Determining rural development priorities using approach hybrid clustering methodologies: a case south sulawesi -Indonesia

Article *in* International Journal of Advanced Technology and Engineering Exploration · July 2023

DOI: 10.19101/IJATEE.2023.10101215

2 authors:

Assoc. Prof. Ir. Muhammad Faisal Ph.D
Universitas Muhammadiyah Makassar

Titik Khawa Abdul Rahman
Asia e University

International Journal of Advanced Technology and Engineering Exploration, Vol 10(103)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2023.10101215

Research Article

Determining rural development priority using approach hybrid clustering methodologies: a case south sulawesi - Indonesia

Muhammad Faisal[1, 2*] and Titik Khawa Abdul Rahman[1]
Department of ICT, AeU University, Wisma Subang Jaya, JL.SS 15/4 No.106 Selangor, 47500, Malaysia [1]
Department of Computer Science, STMIK Profesional Makassar, JL A.P Pettarani No. 27, Makassar, Indonesia [2]

Received: 08-March-2023; Revised: 21-June-2023; Accepted: 24-June-2023

Abstract
This study aims to develop village clustering techniques that facilitate the government's placement of experts in villages identified as development priorities. The clustering process utilizes indicators such as the community standard of living index (CSLI), head of family, and number of residents. The CSLI was constructed by involving 900 respondents from 30 villages in South Sulawesi Province, Indonesia. The clustering technique is constructed using a hybrid approach of self-organizing map (SOM), Xie-Beni, and fuzzy c-means (FCM) methodologies. The resulting clusters are categorized into three groups: CSLI-poor, which comprises three sub-clusters of CSLI-poor, CSLI-average, and CSLI-excellent. To determine recommendations for the required field of expertise in each village, the cosine similarity algorithm is employed. Villages classified within the CSLI-poor cluster are considered development priorities. The findings revealed that 36.7% of villages were classified as CSLI-poor, 26.7% as CSLI-average, and 36.7% as CSLI-excellent. Consequently, all villages require experts in the fields of economics, social sciences, and health sciences.

Keywords
Rural, FCM, SOM, Xie-Beni, Clustering, CSLI.

1.Introduction
Sustainable rural development is relevant for various countries, both developing and developed countries [1]. The main problems that occur in the village environment are the readiness of government agencies and the low participation of rural communities [2]. The movement of people to cities is caused by development inequality between villages and cities, so that rural people do not have access to basic needs such as health, education, and employment opportunities [3]. Sustainable development continues to evolve within the framework of global and national governance [4]. A study states that the development gap between rural and urban areas is a critical issue that needs attention [5]. Village cluster development is the key to identifying sustainable rural development to form a comfortable living environment for the community [6]. A study shows that social infrastructure problems and low living standards in rural settlements significantly complicate rural life in modern realities [7], because sustainable rural development only occurs in villages that are easily accessible, so it is not implemented thoroughly.

Therefore, as a preventive measure, it is necessary to implement a strategy so that equitable development can be carried out. This article proposes alternative solutions by maximizing the full involvement of the community in village development decision-making. The first action taken is to involve the community as a source of information relevant to the conditions in a village. Second, grouping villages based on the level of living standards of the community to identify villages that are development priorities. As stated in a study, village grouping techniques can be done by assessing living standard indicators that generally measure human living standards [8]. Finally, involve experts in the implementation of development so that all planning is carried out correctly.

South Sulawesi province implements a village assistance program. The program aims to ensure the successful implementation of village development activities, reduce urbanization, and address development problems in rural areas. This research discusses finding alternative solutions to determining the priority status of villages that need assistance from experts. Village assistance is an activity in the context of developing self-sufficiency and feasibility of

*Author for correspondence

community livelihoods by increasing knowledge, character, talent, and utilization of resources through assistance program policies by the nature of the problems and priorities the village community needs [9].

The village grouping technique is a strategy used to facilitate the government in assigning experts to villages that are prioritized for development, enabling efficient implementation of village development initiatives. Based on the needs analysis that has been found, this study will explain ways that can be done for village mapping based on community standard of living index (CSLI), using hybrid self-organizing map (SOM), Xie-Beni, and methodology fuzzy c-means (FCM). Furthermore, cosine similarity is used to build recommendations for the fields of knowledge needed in certain villages and is used as a guideline for implementing village mentoring programs by experts. Through this research, the author can contribute to building a strategy for implementing equitable village development acceleration through the support and role of information technology in the province of South Sulawesi.

The research paper follows a structured organizational framework. In Section 2, a comprehensive summary of previous investigations is presented, providing the necessary background and context for the study. Section 3 outlines the research methods employed in this study. Section 4 presents the results of the experimental assessment, providing a clear and concise overview of the findings obtained from the study. In Section 5, a detailed discussion is presented, delving into the implications, interpretations, and significance of the results. Finally, Section 6 serves as the conclusion of the paper, summarizing the study's main points.

2.Literature review

This section describes the literature related to village development strategies, including relevant methods used in the process of scoring, clustering, and recommendation. The development of rural areas aims to improve the welfare and economy of rural communities. One of the proposed development strategies is to increase tourism potential. This study measured the level of community readiness using a quantitative approach with primary and secondary data collection methods [10]. Community characteristics that affect the level of readiness based on cross-tabulation analysis and chi-square test include gender, type of work, and community involvement in tourism village development. According to crucial respondents, the community readiness level score calculation shows an average value of 3.9 which shows that most people already have basic knowledge and understanding of tourism village development [11].

All member states of the United Nations adopted the sustainable development goals in 2015 as the central umbrella for balancing social, economic, and environmental sustainability by 2030. This study uses qualitative methods to discuss the requirements for achieving sustainable development goals for planning new territories in Egypt. This study concludes that sustainable development can be implemented through a strategy for dealing with resources and activities, a strategy for dealing with water resources, a strategy for energy, and a strategy for planning for urban formations [12].

Research stated that it is important to focus on rural development issues in order to enable dialogue on development in a regional and/or national perspective [13]. The results of the investigation show that the great advantages of the business can generate added value and influence the increase in the potential of the entity.

Exploration will benefit sustainable rural development and provide practical inspiration for policymakers in rural revitalization [14]. However, data-driven planning involving data analytics has not been used for the overall planning process for village development. This paper proposes the usefulness of gap-based data-driven planning using data analysis tools. The multi-criteria approach of analytic hierarchy process (AHP) and technique for order preference by similarity to ideal solution (TOPSIS) has been applied to identify and prioritise infrastructure gap areas towards 1663 villages in the Indian region. The analysis results divided all villages into three categories: good, fair, and bad [15].

Optimization of rural settlements is considered a valuable solution for stimulating rural vitality. This study develops a rural development index that considers population, land, economy, agriculture, and residential environment to cluster villages in the Chinese region into 5 development types: growth type, maturity, developing transformation, gradual retrogressive and decline type. The results stated that the development of different villages was influenced by demographic, socio-economic, and geographical environments [16].

Clustering is an unsupervised pattern recognition method that can reveal hidden clusters of similar observations and clusters [17]. Clustering is used in many fields, such as pattern recognition, machine learning, image analysis, information retrieval, computer graphics, and bioinformatics, by grouping data sets that have similarities or do not have similarities between each cluster [18].

Clustering algorithms can group regions based on economic potential with mixed attribute data consisting of numerical and categorical data clustering algorithms can group regions based on economic potential with mixed attribute data consisting of numerical and categorical data [19]. This study aims to group villages according to their economic potential in determining village development targets using fuzzy k-prototypes algorithm and modified eskin distance to measure the distance of low, medium, and high economic attributes. The results of this clustering research are used to determine village development targets in increasing the village development index.

Environmental quality (EQ) assessments are carried out to group settlements into sub-regional levels for a better planning framework process [20]. The current research proposes a 3-tier assessment methodology for various spatial dimensions at macro, meso, and micro scales through the cluster analysis process. The result of this study is the formulation of recommendations for an appropriate regional development planning framework [21].

Research stated divides the role of village settlements based on the clustering of production, ecology, undeveloped villages, settlements, and industry using the k-means clustering algorithm and niche-based modeling to help describe various future scenarios of a village [22]. The use of the index in assessing social vulnerability can only describe the general condition of social vulnerability without indicating which factors are dominant in measuring the level of social vulnerability in the community [23].

The approach to identifying social indicators of the circular economy can be made through qualitative and quantitative methods to account for uncertainties related to data collection, assessment, and the number of attributes. This research includes a combined approach of Delphi and fuzzy methods to explore various surveys and social indicators for circular economy experts to reach a consensus on the necessary social measures. These findings show that indicators such as poverty and hunger eradication are priorities for circular economy experts so that several sustainable development goals targets are achieved [24].

SOM algorithm is a neural network model and algorithm that can be used effectively to visualize and explore a text collection [10], Thus making it more suitable for document clustering based on SOM can capture multiple clusters with highly variable statistical properties within a data set [25]. The SOM algorithm defines data dimensions as the number of variables owned by data points to visualize the relationship between data [26]. This study assumes the data through village identity, which collects information from the community about a property. The SOM algorithm produces a layer containing neurons that arrange themselves based on specific input values in a cluster.

The FCM method is a technique of grouping data into a cluster, which is determined based on the degree of membership value with the short distance to the cluster center. The FCM method is a crucial soft clustering model widely applied in various fields [27] and provides a meaningful way to build an information pool [28]. The FCM clustering algorithm is a method that is frequently used in pattern recognition. It has the advantage of giving good modeling results in many cases, although it cannot specify the number of clusters by itself [29].

Intervention in the revitalization and development of art villages is one way to build areas in the countryside, conform to contemporary aesthetics, and meet the needs of modern society [30]. This study used the FCM method and user collaborative filtering to solve the data sparsity problem. The results of experiments on the dataset show that FCM and collaborative-filtering algorithms can solve data sparsity problems and improve the accuracy of recommendations [31].

Xie-Beni is an algorithm that functions to validate the compactness and separation of fuzzy clustering as explained by previous research that the classes resulting from the clustering process need to be validated based on the grouping indicators of the evaluation results in the form of the level of cohesiveness and the degree of separation [32]. Xie-Beni index can calculate the compactness and separation between fuzzy clusters [33], in order to when the Xie-Beni is applied to the clustering method, it can form an optimal cluster area.

Research shows the evaluation of various preprocessing techniques on text classification. This study used several preprocessing techniques on text data, such as tokenization, stopword removal, and stemming. In addition, this study used term frequency and inverse document frequency (TFIDF) with cosine similarity and chi-square for feature extraction. In addition, this study used TFIDF with cosine similarity and chi-square for feature extraction [34].

Research on fuzzy cosine strings in automated ticket classification and processing systems through performance comparison of cosine similarity string algorithms. Optimization is done by supplementing the fuzzy string-matching algorithm with the convolutional neural network binary classifier. The results achieved better keyword classification ratios for the two ticket categories, with relative 69% and 78% [35]. Such an approach allows classification based on word similarity.

3.Method

3.1Research design

In this research, the clustering process was carried out using the SOM, Xie-Beni, and FCM methodologies. The clustering result is the grouping of villages based on the CSLI scores in the high, medium, and low-level categories.

The standard of living index was combined with previous indicators so that each indicator can represent the actual conditions of the community in a village so that the decision-making process on the level of distribution of village development provides more objective information. *Figure 1* illustrates the research design indexing and clustering CSLI. Based on the illustration shown in *Figure 1*, this panel consists of three objectives, namely village input, clustering, and placement of experts in a village.

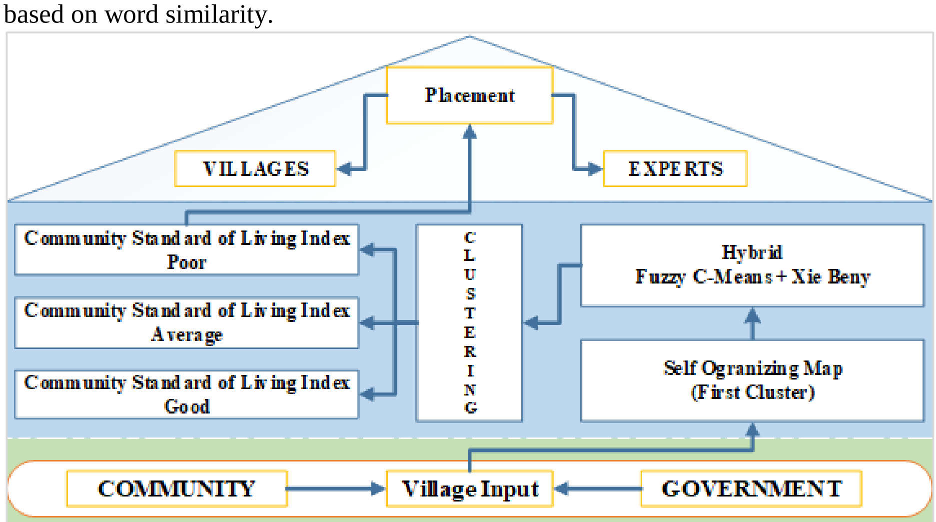

Figure 1 Research design

3.2Village input

The village input process is the initial stage which was carried out by collecting information data, suggestions, and complaints from the community each period. The data collection process was implemented using two methods, including the offline method, namely conducting interviews and discussions with the community, where the results of the interviews were recorded through an application built to store data offline. Then, the online method, namely distributing questionnaires via internet media either through the Google form facility or using a mobile-based application already available on the google play store. As is known, questionnaires containing instruments are distributed to respondents. Then, the calculating the results of assessments by communities using Equation 1.

$$SVj = \frac{\sum \text{Criteria(w)} \times \text{Subcriteria(w)}}{\sum \text{Respondents}} \tag{1}$$

Where
SVj= Scoring Village
Criteria(w)= Weight of Criteria
Subcriteria(w)=Weight of Subcriteria
Respondents=Communities as sample

3.3Clustering the village

The clustering technique defines a collection of classes and places each related object in one class, making it easy to identify objects based on the same characteristics [36]. This research used the clustering process to select village groups prioritizing mentoring. The criteria used are value-weight, number-sample, and number of populations. The uncertainty factor associated with the characteristics of data objects makes it difficult to choose the appropriate number of clusters, especially when dealing with data objects with high dimensions, varying data sizes, and density [37]. Therefore, cluster index validation was carried out based on the smallest cluster value to determine the optimal clusters.

Research mentioned that clustering techniques based on fuzzy and SOM are relatively more efficient than traditional approaches in revealing hidden structures in data sets because segments derived from SOM have more ability to provide exciting insights for retrieval data-driven decisions [38]. The combined use of SOM and FCM grouping has proven to be a powerful tool for determining cluster boundaries [39].

This study built a clustering process using a hybrid FCM, SOM, and Xie-Beni methodologies. The primary considerations for doing a hybrid are: First, a convergent final weight value can be generated through the SOM process, which is transformed into an initial cluster value in the FCM so that the cluster input is static. Secondly, the cluster validity process uses the Xie-Beni through the cluster comparison process, and the smallest index helps determine the selected cluster. So, the data set in each cluster can represent the desired data group, based on the experimental observation that the Xie-Beni rarely underperforms when used to show monotonicity towards cluster determination [40]. The illustration of the hybrid methodology is shown in *Figure 2*.

In general, the stages of the process of determining village clustering include:

(1)Declaration of variable VS, TC, TP, Learning rate, Epsilon.
VS = Village Score
TC = Total of Community
TP = Total of Patriarch
(2)Determine the village score(VS) from the community for each village following Equation 2.
(3)Initiation of output neuron output each y1, y2 ..yn.
(4)The initiation weight is as follows:

$$r_{ij} = \frac{X_{ij}}{Max(X_{ij})} \tag{2}$$

(5)Finding the shortest distance from each output neuron to the input data using the Euclidian distance formula as follows Equation 3:

$$D_j = \sum_j (w_{ji} - x_j)2 \tag{3}$$

Each weight w_{ij} is updated by neighbouring weights using the formula with the following Equation 4.

$$W_{ij}(new) = W_{ij}(old) + \alpha\left(X_i\text{-}W_{ij}(old)\right) \tag{4}$$

(6)Update bias weight (error).
(7)Repeat steps 6 and 7 until there is no weight update or it has reached a stop condition or error in the smallest
(8)Save the convergent weight, Initiation cluster
(9)Determine the total of clusters built
(10)Initialize the initial partition matrix
(11)Calculate the centroid value using the formula Equation 5.

$$Vkj = \frac{\sum_{i=1}^{n}((\mu ik)^{w} * \ Xij)}{\sum_{ij}^{n}(\mu ik)^{w}} \tag{5}$$

Calculating the value of membership degrees using the formula Equation 6.

$$Qj = \sum_{k=1}^{c} \mu \ ik. \tag{6}$$

(12)Calculating the value object function using the formula Equation 7.

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{n} (x_{ik} - v_{kj}\right](\mu_{ik})^{w}\right) \tag{7}$$

(13)Checking the value of the convergent cluster, If it has not converged, calculate the change in the matrix using the formula Equation 8.

$$M_{ik} = \frac{[\sum_{j=1}^{m}(x_{ik} - v_{kj})^2]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}[\sum_{j=1}^{m}(x_{ik} - v_{kj})^2]^{\frac{-1}{w-1}}} \tag{8}$$

If the cluster values converge, then do the optimal cluster validation using the formula Equation 9 [41].

$$XB = \frac{\sum_{j=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} ||xj\text{-}vi||^2}{n^* \min_{i,j} ||v_i\text{-}v_j||^2} \tag{9}$$

Determine the cluster of selected
(14)Save the data clusters

After determining the score for each village, the following process combines the score, the number of the patriarch, and the number of residents to serve as input for the clustering process. Furthermore, from the clustering process, an index of the CSLI was generated, divided into three categories. Firstly, CSLI excellent is the village with a high level of community welfare to create prosperity in life. Secondly, the CSLI average is the village has a moderate level of community welfare, so it is seen that it still requires assistance from the government in several matters relating to improving community welfare. Thirdly, CSLI-poor is village with a low level of community welfare, so it is seen as in dire need of assistance from the government in all fields related to improving community welfare.

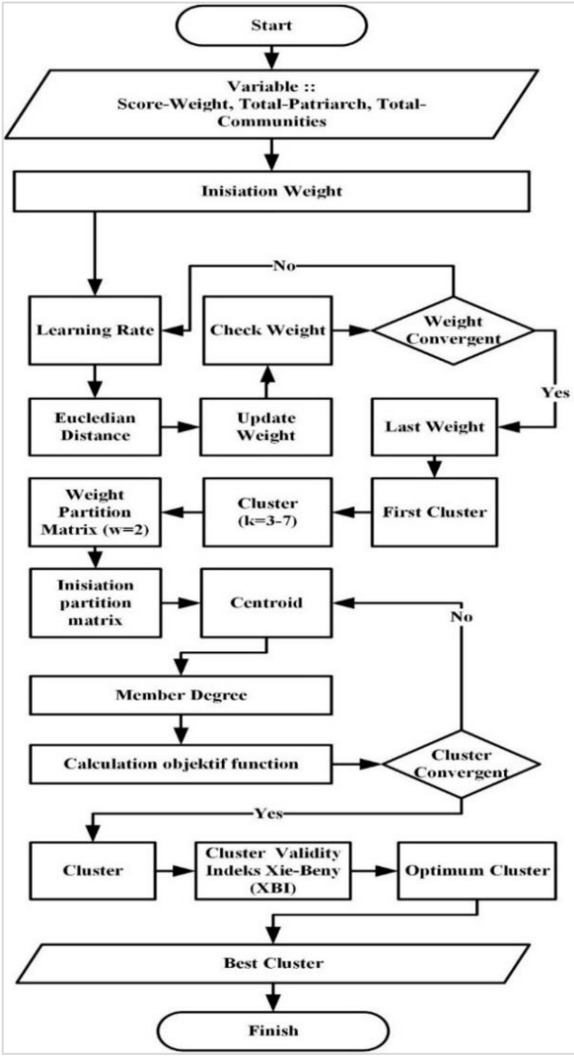

Figure 2 Flowchart village clustering

3.4Placement the expert

The placement of village assistants is adjusted to their areas of expertise and problems that occur in a village so that they can be more productive and synergize with the community to implement their expertise to produce solutions. The first stage was to identify existing problems in a village as village input by collecting information on current conditions through instrument distribution, then comparing village input data with keywords to obtain instructions or guidelines to adjust the placement of experts in a village.

The testing process compares actual and text mining data, and then validates decisions based on needs conditions in a village area. An illustration of the recommendation process is seen in *Figure 3*.

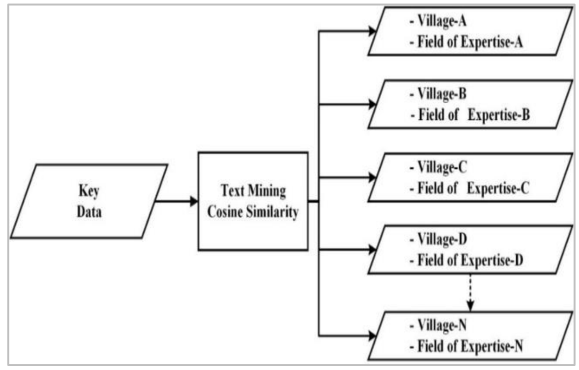

Figure 3 Placement of the expert for each village

Figure 3 illustrates the flow diagram of the process for the placement of experts based on the results of searching for community complaint data when assessing VS. Cosine similarity method using TFIDF in the process of weighting on each word. Generally, the cosine similarity value is used to find the degree of similarity between two sets of elements [42]. The formula used by cosine similarity in calculating similarity is as follows Equation 10.

$$S = \frac{A \bullet B}{|A||B} = \frac{\sum_{i=1}^{n} Aj \text{ x } Bj}{\sqrt{\sum_{i=1}^{n}(Ai)^2 \text{ x } \sqrt{\sum_{i=1}^{n}(Bi)^2}}} \tag{10}$$

Descriptions:
S= Measurement results
A = Vector A, which will be compared the similarity
B = Vector B, whose similarity will be compared
A • B = dot product between vector A and vector B
|A| = length of vector A
|B| = length of vector B
|A||B| = cross product between |A| and |B|

In the context of decision-making, text mining can be used to extract important information from a collection of texts or documents and help policymakers make informed decisions based on the data.

4.Results

4.1Village score(VS)

Criteria weighting analysis is determining multi-criteria weights in decision-making through determining weights for needs at the criteria evaluation stage, where one way to determine weights is to determine the priority order of criteria and use surrogate weights to determine weights according to the number of criteria used [43]. In this research, the VS is a step taken to collect village scores which are measured based on the opportunities for community welfare in each village, which in this study involved 900 people as respondents from 30 villages in South Sulawesi province. In this study, 13 criteria were used which contained 75 sub-criteria, where weights were assigned to each criteria and sub-criteria. The list of criteria and sub-criteria used is shown in *Table 1*. The calculation of the scoring village is determined based on the weight of the criteria and subcriteria that have been determined, while the number of respondents is the number of samples from each village that inputs data on the instrument document. An example of the results of the assessment carried out on one of the respondents is seen in *Table 2*.

Table 1 Criteria and sub-criteria

Criteria	Weight	Sub-citeria	Weight
Ownership Status of residential buildings	0.3144	Privately owned	0.437956
		Contract	0.218978
		Rent-free	0.145985
		Service	0.109489
		Other	0.087591
Status of ownership of the land of residence	0.1572	Privately owned	0.480000
		Someone else's	0.240000
		State lands	0.160000
		Other	0.120000
Types of the widest flooring	0.1048	Marble	0.341417
		Ceramics	0.170709
		Vinyl	0.113806
		Tiles/tiles/terrazzo	0.085354
		High-quality wood/board	0.068283
		Cement/red brick	0.056903
		Bamboo	0.048774
		Low-quality wood/board	0.042677
		Soil	0.037935
		Other	0.034142
The widest types of walls	0.0786	Wall	0.385675
		Bamboo/wire matting stucco	0.192837
		Wood	0.128558
		Bamboo matting	0.096419
		Logs	0.077135
		Bamboo	0.064279
		Other	0.055096
The widest types of roofs	0.0628	Concrete/concrete roof tiles	0.341417
		Ceramic tiles	0.170709
		Metal tiles	0.113806
		Clay tiles	0.085354
		Asbestos	0.068283
The widest types of roofs	0.0524	Zinc	0.056903
		Shingles	0.048774
		Bamboo	0.042677
		Hay	0.037935
Widest wall quality	0.0449	Good/high quality	0.666667
		Ugly/low quality	0.333333
Sources of drinking water	0.0393	Branded bottled water	0.322247
		Refillable water	0.161123
		Plumbing meter	0.107416
		Plumbing retail	0.080562
		Bore/pump wells	0.064449
		Protected wells	0.053708
		Shielded wells	0.046035
		Protected springs	0.040281
		Unprotected springs	0.035805
		River/lake/reservoir water	0.032225

Criteria	Weight	Sub-citeria	Weight
		Rainwater	0.029295
		Other	0.026854
How to get drinking water	0.0349	Buying retail	0.545455
		Client	0.272727
		Not buying	0.181818
Use of defecation facilities	0.03144	Alone	0.480000
		Together	0.240000
		Common	0.160000
		Nothing	0.120000
Main sources of illumination	0.02858	PLN electricity	0.545455
		Non-PLN electricity	0.272727
		Non-electricity	0.181818
Installed electrical power (PLN)	0.0262	450 watts	0.068027
		900 watts	0.081633
		1300 watts	0.102041
		2200 watts	0.136054
		>2200 watts	0.204082
		Not using PLN	0.408163
Fuel for cooking	0.0241	Electricity	0.353486
		Gas>3kg	0.176743
		Gas 3 kg	0.117829
		Gas Kota/biogas	0.088371
		Kerosene	0.070697
		Briquettes	0.058914
		Charcoal	0.050498
		Firewood	0.044186
		Not cooking at home	0.039276

Table 2 Example format of communities scoring

Criteria	C_W	Sub-criteria	S_W	$C_W \times S_W$
Ownership status of residential buildings	0.3144	Privately owned	0.437956	0.137693
Types of the widest flooring	0.1048	Boat ownership	0.24	0.025152
The widest types of walls	0.0786	Wall	0.385675	0.030314
The widest types of roofs	0.0628	Concrete	0.341417	0.021441
The widest types of roofs	0.0524	Zinc	0.056903	0.002982
Widest wall quality	0.0449	Good/high quality	0.666667	0.029933
Sources of drinking water	0.0393	Branded bottled	0.322247	0.012664
How to get drinking water	0.0349	Buying retail	0.545455	0.019036
Use of defecation facilities	0.03144	Ownership of	0.48	0.015091
Main sources of illumination	0.02858	Electricity	0.545455	0.015589
Installed electrical power (PLN)	0.0262	1300 watts	0.102041	0.002673
Fuel for cooking	0.0241	Gas>3kg	0.176743	0.00426
SCORE				0.02641

Table 2 displays the process of calculating the VS using Equation 1. It is written in the example below:

$$SVj = \frac{\sum \text{Criteria(w)} \times \text{Subcriteria(w)}}{\sum \text{Respondents}} = 0.02641$$

After the data collection stage is carried out, the next stage is to calculate the assessment by the community. The VS is shown in *Table 3*. *Table 3* shows the results of calculating village scores for all villages, where each village gets different values. If a village gets a high score, it means that the community gives an assessment that describes the community as having a good level of welfare. Furthermore, if a village gets a low score, it is estimated that the welfare level of the community is in poor condition.

Table 3 Village score (VS)

ID	Village	Score
1	Bonto Cinde	0.33073
2	Lonrong	0.3186
3	Bajiminasa	0.219
4	Borongloe	0.47335
5	Bonto Karaeng	0.31336
6	Bonto Tappalang	0.32778
7	Bonto Lojong	0.32924
8	Tamatto	0.33464
9	Kahaya	0.29003
10	Bialo	0.30667
11	Bonto Minasa	0.36471
12	Bonto Marannu	0.3453
13	Tambangan	0.37964
14	Darubiah	0.321
15	Welado	0.20524
16	Amali Riattang	0.38065
17	Mallari	0.4589
18	Bana	0.5255
19	Bulusirua	0.31047
20	Laoni	0.26416
21	Pusungnge	0.35804
22	Praja Maju	0.36849
23	Data	0.31064
24	Nusa	0.45794
25	Panyili	0.39561
26	Palajau	0.3339
27	Tanjonga	0.31983
28	Pallantikang	0.33649
29	Garassikang	0.3339
30	Pappalluang	0.3339

4.2Initiation weight

A research mentioned that the use of weights had proven its superior ability to dynamically improve the overall performance of the algorithm [44], as in the stages of the algorithm that has been designed, there was a process of loading the weights of the features used. The first step is to choose the largest value from each input criteria, as seen in *Table 4.*

Table 4 The village input

Village	Score(Vs)	Patriarch (Vp)	Residents (Vr)
Bonto Cinde	0.33073	755	2110
Lonrong	0.3186	976	3250
Bajiminasa	0.219	1065	3527
Borongloe	0.47335	***1398***	***4198***
Bonto Karaeng	0.31336	519	1808
Bonto Tappalang	0.32778	523	1518
Bonto Lojong	0.32924	1044	3351
Tamatto	0.33464	912	1945
Kahaya	0.29003	373	1291
Bialo	0.30667	1009	3418
Bonto Minasa	0.36471	350	1000
Bonto Marannu	0.3453	561	1743
Tambangan	0.37964	1100	4089
Darubiah	0.321	836	3172
Welado	0.20524	582	2543
Amali Riattang	0.38065	293	500

Village	Score(Vs)	Patriarch (Vp)	Residents (Vr)
Mallari	0.4589	698	2858
Bana	***0.5255***	599	2434
Bulusirua	0.31047	335	720
Laoni	0.26416	200	739
Pusungnge	0.35804	162	709
Praja Maju	0.36849	319	1394
Data	0.31064	212	920
Nusa	0.45794	429	1675
Panyili	0.39561	320	1340
Palajau	0.3339	1172	3518
Tanjonga	0.31983	347	1200
Pallantikang	0.33649	501	2285
Garassikang	0.3339	845	2972
Pappalluang	0.3339	1014	3395

After making a comparison of the input data on the criteria, the largest value is obtained from each criteria, namely:
Max(VS) = 0.5255
Max(VP) = 1398
Max(VR)= 4198

After obtaining the maximum value of each criteria, the next process is to calculate the normalization of the input criteria using Equation 2.

It is written in example $0.629363 = \frac{0.33073}{0.5255}$. The result is shown in *Table 5*. The next step is calculating the distance between normalized values to get the best matching unit (BMU). It is stated that criteria normalization can significantly improve both processing efficiency and accuracy [45]. Calculating the distance between normalized values is obtained using Equation 3, which is guided by the distance from the smallest normalized value, where the results are shown in *Table 6*.

Table 5 Normalization of criteria weight

Village	Score			Patriarch			Residents		
	VS	**Max(VS)**	**RVS**	**VP**	**Max(VP)**	**RVP**	**VR**	**Max(VR)**	**RVR**
Bonto Cinde	*0.33073*		*0.629363*	*755*		*0.540057*	*2110*		*0.50262*
Lonrong	*0.3186*		*0.60628*	*976*		*0.69814*	*3250*		*0.77417*
Bajiminasa	0.219		0.416746	1065		0.761803	3527		0.84016
Borongloe	0.47335		0.900761	1398		1	4198		1
Bonto Karaeng	0.31336		0.596308	519		0.371245	1808		0.43068
Bonto	0.32778		0.623749	523		0.374106	1518		0.36160
Bonto Lojong	0.32924		0.626527	1044		0.746781	3351		0.79823
Tamatto	0.33464		0.636803	912		0.652361	1945		0.46331
Kahaya	0.29003		0.551912	373		0.26681	1291		0.30752
Bialo	0.30667		0.583578	1009		0.721745	3418		0.81419
Bonto Minasa	0.36471		0.694025	350		0.250358	1000		0.23820
Bonto Marannu	0.3453	*0.5255*	0.657088	561	*1398*	0.401288	1743	*4198*	0.4151
Tambangan	0.37964		0.722436	1100		0.786838	4089		0.97403
Darubiah	0.321		0.610847	836		0.597997	3172		0.75559
Welado	0.20524		0.390561	582		0.416309	2543		0.60576
Amali riattang	0.38065		0.724358	293		0.209585	500		0.11910
Mallari	0.4589		0.873264	698		0.499285	2858		0.680
Bana	0.5255		1	599		0.428469	2434		0.5798
Bulusirua	0.31047		0.590809	335		0.239628	720		0.1715
Laoni	0.26416		0.506489	200		0.143062	739		0.17603
Pusungnge	0.35804		0.681332	162		0.11588	709		0.16889
Praja maju	0.36849		0.701218	319		0.228183	1394		0.33206
Data	0.31064		0.591132	212		0.151645	920		0.2191

Village	Score			Patriarch			Residents		
	VS	Max(VS)	RVS	VP	Max(VP)	RVP	VR	Max(VR)	RVR
Nusa	0.45794		0.871437	429		0.306867	1675		0.399
Panyili	0.39561		0.752826	320		0.228898	1340		0.3192
Palajau	0.3339		0.635395	1172		0.83834	3518		0.83801
Tanjonga	0.31983		0.60862	347		0.248212	1200		0.28585
Pallantikang	0.33649		0.640324	501		0.358369	2285		0.54430
Garassikang	0.3339		0.635395	845		0.604435	2972		0.70795
Pappalluang	0.3339		0.635395	1014		0.725322	3395		0.80871

Respondents' village patriarch (RVP), Village score (VS), Village patriarch (VP), Village residents (VR), Respondents village residents (RVR), Respondents village score (RVS)

Table 6 Determining short distance

Village	Distance			Minimum value	BMU
	X1	X2	X3		
Bonto Cinde	0.35115	0.202692	0.089079	*0.089079*	3
Lonrong	*0.11257*	0.489057	0.142394	*0.112572*	1

The next step is to form weights, where the determination of initial weights randomly is only used for the first village data. Then, build the weighting against other data until all village data has weights, where weight formation is done by involving the learning rate function. Initial weights are randomly shown in *Table 7.*

Table 7 Initial weight randomly

FIRST WEIGHT (RANDOM)	**X1**	0.39	0.934	0.875
	X2	0.94	0.553	0.177
	X3	0.849	0.338	0.506
LEARNING RATE		0.6		
ITERATION		1		

FCM is sensitive to the initial weight and number of clusters so it needs to be determined manually in advance[46]. After the initial weight is determined randomly, the next step is calculating the BMU environment. The SOM shows the data selection based on Euclidean distance from each BMU [47]. The result is shown in *Table 8.*

Table 8 Calculating of BMU weight

Village	BMU	Weight		
		X1	X2	X3
Bonto Cinde	1	0.39	0.934	0.875
	2	0.94	0.553	0.177
	3	*0.717218*	*0.459234*	*0.503972*
Lonrong	*1*	*0.519768*	*0.792484*	*0.814507*
	2	0.94	0.553	0.177
	3	0.717218	0.459234	0.503972

In *Table 8.* data are shown, which simulate the results of calculating the weighting of the BMU. It is a written example on bonto cinde data : *0.717218 = 0.849 + (0.6 × (0.629 – 0849)).* Then, Longrong's data will make the weighting results on bonto cinde the previous weight. It is written in example *0.519768 = 0.39 + (0.6 ×(0.60628 – 0.39)).* The changing of the weights processing is repeated until it reaches the learning rate limit, and convergent data is formed.

4.3Generate convergence data

As explained in *Figure 2*, there are three input variables in the village clustering process: VS, total of the community, and total of the patriarch. The VS variable is obtained based on the results of the input village but on other variables obtained from the central statistical agency of South Sulawesi. In general, convergent data is determined in the clustering method, such as that found in the SOM. Clustering is a technique for determining correlated data groups in a dataset and is used to partition related data into derived groups [48]. In *Table 9* is shown the last weight for the pappaluang on iteration 16.

Changes in the learning value occur continuously as long as the iteration is run, where at each iteration stage, the learning rate value is written with the Equation 10.

$$\text{NewLR} = \frac{\text{LR}}{2} \tag{10}$$

$$\text{NewLR} = (0.0000183105\ )\,/\,2\ = 0.000009155$$

The learning rate is adaptively updated after random initialization at the beginning of the training process [49]. Changes in weight changes and learning rate values are shown in *Table 10.*

Table 9 The changing of weight on iteration 16

LAST WEIGHT (PAPPALLUANG)	**X1**	0.694652253	0.42576953	0.500538268
	X2	0.64434515	0.207977743	0.240877265
	X3	0.644975293	0.734150337	0.820825426
LEARNING RATE		0.0000183105		
ITERATION				16

Table 10 The changing data on iteration 17

Village	**BMU**	**Weight**			**Learning rate**	**iteration**
		X1	**X2**	**X3**		
Bonto Cinde	1	0.694651655	0.425770577	0.500538287	0.000009155	17
	2	0.64434515	0.207977743	0.240877265		
	3	0.644975293	0.734150337	0.820825426		
Garassikang	1	*0.694652*	*0.42577036*	*0.500538064*		
	2	*0.644344777*	*0.207977765*	*0.240876613*		
	3	*0.644974682*	*0.734151699*	*0.820826478*		
Pappalluang	1	*0.694652*	*0.42577036*	*0.500538064*		
	2	*0.644344777*	*0.207977765*	*0.240876613*		
	3	*0.644974594*	*0.734151618*	*0.820826367*		

Based on the data in *Table 10* it shows that there is no significant difference between the final and previous weight data, besides that the table shows the epsilon value is smaller than the learning rate (0.000001 < 0.00000915), so the iteration process is stopped.

After the convergent values are generated, the next step is to convert the convergent cluster values for each village into inputs at the village clustering stage so that it can be seen that the SOM method is used to construct the actual input values in the clustering process. Convergence data for each village are shown in *Table 11*.

Table 11 Convergence data for villages

VILLAGE	**X1**	**X2**	**X3**
Bonto Cinde	*0.017328762*	*0.179010676*	*0.13917040*
Lonrong	0.156873599	0.526118077	0.004970042
Bajiminasa	0.305493335	0.71766551	0.053227052
Borongloe	0.621682569	1.269315763	0.168207371
Bonto Karaeng	0.0175245	0.064989182	0.28628427
Bonto tappalang	0.026999735	0.042596925	0.340972847
Bonto Lojong	0.196314946	0.601276716	0.001010097
Tamatto	0.05607509	0.247011862	0.134570693
Kahaya	0.08289512	0.016447244	0.490546281
Bialo	0.19832076	0.59634429	0.003967544
Bonto Minasa	0.099586117	0.004271281	0.575907165
Bonto Marannu	0.009292937	0.067918142	0.275480705
Tambangan	0.355343639	0.878696505	0.03224894
Darubiah	0.101741487	0.418173252	0.023958171
Welado	0.103632192	0.240949916	0.212002311
Amali riattang	0.193109535	0.021233449	0.773887178
Mallari	0.06980345	0.330795485	0.126885905
Bana	0.099529538	0.28997513	0.277578991
Bulusirua	0.153692441	0.008679574	0.669100558
Laoni	0.220631924	0.027422749	0.784321819
Pusungnge	0.206201414	0.0150322	0.808604623

VILLAGE	X1	X2	X3
Praja maju	0.067468581	0.011958154	0.498058423
Data	0.165040068	0.006476825	0.70422637
Nusa	0.055701845	0.08635345	0.411796151
Panyili	0.07502643	0.018340662	0.518543333
Palajau	0.287618115	0.754016183	0.011242553
Tanjonga	0.085019433	0.004917738	0.523660954
Pallantikang	0.009410391	0.11470361	0.217698786
Garassikang	0.078454699	0.375421678	0.029658469
Pappalluang	0.188217669	0.59016938	0.000316343

The next stage is to determine convergent data as input data in forming village clusters using the FCM method.

4.4Optimal cluster

The process of evaluating the effect of the clustering results generated by the clustering algorithm is done using the clustering validity index method [50]. To develop an understanding of the relative quality of the clusters and an overview of the data configuration, each cluster is built based on a comparison of tightness and separation to identify the object that has the best position within a cluster [51]. A research stated that the grouping validity function is an index used to assess the accuracy of the grouping results [52]. Then, in calculating the optimal number of clusters, one of the cluster validity indices that can be used is the Xie-Beni, which is done by comparing the compactness values as a separation in the FCM method using the Xie-Beni [53].

Xie-Beni is an algorithm that functions to validate the compactness and separation of fuzzy clustering because classes resulting from the clustering process need to be validated based on the grouping indicators of the evaluation results in the form of the level of cohesiveness and the degree of separation. Xie-Beni can calculate the compactness and separation between fuzzy clusters in order to when Xie-Beni is applied to the clustering method, it can form an optimal cluster area. The optimal cluster is determined by comparing the objective value to the number of clusters using Equation 9 until the optimum number of clusters is determined which has the smallest value. The results of testing the number of 3 - 9 clusters are shown in *Table 12*.

Table 12 Optimum cluster using Xie-Beni

Number of clusters	XIE-BENI Value	Iteration
3	*0.000000067693*	*27*
4	0.000000067947	38
5	0.000000089138	41
6	0.000000070658	28
7	0.000000098742	44
8	0.000000098648	51
9	0.000000077329	17

Based on the test results obtained, it is determined that the optimal number of clusters to be used is 3, with the smallest value *0.000000067693*.

4.5Determining cluster of village

After obtaining all the elements needed in the village clustering process, the last step is to calculate the objective value of each village using the FCM method with the provisions that the weight = 2, the number of iterations is 100, and the epsilon value = 0.000001. In this study, the initial weight of the FCM was determined based on the convergent value in *Table 11*, in examples 0.017328762^2, 0.179010676^2, 0.1391704^2. Next, determine the membership value using Equation 4, as it is known that the fuzzy membership function plays an important role in fuzzy set theory [54]. The membership value calculation results are shown in *Table 13*.

Table 13 MiU squared

	MiU SQUARED	
M1	**M2**	**M3**
0.0003	*0.03204*	*0.01937*
0.0246	0.27680	0.00002
0.0933	0.51504	0.00283
0.3865	1.61116	0.02829
0.0003	0.00422	0.08196
0.0007	0.00181	0.11626
0.0385	0.36153	0.00000
0.0031	0.06101	0.01811
0.0069	0.00027	0.24064
0.0393	0.35563	0.00002
0.0099	0.00002	0.33167
0.0001	0.00461	0.07589
0.1263	0.77211	0.00104

MiU SQUARED		
M1	**M2**	**M3**
0.0104	0.17487	0.00057
0.0107	0.05806	0.04494
0.0373	0.00045	0.59890
0.0049	0.10943	0.01610
0.0099	0.08409	0.07705
0.0236	0.00008	0.44770
0.0487	0.00075	0.61516
0.0425	0.00023	0.65384
0.0046	0.00014	0.24806
0.0272	0.00004	0.49593
0.0031	0.00746	0.16958
0.0056	0.00034	0.26889
0.0827	0.56854	0.00013
0.0072	0.00002	0.27422
0.0001	0.01316	0.04739
0.0062	0.14094	0.00088
0.0354	0.34830	0.00000
∑M1 = 1.0900	**∑M2= 5.50316**	**∑M3= 4.87545**

After forming the MiU squared, then calculate the membership value for each cluster C1, C2, and C3 by multiplying each value on the village criteria with the MiU quadrate value. It is written in the example below:

MiU C11 ⇒ 0.00001 = *0.017328762* × *0.0003*
MiU C12 ⇒ 0.00005 = *0.179010676* ×*0.0003*
MiU C13 ⇒ 0.00004 = *0.139170400* × *0.0003*

MiU C21 ⇒ 0.00056 = *0.017328762* × *0.03204*
MiU C22 ⇒ 0.00574 = *0.179010676* × *0.03204*
MiU C23 ⇒ 0.00446 = *0.139170400* × *0.03204*

MiU C31 ⇒ 0.00034 = *0.017328762* ×*0.01937*
MiU C32 ⇒ 0.00347 = *0.179010676* ×*0.01937*
MiU C33 ⇒ 0.00270 = *0.139170400* × *0.01937*

The MiU squared calculation results are shown in *Table 14.*

Table 14 Clustering MiU squared

MiU SQUARED C1			**MiU SQUARED C2**			**MiU SQUARED C3**		
0.00001	*0.00005*	*0.00004*	*0.00056*	*0.00574*	*0.00446*	*0.00034*	*0.00347*	*0.00270*
0.00386	0.01295	0.00012	0.04342	0.14563	0.00138	0.00000	0.00001	0.00000
0.02851	0.06698	0.00497	0.15734	0.36963	0.02741	0.00087	0.00203	0.00015
0.24027	0.49058	0.06501	1.00163	2.04507	0.27101	0.01759	0.03591	0.00476
0.00001	0.00002	0.00009	0.00007	0.00027	0.00121	0.00144	0.00533	0.02346
0.00002	0.00003	0.00025	0.00005	0.00008	0.00062	0.00314	0.00495	0.03964
0.00757	0.02317	0.00004	0.07097	0.21738	0.00037	0.00000	0.00000	0.00000
0.00018	0.00078	0.00042	0.00342	0.01507	0.00821	0.00102	0.00447	0.00244
0.00057	0.00011	0.00337	0.00002	0.00000	0.00013	0.01995	0.00396	0.11804
0.00780	0.02345	0.00016	0.07053	0.21208	0.00141	0.00000	0.00001	0.00000
0.00099	0.00004	0.00571	0.00000	0.00000	0.00001	0.03303	0.00142	0.19101
0.00000	0.00001	0.00002	0.00004	0.00031	0.00127	0.00071	0.00515	0.02091
0.04487	0.11095	0.00407	0.27436	0.67845	0.02490	0.00037	0.00091	0.00003
0.00105	0.00433	0.00025	0.01779	0.07313	0.00419	0.00006	0.00024	0.00001
0.00111	0.00259	0.00228	0.00602	0.01399	0.01231	0.00466	0.01083	0.00953
0.00720	0.00079	0.02886	0.00009	0.00001	0.00035	0.11565	0.01272	0.46348
0.00034	0.00161	0.00062	0.00764	0.03620	0.01388	0.00112	0.00533	0.00204
0.00099	0.00287	0.00275	0.00837	0.02438	0.02334	0.00767	0.02234	0.02139
0.00363	0.00021	0.01581	0.00001	0.00000	0.00005	0.06881	0.00389	0.29955
0.01074	0.00133	0.03818	0.00017	0.00002	0.00059	0.13572	0.01687	0.48248
0.00877	0.00064	0.03438	0.00005	0.00000	0.00018	0.13482	0.00983	0.52870
0.00031	0.00005	0.00227	0.00001	0.00000	0.00007	0.01674	0.00297	0.12355
0.00450	0.00018	0.01918	0.00001	0.00000	0.00003	0.08185	0.00321	0.34925
0.00017	0.00027	0.00128	0.00042	0.00064	0.00307	0.00945	0.01464	0.06983
0.00042	0.00010	0.00292	0.00003	0.00001	0.00017	0.02017	0.00493	0.13943
0.02379	0.06238	0.00093	0.16352	0.42869	0.00639	0.00004	0.00010	0.00000
0.00061	0.00004	0.00379	0.00000	0.00000	0.00001	0.02331	0.00135	0.14360
0.00000	0.00001	0.00002	0.00012	0.00151	0.00286	0.00045	0.00544	0.01032
0.00048	0.00231	0.00018	0.01106	0.05291	0.00418	0.00007	0.00033	0.00003

MiU SQUARED C1			MiU SQUARED C2			MiU SQUARED C3		
0.00667	0.02091	0.00001	0.06556	0.20556	0.00011	0.00000	0.00000	0.00000
∑C1			**∑C2**			**∑C2**		
0.40543	**0.82974**	**0.23797**	**1.90328**	**4.52676**	**0.41419**	**0.69903**	**0.18263**	**3.04634**

A research mentioned that fuzzy cluster centroid is utilized to represent the data points distributed in each class [55]. Therefore the establishment of a cluster centre will be the stage before the objective value is formed Equation 5. It is written in example below:

Centroid(C11) ⇒ *0.371941247* = 0.40543 / 1.0900

Centroid(C21) ⇒ *0.345851615* = 1.90328 / 5.50316

Centroid(C31) ⇒ *0.143377134* = 0.69903 / 4.87545

The process of establishing centroid values is carried out for all clusters, as shown in *Table 15*.

Table 15 Cluster centroid

Cluster	*C1*	*0.371941247*	*0.761195402*	*0.218308827*
	C2	*0.345851615*	0.822575846	0.075263541
	C3	*0.143377134*	0.037459929	0.624831956

In the FCM method, data grouping is carried out for each data to be assigned to a cluster determined by the degree of membership [56]. After forming the centroid cluster, the next step is forming membership degree, where the degree of dependency can measure interactions among components by weight assignment [57], as performed calculations use Equation 6, for example:

$0.00014 = ((0.017328762 - 0.371941247)^2 + (0.179010676 - 0.761195402)^2 + (0.1391704 - 0.218308827)^2) \times 0.0003$

$0.01686 = ((0.017328762 - 0.345851615)^2 + (0.179010676 - 0.822575846)^2 + (0.1391704 - 0.075263541)^2) \times \mathit{0.03204}$

$0.00526 = ((0.017328762 - 0.143377134)^2 + (0.179010676 - 0.037459929)^2 + (0.1391704 - 0.624831956)^2) \times 0.01937$

The results of calculating the membership degree formation are shown in *Table 16.* Based on the accumulation of the membership degree value, the objective value is calculated using Equation 7, by subtracting the new objective value from the previous objective = 0, so the result that is obtained in the first iteration is 1.42178. In *Table 17*. display the objective values between 1 to 27.

Table 16 Membership degree

Membership degree			∑ MEMBERSHIP DEGREE
0.00014	0.01686	0.00526	0.02227
0.00362	0.03558	0.00002	0.03921
0.00313	0.00676	0.00231	0.01220
0.12486	0.45805	0.05531	0.63822
0.00019	0.00307	0.01075	0.01401
0.00047	0.00142	0.01095	0.01284
0.00399	0.02778	0.00000	0.03178
0.00117	0.02555	0.00529	0.03200
0.00489	0.00024	0.00533	0.01046
0.00406	0.02775	0.00001	0.03182
0.00769	0.00002	0.00180	0.00950
0.00005	0.00333	0.01070	0.01408
0.00615	0.00393	0.00115	0.01123
0.00236	0.03948	0.00029	0.04214
0.00368	0.02413	0.00959	0.03740
0.03312	0.00052	0.01495	0.04859
0.00139	0.03509	0.00546	0.04195
0.00297	0.03240	0.01435	0.04972
0.01930	0.00008	0.00130	0.02068
0.04292	0.00087	0.01938	0.06317
0.03966	0.00027	0.02499	0.06492
0.00333	0.00013	0.00558	0.00904
0.02311	0.00005	0.00383	0.02699
0.00184	0.00551	0.00940	0.01676
0.00411	0.00031	0.00439	0.00881
0.00414	0.00693	0.00012	0.01119
0.00540	0.00002	0.00403	0.00946
0.00005	0.00835	0.00899	0.01739
0.00167	0.03855	0.00042	0.04063
0.00392	0.02942	0.00000	0.03334
Objective Value			**1.42178**

Table 17 Objective value of each iteration

Iteration	Objective	Processing
1	*1.421780*	*Continue*
2	0.195269933571	Continue
3	0.196861575861	Continue
4	0.103373977067	Continue
5	0.049904074371	Continue
6	0.024730099495	Continue
7	0.014151418316	Continue
8	0.008485325401	Continue
9	0.005009938180	Continue
10	0.002863101660	Continue
11	0.001588891803	Continue
12	0.000863151028	Continue
13	0.000462262770	Continue
14	0.000245289517	Continue

Iteration	Objective	Processing
15	0.000129382930	Continue
16	0.000067980200	Continue
17	0.000035625789	Continue
18	0.000018637511	Continue
19	0.000009738475	Continue
20	0.000005084314	Continue
21	0.000002652893	Continue
22	0.000001383654	Continue
23	0.000000721452	Continue
24	0.000000376094	Continue
25	0.000000196029	Continue
26	0.000000102163	Continue
27	*0.000000053240*	*Stop*

FCM is used to calculate the membership value owned by FCM in selected clusters based on convergent decisions, where clusters with high membership values are clusters with high importance values [58]. In some cases or applications, FCM is more effective, robust, and consistent in performance as compared to other clustering algorithms [59]. Based on the data in *Table 17*, it is known that the cluster formation process stops at the 27th iteration, with a different value of *0.000000053240* or smaller than epsilon. Determining the convergent value for iteration 1 is shown in *Table 18*.

Table 18 Convergent data on iteration 1

Data convergent(Mik) Iterasi 1			∑Mik	New member		
C1	C2	C3		$\frac{\mathbf{Mik(C1)}}{\mathbf{\sum Mik}}$	$\frac{\mathbf{Mik(C2)}}{\mathbf{\sum Mik}}$	$\frac{\mathbf{Mik(C3)}}{\mathbf{\sum Mik}}$
2.12336	*1.90046*	*3.67928*	7.70311	*0.275650*	*0.246714*	*0.477636*
6.80138	7.77961	1.60463	16.18563	0.420211	0.480650	0.099139
29.79545	76.21617	1.22595	107.23757	0.277845	0.710723	0.011432
3.09533	3.51744	0.51157	7.12434	0.434473	0.493721	0.071806
1.62619	1.37691	7.62130	10.62439	0.153062	0.129599	0.717340
1.53748	1.28101	10.62179	13.44028	0.114393	0.095311	0.790295
9.64903	13.01268	1.40876	24.07048	0.400866	0.540608	0.058526
2.69420	2.38799	3.42595	8.50814	0.316661	0.280672	0.402667
1.40388	1.12177	45.18294	47.70859	0.029426	0.023513	0.947061
9.68409	12.81572	1.42685	23.92666	0.404741	0.535625	0.059634
1.29034	1.01946	184.74813	187.05793	0.006898	0.005450	0.987652
1.62492	1.38338	7.09459	10.10288	0.160837	0.136929	0.702234
20.53377	196.46818	0.90599	217.90794	0.094231	0.901611	0.004158
4.37743	4.42941	1.96957	10.77642	0.406205	0.411028	0.182767
2.91813	2.40583	4.68568	10.00965	0.291532	0.240351	0.468117
1.12588	0.86689	40.07362	42.06639	0.026764	0.020608	0.952628
3.51013	3.11803	2.94630	9.57446	0.366614	0.325661	0.307725
3.33589	2.59558	5.36852	11.30000	0.295212	0.229698	0.475090
1.22380	0.95058	345.49195	347.66632	0.003520	0.002734	0.993746
1.13419	0.86903	31.73992	33.74314	0.033612	0.025754	0.940633
1.07218	0.82684	26.16275	28.06177	0.038208	0.029465	0.932327
1.36552	1.09487	44.47612	46.93651	0.029093	0.023327	0.947580
1.17852	0.91382	129.32079	131.41313	0.008968	0.006954	0.984078
1.68675	1.35233	18.03043	21.06951	0.080056	0.064184	0.855759
1.36961	1.09094	61.21961	63.68016	0.021508	0.017132	0.961361
19.98466	82.03274	1.09800	103.11540	0.193809	0.795543	0.010648
1.33776	1.06649	68.02635	70.43059	0.018994	0.015142	0.965864
1.82023	1.57589	5.27228	8.66840	0.209985	0.181797	0.608219
3.69625	3.65594	2.11567	9.46785	0.390400	0.386142	0.223458
9.04773	11.83736	1.43365	22.31874	0.405387	0.530377	0.064235

Based on *Table 18* is shown the result of determining the convergent cluster for each village based on the latest member value using Equation 8. It is written in the example :

$$2.12336 = ((0.017328762 - 0.371941247)^2 + (0.179010676 - 0.761195402)^2 + (0.1391704 - 0.218308827)^2)^{(-1/(2-1))}$$

$1.90046 = ((0.017328762 - 0.345851615)^2 + (0.179010676 - 0.822575846)^2 + (0.1391704 - 0.075263541)^2)^{(-1/(2-1))}$

$3.67928 = ((0.017328762 - 0.143377134)^2 + (0.179010676 - 0.037459929)^2 + (0.1391704 - 0.624831956)^2)^{(-1/(2-1))}$

Furthermore, the final results of the new member in the 27th iteration were obtained, which became a guideline for determining clusters for 30 villages. It is shown in *Table 19.*

Table 19 Result member value on the last iteration

Village	C1	C2	C3
Bonto cinde	0.9590	0.0180	0.0230
Lonrong	0.2330	0.7150	0.0520
Bajiminasa	0.0060	0.9910	0.0020
Borongloe	0.2080	0.6440	0.1480
Bonto karaeng	0.7560	0.0460	0.1980
Bonto tappalang	0.5970	0.0530	0.3500
Bonto lojong	0.0770	0.9020	0.0220
Tamatto	0.9680	0.0170	0.0150
Kahaya	0.1290	0.0230	0.8480
Bialo	0.0810	0.8970	0.0220
Bonto minasa	0.0170	0.0040	0.9790
Bonto marannu	0.7770	0.0450	0.1780
Tambangan	0.0610	0.9100	0.0300
Darubiah	0.5620	0.3600	0.0780
Welado	0.9780	0.0100	0.0130
Amali riattang	0.0630	0.0240	0.9130
Mallari	0.8750	0.0810	0.0450
Bana	0.8930	0.0450	0.0620
Bulusirua	0.0100	0.0030	0.9860
Laoni	0.0760	0.0300	0.8940
Pusungnge	0.0820	0.0330	0.8850
Praja maju	0.1240	0.0220	0.8540
Data	0.0250	0.0080	0.9670
Nusa	0.4360	0.0480	0.5160
Panyili	0.0860	0.0160	0.8980
Palajau	0.0070	0.9900	0.0030
Tanjonga	0.0710	0.0140	0.9150
Pallantikang	0.9070	0.0270	0.0660
Garassikang	0.6800	0.2450	0.0750
Pappalluang	0.0960	0.8780	0.0260

The convergence method is generally applied to fuzzy as a barrier that the algorithm has achieved optimal cluster results [60]. The general criteria of iteration termination requirements are to measure the total change in membership value or cluster center change between two iterations, if the cluster center change has passed the pre-defined threshold limit, then the iteration process continues. However, there were no significant changes, the algorithm was declared convergent and the iteration ended.

Based on the data shown in *Table 19*, the cluster mapping for each village is determined, as shown in *Figure 4.*

A comparison of research results between hierarchical and non-hierarchical clustering methods was carried out in order to get better results. As it is known that one of the non-hierarchical clustering algorithms is the k-means method [61], while hybrid SOM and FCM are the hierarchical clustering method. A comparison of hierarchical and non-hierarchical clustering methods can be made to find the most efficient algorithm [62]. The comparison results shown in *Figure 4* show that village data still has a reasonably wide distance. In contrast, in *Figure 5* on hybrid SOM and FMC, village data is seen forming clusters with closer distances, especially in cluster 3 as a priority cluster.

Based on the cluster of village data shown in *Figure 5*, it can be used as a guide and guide that there are 11 villages that will become development priorities in cluster 3, which are assumed to be a group of villages with the CSLI category at the poor level. Based on the information that has been collected stating that the villages in cluster 3 are most villages located in mountainous areas, it explains that there are similarities in community life patterns between each village, which is seen from the location of the distance value between villages from the center of cluster 3. While in clusters 1 and 2 the distance can be seen between villages is more tenuous than the center of the cluster.

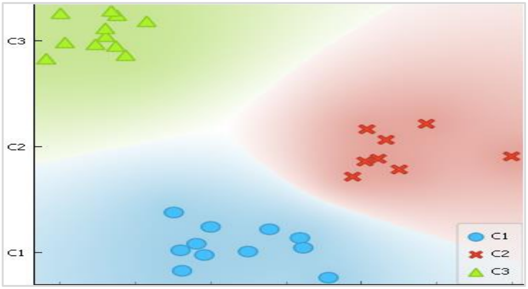

Figure 4 Clustering K-means, Silhouette

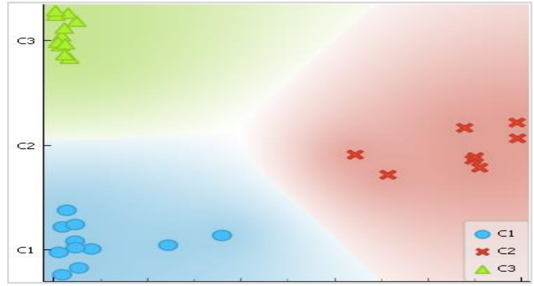

Figure 5 Clustering of SOM, FCM, Xie-Beni

4.6Recommendation the expert for the village

In this study, decisions regarding recommendations for village assistance experts are set based on the geographical location of a village. In general, the province of South Sulawesi consists of mountains and coasts.

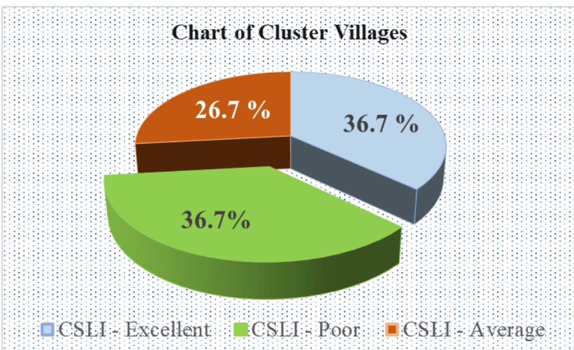

Figure 6 Chart of cluster villages

The illustration shown in *Figure 6* is that there are 26.7% of 30 villages in the CSLI-Average group, namely villages with a medium level of village development, then the number of villages in the CSLI-Excellent and CSLI-poor groups is 36.7%, or each consists of 11 villages. The grouping results show that villages in the CSLI-poor group deserve to be a development priority because many villages are still at low welfare levels.

The datasets described the properties essential for performing relevant research agar identify relevant datasets [63]. The recommendation process for placing experts in CSLI-poor cluster villages was determined based on a dataset of community comments using the cosine similarity method. The dataset is shown in *Table 20.*

Table 20 Dataset community comments

ID	*Query*
Q	*agriculture*
D1	Needed field expertise economic
D2	Needed field expertise agriculture
D3	Needed field expertise forrestry
D4	services in the government need to enhance needed field expertise social

The input data is the data from the query from the database, then the query is determined as a Dataset. Details on dataset variations are in *Table 20*. The next step is to build the TFIDF value that is shown in *Table 21*.

Table 21 Building TFIDF value

No	Term	TF					DF	IDF
		Q	D1	D2	D3	D4		
1	Needed	0	1	1	1	0	3	0.22185
2	agriculture	*1*	0	1	0	0	*2*	*0.39794*
3	economic	0	1	0	0	0	1	0.69897
..	..	..	..	..	..	..	..	..
15	to	0	0	0	0	1	1	0.69897

Based on the data in *Table* 21, it is known that the term frequency (TF) value is obtained from the number of words or the number of occurrences of words in each document, while the document frequency (DF) value is obtained from calculating the number of documents containing certain words using the formula IDF= $tfij$ x log n/df.

It is written as follows:

n = number of document = 5

DF= 2

IDF *0.39794* = log (5 / 2)

Table 22 Counting of TFIDF weight for term

No.	TERM	Q	D1	D2	D3	D4
1	Needed	0	0.22185	0.22185	0.22185	0
2	agriculture	*0.39794*	0	0.39794	0	0
3	economic	0	0.69897	0	0	0

No.	TERM	Q	D1	D2	D3	D4
..	..	..	..	..	..	..
15	to	0	0	0	0	0.69897

The TFIDF value is generated based on the calculation formula $wij = tfij \ x \ idfi$. *Table 21* is an example of applying the results of calculating the TFIDF value: Wij = 0.39794 = 1 × 0.39794. The process of calculating the TFIDF weight value is carried out using data in *Tables 21* and *22*. The next step is calculating the dot product value shown in *Table 23*. The dot product calculates the cosine similarity method to measure the similarity between two vectors in a vector space. It is written in examples: 0.158356244 = *0.39794* ×*0.39794*. Then do calculations on vector values that have keywords and documents. It is shown in *Table 24*. Key and document vector values are generated based on the formula: V_{QD} = (TFIDF (Q, D)) ^ 2. In *Table 23* is written in an example of the application of the results of calculating values V_{QD} = *0.15835624* = (0.39794 ^ 2). The final stage is to calculate the similarity value between the keyword and the document using Equation 10. The calculation results are shown in *Table* 25. In the context of cosine similarity, vector sets can be text vectors representing documents or feature vectors representing numerical data [64], as evidenced by the calculation results obtained, the largest value of cosine similarity is found in document D2, namely: *0.83641316*. It is shown as follows: S=*0.8364131*=*0.158356244*/(0.39794 ×0.47576965), as evidenced by the calculation results obtained, the largest value of cosine similarity is found in document D2.

Based on the results of calculations against proposed data from the community, the field of science is recommended according to the needs of the village. *Table 26* shows village data on the CSLI-poor group and the required fields of expertise. The data shown in the table is the result of observations and communications made to the head of the village so that the latest conditions can be known regarding the scientific needs of experts placed in villages designated as priority development areas. It is Seen in *Table 26* that most of the villages in cluster CSLI-poor are located in mountainous areas, indicating that people tend to be people at low welfare levels who live in areas with poor development levels.

Table 23 Counting dot product value

No.	Term	D1	D2	D3	D4
1	Needed	0	0	0	0
2	agriculture	0	*0.158356244*	0	0
3	economic	0	0	0	0
..	..	..	..	..	..
15	to	0	0	0	0
SUM (Q * D):		**0**	**0.15835624**	**0**	**0**

Table 24 Counting of vector key and document

No	Term	Q	D1	D2	D3	D4
1	Needed	0	0.04922	0.04921742	0.04921742	0
2	agriculture	*0.15835624*	0.00000	0.15835624	0	0
3	economic	0	0.48856	0	0	0
4	enhance	0	0.00000	0	0	0.48856
5	expertise	0	0.00939	0.00939155	0.00939155	0.00939
6	field	0	0.00939	0.00939155	0.00939155	0.00939
7	forestry	0	0	0	0.48855906	0.00000
8	government	0	0	0	0	0.48856
9	in	0	0	0	0	0.48856
10	need	0	0	0	0	0.48856
11	needed	0	0	0	0	0.48856
12	services	0	0	0	0	0.48856
13	social	0	0	0	0	0.48856

No	Term	Q	D1	D2	D3	D4
14	the	0	0	0	0	0.48856
15	to	0	0	0	0	0.48856
SUM (TFIDF (Q, D))2		0.15835624	0.55656	0.22635676	0.55655958	4.41581464
SQRT (SUM (TFIDF (Q, D))2)		0.39794	0.74603	0.47576965	0.74602921	2.10138398

Table 25 Counting of cosine similarity

ID	Text	Cosine	Similarity percentage
D1	Needed field expertise economic	0	0%
D2	*Needed field expertise agriculture*	*0.8364131*	*83.64%*
D3	Needed field expertise forestry	0	0%
D4	services in the government need to enhance needed field expertise social	0	0%

Table 26 Recommendation expert for the village

ID	Villages	Demography	Field expertise							
			Economic	Agriculture	Forestry	Social	Health	Regional Planning	Computer	Marine
	Kahaya	Mountains	√	√	√	√	√	x	x	x
	Amali Riattang		√	√	√	√	√	x	x	x
	Bulusirua		√	√	√	√	√	x	x	x
	Tanjonga		√	√	√	√	√	x	x	x
	Data		√	√	√	√	√	x	x	x
	Nusa		√	√	√	√	√	x	x	x
	Praja Maju		√	√	√	√	√	x	x	x
	Panyili	Plains	√	√	x	√	√	√	√	x
	Bonto Minasa		√	√	x	√	√	√	√	x
	Laoni	Coast	√	x	x	√	√	√	√	√
	Pusungnge		√	x	x	√	√	√	√	√

5.Discussion

This study proposes a technique to determine development priority villages using hybrid clustering techniques. The stages carried out in the application of clustering techniques include: Firstly, determine the weighting of all criteria and sub-criteria. We found that each weight assigned to the criteria and sub-criteria influenced the assessment results by respondents. Secondly, build CSLI based on the VS from 900 responses. The input used to build CSLI comes from questionnaire answers filled out by all respondents, where the data is shown in *Table 3*. We found that answers from the respondent can be used to measure village CSLI levels. Therefore, the success of this study depends heavily on the correctness of information and data from respondents. Thirdly, build village clusters using hybrid SOM, FCM, and Xie-Beni methodologies. The SOM method is used to build the initial weight of the cluster based on 3 attributes: VS, head of family, and number of residents. We found that although SOM produces cluster-shaped output, the drawback found is that the use of random initial weights will affect the accuracy of the result. Therefore, the process output in SOM is used as the initial weight in the FCM method so that it produces better village grouping output through the advanced clustering process. We designated villages in cluster 3 as development priority villages representing village groups at the CSLI-poor level, where the data is shown in *Table 19*. We compared hybrid SOM, FCM, and Xie-Beni with k-means and Silhouette. As shown in *Figures 4* and *5* that when viewed based on the distribution of data and the closest distance value from the cluster center, it is stated that the hybrid SOM-

FCM-Xie-Beni produces a better village cluster than K-means and Sillout. Finally, build scientific recommendations needed in a village. We found that comments and suggestions from respondents at the stage of filling out questionnaires can be a source of information about the needs of the field of science in a village. The recommendation process uses the cosine similarity method by calculating the percentage of word similarity values between comment data and the field of science, that is the keyword. The data shown in *Table 26* illustrate that all villages need experts in field expertise of economics, social and health sciences.

5.1Limitations of work

Several problems and limitations may occur during the process of project development. It takes a single method that can be used to validate and evaluate so that criteria can be determined dynamically. It takes a methodology to build the weight of the criteria to describe the preferences or level of importance given to each criterion. The number of respondents involved in the VS process should be determined through relevant methods to estimate the minimum sample size required to be represented by characteristics or variations within a village. Hybrid SOM, FCM, and Xie-Beni methods are often more computationally complex. Hybrid algorithms typically involve combining multiple techniques or incorporating additional parameters, which can increase the computational load and may require more expertise to apply and interpret. Although cosine similarity is a popular method and is often used in text analysis and data mining, it cannot capture the exact meaning of words, making it possible to erroneously compare texts or documents that have similar words but differ in meaning or context.

A complete list of abbreviations is shown in *Appendix I*.

6.Conclusion and future work

Based on the stages and results obtained, this study concludes that hybrid clustering techniques can used to build village groupings based on predetermined categories. Rural communities can be a source of information in the development equity program. The results of the village grouping showed that from 30 villages as a sample, there are 11 villages or 36.7% in cluster CSLI-Excellent consisting of groups of good community living standards. There are 8 villages or 26.7% in cluster CSLI-Average consisting of sufficient community living standards, while there are 11 villages or 36.7% in cluster CSLI-poor consisting of groups of Low community living standards. Villages located in cluster CSLI-poor can be proposed as priority development villages to interested parties such as the South Sulawesi government. The main limitations of the research presented are the lack of data for a longer period, and the number of villages sample is relatively less. For future studies, it is recommended to involve more data from different village condition areas and consider using more specific parameters to represent all aspects related to village development conditions.

Acknowledgment

We would like to express our gratitude for the generous support received from STMIK Profesional Makassar and Yayasan Dipanegara, who provided funding for this research project. Additionally, we extend our appreciation to our colleagues at AeU University for their valuable feedback and guidance throughout the study. Their contributions were instrumental in the successful completion of this research. Without their support, this study would not have been possible.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contribution statement

Muhammad Faisal: Conceptualization, finalization of device dimensions, result generation and interpretation and writing. **Titik Khawa Abdul Rahman:** Review, conceptualization , interpretation, writing and editing.

References

[1] Harbiankova A, Gertsberg L. Information model for sustainable rural development. Energies. 2022; 15(11):4009.

[2] Purba JT, Budiono S, Hariandja ES, Pramono R. Sustainability strategy to alleviate poverty through education, energy, GRDP, and special funds: evidence from Indonesia. Planning. 2023; 18(5):1397-406.

[3] Wilonoyudho S, Rijanta R, Keban YT, Setiawan B. Urbanization and regional imbalances in Indonesia. The Indonesian Journal of Geography. 2017; 49(2):125-32.

[4] De JE, Vijge MJ. From millennium to sustainable development goals: evolving discourses and their reflection in policy coherence for development. Earth System Governance. 2021; 7:100087.

[5] Andari RN, Ella S. Developing a smart rural model for rural area development in Indonesia. Journal Borneo Administrator. 2019; 15(1):41-58.

[6] Shcherbina EV, Gorbenkova EV. Cluster approach in rural settlement development. In IOP conference series: materials science and engineering 2020 (pp. 1-6). IOP Publishing.

[7] Tatarova S, Zateeva N. Problems of rural development: republics of Buryatia and Tyva. Bulletin of Kemerovo

State University Series Political Sociological and Economic Sciences. 2021; 2021(1):58-65.

[8] Birčiaková N, Stávková J, Antošová V. Evaluating living standard indicators. Danube. 2015; 6(3):175-88.

[9] Wahyudianto H. Village models in the future and their guidance policies. Matra Update: Journal of Policy Innovation. 2020; 4(1):47-57.

[10] Anderson-carpenter KD, Watson-thompson J, Jones MD, Chaney L. Improving community readiness for change through coalition capacity building: evidence from a multisite intervention. Journal of Community Psychology. 2017; 45(4):486-99.

[11] Prayitno G, Dinanti D, Wardani LE, Sania DP. The levels of community readiness and community characteristics in the development of tourism village (Bangelan village, Malang regency, Indonesia). International Journal of Sustainable Development & Planning. 2022; 17(4): 1181-8.

[12] Fawzy NA, Ammer SA, Elmouety EA. Requirements for reaching the sustainable goals of development to plan new areas in Arab Republic of Egypt. International Journal of Sustainable Development & Planning. 2021; 16(3): 579-90.

[13] Kusio T, Fiore M. Which stakeholders sector matters in rural development? that is the problem. Energies. 2022; 15(2):454.

[14] Xiang S, Zhou M, Huang L, Shan L, Wang K. Assessing the dynamic land utilization efficiency and relevant driving mechanism in in-situ urbanized rural areas: a case study of 1979 administrative villages in Hangzhou. Environmental Impact Assessment Review. 2023; 101:107111.

[15] Kapoor N, Ahmad N, Nayak SK, Singh SP, Ilavarasan PV, Ramamoorthy P. Identifying infrastructural gap areas for smart and sustainable tribal village development: a data science approach from India. International Journal of Information Management Data Insights. 2021; 1(2):100041.

[16] Li Y, He J, Yue Q, Kong X, Zhang M. Linking rural settlements optimization with village development stages: a life cycle perspective. Habitat International. 2022; 130:102696.

[17] Hämäläinen J, Jauhiainen S, Kärkkäinen T. Comparison of internal clustering validation indices for prototype-based clustering. Algorithms. 2017; 10(3):1-14.

[18] Verma H, Gupta A, Kumar D. A modified intuitionistic fuzzy c-means algorithm incorporating hesitation degree. Pattern Recognition Letters. 2019; 122:45-52.

[19] Prasetyo H. Grouping areas according to economic potential using a modified fuzzy k-prototypes algorithm to determine village development targets. Journal of Technology and Computer Systems. 2022; 10(1):46-52.

[20] Sonu TS, Bhagyanathan A. The impact of upstream land use land cover change on downstream flooding: a case of Kuttanad and Meenachil River Basin, Kerala, India. Urban Climate. 2022; 41:101089.

[21] Krishnan VS, Firoz CM. Assessment of regional environmental quality using cluster analysis. Environmental Development. 2023; 45:100832.

[22] Yu CA, Li G, Cao Y, Wang J, Fang X, Zhou L, et al. Distinct types of restructuring scenarios for rural settlements in a heterogeneous rural landscape: application of a clustering approach and ecological niche modeling. Habitat International. 2020; 104:102248.

[23] Nasution BI, Kurniawan R, Siagian TH, Fudholi A. Revisiting social vulnerability analysis in Indonesia: an optimized spatial fuzzy clustering approach. International Journal of Disaster Risk Reduction. 2020; 51:101801.

[24] Padilla-rivera A, Do CBB, Arcese G, Merveille N. Social circular economy indicators: selection through fuzzy delphi method. Sustainable Production and Consumption. 2021; 26:101-10.

[25] Merényi E, Taylor J. Empowering graph segmentation methods with SOMs and CONN similarity for clustering large and complex data. Neural Computing and Applications. 2020; 32:18161-78.

[26] Qian J, Nguyen NP, Oya Y, Kikugawa G, Okabe T, Huang Y, et al. Introducing self-organized maps (SOM) as a visualization tool for materials research and education. Results in Materials. 2019; 4:1-14.

[27] Chen Y, Qin B, Liu T, Liu Y, Li S. The comparison of SOM and K-means for text clustering. Computer and Information Science. 2010; 3(2):268-74.

[28] Zhang S, Li X, Zong M, Zhu X, Cheng D. Learning k for KNN classification. ACM Transactions on Intelligent Systems and Technology. 2017; 8(3):1-9.

[29] Ding Y, Fu X. Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm. Neurocomputing. 2016; 188:233-8.

[30] Mu S, Chen X. A study of participatory art interventions based on the context of rural revitalisation. International Journal of Education and Humanities. 2022; 5(1):32-5.

[31] Liang J. Value analysis and realization of artistic intervention in rural revitalization based on the fuzzy clustering algorithm. Scientific Programming. 2022; 2022:1-9.

[32] Kusrini K. Grouping of retail items by using K-Means clustering. Procedia Computer Science. 2015; 72:495-502.

[33] Singh M, Bhattacharjee R, Sharma N, Verma A. An improved xie-beni index for cluster validity measure. In fourth international conference on image information processing 2017 (pp. 1-5). IEEE.

[34] Makkar K, Kumar P, Poriye M, Aggarwal S. Improvisation in opinion mining using data preprocessing techniques based on consumer’s review. International Journal of Advanced Technology and Engineering Exploration. 2023; 10(99):257-77.

[35] Pikies M, Ali J. Analysis and safety engineering of fuzzy string matching algorithms. ISA Transactions. 2021; 113:1-8.

[36] Kaur P, Kaur K. Clustering techniques in data mining for improving software architecture: a review.

International Journal of Computer Applications. 2016; 139(9): 35-39.

[37] Ezugwu AE, Agbaje MB, Aljojo N, Els R, Chiroma H, Abd EM. A comparative performance study of hybrid firefly algorithms for automatic data clustering. IEEE Access. 2020; 8:121089-118.

[38] Arunachalam D, Kumar N. Benefit-based consumer segmentation and performance evaluation of clustering approaches: an evidence of data-driven decision-making. Expert Systems with Applications. 2018; 111:11-34.

[39] Lee KJ, Yun ST, Yu S, Kim KH, Lee JH, Lee SH. The combined use of self-organizing map technique and fuzzy c-means clustering to evaluate urban groundwater quality in Seoul metropolitan city, South Korea. Journal of Hydrology. 2019; 569:685-97.

[40] Hou G, Xiong J, Zhou G, Gong L, Huang C, Wang S. Coordinated control system modeling of ultra-supercritical unit based on a new fuzzy neural network. Energy. 2021; 234:121231.

[41] Mota VC, Damasceno FA, Leite DF. Fuzzy clustering and fuzzy validity measures for knowledge discovery and decision making in agricultural engineering. Computers and Electronics in Agriculture. 2018; 150:118-24.

[42] Verma R, Mittal A. Multiple attribute group decision-making based on novel probabilistic ordered weighted cosine similarity operators with Pythagorean fuzzy information. Granular Computing. 2023; 8(1):111-29.

[43] Sudipa IG, Aryati KS. Weighting approach with surrogate weighting procedures for the simple additive weighting method in multicriteria decision making. International Journal of Natural Science and Engineering. 2019;3(3):113-21

[44] Ishtaiwi A, Alshahwan F, Jamal N, Hadi W, Abuarqoub M. A dynamic clause specific initial weight assignment for solving satisfiability problems using local search. Algorithms. 2021; 14(1):1-14.

[45] Yin C, Zhang B, Liu W, Du M, Luo N, Zhai X, et al. Geographic knowledge graph attribute normalization: improving the accuracy by fusing optimal granularity clustering and co-occurrence analysis. ISPRS International Journal of Geo-Information. 2022; 11(7):360.

[46] Huang S, Guo Y, Yang N, Zha S, Liu D, Fang W. A weighted fuzzy c-means clustering method with density peak for anomaly detection in IoT-enabled manufacturing process. Journal of Intelligent Manufacturing. 2021; 32:1845-61.

[47] Hwang P, Kim Y. Data cleaning of sound data with label noise using self organizing map. In 16th international conference on ubiquitous information management and communication (IMCOM) 2022 (pp. 1-5). IEEE.

[48] Yogapriya P, Geetha P. Dengue disease detection using K-means, hierarchical, Kohonen-SOM clustering. International Journal of Innovative Technology and Exploring Engineering. 2019; 8(10):904-7.

[49] Jamil A, Hameed AA, Orman Z. A faster dynamic convergency approach for self-organizing maps. Complex & Intelligent Systems. 2023; 9(1):677-96.

[50] Zhu E, Ma R. An effective partitional clustering algorithm based on new clustering validity index. Applied Soft Computing. 2018; 71:608-21.

[51] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 1987; 20:53-65.

[52] Wang HY, Wang JS, Wang G. Combination evaluation method of fuzzy c-mean clustering validity based on hybrid weighted strategy. IEEE Access. 2021; 9:27239-61.

[53] Sitepu KA, Sitompul OS, Situmorang Z. Analysis of fuzzy c-means and analytical hierarchy process (AHP) models using xie-beni index. In international conference of computer science and information technology 2019 (pp. 1-6). IEEE.

[54] Deng J, Deng Y. Information volume of fuzzy membership function. International Journal of Computers Communications & Control. 2021; 16(1):1-14.

[55] Wu T, Zhou Y, Xiao Y, Needell D, Nie F. Modified fuzzy clustering with segregated cluster centroids. Neurocomputing. 2019; 361:10-8.

[56] Rahmalia D. Optimizing the membership degree of fuzzy inference system (FIS) and fuzzy clustering means (FCM) in weather data using firefly algorithm. CAUCHY: Journal of Pure Mathematics and Applications. 2021; 6(4):169-80.

[57] Kasmi O, Baina A, Bellafkih M. Fuzzy logic based clustering algorithm for management in critical infrastructure. Cluster Computing. 2021; 24(1):433-58.

[58] Obiedat M, Al-yousef A, Khasawneh A, Hamadneh N, Aljammal A. Using fuzzy c-means for weighting different fuzzy cognitive maps. International Journal of Advanced Computer Science and Applications. 2020; 11(5):545-51.

[59] Azam MH, Hasan MH, Kadir SJ, Hassan S. Prediction of sunspots using fuzzy logic: a triangular membership function-based fuzzy c-means approach. International Journal of Advanced Computer Science and Applications. 2021; 12(2):357-62.

[60] Yapalıa R, Coskunb H, Gurdalc U. Statistical convergence on L− fuzzy normed space. FILOMAT. 2023; 37(7):2077-85.

[61] Alamtaha Z, Djakaria I, Yahya NI Implementation of Hierarchical Clustering and Non-Hierarchical Clustering Algorithms for Grouping Social Media Users. ESTIMASI: Journal of Statistics and its Application. 2023: 33-43.

[62] Ahuja R, Agarwal PP, Kumar T. Clustering methods analysis in the E-learning. In proceedings of the 3rd international conference on communications and cyber physical engineering 2021 (pp. 1297-308). Springer Singapore.

[63] Omelina L, Goga J, Pavlovicova J, Oravec M, Jansen B. A survey of iris datasets. Image and Vision Computing. 2021; 108:104109.

[64] Zhang R, Li T, Liu C. Joint inversion of multiphysical parameters based on a combination of cosine dot-gradient and joint total variation constraints. IEEE Transactions on Geoscience and Remote Sensing. 2021; 60:1-10.

Muhammad Faisal obtained his S.SI degree from the Information System Department at STMIK Profesional MAKASSAR, Makassar, Indonesia, in 2011. He further pursued his education and earned an M.T degree from the Masters in Information Technology program at the Postgraduate School, Hasanuddin University, Makassar, Indonesia, in 2014. Presently, he holds the position of a lecturer and researcher at the Department of Computer Science, STMIK Profesional Makassar, Indonesia. His current research interests encompass various fields, including Mobile Programming, Data Mining, Decision Support System, Expert System, Web Applications, and Image Processing.
Email: muhfaisal@stmikprofesional.ac.id.

Titik Khawa Abdul Rahman earned a B.Sc. in Electrical and Electronics Engineering from Loughborough University of Technology, UK. She completed her Ph.D. in Power Systems at the University of Malaya. Currently, she holds the position of Deputy Vice Chancellor at Asia e University. Her areas of expertise include artificial intelligence, biologically inspired optimization, machine learning, power system analysis and operation, and smart grid technologies.
Email: titik.khawa@aeu.edu.my

Appendix I

S. No.	Abbreviation	Description
1	AHP	Analytic Hierarchy Process
2	BMU	Best Matching Unit
3	CSLI	Community Standard of Living Index
4	DF	Document Frequency
5	EQ	Environmental Quality
6	FCM	Fuzzy C-Means
7	RVP	Respondents Village Patriarch
8	RVR	Respondents Village Residents
9	SOM	Self-Organizing Map
10	TC	Total of Community
11	TFIDF	Term Frequency and Inverse Document Frequency
12	TF	Term Frequency
13	TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
14	TP	Total of Patriarch
15	VP	Village Patriarch
16	VR	Village Residents
17	VS	Village Score