Comparison of K-Medoids and Self Organizing Maps Algorithm in Grouping Hydrometeorological Natural Disasters in Java Island

Novendri Isra Asriny*, Desy Rizki Ramadhanty, Muthia Dishanur Izzati, Matiin Laugiwa Prawira Putra, Annisa Shofura, Chairani Uli Albi, Edy Widodo
Universitas Islam Indonesia

E-mail: *17611040@students.uii.ac.id

Abstract. Indonesia is a country that is prone to disasters. Based on data from the National Disaster Management Authority, there are 6868 disasters in the last 3 years. Changes in weather, temperature, and wind direction that are quite extreme in Indonesia lead to one of the disasters namely hydrometeorological disasters such as floods, landslides, tornadoes, forest fires, and drought. According to the National Coordinating Agency for Disaster Management, this type of disaster is quite common in Indonesia. Over the past 3 years, Java Island received the highest number of disasters. The National Disaster Management Authority states that tornadoes contribute 21% of all disasters in Indonesia. Therefore, the researcher wants to grouping the cases of hydrometeorological natural disaster in Indonesia, especially in Java. This research is to determine the areas that are the focus of the government in disaster mitigation efforts using K-Medoids Clustering and Self Organizing Maps. Based on the results of this study, it was found that the suitable method used was the Self Organizing Maps Method because it was seen based on a smaller standard deviation value. Cluster one there are 108 members who are groups with relatively low disaster vulnerability, while cluster two there are 10 members are city and district groups that have relatively high disaster vulnerability.

1. Introduction
Disasters are occasions that undermine and disturb people's lives caused by normal, non-natural, and human variables that result in fatalities, natural harm, property misfortunes, and mental impacts [11]. Indonesia is a disaster-prone country caused by several factors. Based on data from the National Disaster Management Authority, the number of natural disasters in Indonesia is 6868 events [5].

Indonesia is a tropical locale with two seasons, to be specific warm and rain. Changes in climate, temperature, and wind heading are extraordinary sufficient to cause disaster, one of which is hydrometeorological calamities such as landslide, tornadoes, floods woodland fires, and dry spell [4]. These types of disasters occur quite frequently in Indonesia according to disaster data from the National Coordinating Agency for Handling Disaster (BAKORNAS PB).

Over the past 3 years, Java Island received the highest number of disasters. A tornado is the most frequent disaster, from BNPB data, it shows that a tornado contributed 21% of all disasters in Indonesia [12]. Therefore, the researcher wants to group of hydrometeorological natural disaster cases in Indonesia, especially in Java to determine areas that are the focus of the government in disaster mitigation efforts using K-Medoids Clustering and Self Organizing Maps which then can be identified groups of disaster-prone areas. In this research, a K-Medoids Clustering and Self Organizing Maps
2. Literature Study

2.1. K-Medoids Clustering

The K-Medoids algorithm or what is known as the PAM (Partitioning Around Medoid) algorithm was developed by Leonard Kaufman and Peter J. Rousseeuw, which is similar to K-means because of the two algorithms it is partitional which breaks the dataset into groups [22]. This method is based on the use of medoids rather than observations of the mean held by each cluster, to reduce the sensitivity of the partition for the extreme values that exist in the dataset [21]. K-Medoids is present to overcome the weaknesses of K-Means that are sensitive to outliers due to an object with a large value that may deviate substantially from the distribution of data [13]. K-Medoids uses the partition grouping method to group n sets of objects into some k clusters. This algorithm uses objects in a collection of objects that represent a cluster. The objects that represent a cluster are called medoids. Medoid is an object that is located centrally in a cluster so that it is robust to outliers.

2.2. Self-Organizing Maps (SOM)

Clusters validation that researchers use is internal validation, because this validation is based on evaluating the results of clustering in a quantitative concept in which the agricultural sector production data is also in the form of quantitative data. There are several methods or methods for internal validation, including connectivity, silhouette values, and Dunn index. Dunn index is the ratio of the smallest distance between observations in different clusters with the largest distance in each data cluster. Dunn index is obtained from the division between the smallest distance between observations in different clusters with the largest distance in each data cluster. Silhouette index is calculated as the degree of confidence in the process of clustering on observation with a cluster that is said to form well when the index value approaches 1 and the conditions conversely if the index value approaches the number -1. The value of silhouette in the range of -1 to 1. Connectivity index is seen based on characteristics that have the closest distance (neighbor). This is a parameter in determining the number of neighbors who contribute to the measurement of connectivity.

2.3. Silhouette Coefficient

Silhouette coefficient is used to see quality and cluster strength, how well an object is positioned in a cluster. This method is a combination of cohesion and separation method. Calculation stages of the Silhouette coefficient are as follows [1]:
1. Calculate the average distance from observation, for example i with all the other observations
   that are in one cluster:
   \[
   a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j)
   \]  
   (1)

2. Calculate the average distance from observation i to all observations in other clusters, and take
   the smallest value:
   \[
   d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j)
   \]  
   (2)

   Where \(d(i, C)\) is the average distance of observation i with all objects in other cluster C where \(A \neq C\).

3. The Silhouette Coefficient values are:
   \[
   s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}
   \]  
   (3)

2.4. Within Sum Square (WSS) Method

   The WSS method is a method used to measure the average square distance from all points in the cluster
to the cluster center point. Clusters with low WSS values have better closeness between objects than
clusters with higher WSS values. The formula used in WSS is,
   \[
   WSS = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2
   \]  
   (4)

   where WSS is the number of squares in, \(k\) is the number of clusters, \(n_i\) is the sample size of cluster i,
   \(X_{ij}\) is the j measurement of the i cluster, and \(X_{ij}\) is the average value of the overall distance of the data
[9].

2.5. Standard Deviation (SD)

   Standard deviation is a statistical value that has the benefit of determining the distribution of data in the
sample, as well as how close the individual data is to the average sample value [10]. If the deviation
value shows a large value, it can be seen that the individual data points are far from the average sample
value [16]. Mathematically, the SD is [7]:
   \[
   s = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}}
   \]  
   (5)

   where
   \(s\) = Sample Standard Deviation
   \(X\) = Individual Value
   \(\bar{X}\) = Sample Mean
   \(n\) = Sample Size

3. Methods

3.1. Types of Data Sources

   The population in this study is data on the incidence of hydrometeorological natural disasters in Java,
which are available on the website of the National Disaster Management Authority (BNPB). Meanwhile,
12 variables were sampled, consisting of provincial, district/city, flood, landslide, tornado, drought, and
forest and land fires (Table 1).

3.2. Research Stages
The methods used in this study are the K-Medoids Method and Self Organizing Maps. Figure 1 shows a flowchart that will be used in the study.

Table 1. Research Variable

| No | Variable         | Dimension                          | Source  | Measurement |
|----|------------------|------------------------------------|---------|-------------|
| 1  | Province         | Province on the island of Java     | BNPB    | Nominal     |
| 2  | District/city    | District/city of each province in Java Island | BNPB    | Nominal     |
| 3  | Flood            | Number of flood disasters          | BNPB    | Interval    |
| 4  | Landslide        | Number of landslide disasters      | BNPB    | Interval    |
| 5  | Tornado          | Number of tornado disasters        | BNPB    | Interval    |
| 6  | Drought          | Number of drought disasters        | BNPB    | Interval    |
| 7  | Forest and Land Fires | Number of forest and land fires disasters | BNPB    | Interval    |

Figure 1. Research Flowchart
Based on Figure 1, the following are the stages of data collection and processing carried out:

1. Perform data preprocessing.
2. Determining the number of clusters that will be used using the Silhouette Score method. The silhouette score coefficient is a method used to evaluate clusters and see the quality of data placement in a cluster based on the Sum Square of Error (SSE) value. Silhouette Score is important to see if the resulting cluster is of good quality. The stages of calculating the silhouette coefficient are as follows [14].
   a. Calculate the average distance of object $i$ to all objects in the group. We call this the mean distance $i$.
   b. Calculate the average distance of object $i$ to all objects in the other cluster we call $b(i)$, and take the smallest value.
   c. Silhouette coefficient values are obtained from:
      \[
      S(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))}
      \]  
      and can be written
      \[
      S(i) = \begin{cases} 
      1 - \frac{a(i)}{b(i)}, & \text{if } a(i) < b(i) \\
      0, & \text{if } a(i) = b(i) \\
      \frac{b(i)}{a(i)} - 1, & \text{if } a(i) > b(i) 
      \end{cases}
      \]  
      where:
      $S(i) = \text{Silhouette Score}$
      $a(i) = \text{Average distance between data } i \text{ and all objects in the cluster}$
      $b(i) = \text{Average distance between data } i \text{ to all objects in other clusters}$
      The range of values of the silhouette coefficient is -1 to 1. If the value of the silhouette coefficient is close to 1 then the object is in the right cluster. Meanwhile, if it is around 0 then the object can be between the clusters, and if the result is negative then the object may be in the wrong cluster [18]. The best number of clusters or the number of optimal clusters is the number of clusters with the highest average silhouette score where the average is taken from the silhouette value of each cluster.
3. Using K-Medoids cluster technique and Self-Organizing Maps.
   a. Stages in the K-Medoids cluster technique [13]:
      i. Initialize the cluster center (medoid)
      ii. Calculates Euclidean distances
         \[
         D_{ik} = \sqrt{\sum_{i=1}^{m} (x_{ij} - c_{kj})^2}
         \]  
         \[
         D_{ik}(x_i, x_j) = \text{EuclideanDistance} \\
         x_i = \text{Data} \ (i) \\
         x_j = \text{Data} \ (j) \\
         x_{ij} = \text{Data} \ \text{(i)attribute} \ (j) \\
         c_{kj} = \text{Data} \ \text{(k)attribute} \ (j)
         \]
      iii. Calculates the total distance
      iv. Repeat steps (i-iii) until the new total distance minus the old total distance $> 0$
   b. Stages in the Self-Organizing Maps cluster technique [20]:
      i. Initialize a number of weights with random values
For each sample randomly selected in the dataset,

1. Looking for the weighting closest to the Euclidean distance which is then called the BMU (Best Matching Unit)
2. Updates weight with formula

\[ w_{ij} = w_{ij} + \alpha (x - w_{ij}) \]  

3. Perform iterations to update the weights until they reach the predetermined iteration value and reduce the learning rate (\( \alpha \)).
4. Repeat from step 2

c. Comparing the results of clustering using the K-medoids and SOM methods based on the standard deviation ratio between clusters. The standard deviation ratio (R) was successfully used by [17] to determine the best method between the complete linkage and average linkage methods. The R value is a comparison between the standard deviation in the \( S_w \) group and the standard deviation between the \( S_b \) groups, the minimum R value indicates the best cluster method algorithm. The following equation is used to calculate the values of \( S_w \) and \( S_b \):

\[
S_k = \sqrt{\frac{\sum_{k=1}^{n}(x_k - \bar{x}_k)^2}{n - 1}}
\]

\[
S_w = \frac{1}{n} \sum_{k=1}^{n} S_k
\]

\[
S_b = \sqrt{\frac{\sum_{k=1}^{n}(\bar{x}_k - \bar{x})^2}{n - 1}}
\]

\[
R = \frac{S_w}{S_b}
\]

\[
\text{Number of Hydrometeorological Disasters}
\]

**Figure 2.** Descriptive analysis results
4. Results and Discussion

4.1. Descriptive Analysis
Researchers will conduct a descriptive analysis of the data on the number of hydrometeorological disasters. Descriptive analysis was carried out to find a general description of the development of hydrometeorological disasters during the 2017-2019 period.

Based on Figure 2, there are 20 districts with the highest number of hydrometeorological disasters among other districts in Java. Cilacap Regency is the district with the highest number of hydrometeorological disasters on Java Island of 195 incidents during the 2017-2019 period.

4.2. Determination of Cluster Optimal Number
The formation of the number of clusters is obtained using the silhouette and elbow method approach to see the comparison of Sum Square Error (SSE) values to get the optimal cluster formed.

![Graph of the Silhouette (a) and WSS (b) method](figure3.png)

**Figure 3. Graph of the Silhouette (a) and WSS (b) method**

Based on the SSE silhouette and WSS graph output in Figure 3, an optimal cluster is obtained, namely, the two clusters seen from SSE in the silhouette method in the second cluster are the highest average silhouette values and if viewed based on the value within the sum square in the plot of the WSS method then seen in cluster two has a significant change to cluster three. So that the optimal cluster in

![Output fan diagram](figure4.png)

**Figure 4. Output fan diagram**
this study uses the optimal cluster of two clusters on the number of Hydrometeorological disasters in Java.

4.3. Clustering Analysis with Self Organizing Maps

A model with an R program is obtained in the form of a fan diagram from the analysis of the SOM algorithm. The researcher uses a fan diagram with a hexagonal display with a 5 x 5 grid. The fan diagram is an illustration of the distribution of mapping variables. In the fan diagram below there are 2 different colors, which show the results of the clusters of each district in Java. Each color in the fan diagram has different characteristics. The SOM algorithm can be seen in Figure 4.

The researcher made a table containing the results of the clustering of regencies in Java, based on the number of hydrometeorological disasters.

| Cluster | Total | Cluster Member |
|---------|-------|----------------|
| 1       | 108   | Bandung, Bandung City, Bangkalan, Banjar City, Bantul, Banyuwangi, Batang, Batu City, Bekasi, Bekasi City, Blitar, Blitar City, Blora, Bojonegoro, Bondowoso, Boyolali, Brebes, Ciamis, Cianjur, Cilegon City, Cimahi City, Cirebon, Cirebon City, Demak, Depok City, East Jakarta, Garut, Gresik, Grobogan, Gunung Kidul, Indramayu, Jember, Jepara, Jombang, Karanganyar, Karawang, Kebumen, Kediri, Kediri City, Kendal, Kepulauan Seribu, Klaten, Kudus, Kulon Progo, Kuningan, Lamongan, Lebak, Lumajang, Madiun, Madiun City, Magelang City, Magetan, Main Jakarta, Majalengka, Malang, Malang City, Mojokerto, Nganjuk, Ngawi, North Jakarta, Pacitan, Pamekasan, Pandeglang, Pangandaran, Pasuruan, Pasuruan City, Pati, Pekalongan, Pekalongan City, Pemalang, Ponorogo, Probolinggo, Probolinggo City, Purbalingga, Purwakarta, Purworejo, Rembang, Salatiga City, Sampang, Semarang, Serang, Serang City, Sidoarjo, Situbondo, Sleman, South Jakarta, South Tangerang City, Sragen, Subang, Sukabumi City, Sukoharjo, Sumedang, Sumenep, Surabaya City, Surakarta City, Tangerang, Tangerang City, Tasikmalaya, Tasikmalaya City, Tegal, Tegal City, Trenggalek, Tuban, Tulungagung, West Bandung, West Jakarta, Wonosobo, Yogyakarta City |
| 2       | 10    | Banjarnegara, Banyumas, Bogor, Bogor City, Cilacap, Magelang, Semarang City, Sukabumi, Temanggung, Wonogiri |

From the cluster using the Self Organizing Maps algorithm, the result is that cluster one consists of 108 cities and regencies that have a high similarity value and there are 10 cities and regencies in cluster two in Java that have high similarity.

The researcher segmented the calculation of the average occurrence value of each cluster group and the results were summarized as in Table 3.

| Variable   | Cluster 1 | Cluster 2 |
|------------|-----------|-----------|
| Flood      | 6.592593  | 16.9      |
| Landslide  | 7.490741  | 61.8      |
| Tornado    | 10.26852  | 49.3      |
| Drought    | 1         | 2.4       |
| Forest Fires | 1.898148  | 3.8       |

Table 3 shows the results of the average value of disaster events, cluster 1 is a cluster with areas that are not too vulnerable to hydrometeorological disasters. Meanwhile, cluster 2 is a cluster with areas that are prone to hydrometeorological disasters. Based on the calculation of the average of each group
obtained in the first cluster tornado disaster to be the disaster that has the highest average means that the disaster occurs more frequently, while in the second cluster it is found that landslide disasters are more frequent because they have an average value higher.

4.4. Clustering Analysis with K-Medoids
Using cluster analysis with the K-Medoids algorithm the results are in the form of a cluster plot consisting of two clusters as in Figure 5.

![Cluster plot](image)

**Figure 5.** K-Medoids clustering output

| Cluster | Total | Cluster Member |
|---------|-------|----------------|
| 1       | 86    | Bandung, Bandung City, Bangkalan, Banjar City, Bantul, Banyuwangi, Batang, Batu City, Bekasi, Bekasi City, Blitar, Blitar City, Blora, Bojonegoro, Bondowoso, Ciamis, Cianjur, Cilegon City, Cimahi City, Cirebon, Cirebon City, Demak, Depok City, East Jakarta, Garut, Gresik, Grobogan, Gunung Kidul, Indramayu, Jombang, Karawang, Kediri, Kediri City, Kendal, Kepulauan Seribu, Klaten, Kudus, Kulon Progo, Lamongan, Lebak, Lumajang, Magelang, Magetan, Main Jakarta, Malang, Malang City, Mojokerto, Nganjuk, Ngawi, North Jakarta, Pacitan, Pamekasan, Pandeglang, Pasuruan, Pasuruan City, Pekalongan, Pekalongan City, Probolinggo, Probolinggo City, Purwakarta, Rembang, Salatiga City, Sampang, Serang City, Sleman, South Jakarta, Subang, Sukabumi City, Sukoharjo, Sumedang, Sumenep, Surabaya City, Surakarta City, Tangerang, Tangerang City, Tangerang South City, Tasik Malaya, Tegal, Tegal City, Tuban, Tulungagung, Yogyakarta City |
| 2       | 32    | Banjarnegara, Bayumas, Bogor, Bogor City, Boyolali, Brebes, Cilacap, Jember, Jepara, Karanganyar, Kebumen, Kuningan, Magelang, Majalengka, Pangandaran, Pati, Pemalang, Ponorogo, Purwakarta, Semarang, Semarang City, Serang, Sidoarjo, Situbondo, Sragen, Sukabumi, Tasikmalaya City, Temanggung, Trenggalek, West Bandung, Wonogiri, Wonosobo |
K-Medoids output shows that there are two clusters obtained based on the calculation of similarity between observations of the number of hydrometeorological disasters that occurred in Java. Table 4 shows the results of the profiling are shown by clustering.

From the cluster using the K-Medoids algorithm, the result is that cluster one consists of 86 cities and regencies that have a high similarity value and there are 32 cities and regencies in cluster two in Java that have high similarity.

The researcher segmented the calculation of the average occurrence value of each cluster group and the results were summarized as in Table 5.

Table 5 shows the results of the average value of disaster events, cluster 1 is a cluster with areas that are not too vulnerable to hydrometeorological disasters. Meanwhile, cluster 2 is a cluster with areas that are prone to hydrometeorological disasters. Based on the calculation of the average of each group obtained in the first cluster tornado disaster to be the disaster that has the highest average means that the disaster occurs more frequently, while in the second cluster it is found that landslide disasters are more frequent because they have an average value higher.

| Variable  | Cluster 1 | Cluster 2 |
|-----------|-----------|-----------|
| Flood     | 5.58      | 12.5      |
| Landslide | 3.92      | 34.1      |
| Tornado   | 6.27      | 33.2      |
| Drought   | 0.872     | 1.78      |
| Forest Fires | 1.19     | 4.41      |

4.5. Calculation the Standard Deviation of the Self Organizing Maps Algorithm

From the grouping process using the SOM method described previously, 2 clusters have been obtained. Then the calculation of standard deviation in groups (S_h) and standard inter-group deviation (S_b) will then be made in the SOM method as follows:

1. Standard deviation in groups (S_w)
   Before calculating the S_w value the first k-group standard deviation is calculated. The following is an example of calculating the k-group standard deviation (S_k), by looking at the appendix, following the calculation of the standard deviation of cluster 1, where \( \bar{x}_1 = 5.45 \), and \( \bar{x}_2 = 26.84 \).

   \[
   S_1 = \sqrt{\frac{(0.2-5.45)^2 + (1.6-5.45)^2 + (2.2-5.45)^2 + \cdots + (0.2-5.45)^2}{108}} = 4.811416
   \]

   \[
   S_2 = \sqrt{\frac{(34.8-26.84)^2 + (20.2-26.84)^2 + (17-26.84)^2 + \cdots + (28.8-26.84)^2}{10}} = 6.255749
   \]

   For overall results the \( S_k \) value is as follows:

   | Cluster | Standard Deviation |
   |---------|-------------------|
   | 1       | 4.811416          |
   | 2       | 6.255749          |

   After getting the k-standard deviation value, then the standard deviation in the group can be calculated as follows:

   \[
   S_w = \frac{4.811416 + 6.255749}{2} = 5.533583
   \]

2. Standard deviation between groups (S_b)
The standard deviation between groups \((S_b)\) can be calculated by first calculating the average of the whole group \(\bar{x}\) as follows:

\[
\bar{x} = \frac{5.45 + 26.84}{2} = 16.145
\]

After obtaining the average of the whole group \(\bar{x}\), the standard deviation between groups \((S_b)\) can be calculated as follows:

\[
S_b = \sqrt{\frac{(5.45 - 16.145)^2 + (26.84 - 1.145)^2}{2}} = 10.695
\]

3. Ratio \(S_w\) and \(S_b\)

According to Barakbah and Arai (2007) in Laraswati (2014), from the minimum ratio of \(S_w\) to \(S_b\), in this case, a good method is seen from the minimum ratio value, as follows [3]:

\[
S = \frac{5.533583}{10.695} \times 100\% = 0.517399
\]

4.6. Calculation the Standard Deviation of the K-Medoids Method

From the grouping process using the K-Medoids method previously described, 2 clusters were obtained. Then the calculation of standard deviation in groups \((S_w)\) and inter-group deviation \((S_b)\) in the K-Medoids method will then be performed as follows:

1. Standard deviation in groups \((S_w)\)

Before calculating the \(S_w\) value the first k-group standard deviation is calculated. The following is an example of calculating the k-group standard deviation \((S_k)\), by looking at the appendix, following the calculation of the standard deviation of cluster 1, where \(\bar{x}_1 = 3.565\), and \(\bar{x}_2 = 17.2\).

\[
S_1 = \sqrt{\frac{(0.2 - 3.565)^2 + (1.6 - 3.565)^2 + (2.2 - 3.565)^2 + \cdots + (0.2 - 3.565)^2}{86}} = 2.947477
\]

\[
S_2 = \sqrt{\frac{(34.8 - 1.2)^2 + (20.2 - 17.2)^2 + (17 - 17.2)^2 + \cdots + (28.8 - 17.2)^2}{32}} = 7.898101
\]

For overall results the \(S_k\) value is as follows:

| Table 7. Standard deviation of the K-Medoids method |
|-----------------------------------------------|
| Cluster | Standard Deviation |
|---------|-------------------|
| 1       | 2.947477          |
| 2       | 7.898101          |

After getting the k-standard deviation value, then the standard deviation in the group can be calculated as follows:

\[
S_w = \frac{2.947477 + 7.898101}{2} = 5.422789
\]

2. Standard deviation between groups \((S_b)\)

The standard deviation between groups \((S_b)\) can be calculated by first calculating the average of the whole group \(\bar{x}\) as follows:

\[
\bar{x} = \frac{3.565 + 17.2}{2} = 10.38256
\]

After obtaining the average of the whole group \(\bar{x}\), the standard deviation between groups \((S_b)\) can be calculated as follows:
\[ S_b = \sqrt{\frac{(3.565 - 3.38256)^2 + (17.2 - 10.38256)^2}{2}} = 6.817442 \]

3. Ratio \( S_w \) and \( S_b 

According to Barakbah and Arai (2007) in Laraswati (2014), from the minimum ratio of \( S_w \) to \( S_b \) in this case a good method is seen from the minimum ratio value, as follows [3]:

\[ S = \frac{5.422789}{6.817442} \times 100\% = 0.795429 \]

4.7. Comparison of Standard Deviation of Self Organizing Maps and K-Medoids Method

According to Barakbah and Arai (2007) in Laraswati (2014), from the minimum ratio of \( S_w \) to \( S_b \) in this case, a good method is seen from the minimum ratio value because the smaller the odds ratio \( S_w \) and \( S_b \) mean that the more similarities between observations or observations the closer to the cluster center point. [3]

| No | Method               | \( S_w \)   | \( S_b \)   | Ratio \( S_w \) and \( S_b \) |
|----|----------------------|------------|------------|-------------------------------|
| 1  | Self-Organizing Maps (SOM) | 5.533583  | 10.695     | 0.517399                      |
| 2  | K-Medoids            | 5.422789   | 6.817442   | 0.795429                      |

Based on Table 8, the standard deviation ratio value shows that the Self Organizing Maps (SOM) method has better performance among the K-Medoids method. This can be seen from the value of the ratio of \( \sigma_w \) to \( \sigma_b \) in the Self Organizing Maps (SOM) method which is smaller than the K-Medoids method.

5. Conclusion

Based on the Self Organizing Maps method, it is found that the first cluster of 108 members is a group with relatively low disaster vulnerability, while the second cluster of 10 members is a group of cities and districts that have relatively high vulnerability to hydrometeorological disasters. Whereas in the K-Medoids method, the results obtained from cluster one as many as 86 members are groups with relatively low disaster vulnerability, while cluster two as many as 32 members are groups of cities and districts that have relatively high vulnerability to hydrometeorological disasters. The best clustering method in this study is the Self Organizing Maps method with a standard deviation ratio value of 0.517399. Where this value is smaller than the K-Medoids method.

6. Recommendation

This study obtained results where there are several areas on the island of Java that have a high level of risk of hydrometeorological disasters, this can be overcome by mitigating hydrometeorological disasters in areas that are the focus of hydrometeorological disasters, so that damage or loss of life caused by hydrometeorological disasters can be minimized.

References

[1] Anggara M, Sujiani H and Nasution H 2016 Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness Jurnal Sistem dan Teknologi Informasi (JUSTIN) vol 1 no 1
[2] Bação F L 2010 Introduction to Kohonen’s Self-Organizing Maps (Lisboa: Institutst Superior de Estatistica E Gestao de Informacao Universidade Nova de Lisboa Portugal)
[3] Barakbah A R and Kohei A 2004 Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering Retrieved from http://lecturer.eepis-its.edu/~ridho/papers/Barakbah_IES_2004.pdf
[4] BNPB 2019 Potensi Ancaman Bencana Retrieved from https://bnpb.go.id/potensi-ancaman-
bencana

[5] BNPB 2020 Tabel Data Informasi Bencana Indonesia Retrieved from https://bnpb.cloud/dibi/tabel1b

[6] Chrisnanto Y H 2015 Penerapan algoritma Partitioning Around Medoids PAM clustering untuk melihat gambaran umum kemampuan akademik mahasiswa SENTIKA 2015

[7] Dawson-Sanders B and Trapp R G 1990 Basic and Clinical Biostatistics (Norwalk, Connecticut: Appleton & Lange)

[8] Halim N N and Widodo E 2017 Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps SOM Seminar Nasional Integrasi Matematika dan Nilai-Nilai Islami (Malang: Mathematics Department) pp 188–94

[9] Hardi M P W, Kusuma D A and Basuki P 2019 pengelompokan topik dokumen berbasis text mining dengan algoritme k-means: studi kasus pada dokumen kedutaan besar Australia Jakarta Visi Pustaka 21(1) pp 67–76

[10] Hidayat R N, Sabri L M and Awaluddin M 2019 Analisis desain jaring GNSS berdasarkan fungsi presisi (studi kasus : titik geoid geometri Kota Semarang) Jurnal Geodesi Undip 8(1) pp 48–55

[11] Hilmi E, Hendarto E, Riyanti and Sahri A 2012 Analisis potensi bencana abrasi dan tsunami di pesisir Cilacap Jurnal Penanggulangan Bencana 35

[12] Iryanthony S B 2015 Pengembangan modul kesiapsiagaan bencana angin puting beliung untuk mahasiswa pendidikan geografi Unnes Jurnal Geografi

[13] Jiawei H and Kamber M 2006 Data Mining: Concept and Techniques Second Edition (San Francisco: Morgan Kaufmann)

[14] Kaufman L and Rousseeuw P J 1987 Clustering by means of Medoids Statistical Data Analysis Based on the Norm and Related Methods ed Y Dodge (Amsterdam: North-Holland) pp 405–16

[15] Mikhail and Gracie 1981 Analysis and Adjustment of Survey Measurement (New York: Van Nostrand Reinhold Company Inc)

[16] Mulyadi S 2018 Perbandingan Metode Fuzzy C-Means Terbobot dengan Fuzzy C-Shell Studi Kasus: Kejadian Bencana di Indonesia per-Kabupaten (Yogyakarta: Universitas Islam Indonesia)

[17] Ningisih S, Sri W and Nasution Y N 2016 Perbandingan kinerja metode complete linkage dan average linkage dalam menentukan hasil analisis cluster: studi kasus produksi palawija Provinsi Kalimantan Timur 2014/2015 Prosiding Seminar Sains dan Teknologi FMIPA UNMUL vol 1 no 1 pp 46–50

[18] Prasetyo E 2014 Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab (Yogyakarta: ANDI)

[19] Sandy A E H 2019 Implementasi K-Medoids algoritma pengelompokan untuk pola kelompok kejahatan di Yogyakarta Int. J. of Applied Sciences and Smart Technologies vol 1 no 1

[20] Septianusa and Supriyaniingsih 2014 Karakteristik dan segmentasi pertanian padi menggunakan algoritma Self Organizing Maps SOMs Kohonen Prosiding Seminar Nasional Menuju Masyarakat Madani dan Lestari

[21] Vercellis 2009 Sistem Informasi (Yogyakarta: Lokomedia)

[22] Wira B, Budianto A E and Wiguna A S 2019 Implementasi metode K-Medoids clustering untuk mengetahui pola pemilihan program studi mahasiswa baru tahun 2018 di Universitas Kanjuruhan Malang Jurnal Terapan Sains & Teknologi 1(3) pp 54–69