2023 IEEE 9th Information Technology International Seminar (ITIS)
Batu Malang, Indonesia, October 18 – 20, 2023

Clustering Coping Capacity using HCSS

Ani Apriani
Geological Engineering Department
Institute Teknologi Nasional Yogyakarta
Yogyakarta, Indonesia
aniapriani@itny.ac.id

Iwan Tri Riyadi Yanto
Information System Department
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
yanto.itr@is.uad.ac.id

***Abstract*—Coping capacity, or the capability of individuals, communities, and societies to successfully respond to and recover from bad events, is a critical component of disaster risk reduction. Formulating successful disaster preparedness, response, and recovery plans requires understanding coping capacity. Building resilient communities that can face landslide difficulties and lowering susceptibility needs assessing and improving landslide coping capacity. Clustering based on the soft set method is used in this study to identify separate groups of communities with comparable coping ability features and to enable focused interventions and resource allocation. The experiment provides 3 clusters based on the Dunn Index value. The disaster risk knowledge score of C1 and C2 are obtained at 3 on average, while C2 is obtained at 1.67. The mitigation knowledge scores for C1,c2 and C2 are obtained 1.67, 3 and 2, respectively. Investigators or the government can use this to provide recommendations to determine different treatments and create sensible regulations to promote catastrophe risk reduction and neighbourhood safety in the area.**

Keywords—Coping Capacity; Landslide; Clustering; Soft set.

I. INTRODUCTION

Coping capacity is the ability of individuals, groups, and society to effectively respond to and recover from unpleasant events, shocks, or disasters. It is a multidimensional notion that incorporates many factors, resources, and capacities that enable individuals and communities to withstand, adapt to, and recover from the effects of catastrophes or crises. [1].

Physical infrastructure, social networks, economic resources, knowledge, skills, and governance systems are all components of coping capability. These characteristics jointly affect a community's ability to foresee, absorb, and recover from a disaster's effects, lowering vulnerability and increasing resilience [2].

Understanding coping ability is critical for developing successful preparedness, response, and recovery strategies in the context of catastrophe risk reduction and management. Communities with greater coping capacity are better prepared to handle and mitigate the adverse effects of disasters, resulting in less loss of life, property, and livelihoods. [3].

The ability of a community or region to effectively respond to and recover from landslide disasters is referred to as landslide coping capacity. It includes various elements and resources that help mitigate the impact of landslides and facilitate a quick and efficient recovery. [4].

It is critical for disaster risk reduction initiatives to assess and improve landslide coping capacity. Policymakers and disaster management organizations may conduct targeted interventions, distribute resources more efficiently, and build more resilient communities capable of dealing with the problems posed by landslides by identifying the strengths and weaknesses of a community's coping capacity [5].

One of the challenges that may have various coping capability profiles due to varying socioeconomic, geographical, and cultural contexts is heterogeneity among communities or different communities. Clustering enables us to identify discrete groupings of communities with comparable coping capacity features, allowing us to take a more customized approach to disaster risk reduction. [6], [7].

Clustering algorithms are unsupervised learning methods used in data analysis and machine learning to group comparable data points based on similarities or characteristics. These techniques seek to uncover underlying patterns and structures in datasets and organize data into clusters, with data points within one cluster being more similar to those in other clusters. Pattern recognition, picture segmentation, data mining, and customer segmentation are all examples of where clustering is applied [8].

Several clustering methods have been proposed. Fuzzy k-modes were proposed by Xu et al. et al. [10]. The matching dissimilarity metric is used. Kim et al. [11improved the performance of fuzzy k-modes by substituting fuzzy centroids (FC) with hard centroids due to the potential for artefacts associated with the use of hard centroids. It is a non-parametric strategy that works by reducing the sum of squared errors within clusters. The Fuzzy k-partitioning (FkP) technique, developed by Miin-Shen et al. [12], is a parametric approach based on the likelihood function of multivariate multinomial distributions.

Furthermore, the FkP technique can be viewed as a fuzzy-based clustering algorithm given categorical data. In contrast, almost all previously reported fuzzy categorical data clustering techniques describe data sets as binary values. In contrast, categorical data has multi-valued qualities that can be represented as multi-soft [13]. When displaying categorical data without converting it to binary values, using a multi-soft set for multi-valued attributes provides advantages. Yanto et al. introduced HCSS, a clustering technique for categorical data via multinomial distribution based on soft set theory, based on these advantages [9]. As a result, this research aims to experiment to test the feasibility of grouping the Coping capacity dataset using HCSS.

II. LITERATURE REVIEW

A. Coping capacity

The ability of a community or region to effectively respond to and recover from landslide disasters is referred to as landslide coping capacity. [14]. Coping capacity is a multifaceted concept that considers various characteristics and resources that might assist communities in mitigating the

979-8-3503-0683-5/23/$31.00 ©2023 IEEE

impact of landslides and recovering more quickly and efficiently when such occurrences occur. [15].

A community's coping capacity in landslides may include the following factors.:

1. Early Warning Systems: Effective early warning systems can offer communities and authorities timely warnings about potential landslide dangers, allowing evacuation and preparedness steps.
2. Retaining walls, slope stabilization techniques, and drainage systems, among other things, can lessen a community's vulnerability to landslides.
3. Emergency Response and Preparedness: Having well-prepared emergency response plans and trained personnel can help to ensure efficient and coordinated activities during and after a landslide.
4. Social Networks and Communication: During a landslide, strong social networks and community engagement can aid in the spread of information and resources.
5. Financial Resources: Access to financial resources and insurance can help communities rebuild and restore their livelihoods following a landslide.
6. Land Use Planning: Proper land use planning can assist building and development in avoiding high-risk landslide zones, and decreasing exposure to potential dangers.
7. Knowledge and Awareness: Individuals and communities can be empowered to proactively reduce risks by increasing their awareness and knowledge about landslides, including their causes and preventive methods.

B. *Hard Clustering Based on Soft Set (HCSS)*

The data is analyzed using the clustering technique based on a soft set. The method uses a multinomial distribution function to find the highest probability and multi-soft set based on decomposing the data into several sets with similar values [9].

Let $S = (U, A, V, f)$ be a categorical-valued information system, where $U = \{u_1, u_2, \dots, U_n\}$ is a finite set of instances, $A = \{a_1, a_2, \dots a_m\}$ is a finite set of attributes, V is values set of each attribute A, f is a mapping function

$f\colon (U, A) \to V$ and $S = \left(U, \quad ai, V_{a_i}, f\right), i = 1,2,\cdots,|A|$ Boolean-valued information system, it can be decomposed to be a multi-boolean information system as

$$S = (U, A, V, f) = \begin{cases} S^1 = \left(U, a_1, V_{\{0,1\}}, f\right) \Leftrightarrow (F, a_1) \\ S^2 = \left(U, a_2, V_{\{0,1\}}, f\right) \Leftrightarrow (F, a_2) \\ \qquad \vdots \\ S^{|A|} = \left(U, a_{|A|}, V_{\{0,1\}}, f\right) \Leftrightarrow (F, a_{|A|}) \end{cases} = \left((F, a_1), (F, a_2), \cdots, (F, a_{|A|})\right)$$

Then, $(F, E) = \left((F, a_1), (F, a_2), \cdots, (\quad | \ |)\right)$ F, can be defined as a multi-soft set over universe U representing a categorical-valued information system $S = (U, A, V, f)$.

The multivariate multinomial distribution of multi soft set can be defined as :

$$Maximize\ L_{CML}(z, \lambda) = \sum_{i=1}^{|U|} \sum_{k=1}^{K} z_{ik} \sum_{j=1}^{|A|} \sum_{l=1}^{|a_j|} \ln\left(\lambda_{kjl}^{u_i}\right)^{|F, a_{jl}|}$$

Subject to

$$\sum_{k=1}^{K} z_{ik} = 1, for\ i = 1,2, \dots, |U|.$$

$$\sum_{l=1}^{|a_j|} \lambda_{kjl} = 1.$$

The maximization of the objective function $L_{CML}(z, \lambda)$ can be obtained by updating the equation as follows:

$$\lambda_{kjl} = \frac{\sum_{u_i \in (F, a_{jl})} z_{ik}(u_i)}{|U|}$$

$$z_{ik} = \begin{cases} 1 \quad if \quad \sum_{j=1}^{|A|} \ln \lambda_{kjl}^{u_i} = \max_{1 \le k' \le K} \sum_{j=1}^{|A|} \ln \lambda_{k'jl}^{u_i} \\ 0 \qquad\qquad otherwise \end{cases}$$

where $U = \{u_1, u_2, \dots, U_n\}$ is finite set of instance, $A = \{a_1, a_2, \dots a_m\}$ is finite set of attribute. $(F, E) = \left((F, a_1), (F, a_2), \cdots, (F, a_{|A|})\right)$ can be defined as a multi-soft set over universe U as in [16], where $(F, a_1), \cdots, \left(F, a_{|A|}\right) \subseteq (F, A)$ and $\left(F, a_{j_1}\right), \cdots, \left(F, a_{j_{|a_j|}}\right) \subseteq (F,\ a_j)$.

The HCSS algorithm is shown in Algorithm 1.

Step 1. Fix $2 \le K \le |U|$ and fix $\varepsilon > 0$ and Max iter.
Give initials $z_{ik}^{(0)}$ and let $t = 1$.
Step 2. Compute $\lambda_{kjl}^{(t)}$ with $z_{ik}^{(t-1)}$
Step 3. Update $z_{ik}^{(t)}$ with $\lambda_{kjl}^{(t)}$
Step 4. Compare $z_{ik}^{(t)}$ to $z_{ik}^{(t-1)}$ in a convenient norm
IF $||z_{ik}^{(t)} - z_{ik}^{(t-1)}|| < \varepsilon$ or $t =$ Max iter, THEN Stop
ELSE t=t+1 and return to Step 2.

Algorithm 1. The HCSS algorithm

III. RESULTS AND DISCUSSION

A. *Analysis and Data Description*

The research was conducted on communities adjacent to the landslide points. Then analyzed based on data collected by distributing questionnaires, to as many as 40 respondents as a research sample. There are 14 landslide points in Sidoharjo Village as shown in Figure 1.

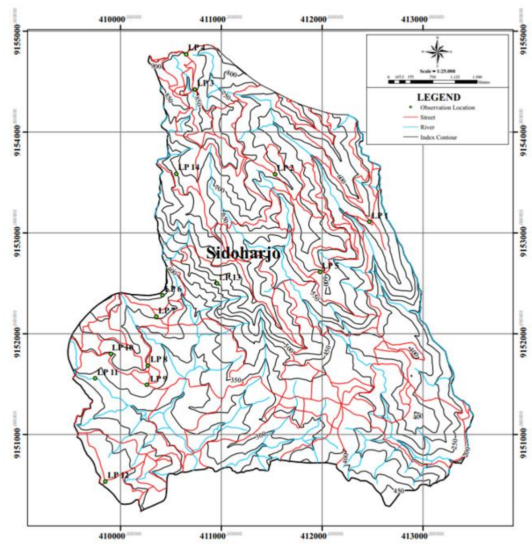

Fig 1. Distribution Map of Landslide Point

The variable of the coping capacity consists of mitigation knowledge and disaster risk knowledge. The mitigation knowledge involves knowledge level, Action plan and Local culture. Meanwhile, the disaster risk knowledge involves Landslide Risk Reduction and Rescue from Disaster. The description of the variable is shown in Table 2.

TABLE 1. DATA DESCRIPTION

No	Category	Landslide Risk Reduction		Rescue from Disaster	
		Frequency	**Percentage**	**Frequency**	**Percentage**
1.	Very Dissatisfied	4	10.0%	6	15%
2.	Not satisfied	20	50.0%	15	37.5%
3.	Enough	12	30.0%	15	37.5%
4.	Satisfied	4	10.0%	4	10.0%
5.	Very satisfied	0	0.0%	0	0.0%
	Total	40	100%	40	100%

Based on Table 2, the general knowledge of landslide risk reduction is mainly in the dissatisfied category, with as many as 20 people (50.0%) and a few respondents who stated that they were satisfied with their knowledge about landslide risk reduction as many as 4 people (10.0%). Most of the community's knowledge about disaster relief was quite satisfactory and dissatisfied, 15 people (37.5%) respectively. At least in the satisfied category, namely 4 people (10.0%).

Based on Table 2 the capacity of the community with parameters of the level of knowledge about landslides is mainly in the sufficient category of 20 people (50.0%), in the good category 12 people (30.0%) and knowledge in the less category is 4 people (10%) . The capacity of the community with the parameters of the action plan is mainly in the sufficient category, with as many as 28 people (70.0%), in the less category as many as 7 people (17.5%) and in the good category as many as 5 people (12.5%). The capacity of the community with local culture parameters is mainly in the good category, namely 34 people (85.0%), in the sufficient category 4 people (10.0%) and in the less category 2 people (5.0%).

TABLE 2 PREDICTOR VARIABLE STATISTICS

No	Category	Knowledge level		Action plan		Local culture	
		Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
1.	Not enough	4	10.0%	7	17.5%	2	5.0%
2.	Enough	20	50.0%	28	70.0%	4	10.0%
3.	Well	12	30.0%	5	12.5%	34	85.0%
	Total	40	100%	40	100%	40	100%

B. *Cluster Analysis*

The clustering result is analyzed using the Dunn Index concerning the number of clusters to determine the desirable number. In the clustering phase, The Dunn index is calculated with the variation number of clusters from 2 to 10. The Dunn index coefficients are summarized in Figure 2. Figure 2 shows that the best number of clusters is 9 on the first level and 3-8 on the second level because it has a higher Dunn index. The data is clustered into 3 clusters using the proposed technique for this case.

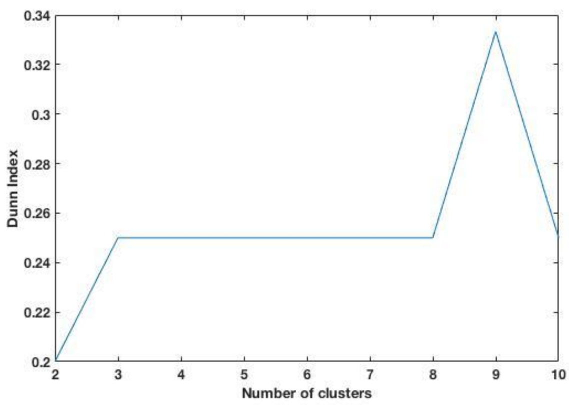

Fig 2. The Dunn index coefficients

Based on the Table 3. The data has been grouped into three distinct clusters (C1, C2, and C3). Each cluster represents a group of areas with similar characteristics regarding disaster risk knowledge and mitigation knowledge.

TABLE 3. CLUSTER RESULTS

Cluster		**C1**	**C2**	**C3**
Area	Wonogiri	0	2	1
	Nglambor	1	2	5
	Keweron	1	2	3
	Blender	0	4	3
	madugondo	1	2	1
	nyemani	0	2	1
	sulur	2	0	5
	serbo	1	1	0
	sum	6	15	19
disaster risk knowledge	average	3	3	1,67
	Level	High	High	Low
mitigation knowledge	average	1.67	3	2
	Level	Low	High	Low

Cluster C1 includes the areas of Wonogiri, Blender, Nyemani, and Serbo. In this cluster, the average disaster risk knowledge is 3, indicating that these areas possess a relatively high level of knowledge about disaster risks. However, the average mitigation knowledge is lower, with a score of 1.67. This suggests that while the residents in these areas are aware of potential risks, they may need more education and support in implementing effective mitigation measures to reduce the impact of disasters.

Cluster C2 comprises the areas of Nglambor, Keweron, and Madugondo. This cluster demonstrates high levels of both disaster risk knowledge and mitigation knowledge, with an average score of 3 for both variables. The residents in these areas are well-informed about disaster risks and understand effective measures to mitigate potential hazards. They are better equipped to respond to disasters and reduce vulnerabilities, making them relatively resilient communities.

Cluster C3 includes the area of Sulur, and it stands out as the cluster with the lowest scores in both disaster risk knowledge and mitigation knowledge. The average disaster risk knowledge is 1.67, indicating a lower level of awareness and understanding of potential risks. Additionally, the average mitigation knowledge is 2, suggesting that residents in this area may have some basic knowledge about mitigation strategies, but there is room for improvement. For this cluster, there is a need for targeted educational initiatives and capacity-building programs to enhance disaster preparedness and risk reduction efforts.

IV. CONCLUSION

The clustering analysis provides valuable insights into the distribution of disaster risk knowledge and mitigation knowledge across different areas in Wonogiri. It highlights the varying levels of awareness and preparedness among the surveyed communities. By understanding each cluster's specific strengths and weaknesses, local authorities and disaster management teams can develop tailored interventions and educational campaigns. These efforts can empower communities to become more proactive in disaster preparedness, enhance their resilience, and reduce the potential impacts of future disasters. The findings from this analysis serve as a valuable foundation for future research and the development of effective policies aimed at fostering disaster risk reduction and community safety in the region.

REFERENCES

[1] M. Parsons *et al.*, “Top-down assessment of disaster resilience: A conceptual framework using coping and adaptive capacities,” *Int. J. Disaster Risk Reduct.*, vol. 19, pp. 1–11, 2016.

[2] K. T. Ton, J. C. Gaillard, C. E. Adamson, C. Akgungor, and H. T. Ho, “Expanding the capabilities of people with disabilities in disaster risk reduction,” *Int. J. Disaster Risk Reduct.*, vol. 34, pp. 11–17, 2019.

[3] M. S. Uddin, C. E. Haque, M. N. Khan, B. Doberstein, and R. S. Cox, “‘Disasters threaten livelihoods, and people cope, adapt and make transformational changes’: Community resilience and livelihoods reconstruction in coastal communities of Bangladesh,” *Int. J. Disaster Risk Reduct.*, vol. 63, p. 102444, 2021.

[4] H. Tjahjono, Suripin, and Kismartini, “Community Capacity in the Face of Landslide Hazards in the Southern of Semarang City,” *E3S Web Conf.*, vol. 31, pp. 1–7, 2018.

[5] A. A. Shah *et al.*, “Current capacities, preparedness and needs of local institutions in dealing with disaster risk reduction in Khyber Pakhtunkhwa, Pakistan,” *Int. J. Disaster Risk Reduct.*, vol. 34, pp. 165–172, 2019.

[6] U. Maulik and S. Bandyopadhyay, “Genetic algorithm-based clustering technique,” *Pattern Recognit.*, vol. 33, no. 9, pp. 1455–1465, Sep. 2000.

[7] D. Nicholson, O. A. Vanli, S. Jung, and E. E. Ozguven, “A spatial regression and clustering method for developing place-specific social vulnerability indices using census and social media data,” *Int. J. Disaster Risk Reduct.*, vol. 38, p. 101224, 2019.

[8] I. T. R. Yanto, R. Setiyowati, M. M. Deris, and N. Senan, “Fast Hard Clustering Based on Soft Set Multinomial Distribution Function BT - Recent Advances in Soft Computing and Data Mining,” 2022, pp. 3–13.

[9] I. Tri, R. Yanto, R. Saedudin, S. Novita, M. Mat, and N. Senan, “Soft Set Multivariate Distribution for Categorical Data Clustering,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 11, no. 5, pp. 1841–1846, 2021.

[10] M. K. Ng, “A fuzzy k-modes algorithm for clustering categorical data,” *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 4, pp. 446–452, 1999.

[11] D.-W. Kim, K. H. Lee, and D. Lee, “Fuzzy clustering of categorical data using fuzzy centroids,” *Pattern Recognit. Lett.*, vol. 25, no. 11, pp. 1263–1271, Aug. 2004.

[12] M.-S. Yang, Y.-H. Chiang, C.-C. Chen, and C.-Y. Lai, “A fuzzy k-partitions model for categorical data and its comparison to the GoM model,” *Fuzzy Sets Syst.*, vol. 159, no. 4, pp. 390–405, 2008.

[13] J. Yang and Y. Yao, “Semantics of soft sets and three-way decision with soft sets,” *Knowledge-Based Syst.*, vol. 194, p. 105538, 2020.

[14] E. Alam, “Landslide hazard knowledge, risk perception and preparedness in southeast Bangladesh,” *Sustain.*, vol. 12, no. 16, 2020.

[15] L. Antronico, F. De Pascale, R. Coscarelli, and G. Gullà, “Landslide risk perception, social vulnerability and community resilience: The case study of Maierato (Calabria, southern Italy),” *Int. J. Disaster Risk Reduct.*, vol. 46, p. 101529, 2020.

[16] T. Herawan, M. M. Deris, and J. H. Abawajy, “Matrices Representation of Multi Soft-Sets and Its Application,” in *Computational Science and Its Applications -- ICCSA 2010: International Conference, Fukuoka, Japan, March 23-26, 2010, Proceedings, Part III*, D. Taniar, O. Gervasi, B. Murgante, E. Pardede, and B. O. Apduhan, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 201–214.