International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Jul-2024)

http://dx.doi.org/10.12785/ijcds/160120

Disaster Event, Preparedness, and Response in Indonesian Coastal Areas: Data Mining of Official Statistics

Gunawan Gunawan

Faculty of Engineering, University of Surabaya, Surabaya, Indonesia

Received 9 Feb. 2024, Revised 9 Apr. 2024, Accepted 16 Apr. 2024, Published 1 Jul. 2024

Abstract: Coastal areas are vulnerable to disasters such as tsunamis, floods, large waves, and hurricanes. Most studies on disasters in coastal areas are based on surveys for specific areas, but studies investigating disasters on a country-wide level are few. Applying data analytics to disaster management is critical to reducing the impact of disasters. This study aims to classify provinces based on disaster events, disaster preparedness, and response capacity in coastal villages through cluster analysis, principal component analysis (PCA), and a combination of PCA and cluster analysis. This secondary study applies data mining techniques to official statistics in Indonesia. Data mining was performed with Python Scikit-learn and Tableau analytical software. The unit of analysis is all provinces of Indonesia as an archipelago country. The cluster analysis optimally produced two clusters with 6 (18%) and 27 (82%) provinces. The small cluster, named the high-intensity cluster, has a higher intensity of disaster events, preparedness, and response than the large cluster, named the low-intensity cluster. The low-intensity cluster has a higher percentage of coastal villages (25%) than the high-intensity cluster (10%). The results of the PCA are used to classify regions through geographic heat maps and scatter plots. Additionally, the combination of multiple principal component analysis and cluster analysis produced three clusters with 6 (18%), 10 (30%), and 17 (52%) provinces. However, the cluster model from cluster analysis alone provides a better separation between clusters than the combination of PCA and cluster analysis. Ultimately, cluster analysis and PCA can be used independently, and both methods are complementary to exploring regional classification. The results of this study recommend improvements in disaster preparedness and response for coastal villages, especially provinces with a high percentage of coastal villages.

Keywords: Data Analytic, Hazard, Rural, Seaside, Sustainability, Tsunami

1. Introduction

Disasters can occur in any region or country at any time, causing many fatalities and damage. One of the most destructive disasters, but one that is difficult to predict, is a tsunami that strikes coastal areas. The most devastating coastal disaster recorded in history is the 2004 Indian Ocean tsunami, which, claimed more than 200 thousand lives and affected 14 Indian Ocean countries [1]. The 2011 Great East Japan Earthquake and Tsunami also killed over 20,000 people, forced more than 400,000 people to evacuate, and destroyed a massive number of buildings and infrastructure [2]. Tsunamis and other coastal disasters, such as floods, storm surges, storms, and earthquakes, cause extensive destruction to people's lives, public infrastructure, business facilities, and the environment. The geographic and topographic profiles of coastal areas are different, and these conditions are also related to the disaster risks they face. By considering these risks and conditions, disaster management in coastal areas is critical.

There are various studies on coastal disasters in many countries. For example, a prior study investigated disaster awareness among communities on a coastal island in Taiwan using qualitative methods [3]. Another qualitative study investigated disaster problems and lessons covering some villages in Bangladesh [4]. Another study analyzed the coastal vulnerability index in Thailand using mixed secondary data sources [5]. Some recent studies employed data analytic techniques. For example, a study about storm surge disasters in coastal areas in China used secondary data and machine learning techniques to cover a large area [6]. Another study investigated coastal regions in Portugal using Cluster Analysis and Principal Component Analysis [7]. Those studies used secondary data and data analytics to cover a large coastal area.

Many studies on disasters in coastal areas in Indonesia are based on specific regions. For example, a study of disaster risk assessment using a semi-quantitative approach and a back-casting method in a coastal village in Jakarta [8]. Another study focused on the northern coastal area of Central Java province and used the official statistics and Geographical Information Systems technique [9]. Using a qualitative approach, another research investigated climate

E-mail address: gunawan@staff.ubaya.ac.id

change disasters on the Northern Coast of Java Island, Indonesia [10]. Those studies provided detailed information and understanding about particular regions but did not have generalizations for other regions or even the whole of Indonesia. This shortcoming limits our understanding of disaster management decisions in coastal areas.

Advances in digital technology, which have influenced various aspects of life, are also expected to help reduce the impact of disasters. The World Risk Report 2022 urged the digitalization of disaster management on three trends: (1) digitalization for data collection and analysis for disaster forecasting and disaster response, (2) digitalization for communication with affected persons, and (3) digitalization to improve cooperation and coordination between parties [11]. Research on disaster relates to the first trend, which is digitalization for disaster management. In this area, data mining is essential for disaster-related data analysis. Data mining represents a process of discovering patterns and other valuable information from a dataset.

The application of data mining in the disaster literature could be classified into three based on its tasks: (1) prediction of disaster events, (2) detection of disaster immediately after it occurs, and (3) disaster management strategies to improve communication and coordination between entities in responding to disasters [12]. Data mining is also widely applied in social behavior studies related to flood disasters [13]. Various data mining techniques used in disaster studies include principal component analysis, cluster analysis, text mining, time series, and temporal analysis. The use of data mining in disaster studies is limited due to some factors, such as the lack of data availability to access and the lack of knowledge to use the appropriate data mining techniques.

In sum, the first issue addressed in this study is limited studies on the coastal areas in Indonesia that cover the whole country. Second, the opportunity to use data analytical techniques to analyze secondary data about coastal area disasters. To address both issues, this study aims to investigate disaster management in Indonesian coastal villages. It achieves this objective by using two statistical techniques: cluster Analysis and Principal Component Analysis. The study's object is coastal villages, and the unit of analysis is provinces in Indonesia.

As an archipelago country with the third longest coastline, Indonesia faces a high vulnerability to disasters in coastal areas [14]. Coastal villages in small districts have little disaster response capacity [15], such as search and rescue, emergency medical assistance, and the establishment of emergency shelters. When a disaster occurs, immediate delivery of health services is a matter of life and death. These services require adequate health facilities (e.g., emergency units, hospitals, community health centers, clinics, drug stores). While vulnerable coastal areas are expected to have better health facilities, health facilities are often poor, especially in rural areas [16].

The remainder of the article is structured as follows. First, the literature on disaster management, principal component analysis, and cluster analysis are reviewed. This is followed by describing the research framework, the dataset, and the analytical methods. The results of the data analysis are then presented. The paper concludes by discussing theoretical and managerial implications, limitations, and directions for further research.

2. Related Works

Following the research purpose described earlier, this study focuses on processing data about disaster management in coastal areas to reveal valuable information. Achieving this goal requires (1) a conceptual framework describing the constructs with their variables and (2) choosing appropriate data analysis techniques. This section discusses these two, namely disaster management as the basis for the conceptual framework and data analysis techniques as the basis for characterizing the research objects.

A. Disaster Management

The United Nations defines a disaster as *"a serious disruption of the functioning of a community or a society at any scale due to hazardous events interacting with conditions of exposure, vulnerability, and capacity, leading to one or more of the following: human, material, economic, and environmental losses and impacts"* [17]. Disasters occur more frequently, and climate change is assumed to be a significant contributing factor. Climate change causes severe weather events such as floods and droughts [18]. As climate change is affected by nature and humans, the term disaster gradually replaces natural disasters. As indicated above, this replacement confirms that the UN defines *disaster* rather than *natural disaster*. The increasing incidence of disasters requires technological assistance, primarily digital technology, in disaster management.

Disasters are commonly classified into three types based on their source. First, meteorological disasters such as floods, thunderstorms, hurricanes, typhoons, snowstorms, drought, and hot waves, are the most regular disasters. Second, geological disasters such as volcanic eruptions, earthquakes, landslides, mudflows, and tsunamis could create extensive casualties. For example, On Dec. 26, 2024, the tsunami devastated the entire coastal community across the Indian Ocean, killing about 230,000 people [19]. The third type is a biological disaster, with an example of the recent COVID-19 outbreak, which took millions of lives worldwide. Coastal disasters come from both meteorological and geological disaster types. As disasters cause loss of life, property damage, social and economic disturbances, or environmental damage, it is critical to implement disaster risk management seriously.

Most governments have established disaster management programs to alleviate the risks and losses to humans and animals caused by disasters. The disaster management cycle is a framework for determining activities according to the stages of a disaster: pre-disaster, during-disaster, and post-disaster. Activities in pre-disaster include prevention, preparedness, and mitigation; during-disaster cover response, rescue, and relief; and post-disaster comprise recovery and development [20]. In general, the disaster management cycle follows four stages: (1) mitigation is to prevent or lessen the potential impact of disasters before they strike, (2) preparedness is the development of strategies, plans, and procedures to cope with potential disasters effectively, (3) response is the immediate effort to minimize the hazards, and (4) recovery is to restore the affected community to normal.

For a global reference, the United Nations Office for Disaster Risk Reduction defines disaster risk management as *the application of disaster risk management policies and strategies to avoid new disaster risks, lessen existing disaster risks, manage residual risk, and contribute to the resilience and reduction of disaster loss* [17]. This framework becomes a global guidance for disaster management implemented by countries. Governments, communities, and organizations must work together to manage disasters. For example, the Indonesian government established the National Disaster Management Agency (*Badan Nasional Penanggulangan Bencana*) and Regional Disaster Management Agencies (*Badan Penanggulangan Bencana Daerah*). These agencies are vital in coordinating the implementation of disaster management, especially for disaster response actions.

Referring to the disaster management cycle, the data analysis in this paper covers disaster preparedness and response. Preparedness is the knowledge and ability that governments, response-recovery agencies, communities, and individuals have developed to effectively anticipate, respond, and recover from the impact of possible, imminent, or current disasters [17]. Preparedness may include planning, training, and education activities for disaster events that cannot be mitigated. The local government prepares for disasters by installing an early warning system, making emergency plans, providing evacuation routes, and disseminating public information.

Furthermore, the response phase refers to action taken immediately before, during, or after disasters to save lives, decrease health impacts, ensure public safety, and respond to people's basic needs [17]. This phase can include providing public and emergency assistance services by the public, private, and community sectors, as well as participating communities and volunteers. Health facilities, for example, are critical to implementing emergency services.

Disasters are one of the leading causes of death in developing countries. Inadequate health facilities to respond to this incident will increase the number of casualties [21]. Health facilities are essential to provide immediate medical assistance and treatment for injuries, trauma, and illnesses resulting from the disaster. Health facilities should also care for people who experience mental disorders due to the impact of disasters. However, mental health management after a disaster, such as post-traumatic stress, is often neglected [22].

B. Analytical Techniques

The analytical techniques addressed are Principal Component Analysis (PCA) and Cluster Analysis because of their relevance to classify regions. PCA is a method to reduce the dimension of large-scale data sets while preserving as much information as possible. PCA is suitable when data are multivariable and correlated, with multiple observations per variable [23]. PCA works by converting the original dimensions (variables) into new dimensions, which are linear combinations of the original ones. These new dimensions are named principal components. Each principal component is orthogonal to each other. PCA is an exploratory method that is also helpful for data pre-processing. In this data pre-processing, fewer dimensions (variables) are input for subsequent statistical analysis or machine learning tools.

Dimensionality reduction makes PCA applicable in various studies and generates practical benefits. For example, PCA was implemented to reduce the number of ecological indicators so that monitoring becomes efficient [24] or to simplify the vulnerability-related variables of heritage buildings for better conservation decisions [25]. PCA is also used in some disaster-related studies. For example, PCA was compared to Expert's method to analyze social vulnerability in Ecuador [13]. PCA was also applied to label the level of damage of sectors after natural disasters in Indonesia [26] or to explore the characteristics and perceptions of the risk of natural disasters in Japan [27].

Cluster analysis is a multivariate data mining technique that groups entities (e.g., individuals, regions, products, events) based on selected characteristics or attributes. Cluster analysis is not a single but a set of statistical tools. The primary classification of cluster analysis is based on a method and algorithm [28]. First, cluster analysis is classified according to various methods: (1) Connectivity-based Clustering, (2) Centroid-based (Partition) Clustering, (3) Density-based (Model-based) Clustering, (4) Distribution-Based Clustering, (5) Fuzzy Clustering and (6) Constraint-based (Supervised) Clustering. Second, cluster analysis is classified based on its algorithms, such as k-means clustering, Hierarchical Clustering, and DBSCAN. These algorithms are developed from clustering methods. For example, the k-means algorithm belongs to the centroid-based method, the hierarchical clustering to connectivity-based clustering, and the DBSCAN algorithm belongs to the density-based method.

The choice of cluster algorithms depends on the characteristics of the data and the purpose of the analysis. The k-means algorithm divides the data into k clusters by minimizing the variance within each cluster. For example, k-means cluster analysis was applied to categorize past earthquakes based on magnitude and consequence [29].

Hierarchical clustering creates a group whose objects are similar to each other and different from the objects of the other group. It visually represents the group in the hierarchical tree called a dendrogram. For example, environmental research used hierarchical clustering to cluster regions [30]. DBSCAN detects clusters as high-density regions separated from low-density regions to discover clusters of any shape and size. For example, this algorithm is used to identify spatial density patterns in urban areas [31].

The use of PCA and cluster analysis requires datasets. One of the valuable data sources is the official data (official statistics) published by the government. Governments collect, process, and publish official data about citizens' lives and regions, such as demographics, social and economic development, living conditions, education, health, business, and the environment. Most official statistics are presented in tables with regions in rows and measures (attributes) in columns. Classifying regions into a few groups based on their similarity is helpful for better understanding the regions and planning public policy. Cluster analysis can identify some groups of regions based on similar characteristics in some measures. Recommendations or decisions could be made for groups rather than individual regions.

Some studies used both PCA and cluster analysis to characterize and classify regions. Those studies could be grouped into two. The first category independently used PCA and cluster analysis to characterize and classify regions. For example, PCA was performed to obtain the principal components used to map the Siberian territories according to the safety of the natural and anthropogenic territory; then, the cluster analysis was performed independently to classify the regions [32].

The second category implemented PCA and used the principal components obtained as input for cluster analysis. This method is useful, especially if analysis involves many variables. It is also useful if collinearity (high correlation level between variables) exists because the uncorrelated principal components of PCA remove collinearity. A data mining study for Indonesian cities implemented PCA to reduce eight variables into two principal components and used them as inputs for cluster analysis [33]. Another study implemented PCA to reduce the dimensions of five variables into three principal components and applied cluster analysis to them. [34]. In both studies, the result of a single PCA was used for the cluster analysis. Rarely are studies that implemented multiple PCAs and used all principal components obtained for cluster analysis.

3. Methods

Reviewing disaster management concepts and data analytic techniques earlier becomes the basis for designing the research framework along with its variables and planning the analysis method. Accordingly, this section presents three parts: research framework, data and variables, and analysis method.

A. Research Framework

This study is included in the secondary quantitative research type. Thus, secondary data is the source of analysis. A research framework was developed based on the disaster management stages, the availability of secondary data, and the rational relationship between constructs composing the framework. The framework was developed to explain the logical relationship between variables rather than to test the hypothesis. Fig. 1 shows three variables: disaster event, disaster preparedness, and disaster response. Disaster events represent various disasters (e.g., sea tidal waves, hurricanes/tornados, earthquakes, and tsunamis) in coastal villages. Disaster preparedness denotes the availability of facilities or efforts to anticipate disasters, such as early warning systems (EWS) and signage/evacuation routes. Furthermore, disaster response in this study refers to the availability of health facilities supporting the community when a disaster happens.

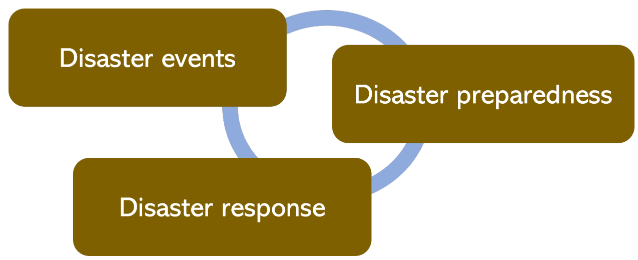

Figure 1. Research framework

B. Data and Variables

This research used secondary data from a report published by the Indonesian Statistics Agency, titled "Marine and Coastal Resources Statistics 2022" [35]. Data about coastal villages and disasters were extracted. The data show that all 34 provinces of Indonesia have coastal areas. Each province is divided into several regencies, districts, and villages. Therefore, a village is the smallest administrative region; consequently, coastal areas also belong to villages.

Table I displays the number of coastal villages (cv) per province and the percentage of coastal villages from the total villages in each province (% cv). The number of coastal villages varies between 17 in Jakarta and 1,040 in Maluku. The percentage of coastal villages ranges from 1% in South Sumatra to 85% in Riau Islands. Jakarta, as a metropolitan city and capital city of Indonesia, has different characteristics from other provinces. Therefore, Jakarta is removed for further analysis. Overall, the percentage of coastal villages in Indonesia is 15%, with 12,510 villages.

The research framework described in Fig. 1 comprises three variables. Based on the secondary data, measures of each variable are presented here. First, disaster events contain ten types: flood, earthquake, wave, wind, landslide, drought, forest fire, flash flood, eruption, and tsunami. Fig. 2 exhibits the frequency of each disaster struck in 2022. Some occurred only in a few provinces, such as tsunamis only

TABLE I. Coastal villages' profile

Province	Code	Coastal village	% Coastal village
Aceh	AC	662	10%
North Sumatra	SU	437	7%
West Sumatra	SB	133	10%
Riau	RI	244	13%
Jambi	JA	28	2%
South Sumatra	SS	31	1%
Bengkulu	BE	184	12%
Lampung	LA	238	9%
Bangka Belitung Island	BB	156	40%
Riau Islands	KR	364	85%
Jakarta	JK	17	6%
West Java	JB	221	4%
Central Java	JT	353	4%
Yogyakarta	YO	33	8%
East Java	JI	666	8%
Banten	BT	146	9%
Bali	BA	175	24%
West Nusa Tenggara	NB	281	24%
East Nusa Tenggara	NT	966	28%
West Kalimantan	KB	162	8%
Central Kalimantan	KT	40	3%
South Kalimantan	KS	161	8%
East Kalimantan	KI	158	15%
North Kalimantan	KU	54	11%
North Sulawesi	SA	760	41%
Central Sulawesi	ST	950	47%
South Sulawesi	SN	520	17%
South East Sulawesi	SG	911	39%
Gorontalo	GO	185	25%
West Sulawesi	SR	154	24%
Maluku	MA	1040	83%
North Maluku	MU	898	75%
West Papua	PB	592	30%
Papua	PA	590	11%

in one province and eruptions in two provinces. Therefore, further analysis selects the top four types of disaster ranging from 9% to 19%: flood, earthquake, wave (e.g., sea tidal wave), and wind (e.g., hurricane, storm, tornado). These four contribute 80% (=7128/8913) of disaster events.

Second, disaster preparedness contains four measures: signs and evacuation routes, safety equipment (e.g., inflatable boats, tents, mask supplies), general early warning systems (EWS), and early warning systems for tsunamis. EWSs for tsunamis are installed in many coastal villages, although actual tsunami disasters are rare. Tsunami awareness emerged after an extensive casualty of tsunami disasters in 2001 in Aceh. Third, disaster response is measured by health facilities covering community health centers, drugstores, and clinics. Hospitals are excluded because only a few villages have them.

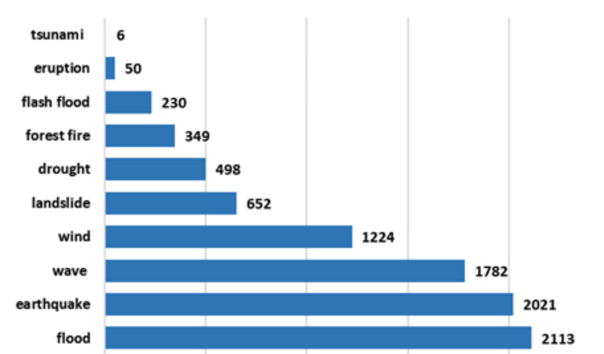

Figure 2. Number of disaster events

The object of the study is coastal villages, and the unit of analysis is an Indonesian province. The way the raw data transformed into a ready-for-analysis measure for each province is illustrated as follows.

- Province name = Aceh
- The number of coastal villages in Aceh (a) = 662
- The number of coastal villages in Aceh experienced earthquake in 2022 (b) = 100
- The percentage of coastal villages in Aceh experienced earthquake in 2022 (a/b in %) = 15.1%
- Earthquake disaster measure for Aceh (a/b*100) = 15.1

So, the measure for earthquake disaster in Aceh is 15.1, which is used for further analysis. By applying the same process, all measures used for data analysis are in the percentage of coastal villages in each province. These are relative measures and comparable among provinces.

The three variables and their measures is presented in Table II, that indicates measure's mean, minimum, and maximum percentages. The mean score in the second column indicates that floods are the most frequent disaster, with an average of 19%, while wind-related disasters are the least, with 9%. For disaster preparedness, on average, 7% of coastal villages in provinces have early warning systems (EWS) for tsunamis. The lower percentage of EWS tsunamis than general EWS is reasonable, as EWS tsunamis are more relevant in villages facing the Indian Ocean and Pacific Ocean but less in villages facing the sea or straits between islands. Furthermore, the highest average, 57%, for the disaster response variable is the availability of health community centers, including inpatient or outpatient care facilities.

C. Analytical Method

The research adopts a data mining methodology with the following generic stages: data pre-processing, modeling, evaluation, and visualization. The analysis adopts exploratory analytical methods to reveal the information from the dataset. The primary analytical methods for modeling are principal component analysis and cluster analysis.

Fig. 3 illustrates the framework of the analysis. Three analyses will be performed: (1) cluster analysis for all measures of three variables, (2) principal component analysis for each of the three variables, and (3) cluster analysis for all principal components obtained. Data mining was performed using the Python Scikit-learn operated in Google Collaboratory. Tableau analytical software was used for visualization, especially in creating quadrant diagrams and geographical heatmaps.

4. Results

Three types of planned analyses were conducted, and the results are presented here in three parts: cluster analysis, principal components analysis, and a combination of principal components analysis and cluster analysis.

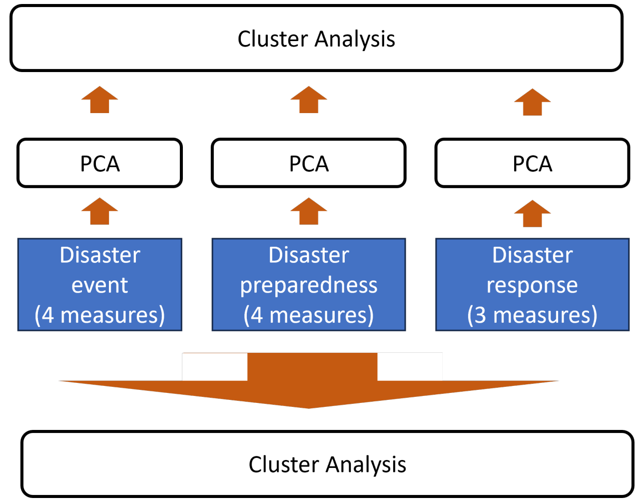

Figure 3. Framework of analysis

A. Classification from Cluster Analysis

This study applies the centroid-based clustering model with the k-means algorithm to group objects into k clusters based on some attributes. The principle of the k-means algorithm is to minimize the sum of square distances between the data and the related cluster centroid. A total of ten variables from disaster events, preparedness, and response become input for cluster analysis. Data were pre-processed with outlier treatment and normalization [0 – 1]. Then, clustering with the k-means algorithm was performed.

The vital question for k-means clustering is the number of clusters (k) to choose. One of the popular methods is the silhouette score, which indicates how well an object lies within its group (cohesiveness) compared to other groups (separation) [36]. The overall mean of the silhouette score for a certain k indicates the overall goodness-of-fit for a particular clustering model. The score varies between -1 and +1, where less than 0.2 is interpreted as a poor model, 0.2 and 0.5 as a fair model, and more than 0.5 as a good model [37].

Table III presents the average Silhouette scores for two to five cluster models. The table shows that the highest mean silhouette score is 0.39 for k=2. Therefore, cluster size is determined as two. Cluster analysis produces two clusters named Cluster A, which comprises six provinces, and Cluster B, which comprises 27 provinces. Table IV presents the normalized mean score of each measure for Clusters A and B. The mean scores for all 11 measures are higher for Cluster A than for Cluster B. Therefore, Cluster A has been named a high-intensity cluster because it has a higher intensity in disaster events, preparedness, and response than Cluster B, which is a low-intensity cluster.

Fig. 4 exhibits a scatter plot of wave disaster (wave), representing disaster events, versus EWS general (EWS), representing disaster preparedness. The figure shows the

TABLE II. Research variables and measures

Variable and measures	Mean (%)	Min (%)	Max (%)
Disaster event			
Flood disaster (flood)	19	4	39
Wave disaster (wave)	16	3	41
Earthquake disaster (earthq)	14	0	92
Wind disaster (wind)	9	1	48
Disaster preparedness			
Signs and evacuation routes (sign)	23	0	85
Safety equipment (safety)	16	3	73
EWS general (EWS)	15	0	100
EWS tsunami (EWStsu)	7	0	58
Disaster response			
Health Community Centers (HCC)	57	19	100
Drugstores (drugstore)	29	3	72
Private clinics (private)	18	0	67

TABLE III. Silhouette scores

number of clusters (k)	Silhouette scores
k = 2	0.39
k = 3	0.21
k = 4	0.21
k = 5	0.21

TABLE IV. Cluster characteristics

Cluster	flood	earthq.	wave	wind
A (6 provinces)	0.72	0.43	0.75	0.58
B (27 provinces)	0.34	0.27	0.24	0.32
Cluster	EWS	EWStsu	safety	sign
A (6 provinces)	0.79	0.82	0.67	0.78
B (27 provinces)	0.25	0.16	0.27	0.25
Cluster	HCC	drugstore	private	
A (6 provinces)	0.55	0.82	0.6	
B (27 provinces)	0.44	0.3	0.25	

distinct positions between the two clusters. The two-digit codes refer to province names in Table I. Six provinces of cluster A (in red) have high-intensity wave disasters (horizontal axis) and high availability of general early warning systems (EWS). On the contrary, 27 provinces of Cluster B (in green) have a low intensity of wave disasters and the availability of EWS.

Moreover, Fig. 5 plots the wave disaster (wave), representing disaster events, versus the private doctor's clinics (private), representing disaster response. Six provinces in cluster A (in red) tend to have a higher intensity of wave disasters and private doctor's clinics than provinces in cluster B (in green).

Fig. 6 displays the dispersion of two clusters through the EWS tsunami (EWStsu), representing disaster preparedness, versus the drugstore (drugstore), representing disaster response. Provinces in cluster A (red) tend to have a higher intensity of EWS tsunamis and drugstores than provinces in cluster B (green).

Fig. 7 depicts the geographic heatmap of both clusters. Cluster A (in blue) consists of six provinces: West Sumatera in Sumatera, West Java, Banten, Central Java, and Yogyakarta in Java, and Bali. Coastal villages in these provinces experienced more frequent overall disasters and have a higher capacity for disaster preparedness and response than those in Cluster B (in orange).

The investigation of the percentage of coastal villages in a province reveals that Cluster B has a higher percentage

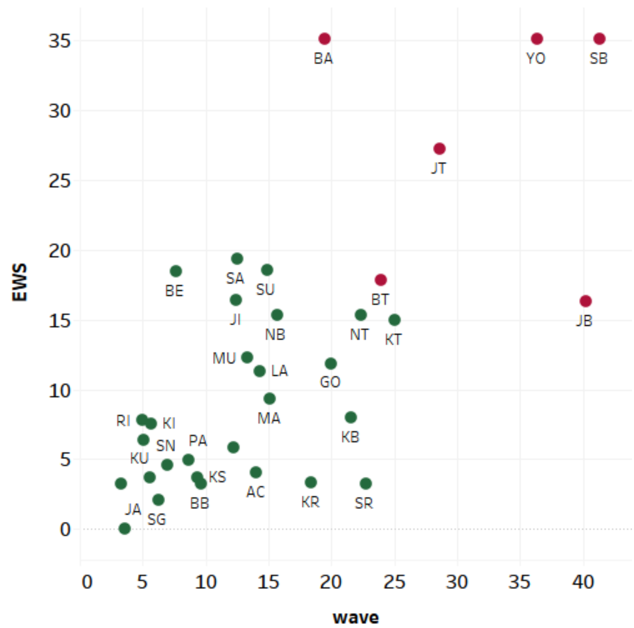

Figure 4. Scatter plot for wave disaster vs. EWS

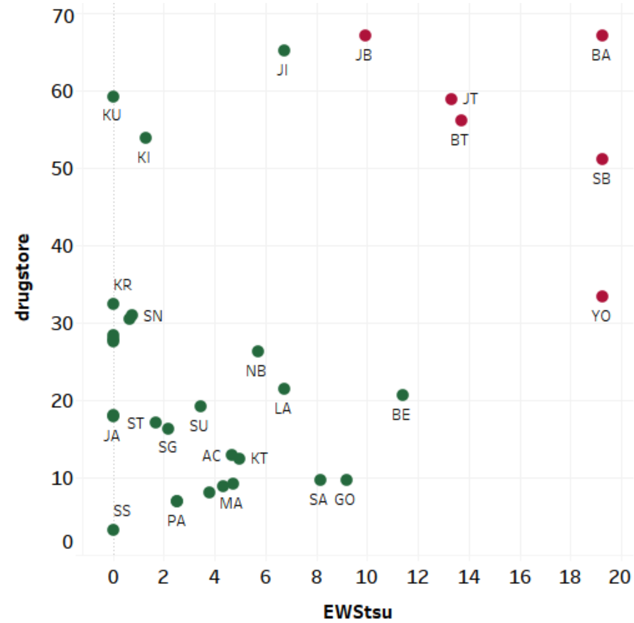

Figure 6. Scatter plot for EWStsunami vs. drugstore

Figure 5. Scatter plot for EWS vs. private doctor's clinic

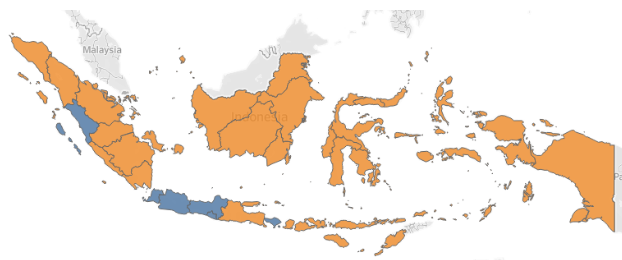

Figure 7. Geographic heatmap of two clusters

of coastal villages (25%) than Cluster A (10%). These numbers indicate that the provinces in Cluster B are more coastal regions than Cluster A. This finding is justified because the coastal provinces of Maluku, Sulawesi, and Riau Islands belong to Cluster B.

B. Classification from Principal Component Analysis

PCA was performed for three variables: disaster events, disaster preparedness, and disaster response. Data were pre-processed with outlier treatment and standardization. Standardization is a process of scaling the data so that the distribution has a mean score of 0 and a standard deviation of 1. By standardization, all measures will have an equal effect in creating principal components.

The main decision in conducting PCA is determining the number of principal components. Some standard methods suggest how many components to keep, such as proportion of variance, eigenvalue, and scree plot. All of them are based on the variance of the data set. This study adopts the proportion of variance by determining the cut-off at 90% to determine the number of principal components retained.

PCA needs to explain the contribution of the original variable to the related principal component. One of the methods is through eigenvectors describing the weight each original variable contributes to the corresponding principal component. Another method is through the correlations between each original variable and the corresponding principal component. The correlation coefficients ranging from +1 to -1 indicate which variables correlate most strongly with each principal component. This paper presents both to interpret the result of PCA.

1) PCA for Disaster Events

The PCA was executed to four measures of disaster events. Three primary components are needed to obtain the minimum 90% of variance. Table V presents the result of PCA with three principal components named disaster1, disaster2, and disaster3. The bottom row of the table indicates that disaster1 covers the most significant variance of 57%, followed by 20% for disaster2 and 15% for disaster3. These three contributed 92% of the total variance.

Table V presents two parts: loadings (eigenvectors) and correlation coefficients. First, PCA loadings are the coefficients of the linear combination of the original variables (measures) from which the principal components (PCs) are constructed. Second, the correlation scores specify the correlation coefficient between the original variables (four measures of disaster events) and the principal component (disaster1 and disaster2). Examining the values of the principal components confirms that each component has zero mean scores and that the correlation between the components is zero (orthogonal).

PCA loadings could be illustrated in diagrams. Fig. 8 displays the four loading vectors of disaster events in the disaster1 vs. disaster2 diagram. PCA loadings indicate how much (the weight) each measure contributes to the corresponding principal components of disaster1 and disaster2. The projection of each vector to the horizontal or vertical axis denotes the weight contribution to each component. The projection of the vector flood to the horizontal axis indicates that the weight contribution of the flood is about 0.6 to the disaster1 component. The contribution of earthquake is about 0.4 to disaster1. The projection to the vertical axis indicates the contribution of each measure to disaster2. For example, earthquake is a dominant factor for disaster2.

Returning to Table V, the correlation scores above 0.5 are written in bold. The first principal component, disaster1, strongly correlated (0.56 – 0.88) with all four measures of disaster events. This component increases with increasing floods, wind, wave-related disasters, and earthquakes. Among the four variables, the correlation of earthquakes is the most minor (0.56). The second principal component, disaster2, has a strong correlation with earthquake disasters and weak correlations with the other measures. Therefore, disaster2 is likely to represent an earthquake disaster. Similarly, the third principal component, disaster3, represents a wind-related disaster. Three components are uncorrelated; therefore, we can interpret that disaster events could be differentiated into three: (1) mixed disasters, (2) earthquake disasters, and (3) wave-related disasters.

Classifying regions using principal components could be made with one or more dimensions. For one dimension, a geographic heatmap exposes the intensity of a particular principal component among regions. For example, Fig. 9 displays the map of provinces for the intensity of disaster1. Darker areas indicate higher disaster intensity. West Sumatra and West Java appear to have high disaster intensity, primarily floods, wind, and waves. Areas with low disaster intensity have lighter colors, such as Papua, West Papua, and South Sumatra.

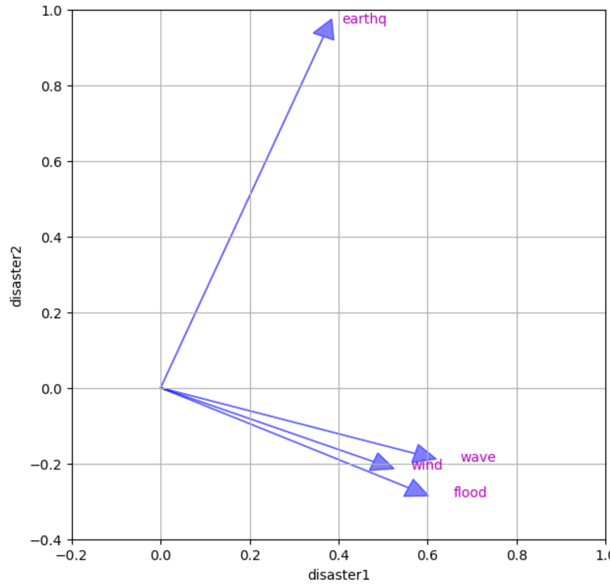

Figure 8. PCA loadings of disaster1 and disaster2

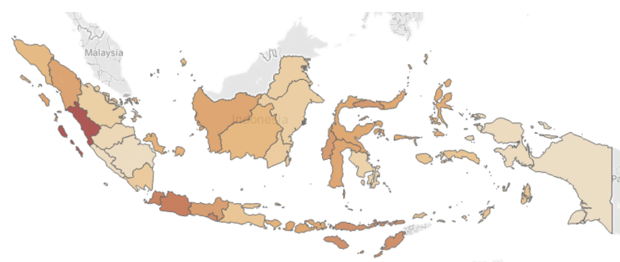

Figure 9. Geographic heatmap for disaster1

Furthermore, regional classification in two dimensions can be done using a scatter plot diagram. Fig. 10 displays a scatter plot with disaster1 vs disaster2. Four quadrants were created based on the average score of each principal component. The lower left quadrant includes ten provinces with low percentages of both types of disasters. This group can be interpreted as having 'good' conditions because the disaster intensity is low. Several provinces in this group include Riau Island (KR), Bangka Belitung Island (BB), North Kalimantan (KU), and Central Kalimantan (KT). The lower right quadrant contains provinces with high-intensity floods, winds, and waves type-disaster but low-intensity earthquakes. In this group, some provinces are West Java (JB), Banten (BT), Yogyakarta (YO), and East Nusa Tenggara (NT). The upper left quadrant comprises provinces with high-intensity earthquakes and low disasters related

TABLE V. PCA for disaster events

Disaster variable	disaster1	disaster2	disaster3
loadings			
flood	0.56	-0.26	-0.39
wind	0.48	-0.2	0.86
wave	0.57	-0.17	-0.34
earthq.	0.37	0.93	0.01
correlation			
flood	**0.85**	-0.24	-0.31
wind	**0.73**	-0.18	0.68
wave	**0.88**	-0.16	-0.27
earthq	**0.56**	**0.85**	0
variance	0.57	0.2	0.15

to floods, waves, and winds. Some provinces included in this group are West Papua (PB), East Java (JI), and Aceh (AC). The upper right quadrant contains six provinces with high intensity of both types of disasters, such as North Maluku (MU), North Sumatera (SU), North Sulawesi (SA), and West Sumatra (SB).

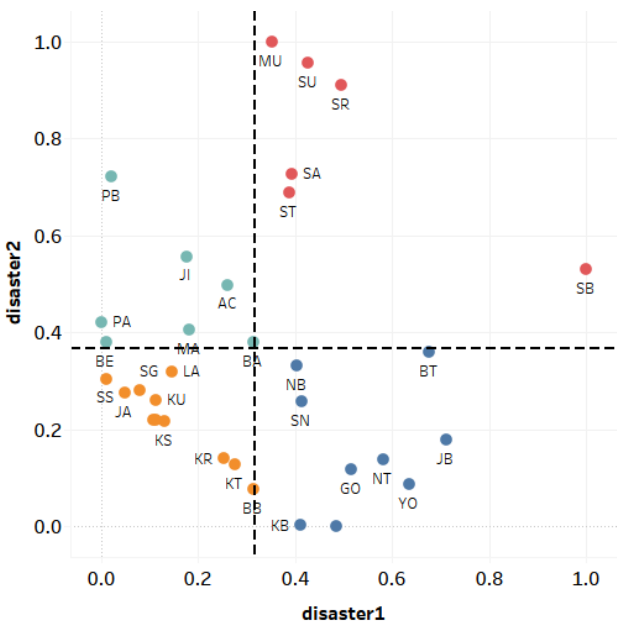

Figure 10. Quadrant diagram of disaster1 vs. disaster2

The following scatter plot visualizes provinces in disaster2 vs. disaster3, shown in Fig. 11. Disaster2 indicates the intensity of earthquakes, while disaster3 is a wind-related disaster. The upper right quadrant comprises provinces with high intensity of earthquake and wind disasters, such as North Sumatera (SU), North Sulawesi (SA), and West Sumatera (SB). The lower left quadrant shows that West Java (JB) and West Kalimantan (KB) belong to provinces with low earthquake and wind disasters. In summary, this classification through the quadrant diagram provides a simple visualization of the regional grouping.

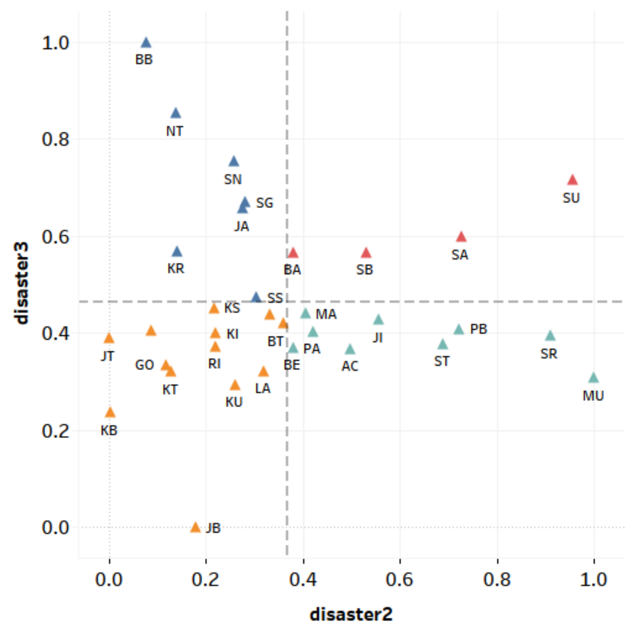

Figure 11. Quadrant diagram of disaster2 vs. disaster3

2) *PCA for Disaster Preparedness*

PCA was performed for four measures of disaster preparedness. Two principal components are needed to obtain the minimum 90% of variance. Table VI presents the result of PCA with two principal components named prepare1 and prepare2, with 78% and 17% variances. Based on the loading scores of prepare1, the weight contribution of the general EWS, the EWS for tsunami, and the signage/evacuation route are medium (around 0.5), while the safety equipment is low. However, prepare2 is dominated by safety equipment. In addition to the loading values, Table VI shows the strength of the correlation of each measure to both principal components. The correlation values of three measures (EWS for tsunami, general EWS, and signage/evacuation) to prepare1 are substantially high (¿0.9). On the contrary, safety equipment strongly correlates (0.76) with the disaster2 component. Therefore, prepare1 could be interpreted as high intensity of EWS, EWS tsunami, and signage/evacuation route, and prepare2 as high intensity of safety equipment.

Fig. 12 exhibits the geographic heatmap of provinces for the intensity of prepare1. Darker areas indicate regions with better preparedness, including Bali and Yogyakarta, as the top tourist destinations. A plausible explanation is that those tourist destinations are more prepared to anticipate disasters. Lighter color areas indicate less capacity for disaster preparedness. Some low-intensity provinces are Sumatra (Jambi, South Sumatera), Sulawesi (West, South, and South-East Sulawesi), and Papua/ West Papua.

Figure 12. Geographic heatmap for the prepare1 component

Furthermore, regional classification in two dimensions, prepare1 vs. prepare2, is visualized in a scatter plot shown in Fig. 13. Four quadrants, based on the average value, classify provinces into four groups (clusters). The concentration of provinces appears on the left side of the vertical line as the average of prepare1. This concentration means that most provinces have a lower capacity for disaster preparedness (prepare1). Among regions with a high capacity of prepare1 but a low capacity of prepare2, the figure indicates West Sumatera (SB). As described above, Yogyakarta (YO) and Bali (BA) have a higher capacity for both preparedness dimensions.

3) PCA for Disaster Response

PCA was performed for disaster response for three measures of health facilities. Two principal components are to keep to obtain more than 90% of the variance explained. Table VII presents a summary of the PCA results. The first component (response1) covers 71%, and the second (response2) covers 25%, so the total variance explained is 96%. PCA loadings indicate that the response1 component is contributed primarily by the drug store and the doctor's clinic. Community health centers primarily contribute to the disaster2. Table VII indicates that the strength of the correlation between drug stores (drugstore) and private physicians' clinics (private) with response1 is significantly high (¿0.9), while community health center (HC) is the leading factor (0.79) for response2.

Fig. 14 exhibits the geographic heatmap for one-dimension classification using the response1 primary component. Darker areas indicate a higher capacity of health facilities in coastal villages. Some provinces are Bali, North

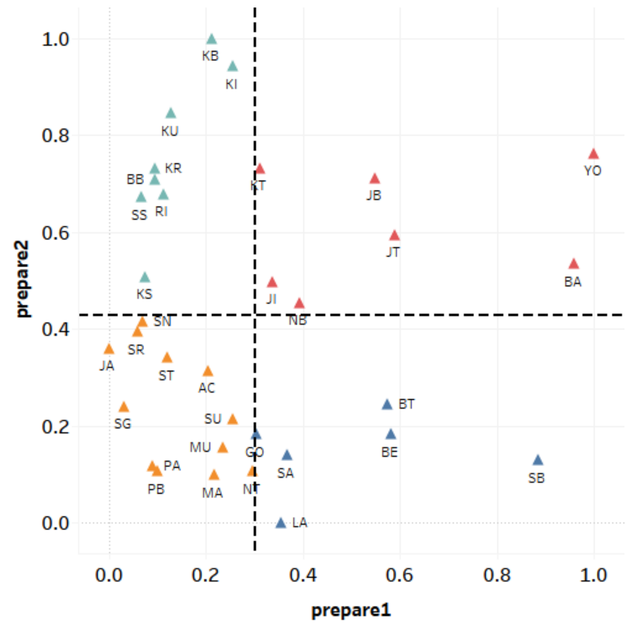

Figure 13. Quadrant diagram of prepare1 vs. prepare2

Kalimantan, and East Kalimantan. Conversely, the lighter color area indicates provinces with less capacity for health facilities in coastal villages. Several provinces are Aceh, South Sumatra, Papua, West Papua, Maluku, and Gorontalo.

Figure 14. Geographic heatmap for the response1 component

Furthermore, regional classification through two-dimensional diagrams is presented in Fig. 15. Provinces in the upper right quadrant have high preparedness (high response1 and response2), especially Yogyakarta (YO) and East Kalimantan (KI). Conversely, the bottom left quadrant comprises provinces with low preparedness, especially South Sumatra (SS), Aceh (AC), and South Kalimantan (KS).

4) Summary of PCAs

The total seven principal components have been produced from multiple PCAs for disaster events, preparedness, and response. Each principal component could identify regions at the top or bottom of that measure. Table VIII presents seven principal components and the top three provinces in each of them. For example, West Sumatra, as the top province in disaster 1, is an area with a high

TABLE VI. PCA for disaster preparedness

Preparedness variable	prepare1	prepare2
loadings		
EWS	0.54	-0.09
EWStsu	0.55	-0.22
safety	0.37	0.91
sign	0.52	-0.33
correlation		
EWS	**0.97**	-0.07
EWStsu	**0.98**	-0.18
safety	**0.67**	**0.76**
sign	**0.93**	-0.28
variance	0.78	0.17

TABLE VII. PCA for disaster response

Response variable	response1	response2
loadings		
HCC	0.43	0.9
drugstore	0.63	-0.38
private	0.65	-0.22
correlation		
HCC	**0.63**	**0.79**
drugstore	**0.93**	-0.34
private	**0.97**	-0.19
variance	0.71	0.25

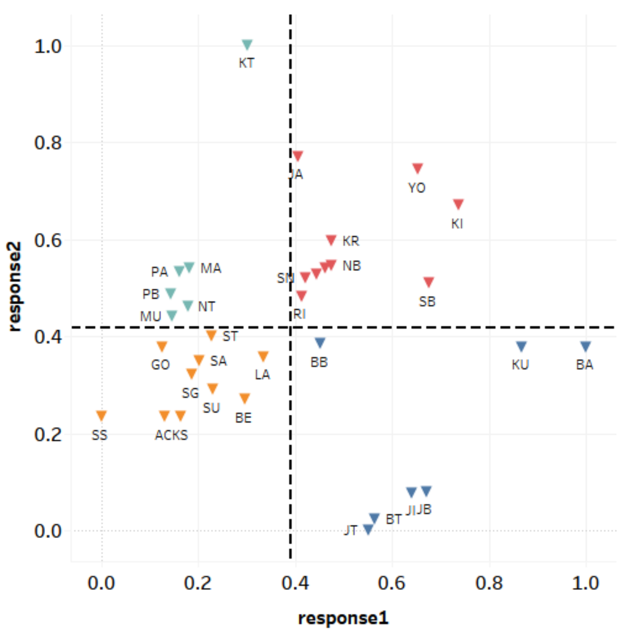

Figure 15. Quadrant diagram of response1 vs. response2

risk of combined disasters. the top province for disaster response. Ultimately, the classification of provinces using PCA is flexible depending on what measures to choose.

C. Classification from the Combination of PCA+CA

This part investigates cluster analysis (CA) using the principal components of three variables, performed in section B, as input. The optimum number of clusters was investigated through the average Silhouette scores. Table IX presents the silhouette scores and cluster size from the k-means method for k=2 to k=5. The highest Silhouette score is 0.24 for k=3. As described above, the silhouette score between 0.2 and 0.5 is a fair model [31]. The highest score of 0.24 is lower than the score of 0.39 obtained from cluster analysis in Section A. The low score indicates that the clusters obtained are weakly separated.

Table IX shows that k-means with k=3 produces three clusters comprising 6, 10, and 17 provinces. Investigation of the provinces within each cluster revealed that six provinces in the first cluster are the same as those in cluster A from the cluster analysis in Section A. Therefore, three clusters are named A, B1, and B2, which refer to Cluster A and B from the previous clustering (Section A). Cluster B1 scores highest for the remaining principal components, except disaster2.

Table X presents the normalized mean scores of seven principal components for each cluster. Cluster A has the highest mean scores in disaster1, preapre1, and response1. Cluster B1 has higher mean scores for the second principal component of preparedness and response than Cluster B2.

TABLE VIII. The top three provinces

Principal component	Province 1	Province 2	Province 3
disaster1	West Sumatra	West Java	Banten
disaster2	North Maluku	North Sumatra	West Sulawesi
disaster3	Bangka Belitung	East Nusa Tenggara	South Sulawesi
prepare1	Yogyakarta	Bali	West Sumatra
prepare2	West Kalimantan	East Kalimantan	North Kalimantan
response1	Bali	North Kalimantan	East Kalimantan
response2	Central Kalimantan	Jambi	Yogyakarta

TABLE IX. Silhouette scores

number of clusters (k)	Silhouette scores	Cluster size
k = 2	0.23	15,18
k = 3	**0.24**	**6,10,17**
k = 4	0.20	6,6,7,14
k = 5	0.20	2,5,6,10,10

TABLE X. Cluster characteristics

cluster	disaster1	disaster2	disaster3	prepare1	prepare2	response1	response2
A (6 provinces)	**0.64**	0.26	0.39	**0.76**	0.50	**0.69**	0.29
B1 (10 provinces)	0.25	0.19	**0.50**	0.17	**0.69**	0.50	**0.59**
B2 (17 provinces)	0.25	**0.51**	0.47	0.22	0.25	0.22	0.36

Furthermore, three clusters of provinces are plotted in a two-dimensional diagram with prepare1 vs. response1, as shown in Fig. 16. Six provinces of cluster A (in red) sit in the upper right quadrant. Cluster B1 (in yellow) resides in the upper left quadrant, and Cluster B2 (in green) in the lower left quadrant. Provinces in clusters B1 and B2 seem differentiated only by response1, not by prepare1.

Fig. 17 displays a geographic map for three clusters. Cluster A (in blue), with the highest intensity of disaster events, preparedness, and response, consists of provinces mainly in Java and Bali. Cluster B1 (in yellow), with the second highest intensity of disaster events, preparedness, and response, covers mainly the provinces of Kalimantan. Cluster B2 (in red), with the lowest intensity, covers most of the provinces in Papua, Sulawesi, and Sumatra.

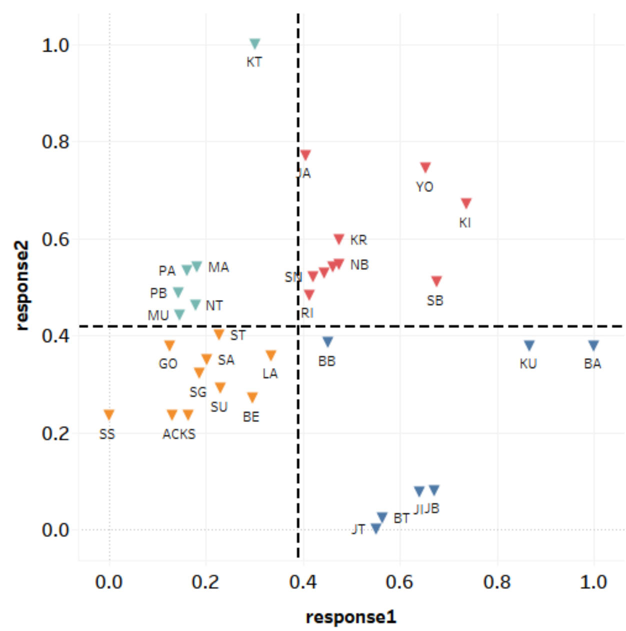

Figure 16. Scatter plot for prepare1 vs. response1

5. Discussion and Conclusion

This study investigated disaster events, preparedness, and response in coastal villages in Indonesian provinces. This research contributes to both disaster management and data analytics literature. For the first area, this research is one of the few studies that investigates coastal villages on a national scale. Most prior studies focused on certain regions or villages (e.g., [8] [9] [10]). For the second area, this study is one of the few studies that applies data analytics for published official data. This study provides a case for exposing rich information from the official data.

The results indicate that, first, the classification of provinces using cluster analysis provides a rigid and simple classification based on disaster management characteristics. Second, principal component analysis becomes an alternative method to classify provinces based on selected principal components (new variables). The application of PCA may

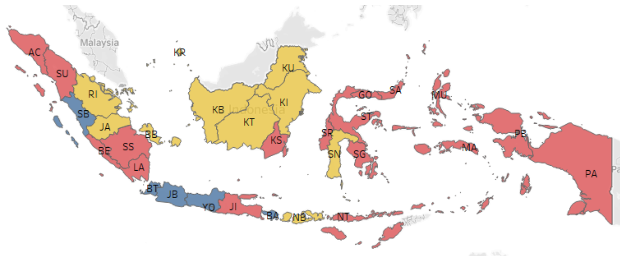

Figure 17. Cluster map

guide to fuller and more informative classification. Third, the combination of principal component analysis and cluster analysis provides an alternative method to cluster analysis alone. This third method is comparable to prior studies [33] [38]. However, the result shows that the cluster model from the cluster analysis alone indicates a better separation between clusters than the combination of PCA and cluster analysis. Therefore, cluster analysis and principal component analysis might be used independently.

Disaster studies are unique and depend on the scope of the disaster model, the type of disaster, object location, data, and analysis methods. Comparison of results between one study and another cannot be firm because of these differences. The method used in this study could be compared partially to other disaster studies. First, this study could be placed among studies that use disaster management phases, such as [39] focused on disaster preparedness and [40] focused on disaster risk reduction. However, the results could not be compared because of different locations, data sources, measures, and analysis methods. Second, this study could be placed among studies that used secondary data and data analytics methods, such as [41] using the typhoon data and [42] using Twitter feeds data from the web and social networks with big data analytics. However, the result could not be compared because of different data sources, measures, and analysis methods. In conclusion, this and other disaster studies are unique and context-based.

The provincial government might use the province classification produced in this study to determine a course of actions in disaster management. The results of this study suggest improvements in coastal village disaster preparation and response, particularly in the province with a high percentage of coastal villages. Although tsunamis are dangerous to coastal villages, data shows that earthquakes are the most frequent disasters. The meteorological and geophysics agencies should provide timely early warning information to society as part of disaster preparedness. Search, rescue, and medical facilities should be improved to reduce the impact of a disaster as a part of disaster response. Finally, this study recommends that governments and other stakeholders use data analytics to support decisions in disaster management and create resilient and sustainable coastal villages.

The primary limitation in this study is that the analysis is based solely on official statistics for a particular year. Data for a different year might produce different results. Furthermore, since the available data objects are provinces, information about which particular coastal villages and regencies could not be obtained.

Further study in the Indonesian context may apply this study method to the dataset for subsequent years so that the change in disaster events, preparedness, and response among regions could be observed. In addition, further studies may be conducted to extend and adapt this method to any country with similar official disaster statistics.

References

[1] K. D. Dissanayaka, N. Tanaka, and T. L. Vinodh, "Integration of Eco-DRR and hybrid defense system on mitigation of natural disasters (Tsunami and Coastal Flooding): a review," *Natural Hazards*, vol. 110, pp. 1–28, 2022.

[2] M. A. Thiri, "Uprooted by tsunami: A social vulnerability framework on long-term reconstruction after the Great East Japan earthquake," *International Journal of Disaster Risk Reduction*, vol. 69, 2022.

[3] S. M. Kao, W. N. Wu, and M. E. Gu, "Coastal disaster protection and hazard perceptions of the residents of Orchid Island, Taiwan," *Marine Policy*, vol. 148, 2023.

[4] G. A. Parvin, R. Dasgupta, M. A. Abedin, M. Sakamoto, B. Ingirige, M. G. Kibria, K. Fujita, M. Basu, R. Shaw, and H. Nakagawa, "Disaster experiences, associated problems and lessons in southwestern coastal Bangladesh: exploring through participatory rural appraisal to enhance resilience," *Sustainable and Resilient Infrastructure*, vol. 8, no. S1, pp. 223–236, 2023.

[5] P. Cheewinsiriwat, U. Langkulsen, V. Lertwattanamongkol, W. Poompongthai, A. Lambonmung, C. Chamchan, S. Boonmanunt, K. Nakhapakorn, and C. Moses, "Assessing Coastal Vulnerability to Climate Change: A Case Study of Nakhon Si Thammarat and Krabi," *Social Sciences*, vol. 13, no. 3, 2024.

[6] S. Zhang, J. Zhang, X. Li, X. Du, T. Zhao, Q. Hou, and X. Jin, "Estimating the grade of storm surge disaster loss in coastal areas of China via machine learning algorithms," *Ecological Indicators*, vol. 136, 2022.

[7] J. L. Barros, P. P. Santos, A. O. Tavares, P. Freire, A. B. Fortunato, A. Rilo, and F. S. Oliveira, "The complexity of the coastal zone: Definition of typologies in Portugal as a contribution to coastal disaster risk reduction and management," *International Journal of Disaster Risk Reduction*, vol. 86, no. February, 2023.

[8] H. P. Rahayu, R. Haigh, D. Amaratunga, B. Kombaitan, D. Khoirunnisa, and V. Pradana, "A micro scale study of climate change adaptation and disaster risk reduction in coastal urban strategic planning for the Jakarta," *International Journal of Disaster Risilience in the Built Environment*, vol. 11, no. 1, pp. 119–133, 2020.

[9] I. Rudiarto, W. Handayani, and J. S. Setyono, "A regional perspective on urbanization and climate-related disasters in the northern coastal region of Central Java, Indonesia," *Land*, vol. 7, no. 34, 2018.

[10] H. Andreas, H. Z. Abidin, D. A. Sarsito, and D. Pradipta, "Adaptation of 'early climate change disaster' to the Northern coast of Java Island Indonesia," *Engineering Journal*, vol. 22, no. 3, 2018.

[11] F. Atwii, L. Kirch, B. Paragi, and K. Radtke, "World Risk Report 2022," Bündnis Entwicklung Hilft, Tech. Rep., 2022. [Online]. Available: http://www.weltrisikobericht.de/uploads/media/WRR_2012_en_online.pdf

[12] S. Goswami, S. Chakraborty, S. Ghosh, A. Chakrabarti, and B. Chakraborty, "A review on application of data mining techniques to combat natural disasters," *Ain Shams Engineering Journal*, vol. 9, no. 3, pp. 365–378, 2018.

[13] A. Bucherie, C. Hultquist, S. Adamo, C. Neely, F. Ayala, J. Bazo, and A. Kruczkiewicz, "A comparison of social vulnerability indices specific to flooding in Ecuador: principal component analysis (PCA) and expert knowledge," *International Journal of Disaster Risk Reduction*, vol. 73, p. 102897, 2022.

[14] World Population Review, "Countries by Coastlines 2023," 2023. [Online]. Available: https://worldpopulationreview.com/country-rankings/countries-by-coastline

[15] W. Handayani, I. Rudiarto, J. S. Setyono, U. E. Chigbu, and A. M. Sukmawati, "Vulnerability assessment: A comparison of three different city sizes in the coastal area of Central Java, Indonesia," *Advances in Climate Change Research*, vol. 8, no. 4, pp. 286–296, 2017.

[16] N. M. Zafri, M. Nurullah, M. N. Neema, and M. Waliullah, "Spatial accessibility to healthcare facilities in coastal region of Bangladesh," *The International Journal of Health Planning and Management*, vol. 36, no. 3, pp. 643–655, 2017.

[17] United Nations, "Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction," Tech. Rep. December, 2016.

[18] K. Furtak and A. Wolińska, "The impact of extreme weather events as a consequence of climate change on the soil moisture and on the quality of the soil environment and agriculture – A review," *Catena*, vol. 231, no. July, 2023.

[19] J. Cels, T. Rossetto, A. W. Little, and P. Dias, "Tsunami preparedness of coastal Sri Lankan schools," in *Proceedings of the International Conference on Natural Hazards and Infrastructure*, no. July, 2022.

[20] B. S. Ngcamu, "Application of the disaster management cycle and climate change: Studying flood disasters in South Africa," *Social Sciences and Humanities Open*, vol. 8, no. 1, 2023.

[21] A. H. Aghapour, M. Yazdani, F. Jolai, and M. Mojtahedi, "Capacity planning and reconfiguration for disaster-resilient health infrastructure," *Journal of Building Engineering*, vol. 26, no. June, 2019.

[22] F. Efendi, R. Indarwati, G. E. Aurizki, I. A. Susanti, and A. E. F. Maulana, "Policymakers' perspectives on responding to the elderly's mental health needs in post-disaster situations," *Journal of Public Health Research*, vol. 11, no. 1, 2022.

[23] F. Kherif and A. Latypova, "Principal component analysis," in *Machine Learning: Methods and Applications to Brain Disorders*, 2020, vol. 1, ch. 12, pp. 209–225.

[24] M. Tripathi and S. K. Singal, "Use of Principal Component Analysis for parameter selection for development of a novel Water Quality Index: A case study of river Ganga India," *Ecological Indicators*, vol. 96, no. Part 1, pp. 430–436, 2019.

[25] I. Turbay, P. Ortiz, and R. Ortiz, "Statistical analysis of principal components (PCA) in the study the vulnerability Heritage Churches," *Procedia Structural Integrity*, vol. 55, pp. 168–176, 2024.

[26] A. T. W. Almais, A. Susilo, A. Naba, M. Sarosa, C. Crysdian, I. Tazi, M. A. Hariyadi, M. A. Muslim, P. M. N. S. A. Basid, Y. M. Arif, M. S. Purwanto, D. Parwatiningtyas, Supriyono, and H. Wicaksono, "Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters," *IEEE Access*, vol. 11, no. May, pp. 74 590–74 601, 2023.

[27] K. Mitsushita, S. Murakoshi, and M. Koyama, "How are various natural disasters cognitively represented?: a psychometric study of natural disaster risk perception applying three-mode principal component analysis," *Natural Hazards*, vol. 116, pp. 977–1000, 2023.

[28] Analytixlabs, "What is Clustering in Machine Learning: Types and Methods," 2022. [Online]. Available: https://www.analytixlabs.co.in/blog/types-of-clustering-algorithms/#Types_of_Clustering_Algorithms

[29] D. D. Atsa'am, T. Gbaden, and R. Wario, "A machine learning approach to formation of earthquake categories using hierarchies of magnitude and consequence to guide emergency management," *Data Science and Management*, vol. 6, no. 4, pp. 208–213, 2023.

[30] D. Mehta, J. Dhabuwala, S. M. Yadav, V. Kumar, and H. M. Azamathulla, "Improving flood forecasting in Narmada river basin using hierarchical clustering and hydrological modelling," *Results in Engineering*, vol. 20, no. November, pp. 1–13, 2023.

[31] X. Tu, C. Fu, A. Huang, H. Chen, and X. Ding, "DBSCAN Spatial Clustering Analysis of Urban "Production–Living–Ecological" Space Based on POI Data: A Case Study of Central Urban Wuhan, China," *International Journal of Environmental Research and Public Health*, vol. 19, no. 9, 2022.

[32] T. G. Penkova, "Principal component analysis and cluster analysis for evaluating the natural and anthropogenic territory safety," in *Procedia Computer Science*, vol. 112. Elsevier B.V., 2017, pp. 99–108.

[33] M. Rais, R. Goejantoro, and S. Prangga, "Optimization of K-Means Cluster with Principal Component Analysis on the Grouping of Districts/Cities on the Island of Kalimantan based on Unemployment Rate Indicator," *Exponensial*, vol. 12, no. 2, pp. 129–136, 2021.

[34] A. E. Haryati and Sugiyarto, "Clustering with Principal Component Analysis and Fuzzy Subtractive Clustering Using Membership Function Exponential and Hamming Distance," *IOP Conference Series: Materials Science and Engineering*, vol. 1077, no. 1, p. 012019, 2021.

[35] BPS, "Statistik Sumber Daya Laut dan Pesisir 2022," Badan Pusat Statistik, Tech. Rep., 2022.

[36] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, no. C, pp. 53–65, 1987.

[37] I. L. Corporal-Lodangco, M. B. Richman, L. M. Leslie, and P. J. Lamb, "Cluster analysis of North Atlantic tropical cyclones," *Procedia Computer Science*, vol. 36, pp. 293–300, 2014.

[38] G. T. Nguyen, S. Y. Liaw, and X. L. Duong, "Readiness of SMEs

for Adopt Big Data: An Empirical Study in Vietnam," *International Journal of Computing and Digital Systems*, vol. 12, no. 1, pp. 509–521, 2022.

[39] M. C. Torrentira Jr. and M. Makilan, "Framework Of Disaster Preparedness Among Coastal Communities In Davao City, Philippines," *International Journal of Food Science and Agriculture*, vol. 2, no. 4, pp. 84–91, 2018.

[40] L. Cumiskey, S. Priest, N. Valchev, C. Viavattene, S. Costas, and J. Clarke, "A framework to include the (inter)dependencies of Disaster Risk Reduction measures in coastal risk assessment," *Coastal Engineering*, vol. 134, no. July 2017, pp. 81–92, 2018.

[41] S. Q. Chen and J. Bai, "Data-driven decision-making model for determining the number of volunteers required in typhoon disasters," *Journal of Safety Science and Resilience*, vol. 4, no. 3, pp. 229–240, 2023.

[42] F. Sufi, "A decision support system for extracting artificial intelligence-driven insights from live twitter feeds on natural disasters," *Decision Analytics Journal*, vol. 5, 2022.

Gunawan Gunawan is a senior lecturer at the Industrial Engineering Department, University of Surabaya, Indonesia. His current research area is data analytics. He completed a Ph.D. in information systems from Loughborough University - UK, an MBA in technology management from the Asian Institute of Technology - Thailand, and a B.Sc. in physics from the Sepuluh Nopember Institute of Technology - Indonesia.