Clustering Provinces with Drought Risk Based on Daily Maximum Temperature

A'yunin Sofro[1,*], Elok Rizqi Auliya[1], Khusnia Nurul Khikmah[1], Danang Ariyanto[1], Riska Wahyu Romadhonia[1], and Hernando Ombao[2]

[1]Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya, East Java, 60231 Indonesia
[2]Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
Email: ayuninsofro@unesa.ac.id (A.S.); elok.rizqi08@gmail.com (E.R.A.); khusniank@gmail.com (K.N.K.); danangariyanto@unesa.ac.id (D.A.); riskaromadhonia@unesa.ac.id (R.W.R.); hernando.ombao@kaust.edu.sa (H.O.)
*Corresponding author

Manuscript received September 20, 2023; revised November 13, 2023; accepted December 21, 2023; published March 22, 2024

***Abstract*—Changes in global weather patterns are sweeping the world, including Indonesia. One of the causes of this change was the El Niño event, where sea surface temperatures in the central Pacific Ocean experienced an increase. Apart from causing temperatures to increase, it also causes the intensity of rainfall to decrease, causing drought disasters. Anticipating natural disasters and disaster mitigation needs to be carried out to reduce their negative impacts. Efforts can be made by identifying areas with a high potential for drought and clustering areas based on the level of potential drought. This article focuses on extreme data from maximum temperatures in 34 provinces in Indonesia. Clustering was performed using the k-means and k-medoids methods and evaluated using the Davies-Bouldin index. Predict the highest maximum temperature in a specific period using the return level. The result shows that the k-means method is more suitable and better implemented by checking on the Davies-Boulding index, which is 0.9945.**

***Keywords*—clustering, drought, k-means, k-medoids, maximum temperature**

I. Introduction

The spike in daily maximum temperatures that occurred in several regions in Indonesia caused the air to become hot. The Meteorology, Climatology and Geophysics Agency (BMKG) has confirmed that this phenomenon is not a heatwave like in several Asian countries, which reported maximum temperatures exceeding 40 Celsius. The highest daily maximum temperature in Indonesia reached 37.2 Celsius. Many factors affect the increase in daily maximum temperature, one of which is the sun's trajectory because the sun's apparent position above Indonesia causes relatively more solar radiation received so that the temperature will increase during the day [1]. In addition, the El Niño problem also influences this because when El Niño occurs, sea surface temperatures warm above average conditions in the central Pacific Ocean [2]. As a result of this warming of the sea surface, cloud growth in the Pacific Ocean increases, and the Indonesian region will experience a reduction in rainfall, which triggers drought in Indonesia. High maximum temperatures will increase evaporation rates [3] and less water on the surface, so the chances of drought will be higher [4]. Drought causes many losses, one of which is the reduced level of water availability [5]. If this continues to happen, it will reduce the availability of water. Continued deterioration of water quality will make it challenging to obtain adequate water to fulfil daily needs, especially for consumption, and will affect health.

Prolonged drought can increase temperatures. High temperatures will cause leaves, twigs, and wood to dry out, triggering fires [6, 7]. In addition, drought can also cause water shortages [8]. Drought will disrupt the natural ecosystem over time if it is not addressed immediately. Therefore, disaster mitigation efforts are needed to minimise the negative impacts. One of the efforts that can be made is to group provinces in Indonesia based on similar characteristics. With this grouping, provinces that have a high risk can prepare and improve appropriate disaster mitigation efforts to reduce the negative impacts that will be caused. Based on this, researchers are interested in clustering the drought disaster risk in Indonesian provinces based on daily maximum temperature.

This study uses daily maximum temperature data, which is time series data. In addition, high daily maximum temperatures are classified as extreme data because they are rare events that significantly impact them. The best approach that can be used for this daily maximum temperature data is extreme value theory using block maxima. Previous research on extreme value theory using the Block Maxima approach was conducted by [9], which predicted deaths caused by pneumonia and influenza. Prediction of head-on collisions when maneuvering using the Block Maxima approach conducted by [10]. Research in [11] analyses the factors that cause land degradation and the risk of desertification using geographic information systems and the triangular fuzzy numbers (TFNs) method. In general, the methods used for clustering that be used in this study are divided into Hierarchical Clustering and Non-Hierarchical Clustering. Non-hierarchical clustering methods often used are k-means and k-medoids because they are simple and fast enough to be efficient. Previous research related to clustering using k-means and k-medoids was conducted by [12], who conducted Indonesian text summarisation, analysis algorithms for big data conducted by [13], and student segmentation based on scholastic achievement conducted by [14].

Based on the background, appropriate and integrated disaster mitigation efforts are needed, especially in high-risk provinces. One such effort is to cluster provinces in Indonesia with drought risk levels based on daily maximum temperatures in Indonesia. Clustering drought risk levels based on daily maximum temperature distribution in each province has not been much researched. In addition, Indonesia is an archipelago with differences in land height that cause different maximum temperatures for each area. Therefore, this research provides an understanding of the utilisation of clustering to support policymakers related to drought management in each province in Indonesia. The tools used to analyse the data in this research also provide extreme value that not too much research provides. In addition, this research is also expected to be one of the sources of information for the community in anticipating drought-related disasters that may occur in each province in Indonesia.

II. Literature Review

A. Extreme Value Theory (EVT)

Extreme Value Theory is a branch of statistics that deals with phenomena or extreme deviations from the median probability distribution. Extreme deviations include infrequent events with an unexpected and significant impact, such as earthquakes. This EVT is important to study because it is indispensable to know extreme phenomena that might occur in the future and can be used as a guide in taking appropriate anticipatory actions. EVT has been widely applied; for example, in the financial sector, it is used to analyze insurance risk; in the climate sector, it is used to predict drought; and it is even used in the health sector to be applied in various epidemiologies [15].

EVT has an approach method that can be applied as Block Maxima (BM) and Peak Over Threshold (POT). Each approach has its way of identifying each extreme value and has a different distribution.

B. Block Maxima (BM)

Block Maxima is one approach to identifying extreme values in a certain period by taking the maximum value from observational data. The data is divided into blocks according to a predetermined period. Periods can be daily, weekly, monthly, quarterly, or yearly. The highest value of each block is called the extreme value. The distribution used by the BM method is Generalized Extreme Value (GEV). Distribution Generalized Extreme Value according to the Fisher and Tippet limit theorem. The GEV distribution includes continuous probability distributions for combining the Gumbel, Fréchet, and Weibull distributions, also known as Type I, II, and III extreme value distributions developed in extreme value theory. Here are the Cumulative Distribution Function (CDF) from GEV.

$$F(x;\mu,\sigma,\xi)=\begin{cases} exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, & \xi\neq 0 \\ exp\left(-exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right), & \xi=0 \end{cases} \quad (1)$$

defined $\left\{x: 1+\xi\left(\frac{x-\mu}{\sigma}\right)>0\right\}$, where $-\infty<\mu<\infty, \sigma>0$ and $-\infty<\xi<\infty$.

whereas the Probability Density Function (PDF) from GEV is as follows.

$$f(x;\mu,\sigma,\xi)= \begin{cases} \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \xi\neq 0 \\ \frac{1}{\sigma} exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right] exp\left(-exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right), \xi=0 \end{cases} \quad (2)$$

In the GEV distribution, there are three parameters, namely location (μ), scale (σ), dan shape (ξ). The location parameter determines the center of the distribution. The scale parameter determines the size of the deviation around the location parameter. While the shape parameter shows the behavior of the tail of the distribution, the greater the value, the fatter the tail will be, which means that the greater the chance of extreme values occurring. The form parameter describes three types of extreme values, namely Gumbel or the distribution of Type I extreme values if the form parameter values are equal to zero ($\xi=0$), Frechet or the distribution of Type II extreme values if the form parameter values are more significant than zero ($\xi>0$), and Weibull or the type extreme value distribution III if the shape parameter value is less than zero ($\xi<0$).

Maximum Likelihood Estimation (MLE) by maximizing the probability function is used to estimate the parameters of the GEV. This is the MLE from GEV.

- For $\xi\neq 0$, Likelihood function is as follow.

$$L(f|x_i;\mu,\sigma,\xi)=\left(\frac{1}{\sigma}\right)^n\prod_{i=1}^n\left\{\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}\right\} exp\left\{-\prod_{i=1}^n\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \quad (3)$$

then log likelihood is as follow.

$$\ln L(f|x_i;\mu,\sigma,\xi)=-n\ln(\sigma)+\sum_{i=1}^n\left[\left(\frac{1}{\xi}-1\right)ln\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]-\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right] \quad (4)$$

Then maximize the likelihood function by deriving against the estimated parameters and equating to zero.

- for $\xi=0$, likelihood function is as follow.

$$L(f|x_i;\mu,\sigma,\xi) = \left(\frac{1}{\sigma}\right)^n exp\left[-\prod_{i=1}^n\left(\frac{x_i-\mu}{\sigma}\right)\right] exp\left(-\prod_{i=1}^n exp\left[-\left(\frac{x_i-\mu}{\sigma}\right)\right]\right) \quad (5)$$

then log likelihood is as follow.

$$\ln L(f|x_i;\mu,\sigma,\xi)=-n\ln(\sigma)-\sum_{i=1}^n\left(\frac{x_i-\mu}{\sigma}\right)-\sum_{i=1}^n exp\left[-\left(\frac{x_i-\mu}{\sigma}\right)\right] \quad (6)$$

Then maximize the function likelihood by deriving against the estimated parameters and equating to zero.

C. K-Means

K-means is a clustering method by dividing n object into k clusters that have been determined so that in one cluster, the similarity rate between members is very high, and with other clusters, the similarity rate is meager. The advantage of k-means is that it is simple and can cluster quickly, making it more efficient [16]. The k-means algorithm is as follows.

1) Determine the number of clusters you want to form.
2) Choose the centroid point randomly, k.
3) Calculates the distance of each object to the centroid. To calculate the distance, you can use Euclidean Distance. The formula is as follows.

$$d_{Euclidean}(x.y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (7)$$

where x is an object and y are centroids.

4) Enter each object into the cluster with the closest distance.
5) Update the centroid point by calculating the average of all objects in each cluster. To calculate the new centroid, you can use the following formula.

$$C_k = \frac{1}{N}\sum_{j=1}^{N} x_j \quad (8)$$

where C_k is the cluster i, N is the number of objects in one cluster.

6) Return to Step 3 until there is no change in the centroid.

D. K-Medoids

K-medoids, known as Partitioning Arounds Medoids (PAM), is one of the partitioning methods because it uses the medoid (median) as the center point of a cluster [17]. The advantage of k medoids is that the clustering results do not depend on the order of the data [18].

The K Medoids algorithm is as follows.

1) Determine the number of clusters formed.
2) Determines the medoid as much as k.
3) Calculates the distance of each object to the medoid. Euclidean distance can be calculated as in Eq. (7).
4) Choose a new medoid candidate from objects in each cluster randomly.
5) Calculates the distance of each object to the new medoid in each cluster.
6) The total deviation (S) is calculated by reducing the new total distance from the old total distance. If $S > 0$, then the clustering process is completed, cluster members are obtained from each medoid. If $S < 0$, then form a group k new object as medoid by changing the object with cluster data.

Return to Stage 3 until there is no change in the medoid.

E. Davies-Bouldin Index

Davies-Bouldin Index (DBI) is used to evaluate the clustering algorithm based on cohesion and separation values [19]. DBI was introduced by David L. Davies and Donald W. Bouldin in 1979. DBI is used to find out how well the clustering scheme is. The smaller the DBI value obtained, the better the clustering results [20]. The DBI value is obtained from the ratio value obtained from the following equation.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i \neq j}(R_{i,j}) \quad (9)$$

where k is the number of clusters formed, $R_{i,j}$ is the value of the ratio. To find the value of the ratio, you can use the following equation.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (10)$$

SSW is cohesion in the clusters formed, and SSB is the cluster separation. To know the *Sum of Square Within* (SSW) can use the equation formulated as follows.

$$SSW_i = \frac{1}{m_i}\sum_{j=1}^{m_i} d(x_j, c_i) \quad (11)$$

where m_i is the amount of data in the cluster i, c_i is the centroid of the cluster i, whereas d is the distance of each object to the centroid. The *Sum of Square Between* (SSB) can use the following equation.

$$SSB_{i,j} = d(c_i, c_j) \quad (12)$$

F. Return Level

The return Level is the maximum value expected to be exceeded once within a certain period [21]. The return level at a particular time is derived from the GEV distribution, expressed in the following equation.

$$z_p = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \{-ln(1-p)\}^{-\hat{\xi}}\right] \quad (13)$$

where p is an opportunity in a given year.

III. Materials and Methods

A. Data Source

In this study, the data used were secondary data taken from the official website of the BMKG Database Center Online Data, namely data online.bmkg.go.id. The data used is daily maximum temperature data in degrees Celsius units taken from 2015 to 2022 in 34 Provinces in Indonesia. The Provinces are Bali, Banten, Bengkulu, DI Yogyakarta, DKI Jakarta, Gorontalo, Jambi, West Java, Central Java, East Java, West Kalimantan, South Kalimantan, Central Kalimantan, East Kalimantan, North Kalimantan, Bangka Belitung Islands, Riau Islands, Lampung, Maluku, North Maluku, Nangroe Aceh Darussalam, West Nusa Tenggara, East Nusa Tenggara, Papua, West Papua, Riau, West Sulawesi, South Sulawesi, Central Sulawesi, Southeast Sulawesi, North Sulawesi, West Sumatra, South Sumatra and North Sumatra.

In general, the daily maximum temperature data used in this research has the characteristics provided in the Table 1.

Table 1. Characteristics of temperature time series data

Descriptive Statistics	**Location (μ)**
Mean	34.44
Median	34.50
Mode	35.00
Standard Deviation	1.21
Variance	1.46
Range	8.30
Minimum	30.50
Maximum	38.80

B. Research Flowchart

The flow chart in this study is as follows.

Based on Fig. 1, the first step is to collect daily maximum temperature data from 34 Provinces in Indonesia. Then for each Province, the extreme values are taken using the EVT method with the BM approach by dividing the data into blocks based on quarterly periods. The extreme values obtained are then estimated using the MLE parameter by following the GEV distribution. After that, Province clustering was carried out using the Non-Hierarchical Clustering namely the K Means and K Medoids methods. To determine the best clustering method using Davies-Bouldin Index. Then determine the return level to show the greatest maximum temperature prediction value of each Province.

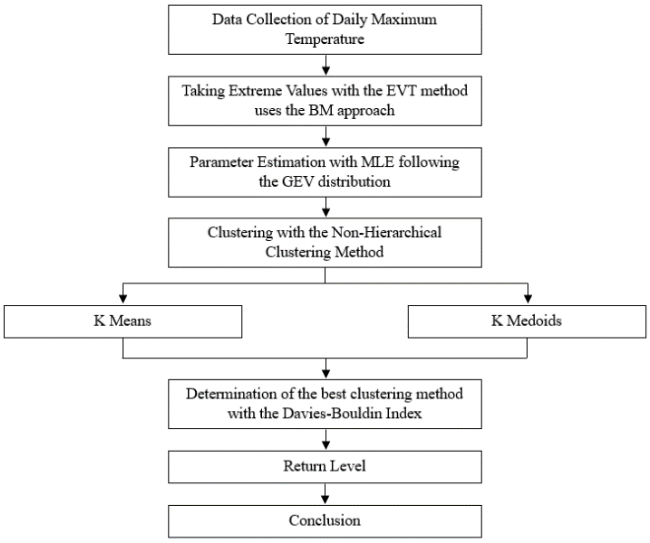

Fig. 1. Research flowchart.

IV. Result and Discussion

This section will explain how to take extreme values from daily maximum temperatures and cluster Provinces with drought risk. method extreme value theory with block maxima approach used in determining extreme values and clustering using the k-means and k-medoids methods. The Davies-Bouldin index determines the clustering method that produces the best clusters. Determine maximum temperature predictions using return levels for the next 25-year, 50-year, and 100-year.

A. Extreme Value Theory using Block Maxima approach

Maximum temperature data is not used as a whole, but only extreme data is taken from each Province by taking the highest value quarterly. To obtain 32 data that will be used from each Province. The data will be estimated using maximum likelihood estimation (MLE) following the generalized extreme value (GEV) distribution. Parameter estimation values for each Province were obtained with the help of RStudio. For example, in the Province of Bali, the parameter estimation results are shown in Table 2.

Table 2. Parameter estimation results using MLE follow the GEV distribution in the Province of Bali

Province	Location (μ)	Scale (σ)	Form (ξ)
Bali	32.29773	1.1650035	−0.293143299

The estimated parameter values are substituted into Eq. (2), and the PDF is obtained as follows.

$$f(x;\mu,\sigma,\xi) = \frac{1}{1.1650035}\left[1 + (-0.293143299)\left(\frac{x_i - 32.29773}{1.1650035}\right)\right]^{-\frac{1}{(-0.293143299)}-1}$$

$$exp\left\{-\left[1+(-0.293143299)\left(\frac{x_i-32.29773}{1.1650035}\right)\right]^{-\frac{1}{(-0.293143299)}}\right\}, \xi \neq 0$$

By doing the same way, a PDF of the GEV will be obtained for each Province. The parameter estimation results for each Province can be seen in the Table 3.

Table 3. Parameter estimation results using MLE follow the GEV of 34 provinces.

No	Province	Location (μ)	Scale (σ)	Form (ξ)
1.	Bali	32.2977	1.1650	−0.2931
2.	Banten	34.3626	0.5988	−0.1561
3.	Bengkulu	33.9741	0.7586	−0.2770
4.	Yogyakarta	33.8368	0.7509	−0.2686
5.	DKI Jakarta	34.8874	0.3465	−0.6220
6.	Gorontalo	34.9104	0.6809	−0.3750
7.	Jambi	34.5898	0.5288	−0.2961
8.	West Java	31.7959	0.6204	0.1617
9.	Central Java	34.7341	1.4310	−0.2358
10.	East Java	35.9384	0.5477	0.1177
11.	West Kalimantan	34.8474	0.4924	−0.4260
12.	South Kalimantan	34.6088	0.6734	0.1039
13.	Central Kalimantan	34.9463	0.3680	0.0034
14.	East Kalimantan	33.3000	0.7341	−0.2196
15.	North Kalimantan	33.8602	0.4680	−0.2557
16.	Bangka Belitung Islands	33.8564	0.6077	−0.4345
17.	Riau Islands	34.0132	0.6090	−0.5869
18.	Lampung	34.3790	0.6507	0.1558
19.	Maluku	32.7563	0.9698	−0.2765
20.	North Maluku	33.1964	0.5410	−0.1121
21.	Nangroe Aceh Darussalam	34.5150	1.2594	−0.2978
22.	West Nusa Tenggara	35.3362	1.3316	−0.3438
23.	East Nusa Tenggara	34.6990	0.8488	−0.1936
24.	Papua	34.5255	0.5641	−0.1444
25.	West Papua	33.4589	0.9050	−0.1780
26.	Riau	35.1755	0.6121	−0.2553
27.	West Sulawesi	33.9028	0.7178	−0.3468
28.	South Sulawesi	34.2274	0.4727	−0.4325
29.	Central Sulawesi	34.1853	0.9138	−0.2024
30.	Southeast Sulawesi	34.1425	1.2575	−1.0000
31.	North Sulawesi	33.9751	0.8066	−0.0221
32.	West Sumatra	32.6356	0.4312	0.2244
33.	South Sumatra	34.5693	0.5375	0.0684
34.	North Sumatra	34.3014	1.1783	−0.1811

Based on Table 2, the location parameters indicate the location of the maximum temperature center point in each Province; for example, the Province of Bali has a maximum temperature center point of 32.29773 degrees Celsius. While the scale parameter shows the maximum temperature variation in Province, for example, Bali Province has a maximum temperature variation of 1.1650035 degrees Celsius. Based on the shape parameters, Provinces follow the Weilbull distribution, namely Bali, Banten, Bengkulu, DI Yogyakarta, DKI Jakarta, Gorontalo, Jambi, Central Java, West Kalimantan, East Kalimantan, North Kalimantan, Bangka Belitung Islands, Riau Islands, Maluku, North Maluku, Nangroe Aceh Darussalam, West Nusa Tenggara, Nusa Tenggara East Southeast, Papua, West Papua, Riau, West Sulawesi, South Sulawesi, Central Sulawesi, Southeast Sulawesi, North Sulawesi, and North Sumatra. This is because the shape parameter's value is negative or $\xi < 0$.

Meanwhile, the Province follows the Frechet distribution and has positive form parameter values or $\xi > 0$, namely DKI Jakarta, West Java, East Java, South Kalimantan, Central Kalimantan, Lampung, South Sulawesi, Southeast Sulawesi, North Sulawesi, West Sumatra, and South Sumatra. Because of Frechet distribution, the chance of occurrence of extreme values is more significant.

B. K-Means

The parameter estimation result data from block maxima maximum temperature is then clustered using the k-means method. You can use the Elbow method to determine the optimal number of clusters to form. The plot of the results is shown in Fig. 2.

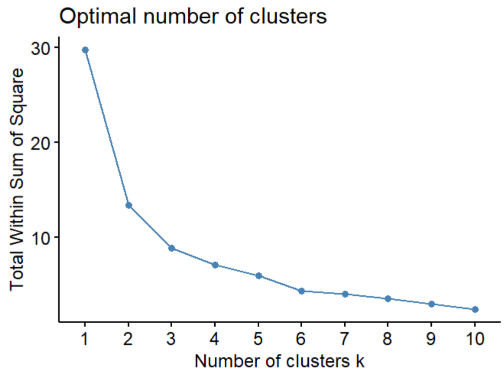

Fig. 2. The optimal cluster plot of the k-means method with the elbow method.

Based on the picture above, the optimal number of clusters that will be formed is $k = 3$, so the plot of the clustering results is shown in Fig. 3.

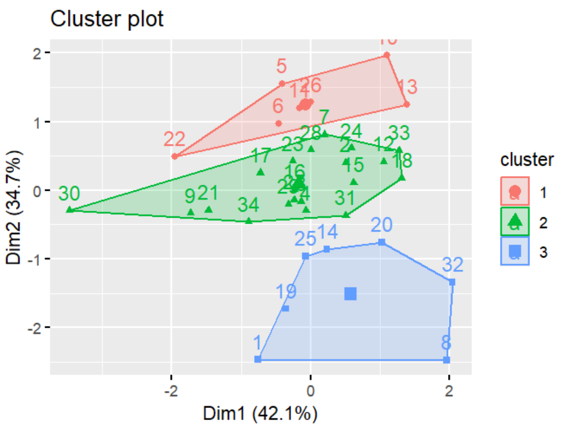

Fig. 3. The of clustering results using the k-means method.

Clustering results using k-means. Namely, Provinces in cluster 1 include DKI Jakarta, Gorontalo, East Java, West Kalimantan, Central Kalimantan, West Nusa Tenggara, and Riau. Provinces in Cluster 2, namely Banten, Bengkulu, DI Yogyakarta, Jambi, Central Java, South Kalimantan, North Kalimantan, Bangka Belitung Islands, Riau Islands, Lampung, Nangroe Aceh Darussalam, East Nusa Tenggara, Papua, West Sulawesi, South Sulawesi, Central Sulawesi, Southeast Sulawesi, North Sulawesi, South Sumatra, and North Sumatra. While the Provinces in Cluster 3 include Bali, West Java, East Kalimantan, Maluku, North Maluku, West Papua, and West Sumatra. The number of members of each cluster can be seen in Table 4.

Table 4. Number of members of each cluster using the k-means method

Cluster	Number of members	Percentage (%)
1	7	20.6
2	20	58.8
3	7	20.6

While the centroid values of each cluster using the k-means method are presented in Table 5.

Table 5. Centroid values for each cluster using the k-means method

Cluster	Location (μ)	Scale (σ)	Form (ξ)
1	35.1488	0.6256	−0.2716
2	34.2629	0.7817	−0.2502
3	32.7871	0.7749	−0.1090

Based on Table 4, the centroid value of Cluster 1 with the average location of the maximum temperature center point is 35.14881 degrees Celsius, and the maximum temperature variation is 0.6256139 degrees Celsius. The centroid value of Cluster 2 with the average location of the maximum temperature center point is 34.26292 degrees Celsius, and the maximum temperature variation is 0.7816592 degrees Celsius. While the centroid value of Cluster 3 with the average location of the maximum temperature center point is 32.78705 degrees Celsius, and the maximum temperature variation is 0.7748633 degrees Celsius. This shows that Cluster 1 is a Province with a high probability of drought, Cluster 2 indicates a Province with a moderate probability of drought, and Cluster 3 indicates a Province with a low probability of drought.

C. K-Medoids

Data from block maxima's maximum temperature results were clustered using the k-medoids method. The Elbow method can be used to determine the optimal number of clusters formed. The plot of the results is shown in Fig. 4.

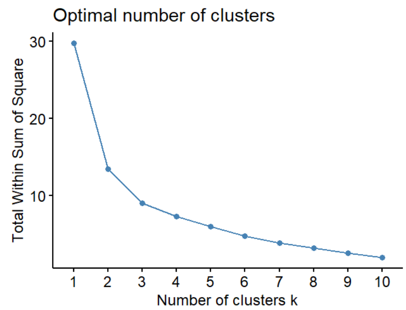

Fig. 4. The optimal cluster plot of the k-medoids method with the Elbow method.

Based on the picture above, the optimal number of clusters formed is $k = 3$. So, the plot of the clustering results is shown in Fig. 5.

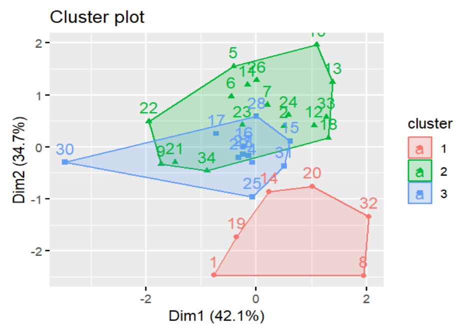

Fig. 5. The plot of clustering results with the k-medoids method.

Clustering results using k-medoids, namely, Provinces in Cluster 1 include Bali, West Java, East Kalimantan, Maluku, North Maluku, and West Sumatra. Provinces in Cluster 2 are Banten, DKI Jakarta, Gorontalo, Jambi, Central Java, East Java, West Kalimantan, South Kalimantan, Central Kalimantan, Lampung, Nangroe Aceh Darussalam, West Nusa Tenggara, East Nusa Tenggara, Papua, Riau, South Sumatra, and North Sumatra. While the Provinces in Cluster 3 include Bengkulu, DI Yogyakarta, North Kalimantan, Bangka Belitung Islands, Riau Islands, West Papua, West Sulawesi, South Sulawesi, Central Sulawesi, Southeast Sulawesi, and North Sulawesi. The number of Provinces in each cluster can be seen in Table 6.

Table 6. The number of members in each cluster using the k-medoids method

Cluster	Number of members	Percentage (%)
1	6	17.6
2	17	50
3	11	32.4

While the medoid values of each cluster using the k-medoids method are presented in Table 7.

Table 7. Medoid values for each cluster using the k-medoids method

Cluster	Location (μ)	Scale (σ)	Form (ξ)
1	32.7563	0.9698	−0.2765
2	34.6990	0.8488	−0.1936
3	33.9028	0.7178	−0.3468

Based on Table 6, the medoid value of cluster 1 with an average location of the maximum temperature center point of 32.75625 degrees Celsius and the maximum temperature variation is 0.9698366 degrees Celcius. The medoid value of Cluster 2 with an average maximum temperature center point location of 34.69899 degrees Celsius, and the maximum temperature variation is 0.8488470 degrees Celcius. The medoid value of Cluster 3 with an average maximum temperature center point location of 33.90279 degrees Celsius, and the maximum temperature variation is 0.7177741 degrees Celcius. This indicates that Cluster 1 is a Province with a low risk of drought, Cluster 2 is a Province with a high risk, and Cluster 3 is a Province with moderate risk.

D. Davies-Bouldin Index

To find the best clustering scheme using the k-means and k-medoids methods, you can use the Davies-Bouldin index to find the best clustering scheme between clustering using the k-means and k-medoids methods. The DBI value of each method can be seen in Table 7.

Based on Table 8, the DBI value for the k-means method is lower than the DB value for the k-medoids method. A lower DBI value indicates a better clustering scheme, so the k-means method is more suitable and better implemented in this study. Based on the validation test carried out, the visualization of the clustering results obtained is presented in Fig. 6.

Table 8. Davies-Bouldin index value

Clustering method	Davies-Boulding index
k-means	0.9945
k-medoids	1.0807

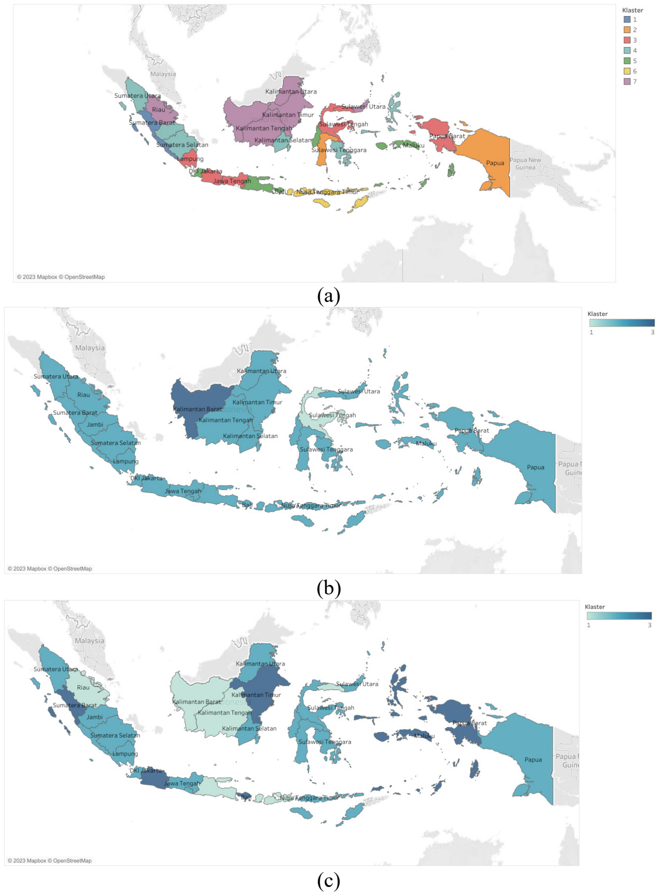

Fig. 6. Maps plot of the best clustering result. (a). Results of k-means clustering block maxima rainfall. (b). Results of k-means clustering block minima maximum temperature (c). Results of k-means clustering block maxima maximum temperature.

E. Return Level

After getting the estimation results. The data is used to calculate each province's return level maximum temperature. Return level data for the highest maximum temperatures at 25-year, 50-year, and 100-year are presented in Table 9.

The Table 9 shows the prediction of the lowest maximum temperature in each province. For example, in the province of Bali, the maximum temperature in the 25 years is predicted to be 34.71580 degrees Celsius. In 50 years, it is predicted to be 35.00575 degrees Celsius. In 100 years, it is predicted to be 35.24010 degrees Celsius.

Table 9. Return level highest maximum temperature of 34 provinces

No	Province	Return level		
		25-year	50-year	100-year
1.	Bali	34.7158	35.0058	35.2401
2.	Banten	35.8703	36.1124	36.3279
3.	Bengkulu	35.5835	35.7834	35.9468
4.	Yogyakarta	35.4484	35.6523	35.8199
5.	DKI Jakarta	35.3682	35.3952	35.4125
6.	Gorontalo	36.1790	36.3059	36.4028
7.	Jambi	35.6830	35.8132	35.9183
8.	West Java	34.3947	35.1699	36.0317
9.	Central Java	37.9480	38.3843	38.7514
10.	East Java	38.0658	38.6511	39.2820
11.	West Kalimantan	35.7074	35.7840	35.8404
12.	South Kalimantan	37.1639	37.8491	38.5806
13.	Central Kalimantan	36.1300	36.3920	36.6527
14.	East Kalimantan	34.9868	35.2238	35.4255
15.	North Kalimantan	34.8828	35.0158	35.1262

No	Province	Return level		
		25-year	50-year	100-year
16.	Bangka Belitung Islands	34.9066	34.9983	35.0655
17.	Riau Islands	34.8921	34.9458	34.9812
18.	Lampung	37.0766	37.8727	38.7540
19.	Maluku	34.8154	35.0715	35.2809
20.	North Maluku	34.6505	34.9061	35.1407
21.	Nangroe Aceh Darussalam	37.1125	37.4208	37.6691
22.	West Nusa Tenggara	37.9197	38.1967	38.4128
23.	East Nusa Tenggara	36.7232	37.0238	37.2843
24.	Papua	35.9705	36.2082	36.4214
25.	West Papua	35.6660	36.0046	36.3013
26.	Riau	36.5136	36.6878	36.8323
27.	West Sulawesi	35.2898	35.4376	35.5526
28.	South Sulawesi	35.0463	35.1181	35.1708
29.	Central Sulawesi	36.3371	36.6507	36.9208
30.	Southeast Sulawesi	35.3487	35.3746	35.3874
31.	North Sulawesi	36.4659	36.9904	37.5030
32.	West Sumatra	34.7260	35.3213	35.9799
33.	South Sumatra	36.4910	36.9730	37.4749
34.	North Sumatra	37.1621	37.5981	37.9793

V. Discussion of the Disaster Risk in Indonesia

This study provides the results of clustering provinces in Indonesia with drought risk based on their maximum temperatures. There are 34 points in Indonesia which are used as point references. This research shows that Provinces with drought risk based on maximum temperature are divided into three Provinces with high, medium, and low drought risk. West Papua Province is a province that has a low risk of drought. This is due to the intensity of rain each month in the province when the dry season is still moderate, so the maximum temperature is lower than in other Provinces. Because of this, the risk of drought is low. However, this condition is inversely proportional to East Java Province, which has the highest drought risk. The results of this study support previous research, which stated that East Java Province has a very high average level of drought [22]. Land misuse is a plausible cause that puts East Java Province at high drought risk. One of the areas in East Java Province where land abuse is often found is Mojokerto. Many lands have been converted into residential areas, the development of tourist attractions, and even the conversion of forest land to agriculture.

In addition, previous research on [11] found that temperature is one factor that caused land degradation by using GIS with triangular fuzzy numbers. This research provided more advantages to the analysis approach by using extreme value. Focuses on extreme value can be a reference for indications of disaster occurrence. Clustering is based on the similarity of extreme value characters. Then, the clustering result can be visualized more and more manageably for the same province in a cluster. Therefore, the Indonesian government needs to consider the steps and efforts to be taken to prevent and overcome drought disasters so that they do not cause more negative impacts. This study used several clustering methods to get the best clustering results. The clustering method with the k-means method gives the best clustering results in clustering provinces with drought risk based on their maximum temperature. Besides that, the results of the return level can also provide maximum temperature prediction information in the future. It is hoped that this will become a consideration for the government in making appropriate policies to deal with this drought disaster.

VI. Conclusion

In clustering Provinces with drought risk based on maximum temperature, Davies-Bouldin index values from the k-means method are more diminutive than k-medoids. This indicates that the k-means clustering scheme is better. The optimal clusters that are formed are:

- as many as 3 clusters with Cluster 1, namely Provinces with a high probability of drought, there are as many as 7 Provinces;
- Cluster 2, namely Provinces with a moderate possibility of drought, there are 20 Provinces and
- Cluster 3, namely, Provinces with a low probability of drought, has as many as 7 Provinces.

The results of this study are likely helpful for related parties in Indonesia in carrying out efforts to prevent and control drought disasters. In future studies, the clustering method can use other clustering methods and add other influential factors expected to produce even more accurate results.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

AS and ERA have conducted the research, analyzed the data, and wrote the paper; KNK revised the manuscript; DA revised the manuscript; RWR revised the manuscript; HO research consultant; all authors had approved the final version

References

[1] V. A. Tartakovsky, N. N. Cheredko, V. G. Maximov, and Y. V. Volkov, "How does the Sun affect the surface temperature?" *IOP Conf Ser Earth Environ Sci*, vol. 211, no. 1, 2018. doi: 10.1088/1755-1315/211/1/012002

[2] D. R. Thielen *et al.*, "Effect of extreme El Niño events on the precipitation of Ecuador," *Natural Hazards and Earth System Sciences*, vol. 23, no. 4, pp. 1507–1527, 2023. doi: 10.5194/nhess-23-1507-2023

[3] D. F. Balting, A. AghaKouchak, G. Lohmann, and M. Ionita, "Northern Hemisphere drought risk in a warming climate," *NPJ Clim Atmos Sci*, vol. 4, no. 1, pp. 1–13, 2021. doi: 10.1038/s41612-021-00218-2

[4] C. Li, B. Adu, J. Wu, G. Qin, H. Li, and Y. Han, "Spatial and temporal variations of drought in Sichuan Province from 2001 to 2020 based on modified temperature vegetation dryness index (TVDI)," *Ecol Indic*, vol. 139, no. March, p. 108883, 2022. doi: 10.1016/j.ecolind.2022.108883

[5] C. Payus *et al.*, "Impact of extreme drought climate on water security in North Borneo: Case study of Sabah," *Water (Switzerland)*, vol. 12, no. 4, pp. 1–19, 2020. doi: 10.3390/W12041135

[6] B. A. Cassell, R. M. Scheller, M. S. Lucash, M. D. Hurteau, and E. L. Loudermilk, "Widespread severe wildfires under climate change lead to increased forest homogeneity in dry mixed-conifer forests," *Ecosphere*, vol. 10, no. 11, 2019. doi: 10.1002/ecs2.2934

[7] C. E. Reid and M. M. Maestas, "Wildfire smoke exposure under climate change," *Curr Opin Pulm Med*, vol. 25, no. 2, pp. 179–187, 2019. doi: 10.1097/mcp.0000000000000552

[8] M. D. Peña-Guerrero, A. Nauditt, C. Muñoz-Robles, L. Ribbe, and F. Meza, "Drought impacts on water quality and potential implications for agricultural production in the Maipo River Basin, Central Chile,"

Hydrological Sciences Journal, vol. 65, no. 6, pp. 1005–1021, 2020. doi: 10.1080/02626667.2020.1711911

[9] M. Thomas *et al.*, "Applications of extreme value theory in public health," *PLoS One*, vol. 11, no. 7, pp. 3–9, 2016. doi: 10.1371/journal.pone.0159312

[10] C. L. Azevedo and H. Farah, "Using Extreme Value Theory for the Prediction of Head-On Collisions during Passing Maneuvres," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2015-Octob, pp. 268–273, 2015. doi: 10.1109/ITSC.2015.53

[11] S. Wijitkosum, "Factor influencing land degradation sensitivity and desertification in a drought prone watershed in Thailand," *International Soil and Water Conservation Research*, vol. 9, no. 2, pp. 217–228, 2021.

[12] K. K. Purnamasari, "K-Means and K-Medoids for Indonesian Text Summarization," *IOP Conf Ser Mater Sci Eng*, vol. 662, no. 6, 2019. doi: 10.1088/1757-899X/662/6/062013

[13] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm for Big Data," *Phys Procedia*, vol. 78, no. December 2015, pp. 507–512, 2016. doi: 10.1016/j.procs.2016.02.095

[14] U. Badhera, A. Verma, and P. Nahar, "Applicability of K-medoids and K-means algorithms for segmenting students based on their scholastic performance," *Journal of Statistics and Management Systems*, vol. 25, no. 7, pp. 1621–1632, 2022. doi: 10.1080/09720510.2022.2130566

[15] W. Cao and Z. Zhang, "New extreme value theory for maxima of maxima," *Stat Theory Relat Fields*, vol. 5, no. 3, pp. 232–252, 2021. doi: 10.1080/24754269.2020.1846115

[16] I. Davidson, "Understanding K-means non-hierarchical clustering Understanding K-Means Non-hierarchical Clustering," *SUNY Albany-Techical Report*, vol. 02, no. 2, 2015.

[17] E. A. Maharaj, A. M. Alonso, and P. D'Urso, "Clustering seasonal time series using extreme value analysis: An application to Spanish temperature time series," *Commun Stat Case Stud Data Anal Appl*, vol. 1, no. 4, pp. 175–191, 2015. doi: 10.1080/23737484.2016.1179140

[18] N. Sureja, B. Chawda, and A. Vasant, "An improved K-medoids clustering approach based on the crow search algorithm," *Journal of Computational Mathematics and Data Science*, vol. 3, no. July 2021, p. 100034, 2022. doi: 10.1016/j.jcmds.2022.100034

[19] B. Jumadi Dehotman Sitompul, O. Salim Sitompul, and P. Sihombing, "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm," *J Phys Conf Ser*, vol. 1235, no. 1, pp. 6–12, 2019. doi: 10.1088/1742-6596/1235/1/012015

[20] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *J Intell Inf Syst*, vol. 17, no. 2–3, pp. 107–145, 2001. doi: 10.1023/A:1012801612483

[21] P. E. Kindermann *et al.*, "Return level analysis of the hanumante river using structured expert judgment: A reconstruction of historical water levels," *Water (Switzerland)*, vol. 12, no. 11, pp. 1–29, 2020. doi: 10.3390/w12113229

[22] D. Witarsih, "Pemetaan Tingkat Kekeringan Berdasarkan Parameter Indeks TVDI Data Citra Satelit Landsat-8 (Studi Kasus : Provinsi Jawa Timur)," *Jurnal Teknik ITS*, vol. 5, no. 2, 2016. doi: 10.12962/j23373539.v5i2.18118