SUDEIJ
Sustainable Urban Development and Environmental Impact Journal
SUDEIJ 3(1): 70–86
ISSN 3062-8997

Institute for Advanced Science, Social and Sustainable Future
MORALITY BEFORE KNOWLEDGE

Identification and clustering of drought-prone areas based on geographical, climatic, and socio-economic indicators: Supporting sustainable environmental management policies

Zainal Mu'arif[1,*]
[1] *Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Tadulako, Palu City, Central Sulawesi 94148, Indonesia.*
*Correspondence: zainalmuarifstory@gmail.com

Received Date: January 20, 2026 Revised Date: February 20, 2026 Accepted Date: February 24, 2026

ABSTRACT

Background: Drought is one of the most crucial environmental issues with widespread impacts across various regions of the world, including Indonesia, where changes in rainfall patterns and land use exacerbate the condition. This study aims to identify and cluster drought-prone areas in Central Sulawesi based on geographic, climatic, and socio-economic indicators, with 22 variables representing these three indicators. **Methods:** The analysis uses a quantitative approach based on data mining through the K-Means Clustering technique. Secondary data from 2019 to 2025 were integrated from multiple agencies, including BPS, BMKG, and BNPB, while the optimal number of clusters was determined using the Silhouette method executed via RStudio. **Findings:** The analysis results show three clusters with different levels of drought vulnerability, namely Cluster 1 (high drought), Cluster 2 (moderate drought), and Cluster 3 (low drought). Cluster 1 is characterized by high temperatures, low rainfall, and intensive mining activities. Cluster 2 has moderate rainfall and better environmental conditions. Cluster 3 shows relatively stable hydrological and socio-economic conditions. Climatic factors, particularly rainfall are the most influential indicators of drought vulnerability. Geographical factors such as irrigated areas and the extent of forests and water bodies also contribute, as do socioeconomic factors such as population density, poverty levels, and access to clean water. **Conclusion:** This analysis provides a spatial overview of the distribution of drought risk and serves as a scientific basis for policy formulation. The analysis then provides policy recommendations, including irrigation development and water conservation in moderately vulnerable areas, sustainable resource management in low-risk areas, and green economy development in safe areas to support sustainable environmental management. **Novelty/Originality of this article:** The novelty of this study lies in its integrated multidimensional approach, combining geographic, climatic, and socio-economic indicators through K-Means Clustering to map drought vulnerability in Central Sulawesi.

KEYWORDS: drought; clustering analysis; sustainable environmental management.

1. Introduction

Drought is recognized as one of the most widespread and persistent environmental disasters affecting many regions across the globe, particularly those with complex geographical structures and diverse climatic conditions (Orimoloye et al., 2022; Wang et al., 2025). It can be broadly defined as a prolonged period of abnormally low precipitation or a significant reduction in water supply and humidity compared to the average or

Cite This Article:
Mu'arif, Z. (2026). Identification and clustering of drought-prone areas based on geographical, climatic, and socio-economic indicators: Supporting sustainable environmental management policies. *Sustainable Urban Development and Environmental Impact Journal*, *3*(1), 70-86. https://doi.org/10.61511/sudeij.v3i1.2026.3150

expected levels during a specific time frame (Kuswanto et al., 2021). Unlike other natural disasters that occur suddenly, drought develops gradually, often making its onset difficult to detect and its impacts long-lasting (Minea & Albulescu, 2025; Mens et al., 2022). This slow progression, however, can lead to extensive ecological, agricultural, and socio-economic consequences once critical water thresholds are surpassed (Yang et al., 2023).

The increasing intensity and frequency of drought events in recent decades are strongly associated with global climate change, which has altered rainfall distribution, increased surface temperatures, and disrupted hydrological cycles (Savelli et al., 2022). Changes in these climatic patterns have resulted in reduced groundwater recharge, diminished river flow, and declining soil moisture, thereby amplifying the risks to agricultural systems and water resource management (Biswas et al., 2025). Consequently, regions dependent on rain-fed agriculture and natural water sources face greater exposure to crop failures, loss of livelihood, and food insecurity. These climate-induced drought dynamics highlight the urgent need for adaptive environmental management and sustainable land-use practices (Van Loon et al., 2024).

In Indonesia, drought poses a particularly pressing environmental challenge due to the country's archipelagic nature and vast ecosystem diversity. The phenomenon has far-reaching implications for national food security, ecosystem stability, and the socio-economic well-being of local communities (Akhmad et al., 2025; Juliannisa et al., 2025; Mu'arif et al., 2024). One region increasingly showing signs of vulnerability to drought is Central Sulawesi Province, where declining water availability, reduced agricultural productivity, and growing rural vulnerability have become evident indicators of environmental stress. These emerging conditions underscore the importance of identifying and mapping drought-prone areas through scientific approaches such as data-driven clustering analysis, which can support the development of effective drought mitigation and sustainable environmental management strategies.

According to Wijitkosum (2025), drought is a multidimensional phenomenon influenced by complex interactions between climatic, geographical, soil, and human activity factors. The mutually reinforcing relationship between climate variability and land use change has the potential to exacerbate drought risk, especially in tropical regions with high environmental pressure. The study also confirms that a combination of climatological factors (such as rainfall and drought index), edaphic factors (soil texture and drainage), and anthropogenic factors (land use and land cover) play a major role in determining a region's vulnerability to drought.

Although the phenomenon of drought in Central Sulawesi is becoming increasingly apparent, research that systematically integrates geographical, climatic, and socio-economic indicators to identify and classify drought-prone areas is still very limited. Most existing research in Indonesia still focuses on drought monitoring using a hydrological or meteorological approach alone, without considering the socio-economic factors that also influence the level of community vulnerability. This situation highlights the urgent need for a more comprehensive, data-driven analytical approach to understand the multidimensional aspects of vulnerability to drought (En-Nagre et al., 2024; Shim et al., 2021). Such an approach is expected to provide a strong scientific basis for supporting the formulation of sustainable, evidence-based environmental management policies.

This study will use a Data Mining approach and K-Means Clustering technique based on its ability to process large and complex data to find hidden patterns without requiring labels or initial assumptions about data structure (Doan et al., 2023; Awad & Hamad, 2022; Wani, 2024). This method is effective in environmental spatial studies because it can produce homogeneous regional groupings based on similarities in characteristics between variables (Roy et al., 2024; Doan et al., 2023). Compared to conventional statistical methods, clustering techniques in Data Mining are more adaptive to multidimensional data and have high accuracy in detecting patterns of interrelationships between indicators (En-Nagre et al., 2024; Faizan et al., 2020). In particular, K-Means Clustering was chosen because of its efficiency, ease of interpretation, and ability to produce centroids that represent the distinctive profiles of each regional group (Lim et al., 2021; Li et al., 2023). This allows researchers to assess the differences in drought vulnerability levels between clusters and determine the dominant factors causing them. Furthermore, this method is also suitable for the formulation of area-based policies, as the results can be visualized spatially to support decision-making at the regional level.

This study aims to identify and cluster drought-prone areas in Central Sulawesi Province using a quantitative approach based on data mining through the K-Means Clustering technique. By utilizing a set of geographical, climatic, and socio-economic indicators, this study seeks to reveal patterns and similarities in characteristics between regions based on their vulnerability to drought. The clustering results are expected to reveal the dominant factors that cause one region to be more vulnerable to drought than others. In addition, the findings of this study are expected to provide a scientific basis for the formulation of adaptive and sustainable environmental management strategies in Central Sulawesi, in order to increase the region's resilience to the impacts of climate change in the future.

2. Methods

2.1 Study area

Central Sulawesi Province, geographically located between 2°22′ North Latitude and 3°48′ South Latitude, and between 119°22′ and 124°22′ East Longitude, is one of the provinces in eastern Indonesia. Its strategic location along the equator plays a significant role in shaping the region's climate characteristics, ecological dynamics, and biodiversity patterns. As part of the Indonesian archipelago, the province boasts diverse topographic features, ranging from coastal lowlands to high mountains, contributing to its complex environmental and climatic conditions. The diversity of altitudes and landforms not only influences temperature variations and rainfall distribution but also determines the types of vegetation and agricultural activities that can thrive in each region (Mutia et al., 2023).

Central Sulawesi is characterized by a tropical climate heavily influenced by monsoon wind patterns, resulting in two distinct seasons throughout the year: the dry season, which generally occurs between May and October, and the rainy season, which typically lasts from November to April (Badan Metereologi Klimatologi Geofisika, 2025). This seasonal shift is accompanied by significant fluctuations in humidity, temperature, and rainfall intensity. The predominance of tropical climate conditions ensures relatively high humidity levels, with relatively stable temperatures, generally ranging between 25°C and 30°C throughout the year. This temperature consistency supports the growth of diverse ecosystems and agricultural activities, although it also increases the region's sensitivity to climate anomalies, such as prolonged dry seasons or unseasonal rainfall patterns.

In terms of rainfall, Central Sulawesi records an average annual rainfall ranging from 800 to 3,000 millimeters, a figure that reflects significant spatial variability across the province. Some inland and coastal areas experience moderate rainfall, while others, particularly those located under the mountain ranges, tend to receive much lower rainfall. Compared with most other provinces in Indonesia, Central Sulawesi is considered to have relatively lower rainfall levels, making it more vulnerable to the impacts of drought and fluctuations in water availability (Alfiandy & Permana, 2020). These climatic characteristics, combined with geographic diversity and human land-use activities, highlight the importance of understanding spatial variations in environmental conditions across the province to support effective water resource management, agricultural planning, and disaster risk reduction efforts (Wijitkosum, 2025).

Central Sulawesi Province has 13 districts/cities with an area of 61605,718 km^2 and a population of 3156,1 thousand at the end of 2024 (Badan Pusat Statistik Sulawesi Tengah, 2025). The study area has four meteorological stations that observe the climate. The topography of the study area is geographically diverse (lowlands, mountains, valleys), which affects rainfall distribution in various areas. In addition, the climate in the study area is influenced by global phenomena such as El Niño and La Niña, which affect rainfall patterns and temperatures from year to year (Bahtiar & Yulfiah, 2025).

2.2 Dataset

The data used in this study is secondary data obtained from the Central Sulawesi Statistics Agency (Badan Pusat Statistik Sulawesi Tengah, 2025), the Central Sulawesi Meteorology, Climatology and Geophysics Agency (Badan Metereologi Klimatologi Geofisika, 2025), the National Disaster Management Agency (Badan Nasional Penanggulangan Bencana, 2025) and the Central Sulawesi Energy and Mineral Resources Agency (Dinas Energi dan Sumber Daya Mineral Sulawesi Tengah, 2025). The data used is from the period 2019–2025, or the last 6 years. The research objects used are 13 regencies/cities in Central Sulawesi, as presented in the following Table 1.

Table 1. Total area of study

No	Regencies/City (areas)
1	Banggai
2	Banggai Kepulauan
3	Banggai Laut
4	Buol
5	Donggala
6	Kota Palu
7	Morowali
8	Morowali Utara
9	Parigi Moutong
10	Poso
11	Sigi
12	Tojo Una-Una
13	Tolitoli

The research variables used consist of three indicators that influence drought vulnerability, namely geographical indicators: describing natural conditions and landforms that influence drought risk spatially and hydrologically, climate indicators: describing atmospheric and weather conditions that directly influence drought events, and socio-economic indicators: describing the social and economic conditions of communities that influence vulnerability and adaptive capacity to drought. From these three indicators, a total of 22 variables is used and presented in Table 2.

Table 2. Indicator research variable

Variable	Code	Unit	Describe
Geographic			
Area size	X_1	km^2	The size of the administrative area affects the scale of risk distribution.
Number of villages	X_2	Total	Indicators of the level of regional fragmentation and the number of social units affected.
Elevation	X_3	MDPL	Determining potential rainfall and temperature (higher areas are usually cooler, rainfall varies).
Number of irrigation areas	X_4	Total	Surface water infrastructure that helps mitigate the effects of drought.
Total forest and water area	X_5	Total	Describing the ecological carrying capacity and natural capacity of the region to maintain water availability.
Climate			
History of drought	X_6	Total	Historical frequency of drought; key indicator of vulnerability.

History of extreme weather	X_7	Total	The frequency of climate disasters related to drought causes meteorological droughts.
History of forest and land fires	X_8	Total	Influencing the impact and exacerbating drought conditions
Temperature	X_9	Average (degrees celsius)	Affecting evaporation and water requirements.
Air humidity	X_{10}	Average (%)	Indicator of water vapor content in the atmosphere
Wind speed	X_{11}	Average (knot)	Influencing evaporation and moisture distribution.
Air pressure	X_{12}	Average (mbar)	Related to weather patterns and atmospheric circulation
Rainfall amount	X_{13}	Average (mm/year)	Key water availability indicators.
Number of rainy days	X_{14}	Average (day)	Shows the temporal distribution of rainfall (more detailed than total rainfall).
Duration of sunshine	X_{15}	Average (hour)	Affecting temperature and evapotranspiration.
Socio-Economic			
Population	X_{16}	Total	Pressure on water resources.
Population density	X_{17}	km^2	Population concentration in certain areas; affects social vulnerability.
Main source of drinking water (spring)	X_{18}	Percentage (%)	Demonstrating dependence on water sources, namely springs.
Number of poor people	X_{19}	Total	Absolute indicators of economic vulnerability.
Percentage of poor people	X_{20}	Percentage (%)	Relative (proportional) indicators of the economic condition of the community.
Access to safe drinking water	X_{21}	Percentage (%)	Measuring the level of infrastructure and basic services that affect resilience to drought.
Number of mining activities	X_{22}	Total	Demonstrating human pressure on natural resources because mining can cause environmental degradation.

2.3 Cluster analysis

Cluster analysis is a fundamental unsupervised learning technique in data mining and statistical modeling that groups or categorizes a set of objects into clusters based on their similarities or differences in a multidimensional feature space (Wani, 2024). The primary goal of this approach is to ensure that objects within the same cluster exhibit a high degree of similarity in their characteristics, while objects within different clusters exhibit significant differences (Yin et al., 2024). Through this process, researchers can uncover hidden patterns, structures, or relationships in complex datasets that lack predefined labels, allowing for more meaningful interpretations of data behavior (Chaudhry et al., 2023; Mason et al., 2025; Mutambik, 2024). This method is widely applied across various disciplines, including environmental studies, socioeconomic research, biology, marketing, and machine learning, to identify natural groupings or segmentations that help simplify and interpret large datasets (Oyewole & Thopil, 2023).

Among the various clustering algorithms available, K-Means Clustering is one of the most popular and efficient methods due to its conceptual simplicity and computational effectiveness. The K-Means algorithm is a non-hierarchical clustering technique that partitions data into a predetermined number of clusters (K) based on the proximity of each data point to the nearest cluster centroid (Iqbal et al., 2022). This process begins with the selection of K initial centroids, followed by an iterative procedure that assigns each observation to the nearest centroid and updates the centroid positions based on the mean of the assigned data points. This iterative optimization continues until the centroids stabilize, indicating that the algorithm has reached convergence. As a result, the data is divided into clusters where internal similarity within each cluster is maximized and differences between clusters are minimized (Gallego et al., 2025).

Simply put, the K-Means Clustering method ensures that data points with similar values or characteristics are grouped in the same cluster, while data points with significantly different values are placed in separate clusters (Tabianan et al., 2022). This approach allows researchers to visualize and interpret patterns that may be hidden in large or multidimensional datasets. The primary goal is to generate clusters that are internally homogeneous, meaning that data points within each cluster are as similar as possible and externally heterogeneous, thus ensuring clear boundaries between clusters. Therefore, the use of K-Means Clustering provides a structured and interpretable framework for understanding the underlying data distribution and facilitates decision-making processes in a variety of applications, from environmental classification to socio-economic segmentation (Yang et al., 2021; Pant et al., 2025; Diggle, 2023).

Data processing and analysis in this study used RStudio programming (R version 4.2.1), along with several important libraries (Alfonso et al., 2023; Starbuck, 2023). Several libraries in R have different functions according to the needs of data analysis and mapping (Farkas, 2019). The 'factoextra' library is used to visualize the results of multivariate analysis such as K-Means clustering (Zagajewski et al., 2021), while 'mapvalues' from the 'plyr' library functions to replace or map certain values in a variable The 'plyr' library itself is used to group and summarize data (Yousif et al., 2020). Meanwhile, several other libraries focus on spatial data processing. 'rgdal', 'sf', and 'terra' are used to read, manage, and analyze spatial data (both vector and raster). The 'tmap' library is used to create interactive and static thematic maps, and 'tmaptools' provides additional functions such as geocoding or map reprojection. The 'lwgeom' package supports advanced geometry operations for 'sf', such as calculating area, distance, or creating more complex spatial buffers (Sanjari et al., 2023; Alfonso et al., 2023). The data analysis steps in RStudio programming for the k-means clustering algorithm in this study are as follows: importing the dataset into the program and then perform data cleaning; and exploring the data through descriptive statistics, namely the minimum, mean, median, and maximum values (Bulanov et al., 2021; Starbuck, 2023).

$$\mu = \frac{\sum x_i}{n} \quad \text{(Eq. 1)}$$

Where μ represents the meaning of the data set, x_i denotes the value of each data point, n is the total number of data points, and Σ denotes the sum of all data values included in the calculation.

$$Q_2 = Tb \left(\frac{\frac{1}{2}n - f_k}{f_i}\right) p \quad \text{(Eq. 2)}$$

Where Q_2 represents the median, Tb denotes the lower boundary of the median class, n is the total frequency, f_k is the cumulative frequency up to the median class, f_i is the frequency of the median class, and p represents the class interval width. Standardize data, as each data scale/unit is different (Cuevas-Diaz Duran et al., 2024).

$$Z = \frac{x_i - \mu}{s} \quad \text{(Eq. 3)}$$

Here, Z represents the standardized data value (Z-score), x_i denotes the original value of the i-th data point, μ is the mean of the data set, and s represents the standard deviation. Determination of the optimum cluster using the silhouette methods through plots (Hartama & Oktaviani, 2025).

$$S(i) = \frac{b(i) - a(i)}{max\ (a(i), b(i))} \quad \text{(Eq. 4)}$$

Where $S(i)$ represents the Silhouette Coefficient for object i, $a(i)$ denotes the average distance between object i and all other objects in the same cluster, $b(i)$ is the minimum average distance between object i and objects in other clusters, and max refers to the maximum value between $a(i)$ and $b(i)$ used in the calculation of the Silhouette Coefficient. The analysis steps are as follows: first perform K-Means Clustering analysis, then identify the results of the plot and clusters formed and the acquisition of each cluster member. Second, visualize areas based on established clusters. Third provide recommendations for sustainable environmental management policies based on the conditions of each areas in each cluster.

3. Results and Discussion

3.1 Descriptive statistics

Descriptive statistics are used to provide a concise yet informative summary of research data, allowing for an initial understanding of the general characteristics, trends, and variations within the data set before proceeding to further analysis. This step involves calculating key statistical measures such as the mean, median, minimum, and maximum, which collectively provide insight into the central tendency, dispersion, and overall distribution of each variable. By presenting these descriptive values, researchers can identify potential data patterns, detect anomalies, and assess whether the data is suitable for further analysis, such as clustering. A detailed summary of these descriptive statistics for all variables used in the study is presented in Table 3.

Table 3. Descriptive statistics of research variable

Variable	Minimum	Mean	Median	Maximum
X_1	356.4	4738.9	5126.6	8736
X_2	46	155.4	144	337
X_3	3	63.31	50	183
X_4	0	2.308	1	8
X_5	14162	330096	302311	735171
X_6	0	0.1538	0	1
X_7	2	5.077	3	15
X_8	0	2.462	2	7
X_9	33.21	34.07	33.65	35.84
X_{10}	89.25	94.02	93.54	98.58
X_{11}	5.5	7.315	6.27	11.42
X_{12}	1006	1011	1012	1013
X_{13}	67.41	180	226.95	248.71
X_{14}	17	29.38	25	60
X_{15}	4.8	7.115	4.9	18.3
X_{16}	74940	242777	236020	464770
X_{17}	15	180	52	1115
X_{18}	1.59	24.96	29.17	54,53
X_{19}	11160	29213	26120	74570
X_{20}	5.94	12.29	12.45	16.36
X_{21}	73.05	87.6	88.92	99.25
X_{22}	0	23.31	10	79

Based on Table 3 above, the descriptive statistics for each variable can be interpreted as follows: For example, for the variable of rainfall amount (X_{13}), it is known that the minimum value of 67.41 mm/year is located in Donggala, the mean value is 180 mm/year, the median value is 226.85, and the maximum value is 248.71, which is located in Morowali Utara, Poso, and Tojo Una-Una. The same interpretation applies to all other variables.

3.2 Data standardization

Data standardization is performed as an important preprocessing step because the variables used in the analysis have different scales and units of measurement, potentially biasing the clustering results. This process ensures that each variable contributes equally to the clustering algorithm by transforming the data so that all variables have comparable scales, typically with a mean of zero and a standard deviation of one. Through standardization, variables measured over a larger numerical range are prevented from dominating variables with a smaller range, resulting in more balanced and reliable clustering results. The results of the data standardization process, including adjusted values for each variable, are presented in Table. 4, which illustrates how the transformation normalized the dataset and prepared it for subsequent K-Means clustering analysis.

Table 4. Data standardization

X_1	X_2	X_3	...	X_{20}	X_{21}	X_{22}
1.34628478	2.29249323	1.14035384	...	-1.8947159	0.43448142	-0.1219731
-0.90180444	-0.14370563	0.75217853	...	0.00966432	1.05177743	-0.7857336
-1.55193332	-1.12828341	-0.7265845	...	0.49237183	-0.21239335	-0.8594848
-0.38899316	-0.5097666	-0.0056875	...	0.26093673	-0.89913516	-0.6013557
0.14853067	0.14661858	-0.5232546	...	0.99491664	0.23257418	2.0536865
-1.67905334	-1.38073925	0.01279699	...	-2.0997013	1.498031	0.5417874
-0.09655083	-0.28255635	-1.1147598	...	-0.2449143	0.16955855	1.7586818
1.53138042	-0.37091589	-0.2459865	...	-0.1126656	-0.38857992	0.8736677
0.40867999	1.61086246	-0.6341618	...	0.6312329	-0.76024356	-0.6751069
1.07543561	0.18448696	-0.8374917	...	0.64115155	0.97461543	-0.7857336
0.18640194	0.2728465	2.21245708	...	-0.0762972	-1.19106473	-0.7488580
0.31950672	-0.11846005	0.99247753	...	1.34537552	0.96175509	-0.4907289
-0.39788505	-0.57288056	-1.0223371	...	0.05264513	-1.87137637	-0.1588486

3.3 Determination of the optimum cluster

Optimum clustering is a crucial step in the clustering analysis process because it determines the most appropriate number of clusters and best represents the natural clustering structure in a dataset. The primary goal of this stage is to ensure that the clustering results accurately capture the underlying patterns and relationships among data points, thus preventing overfitting or underfitting. Essentially, optimal clustering helps researchers find a balance between too few clusters, which can oversimplify the data, and too many clusters, which can complicate interpretation and reduce generalizability.

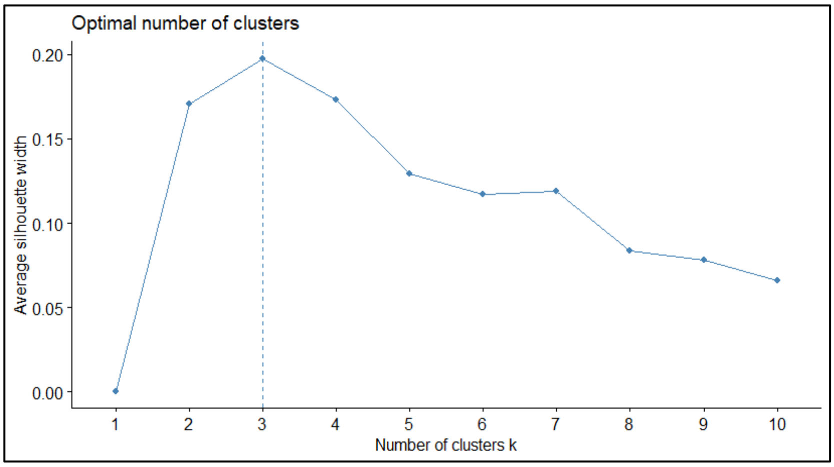

Fig. 1. Plot silhouette

This determination process focuses on identifying the point at which adding additional clusters no longer significantly improves the model's performance in explaining data variability, a concept often referred to as the point of diminishing returns. To achieve this, various evaluation methods can be applied, one of the most widely used being the silhouette method, which measures how well each data point within a given cluster fits relative to the others. A higher average silhouette value indicates a more appropriate and stable clustering configuration. By testing different numbers of clusters and comparing their silhouette scores, researchers can select the optimal solution that balances model simplicity and interpretative accuracy. The results of this evaluation are typically visualized in an optimal cluster plot, as shown in Fig. 1, which displays the relationship between the number of clusters and the level of explained variance. This plot not only supports the identification of the most stable and meaningful cluster configuration but also strengthens the overall reliability of the clustering analysis. Based on (Fig. 1), it is known that the optimum number of clusters for area clustering is 3.

3.4 Cluster results

After determining the optimal number of clusters through an evaluation process such as the Silhouette Method, the K-Means clustering algorithm is run to group the data into their respective categories based on their characteristic similarities. This algorithm iteratively assigns each data point to the nearest cluster centroid and recalculates the centroid positions until convergence is achieved, ensuring that intra-cluster similarities are maximized while inter-cluster differences are minimized. As a result, the analysis produces the following cluster plot (Fig. 2), which visually represents the distribution of data points within each cluster and illustrates how well objects are separated among the predefined groups.

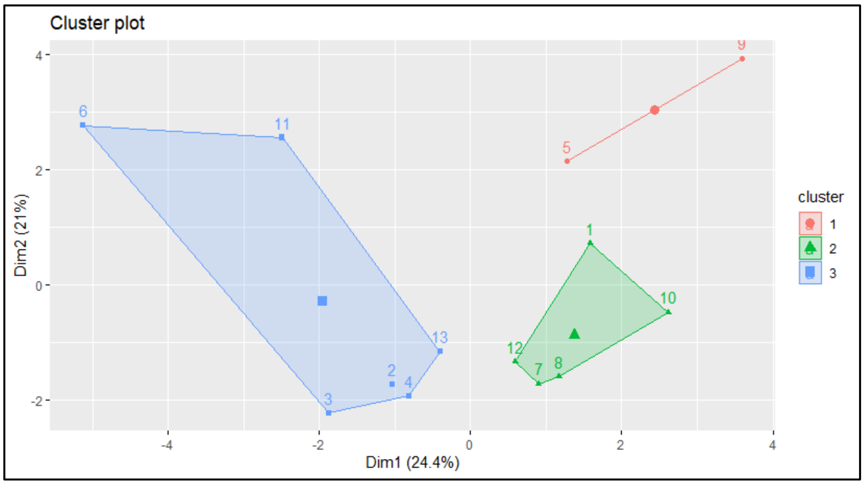

Fig. 2. Cluster plot

The K-Means Clustering plot results show the formation of three clusters with fairly clear separation patterns. Cluster 1 consists of a small number of objects and is located far away from the other clusters, indicating that this group has the most different or extreme characteristics compared to the other clusters. Cluster 2 has closely related members and forms a compact group, reflecting the uniformity of characteristics between objects within it. Meanwhile, cluster 3 has the most members and covers a wider area, indicating that there is variation in values within the group but it is still homogeneous enough to be classified together.

3.5 Visualization of cluster results

Cluster visualization using a map of Central Sulawesi Province was conducted to provide a clearer and more comprehensive understanding of the spatial distribution and characteristics of each cluster obtained from the analysis. By displaying the clustering results in a geographic context, it is easier to observe how different regions are grouped based on similarities in selected indicators. This approach not only aids in interpreting the analysis results but also enables policymakers and researchers to identify regional patterns, prioritize areas for intervention, and make more informed decisions regarding environmental and development planning in Central Sulawesi Province.

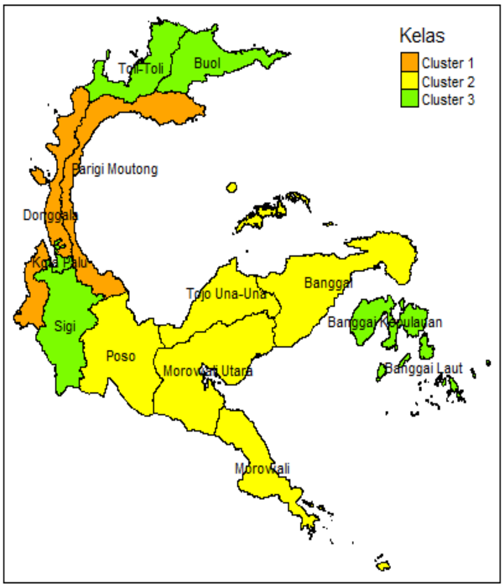

Fig. 3. Cluster visualization

The visualization results are as follows (Fig. 3). The following are the members of each cluster in Table 5. Based on the analysis results, the characteristics of each cluster and its area can be interpreted as follows. Cluster 1 is an area prone to high drought, consisting of two areas, characterized by a large land area with low rainfall and few rainy days.

Table 5. Final results of each cluster member

Cluster	Regencies/City (area)	Charateristics	Category
1	Donggala and Parigi Moutong.	High temperatures, low rainfall, and high mining activity.	High drought.
2	Banggai, Morowali, Morowali Utara, Poso, and Tojo Una-Una.	Extensive forests, moderate rainfall, and mining activity.	Moderate drought.
3	Banggai Kepulauan, Banggai Laut, Buol, Kota Palu, Sigi, and Tolitoli.	High rainfall and good water quality.	Low drought.

The air temperature is relatively high (around 35°C), and air pressure tends to be low, indicating hot and dry conditions. Access to clean drinking water is fairly good but not yet evenly distributed. Mining activity is quite high (especially in Donggala). Forest cover is still extensive, but there is potential for environmental degradation due to mining. Cluster 2 is an area with moderate drought risk, consisting of five areas, characterized by a large land area with extensive forest and water cover. Rainfall is moderate to high and air humidity is high. Some areas experience extreme weather and forest fires, but not significantly. Access to safe drinking water is good (80–95%). Mining activities exist, but are not as extreme as in Cluster 1. Cluster 3 is an area with low drought risk, consisting of six areas, characterized by low to zero frequency of drought events. Rainfall and number of rainy days are high. Air temperature is more stable, humidity is high. Water infrastructure and access to drinking water are very good (even >95% in Kota Palu and Sigi). Mining activity is relatively small, and socio-economic pressure is also low.

3.5 Policy recommendations

The results of the analysis for each cluster have identified the members of each regencies/city in Central Sulawesi with the characteristics and conditions of the area, so that policy recommendations can then be made to support sustainable environmental management as follows Table. 6.

Table 6. Policy recommendations

Cluster	Regencies/City (area)	Policy recommendations
1	Donggala and Parigi Moutong.	Optimization of irrigation systems, village reservoirs, and rainwater management. Rehabilitation of forests and water catchment areas to maintain groundwater reserves. Increased community capacity in water management and water-efficient agriculture. Integration of spatial and weather data for regional development planning.
2	Banggai, Morowali, Morowali Utara, Poso, and Tojo Una-Una.	Maintain existing environmental conditions through water and forest conservation policies. Implement sustainable agricultural systems such as agroforestry. Develop an early warning system for extreme climate change. Establish local inter-regional water management institutions.
3	Banggai Kepulauan, Banggai Laut, Buol, Kota Palu, Sigi, and Tolitoli.	Focus on conservation and prevention of environmental degradation. Development of a green economy through eco-tourism or sustainable agriculture. Monitoring of mining expansion and land conversion. Replication of good water and land management practices to other regions.

In Cluster 1 (high drought), this region is characterized by a high intensity of drought and extreme weather events, low rainfall, and limited irrigation networks and access to clean drinking water. Therefore, policy recommendations focus on increasing drought resilience and water resource conservation. Local governments are advised to construct and rehabilitate irrigation networks at the village level and develop integrated watershed-based water management systems. Water resource conservation efforts through reforestation and reforestation in critical lands need to be strengthened to maintain groundwater reserves. Furthermore, the implementation of climate-adaptive agricultural technologies such as drip irrigation and drought-resistant crop varieties is an important strategy to reduce the risk of crop failure. Increasing community literacy on climate change adaptation and developing drought risk maps can support local preparedness. Finally, diversification of the rural economy based on local resources, such as ecotourism and environmentally friendly businesses, can reduce community dependence on the agricultural sector, which is vulnerable to climate change.

Furthermore, Cluster 2 (moderate drought) describes regions with relatively stable climates, extensive forest cover, and adequate access to water. However, the threat of environmental degradation due to mining and land conversion still needs to be addressed. Policy recommendations for this cluster include strengthening forest and water catchment area protection policies as the primary bulwark for long-term water supply. The government needs to tighten oversight of mining activities and ensure the implementation of post-mining land reclamation. Furthermore, encouraging the use of renewable energy and environmentally friendly technologies can support the transition to a green economy. In the agricultural sector, the implementation of agroforestry and conservation agriculture systems is crucial for maintaining soil moisture and production stability. Regular monitoring of local climate conditions and the development of community-based water governance are also recommended to enable communities to play an active role in preserving water resources. Finally, collaboration between government, academics, and non-governmental organizations in environmental research and management needs to be strengthened to produce scientifically based policies.

Cluster 3 (low drought) indicates areas with high humidity and rainfall, adequate irrigation infrastructure, and access to safe drinking water. The policy focus for this cluster is not on drought mitigation, but rather on efforts to maintain stable and sustainable environmental conditions. Regional governments need to implement long-term environmental conservation policies, including protecting green areas and controlling land conversion. The development of green infrastructure such as city parks, green open spaces, and infiltration wells can help maintain the carrying capacity of urban and rural environments. Spatial data integration and digital monitoring systems need to be developed to monitor land-use changes and detect environmental degradation early. Furthermore, increasing community capacity to mitigate climate disasters, particularly floods and extreme weather, is crucial for maintaining ecosystem balance. Circular economic development and sustainable waste management are also recommended to support green development principles, while biodiversity conservation and water conservation need to be strengthened to maintain ecosystem stability. Areas within this cluster can also serve as models or best practices for sustainable environmental management that can be replicated in other areas of Central Sulawesi.

Overall, the results of this clustering provide a scientific basis for formulating drought adaptation and mitigation policies at the regional level. Each cluster requires a different approach based on its environmental and socio-economic characteristics. Synergy between regional governments, technical agencies, academics, and communities is key to creating an environmental management system that is responsive to climate change, efficient in natural resource utilization, and inclusive for community well-being.

4. Conclusions

This study highlights conditions in Central Sulawesi by identifying and clustering areas based on their vulnerability to drought using geographic, climatic, and socio-economic indicators. A total of 22 variables were used, covering the three main indicators mentioned above. The analysis was conducted using the K-Means Clustering method and RStudio programming. The analysis resulted in three main clusters representing different levels of vulnerability, namely Cluster 1 (high drought), Cluster 2 (moderate drought), and Cluster 3 (low drought).

Cluster 1 consists of two areas, Donggala and Parigi Moutong, which are characterized by relatively high extreme weather, low rainfall, and high mining activity. Cluster 2 consists of 5 areas, namely Banggai, Morowali, Morowali Utara, Poso, and Tojo Una-Una, which have relatively better conditions with stable rainfall, extensive forest cover, and limited mining activity. Meanwhile, Cluster 3 consists of 6 areas, namely Banggai Kepulauan, Banggai Laut, Buol, Kota Palu, Sigi, and Tolitoli, which are the safest areas, with high rainfall, good air humidity, and adequate access to drinking water.

These findings show that climate factors such as rainfall, temperature, and air humidity are the most influential indicators in distinguishing the level of drought vulnerability between regions. In addition, socio-economic conditions also play an important role in strengthening or weakening the community's ability to adapt to drought. Regions with high poverty rates and limited access to drinking water tend to be more vulnerable, even though they have similar geographical conditions to other regions.

Based on the clustering results, each group of regions requires different policy strategies. High drought clusters (cluster 1) require priority in the development of irrigation facilities, water conservation, and strengthening community capacity to cope with climate change. The moderate drought cluster (cluster 2) needs to be directed towards efforts to maintain the sustainability of existing natural resources, while the low drought (cluster 3) should focus on conservation and green economic development activities to prevent environmental degradation. Overall, the results of this study can be used as a starting point for local governments in formulating data-based policies related to water resource management and drought mitigation. The data-driven approach using the K-Means Clustering technique provides a spatial overview of drought vulnerability distribution and supports sustainable environmental development efforts. Further research is recommended to integrate more detailed temporal and spatial data, such as vegetation indices or land cover changes, to enable more comprehensive drought risk mapping.

Acknowledgement

The author sincerely thanks all parties involved in this research for their contributions; and would like to express my gratitude to the relevant agencies for providing open access data and methodological insights.

Author Contribution

Z.M., Writing – review & editing, Writing – original draft, Methodology, Data Analysis, Programming, Conceptualization.

Data Availability Statement

The dataset used is sourced from the Central Sulawesi Statistics Agency (Badan Pusat Statsitik Sulawesi Tengah) in the publication "Sulawesi Tengah Dalam Angka 2025" and all publications from each regency/city, as well as data from the Central Sulawesi Energy and Mineral Resources Agency (Dinas Energi dan Sumber Daya Mineral Sulawesi Tengah) in the publication "Buku Data dan Statistik DESDM 2025", and data from Climatology and Geophysics Agency (Badan Metereologi Klimatologi dan Geofosika) and National Disaster Management Agency (Badan Penanggulangan Bencana Nasional Sulawesi Tengah).

Conflicts of Interest

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Generative AI Use

During the preparation of this work, the author used Grammarly and DeepL to help improve the grammar, clarity, translation, and academic tone of the manuscript. After using these tools, the author reviewed and edited the content as necessary and is fully responsible for the content of the publication.

Open Access

References

Akhmad, R., Sumarmi, S., Astina, I. K., & Wagistina, S. (2025). A sustainability trilogy approach for drought risk prevention: Case study in Indonesia. *Jàmbá Journal of Disaster Risk Studies*, 1–12. https://doi.org/10.4102/jamba.v17i1.1811

Alfiandy, S., & Permana, D. S. (2020). Trend of Rainfall Based on BMKG Observation Data and MERRA-2 NASA Reanalysis in Central of Sulawesi Province. *Jurnal Sains & Teknologi Modifikasi Cuaca*, *21*(2), 63–72. https://ejournal.brin.go.id/JSTMC/article/view/1205

Alfonso, L., Martos, P., García-vico, Á. M., González, P., & Carmona, C. J. (2023). Clustering : an R library to facilitate the analysis and comparison of cluster algorithms. *Progress in Artificial Intelligence*, *12*(1), 33–44. https://doi.org/10.1007/s13748-022-00294-2

Awad, F. H., & Hamad, M. M. (2022). Improved k-Means Clustering Algorithm for Big Data Based on Distributed SmartphoneNeural Engine Processor. *Electronics (Switzerland)*, *11*(6). https://doi.org/10.3390/electronics11060883

Badan Metereologi Klimatologi Geofisika. (2025). *Iklim Sulawesi Tengah*. Badan Metereologi Klimatologi Geofisika. https://www.bmkg.go.id

Badan Nasional Penanggulangan Bencana. (2025). *PUSDALOPS-PB SULTENG*. Badan Nasional Penanggulangan Bencana. https://www.dashboard.core.pusdalops-bpbdsulteng.com

Badan Pusat Statistik Sulawesi Tengah. (2025). *Sulawesi Tengah Dalam Angka 2025* (BPS Sulawesi Tengah (Ed.); Vol. 55). BPS Sulawesi Tengah. https://sulteng.bps.go.id

Bahtiar, A., & Yulfiah, Y. (2025). Impacts of La Niña on Sea Surface Temperature, Chlorophyll-a, and Fishery Productivity in Northern East Java Ocean. *Journal of Earth and Marine Technology (JEMT)*, *5*(2), 171–183. https://doi.org/10.31284/j.jemt.2025.v5i2.7905

Biswas, A., Sarkar, S., Das, S., Dutta, S., Roy Choudhury, M., Giri, A., Bera, B., Bag, K., Mukherjee, B., Banerjee, K., Gupta, D., & Paul, D. (2025). Water scarcity: A global hindrance to sustainable development and agricultural production – A critical review of the impacts and adaptation strategies. *Cambridge Prisms: Water*, *3*. https://doi.org/10.1017/wat.2024.16

Bulanov, N., Suvorov, A. Y., Blyuss, O. B., Munblit, D. B., Butnaru, D. V., Nadinskaia, M. Y., & Zaikin, A. A. (2021). Basic principles of descriptive statistics in medical research. *Sechenov Medical Journal*, *12*(3), 4–16. https://doi.org/10.47093/2218-7332.2021.12.3.4-16

Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D. L. R., Thompson, E. B., & Ashraf, I. (2023). A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. *Symmetry*, *15*(9), 1–44. https://doi.org/10.3390/sym15091679

Cuevas-Diaz Duran, R., Wei, H., & Wu, J. (2024). Data normalization for addressing the challenges in the analysis of single-cell transcriptomic datasets. *BMC Genomics*, *25*(1), 1–18. https://doi.org/10.1186/s12864-024-10364-5

Diggle, P. J. (2023). Data sets. *Time Series*, 228–244. https://doi.org/10.1093/oso/9780198522065.005.0001

Dinas Energi dan Sumber Daya Mineral Sulawesi Tengah. (2025). *Buku Data dan Statistik Energi dan Sumber Daya Mineral Sulawesi Tengah* (DESDM Sulteng (Ed.)). https://desdm.sultengprov.go.id

Doan, Q. Van, Amagasa, T., Pham, T. H., Sato, T., Chen, F., & Kusaka, H. (2023). Structural k-means (S k-means) and clustering uncertainty evaluation framework (CUEF) for mining climate data. *Geoscientific Model Development*, *16*(8), 2215–2233. https://doi.org/10.5194/gmd-16-2215-2023

En-Nagre, K., Aqnouy, M., Ouarka, A., Ali Asad Naqvi, S., Bouizrou, I., Eddine Stitou El Messari, J., Tariq, A., Soufan, W., Li, W., & El-Askary, H. (2024). Assessment and prediction of meteorological drought using machine learning algorithms and climate data. *Climate Risk Management*, *45*(October 2023), 100630. https://doi.org/10.1016/j.crm.2024.100630

Faizan, M., Zuhairi, M. F., Ismail, S., & Sultan, S. (2020). Applications of Clustering Techniques in Data Mining: A Comparative Study. *International Journal of Advanced Computer Science and Applications*, *11*(12), 146–153. https://doi.org/10.14569/IJACSA.2020.0111218

Farkas, G. (2019). Possibilities of using raster data in client-side web maps. *Transactions in GIS, 24*(1). https://doi.org/10.1111/tgis.12588

Gallego, V., Freixes, A., & Lingan, J. (2025). Applying Machine Learning in Marketing: An Analysis Using the NMF and K-Means Algorithms. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* , *14778 LNCS*, 14–26. https://doi.org/10.1007/978-3-031-78238-1_2

Hartama, D., & Oktaviani, S. (2025). Optimization of K-Means and K-Medoids Clustering Using Dbi Silhouette Elbow on Student Data. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, *11*(2), 289–296. https://doi.org/10.33330/jurteksi.v11i2.3531

Iqbal, A., Shil, A., Chowdhury, M. J. M., Moni, M. A., & Sarker, I. H. (2022). An Improved K-means Clustering Algorithm Towards an. *Annals of Data Science*. https://doi.org/10.1007/s40745-022-00428-2

Juliannisa, I. A., Rahma, H., Mulatsih, S., & Fauzi, A. (2025). Regional Vulnerability to Food Insecurity: The Case of Indonesia. *Sustainability (Switzerland)*, *17*(11), 1–20. https://doi.org/10.3390/su17114800

Kuswanto, H., Puspa, A. W., Ahmad, I. S., & Hibatullah, F. (2021). Drought Analysis in East Nusa Tenggara (Indonesia) Using Regional Frequency Analysis. *Scientific World Journal*, *2021*. https://doi.org/10.1155/2021/6626102

Li, M., Frank, E., & Pfahringer, B. (2023). Large scale K-means clustering using GPUs. *Data Mining and Knowledge Discovery*, *37*(1), 67–109. https://doi.org/10.1007/s10618-022-00869-6

Lim, Z. Y., Ong, L. Y., & Leow, M. C. (2021). A review on clustering techniques: Creating better user experience for online roadshow. *Future Internet*, *13*(9). https://doi.org/10.3390/fi13090233

Mason, L., Hicks, B., & Almeida, J. S. (2025). ClusterRadar: An interactive web-tool for the multi-method exploration of spatial clusters over time. *PLoS ONE*, *20*(5 May), 1–23. https://doi.org/10.1371/journal.pone.0322393

Mens, M. J. P., Van Rhee, G., Schasfoort, F., & Kielen, N. (2022). Integrated drought risk assessment to support adaptive policymaking in the Netherlands. *Natural Hazards and Earth System Sciences*, *22*(5), 1763–1776. https://doi.org/10.5194/nhess-22-1763-2022

Minea, I., & Albulescu, A. C. (2025). Drought Over Time: an Investigation of the Impact of Meteorological Drought on Groundwater and Surface Water in the East of Romania. *Earth Systems and Environment*, *9*(3), 1959–1981. https://doi.org/10.1007/s41748-025-00725-9

Mu'arif, Z., Afriza, D. A., Aulia, F., Anggelina E, M. P., & Gamayanti, N. F. (2024). Analysis of Priority Areas for Handling Stunting Cases in Sigi Regency Using the Topsis Method Based on Web Dashboard. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *18*(3), 1411–1422. https://doi.org/10.30598/barekengvol18iss3pp1411-1422

Mutambik, I. (2024). An Entropy-Based Clustering Algorithm for Real-Time High-Dimensional IoT Data Streams. *Sensors*, *24*(22). https://doi.org/10.3390/s24227412

Mutia, R., Mardani, M., Fathia, S., & Kusmita, T. (2023). Identifikasi Sesar dengan Menggunakan Metode Gaya Berat (Studi Kasus: Sulawesi Tengah). *Jurnal Riset Fisika Indonesia*, *3*(2), 1–6. https://doi.org/10.33019/jrfi.v3i2.3552

Orimoloye, I. R., Belle, J. A., Orimoloye, Y. M., Olusola, A. O., & Ololade, O. O. (2022). Drought: A Common Environmental Disaster. *Atmosphere*, *13*(1). https://doi.org/10.3390/atmos13010111

Oyewole, G. J., & Thopil, G. A. (2023). Data clustering: application and trends. In *Artificial Intelligence Review* (Vol. 56, Issue 7). Springer Netherlands. https://doi.org/10.1007/s10462-022-10325-y

Pant, Y. R., Leigh, L., & Fajardo Rueda, J. (2025). Improving K-Means Clustering: A

Comparative Study of Parallelized Version of Modified K-Means Algorithm for Clustering of Satellite Images. *Algorithms*, *18*(8). https://doi.org/10.3390/a18080532

Roy, S. K., Morshed, A., Mojumder, P., Hasan, M. M., & Islam, A. K. M. S. (2024). Innovative trend analysis technique with fuzzy logic and K-means clustering approach for identification of homogenous rainfall region: A long-term rainfall data analysis over Bangladesh. *Quaternary Science Advances*, *15*(June), 100227. https://doi.org/10.1016/j.qsa.2024.100227

Sanjari, E., Majidian Dehkordi, F., & Raeisi Shahraki, H. (2023). Clustering Undergraduate Students Based on Academic Burnout and Satisfaction from the Field Using Partitioning around Medoid. *Computational and Mathematical Methods in Medicine*, *2023*(c). https://doi.org/10.1155/2023/8898939

Savelli, E., Rusca, M., Cloke, H., & Di Baldassarre, G. (2022). Drought and society: Scientific progress, blind spots, and future prospects. *Wiley Interdisciplinary Reviews: Climate Change*, *13*(3), 1–25. https://doi.org/10.1002/wcc.761

Shim, I., Kim, H., Hong, B., An, J., & Hwang, T. (2021). Drought vulnerability assessment and cluster analysis of island areas taking Korean island areas at Eup (Town) and Myeon (subcounty) levels as study targets. *Water (Switzerland)*, *13*(24). https://doi.org/10.3390/w13243657

Starbuck, C. (2023). The Fundamentals of People Analytics. In *The Fundamentals of People Analytics*. https://doi.org/10.1007/978-3-031-28674-2

Tabianan, K., Velu, S., & Ravi, V. (2022). K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data. *Sustainability (Switzerland)*, *14*(12), 1–15. https://doi.org/10.3390/su14127243

Van Loon, A. F., Kchouk, S., Matanó, A., Tootoonchi, F., Alvarez-Garreton, C., Hassaballah, K. E. A., Wu, M., Wens, M. L. K., Shyrokaya, A., Ridolfi, E., Biella, R., Nagavciuc, V., Barendrecht, M. H., Bastos, A., Cavalcante, L., de Vries, F. T., Garcia, M., Mård, J., Streefkerk, I. N., … Werner, M. (2024). Review article: Drought as a continuum – memory effects in interlinked hydrological, ecological, and social systems. *Natural Hazards and Earth System Sciences*, *24*(9), 3173–3205. https://doi.org/10.5194/nhess-24-3173-2024

Wang, Q., Yang, X., Qu, Y., Qiu, H., Wu, Y., Qi, J., Song, H., Chen, Y., Chu, H., & Zeng, J. (2025). Global Climate Change Exacerbates Socioeconomic Drought Severity Across Vegetation Zones During 1901–2018. *International Journal of Disaster Risk Science*, *16*(2), 291–306. https://doi.org/10.1007/s13753-025-00631-8

Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*, *10*, 1–45. https://doi.org/10.7717/PEERJ-CS.2286

Wijitkosum, S. (2025). Integrated spatial analysis of drought risk factors using agglomerative hierarchical clustering and correlation. *Environmental Advances*, *21*(June), 100646. https://doi.org/10.1016/j.envadv.2025.100646

Yang, J., Wang, Y. K., Yao, X., & Lin, C. T. (2021). Adaptive Initialization Method for K-Means Algorithm. *Frontiers in Artificial Intelligence*, *4*(November), 1–13. https://doi.org/10.3389/frai.2021.740817

Yang, X., Liao, X., Di, D., & Shi, W. (2023). A Review of Drought Disturbance on Socioeconomic Development. *Water (Switzerland)*, *15*(22). https://doi.org/10.3390/w15223912

Yin, H., Aryani, A., Petrie, S., Nambissan, A., Astudillo, A., & Cao, S. (2024). *A Rapid Review of Clustering Algorithms*. 1–25. http://arxiv.org/abs/2401.07389

Yousif, A., Drou, N., Rowe, J., Khalfan, M., Gunsalus, K. C., & Gunsalus, K. C. (2020). NASQAR: A web-based platform for high-throughput sequencing data analysis and visualization. *BMC Bioinformatics*, *21*(1), 1–14. https://doi.org/10.1186/s12859-020-03577-4

Zagajewski, B., Kluczek, M., Raczko, E., Njegovec, A., Dabija, A., & Kycko, M. (2021). Comparison of random forest, support vector machines, and neural networks for post-disaster forest species mapping of the krkonoše/karkonosze transboundary biosphere reserve. *Remote Sensing*, *13*(13). https://doi.org/10.3390/rs13132581

Biography of Author

Zainal Mu'arif, Independent Researcher or Alumni of the Department of Statistics, Faculty of Mathematics and Natural Sciences, Tadulako University, Bachelor of Statistics graduate, field of expertise; statistics and data science focusing on social, economic, environmental issues, and data-driven policies.

- Email: zainalmuarifstory@gmail.com
- ORCID: 0009-0008-9260-4720
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: N/A