Jurnal Varian
Vol. 7, No. 1, October 2023, pp 99–114
e-ISSN: 2581-2017

Comparison K-Means and Fuzzy C-Means in Regencies/Cities Grouping Based on Educational Indicators

Gerald Claudio Messakh, Memi Nor Hayati, Sifriyani Sifriyani
Universitas Mulawarman, Samarinda, Indonesia

Article Info

Article history:

Received : 03-27-2023
Revised : 08-20-2023
Accepted : 10-26-2023

Keywords:

Education Indicators;
Fuzzy C-Means;
K-Means;

ABSTRACT

Cluster analysis is an analysis that aims to classify data based on the similarity of specific characteristics. The methods used in this research are K-Means and Fuzzy C-Means (FCM). K-Means is a partition-based non-hierarchical data grouping method. FCM is a clustering technique in which the existence of each data is determined by the degree of membership. The purpose of this study is to classify regencies/cities in Kalimantan based on education indicators in 2021 using K-Means and FCM and find out which method is better to use between K-Means and FCM based on the standard deviation ratio so it can be used efficiently and effectively for decision making by the government to advance the level of education on the island of Kalimantan. Based on the results of the analysis, it's concluded that K-Means is the better method with the ratio of the standard deviation within a cluster against the standard deviation between clusters of 0.6052 which produces optimal clusters of 2 clusters, namely the first cluster consisting of 14 Regencies/Cities, while the second cluster consists of 42 Regencies/Cities in Kalimantan.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v7i1.2879

Corresponding Author:

Gerald Claudio Messakh,
Departement of Mathematics, Universitas Mulawarman.
Email: geraldmessakh@gmail.com

How to Cite:
Messakh, G. C., Hayati, M. N., & Sifriyani, S. (2023). Comparison K-Means and Fuzzy C-Means in Regencies/Cities Grouping Based on Educational Indicators. *Jurnal Varian*. *7*(1). 99-114.

A. INTRODUCTION

Cluster analysis is a data mining method for identifying a set of objects that have certain common characteristics and can be separated from other clusters. Cluster analysis is divided into two, hierarchical and non-hierarchical (Suyanto, 2018). The hierarchical approach has a weakness, if one of the mergers or splits is carried out in the wrong place, the optimal cluster will not be obtained (Jollyta et al., 2020). The advantage of the non-hierarchical method is that it can perform analysis with a larger number of samples compared to the hierarchical method, several methods included in the non-hierarchical method are K-means and Fuzzy C-means (Triyanto, 2015).

K-Means is a data clustering techniques that divides objects into C clusters by allocating each object to the nearest centroid (Siregar, 2016). With K-Means, data objects will only be one cluster members and hard to achive convergence. Therefore, a comparison is made with other clustering methods that use fuzzy logic, because in its application a data object can be between two or more clusters (Jang and Sun, 1995). One of the fuzzy grouping algorithms is Fuzzy C-Means (FCM). FCM is an object or data clustering technique where the existence of each data in a cluster is determined by the membership function. An object is more likely to become a member of a group if it has the highest membership level (Kusuma et al., 2015).

K-Means and FCM can be applied in various sector such as health and social as previously done by (Nurmin et al., 2022) applied FCM method to classify regencies/cities in Kalimantan based on indicators of people's welfare. Another research was conducted by (Abdullah et al., 2022) applied K-Means clusterng for province clustering in Indonesia of the risk of the COVID-19 based on COVID-19 data (Mahmudi et al., 2021). Based on previous research, the difference between this research and previous research that compared K-Means and FCM is this study used the validity index of the standard deviation ratio to determine the optimal number of clusters and determine the best method at once where according to (Purnamasari et al., 2014) the selection of the method that produces the best quality clustering is done by taking into account the value of the ratio the average standard deviation within clusters and the standard deviation between clusters.

According to the (Direktorat Statistik Kesejahteraan Rakyat, 2021) one of indicator used to see educational attainment is the average length of schooling. In 2021, the average length of schooling for residents aged 15 and over in Kalimantan is 8.41 years. It is 0.56 lower than Indonesia's RLS or equivalent to 8.97, where the regencies/city with the highest RLS is Palangkaraya city while the lowest is Kayong Utara regencies. Of the five Kalimantan provinces, East Kalimantan has the largest RLS of 9.44, while West Kalimantan has the smallest RLS of 7.35. However, this number continues to increase over time.

Even though there have been many achievements and programs that have produced positive outputs in the field of education, there are still many gaps and challenges that need to be resolved in the development of education so that the targets for various educational indicators can be met by the end of 2025. The importance of the role of education in the progress of the nation, measurement, and calculation of indicators educational indicators need to be carried out to see the extent of educational equity. To know the distribution of education or educational level characteristics, grouping is done using cluster analysis. Based on the description above, this study will classify regencies/cities in Kalimantan based on educational indicators using K-Means and FCM. The purpose of this study is to get the best clustering results using K-Means and FCM based on standard deviation ratio and find out which method is better betwen for grouping regencies/cities in Kalimantan based on education indicators. The contribution of this research is to create a view or data/information that is more effective and efficient for the government and society so that they can consider the policies that will be taken to increase the level of education in Kalimantan.

B. RESEARCH METHOD

The research design used is non-experiment. The data used in this study is secondary data, namely education indicators obtained from the Badan Pusat Statistik website. The sample for this research is all regencies/cities on Kalimantan Island in 2021, totaling 56 regencies/cities. The determination of the educational indicator variables used uses references to the calculation of HDI educational aspects at BPS and follows previous research by (Ls et al., 2021). Table 1 shows the variables used in this study.

Table 1. Variable Used

X_k	**Description**	**Unit**
X_1	Expected length of schooling	Years
X_2	Average Length of School	Years
X_3	Number of Elementary Schools	Schools
X_4	Number of Middle Schools	Schools
X_5	Number of High Schools	Schools
X_6	Number of Elementary School Teachers	Peoples
X_7	Number of Middle School Teachers	Peoples
X_k	Number of High School Teachers	Peoples
X_9	Number of Elementary Students	Peoples
X_{10}	Number of Middle School Students	Peoples
X_{11}	Number of Senior High School Students	Peoples

DOI: https://doi.org/10.30812/varian.v7i1.2879

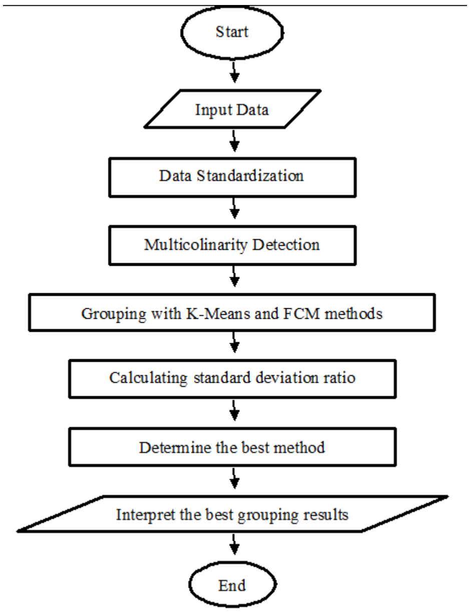

Figure 1. Flowchart of research analysis steps

The stages of data analysis in this study use the help of software R below :

1. Standardize data
   In cluster analysis, large difference in values between variables can cause the distance calculation to become unstable, so it is necessary to standardize the data by reducing the range of data (Hidayatullah et al., 2014). One of the algorithms that can be used to standardize data is the Min-Max algorithm. This algorithm is formulated below (Suyanto, 2018) :

$$x'_{i,k} = \frac{x_{i,k} - x_{kmin}}{x_{kmax} - x_{kmin}} \tag{1}$$

   description :
   $x'_{i,k}$: standardization of the i data for the k variable
   $x_{i,k}$: data-i of the k variable
   x_{kmin} : minimum data of the k variable
   x_{kmax} : maximum data of the k variable
2. Detect multicollinearity. Cluster analysis has two assumptions, representative sample and non-multicollinearity (Ghozali, 2016). Multicollinearity is a situation where there is a robust linear relationship between variables. One way to know if occur multicollinearity is to look at the Variance Inflation Factor (VIF) value.

$$VIF_k = \frac{1}{1 - R_k^2}, \quad k = 1, 2, \ldots, p \tag{2}$$

3. Grouping the observed objects using K-Means method : K-Means is a partition-based method that separates into C clusters different from assigning each data to the nearest centroid. The centroid is obtained by the average value of the variables of all objects in cluster. The results of clustering using K-Means depend on the initial centroid value that has been used. Giving different initial values can produce different groups. The steps for the K-Means method below (Kakushadze and Yu, 2017):
   (a) Determine the number of clusters (C).
   (b) Determine the centroid v_{ck} randomly from the object of observation.
   (c) Calculate the euclidean distance for each observation object to the centroid.

$$d(x'_i, v_c) = \sqrt{\sum_{k=1}^{p} \left(x'_{i,k} - v_{c,k}\right)^2 m} \tag{3}$$

   description :
   $d(x_i, v_c)$: euclidean distance between of the i observation data and center of the c cluster
   $v_{c,k}$: centroid in the c cluster on the k variable

DOI: https://doi.org/10.30812/varian.v7i1.2879

(d) Assigns each object to the cluster with the most similar object, based on the closest distance between objects to each centroid.

(e) Updating the centroid is with calculating the average value of each object for each cluster .

$$v_{c,k}^{t} = \sum_{i=1}^{n_c} \frac{x'_{i,k,c}}{n_c} \tag{4}$$

description :

$v_{c,k}^{t}$: centroid in the c cluster on the k variable in the t iteration

$x'_{i,k,c}$: standardization of the i data for the k variable into the c cluster

n_c : number of data in c cluster

(f) Repeat steps 3, 4, and 5 until no more cluster members change cluster.

4. Grouping the observed objects using Fuzzy C-Means method : Fuzzy C-Means (FCM) is a clustering method where the existence of each data in a cluster is determined by the degree of membership. FCM begins by determining the centroid which will mark the average location for each cluster. By repairing the membership degree of each data centroid repeatedly, the centroid will go to the right place.. As a result of the degree of membership, data points can belong to more than one cluster. The steps in the FCM method below (Kusuma et al., 2015):

(a) Determine number of clusters (c), rank (m), maksimum iteration (MaxIter), smallest expected error (ε), initial objective function $(P_0 = 0)$, is a fuction to be optimized.

(b) Generate random numbers as the initial elements of the initial membership matrix U.

(c) Calculating the center of the c cluster with the following equation :

$$v_{ck} = \frac{\sum_{i=1}^{n} \left((u_{i,c})^m x'_{i,k} \right)}{\sum_{i=1}^{n} (\mu_{i,c})^m} \tag{5}$$

(d) Calculating the objective function in the t iteration with the following equation :

$$P_t = \sum_{i=1}^{n} \sum_{c=1}^{C} \left(\left[\sum_{k=1}^{P} (x'_{i,k} - v_{c,k})^2 \right] (\mu_{i,c})^m \right) \tag{6}$$

(e) Calculating of membership matrix changes with the following equation :

$$\mu_{ik} = \frac{\left[\sum_{k=1}^{P} \left(x'_{i,k} - v_{c,k} \right)^2 \right]^{\frac{1}{m-1}}}{\sum_{c=1}^{C} \left[\sum_{k=1}^{P} \left(x'_{i,k} - v_{c,k} \right)^2 \right]^{-\frac{1}{m-1}}} \tag{7}$$

(f) Stop condition calculation if $|P_t - P_{t-1}| < \varepsilon$ atau $t > MaxIter$ then stop. If not $t := t + 1$, repeat step 3 to step 5

5. Determine the best method based on the value of the standard deviation ratio.

According to a grouping method that can be used to form clusters be told to have good performance if it has a minimum standard deviation within the cluster to the standard deviation between clusters. Standard deviation within cluster (S_w) and standard deviation between clusters (S_b) can be calculated using following equation (Barakbah and Arai, 2004):

$$S_w = \frac{1}{C} \sum_{c=1}^{C} S_c \tag{8}$$

$$S_{c,k} = \sqrt{\frac{1}{n_c - 1} \sum_{i=1}^{n_c} \left(x'_{i,k} - \bar{x}_{c,k} \right)^2} \tag{9}$$

$$S_c = \frac{1}{p} \sum_{k=1}^{p} S_{c,k} \tag{10}$$

$$\bar{x}_c = \frac{1}{p} \sum_{k=1}^{p} \bar{x}_{c,k} \tag{11}$$

DOI: https://doi.org/10.30812/varian.v7i1.2879

$$S_b = \sqrt{\frac{1}{C-1}\sum_{c=1}^{C}\left(\bar{x}_c - \bar{x}\right)^2} \tag{12}$$

$$\bar{x} = \frac{1}{C}\sum_{c=1}^{C}\bar{x}_c \tag{13}$$

description :

S_w : standard deviation within cluster

$S_{c,k}$: standard deviation of the c cluster on k variabel

$\bar{x}_{c,k}$: average of the c cluster on k variable

S_c : standard deviation of the c cluster

S_b : standard deviation between cluster

$\bar{x}_c$: average of the c cluster

$\bar{x}$: average of the entire cluster

C : number of clusters

6. Interpret the best grouping results based on the value of the smallest standard deviation ratio.

C. RESULTS AND DISCUSSION

1. Data Standardization

The result of calculating data standardization for each variable can be seen in Table 2 below :

Table 2. Standardization Data

Data	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
1	0,372	0,127	0,800	0,633	0,475	0,608	0,680	0,533	0,809	0,770	0,583
2	0,237	0,142	0,477	0,338	0,277	0,305	0,294	0,276	0,352	0,338	0,274
3	0,321	0,200	0,826	0,457	0,411	0,423	0,448	0,398	0,473	0,478	0,455
4	0,378	0,185	0,394	0,355	0,326	0,323	0,379	0,328	0,336	0,327	0,302
5	0,107	0,249	0,855	0,521	0,340	0,484	0,440	0,322	0,572	0,495	0,424
6	0,163	0,261	1,000	0,675	0,482	0,662	0,754	0,535	0,668	0,606	0,516
7	0,222	0,192	0,813	0,568	0,418	0,502	0,542	0,462	0,589	0,542	0,516
8	0,232	0,274	0,747	0,457	0,270	0,379	0,428	0,268	0,319	0,299	0,254
9	0,184	0,151	0,409	0,265	0,177	0,187	0,200	0,205	0,248	0,222	0,229
10	0,000	0,162	0,434	0,457	0,262	0,273	0,301	0,204	0,235	0,160	0,195
11	0,171	0,000	0,159	0,154	0,121	0,127	0,144	0,107	0,117	0,123	0,117
12	0,689	0,178	0,972	1,000	1,000	0,766	0,839	0,728	0,764	0,770	0,646
13	0,980	0,800	0,323	0,436	0,716	0,515	0,686	0,985	0,762	0,807	1,000
14	0,444	0,341	0,151	0,154	0,227	0,182	0,237	0,313	0,276	0,296	0,331
15	0,291	0,339	0,447	0,295	0,241	0,422	0,415	0,378	0,394	0,303	0,272
16	0,199	0,261	0,445	0,291	0,284	0,410	0,367	0,358	0,387	0,334	0,287
17	0,398	0,272	0,830	0,517	0,426	0,746	0,767	0,606	0,602	0,522	0,443
18	0,314	0,278	0,564	0,389	0,305	0,440	0,438	0,417	0,347	0,320	0,265
19	0,199	0,318	0,309	0,124	0,099	0,256	0,191	0,175	0,195	0,144	0,133
20	0,309	0,316	0,479	0,218	0,163	0,399	0,339	0,282	0,225	0,216	0,154
21	0,265	0,361	0,508	0,214	0,206	0,430	0,388	0,350	0,255	0,244	0,223
22	0,446	0,314	0,447	0,235	0,213	0,450	0,364	0,340	0,225	0,217	0,218
23	0,439	0,561	0,425	0,321	0,206	0,369	0,426	0,360	0,296	0,253	0,263
24	0,334	0,354	0,338	0,325	0,270	0,388	0,434	0,395	0,420	0,348	0,358
25	0,332	0,303	0,343	0,137	0,128	0,265	0,179	0,179	0,132	0,095	0,102
26	0,707	0,759	0,545	0,363	0,411	0,680	0,731	0,887	0,718	0,698	0,798
27	0,931	0,897	0,134	0,115	0,241	0,256	0,295	0,410	0,290	0,334	0,338
28	0,398	0,456	0,345	0,286	0,227	0,284	0,317	0,317	0,348	0,313	0,275
29	0,429	0,387	0,696	0,509	0,355	0,604	0,547	0,463	0,586	0,550	0,430
30	0,449	0,287	0,891	0,671	0,411	0,734	0,628	0,474	0,450	0,411	0,323
31	0,380	0,532	0,289	0,291	0,248	0,252	0,221	0,241	0,119	0,140	0,131
32	0,339	0,514	0,300	0,150	0,156	0,281	0,214	0,238	0,146	0,148	0,146
33	0,245	0,376	0,042	0,034	0,050	0,053	0,055	0,075	0,037	0,015	0,032
34	0,337	0,437	0,164	0,150	0,128	0,106	0,095	0,111	0,089	0,075	0,073
35	0,212	0,352	0,272	0,252	0,142	0,235	0,219	0,147	0,202	0,162	0,097
36	0,418	0,483	0,353	0,338	0,234	0,225	0,247	0,213	0,189	0,191	0,155
37	0,319	0,394	0,332	0,197	0,213	0,251	0,199	0,227	0,143	0,134	0,116
38	0,186	0,574	0,285	0,209	0,106	0,216	0,173	0,132	0,151	0,131	0,109
39	0,426	0,583	0,238	0,120	0,114	0,225	0,193	0,196	0,105	0,102	0,096
40	0,151	0,289	0,285	0,252	0,163	0,202	0,178	0,160	0,139	0,127	0,122

DOI: https://doi.org/10.30812/varian.v7i1.2879

Data	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
41	0,967	1,000	0,225	0,222	0,326	0,319	0,350	0,579	0,315	0,337	0,411
42	0,531	0,503	0,385	0,350	0,305	0,377	0,425	0,431	0,355	0,322	0,177
43	0,472	0,486	0,340	0,244	0,312	0,352	0,330	0,305	0,210	0,191	0,101
44	0,620	0,583	0,906	0,812	0,872	1,000	1,000	1,000	1,000	0,950	0,520
45	0,441	0,619	0,387	0,385	0,333	0,484	0,460	0,400	0,555	0,424	0,172
46	0,551	0,637	0,270	0,235	0,262	0,341	0,339	0,391	0,342	0,302	0,204
47	0,357	0,425	0,151	0,141	0,156	0,168	0,190	0,196	0,206	0,200	0,087
48	0,367	0,392	0,017	0,017	0,071	0,017	0,050	0,054	0,007	0,015	0,008
49	0,778	0,888	0,342	0,316	0,390	0,529	0,590	0,589	0,763	0,794	0,374
50	1,000	0,811	0,417	0,517	0,674	0,693	0,936	0,946	0,963	1,000	0,589
51	0,510	0,868	0,059	0,094	0,142	0,125	0,154	0,254	0,188	0,200	0,168
52	0,546	0,613	0,149	0,124	0,135	0,152	0,174	0,146	0,078	0,089	0,078
53	0,469	0,575	0,221	0,248	0,192	0,195	0,219	0,233	0,177	0,164	0,162
54	0,268	0,505	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
55	0,378	0,390	0,230	0,192	0,199	0,254	0,223	0,264	0,226	0,221	0,194
56	0,730	0,719	0,083	0,090	0,163	0,182	0,217	0,251	0,267	0,261	0,250

2. Multicollinearity detection

Table 3. VIF Value

Variable	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
VIF	8,667	5,197	19,804	29,219	28,284	42,184	61,376	64,526	124,905	125,110	13,594

Based on Table 3, there are 9 variables with VIF bigger than 10, which means there is multicollinearity between the variables and cannot be continued into the grouping process. To overcome multicollinearity, this can be done by removing the variable with the biggest VIF value and then regressing the remaining variables. After removing several variables and performing regression calculations, the final VIF can be seen in Table 4 :

Table 4. VIF Value

Variable	x_1	x_2	x_3	x_5	x_9	x_{11}
VIF	6,222	4,170	4,450	7,383	7,741	4,731

Based on Table 4, it can be seen that all the remaining variables have a VIF value of less than 10, so it can be said that there is no multicollinearity between the variables and can proceed to the grouping process using variables that are not excluded.

3. K-Means Clustering

1. Determining Number of Cluster
   In this study, the number of clusters to be used is $C = 2, 3, 4, 5$ and 6. As an example of the calculations in this study, it is done using $C = 2$.
2. Choose Initial Centroid
   The initial centroids that were selected were the 18th and 36th can be seen in Table 5:

Table 5. Initial Centroid

Variable	$v_{1,k}$	$v_{2,k}$	Variable	$v_{1,k}$	$v_{2,k}$
X_1	0,314	0,418	X_5	0,305	0,234
X_2	0,278	0,483	X_9	0,347	0,189
X_3	0,564	0,353	X_{11}	0,265	0,155

3. Calculating the Distance of All Observational Data with the Initial Centroid
   The entire calculation results can be seen in Table 6 below:

Table 6. Eulidean Distance to Each Initial Centroid

Data	Euclidean Distance of Data to Initial Centroid		Data	Euclidean Distance of Data to Initial Centroid	
	Centroid 1	Centroid 2		Centroid 1	Centroid 2
1	0,652	0,976	29	0,360	0,612
2	0,181	0,455	30	0,388	0,676
3	0,372	0,718	31	0,467	0,117
4	0,208	0,379	32	0,450	0,135
5	0,453	0,796	33	0,709	0,459
6	0,641	0,977	34	0,567	0,269
7	0,461	0,812	35	0,421	0,280
8	0,206	0,512	36	0,374	0,000
9	0,290	0,425	37	0,373	0,149
10	0,385	0,538	38	0,533	0,294
11	0,609	0,593	39	0,579	0,220
12	1,058	1,307	40	0,434	0,350
13	1,287	1,301	41	1,042	0,822
14	0,455	0,316	42	0,371	0,217
15	0,156	0,319	43	0,406	0,111
16	0,174	0,406	44	1,057	1,247
17	0,435	0,750	45	0,466	0,405
18	0,000	0,374	46	0,527	0,273
19	0,403	0,310	47	0,517	0,243
20	0,236	0,249	48	0,743	0,453
21	0,180	0,268	49	0,911	0,834
22	0,241	0,209	50	1,182	1,196
23	0,357	0,189	51	0,839	0,502
24	0,270	0,346	52	0,689	0,321
25	0,392	0,240	53	0,532	0,175
26	0,905	0,960	54	0,809	0,512
27	0,981	0,725	55	0,399	0,172
28	0,305	0,202	56	0,791	0,495

4. Placing Observation Data to the Nearest Centroid
   The results of data allocation can be seen in Table 7 below :

Table 7. Results of Placement of Each Data to the Nearest Centroid

Data	Euclidean Distance of Data to Initial Centroid		Cluster Allocation	Data	Euclidean Distance of Data to Initial Centroid		Cluster Allocation
	Centroid 1	Centroid 2			Centroid 1	Centroid 2	
1	0,652	0,976	1	29	0,360	0,612	1
2	0,181	0,455	1	30	0,388	0,676	1
3	0,372	0,718	1	31	0,467	0,117	2
4	0,208	0,379	1	32	0,450	0,135	2
5	0,453	0,796	1	33	0,709	0,459	2
6	0,641	0,977	1	34	0,567	0,269	2
7	0,461	0,812	1	35	0,421	0,280	2
8	0,206	0,512	1	36	0,374	0,000	2
9	0,290	0,425	1	37	0,373	0,149	2
10	0,385	0,538	1	38	0,533	0,294	2
11	0,609	0,593	2	39	0,579	0,220	2
12	1,058	1,307	1	40	0,434	0,350	2
13	1,287	1,301	1	41	1,042	0,822	2
14	0,455	0,316	2	42	0,371	0,217	2
15	0,156	0,319	1	43	0,406	0,111	2
16	0,174	0,406	1	44	1,057	1,247	1
17	0,435	0,750	1	45	0,466	0,405	2
18	0,000	0,374	1	46	0,527	0,273	2
19	0,403	0,310	2	47	0,517	0,243	2
20	0,236	0,249	1	48	0,743	0,453	1
21	0,180	0,268	1	49	0,911	0,834	2
22	0,241	0,209	2	50	1,182	1,196	1
23	0,357	0,189	2	51	0,839	0,502	2
24	0,270	0,346	2	52	0,689	0,321	2
25	0,392	0,240	2	53	0,532	0,175	2
26	0,905	0,960	2	54	0,809	0,512	2
27	0,981	0,725	2	55	0,399	0,172	2
28	0,305	0,202	2	56	0,791	0,495	2

Based on Table 7, the euclidean for the 1st observation data to the center of cluster 1 is smaller than the euclidean of the 1st observation data to the center of cluster 2 so that the 1st observation data included in the membership of a cluster 1 and so on up to the 56th observation data. Based on the placement results, cluster 1 consisted of 24 regencies/cities while cluster 2 consisted of 32 regencies/cities.

5. Updating the Centroid
The results of the calculation the new centroid can be seen in Table 8 :

Table 8. New Centroid

Variable	$v^1_{1,k}$	$v^1_{2,k}$	Variable	$v^1_{1,k}$	$v^1_{2,k}$
X_1	0,314	0,418	X_5	0,305	0,234
X_2	0,278	0,483	X_9	0,347	0,189
X_3	0,564	0,353	X_{11}	0,265	0,155

Based on the calculation results in Table 8, it can be seen that there is a difference between the new centroid and the previous centroid, so the grouping is continued to the next iteration.

6. Repeat steps c, d and e until there is no change in the centroid from the previous centroid
Based on the calculation results, the clustering is stop at the 5th iteration, where there is no change in the cluster membership. So that the new centroid will be the same as the old centroid. The results of grouping the K-Means method with $C = 2$ can be seen in Table 9 :

Table 9. K-Means Clustering Results

Cluster	Number of Members	Clusters Member
1	14	Sambas, Landak, Sangau, Ketapang, Sintang, Kubu Raya, Pontianak, Banjar, Banjarmasin, Kotawaringin Timur, Kapuas, Kutai Kartanegara, Balikpapan, Samarinda.
2	42	Bengkayang, Mempawah, Kapuas Hulu, Sekadau, Melawi, Kayong Utara, Singkawang, Tanah Laut, Kotabaru, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Banjar Baru, Kotawaringin Barat, Barito Selatan, Barito Utara, Sukamara, Lamandau, Seruyan, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Palangkaraya, Paser, Kutai Barat, Kutai Timur, Berau, Penajam Paser Utara, Mahakam Ulu, Bontang, Malinau, Bulungan, Tana Tidung, Nunukan, Tarakan.

Table 10. K-Means Clustering Results Using $C = 3$

Cluster	Number of Members	Clusters Member
1	6	Pontianak, Banjarmasin, Banjar Baru, Palangka Raya, Balikpapan, Samarinda.
2	11	Sambas, Landak, Sanggau, Ketapang, Sintang, Kapuas Hulu, Kubu Raya, Banjar, Kotawaringin Timur, Kapuas, Kutai Kartanegara.
3	39	Bengkayang, Mempawah, Sekadau, Melawi, Kayong Utara, Singkawang, Tanah Laut, Kotabaru, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Kotawaringin Barat, Barito Selatan, Barito Utara, Sukamara, Lamandau, Seruyan, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Paser, Kutai Barat, Kutai Timur, Berau, Penajam Paser Utara, Mahakam Ulu, Bontang, Malinau, Bulungan, Tana Tidung, Nunukan, Tarakan.

Table 11. K-Means Clustering Results Using $C = 4$

Cluster	Number of Members	Clusters Member
1	10	Sambas, Landak, Sanggau, Ketapang, Sintang, Kubu Raya, Banjar, Kotawaringin Timur, Kapuas, Kutai Kartanegara.
2	23	Singkawang, Tabalong, Kotawaringin Barat, Barito Selatan, Barito Utara, Sukamara, Lamandau, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Paser, Kutai Barat, Kutai Timur, Berau, Penajam Paser Utara, Mahakam Ulu, Bontang, Malinau, Bulungan, Tana Tidung, Nunukan, Tarakan.
3	17	Bengkayang, Mempawah, Kapuas Hulu, Sekadau, Melawi, Kayong Utara, Tanah Laut, Kotabaru, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tanah Bumbu, Balangan, Seruyan, Murung Raya.
4	6	Pontianak, Banjarmasin, Banjar Baru, Palangka Raya, Balikpapan, Samarinda

Table 12. K-Means Clustering Results Using $C = 5$

Cluster	Number of Members	Clusters Member
1	5	Pontianak, Banjarmasin, Kutai Kartanegara, Balikpapan, Samarinda.
2	16	Bengkayang, Mempawah, Kapuas Hulu, Sekadau, Melawi, Tanah Laut, Kotabaru, Barito Kuala, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Kotawaringin Barat, Paser, Kutai Timur.
3	9	Sambas, Landak, Sanggau, Ketapang, Sintang, Kubu Raya, Banjar, Kotawaringin Timur, Kapuas.
4	21	Kayong Utara, Singkawang, Tapin, Balangan, Barito Selatan, Barito Utara, Sukamara, Lamandau, Seruyan, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Kutai Barat, Penajam Paser Utara, Mahakam Ulu, Malinau, Bulungan, Tana Tidung, Nunukan.
5	5	Banjar Baru, Palangka Raya, Berau, Bontang, Tarakan.

Table 13. K-Means Clustering Results Using $C = 6$

Cluster	Number of Members	Clusters Member
1	4	Banjar Baru, Palangka Raya, Bontang, Tarakan.
2	8	Sukamara, Lamandau, Gunung Mas, Barito Timur, Penajam Paser Utara, Mahakam Ulu, Malinau, Tana Tidung.
3	13	Singkawang, Tabalong, Kotawaringin Barat, Barito Selatan, Barito Utara, Katingan, Pulang Pisau, Paser, Kutai Barat, Kutai Timur, Berau, Bulungan, Nunukan.
4	5	Pontianak, Banjarmasin, Kutai Kartanegara, Balikpapan, Samarinda.
5	17	Bengkayang, Mempawah, Kapuas Hulu, Sekadau, Melawi, Kayong Utara, Tanah Laut, Kotabaru, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tanah Bumbu, Balangan, Seruyan, Murung Raya.
6	9	Sambas, Landak, Sanggau, Ketapang, Sintang, Kubu Raya, Banjar, Kotawaringin Timur, Kapuas.

4. Fuzzy C-Means Clustering

1. Specifies parameter values

   In this study, number of clusters $(C) = 1, 2, 3, 4, 5$ and 6, $m = 2$, $MaxIter = 1.000$, $\varepsilon = 10^{-5}$ and $P_0 = 0$. As an example of calculations, it is done using $C = 6$.
2. Generate Random Numbers

   Generate random numbers μ_{ic} as an element of the initial membership matrix U . The initial membership value is presented in Table 14 :

Table 14. Initial Membership Value

No	Regencies/Cities	Initial Membership Value					
		(μ_{i1})	(μ_{i2})	(μ_{i3})	(μ_{i4})	(μ_{i5})	(μ_{i6})
1	Sambas	0,239	0,106	0,191	0,300	0,033	0,130
2	Bengkayang	0,194	0,303	0,182	0,099	0,198	0,025
3	Landak	0,151	0,249	0,076	0,233	0,073	0,219
4	Mempawah	0,225	0,093	0,145	0,191	0,099	0,247
5	Sanggau	0,240	0,162	0,223	0,015	0,174	0,186
6	Ketapang	0,252	0,233	0,228	0,187	0,043	0,057
7	Sintang	0,073	0,335	0,087	0,097	0,248	0,160
8	Kapuas Hulu	0,078	0,162	0,256	0,150	0,274	0,081
9	Sekadau	0,106	0,233	0,032	0,283	0,212	0,133
10	Melawi	0,207	0,179	0,111	0,102	0,182	0,219
11	Kayong Utara	0,077	0,318	0,032	0,227	0,059	0,286
12	Kubu Raya	0,201	0,222	0,010	0,034	0,336	0,198
13	Kota Pontianak	0,191	0,151	0,218	0,225	0,017	0,199
14	Kota Singkawang	0,028	0,099	0,236	0,222	0,210	0,203
15	Tanah Laut	0,127	0,174	0,098	0,307	0,104	0,190
16	Kotabaru	0,299	0,114	0,155	0,125	0,072	0,235
17	Banjar	0,185	0,211	0,129	0,191	0,148	0,136
18	Barito Kuala	0,129	0,182	0,129	0,131	0,182	0,248
19	Tapin	0,294	0,137	0,074	0,265	0,125	0,105
20	Hulu Sungai Selatan	0,160	0,060	0,319	0,113	0,100	0,249
21	Hulu Sungai Tengah	0,057	0,087	0,264	0,183	0,251	0,158
22	Hulu Sungai Utara	0,245	0,191	0,131	0,172	0,193	0,068
23	Tabalong	0,101	0,130	0,220	0,056	0,223	0,270
24	Tanah Bumbu	0,108	0,015	0,221	0,209	0,248	0,200
25	Balangan	0,297	0,174	0,010	0,239	0,055	0,225
26	Kota Banjarmasin	0,139	0,145	0,217	0,229	0,151	0,121
27	Kota Banjar Baru	0,058	0,342	0,211	0,142	0,156	0,091
28	Kotawaringin Barat	0,079	0,028	0,135	0,206	0,183	0,369

No	Regencies/Cities	Initial Membership Value					
		(μ_{i1})	(μ_{i2})	(μ_{i3})	(μ_{i4})	(μ_{i5})	(μ_{i6})
29	Kotawaringin Timur	0,155	0,224	0,178	0,155	0,126	0,162
30	Kapuas	0,181	0,057	0,205	0,257	0,145	0,155
31	Barito Selatan	0,156	0,068	0,230	0,044	0,289	0,212
32	Barito Utara	0,188	0,150	0,169	0,124	0,073	0,296
33	Sukamara	0,217	0,154	0,015	0,197	0,244	0,174
34	Lamandau	0,151	0,253	0,262	0,024	0,172	0,139
35	Seruyan	0,346	0,135	0,097	0,038	0,049	0,335
36	Katingan	0,100	0,256	0,162	0,063	0,234	0,185
37	Pulang Pisau	0,077	0,039	0,164	0,029	0,416	0,275
38	Gunung Mas	0,012	0,204	0,254	0,133	0,272	0,124
39	Barito Timur	0,164	0,046	0,268	0,100	0,168	0,255
40	Murung Raya	0,046	0,121	0,316	0,316	0,081	0,121
41	Palangka Raya	0,124	0,286	0,149	0,006	0,360	0,075
42	Paser	0,336	0,011	0,148	0,029	0,329	0,148
43	Kutai Barat	0,279	0,024	0,109	0,150	0,229	0,211
44	Kutai Kartanegara	0,066	0,099	0,091	0,329	0,095	0,321
45	Kutai Timur	0,086	0,058	0,265	0,268	0,144	0,179
46	Berau	0,133	0,113	0,102	0,198	0,176	0,278
47	Penajam Paser Utara	0,208	0,115	0,195	0,191	0,169	0,122
48	Mahakam Ulu	0,109	0,125	0,214	0,250	0,070	0,232
49	Balikpapan	0,077	0,198	0,161	0,090	0,232	0,243
50	Samarinda	0,049	0,117	0,242	0,269	0,193	0,129
51	Bontang	0,045	0,045	0,225	0,270	0,388	0,028
52	Malinau	0,362	0,013	0,175	0,379	0,013	0,060
53	Bulungan	0,043	0,009	0,462	0,410	0,034	0,043
54	Tana Tidung	0,147	0,020	0,217	0,219	0,208	0,190
55	Nunukan	0,252	0,189	0,027	0,186	0,225	0,121
56	Tarakan	0,199	0,146	0,292	0,064	0,187	0,111

3. Calculating Centroids
   The initial centroid calculation uses equation (5) for $k = 1, 2, 3, 5, 9, 11$. The entire updated centroid can be seen in table 15 :

Table 15. New Centroid

Variable	Centroid					
	1	2	3	4	5	6
X_1	0,379	0,416	0,426	0,419	0,443	0,390
X_2	0,390	0,403	0,486	0,439	0,486	0,415
X_3	0,413	0,449	0,368	0,395	0,383	0,405
X_5	0,268	0,300	0,254	0,287	0,286	0,287
X_9	0,319	0,367	0,320	0,363	0,320	0,344
X_{11}	0,247	0,314	0,258	0,277	0,257	0,257

4. Calculating the Objective Function
   Based on the calculation results, the objective function values on $t = 1$ is $3,475$. Initial objective function value P_0 is 0 so $|P_1 - P_0| = 3,475 > \varepsilon = 10^{-5}$ because the change in the objective function is still greater than the value ε, then the process continues to the next iteration.

Table 16. Updated Membership Value

No	Regencies/Cities	μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}	μ_{i5}	μ_{i6}
1	Sambas	0,165	0,203	0,143	0,172	0,147	0,170
2	Bengkayang	0,227	0,192	0,118	0,155	0,116	0,192
3	Landak	0,173	0,219	0,132	0,165	0,139	0,172
4	Mempawah	0,224	0,204	0,109	0,159	0,111	0,193
5	Sanggau	0,176	0,202	0,139	0,166	0,141	0,175
6	Ketapang	0,169	0,198	0,144	0,168	0,149	0,171
7	Sintang	0,171	0,211	0,136	0,168	0,141	0,173
8	Kapuas Hulu	0,200	0,202	0,130	0,156	0,132	0,181
9	Sekadau	0,233	0,162	0,137	0,152	0,128	0,188
10	Melawi	0,211	0,164	0,144	0,157	0,138	0,186
11	Kayong Utara	0,199	0,153	0,159	0,159	0,151	0,178
12	Kubu Raya	0,162	0,189	0,152	0,170	0,160	0,167
13	Kota Pontianak	0,152	0,176	0,166	0,173	0,172	0,161
14	Kota Singkawang	0,169	0,134	0,189	0,173	0,169	0,167
15	Tanah Laut	0,262	0,189	0,093	0,151	0,086	0,219

No	Regencies/Cities	μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}	μ_{i5}	μ_{i6}
16	Kotabaru	0,230	0,189	0,114	0,156	0,109	0,202
17	Banjar	0,164	0,218	0,135	0,170	0,143	0,170
18	Barito Kuala	0,230	0,230	0,099	0,146	0,103	0,192
19	Tapin	0,222	0,132	0,168	0,150	0,146	0,182
20	Hulu Sungai Selatan	0,269	0,138	0,140	0,138	0,128	0,187
21	Hulu Sungai Tengah	0,278	0,150	0,126	0,136	0,116	0,194
22	Hulu Sungai Utara	0,271	0,143	0,135	0,139	0,132	0,180
23	Tabalong	0,099	0,087	0,312	0,134	0,257	0,112
24	Tanah Bumbu	0,175	0,202	0,117	0,199	0,107	0,201
25	Balangan	0,225	0,128	0,171	0,146	0,153	0,177
26	Kota Banjarmasin	0,146	0,184	0,165	0,175	0,171	0,159
27	Kota Banjar Baru	0,143	0,150	0,191	0,169	0,194	0,152
28	Kotawaringin Barat	0,109	0,064	0,336	0,192	0,151	0,148
29	Kotawaringin Timur	0,153	0,237	0,131	0,174	0,140	0,166
30	Kapuas	0,170	0,214	0,136	0,166	0,146	0,170
31	Barito Selatan	0,167	0,105	0,223	0,143	0,205	0,156
32	Barito Utara	0,179	0,106	0,225	0,143	0,188	0,159
33	Sukamara	0,183	0,135	0,186	0,158	0,170	0,169
34	Lamandau	0,183	0,122	0,199	0,153	0,177	0,166
35	Seruyan	0,216	0,125	0,175	0,151	0,152	0,182
36	Katingan	0,171	0,089	0,242	0,132	0,215	0,151
37	Pulang Pisau	0,222	0,113	0,183	0,142	0,163	0,177
38	Gunung Mas	0,181	0,123	0,201	0,152	0,175	0,168
39	Barito Timur	0,162	0,115	0,222	0,150	0,198	0,154
40	Murung Raya	0,217	0,136	0,164	0,152	0,149	0,182
41	Palangka Raya	0,144	0,156	0,186	0,169	0,191	0,154
42	Paser	0,104	0,100	0,209	0,170	0,286	0,132
43	Kutai Barat	0,157	0,100	0,214	0,145	0,230	0,155
44	Kutai Kartanegara	0,157	0,185	0,156	0,173	0,164	0,166
45	Kutai Timur	0,129	0,147	0,181	0,191	0,192	0,161
46	Berau	0,111	0,106	0,245	0,161	0,247	0,130
47	Penajam Paser Utara	0,181	0,115	0,205	0,155	0,177	0,168
48	Mahakam Ulu	0,178	0,134	0,189	0,158	0,175	0,166
49	Balikpapan	0,141	0,166	0,176	0,177	0,183	0,157
50	Samarinda	0,150	0,175	0,166	0,175	0,173	0,161
51	Bontang	0,146	0,131	0,209	0,162	0,199	0,153
52	Malinau	0,155	0,123	0,213	0,156	0,200	0,153
53	Bulungan	0,142	0,103	0,247	0,148	0,218	0,143
54	Tana Tidung	0,175	0,135	0,191	0,158	0,176	0,166
55	Nunukan	0,197	0,103	0,210	0,150	0,170	0,171
56	Tarakan	0,140	0,136	0,207	0,166	0,202	0,149

After calculating the change in the membership matrix, the next step is to recalculate the centroid, and objective function with the updated centroid and calculate the change matrix again. Iteration stops when $|P_t - P_{t-1}| < 10^{-5}$ or $t > 1000$. Based on the calculation results, iteration stops at the 66th iteration. The final results of the membership values of 56 Regencies/Cities are shown in Table 17 :

Table 17. Final Membership Value

No	Regencies/Cities	Final Membership Value						Followed Clusters
		μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}	μ_{i5}	μ_{i6}	
1	Sambas	0,103	0,058	0,044	0,692	0,068	0,036	4
2	Bengkayang	0,771	0,076	0,064	0,061	0,012	0,016	1
3	Landak	0,067	0,028	0,020	0,857	0,015	0,012	4
4	Mempawah	0,665	0,137	0,091	0,064	0,018	0,025	1
5	Sanggau	0,136	0,059	0,046	0,703	0,031	0,025	4
6	Ketapang	0,085	0,047	0,036	0,766	0,040	0,026	4
7	Sintang	0,044	0,020	0,015	0,899	0,013	0,009	4
8	Kapuas Hulu	0,461	0,135	0,098	0,241	0,029	0,036	1
9	Sekadau	0,623	0,131	0,154	0,055	0,015	0,023	1
10	Melawi	0,481	0,162	0,196	0,098	0,026	0,036	1
11	Kayong Utara	0,275	0,202	0,370	0,072	0,031	0,051	3
12	Kubu Raya	0,138	0,107	0,082	0,340	0,239	0,094	4
13	Kota Pontianak	0,050	0,054	0,041	0,062	0,691	0,102	5
14	Kota Singkawang	0,216	0,383	0,269	0,045	0,027	0,061	2
15	Tanah Laut	0,831	0,079	0,044	0,028	0,007	0,011	1
16	Kotabaru	0,841	0,059	0,045	0,037	0,008	0,011	1
17	Banjar	0,027	0,014	0,009	0,933	0,010	0,007	4

No	Regencies/Cities	Final Membership Value						Followed Clusters
		μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}	μ_{i5}	μ_{i6}	
18	Barito Kuala	0,785	0,079	0,048	0,062	0,011	0,015	1
19	Tapin	0,228	0,205	0,503	0,029	0,012	0,023	3
20	Hulu Sungai Selatan	0,529	0,226	0,172	0,039	0,012	0,022	1
21	Hulu Sungai Tengah	0,688	0,153	0,095	0,037	0,010	0,017	1
22	Hulu Sungai Utara	0,415	0,341	0,154	0,044	0,016	0,030	1
23	Tabalong	0,169	0,590	0,128	0,041	0,021	0,052	2
24	Tanah Bumbu	0,528	0,222	0,119	0,069	0,025	0,038	1
25	Balangan	0,179	0,254	0,510	0,025	0,010	0,022	3
26	Kota Banjarmasin	0,071	0,072	0,048	0,101	0,595	0,114	5
27	Kota Banjar Baru	0,008	0,016	0,011	0,005	0,011	0,949	6
28	Kotawaringin Barat	0,250	0,533	0,132	0,035	0,016	0,035	2
29	Kotawaringin Timur	0,175	0,093	0,056	0,576	0,057	0,044	4
30	Kapuas	0,156	0,083	0,055	0,625	0,043	0,038	4
31	Barito Selatan	0,066	0,633	0,256	0,014	0,008	0,023	2
32	Barito Utara	0,067	0,509	0,387	0,013	0,007	0,018	2
33	Sukamara	0,093	0,168	0,661	0,026	0,016	0,038	3
34	Lamandau	0,021	0,069	0,895	0,005	0,003	0,008	3
35	Seruyan	0,146	0,187	0,621	0,020	0,009	0,018	3
36	Katingan	0,032	0,894	0,058	0,006	0,003	0,007	2
37	Pulang Pisau	0,131	0,375	0,451	0,018	0,008	0,017	3
38	Gunung Mas	0,123	0,293	0,504	0,029	0,015	0,036	3
39	Barito Timur	0,073	0,423	0,426	0,020	0,013	0,045	3
40	Murung Raya	0,231	0,195	0,501	0,034	0,014	0,026	3
41	Palangka Raya	0,025	0,042	0,028	0,018	0,046	0,841	6
42	Paser	0,175	0,550	0,130	0,048	0,028	0,069	2
43	Kutai Barat	0,088	0,719	0,135	0,019	0,011	0,029	2
44	Kutai Kartanegara	0,113	0,096	0,071	0,261	0,361	0,097	5
45	Kutai Timur	0,221	0,332	0,150	0,104	0,070	0,124	2
46	Berau	0,116	0,507	0,153	0,041	0,035	0,148	2
47	Penajam Paser Utara	0,045	0,174	0,753	0,009	0,005	0,014	3
48	Mahakam Ulu	0,095	0,201	0,606	0,029	0,019	0,051	3
49	Balikpapan	0,087	0,116	0,074	0,082	0,336	0,306	5
50	Samarinda	0,029	0,032	0,023	0,037	0,818	0,062	5
51	Bontang	0,103	0,283	0,227	0,045	0,049	0,293	6
52	Malinau	0,093	0,386	0,356	0,031	0,025	0,110	2
53	Bulungan	0,061	0,654	0,216	0,016	0,011	0,042	2
54	Tana Tidung	0,106	0,213	0,560	0,035	0,023	0,063	3
55	Nunukan	0,106	0,504	0,349	0,015	0,008	0,019	2
56	Tarakan	0,075	0,200	0,129	0,035	0,045	0,517	6

Based on the membership values in Table 17, the 1st observation data has a membership value of 0.103 in the first cluster, 0.058 in the second cluster, 0.044 in the third cluster, 0.692 in the fourth cluster, 0.068 in the fifth cluster and 0.036 in the sixth cluster. The largest membership value is 0.692 so the 1st observation data is most appropriate as a member of the fourth cluster. The determination of the clusters carried out in the same way for the 2nd observation data to the 56th observation data.

Table 18. FCM Clustering Results Using $C = 2$

Cluster	Number of Members	Clusters Member
1	12	Bengkayang, Mempawah, Kapuas Hulu, Sekadau, Melawi, Tanah Laut, Kotabaru, Barito Kuala, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tanah Bumbu.
2	13	Singkawang, Tabalong, Kotawaringin Barat, Barito Selatan, Barito Utara, Katingan, Paser, Kutai barat, Kutai Timur, Berau, Malinau, Bulungan, Nunukan.
3	13	Kayong Utara, Tapin, Balangan, Sukamara, Lamandau, Sruyan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Penajam Paser Utara, Mahakam Ulu, Tana Tidung.
4	9	Sambas, Landak, Sanggau, Ketapang, Sintang, Kubu Raya, Banjar, Kotawaringin Timur, Kapuas.
5	5	Pontianak, Banjarmasin, Kutai Kartanegara, Balikpapan, Samarinda.
6	4	Banjar Baru, Palangka Raya, Bontang, Tarakan.

Based on Table 18, the final results of the grouping of regencies/cities on Kalimantan Island are based on education indicators in 2021 using the FCM method with $C = 6$ obtained the first cluster consisting of 12 Regencies/Cities, the second cluster consists of 13 Regencies/Cities, the third cluster consisting of 13 Regencies/Cities, the fourth cluster consists of 9 Regencies/Cities, the fifth cluster consisting of 5 Regencies/Cities and the sixth cluster consists of 4 Regencies/Cities. Calculations carried out with the same steps using C = 2 to 5, and the results of grouping with $C = 2, 3, 4$ and 5 as follows

Table 19. FCM Clustering Results Using $C = 2$

Cluster	Number of Members	Clusters Member
1	15	Sambas, Landak, Sangau, Ketapang, Sintang, Kapuas Hulu, Kubu Raya, Pontianak, Banjar, Banjarmasin, Kotawaringin Timur, Kapuas, Kutai Kartanegara, Balikpapan, Samarinda.
2	41	Bengkayang, Mempawah, Sekadau, Melawi, Kayong Utara, Singkawang, Tanah Laut, Kotabaru, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Banjar Baru, Kotawaringin Barat, Barito Selatan, Barito Utara, Sukamara, Lamandau, Seruyan, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Palangkaraya, Paser, Kutai Barat, Kutai Timur, Berau, Penajam Paser Utara, Mahakam Ulu, Bontang, Malinau, Tana Tidung, Nunukan, Tarakan.

Table 20. FCM Clustering Results Using $C = 3$

Cluster	Number of Members	Clusters Member
1	12	Sambas, Bengkayang, Landak, Sanggau, Ketapang, Sitang, Kapuas Hulu, Kubu Raya, Banjar, Barito Kuala, Kotawaringin Timur, Kapuas.
2	7	Pontianak, Banjarmasin, Banjar Baru, Palangka Raya, Kutai Kartanegara, Balikpapan, Samarinda.
3	37	Mempawah, Sekadau, Melawi, Kayong Utara, Singkawang, Tanah Laut, Kotabaru, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Kotawaringin Barat, Barito Selatan, Barito Utara, Sukamara, Lamandau, Seruyan, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Paser, Kutai Barat, Kutai Timur, Berau, Penajam Paser Utara, Mahakam Ulu, Bontang, Malinau, Bulungan, Tana Tidung, Nunukan, Tarakan.

Table 21. FCM Clustering Results Using $C = 4$

Cluster	Number of Members	Clusters Member
1	7	Pontianak, Banjarmasin, Banjar Baru, Palangka Raya, Kutai Kartanegara, Balikpapan, Samarinda.
2	19	Singkawang, Barito Selatan, Barito Utara, Sukamara, Lamandau, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Kutai Barat, Berau, Penajam Paser Utara, Mahakam Ulu, Bontang, Malinau, Bulungan, Tana Tidung, Nunukan, Tarakan.
3	9	Sambas, Landak, Sanggau, Ketapang, Sintang, Kubu Raya, Banjar, Kotawaringin Timur, Kapuas.
4	21	Bengkayang, Mempawah, Kapuas Hulu, Sekadau, Melawi, Kayong Utara, Tanah Laut, Kotabaru, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Kotawaringin Barat, Seruyan, Murung Raya, Paser, Kutai Timur.

Table 22. FCM Clustering Results Using $C = 5$

Cluster	Number of Members	Clusters Member
1	9	Sambas, Landak, Sanggau, Ketapang, Sintang, Kubu Raya, Banjar, Kotawaringin Timur, Kapuas.
2	18	Kayong Utara, Singkawang, Tapin, Balangan, Barito Selatan, Barito Utara, Sukamara, Lamandau, Seruyan, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Penajam Paser Utara, Mahakam Ulu, Tana Tidung, Nunukan.
3	11	Tabalong, Banjar Baru, Palangka Raya, Paser, Kutai Barat, Kutai Timur, Berau, Bontang, Malinau, Bulungan, Tarakan.
4	13	Bengkayang, Mempawah, Kapuas Hulu, Sekadau, Melawi, Tanah Laut, Kotabaru, Barito Kuala, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tanah Bumbu, Kotawaringin Barat.
5	5	Pontianak, Banjarmasin, Kutai Kartanegara, Balikpapan, Samarinda.

5. Standard Deviation Ratio

Calculation of the ratio of standard deviation within groups (S_w) and standard deviation between groups (S_b) FCM method with $C = 2, 3, 4, 5, 6$ using equations (8) to (13) obtained the complete calculation results can be seen in Table 23 below :

Table 23. Standard Deviation ratio of K-Means and FCM

K-Means		FCM	
Number of Clusters	Standard Deviation Ratio	Number of Clusters	Standard Deviation Ratio
2	0,605	2	0,605
3	0,847	3	0,847
4	0,747	4	0,747
5	0,679	5	0,679
6	0,617	6	0,617

Based on Table 23, it can be seen that the grouping K-Means with $C = 2$ has a smaller standard deviation ratio compared to groups with $C = 3, 4, 5$ dan 6. it shows that the results of grouping with $c = 2$ are better than the results of grouping with $C = 3, 4, 5$ dan 6. While grouping FCM with $C = 6$ has a smaller standard deviation ratio compared to groups with $C = 2, 3, 4$ dan 5. it shows that the results of grouping with C= 6 are better than the results of grouping with $C = 2, 3, 4$ dan 5.

6. Best Method Interpretation

Based on the results of calculation standard deviation ratio of K-Means is smaller than FCM, indicates that the K-Means method is more appropriate for grouping regencies/cities in Kalimantan based on the education indicator. After the group formed, the next step is calculate the average value of all variables for each cluster. The average calculation results can be seen in Table 24 below :

Table 24. Variable Average of Each Cluster

Cluster	Number of Members	Variable Average					
		X_1	X_2	X_3	X_4	X_5	X_6
1	14	13,20	8,38	417	78	65.045	21.972
2	42	12,64	8,42	189	32	23.785	8.058

Based on the best grouping results, cluster 1 consists of 14 regencies/cities where 3 of the 14 members of cluster 1 are provincial capitals and there are several members classfified as a big city on the Kalimantan such as Balikpapan. While cluster 2 consists of 42 regencies/cities on the island of Kalimantan mostly members of cluster 2 are regencies/cities with variable averages smaller than the regencies/cities that are members of cluster 1. This can be seen from Table 15 where the average variable value of the expected length of schooling (X_1), variable Number of elementary schools (X_3), number of high schools (X_5), variable number of elementary students (X_9), and variable number of senior high school students (X_{11}) in cluster 2 is smaller than the regencies/city that is a member of cluster 1. However, the variable average length of schooling (X_2) for cluster 2 has a better value than cluster 1.

The findings of this study can be used as information for government agencies interested in making policies related to education indicators on the island of Kalimantan. especially the districts/cities in cluster 2 so that they can be used as evaluation material in increasing the level of education in Kalimantan. Compared with research conducted by (Ls et al., 2021) which grouped based on educational indicators using the ward method, this study compared 2 methods, K-Means and FCM methods so that more varied results were obtained while at the same time being able to find out which method was more effective used in grouping. Similar to the research conducted (Putri and Dwidayati, 2021), grouping using the K-Means method has better grouping results than the FCM method.

D. CONCLUSION AND SUGGESTION

Based on the results of research and discussion, the conclusions that can be drawn are grouping the K-Means with C = 2, 3, 4, 5 dan 6 based on the value of the standard deviation ratio shows that the K-Means method with $C = 2$ has better grouping results compared to the others. Grouping the FCM with C = 2, 3, 4, 5 dan 6 based on the standard deviation ratio shows that the FCM method with C = 6 has better grouping results than the others C. Based on the calculation of the value of the standard deviation ratio of the K-Means method of 0.605 while the value of the standard deviation ratio of the FCM method is 0.624, it can be concluded that the better method between K-Means and FCM for grouping regencies/cities in Kalimantan based on the year of education indicator 2021 is the K-Means method with $C = 2$ cluster.

Further researchers can use non-hierarchical grouping algorithms development of the K-means method, namely K-Harmonic Means and development of the Fuzzy C-means method, namely Subtractive Fuzzy C-Means.

DECLARATIONS

AUTHOR CONTIBUTION

All authors contributed to the writing of this article.

FUNDING STATEMENT

-

COMPETING INTEREST

The authors declare no conflict of interest in this article.

REFERENCES

Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., and Hidayat, R. (2022). The application of k-means clustering for province clustering in indonesia of the risk of the covid-19 pandemic based on covid-19 data. *Quality & Quantity*, 56(3):1283–1291. https://doi.org/10.1007/s11135-021-01176-w.

Barakbah, A. R. and Arai, K. (2004). Determining constraints of moving variance to find global optimum and make automatic clustering. In *Proc. Industrial Electronics Seminar (IES)*, pages 409–413.

Direktorat Statistik Kesejahteraan Rakyat (2021). Statistik kesejahteraan rakyat.

Ghozali, I. (2016). *Aplikasi Analisis Multivariete SPSS 23*. Badan Penerbit Universitas Diponegoro.

Hidayatullah, K. H. et al. (2014). Analisis klaster untuk pengelompokan kabupaten/kota di provinsi jawa tengah berdasarkan indikator kesejahteraan rakyat. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 2(1). https://.26714/jsunimus.2.1.2014%25p.

Jang, J.-S. and Sun, C.-T. (1995). Neuro-fuzzy modeling and control. *Proceedings of the IEEE*, 83(3):378–406.

Jollyta, D., Ramdhan, W., and Zarlis, M. (2020). *Konsep Data Mining Dan Penerapan*. Deepublish.

Kakushadze, Z. and Yu, W. (2017). K-means and cluster models for cancer signatures. *Biomolecular detection and quantification*, 13:7–31. https://10.1016/j.bdg.2017.07.0001.

Kusuma, D. T. et al. (2015). Prototipe komparasi model clustering menggunakan metode k-means dan fcm untuk menentukan strategi promosi: Study kasus sekolah tinggi teknik-pln jakarta. *Jurnal TICom*, 3(3). https://10.13140/RG.2.2.35612.08326.

Ls, D., Lesnussa, Y. A., Talakua, M. W., and Matdoan, M. Y. (2021). Analisis klaster untuk pengelompokkan kabupaten/kota di provinsi maluku berdasarkan indikator pendidikan dengan menggunakan metode ward. *Jurnal Statistika dan Aplikasinya*, 5(1):51–60. https://doi.org/10.21009/JSA.05105.

Mahmudi, M., Goejantoro, R., and Amijaya, F. D. T. (2021). Perbandingan metode c-means dan fuzzy c-means pada pengelompokan kabupaten/kota di kalimantan berdasarkan indikator ipm tahun 2019. *EKSPONENSIAL*, 12(2):193–200. https://doi.org/10.30872/eksponesial.v12i2.814.

Nurmin, D., Hayati, M. N., and Goejantoro, R. (2022). Penerapan metode fuzzy c-means pada pengelompokan kabupaten/kota di pulau kalimantan berdasarkan indikator kesejahteraan rakyat tahun 2020. *EKSPONENSIAL*, 13(2):189–196. https://doi.org/10.30872/eksponensial.v13i2.1068.

Purnamasari, S. B., Yasin, H., and Wuryandari, T. (2014). Pemilihan cluster optimum pada fuzzy c-means (studi kasus: Pengelompokan kabupaten/kota di provinsi jawa tengah berdasarkan indikator indeks pembangunan manusia). *Jurnal Gaussian*, 3(3):491–498. https://10.14170/j.gauss.3.3.491-498.

Putri, A. L. R. and Dwidayati, N. (2021). Analisa perbandingan k-means dan fuzzy c-means dalam pengelompokan daerah penyebaran covid-19 indonesia. *Unnes Journal of Mathematics*, pages 50–55.

Siregar, Y. (2016). Analisis perbandingan algoritma fuzzy c-means dan k-means. In *Annual Research Seminar*, volume 2, pages 151–155.

Suyanto, S. (2018). Machine learning tingkat dasar dan lanjut. *Informatika Bandung*.

Triyanto, W. A. (2015). Algoritma k-medoids untuk penentuan strategi pemasaran produk. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 6(1):183–188. https://doi.org/10.24176/simet.v6i1.254.