, high, and extreme). The seventh stage is assigning a plan to a cluster by interpreting cluster characteristics. The final stage is mapping each cluster to the region with a dengue contingency plan. The detail of the proposed method is explained step by step in the following sub-sections as follows.

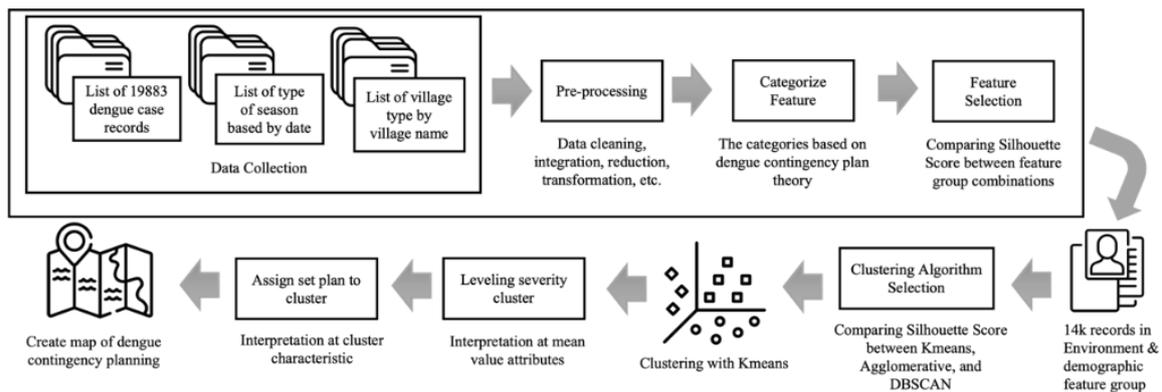


Figure 1. The framework of the proposed model

### Dataset Description

Although our experiments were explored using local data, the proposed method applies to other dengue cases. The used dataset includes DHF, DF, and DSS cases that occurred in Semarang, with 177 villages in 16 sub-districts. Dengue case data from the Disease Prevention and Eradication Council “Pencegahan dan Pengendalian Penyakit (P2P)” were from 2016–2020. The data include patient information, such as patient demographics and timelines, test results, medical history, and surveillance and PSN reports. This data is an archive belonging to the Disease Prevention and Eradication Council, which is indeed stored for the needs of monitoring dengue cases in Semarang City every year. The Disease Prevention and Eradication Council conducts surveillance at home when dengue cases occur and in all hospitals in Semarang City and stores the information in the dataset. The Meteorology, Climatology, and Geophysics Council has provided climate data, specifically from the Semarang climatology station. Rainfall records in climate data helped determine the month’s season and related to the increasing dengue cases [13]. Our demographic data came from the Central Bureau of Statistics with the village type to determine where dengue-infected patients live since urban/rural area has differences, such as population characteristics, livelihoods, and infrastructure, which can directly affect the dengue cases. Our data has 49 attributes with 19,881 records.

Table 1. Attributes information of dataset after pre-processing with their relation to dengue contingency plan (P1: surveillance system, P2: alert algorithm, P3: managerial capacity, P4: vector control, P5: clinical services)

Attributes	Type	Category	P1	P2	P3	P4	P5	References
A1	Age	Numerical	Physical	v	v		v	[2]
A2	Village name	Categorical	Environmental	v	v	v	v	[2]
A3	District name	Categorical	Environmental	v	v	v	v	[2]
A4	Hospital name	Categorical	Environmental				v	[14]
A5	Distance address to the hospital	Numerical	Demography			v		[2]
A6	Number of houses surveyed	Numerical	Environmental		v		v	[14]
A7	Number of houses with larvae	Numerical	Environmental		v		v	[15]
A8	Additional cases of fever	Numerical	Environmental				v	[15]
A9	Additional cases of DF	Numerical	Environmental				v	[15]
A10	Additional case of DHF	Numerical	Environmental				v	[15]
A11	Additional case of DSS	Numerical	Environmental				v	[15]
A12	Type of village	Categorical	Demography			v	v	[16]
A13	Type of season	Categorical	Demography	v	v		v	[16]
A14	Sick date	Categorical	Physical	v	v	v	v	[2]
A15	Work status	Categorical	Physical			v		[2]
A16	Gender	Categorical	Physical					[2]
A17	Diagnosis	Categorical	Physical	v	v		v	[2]

### **Data Preprocessing**

Due to the staff manual record process, actual data from the Disease Prevention and Eradication Council “Pencegahan dan Pengendalian Penyakit (P2P)” Semarang City, Indonesia, often has flaws with many instances of missing values, wrong format, noise, null values, and irrelevant attributes. Clearing them requires pre-processing that can dramatically increase the quality of attribute extraction and the results of data analysis [17]. Our experiment was performed in a Graphics Processing Unit (GPU) based cloud instance with the specification four-core CPU, 16GB Random Access Memory (RAM), and Vega 10. The cleaning steps are deleting rows or columns with null values conditioned to if the number of rows in a column is more than 90% "null" and dropping 11 attributes from 49 attributes. Since data inconsistencies create variations, we reformatted data for the age attribute under zero by calculating manually using the birth date attribute. We modified the village-name attribute by removing invalid values and retaining only processed data with a precise location.

We encoded categorical data using manual encoding with values starting at 0 and dropped 21 uncategorized attributes from 38 attributes since they contain irrelevant yet redundant information. Table 1 summarizes the remaining attributes after preprocessing data and categorizes them into several feature groups Physical, Environmental, and Demography. The last process is feature selection by searching for the best silhouette score from clustering results in every feature group after normalizing with the min-max method to produce balance values ranging from 1 to 0.

### **Assign Plans to Cluster**

After achieving the best combination of attribute groups, the best algorithm search process was carried out from K-Means, Agglomerative, and DBSCAN. The clustering algorithm is used to group cases in the same year. The clustering results carried out an assessment of the severity level based on the mean value attributes by clustering across years. Implementing the plans is unnecessary unless it becomes extreme if the severity level is low. Our steps are listed as pseudocode in Pseudocode 1.

We produce a map from the results of clustering and levelling. The evaluation was supported by visualizing the Semarang map equipped with village boundaries using the Figma interface design tool for our experiments area of each village is assigned a colour corresponding to the plan per village cluster. Villages with the same plan will be visualized with the same colour on the map, making reading the DHF contingency plan map easier.

## **RESULT AND DISCUSSION**

### **Experiment on Feature Group Attribute**

Our preprocessing methods form the final data into 17 attributes and 14,462 records (9 in the environmental group, 5 in the physical group, and 3 in the demographic group, as seen in Table 1). The optimal attribute combination was obtained with at least two categories in one subset from clustering iterations of 2 to 10 for the number of clusters. As for the other parameters, we used the standard parameters in the Python library sklearn.

The attribute comparison results appear in Table 2. Overall, the highest silhouette score was most on the feature group environmental and demography, closest to 1. The lowest silhouette score could be seen in the feature group demographic and physical in all four combinations. In terms of environmental and demographic combination, the 2018 silhouette score reaches the highest at 0.915. The score for 2017 was also high (0.910), which is the same as in 2016. 2019 and 2020 results were 0.885 and 0.870, respectively.

On the other hand, the lowest silhouette score was on feature group demographic and physical, accounting for under 0.41% in all years. Only 2017 on this combination reached 0.413 of the silhouette score feature group combination, which was around double that of 2016 (0.29%). Previous work with dengue fever data shows that demographic location and environment are vital in determining the spread of DHF [18], [19]. The results are like Table 2 for the environment and demographic attribute groups. In addition, the clusters formed using environmental and demographic attribute groups are not overfitting.

---

**INPUT:**

- a. List of type of village by village name
- b. List of type of season by recording date
- c. List of dengue surveillance cases containing patient demographics, timeline, test result, and medical history (19K).

**OUTPUT:**

- a. List of cluster cases with recommended dengue contingency set plan
- b. Map of city labelled with recommended dengue contingency set plan

**PROCEDURE:**

**Step #01:** Pre-process (cleaning, integration, reduction, transformation) cases (14K)

**Step #02:** FOR each year in data DO #iteration for finding the best feature group combination

**Step #03:** Select the feature group combination

**Step #04:** FOR number of cluster ( $n\_cluster$ ) = 2 ... 10 DO

**Step #05:** KMeans Clustering ( $n\_cluster$ )

**Step #06:** Calculate Silhouette Score (clustering result)

**Step #07:** WHILE all feature group combinations are executed all

**Step #08:** Find the highest Silhouette Score from steps 3 to 7 #to choose the best feature group combination

**Step #09:** Select the best feature group combination

**Step #10:** FOR each year in data DO #iteration to find the best clustering method

**Step #11:** FOR number of cluster ( $n\_cluster$ ) = 2 ... 10 DO

**Step #12:** KMeans Clustering ( $n\_cluster$ ), Agglomerative Clustering (), DBSCAN Clustering ()

**Step #15:** Find the highest Silhouette Score from steps 9 to 13 #to choose the best clustering method

**Step #16:** Create a graph with mean value attributes for each cluster and year in Figure 2.

**Step #17:** FOR each year in data DO #iteration to find the severity level

**Step #18:** FOR each cluster in data DO

**Step #19:** FOR each feature in the data DO

**Step #20:** Normalize()

**Step #21:** Mean ()

**Step #22:** IF ( $A9 \leq 0.265$  and  $A6 \geq 0.506$ ): low severity level

**Step #23:** ELSE IF ( $A9 \leq 0.265$  and  $A6 \leq 0.506$ ): moderate severity level

**Step #24:** ELSE IF ( $A9 \geq 0.265$  and  $A7 \leq 0.257$  and  $A6 > 0$ ): moderate severity level

**Step #25:** ELSE IF ( $A9 \geq 0.265$  and  $A7 \geq 0.257$ ): high severity level

**Step #26:** ELSE IF ( $A9 \geq 0.265$  and  $A7 \geq 0.387$ ): extreme severity level

**Step #27:** ELSE IF ( $A9 \geq 0.265$  and  $A6 \leq 0$ ): extreme severity level

**Step #28:** Make a cluster characteristic analysis every year in Table 4.

**Step #29:** FOR each year in data DO #iteration for find match set plan

**Step #30:** FOR each cluster in data DO

**Step #31:** IF (severity = extreme & village type = not rural): set plan 5

**Step #32:** ELSE IF (severity = extreme & village type = rural): set plan 1

**Step #33:** ELSE IF (severity = high): set plan 1

**Step #34:** ELSE IF (severity = moderate): set plan 4

**Step #35:** ELSE IF (severity = moderate & village type = rural): set plan 2

**Step #36:** ELSE IF (severity = moderate & village type = not rural): set plan 3

**Step #37:** ELSE IF (severity = moderate & village type = not rural & season = rainy): set plan 1

**Step #38:** ELSE IF (severity = low): set plan 4

**Step #39:** Build a map from the cluster with a set plan label in Figure 3.

---

Pseudocode 1. Pseudocode for clustering task in the proposed framework

**Experiment with the Optimal Clustering Method**

The following experiment identifies the optimal clustering method of K-Means, Agglomerative, and DBSCAN, using data from 2016 (the year with the most cases) and 2018 (the year with the least). Each algorithm's capabilities are assessed and compared for both years. Aside from cluster number and standard sklearn parameters, the Agglomerative algorithm uses Euclidean for its computational metrics and Ward for linkage while testing on DBSCAN also performed iterations on 0.01–0.02 epsilon. The comparison appears in Table 3.

**Table 2. Comparison of silhouette score values in a combination of feature groups**

<b>Feature Group Combination</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>
All Features	0.317 (4 Cluster)	0.312 (4 Cluster)	0.257 (4 Cluster)	0.318 (4 Cluster)	0.320 (4 Cluster)
Environmental Demography	0.910 (3 Cluster)	0.910 (4 Cluster)	0.915 (5 Cluster)	0.885 (5 Cluster)	0.870 (3 Cluster)
Environmental Physical	0.705 (8 Cluster)	0.700 (8 Cluster)	0.711 (8 Cluster)	0.645 (8 Cluster)	0.627 (8 Cluster)
Demographic Physical	0.292 (4 Cluster)	0.413 (4 Cluster)	0.299 (4 Cluster)	0.328 (4 Cluster)	0.403 (4 Cluster)

**Table 3. Comparison of silhouette score between clustering methods (k-means, agglomerative and DBSCAN)**

<b>Iteration</b>	<b>2016</b>			<b>2018</b>		
	<b>K-Means</b>	<b>Agglomerative</b>	<b>DBSCAN</b>	<b>K-Means</b>	<b>Agglomerative</b>	<b>DBSCAN</b>
2	0.694	0.684	0.259	0.792	0.684	0.449
3	0.667	0.662	0.299	0.798	0.662	0.200
4	0.657	0.628	0.299	0.783	0.628	0.200
5	0.684	0.664	0.303	0.549	0.664	2.000
6	0.690	0.677	0.139	0.514	0.677	0.202
7	0.699	0.682	0.038	0.547	0.682	0.210
8	0.706	0.683	0.038	0.419	0.683	2.000

Overall, K-Means produces better results than Agglomerative and DBSCAN, performed on 2016 and 2018 data. The lowest score could be seen in the DBSCAN method. In 2016 data, K-Means achieved the most, 0.706 at 8 clusters. Agglomerative was also high (0.684 at 2 clusters) than DBSCAN resulting in at least 0.303 at 5 clusters. In 2018 data had the same result, K-Means also achieved the highest. But found out the 3 clusters reached better results than the 8 clusters performed using K-Means (0.789). The difference in the optimal number of clusters is due to the greater diversity of data in 2016. For equality testing, the following process is carried out using 3 clusters.

On the other hand, previous studies also found that the Hierarchical Clustering (Agglomerative) method did not work well on extensive data. This condition occurs because the complexity of K-Means is linear and Hierarchical Clustering is quadratic. Also, other studies show that K-Means is superior to DBSCAN, with faster computation time [20].

### **Levelling Severity Cluster Result**

The formation of case area clusters has been carried out using K-Means. To see the characteristics of each cluster, decision-makers can see the average normalization value of the relevant attributes to be used as the basis for decision-making. The severity of a cluster can be seen in the attributes A6, A7, and A9. Attribute A9 shows DD cases had a higher value than DSS and DHF cases. At the same time, attribute A7 is used for mosquito population vigilance which will cause potential new cases of DHF, DD, and DSS in the cluster. To get the severity level, K-Means in the three attributes are performed. K-Means are completed in 4 number clusters to separate cases into four levels (low, moderate, high, and extreme). Details of the result and the severity level of each cluster are seen in Table 4.

Figure 2 illustrates the proportion of attributes and recommended plan set in the normalized mean value for every cluster. Information is divided into five plans set on the radar chart.

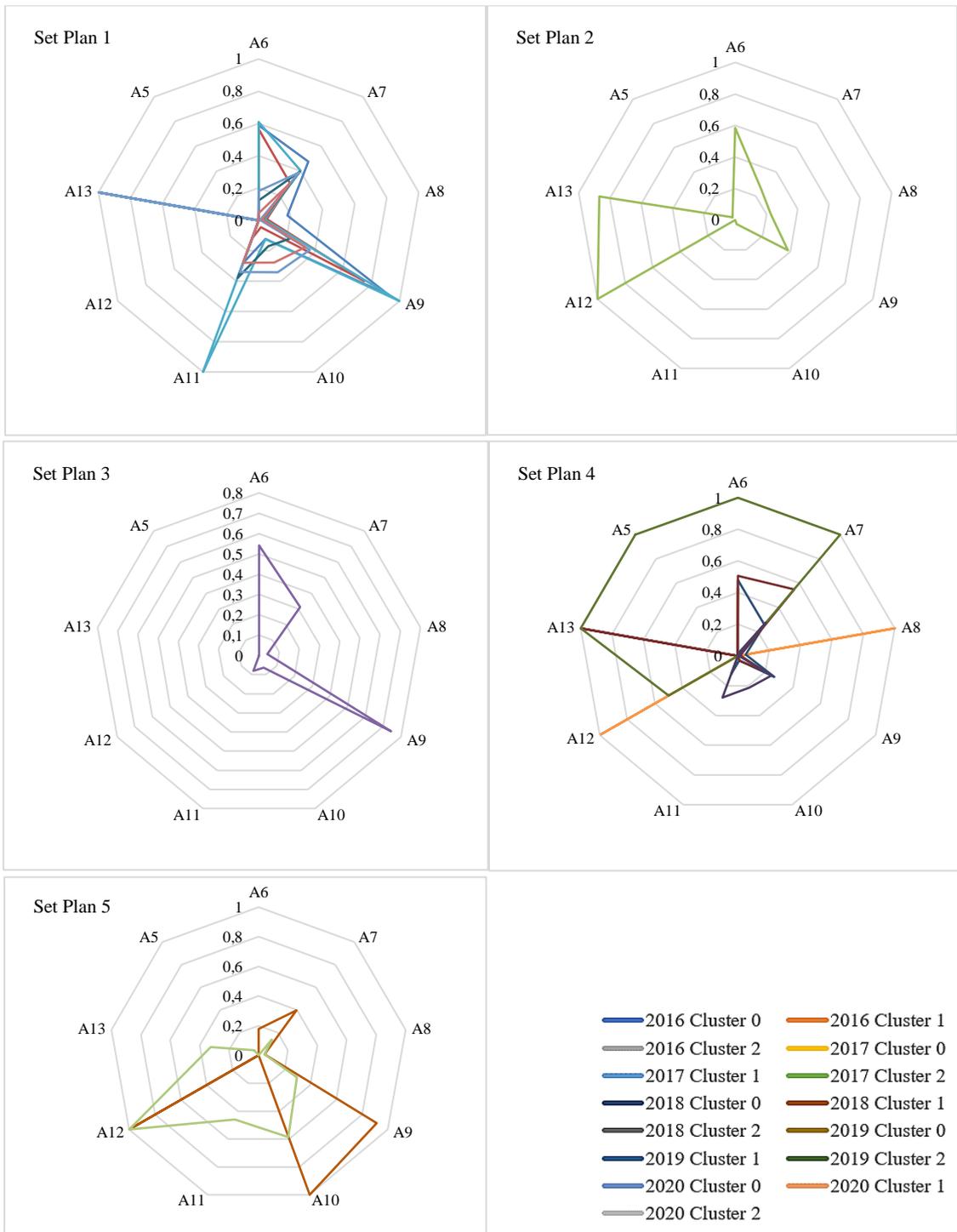


Figure 2. The normalized mean value of all attributes

Overall, it can be seen that the most contributed attribute was A9, A12, and A13, while other attributes accounted for the minimum contribution. Furthermore, every plan set has a different proportion of attribute strength. Set plan 4 has more dominant attributes than others. But only A12 and A13 that similar to plan sets 1, 2, and 5. Both plans, sets 2 and 3, only have limited attribute contributions, so it doesn't have many aspects to consider.

Table 4. Summary cluster characteristics and recommended plan set

Year	Cluster	Cluster Characteristic	Plan Set
2016	0	Extreme severity, rainy season, not rural	1 (P1, P2, P4)
	1	High severity, dry season, not rural	1 (P1, P2, P4)
	2	Moderate severity, rural	2 (P1, P4)
2017	0	Moderate severity, dry season, not rural	3 (P1, P2)
	1	Extreme severity, rainy season, not rural	1 (P1, P2, P4)
	2	Moderate severity, dry season, rural, a high number of additional fever cases	4 (P1)
2018	0	Moderate severity, dry season, not rural	4 (P1)
	1	Moderate severity, rainy season, not rural	4 (P1)
	2	Low severity, a high number of houses with larvae, rainy season	4 (P1)
2019	0	Moderate severity, dry season, not rural	4 (P1)
	1	Moderate severity, rainy season, not rural	1 (P1, P2, P4)
	2	Extreme severity, dry season, rural	5 (P1-P5)
2020	0	High severity, rainy season, not rural	1 (P1, P2, P4)
	1	High severity, dry season, not rural	1 (P1, P2, P4)
	2	Extreme severity, rural, far away from the hospital	5 (P1-P5)

On the other view, each set plan has the same graphic area, the visual area for set plan 5 is at the bottom, set plan 4 is on the left, set plan 3 is on the right, and set plan 2 is on the left. The similar relationship between the graph areas proves an effect of value on each attribute in determining the selected plan. The larger the graphic area displayed, the more plans are needed.

The set plan was formed based on a collection of contingency plans, and Table 4 shows the result of planning based on cluster conditions. Determination of the right set plan can be seen in each of the characteristics of each cluster. The following is a summary of cluster characteristics in 5 set plans:

- Set plan 1: extreme severity level & rural; or high severity level;
- b. Set plan 2: moderate severity level & rural;
  - c. Set plan 3: moderate severity level & not rural;
  - d. Set plan 4: moderate severity level; or low severity level;
  - e. Set plan 5: extreme severity level & not rural;

For instance, most clusters are advised that applying surveillance systems is fundamental in the dengue contingency plan. The plan recommends an alert algorithm plan for clusters with moderate-to-extreme severity cases. Regarding the type of season, if it rains, prioritize the algorithm plan as an alert measure. Meanwhile, managerial capacity and clinical service are the least recommended plans because they are only for clusters with higher fever cases (i.e., DHF, DF, and DSS) than other clusters. For vector control, it is recommended that clusters have moderate to extreme severity characteristics during the rainy season.

Moreover, the worse the cluster condition becomes the more recommended plans. Evidence of this appears in 2018 and 2019. In 2018, the number of dengue cases in Semarang was low. In addition, the attributes that become the parameters for evaluating other plans also have good values, so the recommended plan was only a surveillance system. On the other hand, in 2019, the number of cases in Semarang began to rise again. Furthermore, many of the parameters' attributes had poor values, so more than one plan was recommended.

### Visualizing with Map

After establishing all the clusters and acquiring set plans and rules, the data was visualized on the Semarang map, consisting of village boundaries and colours reflecting established plans. Figure 2 shows that five sets of plans are recommended for the dengue contingency plans, with differing colours for each plan. Each village is grouped, and coloured based on the clustering results for each village. One of the maps appears in Figure 4. We see that set plan number 1 is recommended for many rural parts or villages in Semarang,

and plan number 5 is rarely recommended. Set plan one consists of a surveillance system, alert algorithm, and vector control. This case is accurate, as the government tends to execute this plan most of the time. Set plan two consists of only a surveillance system, only advised in rural areas where dengue cases are rare for a specific year, and the need for vector control and alert algorithms is minimal. Set plan five consists of all plans recommended when the area has a high severity level. Meanwhile, the undefined label signifies that the cluster area had no DHF case records, so it was not included in the data processing.

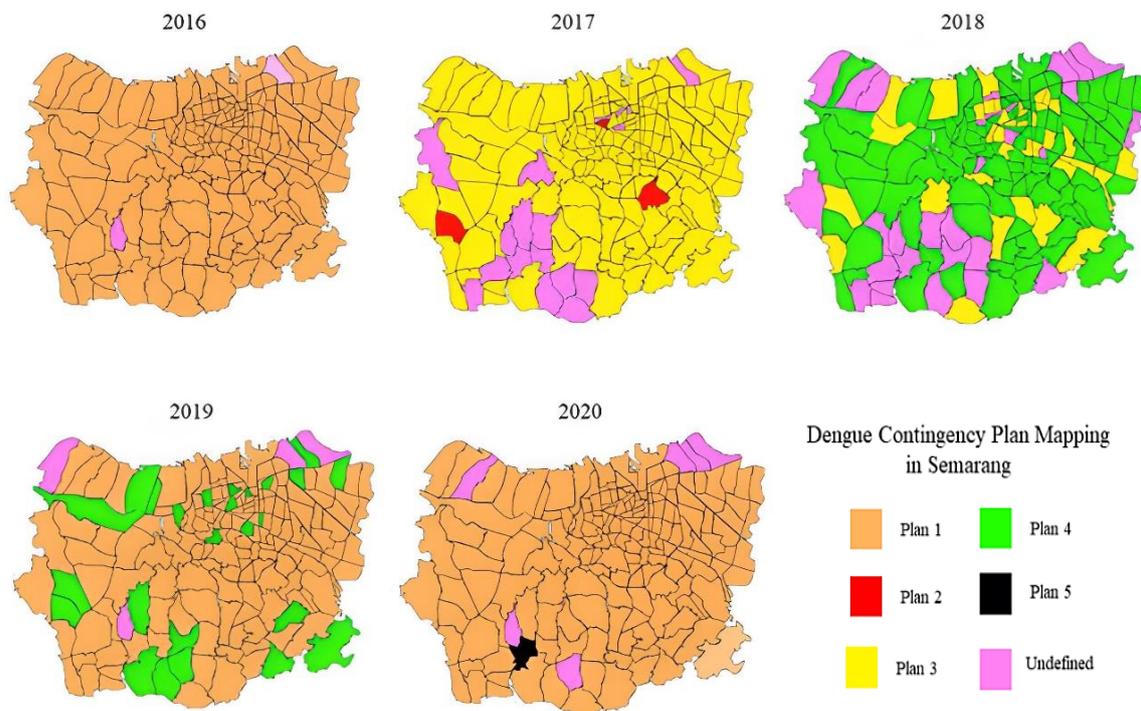


Figure 3. Visualization of Semarang map with a dengue contingency plan

### Discussion

Dengue has occurred in Indonesia for years but has not experienced a significant decrease. The government looks slow in dealing with it. This is because local regional policies are more inclined to dengue preventive measures [3]. So, with this, it is necessary to have a comprehensive dengue contingency plan mapping considering all aspects that affect the increase in morbidity. This study discusses how to determine a dengue contingency plan using clustering. Many previous studies only discussed how to predict dengue and its prevention using clustering and classification methods, but no one has used it to determine the dengue contingency plan from cluster analysis. This research consists of four test scenarios. The first scenario calculates the silhouette score in determining the selection of the feature group combination that will be used. Of the four combinations of all the features, environmental demography, environmental physical, and physical demography, it was found that the combination of features of the environmental demographic group was the most suitable for use in this data.

The second scenario searches for a clustering algorithm with the best silhouette score results. K-Means, Agglomerative, and DBSCAN found that K-Means got better results than others. K-Means performs very well on extensive data, while agglomerative does not. K-Means also has a faster computation because it has simpler calculations than DBSCAN. The third scenario assesses the severity level of each cluster area. This severity level is done by performing K-Means on the mean value attributes of each cluster. From the cluster results, four levels of severity were found, namely low, moderate, high, and extreme. This can facilitate the next process in doing the mapping. The last scenario is to map areas with the same characteristics using the same colouring. From the mapping results, it is known that the plan results are in accordance with the conditions of dengue in the city of Semarang. This can be seen from the increasing

number of cases in the region, the more plans are recommended, and if on the contrary there are fewer recommended plans.

## CONCLUSION

This study presents mapping dengue contingency plans using clustering. Experimental results show that the K-Means are superior to Agglomerative and DBSCAN because they have a higher silhouette score than the others. The proposed method could recognize demographics and the environment as the relevant attributes that map the dengue contingency plan. This study also identifies an area experiencing severe, moderate, and good levels that will adjust the plan accordingly. The better the area's condition, the fewer plans will be recommended. The worse an area is, the more plans will be recommended. For future work, the results of this study can be used as starting points for identifying and developing logical rules based on dengue contingency plan mapping.

## REFERENCES

- [1] KEPMENKES, "Tentang Pemberantasan Penyakit Demam Berdarah Dengue," No. 581/MENKES/SK/VII/1992, 1992.
- [2] Kementerian Kesehatan Republik Indonesia, "Profil Kesehatan Indonesia 2019," Jakarta, 2020.
- [3] S. Runge-Ranzinger *et al.*, "Dengue Contingency Planning: From Research to Policy and Practice.," *PLoS Negl. Trop. Dis.*, vol. 10, no. 9, Sep. 2016.
- [4] M. J. Watts, P. Kotsila, P. G. Mortyn, V. Sarto I Monteys, and C. Urzi Brancati, "Influence of Socio-Economic, Demographic and Climate Factors on The Regional Distribution of Dengue in The United States and Mexico.," *Int. J. Health Geogr.*, vol. 19, no. 1, p. 44, Nov. 2020.
- [5] B. E. E. Martina, P. Koraka, and A. D. M. E. Osterhaus, "Dengue Virus Pathogenesis: An Integrated View.," *Clin. Microbiol. Rev.*, vol. 22, no. 4, pp. 564–581, Oct. 2009.
- [6] I. H. Susilowati *et al.*, "Descriptive Analysis of the Contingency Plan in Cilegon City, Indonesia.," *KnE Life Sci.*, vol. 4, no. 5, pp. 532–543, 2018.
- [7] A. Fariza, Mu'Arifin, and D. W. Astuti, "Spatial-Temporal Visualization of Dengue Hemorrhagic Fever Vulnerability in Kediri District, Indonesia, Using K-means Algorithm.," in *2021 Int. Conf. Data Softw. Eng.*, 2021, pp. 1–6.
- [8] S. Vandhana and J. Anuradha, "Dengue Prediction Using Hierarchical Clustering Methods," in *Des. Digit. Glob. World*, pp. 157–168, 2018.
- [9] G. M. Nandana, S. Mala, and A. Rawat, "Hotspot Detection of Dengue Fever Outbreaks Using DBSCAN Algorithm," in *2019 9th Int. Conf. Cloud Comput. Data Sci. Eng.*, 2019, pp. 158–161.
- [10] W. H. Organization, *Technical handbook for dengue surveillance, outbreak prediction/detection and outbreak response*. World Health Organization, 2016.
- [11] T. B. Tunggal Satoto, A. H. Dwiputro, R. N. Risdiyanto, A. U. F. Hakim, N. A. Pascawati, and A. Diptyanusa, "Prediction Model of Dengue Hemorrhagic Fever Transmission to Enhance Early Warning System in Gergunung Village, Klaten District, Central Java," *J. Med. Sci.*, vol. 51, no. 03, pp. 258–269, 2019.
- [12] H. L. Dangiran and others, "Spatial Analysis of Dengue Hemorrhagic Fever Incidence in Salatiga City, Central Java, Indonesia," *Adv. Sci. Lett.*, vol. 23, no. 4, pp. 3376–3379, 2017.
- [13] W. Anggraeni *et al.*, "Prediction of Dengue Fever Outbreak Based on Climate and Demographic Variables Using Extreme Gradient Boosting and Rule-Based Classification," in *2021 IEEE 9th Int. Conf. Serious Games Appl. Health*, 2021, pp. 1–8.
- [14] M. Arif and S. S. Lestari, "Manajemen Komunikasi Bencana Pemerintah Desa dalam Penanganan Covid-19," *J. Commun. Stud.*, vol. 3, no. 1, p. 35, 2021.
- [15] Kemenkes RI, "Pusat Data dan Surveilans Epidemiologi Kementerian Kesehatan Republik Indonesia," 2010.
- [16] M. R. Ridha, L. Indriyati, A. Tomia, and Juhairiyah, "The Influence of Climate on the Incidence of Dengue Hemorrhagic Fever in Ternate City," *Spirakel*, vol. 11, no. 2, pp. 53–62, 2019.
- [17] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," *Front. Energy Res.*, vol. 9, no. March, pp. 1–17, 2021.
- [18] T. W. Kesetyaningsih, S. Andarini, Sudarto, and H. Pramoedyo, "Determination Of Environmental Factors Affecting Dengue Incidence In Sleman District, Yogyakarta, Indonesia," *African J. Infect. Dis.*, vol. 12, no. 1, pp. 13–25, 2018.
- [19] A. Setiyoutami, D. Purwitasari, W. Anggraeni, E. M. Yuniarno, and M. H. Purnomo, "Modelling Dengue Spread as Dynamic Networks of Time and Location Changes," *Int. J. Intell. Eng. Syst.*,

- vol. 14, pp. 346–358, 2021.
- [20] N. Shahriar, S. M. Akib Al Faisal, M. M. Pinjor, M. Al Sharif Zobayer Rafi, and A. Rahman Sarkar, “Comparative Performance Analysis of K-Means and DBSCAN Clustering algorithms on various platforms,” in *2019 22nd Int. Conf. Comput. Inf. Technol.*, 2019, pp. 1–6.