



*E-ISSN: 2620 - 4872**Vol.08, No.02**Doi:**10.21009/JKOMA.082.03*

*Received: 13 Dec 2025**Accepted: 22 Dec 2025**Published: 26 Dec 2025*

Keywords:*Clustering Large Applications;
Forest Fires;
Dunn;
Silhouette;****Correspondence Email:***ridyadestriani22@gmail.com*

Forest Fire Clustering in Indonesia Using the Clustering Large Applications (CLARA) Method

Ridya Destriani^{1*}, Muhammad Arib Alwansyah², Sigit Nugroho³, Nurul Hidayati⁴

^{1,3,4}Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Bengkulu, Indonesia

²Department of Mathematics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri, Indonesia

Abstract

Clustering is a process of grouping, observing or grouping classes that have similar objects. One clustering method that handles large amounts of data is clustering large applications (CLARA). This research aims to determine groups of forest fires in Indonesia using the CLARA method and to find out the results of forest fire characteristics and images of forest fire occurrence points in Indonesia. The data used is hotspot data totaling 3,265 events which can be obtained from the NASA LANCE–FIRM MODIS Active Fire website. The variables used to group forest fire events are latitude, longitude, brightness, frp and confidence. So by grouping 3,265 hot spot data by determining the optimum cluster using the Silhouette index and Dunn index values, the optimum cluster results were obtained, namely 2 clusters, the first cluster obtained an average brightness value of 321.18°K, an FRP value of 24.85 MW and a confidence value of 74.18%. Meanwhile, the second cluster has an average brightness value of 313.13°K, frp 11.85 MW and confidence 37.94%.

INTRODUCTION

Forests are one of Indonesia's greatest resources. Land conversion to agricultural land and logging in forest areas have economic impacts on communities, and if not managed properly, they can cause numerous problems. Forests, as a natural resource and a life support system, must be managed sustainably to provide both direct and indirect benefits. One form of forest management believed to meet these criteria is forest management through community empowerment, while also empowering them to act as partners in the protection and development of forest areas [1].

The decline in forest area in Indonesia is partly due to forest fires. These events directly impact the global ecosystem, increasing carbon emissions and losing biodiversity. Forest fires are also the largest environmental disasters in nature, disrupting life [2]. One analysis used to group data based on attribute similarities is clustering. The clustering method is a process of grouping, observing, or classifying objects with similar characteristics [3]. To easily identify the high potential for forest fires, this is a development of previous research. Previous research used K-medoids, while this research uses CLARA, because CLARA is an algorithm for clustering large-scale data containing outliers using sampling techniques and then applying the PAM algorithm by calculating the distance between objects to the cluster center using silhouette and dunn distances.

METHODS

2.1 Clustering

Clustering is a method in data mining that is used to group objects based on similar characteristics into several clusters [4]. In clustering, a measure is used that explains the similarity between data to explain the

structure of simpler groups and comes from more complex data, namely the similarity measure using Euclidean distance [5]. Clustering is included in the basic principle method of measuring the distance or similarity between objects in a clustering database, which is often done as an initial step in the data mining process. Many cluster algorithms have been used by previous researchers such as K-means, Improved K-means, Fuzzy C-means, DBSCAN, K-medoids (PAM), CLARA and Fuzzy Subtractive. Each algorithm has its own advantages and disadvantages.

The cluster analysis process can be carried out with the following steps [6]:

1. The initial clustering process involves measuring the similarity between objects by determining the distance between the data.
2. There are two methods for creating clusters: the hierarchical method and the non-hierarchical method. The main difference between the hierarchical and non-hierarchical methods is in determining the number of clusters. The hierarchical method is used when the number of clusters to be formed is unknown in advance. The non-hierarchical method begins by selecting the desired number of initial clusters.
3. After clusters are formed using the hierarchical or non-hierarchical method, the next step is to interpret the resulting clusters.
4. The final step is to validate the clusters.

2.2 Clustering Large Applications (CLARA)

Clustering Large Applications (CLARA) is a development of the K-medoids algorithm used for clustering large-scale data. CLARA can divide large amounts of data into smaller subsets to analyze the desired number of clusters that can be measured by calculating the average distance of each object's proximity in the data set, then selecting the sample that has the smallest distance [7]. One type of data grouping with the medoid as the center of the cluster. Medoid is an object that is located centrally in a cluster, or in other words, it is an object that represents a member of the data and has the smallest average difference with other members [8].

According to [9], the CLARA analysis process can be carried out with the following steps:

1. Initialize k cluster centers (the number of clusters).
2. Allocate each data object to the nearest cluster using the distance measure equation.
3. Randomly select an object in each cluster as a new candidate medoid.
4. Calculate the distance of each object to the new candidate medoid and assign each object to the nearest candidate medoid.
5. Calculate the total distance, then the difference between the total distances "(S)" by calculating the total new distance - the total old distance. If " $S < 0$," then swap the object with the cluster data to form a new set of k objects as medoids.
6. Repeat steps c through e until there are no changes in medoids, resulting in clusters and their respective members.

2.3 Euclidean Distance

Euclidean distance is a calculation to measure the distance between two points in a plane that studies the relationship between angles and the distance measured in a straight line from one point to another. Although this method is less realistic, it is generally used because it is easy to understand and easy to model [10]. The following is the Euclidean distance formula [11]:

$$D_{Euclidean} = \sqrt{\sum_{m=1}^n (x_{im} - x_{jm})^2} \quad (1)$$

2.4 Silhouette Index

The silhouette index method is a validation measure based on internal criteria. The silhouette index will evaluate the placement of each object in each cluster by comparing the average distance of objects in one cluster and the distance between objects in different clusters [12]. The silhouette coefficient is defined as the average $s(i)$, namely:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

where:

$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)} & , \text{jika } a(i) < b(i) \\ 0 & , \text{jika } a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1 & , \text{jika } a(i) > b(i) \end{cases} \quad (3)$$

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, if it is around 0 then the object can be between 2 clusters, and if the result is negative then the object may be in the wrong cluster. The best number of clusters or the optimal number of clusters is the number of clusters with the highest average silhouette score where the average is taken from the silhouette value of each cluster.

2.5 Indeks Dunn

The Dunn index is the ratio of the smallest distance between observations in different clusters to the largest distance within each data cluster [13]. For each pair of clusters (C_i, C_j) , calculate the distance between them. The distance between clusters is usually measured by the Euclidean distance between the observations within the clusters.

$$d(C_i, C_j) = \text{Distance between clusters } C_i \text{ and } C_j$$

Select the shortest distance between all pairs of distinct clusters. The minimum intercluster distance is:

$$\min_{i \neq j} \{d(C_i, C_j)\} \quad (4)$$

Using the minimum intercluster distance and the maximum intracluster distance calculated, use the Dunn Index formula.

$$D = \frac{\min_{i \neq j} \{d(C_i, C_j)\}}{\max_k \{\delta(C_k)\}} \quad (5)$$

Given that $\min_{i \neq j} \{d(C_i, C_j)\}$ is the shortest distance between distinct clusters and $\max_k \{\delta(C_k)\}$ is the maximum intracluster distance between the clusters. The Dunn Index is obtained by dividing d_{min} by d_{max}

$$Dunn = \frac{d_{min}}{d_{max}} \quad (6)$$

RESULTS AND DISCUSSION

3.1 Descriptive Statistics

This study aims to identify forest fire clusters in Indonesia using the CLARA method, as well as to identify forest fire characteristics and imagery of forest fire occurrence points in Indonesia.

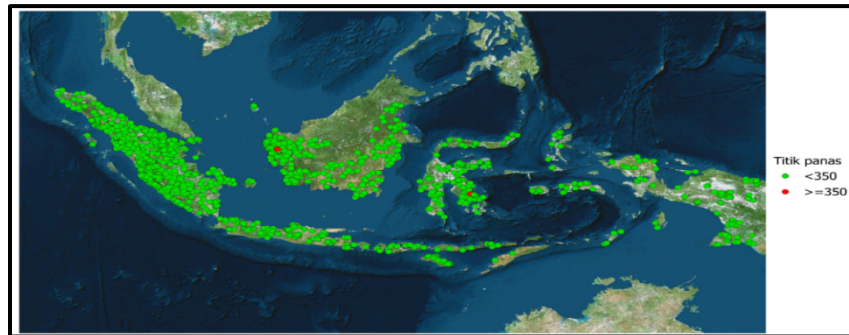


FIGURE 1. Distribution of Forest Fire Hotspots Brightness Data

Based on Figure 1, it can be seen that most of the forest fire incident points in Indonesia in the brightness data have a value of $< 350^{\circ}K$ are symbolized by green dots, while red dots are forest fire incidents with a value of more than $\geq 350^{\circ}K$.

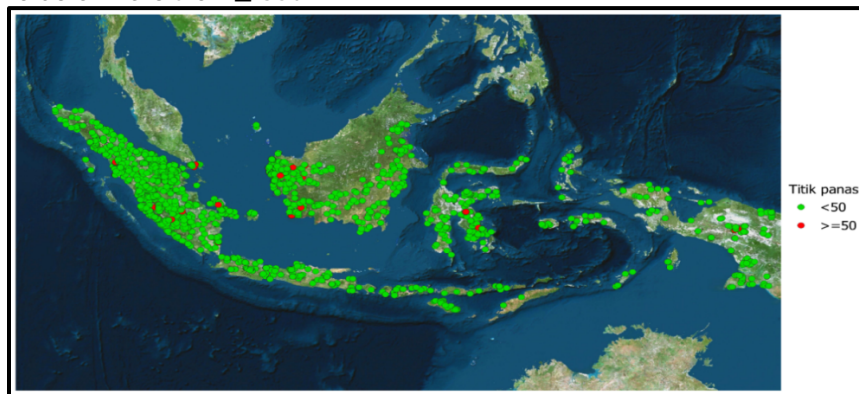


FIGURE 2. Distribution of Forest Fire Hotspots Frp Data

Based on Figure 2, it can be seen that the majority of forest fire incidents in Indonesia are in the FRP data. Forest fire incidents with a value of < 50 MW are symbolized by green dots, while red dots represent forest fire incidents with a value of ≥ 50 MW, which occurred in the islands of Sumatra and Kalimantan. Descriptive statistics for forest fire incidents in Indonesia can also be seen in the following table:

TABLE 1. Descriptive Statistics of Forest Fire Hotspots in Indonesia

Descriptive Statistics	Brightness($^{\circ}K$)	Frp(MW)	Confidence(%)
Mean	340,46	148,91	51,91
Minimum	300,1	2,9	0
Maximum	404,2	425,6	100
Mode	313,2	7,9	0
Standard Deviation	9,65	21,57	23,19

Based on Table 1, the distribution of forest fire hotspots in Indonesia for the brightness variable has a minimum value of 300.1, a maximum value of 404.2, a mode value of 313.2, a standard deviation value of 9.65, and a mean of 340.46. The frp variable has a minimum value of 2.9, a maximum value of 425.6, a mode value of 7.9, a standard deviation value of 21.57, and a mean value of 148.91. The confidence variable has a minimum value of 0, a maximum value of 100, a mode value of 0, a standard deviation value of 23.19, and a mean value of 51.91

3.2 Analysis for Determining the Number of Clusters

The number of clusters was tested using 3,265 test data sets, considering the necessary values. Several experiments using the Dunn index and the resulting silhouette index were conducted to determine the optimal number of clusters. The cluster value to be tested is the k value between 2 and 9. Table 2 shows the results of the Dunn index and silhouette index tests.

TABLE 2. Results of the Cluster Number Test

Total Data	K Value	Dunn Index Value	Silhouette Index Value
3265	2	0.0029	0.3137
	3	0.0008	0.2996
	4	0.0028	0.2748
	5	0.0029	0.1827
	6	0.0027	0.2255
	7	0.0027	0.1847
	8	0.0028	0.2226

Based on Table 2, it can be seen that the results of the testing process present the values of two validation measures indicating the number of clusters selected is increasingly optimal for each validation test measure. It can be seen in the table when the number of clusters 2 and 5 in the dunn index validation test shows the most optimal number of clusters is fairly the same, namely 0.0029. In the Silhouette index test, the number of clusters 2 shows the most optimal number of clusters, namely 0.3137, so we can see that the dunn index and silhouette index values in cluster 2 both obtain optimal values having the best cluster quality with the value of the resulting clustering structure is very good. From these results it can be seen that the quality of the clustering results formed is very good and can be accepted in the use of the value of $K = 2$ with the highest average value of the dunn index and silhouette index.

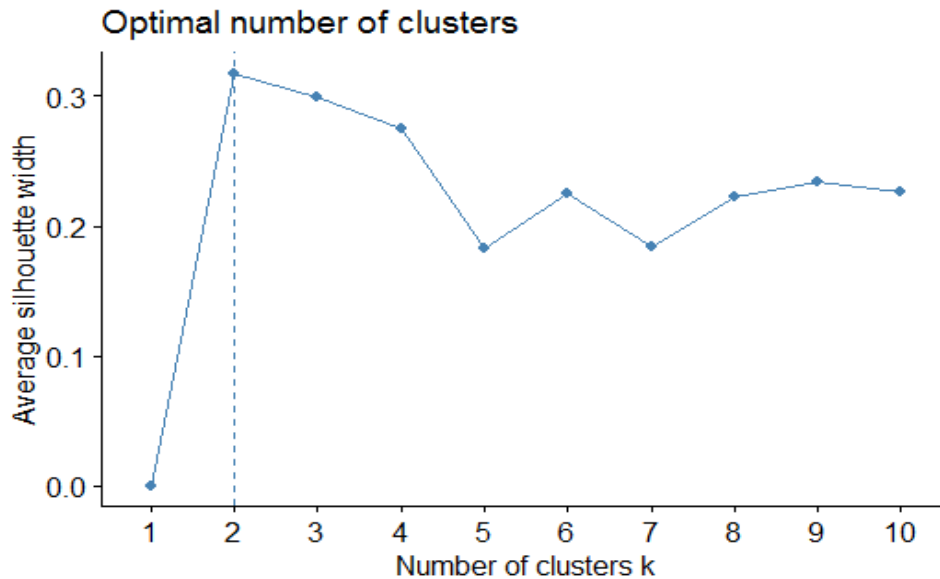


FIGURE 3. Graph of the Effect of the Number of Clusters on the Silhouette Index

Figure 3 shows that the highest silhouette index value, indicated by the dotted line, is obtained when $K = 2$. Therefore, the optimal number of clusters to be formed using the CLARA method is 2. The higher the k value, the lower the resulting silhouette index. This is because the silhouette value approaches zero, indicating that the cluster is closer to other clusters than to its intended cluster.

3.3 Descriptive Statistics of Optimal Cluster Results

Previous test results indicated an optimal number of clusters of two clusters with 3,265 data sets. The descriptive statistics for the optimal clusters for several variables, namely brightness, FRP, and confidence, are as follows:

TABLE 3. Results of the Analysis of Forest Fire Potential Determination

Cluster	Average Brightness(K)	Average FRP(MW)	Average Confidence(%)	Std. Deviation	Number of Data
1	321,18	24,85	74,18	17155,07	1.604
2	313,13	11,85	37,94	18702,89	1.661

Based on Table 3, it can be seen that cluster 1 has 1,604 members with an average brightness value of 321.18°K, frp 24.85 MW, confidence 74.18%, and a standard deviation value of 17155.07. Where the average brightness value is more than 320°K and the average confidence value is more than 70%, and has a fairly high frp value which indicates that cluster 1 is included in the high category of pollution produced, greatly affecting the global climate. While cluster 2 has 1,661 members with an average brightness value of 313.13°K, frp 11.85 MW, confidence 37.94%, and a standard deviation value of 18702.89, where cluster 2 is included in the moderate category so that precautions are needed against fires.

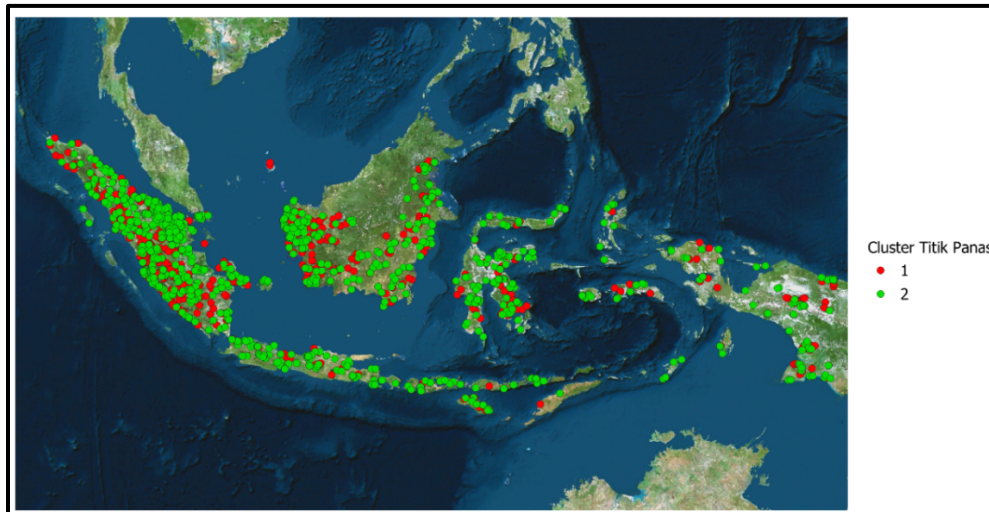


FIGURE 4. Forest Fire Clusters in Indonesia in 2022

Figure 4 shows the first cluster with green dots and the second cluster with red dots. The figure shows that the red dots are more numerous than the green dots. This indicates that the potential for forest fires is almost universal across Indonesia. These points with high fire potential are spread across most of the islands of Sumatra and Kalimantan, as well as several areas in eastern Indonesia. This indicates that the potential for forest fires in Indonesia is almost universal. Therefore, the government must take further action to mitigate forest fires.

CONCLUSION

Based on the results of the cluster analysis on forest fires in Indonesia, it can be concluded that the forest fire data in Indonesia obtained two cluster groups with a silhouette value of 0.3137 and a dunn of 0.0029 on data of 3,265 observations. Cluster 1 has a higher potential for forest fires, with an average brightness value of 321.18°K, FRP 24.85 MW, confidence 74.18% and a standard deviation value of 17155.07. On the other hand, cluster 2 with an average brightness value of 313.13°K, FRP 11.85 MW, confidence 37.94% and a standard deviation value of 18702.89 has a moderate potential for forest fires.

With descriptive statistics, the distribution value of forest fire hotspots in Indonesia on the brightness variable has a minimum value of 300.1, a maximum value of 404.2, a mode value of 313.2, a standard

deviation value of 9.65 and an average of 340.46 while the frp variable has a minimum value of 2.9, a maximum value of 425.6, a mode value of 7.9, a standard deviation value of 21.57 and an average value of 148.91, while the confidence variable has a minimum value of 0, a maximum value of 100, a mode value of 0, a standard deviation value of 23.19 and an average value of 51.91.

REFERENCES

- [1] Kementerian Lingkungan Hidup dan Kehutanan. 2016. Analisa Data Luas Areal Kebakaran. Jakarta. Direktorat Jenderal Planologi Kehutanan dan Tata Lingkungan.
- [2] Hatta. 2020. Deteksi Gen Interleukin-6 (IL-6) pada Sampel Klinis Tuberkulosis dengan Metode *Polymerase Chain Reaction* (PCR). [Skripsi]. Makassar (ID) : Universitas Hassanuddin.
- [3] Wira, B., Budianto, A., E, dan Wiguna, A., S. 2019. Implementasi Metode K-Medoids *Clustering* untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Thun 2018 Di Universitas Kejuruan Malang. *Jurnal Teknologidan Sains*. Vol. 1, No.3.
- [4] Damayanti, I., Hermawan, E., dan Kamilah, K,. 2021. Analisis *Spasial Clustering* Zona Potensi Ikan Konsumsi Air Tawar di Kabupaten Bogor. Krea-TIF: *Jurnal Teknik Informatika*. Vol 9 No. 2 pp. 11-20.
- [5] Johnson, R. A., Wichern, D. W., dan Others. 2002. *Applied multivariate statistical analysis* (Vol. 5). Prentice hall Upper Saddle River, NJ.
- [6] Sitepu, R., Irmeilyana, dan B. Gultom. 2011. Analisis *Cluster* terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumatera Selatan. *Jurnal Penelitian Sains*. Vol. 14, No. 3.
- [7] Zulyanti, T., dan Noeryanti. 2022. Perbandingan pengelompokan usaha mikro Kecil dan menengah di Kabupaten Klaten tahun 2019 dengan metode *K-Means* dan *Clustering Large Application*. *Jurnal Statistika, Institut Sains dan Teknologi*. Vol 7, No. 1, hal 46-59.
- [8] M. A. Alwansyah, D. I. Fadli, D. S. Rini, M. A. Nabhan, and R. Rachmawati, "Grouping of Earthquakes in Sumatra Region Using Partitioning Around Medoids (PAM) and Clustering Large Applications (CLARA) Methods," AIP Conf. Proc., vol. 3285, no. 1, 2025.
- [9] Kaufman, L., dan Rousseeuw, P. J. 2005. *Finding Groups in Data : An Introduction to Cluster Analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- [10] Derisma, F. & Yusya, R. P. 2016. Perancangan Ikat Pinggang Elektronik Untuk Tunanetra Menggunakan Mikrokontroler Dan *Global Positioning System* (Gps) Pada *Smartphone* Android. *Jurnal Teknik Elektro ITP*, 5(2), 130-136.
- [11] M. A. Alwansyah, N. Hidayat, D. I. Fadli, D. S. Rini, and P. Jana, "Spatial zone analysis of earthquake in Sumatra," Proceeding 7Th Int. Conf. Sci. Technol. Interdiscip. Res. (Ic-Star 2021), vol. 2601, p. 020009, 2023.
- [12] Aini, F.N., Palgunadi. S., dan Anggrainingsih. R. 2014. *Clustering* Business Recess Model Petri Net dengan *Complete Linkage*. *Jurnal ITSMART*. Vol 3. No. 2: Hal. 47-51.
- [13] Irwansyah, E., dan Faisal, M. 2015. *Advanced Clustering* Teori dan Aplikasi. Yogyakarta: Deepublish.