BAREKENG: Journal of Mathematics and Its Applications
June 2024 Volume 18 Issue 2 Page 0821–0836
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss2pp0821-0836

APPLICATION OF K-MEANS AND FUZZY C-MEANS ALGORITHMS TO DETERMINE FLOOD VULNERABILITY CLUSTERS (CASE STUDY: KUTAI KARTANEGARA REGENCY)

Desi Nurjanah[1], Indira Anggriani[2*], Primadina Hasanah[3]

[1,2]*Mathematics Department, Institut Teknologi Kalimantan*
[3]*Actuarial Sciences Department, Institut Teknologi Kalimantan*
Balikpapan, 71627, Indonesia

*Corresponding author's e-mail: ***indira@lecturer.itk.ac.id***

Article History:

Received: 4th November 2023
Revised: 4th January 2024
Accepted: 5th March 2024
Published: 1st June 2024

Keywords:

Clustering;
Flood;
Fuzzy C-Means;
K-Means Algorithm.

ABSTRACT

Flooding show situation where areas that are not usually inundated, such as farmland and settlements, and city district areas, become inundated due to water. Floods can to occur when the flow of water on rivers or waste channels overrun its normal measurements. This study describes the K-Means and Fuzzy C-Means Algorithm methods for clustered flood-prone areas built on Districts in Kutai Kartanegara Regency. This research begins with data collection in the character of rainfall, land elevation, the number of victims affected, the quantity of damaged houses, the quantity of damage to facilities and the quantity of flood events. Before the data is processed using these two methods, data normalization will be carried out in a dataset which aims to shape the data into positional values from the same range. K-Means and Fuzzy C-Means are accustomed to identifying groups in each sub-district in Kutai Kartanegara Regency that have a level of vulnerability to floods. At this stage, 3 initial clusters were carried out, namely high, medium, and low vulnerability clusters. The validity test produces a Silhouette Index value of 0.574283589 and a Partition Coefficient Index of 0.78905. The outcome of the K-Means method with the standard deviation within and between clusters are 0.5131 and the Fuzzy C-Means method for the standard deviations within and between clusters is 0.3489. based uppon value of the silhouette index, partition coefficient index and standard deviation within and between clusters it results that Fuzzy C-Means is the best method of this study.

How to cite this article:

D. Nurjanah, I. Anggriani and P. Hasanah., "APPLICATION OF K-MEANS AND FUZZY C-MEANS ALGORITHMS TO DETERMINE FLOOD VULNERABILITY CLUSTERS (CASE STUDY: KUTAI KARTANEGARA REGENCY)," *BAREKENG: J. Math. & App.*, vol. 18, iss. 2, pp. 0821-0836, June, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Kutai Kartanegara Regency is one of the buffer areas for the National Capital City, where the sub-districts included in the buffer zone include several sub-districts which include Loa Janan, Loa Kulu, Sanga-Sanga, and Muara Jawa. As a buffer area this will have a major influence on population growth and population density as well as infrastructure growth which will reduce water catchment areas. According to data based on the 2020 Indonesian Disaster-Prone Index (IRBI) report released by BNPB, Kutai Kartanegara Regency is classified as a medium level disaster alert category [1] .

Kutai Kartanegara Regency is one of the supporting areas of the National Capital City, where the sub-districts included in the buffer zone include several sub-districts which include Loa Janan, Loa Kulu, Sanga-Sanga, and Muara Jawa. As a buffer area this will have a major influence on population growth and population density as well as infrastructure growth which will reduce water catchment areas [2].

Based on the results of the BNPB index, the flood disaster that hit several areas of Kutai Kartanegara Regency was caused by high rainfall and erosion from the Belayan river [3]. When high rainfall causes the overflow of the river's water discharge, it causes flooding and causes the impact received by the flood victims. Floods not only cause inundation in housing and settlements but can also cause damage to public infrastructure facilities and even cause fatalities. Losses will be even greater if economic activity is disrupted.

In the sub-districts in Kutai Kartanegara Regency, the frequency of major floods occurs at several sub-district points, including the Kembang Janggut and Muara Kaman sub-districts. Therefore, it is necessary to cluster the Kutai Kartanegara flood data by sub-district to group the data into several clusters. A previous research study that using the K-Means method for clustering areas prone to natural disasters in Indonesia which divides into 3 cluster groups (high, medium, low) and uses the Silhouette Index to be able to test its validity level [4, 5, 6].

The commonly used clustering methods are K-Means and Fuzzy C-Means. The K-Means algorithm has the advantage that it can be easily understood, implement large-scale data, and reduce the complexity of existing data [7] .In implementing problem-solving, the K-Means algorithm requires numerical attribute data. Meanwhile, Fuzzy C-Means aims to minimize the objective function specified in the grouping process which generally tries to reduce variation within a cluster and increase variation between clusters. The advantages of Fuzzy C-Means are determining optimal cluster points and grouping (clustering) more than one variable simultaneously [8].

Previous research studies used the Fuzzy C-Means and K-Means methods to map crime-prone areas in Semarang City, then validated with the results of processing tests, namely the Partition Coefficient Index Test on Fuzzy C-Means and Silhouette Index in K-Means [9]. The research with the K-Means method used for regional clustering prone to natural disasters in Indonesia which is divided into 3 clustering groups (high, medium, low) and uses the Silhouette Index to test the level of validity [4]. Therefore, a comparison will be carried out between clustering using the K-Means and Fuzzy C-Means algorithms to obtain the optimal clustering method. Clustering is needed so that relevant institutions can carry out disaster management quickly but remain well coordinated and can assist in clustering areas prone to flooding. The data used in this research are factors that influence the occurrence of floods.

Based on these problems, the K-Means and Fuzzy C-Means algorithms will be used for clustering flood-prone areas in Kutai Kartanegara Regency. By testing the validity with the silhouette index on K-Means and the partition coefficient index on fuzzy C-Means. Then by calculating the standard deviation within and between clusters on the two methods to determine the best performance between K-Means and Fuzzy C-Means.

2. RESEARCH METHODS

This research includes several methods for forming clustering and determining the best method. Each stage will be explained in more detail in the following subchapters.

2.1 Data Normalization

The normalization process in the role of data mining is very significant in achieving optimal results. By normalizing, the use of algorithms becomes more efficient. Data normalization in a dataset aims to change the data values into the same range. In data normalization, each value in the dataset will be narrowed down to a scale between 0 to 1, where the smallest value will be represented as 0 and the largest value as 1 **[10]**. To find data values using min-max normalization, calculations are carried out in the following **Equation (1)**

$$v' = \frac{v - v_{min}}{v_{max} - v_{min}} \tag{1}$$

2.2 K-Means

K-Means is a method that requires criteria from clusters, and groups all *i*-th data into cluster criteria by maximizing the similarity between elements in a cluster, while reducing similarity with other clusters. In other words, this algorithm divides the data into k clusters in such a way that the data in one cluster has a high level of accuracy, while the similarity with other clusters is very low **[11, 12]**. The following are the steps for K-Means:

1. Determine the number of clusters.
2. Determine the centroid of each cluster. At the beginning of the iteration, the centroid of each cluster is determined randomly.
3. The next step is to calculate the distance from the object with the centroid. To calculate the distance, **Equation (2)** can be used, namely:

$$D(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\,; i = 1, 2, 3, \dots, n \tag{2}$$

where:

$D(x,y)$ = distance of data to the center of the cluster center
x_i = *i*-th data
y_i = *i*-th *centroid*
n = lots of data

1. Then group the data according to the minimum value with the cluster center in **Equation (3)**.

$$Min \sum_{k=1}^{c} D(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\,; i = 1, 2, 3, \dots, n \tag{3}$$

2. The average member data in the cluster can be used to calculate a new cluster center value, using the equation below with $i = 1, 2, 3, \dots, n$ and $j = 1, 2, 3, \dots, m$:

$$c_{ij} = \frac{\sum_{i=1}^{n} x_{ij}}{n} \tag{4}$$

where:

c_{ij} = the latest centroid in the *j*-th iteration
x_{ij} = *k*-th cluster data
n = lots of *k*-th cluster group data

3. Repeat the calculation steps repeatedly **Equation (2)** until no members from each cluster move to another cluster.

2.3 Silhouette Index

The Silhouette Index (SI) is applied as a tool to confirm the structural strength of data, single clusters or all clusters. This algorithm is often applied to evaluate clusters by collecting cohesion and separation values. The SI value ranges in the interval $(-1, 1)$. According to a study conducted by Struyf, Hubert, and Rousseeuw in 2007 **[13, 14]**, SI provides useful information in measuring the quality and suitability of clusters. The following are the criteria for the silhouette index:

Table 1. Silhouette Index Interpretation

Silhouette Interval	Interpretation
0.71 - 1.0	Strong Structure
0.51 - 0.70	Good Structure
0.26 - 0.50	Sufficient Structure
< 0.25	Bad Structure

The initial step in calculating the silhouette index is to take the average value of the distance between the i-th data and all existing data in the same cluster. The following is the equation $a(i)$ as follows.

$$\boldsymbol{a(i) = \frac{1}{|A|-1}\sum_{j \in A, j \neq i} d(i,j)} \quad (5)$$

where:

$\boldsymbol{A}$: the amount of data in cluster $\boldsymbol{A}$

$\boldsymbol{d(i,j)}$: distance between data i and data in the same cluster

The next step is to calculate the $b(i)$ value which is the minimum value of the distance between the i-th data and all different clusters. This calculation is carried out from cluster 1 to cluster 3. The following is the equation of $\boldsymbol{d(i,C)}$: as follows:

$$\boldsymbol{d(i,C) = \frac{1}{|C|}\sum_{j \in C} d(i,j)\ ; C \neq A} \quad (6)$$

where:

$\boldsymbol{C}$: amount of data in cluster $\boldsymbol{C}$

$\boldsymbol{d(i,j)}$: distance of the i-th data to data in other clusters

Next, after calculating the $\boldsymbol{d(i,C)}$ value for the entire $\boldsymbol{C \neq A}$ cluster. Then determine the minimum value represented as $b(i)$ as follows:

$$\boldsymbol{b(i) = \min d(i,C)} \quad (7)$$

The next step is to find the silhouette coefficient value for each data as follows.

$$\boldsymbol{s(i) = \frac{b(i)-a(i)}{max(a(i),b(i))}} \quad (8)$$

After getting the silhouette value from each data, to calculate the average silhouette value use the following equation:

$$\boldsymbol{SC = \frac{1}{n}\sum_{i=1}^{n} s(i)} \quad (9)$$

Where:
$\boldsymbol{n}$: amount of data
$\boldsymbol{s(i)}$: silhouette value of each data

2.4 Fuzzy C-Means

Fuzzy C-Means Clustering is a cluster analysis approach that considers membership levels as a fuzzy set which is the basis for weights in the clustering process. The following are some of the calculation processes in the Fuzzy C-Means method, as follows [15, 16]:

1. The data to be grouped *(X)* is in the form of a matrix measuring $\boldsymbol{n} \times \boldsymbol{m}$ ($\boldsymbol{n}$ is the number of data samples, m is the attribute of each data). $\boldsymbol{X_{ij}}$ i-th sample data $i = 1, 2, \ldots, n$, j-th variable $j = 1, 2, \ldots, m$.
2. Set the number of clusters ($\boldsymbol{c}$), weight exponent ($\boldsymbol{w}$), maximum number of iterations (MaxIter), desired smallest error tolerance ($\boldsymbol{\xi}$), initial objective function ($\boldsymbol{P_0}$), and initial iteration ($\boldsymbol{t}$).
3. Calculate the number of each row to generate a random number matrix shown in **Equation (10)**.

$$\boldsymbol{Q_i} = \sum_{k=1}^{c} \boldsymbol{\mu_{ik}} = \boldsymbol{1} \tag{10}$$

4. Calculate the centroid value V with $k = 1, 2, \ldots, c$ and $j = 1, 2, \ldots, m$.

$$\boldsymbol{V_{kj}} = \frac{\sum_{i=1}^{n}\left((\boldsymbol{\mu_{ik}})^{w} \times \boldsymbol{X_{ij}}\right)}{\sum_{i=1}^{n}(\boldsymbol{\mu_{ik}})^{w}} \tag{11}$$

where:
V_{kj} = k-th cluster center
μ_{ik} = j-th iteration membership degree
X_{ij} = initial data after data normalization

5. Calculating changes in the partition matrix or increasing the membership level of each data in each cluster using $i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, c$ as shown in **Equation (12)**

$$\boldsymbol{\mu_{ik}} = \frac{\left[\sum_{j=1}^{m}\left(\boldsymbol{X_{ij}} - \boldsymbol{V_{kj}}\right)^{2}\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}\left(\boldsymbol{X_{ij}} - \boldsymbol{V_{kj}}\right)^{2}\right]^{\frac{-1}{w-1}}} ; \boldsymbol{w > 1} \tag{12}$$

6. Calculating the objective function at the t-th iteration, $\boldsymbol{P_t}$ is shown in **Equation (13)**

$$\boldsymbol{P_t} = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(\boldsymbol{X_{ij}} - \boldsymbol{V_{kj}}\right)^{2}\right](\boldsymbol{\mu_{ik}})^{w}\right) \tag{13}$$

Equation (13) explains the distance of data to the cluster center multiplied by the degree of membership and $\boldsymbol{P_t}$ is the total calculation of all clusters.

a. If $(|P_t - P_{t-1}| < \xi)$ or $(t > MaxIter)$ then stop;
b. If not: $t = t + 1$, repeat step 4.

The condition above shows that if the iteration condition is more than the specified maximum iteration, then the iteration stops regardless of the resulting objective function value.

2.5 Partition Coefficient Index

PCI is intended to measure the level of accuracy of membership levels without considering the value of the data which reflects the spread of the data. This index has a value in the interval [0,1], where a value close to 1 indicates better cluster quality [17]. The following is the partition coefficient index equation:

$$\boldsymbol{PCI = \frac{1}{N}\left(\sum_{i=1}^{N}\sum_{j=1}^{K} U_{ij}^{2}\right)} \tag{14}$$

2.6 Cluster Standar Deviation

The clustering process can be carried out using one or more types of clustering methods, with the aim of finding the best performance of the two approaches to objectively evaluate the validity of cluster analysis. Therefore, determining the best method is carried out using and considering the standard deviation within clusters ($\boldsymbol{S_w}$) and between clusters ($\boldsymbol{S_b}$) by calculating and comparing [18].

The standard deviation in the cluster ($\boldsymbol{S_w}$) is calculated using the equation below.

$$\boldsymbol{S_w = K^{-1}\sum_{k=1}^{K} S_k} \tag{15}$$

where:
K = number of clusters built
S_k= standard deviation of the k-th cluster
The following is the standard deviation equation for the k-th cluster.

$$\boldsymbol{S_k = \sqrt{\frac{\sum(x_{ik}-\bar{x}_k)^2}{n-1}}} \tag{16}$$

S_k = standard deviation of the k-th cluster
x_{ik} = the i-th data in cluster k which has been added up by all the variables
$\bar{x}_k$ = average of k clusters
n = the number of data members in the k-th cluster

The standard deviation between clusters S_b, is calculated using the equation below.

$$\boldsymbol{S_b = \left[(K-1)^{-1}\sum_{k=1}^{K}(\bar{X}_k - \bar{X})^2\right]^{\frac{1}{2}}} \tag{17}$$

where:
$\bar{X}_k$ = average of the k-th cluster
$\bar{X}$ = average of the entire cluster
K = number of clusters

3. RESULTS AND DISCUSSION

The results and discussion will explain data normalization, K-Means calculations, silhouette index calculations, fuzzy C-Means calculations, Partition Coefficient Index Calculation, and Standard Deviation Cluster, which will be explained in more detail in the following subsection.

3.1 Data Collection and Preprocessing

In this research, the data preparation stage was carried out before carrying out the K-Means and Fuzzy C-Means algorithm steps. Collecting data on the factors and impacts of floods that occurred in Kutai Kartanegara Regency in 2022 from the Regional Disaster Management Agency, the Central Statistics Agency of Kutai Kartanegara Regency and the Meteorology, Climatology and Geophysics Agency.

Table 2. Descriptive Statistics of Data

Descriptive Statistics	Unit	Minimum	Maximum	Mean	Standard Deviation
X_1 (Rainfall)	mm	2673.9	3265	2745.2	30.775
X_2 (Land Height)	Mdpl	3	95	18.222	4.931
X_3 (Affected Victims)	Person	**0**	25962	3174.8	1574.6
X_4 (Submerged Houses)	Unit	**0**	3739	452.111	229.918
X_5 (Damage to School Facilities)	Unit	**0**	9	0.889	0.523
X_6 (Damage to Facilities in Places of Worship)	Unit	**0**	10	0.944	0.597
X_7 (Damage to Health Facilities)	Unit	**0**	3	0.389	0.200
X_8 (Number of Flood Event)	Event	**0**	27	**4**	**1.686**

3.2 Data Normalization

The purpose of data normalization is to change data that has a different range of values into data with a uniform range of values, so that the data mining process becomes unbiased. In the data normalization process for attributes such as rainfall, land elevation, flood victims, submerged houses, damage to facilities (schools, places of worship, and health facilities), as well as the number of flood events, these values will be normalized to a range of values [0, 1]. This means that the lowest value will be represented as 0 and the highest value as 1, so that all attributes have the same range of values. Then minimum and maximum normalization will be carried out with **Equation (1)** for all variables which can be implemented in **Table 3**.

Table 3. Data Normalization Results

Subdistrict	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Anggana	0.0782	0.0109	0	0	0	0	0	0
Kembang Janggut	0.0782	0.0870	0.52	1	1	1	0.333	1
Kenohan	0.0782	0.0761	0.04	0.0834	0	0	0.333	0.111
Kota Bangun	0.0782	0.0652	0.29	0.4592	0	0	0	0.111
Loa Janan	0.0782	0.1522	0	0	0	0	0	0.074
Loa Kulu	0.0782	0.2935	0	0	0	0	0	0
Marang Kayu	0.0782	0.1304	0.09	0.0735	0	0	0	0.074
Muara Badak	0.0782	0.1413	0.09	0.0864	0.22	0.30	1.00	0.148
Muara Jawa	0.0782	0.0543	0	0	0	0	0	0
Muara Kaman	0.00	0.0109	1.00	0.4343	0.33	0.40	0.667	0.667
Muara Muntai	0.0782	0.0217	0	0	0	0	0	0
Muara Wis	0.0782	0.00	0.01	0.0264	0.22	0	0	0.18
Samboja	1.00	0.1848	0.05	0.013105	0	0	0	0.037
Sanga-Sanga	0.0782	0.1957	0	0	0	0	0	0
Sebulu	0.0782	0.3152	0.12	0	0	0	0	0.037
Tabang	0.00	1.00	0	0	0	0	0	0.222
Tenggarong	0.0782	0.0761	0	0	0	0	0	0
Tenggarong Seberang	0.0782	0.1630	0	0	0	0	0	0

3.3 K-Means Calculation

Based on the data obtained, it contains 8 variables that have been normalized. In the grouping process, there are 3 clusters. In the first stage of calculating the K-Means Clustering approach, namely determining the centroid, this determination is carried out by considering the value of the number of flood events and the height of the land in each sub-district as in **Table 4**.

Table 4. Initial Centroid

Subdistrict	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Muara Kaman	0	0.0108	1	0.4343	0.333	0.4	0.667	0.667
Kembang Janggut	0.0782	0.0869	0.520	1	1	1	0.333	1
Marang Kayu	0.0782	0.1304	0.095	0.0735	0	0	0	0.074

The distance calculation is calculated using **Equation (6)**, the calculation is calculated from region 1 to region 18. Then proceed to the calculation of centroid 2 and 3. The results of grouping data for the 1st iteration by determining the closest (minimum) distance between the three resulting centroids. The next step is to carry out the second iteration by determining the new centroid.

Table 5. First Iteration Clusterization Results

No	Subdistrict	C1	C2	C3	Clusterization Result
1	Anggana	1.5345	2.0947	**0.1849**	3
2	Kembang Janggut	1.2155	**0.9259**	2.0377	2
3	Kenohan	1.1956	1.7546	**0.3592**	3
4	Kota Bangun	1.1109	1.5707	**0.4534**	3
5	Loa Janan	1.3913	1.8416	**0.1427**	3
6	Loa Kulu	1.4107	1.8505	**0.2024**	3
7	Marang Kayu	1.3004	1.7770	**0.0741**	3
8	Muara Badak	**1.0583**	1.6064	1.0778	1
9	Muara Jawa	1.3828	1.8393	**0.1421**	3
10	Muara Kaman	**0.6667**	1.3864	1.4594	1
11	Muara Muntai	1.3822	1.8401	**0.1619**	3
12	Muara Wis	1.3382	1.7203	**0.3302**	3
13	Samboja	1.7068	2.0520	**0.9304**	3
14	Sanga-Sanga	1.3944	1.8422	**0.1366**	3
15	Sebulu	1.3275	1.8214	**0.2052**	3
16	Tabang	1.7123	2.0666	**0.9089**	3
17	Tenggarong	1.3837	1.8390	**0.1317**	3
18	Tenggarong Seberang	1.3905	1.8405	**0.1243**	3

The new centroid for the second iteration is determined based on the sum of all data members in the *k*-th cluster divided by the amount of data in that cluster. The following are the results of calculating the average value from each cluster.

Table 6. Second Iteration Centroid

Second Iteration Centroid	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
C1	0.039	0.547	0.076	0.260	0.278	0.350	0.833	0.407
C2	0.078	0.520	0.087	1	1	1	0.333	1
C3	0.134	0.039	0.183	0.044	0.01	0.000	0.022	0.057

In the next iteration, the same calculations were carried out as in the 1st iteration step, the results in **Table 7** were obtained as follows.

Table 7. First and Second Iteration Clusterization Results

No	Subdistrict	C1	C2	C3	Result Second Iteration
1	Anggana	1.197	2.095	**0.200**	3
2	Kembang Janggut	1.320	**0.943**	2.056	2
3	Kenohan	0.859	1.755	**0.354**	3
4	Kota Bangun	1.005	1.571	**0.518**	3
5	Loa Janan	1.129	1.842	**0.117**	3
6	Loa Kulu	1.145	1.851	**0.140**	3
7	Marang Kayu	1.069	1.777	**0.127**	3
8	Muara Badak	**0.544**	1.606	1.059	1
9	Muara Jawa	1.124	1.839	**0.154**	3
10	Muara Kaman	**0.848**	1.386	1.498	1
11	Muara Muntai	1.125	1.840	**0.182**	3
12	Muara Wis	1.094	1.720	**0.339**	3
13	Samboja	1.478	2.052	**0.868**	3
14	Sanga-Sanga	1.130	1.842	**0.087**	3
15	Sebulu	1.094	1.821	**0.181**	3
16	Tabang	1.472	2.067	**0.860**	3
17	Tenggarong	1.124	1.839	**0.137**	3
18	Tenggarong Seberang	1.127	1.841	**0.088**	3

If the results of the second iteration do not change, the iteration calculation stops.

3.4 Silhouette Index Calculation

The silhouette index calculation is used to find out how strong the data structure is in the K-Means clustering algorithm. Test the strength of the cluster structure in K-Means with the Silhouette Index. The following are the silhouette index results in **Table 8**.

Table 8. Silhoutte Score Calculation Results

Cluster	Subdistrict	$a(i)$	$b(i)$	$s(i)$
1	Muara Badak	1.1643	1.1210	-0.0372
	Muara Kaman	1.1643	1.2582	0.0746
2	Kembang Janggut	2.0961	1.5354	-0.2675
3	Anggana	0.2933	1.3106	0.7762
	Kenohan	0.4295	1.0487	0.5904
	Kota Bangun	0.5994	1.2181	0.5080
	Loa Janan	0.2342	1.2943	0.8190
	Loa Kulu	0.2244	1.3143	0.8292
	Marang Kayu	0.2714	1.2467	0.7823
	Muara Jawa	0.2669	1.3108	0.7964
	Muara Muntai	0.2594	1.3106	0.8020
	Muara Wis	0.4008	1.2310	0.6744
	Samboja	1.0462	1.6216	0.3548
	Sanga-Sanga	0.2417	1.3196	0.8169
	Sebulu	0.3260	1.2872	0.7468
	Tabang	0.9151	1.5891	0.4241
	Tenggarong	0.2339	1.3109	0.8216
	Tenggarong Seberang	0.2303	1.3161	0.8250

Based on **Equation (9)** the final silhouette value (global silhouette) is the overall average value of $s(i)$

$$SI = \frac{1}{n}\sum_{i=1}^{n} s(i) = \frac{10.3371}{18} = 0.574283589$$

It can be seen from **Table 1** that the interpretation of the K-Means calculation results (Kauffman and Rouseeuw criteria table) is that the resulting silhouette value shows the strength structure of the data from the K-Means method used is good.

3.5 Fuzzy C-Means Calculation

The Fuzzy C-Means calculation includes several stages, namely:

1. Formation of data matrix after normalization. data for the ith region ($i = 1,2,\dots,n$), variable or attribute for the j-th ($j = 1,2,\dots,n$) measuring 18×8 as follows:

$$X_{18\times8} = \begin{bmatrix} 0.0782 & 0.0108 & 0 & \cdots & 0 \\ 0.0782 & 0.0869 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.0782 & 0.0761 & 0 & \cdots & 0 \\ 0.0782 & 0.16304 & 0 & \cdots & 0 \end{bmatrix}$$

2. Determination of parameter values used as a calculation process. Number of clusters $(c) = 3$, weight exponent (1.5), maximum number of iterations (10), smallest expected error tolerance (0.05), initial objective function (0), first iteration (1).
3. Generating an initial partition matrix U forms a size of 18×3 (18 is the number of regions, and 3 is the number of clusters).

$$U = \begin{bmatrix} 0.109 & 0.23 & 0.661 \\ 0.55 & 0.25 & 0.2 \\ \vdots & \vdots & \vdots \\ 0.24 & 0.12 & 0.64 \\ 0.01 & 0.02 & 0.97 \end{bmatrix}$$

This matrix is a measure of the degree of membership of the number of data totaling 18 data and the number of clusters totaling 3 clusters.

4. The calculation of the cluster center is based on **Equation (8)** namely by adding up the results of the operation of increasing the degree of cluster membership multiplied by the value of each data then dividing by the total results of increasing the value of the degree of cluster membership. The following is the calculation of the center of cluster one $(k = 1)$.

$$V_{1j} = \frac{\sum_{i=1}^{n}\left((\mu_{i1})^{w} \times X_{ij}\right)}{\sum_{i=1}^{n}(\mu_{i1})^{w}}$$

In this study there are 18 subdistricts so that for the values $i = 1,2,3,\dots,18$, calculations are carried out for all variable attributes. Then calculate the cluster center for every cluster. The calculation process is carried out on all attributes, then the cluster center is obtained as follows.

$$V_{3\times8} = \begin{bmatrix} 0.079 & 0.131 & 0.387 & 0.353 & 0.353 & 0.374 & 0.395 & 0.436 \\ 0.104 & 0.274 & 0.161 & 0.182 & 0.73 & 0.073 & 0.163 & 0.182 \\ 0.126 & 0.126 & 0.024 & 0.023 & 0.030 & 0.017 & 0.025 & 0.046 \end{bmatrix}$$

The cluster center results above show that $V_{3\times8}$ is 3 clusters and 8 variable attributes.

5. Calculation of partition matrix changes or improving the membership level value of each region in the cluster. Calculation of changes to the new partition matrix is carried out. The following is an example of calculations for each cluster, $\mu_{11} = \frac{\left[(0.0782-0.079)^2+(0.01087-0.131)^2+\cdots+(0-0.436)^2\right]^{-2}}{\left[(0.0782-0.079)^2+\cdots+(0-0.436)^2+\cdots+(0-0.1266)^2+\cdots+(0-0.046)^2\right]^{-2}} = 0.00052$. In a similar way, calculations are carried out for all μ_{ii} members in the partition matrix U. Each new membership degree is modified to obtain a better membership level that is better than the previous one by referring to the centroid value V that has been generated. Therefore, we obtain a new partition matrix that has been modified as follows:

$$U = \begin{bmatrix} 0.00052 & 0.01061 & 0.98887 \\ 0.72664 & 0.16831 & 0.10505 \\ \vdots & \vdots & \vdots \\ 0.00012 & 0.00342 & 0.99646 \\ 9.8 \times 10^{-5} & 0.0038 & 0.9961 \end{bmatrix}$$

6. Calculation of the objective function in the t-th iteration. based on **Equation (15)** The following is the objective function of iterations 1 to 5. Calculation are carried out for each cluster. The first step is to search cluster 1 objective function value (P_1) for iteration using **Equation (13)**. The calculation results are obtained $P_1 = 4.5105$ and $P_0 = 0$. The calculation continues to the next iteration until it is found $|P_t - P_{t-1}| < \xi$. The iteration calculation stops at the 5-th iteration and the function value is obtained objective $P_5 = 2.848$ and $P_4 = 2.8724$. Therefore, the iteration condition has been achieved so that Fuzzy C-Means calculation stop. The new partition matrix is used as a determinant of cluster results in Fuzzy C-Means following are the results of clustering in Fuzzy C-Means by looking at the maximum value of each cluster.

Table 9. Fuzzy C-Means Clusterization Results

Subdistrict	Membership Data			Cluster
	Cluster 1	Cluster 2	Cluster 3	
Anggana	0.0006	0.0262	0.9731	3
Kembang Janggut	0.9402	0.0334	0.0265	1
Kenohan	0.0040	0.6170	0.3790	2
Bangun City	0.0297	0.3798	0.5905	3
Loa Janan	0.0001	0.0035	0.9964	3
Loa Kulu	0.0008	0.0639	0.9353	3
Marang Kayu	0.0002	0.0119	0.9879	3
Muara Badak	0.1126	0.6379	0.2495	2
Muara Jawa	0.0003	0.0130	0.9867	3
Muara Kaman	0.8336	0.0995	0.0670	1
Muara Muntai	0.0005	0.0224	0.9771	3
Muara Wis	0.0056	0.1593	0.8351	3
Samboja	0.0444	0.5119	0.4437	2
Sanga-Sanga	0.0001	0.0085	0.9914	3
Sebulu	0.0016	0.1241	0.8743	3
Tabang	0.0445	0.5289	0.4266	2
Tenggarong	0.0002	0.0087	0.9911	3
Tenggarong Seberang	0.0001	0.0044	0.9955	3

In **Table 9** explains that the iteration stops at the 4-th iteraion because of the value objective function is smaller than the expected error value, it can be concluded in calculations using Fuzzy C-Means clustering that the subdistricts are included in cluster 1 namely Kembang Janggut and Muara Kaman, cluster 2 namely Kenohan, Muara Badak, Samboja and Tabang, while cluster 3 namely Anggana, Bangun City, Loa Janan, Loa Kulu, Marang Kayu, Muara Jawa, Muara Muntai, Muara Wis, Sanga-Sanga, Sebulu, Tenggarong and Tenggarong Seberang.

3.6 Partition Coefficient Index Calculation

Test the validaty of clusters in Fuzzy C-Means using Partition Coefficient Index to determine the strength of the data structure in clustering. The results of the partition coefficient calculation using **Equation 14** can be seen in **Table 10** as follows.

Table 10. Partition Coefficient

Regional Data	Membership Degree Data			$\sum U_{ij}^2$
	u_{i1}^2	u_{i2}^2	u_{i3}^2	
1	3.9×10^{-7}	6.89×10^{-4}	9.47×10^{-1}	0.9477
2	8.84×10^{-1}	1.11×10^{-3}	7.01×10^{-4}	0.8857
3	1.58×10^{-5}	3.81×10^{-1}	1.44×10^{-1}	0.5244

Regional Data	Membership Degree Data u_{i1}^2	u_{i2}^2	u_{i3}^2	$\sum U_{ij}^2$
….	….	….	….	….
16	1.98×10^{-3}	2.8×10^{-1}	1.82×10^{-1}	0.4637
17	2.82×10^{-8}	7.56×10^{-5}	9.82×10^{-1}	0.9824
18	4.51×10^{-9}	1.97×10^{-5}	9.91×10^{-1}	0.9910
			∑ PCI	0.78905

In **Table 10** explains that the partition coefficient index result Fuzzy C-Means algortithm is 0.78905. In the calculation of the Fuzzy C-Means algorithm, one of the factors influence is the value of the weighting power (w). Based on research from Wu in 2012 determines and selects fuzzy C-Means weighting ranks, then on this research carried out several experiments by considering values the weighting power of [1.5, 4]. The highest partition coefficient index results in the weighted rank experiment were 1.5. Therefore, Fuzzy C-Means calculations were carried out based on the results of the partition coefficient index.

Interpretation of the results of the Fuzzy C-Means algorithm is needed to analyze clustering results with real data values. The following are the results of grouping regions into cluster results.

Table 11. Fuzzy C-Means Region Cluster Results

Cluster	Regional Member Data
1 (High level of flood vulnerability)	Kembang Janggut and Muara Kaman
2 (Medium level of flood vulnerability)	Kenohan, Muara Badak, Samboja and Tabang
3 (Low level of flood vulnerability)	Anggana, Bangun City, Loa Janan, Loa Kulu, Marang Kayu, Muara Jawa, Muara Kaman, Muara Muntai, Muara Wis, Sanga-Sanga, Sebulu, Tenggarong and Tenggarong Seberang

In **Table 11**, is the result of the Fuzzy C-Means calculation, the region in cluster 1 is included in the high level of vulnerability category, namely Kembang Janggut and Muara Kaman. This categorization is seen from the number of flood events and the number of damaged houses and victims affected, so these two sub-districts have a higher frequency of flood events than other sub-district areas.

Areas that are included in cluster 2 are areas that have frequency the number of flood events is quite large but is still under the cluster 1 area. Clustering is seen from the height of the members of the cluster 2 area, so Tabang District has a height of 95 mdpl, this affects the areas around Tabang District which have a lower altitude.

Cluster 3 is an areas that have a lower frequency of flood events than members of cluster areas 1 and 2. This is influenced by the rainfall that occurs in cluster 3 areas so that members of cluster 3 areas are categorized as having a low level of flood vulnerability.

3.7 Standard Deviation Cluster

In this subchapter we will discuss clustering evaluation by calculating the standard deviation within and between clusters to determine the best clustering method from the results obtained. In the two methods used in clustering, it is necessary to determine the best method and performance, namely by using criteria and standard deviation values, namely standard deviation within clusters (S_w) and standard deviation between clusters (S_b). Determining the best method is based on the ratio of the smallest standard deviation within a cluster (S_w) and the standard deviation between clusters (S_k). The smaller the value (S_w) and the greater the value (S_b), then the method has the best performance **[18]**.

Table 12. Determining the Standard Deviation within clusters (S_w) and the Standard Deviation between clusters (S_b) K-Means Method.

Subdistrict	Total Variable Value	Mean in Cluster	Standar Deviation in Cluster
Cluster 1 (High level of flood vulnerability)			
Muara Badak	0.259	0.349	0.1274
Muara Kaman	0.439		

Subdistrict	Total Variable Value	Mean in Cluster	Standar Deviation in Cluster
Cluster 2 (Medium level of flood vulnerability)			
Kembang Janggut	0.627	0.627	0
Cluster 3 (Low level of flood vulnerability)			
Anggana	0.011	0.0617	0.102622
Kenohan	0.091		
Kota Bangun	0.126		
Loa Janan	0.038		
Loa Kulu	0.046		
Marang Kayu	0.056		
Muara Jawa	0.017		
Muara Muntai	0.012		
Muara Wis	0.066		
Samboja	0.155		
Sanga - Sanga	0.034		
Sebulu	0.070		
Tabang	0.153		
Tenggarong	0.019		
Tenggarong Seberang	0.030		

Based on the **Table 12**, the standard deviation within clusters (S_w) using **Equation (15)** can be obtained 0.102622 and the standard deviation between clusters (S_b) using **Equation (17)** is 0.1999 in K-Means Method.

Table 13. Determining the Standard Deviation within clusters (S_w)and the Standard Deviation between clusters (S_b) Fuzzy C-Means.

Subdistrict	Total Variable Value	Mean in Cluster	Standar Deviation in Cluster
Cluster 1 (High level of flood vulnerability)			
Kembang Janggut	5.109	4.265	1.06552
Muara Kaman	3.512		
Cluster 2 (Medium level of flood vulnerability)			
Kenohan	0.727	1.315	0.9617
Muara Badak	2.070		
Samboja	1.241		
Tabang	1.222		
Cluster 3 (Low level of flood vulnerability)			
Anggana	0.089	0.351	0.863
Kota Bangun	1.009		
Loa Janan	0.304		
Loa Kulu	0.372		
Marang Kayu	0.451		
Muara Jawa	0.133		
Muara Muntai	0.100		
Muara Wis	0.529		
Sanga - Sanga	0.274		
Sebulu	0.560		
Tenggarong	0.154		
Tenggarong Seberang	0.241		

In addition, from the **Table 13** we can get the standard deviation within clusters (S_w) using **Equation (15)** can be obtained 0.963655 and the standard deviation between clusters (S_b) using **Equation (17)** is 0.7619 in Fuzzy C-Means. Below is a comparison of the results of the standard deviation ratio S_w/S_b from the two methods used.

Table 14. Comparison of Methods with Standard Deviation

Standard Deviation Ratio	K-Means	Fuzzy C-Means
S_w/S_b	0.5131	0.3489

Table 14. explains that the Fuzzy C-Means ratio value is smaller than the K-Means ratio results. Therefore, in this clustering it is proven that the Fuzzy C-Means method is superior to K-Means Clustering.

3.8 The Best Interpretation of Clustering Results

Based on the calculation results, the best method is determined by using the standard deviation within the cluster (S_w) and the standard deviation S_b. The following is a comparison of cluster results using two clustering methods as follows.

Table 15. Comparison of Cluster Results

Cluster	Regional Member Data	
	K-Means Algorithm	Fuzzy C-Means Algorithm
1 (High level of flood vulnerability)	Muara Badak and Muara Kaman	Kembang Janggut and Muara Kaman
2 (Medium level of flood vulnerability)	Kembang Janggut	Kenohan, Muara Badak, Samboja and Tabang
3 (Low level of flood vulnerability)	Kenohan, Samboja, Tabang, Anggana, Bangun City, Loa Janan, Loa Kulu, Marang Kayu, Muara Jawa, Muara Kaman, Muara Muntai, Muara Wis, Sanga-Sanga, Sebulu, Tenggarong and Tenggarong Seberang	Anggana, Bangun City, Loa Janan, Loa Kulu, Marang Kayu, Muara Jawa, Muara Kaman, Muara Muntai, Muara Wis, Sanga-Sanga, Sebulu, Tenggarong and Tenggarong Seberang

In **Table 15**, It can be seen that the clustering results with K-Means and Fuzzy C-Means have a data distribution that is not much different. In the data distribution of the two methods, the distance between clusters is quite close. So that in the research the two methods have close clustering results.

4. CONCLUSIONS

Based on the results of tests carried out previously in the previous chapter, the following can be concluded:

1. The results of the clustering evaluation using the K-Means method produced cluster category 1 (high level of vulnerability), namely Muara Badak and Kembang Janggut Districts. Cluster 2 areas (medium level of vulnerability) are Kembang Janggut District and cluster 3 (low level of vulnerability) The areas included are Anggana, Kenohan, Kota Bangun, Loa Janan, Loa Kulu, Marang Kayu, Muara Jawa, Muara Kaman, Muara Muntai, Muara Wis, Samboja, Sanga-Sanga, Sebulu, Tabang, Tenggarong, and Tenggarong Seberang. The results of the clustering evaluation using the Fuzzy C-Means method produced areas in cluster category 1 (high level of vulnerability), namely Muara Kembang Janggut and Muara Kaman Districts. Cluster 2 areas (medium level of vulnerability) namely Kecamatan, Kenohan, Muara Badak, Samboja and Tabang and cluster category 3 (low level of vulnerability) namely Anggana, Kota Bangun, Loa Janan, Loa Kulu, Marang Kayu, Muara Jawa, Muara Kaman, Muara Muntai, Muara Wis, Sanga-Sanga and Sebulu.
2. The best cluster comparison results from this research were determined based on the validity test of the Silhouette Index on K-Means and the Partition Coefficients Index on Fuzzy C-Means and verification of the standard deviation values within and between clusters. The validity test produces a Silhouette Index value of 0.574283589 and a Partition Coefficient Index of 0.78905. The results of the K-Means method with standard deviation within and between clusters are 0.5131 and the Fuzzy C-Means method for standard deviation within and between clusters is 0.3489.

ACKNOWLEDGMENT

Gratitudes are expressed to those who support the implementation of community service activities. Mainly to LPPM ITK as the funder of this activity, who has provided support for the running of this program.

REFERENCES

[1] P. Kukar, "Bupati Kukar Datangi Titik Banjir Kembang Janggut, Ketersediaan Sembako Dipastikan Aman," 2019. [Online].

[2] H. Rahmah, "Kabupaten Kutai Kartanegara Dalam Angka 2023," Balikpapan, 2023.

[3] D. R. a. W. D. U. R. L. Rohmah, "LOMBA DAN SEMINAR MATEMATIKA XXVIII Zonasi Daerah Terdampak Bencana Angin Puting Beliung Menggunakan K-Means Clustering," 2020. [Online]. Available: http://dibi.bnpb.go.id.

[4] E. Oktaviana, CLUSTERING BENCANA ALAM DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS, Surabaya, 2022.

[5] T. S. Madhulatha, "An Overview on Clustering Methods," *IQSR*, vol. 2, no. 4, pp. 719 - 725, 2012.

[6] A. E. Z. Ramadhan, "Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling," *SNTKI*, pp. 219 - 226, 2017.

[7] B. M. N. P. V. Bangoria, "A Survey on Efficient Enhanced K-Means Clustering Algorithm," *International Journal for Scientific Research and Development*, vol. I, no. 9, 2013.

[8] C. L. K. N. I. B. Simbolon, "Clustering lulusan mahasiswa matematika FMIPA Untan Pontianak menggunakan Algoritma Fuzzy C-Means," *Buletin Ilmiah Mat, Stat, dan Terapannya (Bimaster)*, vol. 02, no. 1, pp. 21 - 26, 2013.

[9] A. L. N. B. S. M. A. Hana Sugiastuti F., "PERBANDINGAN METODE FUZZY C-MEANS DAN K-MEANS UNTUK PEMETAAN DAERAH RAWAN KRIMINALITAS DI KOTA SEMARANG," *Elipsoida: Jurnal Geodesi dan Geomatika*, vol. 4, no. 1, pp. 58 - 64, 2021.

[10] T. T. H. a. S. Al-Faraby, "Analisis Churn Prediction pada Data Pelanggan PT. Telekomunikasi dengan Logistic Regression dan Underbagging,"," 2017.

[11] P. P. H. K. K. K. a. D. M. Ahmad Harmain, "NORMALISASI DATA UNTUK EFISIENSI K-MEANS PADA PENGELOMPOKAN WILAYAH BERPOTENSI KEBAKARAN HUTAN DAN LAHAN BERDASARKAN SEBARAN TITIK PANAS," *TEKNIMEDIA: Teknologi Informasi dan Multimedia*, vol. 2, no. 2, pp. 83 - 89, 2022.

[12] D. A. P. D. J. d. I. T. Remawati, "Metode K-Means Untuk Pemetaan Persebaran Usaha Mikro Kecil Dan Menengah," *Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN)*, vol. 9, no. 2, pp. 39 - 46, 2021.

[13] M. H. a. P. R. A. Struyf, "Clustering in an Object-Oriented Environment," *Stat Softw*, vol. 1, no. 4, 1996.

[14] R. L. Rohmah, "Zonasi Daerah Terdampak Bencana Angin Puting Beliung Menggunakan K-Means Clustering dengan Analisis Silhouette 65 Coefficient, Davies Bouldin Index dan Purity," Universitas Islam Negeri Sunan Ampel, 2019.

[15] V. Y. I. I. a. M. H. H. D. L. Rahakbauw, "IMPLEMENTASI FUZZY C-MEANS CLUSTERING DALAM PENENTUAN BEASISWA," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 11, no. 1, pp. 1 - 12, 2017.

[16] H. I. K. H. Saadaki Miyamoto., Algorithms for Fuzzy Clustering, Osaka, 2008.

[17] K.-L. Wu, "Analysis of parameter selections for fuzzy C-Means," *Pattern Recognit*, vol. 45, no. 1, pp. 407 - 415, 2012.

[18] J. R. M. J. a. A. T. D. M. J. Bunkers, "Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique," *J Clim*, vol. 9, no. 1, pp. 130 - 146, 1996.