International Journal of u- and e- Service, Science and Technology
Vol. 9, No. 12, (2016), pp. 291-308
http://dx.doi.org/10.14257/ijunesst.2016.9.12.26

Clustering using K-Means and Fuzzy C-Means on Food Productivity

Adriyendi

IAIN Batusangkar, Indonesia
elektronikpos@gmail.com

Abstract

This paper provided an overview of analysis and implementation clustering for food productivity. Food productivity is determined by food production. Rice is one of staple food in Indonesia. Rice production is influencing adequacy level of national food production. Rice productivity is very important to accomplishment food affordability. Rice productivity per province in Indonesia must be increased, because large population and high consumption. Rice productivity that fluctuates and tends to decrease, need to clustering to determinant category cluster of productivity. Clustering is using K-Means and Fuzzy C-Means. Method improvement of K-Means is modification Intra Cluster Distance and Inter Cluster Distance. Calculate distance (Inter Cluster Distance and Intra Cluster Distance) to evaluate the clustering results and to compare the efficiency of the clustering algorithms. Method improvement of Fuzzy C-Means is modification algorithm, alternative process and iteration. Data processing is using Excel software. Clustering produce three cluster (C_1, C_2, C_3) is convergence. Measurement cluster based on comparison of membership cluster, consistency, and productivity. Membership cluster, there is point data anomaly (x_{22}, x_{23}, x_{29}, x_{33}). Consistency data on K-Means (C_1 = 72.73%, C_2 = 93.75%, C_3 = 100%). Consistency data on Fuzzy C-Means (C_1 = 100%, C_2 = 88.33%, C_3 = 87.50%). Rice Productivity is Cluster 1 (decrease), Cluster 2 (decrease, except 3 provinces), and Cluster 3 (increase, except 1 province). Majority in rice productivity is 70.59%. Result of clustering showed that majority rice productivity on category cluster is low productivity.

Keywords: *Clustering, K-Means, Fuzzy C-Means, Food, Rice Productivity*

1. Introduction

Food production per province in Indonesia influence national food production. Rice is one of staple food in Indonesia. Rice production is determining food productivity. Based on BPS Publication 2015, growth 2015 over 2014 about rice productivity is fluctuates and tends to decrease. Fluctuation in rice productivity is not good for national food production. Good productivity influence adequacy of national food production. Rice productivity on high level is a determinant factor of food affordability. Hence, need to clustering dataset to determinant low productivity and high productivity on rice production per province in Indonesia. Clustering is the process of grouping data objects into similar classes for finding similarities in data and putting similar data into groups namely cluster. Cluster is a collection of data object that are similar in same class and dissimilar to object in other class [1]. Clustering analysis is one of the important technologies in data mining, machine learning, pattern recognition, and many applications. Fuzzy clustering will be a better choice for the data points and clustering algorithms allocated each object to a cluster became most fundamental [2]. K-Means is clustering algorithm that fast, robust, relatively efficient in computational time, simple to implement, and gives comparatively good results if clusters in datasets are distinct or well separated in clustering [3].

ISSN: 2005-4246 IJUNESST

K-Means is high simplicity, fast convergence rate, efficiency, excellent especially dealing with large datasets, local search ability, better performance for spherical cluster, data samples from variant cluster show obvious different, and practical clustering algorithm in many applications [4]. Fuzzy C-Means is a clustering method similar to K-Means by using fuzzy theory to improve clustering results. Fuzzy C-Means is most popular fuzzy clustering method, easily implemented, has obtained satisfactory results, and become an important tool in many applications [5]. Fuzzy C-Means is very effective in image segmentation and clustering algorithm. [6].

2. Related Work

Based on paper [7], clustering using cluster center based K-Means and representative object based Fuzzy C-Means. Fuzzy C-Means Clustering produces close results to K-Means Clustering, but it still requires more computation time than K-Means Clustering. The time complexity of the K-Means Algorithm is O (ncdi) and the time complexity of Fuzzy C-Means Algorithm is O (ndc^2i). It is meaning that K-Means Algorithm seems to be superior then Fuzzy C-Means Algorithm. Based on paper [8], clustering is using K-Means, Fuzzy C-Means, and Possibilistic Fuzzy C-Means for segment in standard and color image. Result shown that Possibilistic Fuzzy C-Means favorable over Fuzzy C-Means and K-Means and provide the better result on noise gray scale images, but it require more computational time than K-Means and Fuzzy C-Means. Based on paper [9], K-Means Algorithm is enough to extract type of tumors from the brain cells, segmentation using Fuzzy C Means for accurate tumor shape extraction of malignant tumor, and both method is gives more accurate result.

In this paper [10] clustering algorithm is K-Means and Fuzzy C-Means. As the number of records increases the time execution of both the technique gets increased but the Fuzzy C-Means performance is found to be better than K-Means Algorithm. The precision, recall and f measure values are more accurate on applying Fuzzy C-Means compared to K-Means Algorithm. The number of data points is evenly distributed in Fuzzy C-Means Algorithm. In this paper [11], a comparative research between Fuzzy Clustering Algorithm and Hard Clustering Algorithm. Fuzzy C-Means is chosen on the behalf of Fuzzy Clustering Algorithm and K-Means Algorithm is chosen on the behalf of Hard Clustering Algorithm. On the basis of experiments, found that the computational time of K-Means Algorithm is less than that of Fuzzy C-Means Algorithm for the Iris Dataset. This research concludes that the K-Means performance is better than Fuzzy C-Means performance in terms computational time.

3. Method Proposed

3.1. Framework

Method proposed can be depictured in framework shown on Figure 1.

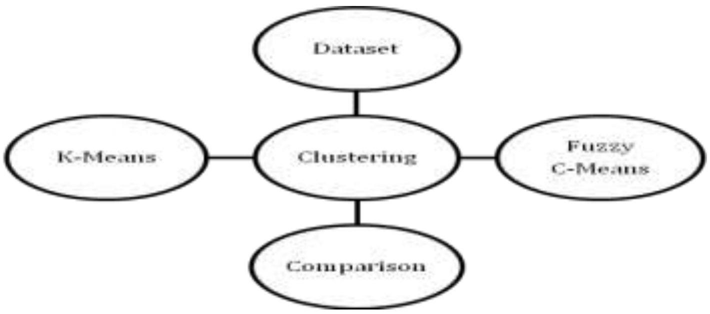

Figure 1. Framework of Clustering

Figure 1 shown framework of clustering using K-Means and Fuzzy C-Means. Dataset based on data source from BPS Publication 2015. Comparison is between K-Means and Fuzzy C-Means based on Membership Cluster, Consistency Data, and Productivity.

Algorithmic steps for K-Means Clustering [11]:

1. Select K points as initial cluster center (centroid).
2. Repeat:
3. Form k clusters by assigning all points to the closest cluster center (centroid).
4. Recompute the cluster center (centroid) of each cluster.
5. Until the cluster center (centroid) do not change.

Advantages of K-Means Clustering:

1. K-Means Clustering is a method of vector quantization that is popular for cluster analysis and works great if clusters are spherical.
2. K-Means Clustering tends to find clusters of comparable spatial extent, while the expectation maximization mechanism allows clusters to have different shapes.
3. K-Means Clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This algorithm aims at minimizing an objective function.
4. K-Means Clustering is one of the simplest algorithm which uses un-supervised learning method to solve known clustering issues.
5. K-Means Clustering works really well with large datasets.
6. K-Means Clustering may produce tighter clusters than hierarchical clustering, especially if the clusters are globular.
7. K-Means Clustering is most of the times computationally faster than hierarchical clustering for a large number of variables, if keep k smalls.
8. K-Means Clustering is faster, because order of time complexity is linear with the number of data.

Algorithmic steps of Fuzzy C-Means Clustering [11]:

1. Initialize U,μ_{ik} *k, m, t,* P_t*, maximum iteration (treshold)* .
2. Select m, initialize the membership function value.
3. Compute the cluster centers.
4. Compute Euclidean distance.
5. Update the membership function.
6. If not converged, go to step 2.

Advantages of Fuzzy C-Means Clustering:

1. Fuzzy C-Means Clustering is un-supervised learning.
2. Gives best result for overlapped dataset and comparatively better then K-Means Algorithm in process of clustering.
3. Unlike K-Means Clustering, where data point must exclusively, belong to one cluster center, here data point is assigned membership to each cluster center, as a result of which data point may belong to more than one cluster center.
4. Fuzzy C-Means Clustering works with all data is convergences.

3.2. K-Means

K-Means Algorithm can be depictured as flowchart for this application, shown on Figure 2.

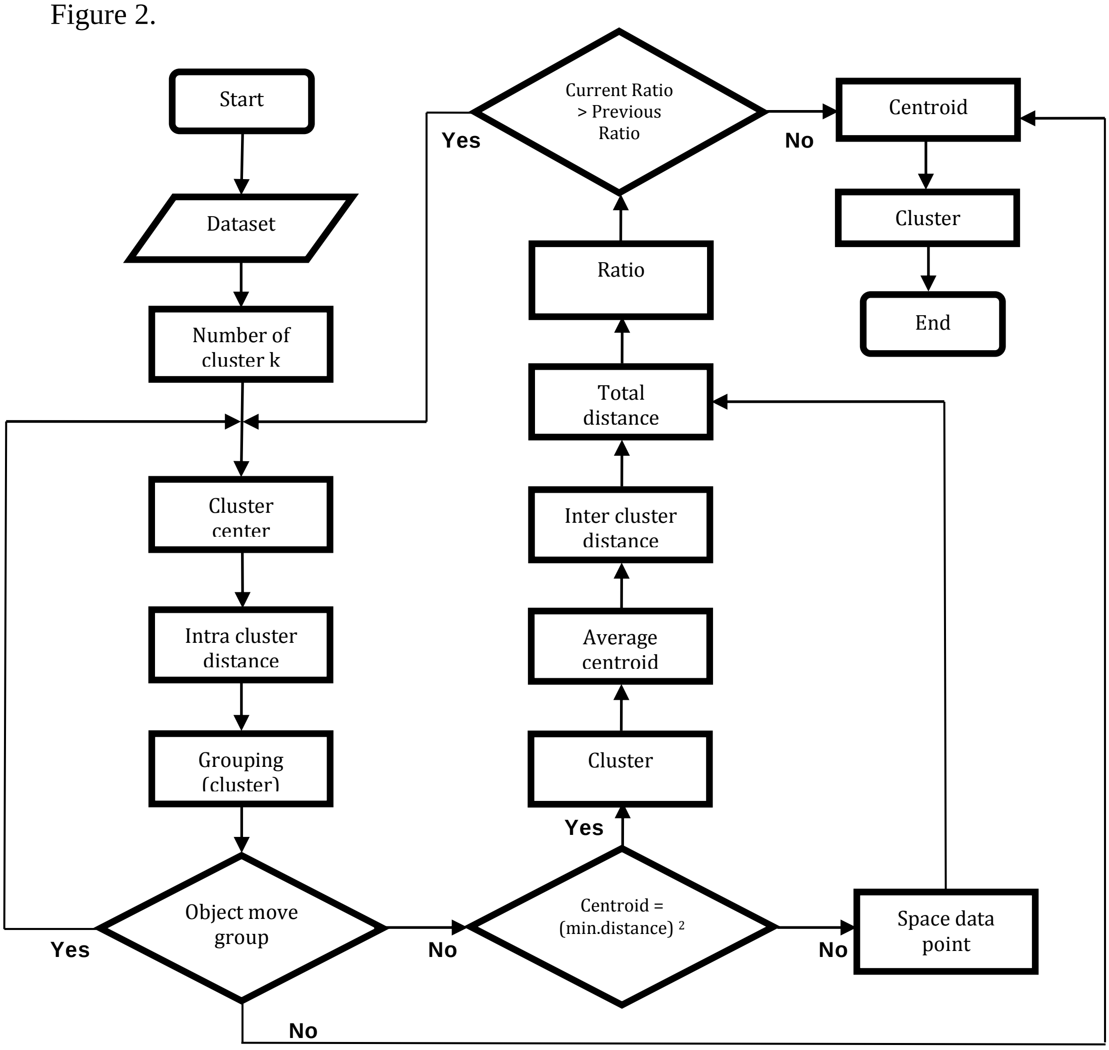

Figure 2. Flowchart K-Means Algorithm

3.3. Method Improvement of K-Means

- Intra cluster distance [12] specifies between the data vector within a cluster using Equation (1).

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}; D = \frac{1}{c} | \sum_{i=1}^{c} \sum_{k=1}^{n} \| x_k - v_i \|^2 \tag{1}$$

with D_e is Euclidean distance, i is number of object, (x, y) is coordinate of object, (s, t) is coordinate of cluster center.

- Inter cluster distance [13] is calculate minimum distance between inter cluster center for measure separation cluster using Equation (2).

$$D_{ij} = \| z_i - z_j \|; D_{inter} = \| v_i - v_j \| \tag{2}$$

with D_{ij} is distance inter cluster center, z_i is cluster center i, dan z_j is cluster center j, v_i is centroids i and v_j is centroids j. Measurement weighted exponent ($1 \leq m < \infty$, give good result: $1.5 \leq m \leq 3.0$), cluster ($2 \leq c < n$), observation data ($1 \leq k \leq N$), and number of object ($1 \leq i \leq c$).

3.4. Fuzzy C-Means

Fuzzy C-Means Algorithm can be depictured as flowchart for this application, display on Figure 3.

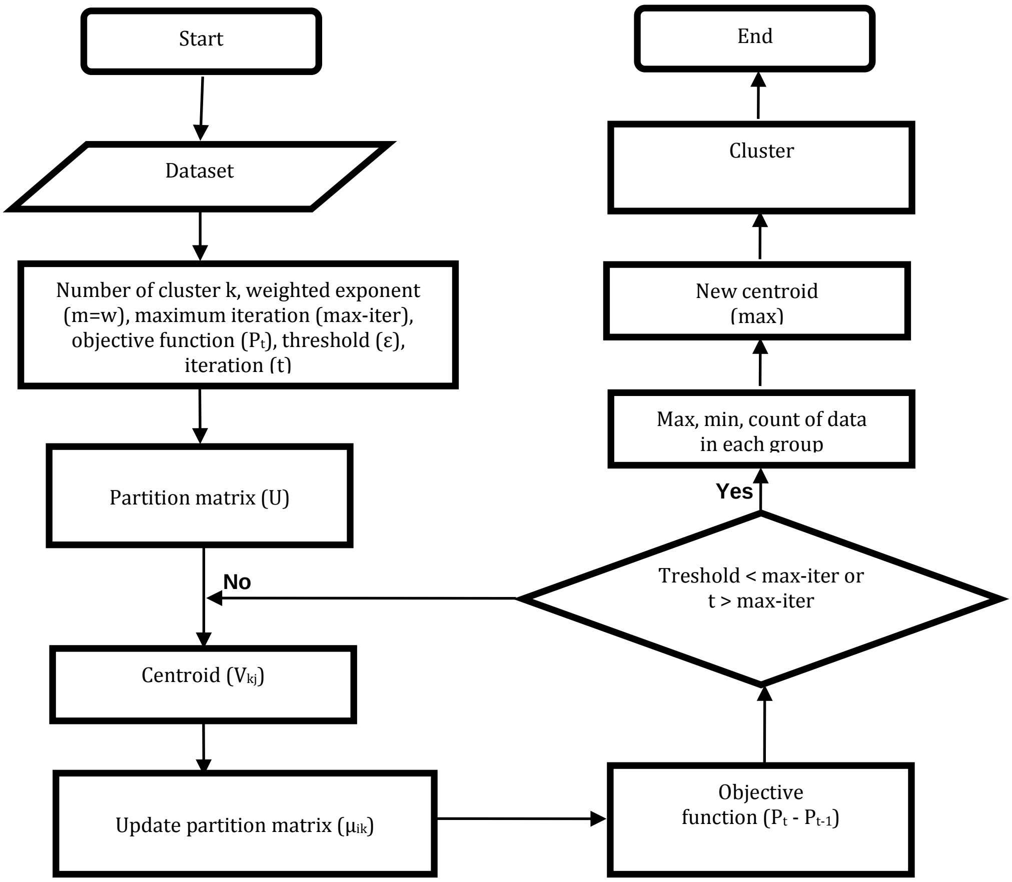

Figure 3. Flowchart Fuzzy C-Means Algorithm

Step by step of Fuzzy C-Means Algorithm based on paper [14]:

- Data (X) → matrix (n x m), X_{ij} → (data i = 1, 2, 3, …, n; attribute j = 1, 2, 3, …, m).
- Constrain → number of cluster k, weighted exponent (fuzzier = w or m), maximum iteration (max-iter), minimum error (threshold = 10^{-6}), objective function (P_t) and iteration (t = 1).
- Initialize $U=[\mu_{ij}]$, partition matrix $U^{(0)}$ and generate random number (μ_{ci}) → i = 1, 2, …, n; k = 1, 2, …, c; using Equation (3).

$$\sum_{i=1}^{c} \mu_{ci} = 1 \qquad (3)$$

- Calculation matrix of cluster center (centroid); V_{kj} (k = 1, 2 …, c; j = 1, 2 …, m) using Equation (4).

$$V_{kj} = \frac{\sum_{i=1}^{n} ((\mu_{ik})^{w} * x_{ij})}{\sum_{i=1}^{n} ((\mu_{ik})^{w})} \qquad (4)$$

- Calculation objective function (P_t) using Equation (5).

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{m} \left(x_{ij} - V_{kj} \right)^2 \right] (\mu_{ik})^{w} \right) \qquad (5)$$

- Update partition matrix (μ_{ik}) using Equation (6).

$$\mu_{ik} = \frac{\left[\sum_{i=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]^{\frac{-1}{w-1}}} \quad (6)$$

- Stopping iteration:
    - $\Delta = U^t - U^{t-1}$; if Δ < threshold (ε), then stop iteration;
    - If ($|P_t - P_{t-1}|$) < ε or (t > max-iter), then stop iteration;
    - If no, then t = t + 1; recalculated centroid (V_{kj}).

3.5. Method Improvement of Fuzzy C-Means

Fuzzy C-Means Clustering modified in this application shown in Table 1.

Table 1. Fuzzy C-Means Clustering

Algorithm	Application	Stop Iteration
- $t \leftarrow 0$	- $\mu \rightarrow \mu^2 \rightarrow \Sigma(\mu^2)$	$\Delta = U^t - U^{t-1}$;
- *repeat*	- $\mu^2.x_i \rightarrow \Sigma(\mu^2.\, x_i)$	- *if* Δ < *threshold* (ε),
- $t \leftarrow t+1$	- $\Sigma(\mu^2.x_i)/\Sigma(\mu^2) \rightarrow V_{kj}$	*then stop iteration;*
- $U_t=F_\partial(V_{t-1})$	- $V_{kj} \rightarrow X_{ij}\,.V_{kj}$	- *If* ($|P_t - P_{t-1}|$) < ε *or*
- $V_t=G_\partial(U_{t-1})$	- $X_{ij}\, V_{kj}\,.\mu^2 \rightarrow L \rightarrow \Sigma(L)$	(t > *max-iter*),
- *until*	- $(X_{ij}.V_{kj})^{-1/w-1} \rightarrow T \rightarrow \Sigma(T)- \Sigma(L.T)$	*then stop iteration;*
- ($t=T$ *or*	- $P_t \rightarrow \Sigma(X_{ij}.V_{kj}).\Sigma(L)\ \Sigma(L.T)$	- *If not, then* $t = t + 1$;
$\lvert\lvert V_t-V_{t-1})\rvert\rvert \leq \varepsilon$	- $(\mu_{ik}) \rightarrow \Sigma(X_{ij}.V_{kj})^{-1/w-1}/$	*recalculated* (V_{kj})
- $(U,V) \leftarrow (U_t, V_t)$	$\Sigma(X_{ij}.V_{kj})^{-1/w-1}$	*update* (μ_{ik})

4. Result and Discussion

4.1. Dataset

The process of clustering using Dataset based on Data Source (Rice Productivity per Province in Publication BPS 2015), display on Table 2.

Table 2. Data Source

(x_i)	Province	Year (Ku/Ha)		Growth 2015 over 2014 (%)	(x_i)	Province	Year (Ku/Ha)		Growth 2015 over 2014 (%)
		2014 (%)	2015 (%)				2014 (%)	2015 (%)	
x_1	Aceh	48,39	50,56	4,48	x_{18}	NTB	48,80	51,71	5,96
x_2	Sumatera Utara	50,62	51,74	2,21	x_{19}	NTT	33,46	35,61	6,43
x_3	Sumatera Barat	50,06	50,25	0,38	x_{20}	Kalimantan Barat	30,35	29,40	-3,13
x_4	Riau	36,35	36,63	0,77	x_{21}	KalimantanTengah	34,57	35,07	1,45
x_5	Jambi	45,53	44,31	-2,68	x_{22}	KalimantanSelatan	42,05	41,87	-0,43
x_6	SumateraSelatan	45,26	48,67	7,53	x_{23}	Kalimantan Timur	42,55	41,20	-3,17
x_7	Bengkulu	40,20	44,92	11,74	x_{24}	Kalimantan Utara	36,05	27,27	-24,36
x_8	Lampung	51,18	51,49	0,61	x_{25}	Sulawesi Utara	48,91	49,05	0,29
x_9	BangkaBelitung	23,62	22,85	-3,26	x_{26}	Sulawesi Tengah	46,54	48,57	4,36
x_{10}	Kepulauan Riau	36,44	36,46	0,05	x_{27}	Sulawesi Selatan	52,17	52,41	0,46
x_{11}	DKI Jakarta	53,86	55,95	3,88	x_{28}	Sulawesi Tenggara	46,84	47,07	0,49
x_{12}	Jawa Barat	58,82	61,22	4,08	x_{29}	Gorontalo	50,20	55,51	10,58
x_{13}	Jawa Tengah	53,57	60,25	12,47	x_{30}	Sulawesi Barat	47,65	49,41	3,69
x_{14}	DI Yogyakarta	57,87	60,65	4,80	x_{31}	Maluku	47,52	55,72	17,26
x_{15}	Jawa Timur	59,81	61,13	2,21	x_{32}	Maluku Utara	34,01	35,11	3,23
x_{16}	Banten	52,95	56,61	6,91	x_{33}	Papua Barat	40,21	42,12	4,75
x_{17}	Bali	60,12	62,14	3,36	x_{34}	Papua	43,09	43,95	2,00
	Indonesia	51,35	53,41	4,01		Indonesia	51,35	53,41	4,01

Table 2 shown Dataset (x_i) about National Food (Rice Productivity) per province (x_i = 34 province), measure (Ku/Ha: Kuintal/Hectare = 100 kg/Ha), year (2014 and 2015 in %), growth (2015 over 2014 in %), and Indonesia (average of national food in %).

4.2. K-Means

Step 1: Dataset based on Data Source in Table 2, can be seen in Table 4.

Step 2: Number of cluster k (C_j = 3) display on Table 3.

Table 3. Cluster for K-Means

Cluster				Description
C_i	x_i	z_i	z_j	Cluster (C_i) = 3
C_1	x_3	50.06	50.25	Data (x_i) = x_3, x_8, x_{27} (random)
C_2	x_8	51.18	51.49	Centroid Cluster (z_i)
C_3	x_{27}	52.17	52.41	Centroid Cluster (z_j)

Step 3: Cluster center or centroid (randomly: x_3, x_8, x_{27}) can be seen in Table 3. Next centroid based on minimum distance in each cluster.

Step 4: Intra cluster distance using Equation (1) showed on Table 4. Implementing in Excel =SQRT((C6-$C44)^2+($D$6-$D44)^2).

Step 5: Grouping based on centroid display on Table 4. Grouping using Excel =IF(F51="C1",C6,""). Grouping based on minimum distance using Excel =IF(MIN(C51:E51)=C51,C$49,IF(MIN(C51:E51)=D51,D$49,IF(MIN(C51:E51)=E51,E$49,""))). If data point move group, back to step 3. If no move group, next step 6.

Step 6: If centroid cluster = (minimum distance) ^{2}, then group in new cluster. Minimum distance using Excel =MIN(C51:E51). Minimum distance ^{2} using Excel =G51*G51. Total minimum distance (wcv) using Excel =SUM(H51:H84), display on Table 4.

Step 7: Average centroid using Excel =AVERAGEIF(I51:I84,"<>"",I51:I84) in Table 4.

Step 8: Inter cluster distance based on Equation (2) using Excel =SQRT((C98-C99)^2+(D98-D99)^2), can be seen in Table 5.

Step 9: Total distance inter cluster (bcv) using Excel =SUM(C141:C143) in Table 5.

Step 10: Ratio to threshold (bcv/wcv) using Excel =C90/H85, can be seen in Table 5.

Step 11: If ratio > threshold or current ratio > previous ratio, then recalculated new centroid using Excel =IF(C92>C93,"next iteration because current ratio > previous ratio (ratio 1 > ratio 0)"," current ratio < previous ratio, stop iteration"), shown on Table 5.

Step 12: Iteration back to Step3, to recalculating new centroid until current ratio ≤ previous ratio. On iteration 6, current ratio ≤ previous ratio, then stop iteration, shown on Table 6.

Step 13: Cluster determine by grouping data point to centroid based on minimum distance, when current ratio ≤ previous ratio. Implementing cluster in Excel =IF(MIN(C321:E321)=C321,C$49,IF(MIN(C321:E321)=D321,D$49,IF(MIN(C321:E321)=E321,E$49,""))), display on Table 7.

Inter-Cluster Distance basically specifies the distance between the cluster center (centroids) of the clusters. The maximum value of Inter-Cluster distance shows that the objects of different clusters are more dissimilar. Intra-Cluster Distance specifies the distance between the data vectors within a cluster.

Table 4. Distance for K-Means

Dataset (x_i)	Distance Intra Cluster (D C_i)						Centroid (C_1)		Centroid (C_2)		Centroid (C_3)	
	D C_1	D C_2	D C_3	C_i	min D	(min D) ^{2}	z_i	z_j	z_i	z_j	z_i	z_j
x_1	1.6985	2.9409	4.2084	C_1	1.6985	2.8850	48.39	50.56				
x_2	1.5918	0.6133	1.6886	C_2	0.6133	0.3761			50.62	51.74		
x_3	0.0000	1.6709	3.0196	C_1	0.0000	0.0000	50.06	50.25				
x_4	19.3253	20.9940	22.3446	C_1	19.3253	373.4685	36.35	36.63				
x_5	7.4702	9.1365	10.4738	C_1	7.4702	55.8045	45.53	44.31				
x_6	5.0534	6.5573	7.8572	C_1	5.0534	25.5364	45.26	48.67				
x_7	11.2084	12.7955	14.1202	C_1	11.2084	125.6285	40.20	44.92				
x_8	1.6709	0.0000	1.3515	C_2	0.0000	0.0000			51.18	51.49		
x_9	38.0767	39.7467	41.0962	C_1	38.0767	1449.8336	23.62	22.85				
x_{10}	19.3822	21.0516	22.4017	C_1	19.3822	375.6685	36.44	36.46				
x_{11}	6.8505	5.2033	3.9227	C_3	3.9227	15.3877					53.86	55.95
x_{12}	14.0385	12.3710	11.0381	C_3	11.0381	121.8386					58.82	61.22
x_{13}	10.5981	9.0802	7.9640	C_3	7.9640	63.4256					53.57	60.25
x_{14}	13.0060	11.3429	10.0194	C_3	10.0194	100.3876					57.87	60.65
x_{15}	14.6095	12.9386	11.5934	C_3	11.5934	134.4080					59.81	61.13
x_{16}	6.9858	5.4173	4.2718	C_3	4.2718	18.2484					52.95	56.61
x_{17}	15.5748	13.9049	12.5648	C_3	12.5648	157.8754					60.12	62.14
x_{18}	1.9285	2.3901	3.4419	C_1	1.9285	3.7192	48.80	51.71				
x_{19}	22.1334	23.7944	25.1457	C_1	22.1334	489.8896	33.46	35.61				
x_{20}	28.6916	30.3621	31.7108	C_1	28.6916	823.2066	30.35	29.40				
x_{21}	21.6881	23.3561	24.7070	C_1	21.6881	470.3725	34.57	35.07				
x_{22}	11.5924	13.2628	14.6118	C_1	11.5924	134.3845	42.05	41.87				
x_{23}	11.7602	13.4299	14.7719	C_1	11.7602	138.3026	42.55	41.20				
x_{24}	26.9139	28.5574	29.8643	C_1	26.9139	724.3605	36.05	27.27				
x_{25}	1.6621	3.3326	4.6816	C_1	1.6621	2.7625	48.91	49.05				
x_{26}	3.9004	5.4823	6.8149	C_1	3.9004	15.2128	46.54	48.57				
x_{27}	3.0196	1.3515	0.0000	C_3	0.0000	0.0000					52.17	52.41
x_{28}	4.5256	6.1945	7.5448	C_1	4.5256	20.4808	46.84	47.07				
x_{29}	5.2619	4.1377	3.6730	C_3	3.6730	13.4909					50.20	55.51
x_{30}	2.5522	4.0972	5.4250	C_1	2.5522	6.5137	47.65	49.41				
x_{31}	6.0310	5.5936	5.7078	C_2	5.5936	31.2885			47.52	55.72		
x_{32}	22.0640	23.7300	25.0814	C_1	22.0640	486.8221	34.01	35.11				
x_{33}	12.7718	14.4270	15.7774	C_1	12.7718	163.1194	40.21	42.12				
x_{34}	9.3953	11.0589	12.4104	C_1	9.3953	88.2709	43.09	43.95				
Total minimum distance ^{2} (wcv)						6632.9695	average centroid (z_i, z_j)					
							40.95	41.46	49.77	52.98	55.49	58.43

Highly superior clusters have high intra-class similarity (intra cluster distance) and low inter-class similarity (inter cluster distance).

Table 5. Ratio for K-Means

Inter cluster distance (D)		C_i	z_i	z_j	C_i	z_i	z_j
C_1-C_2	1.6709	C_1	40.95	41.46	C_1	40.95	41.46
C_1-C_3	3.0196	C_2	49.77	52.98	C_2	49.77	52.98
C_2-C_3	1.3515	C_3	55.49	58.43	C_3	55.49	58.43
bcv	6.0420	ratio 1		0.0009	ratio 1 > ratio 0, next iteration		
wcv	6632.9695	ratio 0		0			

Table 6. Iteration for K-Means

Iteration 1						
C_i	z_i	z_j	inter	D	bcv	wcv
C_1	50.06	50.25	C_1-C_2	1.6709	6.0420	6632.9695
C_2	51.18	51.49	C_1-C_3	3.0196	Ratio 1	0.0009
C_3	52.17	52.41	C_2-C_3	1.3515	Ratio 0	0
Ratio 1 > Ratio 0, Next Iterartion						
Iteration 2						
C_i	z_i	z_j	inter	D	bcv	wcv
C_1	40.95	41.46	C_1-C_2	14.5147	44.7529	1849.0325
C_2	49.77	52.98	C_1-C_3	22.3454	Ratio 2	0.0242
C_3	55.49	58.43	C_2-C_3	7.8928	Ratio 1	0.0009
Ratio 2 > Ratio 1, Next Iterartion						
Iteration 3						
C_i	z_i	z_j	inter	D	bcv	wcv
C_1	37.03	36.91	C_1-C_2	18.2924	60.2352	1178.1934
C_2	48.78	50.94	C_1-C_3	30.1150	Ratio 3	0.0511
C_3	56.71	59.71	C_2-C_3	11.8278	Ratio 2	0.0242
Ratio 3 > Ratio 2, Next Iterartion						
Iteration 4						
C_i	z_i	z_j	inter	D	bcv	wcv
C_1	35.82	35.71	C_1-C_2	18.9171	63.6370	1111.3304
C_2	48.18	50.03	C_1-C_3	31.8183	Ratio 4	0.0573
C_3	56.71	59.71	C_2-C_3	12.9016	Ratio 3	0.0511
Ratio 4 > Ratio 3, Next Iterartion						
Iteration 5						
C_i	z_i	z_j	inter	D	bcv	wcv
C_1	35.42	34.87	C_1-C_2	19.2477	6.4319	1096.2603
C_2	47.69	49.71	C_1-C_3	32.7122	Ratio 5	0.0597
C_3	56.71	59.71	C_2-C_3	13.4720	Ratio 4	0.0573
Ratio 5 > Ratio 4, Next Iterartion						
Iteration 6						
C_i	z_i	z_j	inter	D	bcv	wcv
C_1	35.42	34.87	C_1-C_2	19.2477	65.4319	1096.2603
C_2	47.69	49.71	C_1-C_3	32.7122	Ratio 6	0.0597
C_3	56.71	59.71	C_2-C_3	13.4720	Ratio 5	0.0597
Ratio 6 ≤ Ratio 5, Stop Iterartion						

Clustering algorithms have categories: hierarchical-based algorithms, partition-based algorithms, density-based algorithms and grid based algorithms. Partition-based clustering is cluster center (centroid) based which splits data points into k partition and each partition represents a cluster. K-Means is method can effectively improve the speed and accuracy of clustering, reducing the computational complexity.

Table 7. Membership Cluster for K-Means

xi	C_i	C_1		C_2		C_3		Membership Cluster (C_1)	Membership Cluster (C_2)	Membership Cluster (C_3)
		z_i	z_j	z_i	z_j	z_i	z_j			
x_1	C_2			48.39	50.56				Aceh	
x_2	C_2			50.62	51.74				Sumatera Utara	
x_3	C_2			50.06	50.25				Sumatera Barat	
x_4	C_1	36.35	36.63					Riau		
x_5	C_2			45.53	44.31				Jambi	
x_6	C_2			45.26	48.67				Sumatera Selatan	
x_7	C_2			40.20	44.92				Bengkulu	
x_8	C_2			51.18	51.49				Lampung	
x_9	C_1	23.62	22.85					Bangka Belitung		
x_{10}	C_1	36.44	36.46					Kepulauan Riau		
x_{11}	C_3					53.86	55.95			DKI Jakarta
x_{12}	C_3					58.82	61.22			Jawa Barat
x_{13}	C_3					53.57	60.25			Jawa Tengah
x_{14}	C_3					57.87	60.65			DI Yogyakarta
x_{15}	C_3					59.81	61.13			Jawa Timur
x_{16}	C_3					52.95	56.61			Banten
x_{17}	C_3					60.12	62.14			Bali
x_{18}	C_2			48.80	51.71				Nusa Tenggara Barat	
x_{19}	C_1	33.46	35.61					Nusa Tenggara Timur		
x_{20}	C_1	30.35	29.40					Kalimantan Barat		
x_{21}	C_1	34.57	35.07					Kalimantan Tengah		
x_{22}	C_1	42.05	41.87					Kalimantan Selatan		
x_{23}	C_1	42.55	41.20					Kalimantan Timur		
x_{24}	C_1	36.05	27.27					Kalimantan Utara		
x_{25}	C_2			48.91	49.05				Sulawesi Utara	
x_{26}	C_2			46.54	48.57				Sulawesi Tengah	
x_{27}	C_2			52.17	52.41				Sulawesi Selatan	
x_{28}	C_2			46.84	47.07				Sulawesi Tenggara	
x_{29}	C_2			50.20	55.51				Gorontalo	
x_{30}	C_2			47.65	49.41				Sulawesi Barat	
x_{31}	C_2			47.52	55.72				Maluku	
x_{32}	C_1	34.01	35.11					Maluku Utara		
x_{33}	C_1	40.21	42.12					Papua Barat		
x_{34}	C_2			43.09	43.95				Papua	

Clustering produce 3 clusters. Membership cluster in Cluster 1 is 11 data point. Membership cluster in Cluster 2 is 16 data point. Membership cluster in Cluster 3 is 7 data point.

4.3. Fuzzy C-Means

Initialize Fuzzy C-Means Clustering using Excel software, display on Table 8.

Table 8. Cluster for Fuzzy C-Means

Number of cluster k	3
Maximum iteration (max-iter)	100
Weighted exponent (m=w)	2
Treshold (minimum error = ε)	0.000001

Step 1: Dataset based on Data Source can be seen in Table 2.

Step 2: Number of cluster k (3), max-iter (100), weighted exponent or fuzzier (m=2), threshold (ε = 0.000001 or 10^{-6}), objective function (P_t) and iteration (t=1).

Step 3: Random Number (number > 0 and number < 1, count = 1) based on Equation (3), using Excel =SUM(K7:M7), can be seen on Table 9.

Table 9. Random Number for Fuzzy C-Means

x_i	C_1	C_2	C_3	Count	x_i	C_1	C_2	C_3	Count
x_1	0.5	0.4	0.1	1	x_{18}	0.1	0.6	0.3	1
x_2	0.6	0.3	0.1	1	x_{19}	0.1	0.7	0.2	1
x_3	0.7	0.2	0.1	1	x_{20}	0.1	0.8	0.1	1
x_4	0.8	0.1	0.1	1	x_{21}	0.5	0.4	0.1	1
x_5	0.4	0.1	0.5	1	x_{22}	0.6	0.3	0.1	1
x_6	0.3	0.1	0.6	1	x_{23}	0.7	0.2	0.1	1
x_7	0.2	0.1	0.7	1	x_{24}	0.8	0.1	0.1	1
x_8	0.1	0.1	0.8	1	x_{25}	0.6	0.3	0.1	1
x_9	0.1	0.5	0.4	1	x_{26}	0.8	0.1	0.1	1
x_{10}	0.1	0.6	0.3	1	x_{27}	0.3	0.1	0.6	1
x_{11}	0.1	0.7	0.2	1	x_{28}	0.1	0.1	0.8	1
x_{12}	0.1	0.8	0.1	1	x_{29}	0.1	0.6	0.3	1
x_{13}	0.4	0.1	0.5	1	x_{30}	0.1	0.5	0.4	1
x_{14}	0.3	0.1	0.6	1	x_{31}	0.2	0.1	0.7	1
x_{15}	0.2	0.1	0.7	1	x_{32}	0.4	0.1	0.5	1
x_{16}	0.1	0.1	0.8	1	x_{33}	0.7	0.2	0.1	1
x_{17}	0.1	0.5	0.4	1	x_{34}	0.5	0.4	0.1	1

Step 4: Partition matrix (μ), centroid ($V_{kj} = \Sigma (\mu^2 . x_i) / \Sigma (\mu^2)$ using Equation (4), display on Table 10 and Table 11.

Table 10. Calculation Centroid for Fuzzy C-Means

Matrix (μ)	Excel
μ^2	=POWER(B46,H7)
$\Sigma (\mu^2)$	=SUM(F46:F79)
$\mu^2 . x_i$	=F46*C7
$\Sigma (\mu^2 . x_i)$	=SUM(J46:J79)
$\Sigma (\mu^2 . x_i) / \Sigma (\mu^2)$	=J81/F80

Table 11. Centroid for Fuzzy C-Means

x_i	μ			$μ^2$			$μ^2.x_i$ (C_1)		$μ^2.x_i$ (C_2)		$μ^2.x_i$ (C_3)	
x_1	0.5	0.4	0.1	0.2500	0.1600	0.0100	12.0975	12.6400	7.7424	8.0896	0.4839	0.5056
x_2	0.6	0.3	0.1	0.3600	0.0900	0.0100	18.2232	18.6264	4.5558	4.6566	0.5062	0.5174
x_3	0.7	0.2	0.1	0.4900	0.0400	0.0100	24.5294	24.6225	2.0024	2.0100	0.5006	0.5025
x_4	0.8	0.1	0.1	0.6400	0.0100	0.0100	23.2640	23.4432	0.3635	0.3663	0.3635	0.3663
x_5	0.4	0.1	0.5	0.1600	0.0100	0.2500	7.2848	7.0896	0.4553	0.4431	11.3825	11.0775
x_6	0.3	0.1	0.6	0.0900	0.0100	0.3600	4.0734	4.3803	0.4526	0.4867	16.2936	17.5212
x_7	0.2	0.1	0.7	0.0400	0.0100	0.4900	1.6080	1.7968	0.4020	0.4492	19.6980	22.0108
x_8	0.1	0.1	0.8	0.0100	0.0100	0.6400	0.5118	0.5149	0.5118	0.5149	32.7552	32.9536
x_9	0.1	0.5	0.4	0.0100	0.2500	0.1600	0.2362	0.2285	5.9050	5.7125	3.7792	3.6560
x_{10}	0.1	0.6	0.3	0.0100	0.3600	0.0900	0.3644	0.3646	13.1184	13.1256	3.2796	3.2814
x_{11}	0.1	0.7	0.2	0.0100	0.4900	0.0400	0.5386	0.5595	26.3914	27.4155	2.1544	2.2380
x_{12}	0.1	0.8	0.1	0.0100	0.6400	0.0100	0.5882	0.6122	37.6448	39.1808	0.5882	0.6122
x_{13}	0.4	0.1	0.5	0.1600	0.0100	0.2500	8.5712	9.6400	0.5357	0.6025	13.3925	15.0625
x_{14}	0.3	0.1	0.6	0.0900	0.0100	0.3600	5.2083	5.4585	0.5787	0.6065	20.8332	21.8340
x_{15}	0.2	0.1	0.7	0.0400	0.0100	0.4900	2.3924	2.4452	0.5981	0.6113	29.3069	29.9537
x_{16}	0.1	0.1	0.8	0.0100	0.0100	0.6400	0.5295	0.5661	0.5295	0.5661	33.8880	36.2304
x_{17}	0.1	0.5	0.4	0.0100	0.2500	0.1600	0.6012	0.6214	15.0300	15.5350	9.6192	9.9424
x_{18}	0.1	0.6	0.3	0.0100	0.3600	0.0900	0.4880	0.5171	17.5680	18.6156	4.3920	4.6539
x_{19}	0.1	0.7	0.2	0.0100	0.4900	0.0400	0.3346	0.3561	16.3954	17.4489	1.3384	1.4244
x_{20}	0.1	0.8	0.1	0.0100	0.6400	0.0100	0.3035	0.2940	19.4240	18.8160	0.3035	0.2940
x_{21}	0.5	0.4	0.1	0.2500	0.1600	0.0100	8.6425	8.7675	5.5312	5.6112	0.3457	0.3507
x_{22}	0.6	0.3	0.1	0.3600	0.0900	0.0100	15.1380	15.0732	3.7845	3.7683	0.4205	0.4187
x_{23}	0.7	0.2	0.1	0.4900	0.0400	0.0100	20.8495	20.1880	1.7020	1.6480	0.4255	0.4120
x_{14}	0.8	0.1	0.1	0.6400	0.0100	0.0100	23.0720	17.4528	0.3605	0.2727	0.3605	0.2727
x_{25}	0.6	0.3	0.1	0.3600	0.0900	0.0100	17.6076	17.6580	4.4019	4.4145	0.4891	0.4905
x_{26}	0.8	0.1	0.1	0.6400	0.0100	0.0100	29.7856	31.0848	0.4654	0.4857	0.4654	0.4857
x_{27}	0.3	0.1	0.6	0.0900	0.0100	0.3600	4.6953	4.7169	0.5217	0.5241	18.7812	18.8676
x_{28}	0.1	0.1	0.8	0.0100	0.0100	0.6400	0.4684	0.4707	0.4684	0.4707	29.9776	30.1248
x_{29}	0.1	0.6	0.3	0.0100	0.3600	0.0900	0.5020	0.5551	18.0720	19.9836	4.5180	4.9959
x_{30}	0.1	0.5	0.4	0.0100	0.2500	0.1600	0.4765	0.4941	11.9125	12.3525	7.6240	7.9056
x_{31}	0.2	0.1	0.7	0.0400	0.0100	0.4900	1.9008	2.2288	0.4752	0.5572	23.2848	27.3028
x_{32}	0.4	0.1	0.5	0.1600	0.0100	0.2500	5.4416	5.6176	0.3401	0.3511	8.5025	8.7775
x_{33}	0.7	0.2	0.1	0.4900	0.0400	0.0100	19.7029	20.6388	1.6084	1.6848	0.4021	0.4212
x_{34}	0.5	0.4	0.1	0.2500	0.1600	0.0100	10.7725	10.9875	6.8944	7.0320	0.4309	0.4395
$Σ(μ^2)$				6.2200	5.1100	6.1900	$Σ(μ^2 . x_i)$		$Σ(μ^2 . x_i)$		$Σ(μ^2 . x_i)$	
Iteration 1 (t=1)							270.8034	270.7107	226.7430	234.4091	300.8864	315.9030
$Σ(μ^2 . x_i) / Σ(μ^2)$							43.54	43.52	44.37	45.87	48.61	51.03

Step 5: Objective Function (P_t) using Equation (5) and updating partition matrix ($μ_{ik}$) using Equation (6), shown on Table (12). Result of calculation P_t→$X_{ij}.V_{jk}$, L→$(X_{ij}.V_{jk})$* $μ^2$, Total L→$Σ(X_{ij}.V_{jk})$, LT→$(X_{ij}.V_{jk})^{-1/w-1}$, Total LT→$Σ(X_{ij}.V_{jk})^{-1/w-1}$)→$μ_{ik}$, $Δ P_t$→ $P_t - P_{t-1}$, shown on Table 12 and Table 13.

Table 12. Objective Function for Fuzzy C-Means

x_i	$X_{ij} . V_{kj}$			$(X_{ij} . V_{kj}) * \mu^2$			$\Sigma(X_{ij} . V_{kj}) * \mu^2$	$(X_{ij} . V_{kj})^{1/w-1}$			$\Sigma (X_{ij} . V_{kj})^{1/w-1}$
x_1	73.0712	38.1126	0.2728	18.2678	6.0980	0.0027	24.3685	0.0137	0.0262	3.6658	3.705717
x_2	117.6868	73.4585	4.5441	42.3672	6.6113	0.0454	49.0240	0.0085	0.0136	0.2201	0.242174
x_3	87.8003	51.5101	2.7223	43.0222	2.0604	0.0272	45.1098	0.0114	0.0194	0.3673	0.398146
x_4	99.1687	149.7851	357.7570	63.4680	1.4979	3.5776	68.5434	0.0101	0.0067	0.0028	0.019555
x_5	4.5899	3.7818	54.6946	0.7344	0.0378	13.6737	14.4459	0.2179	0.2644	0.0183	0.500575
x_6	29.4624	8.6131	16.8027	2.6516	0.0861	6.0490	8.7867	0.0339	0.1161	0.0595	0.209558
x_7	13.0917	18.3165	108.0883	0.5237	0.1832	52.9633	53.6701	0.0764	0.0546	0.0093	0.140231
x_8	121.8866	77.8983	6.8204	1.2189	0.7790	4.3650	6.3629	0.0082	0.0128	0.1466	0.167662
x_9	824.0650	960.7035	1418.7844	8.2407	240.1759	227.0055	475.4220	0.0012	0.0010	0.0007	0.002959
x_{10}	100.2555	151.5205	360.4850	1.0026	54.5474	32.4436	87.9936	0.0100	0.0066	0.0028	0.019348
x_{11}	260.9933	191.5680	51.7416	2.6099	93.8683	2.0697	98.5479	0.0038	0.0052	0.0193	0.028378
x_{12}	546.7513	444.2749	208.0217	5.4675	284.3360	2.0802	291.8837	0.0018	0.0023	0.0048	0.008887
x_{13}	380.4558	291.3047	109.5439	60.8729	2.9130	27.3860	91.1720	0.0026	0.0034	0.0091	0.015190
x_{14}	498.7670	400.5559	178.2356	44.8890	4.0056	64.1648	113.0594	0.0020	0.0025	0.0056	0.010112
x_{15}	574.8133	471.1069	227.3953	22.9925	4.7111	111.4237	139.1273	0.0017	0.0021	0.0044	0.008260
x_{16}	259.8742	188.8664	49.9361	2.5987	1.8887	31.9591	36.4465	0.0038	0.0053	0.0200	0.029168
x_{17}	621.5853	512.6143	255.8496	6.2159	128.1536	40.9359	175.3053	0.0016	0.0020	0.0039	0.007468
x_{18}	94.7268	53.6786	0.4931	0.9473	19.3243	0.0444	20.3159	0.0106	0.0186	2.0280	2.057144
x_{19}	164.1661	224.4020	467.3884	1.6417	109.9570	18.6955	130.2942	0.0061	0.0045	0.0021	0.012687
x_{20}	373.3592	467.9752	801.4193	3.7336	299.5041	8.0142	311.2519	0.0027	0.0021	0.0012	0.006063
x_{21}	151.8633	212.7838	451.9409	37.9658	34.0454	4.5194	76.5306	0.0066	0.0047	0.0022	0.013497
x_{22}	4.9439	21.4146	126.9999	1.7798	1.9273	1.2700	4.9771	0.2023	0.0467	0.0079	0.256841
x_{23}	6.3698	25.1546	133.4206	3.1212	1.0062	1.3342	5.4616	0.1570	0.0398	0.0075	0.204241
x_{24}	320.2107	415.3200	722.4622	204.9348	4.1532	7.2246	216.3127	0.0031	0.0024	0.0014	0.006915
x_{25}	59.4154	30.6855	4.0288	21.3896	2.7617	0.0403	24.1915	0.0168	0.0326	0.2482	0.297632
x_{26}	34.4909	11.9743	10.3519	22.0742	0.1197	0.1035	22.2974	0.0290	0.0835	0.0966	0.209106
x_{27}	153.5052	103.5398	14.5768	13.8155	1.0354	5.2476	20.0985	0.0065	0.0097	0.0686	0.084775
x_{28}	23.4902	7.5227	18.8440	0.2349	0.0752	12.0602	12.3703	0.0426	0.1329	0.0531	0.228569
x_{29}	188.0859	126.8399	22.5639	1.8809	45.6624	2.0307	49.5740	0.0053	0.0079	0.0443	0.057519
x_{30}	51.5737	23.2557	3.5574	0.5157	5.8139	0.5692	6.8988	0.0194	0.0430	0.2811	0.343497
x_{31}	164.6362	106.8782	23.1395	6.5854	1.0688	11.3384	18.9926	0.0061	0.0094	0.0432	0.058647
x_{32}	161.5459	223.2135	466.7020	25.8473	2.2321	116.6755	144.7550	0.0062	0.0045	0.0021	0.012813
x_{33}	13.0398	31.4078	150.0009	6.3895	1.2563	1.5000	9.1458	0.0767	0.0318	0.0067	0.115194
x_{34}	0.3829	5.3410	80.6423	0.0957	0.8546	0.8064	1.7567	2.6114	0.1872	0.0124	2.811066
Centroid			43.54	43.52	Obj. function (P_t)		2854.4937	If Δ (P_t) > (ε), then next iteration			
			44.37	45.87	Δ (P_t)		2854.4937				
			48.61	51.03	Min. error (ε)		0.000001				

Step 6: Iteration (t=1) until maximum iteration (t=100). If Δ (P_t) > (ε) or t < max-iter, then back to Step 4. On iteration 43, Δ (P_t) < (ε), stop iteration, next to Step 7.

Step 7: Grouping max. value, min. value, count of data point to new centroid based on max. value of data point in each new centroid become cluster, can be seen on Table 14.

Table 13. Calculation Update Partition Matrix for Fuzzy C-Means

Matrix (μ)	Excel	Matrix (μ)	Excel
$X_{ij}.V_{kj}$	= ((C7-B84) * (C7-B84)) + ((D7-C84) * (D7-C84))	$\Sigma(X_{ij}.V_{kj})^{-1/w-1}$	=SUM(O84:Q84)
$X_{ij}.V_{kj} * \mu^2$	=E84*F46	P_t	=SUM(M84:M117)
$\Sigma (X_{ij}.V_{kj}) * \mu^2$	=SUM(I84:K84)	$\Delta P_t \rightarrow P_t - P_{t-1},$	=ABS(B89-0)
$(X_{ij}.V_{kj})^{-1/w-1}$	=POWER(E84, (-1/(H7-1)))	Cluster (C_i)	=MAX(J3297:Q3297)

Table 14. Membership for Fuzzy C-Means

x_i	Member (μ)			max	C_i	C_1	C_2	C_3
x_1	0.0127	0.9383	0.0490	0.9383	C_2		Aceh	
x_2	0.0300	0.7468	0.2232	0.7468	C_2		Sumatera Utara	
x_3	0.0213	0.8762	0.1025	0.8762	C_2		Sumatera Barat	
x_4	0.9202	0.0623	0.0175	0.9202	C_1	Riau		
x_5	0.0645	0.8853	0.0502	0.8853	C_2		Jambi	
x_6	0.0076	0.9802	0.0123	0.9802	C_2		Sumatera Selatan	
x_7	0.2257	0.6872	0.0871	0.6872	C_2		Bengkulu	
x_8	0.0319	0.7219	0.2463	0.7219	C_2		Lampung	
x_9	0.7866	0.1417	0.0717	0.7866	C_1	Bangka Belitung		
x_{10}	0.9236	0.0596	0.0168	0.9236	C_1	Kepulauan Riau		
x_{11}	0.0139	0.1199	0.8662	0.8662	C_3			DKI Jakarta
x_{12}	0.0084	0.0386	0.9530	0.9530	C_3			Jawa Barat
x_{13}	0.0062	0.0375	0.9563	0.9563	C_3			Jawa Tengah
x_{14}	0.0043	0.0211	0.9745	0.9745	C_3			DI Yogyakarta
x_{15}	0.0120	0.0529	0.9351	0.9351	C_3			Jawa Timur
x_{16}	0.0147	0.1285	0.8568	0.8568	C_3			Banten
x_{17}	0.0158	0.0661	0.9181	0.9181	C_3			Bali
x_{18}	0.0223	0.8606	0.1171	0.8606	C_2		Nusa Tenggara Barat	
x_{19}	0.9796	0.0154	0.0050	0.9796	C_1	Nusa Tenggara Timur		
x_{20}	0.9414	0.0415	0.0171	0.9414	C_1	Kalimantan Barat		
x_{21}	0.9862	0.0105	0.0033	0.9862	C_1	Kalimantan Tengah		
x_{22}	0.2953	0.6210	0.0837	0.6210	C_2		Kalimantan Selatan	
x_{23}	0.3122	0.6027	0.0851	0.6027	C_2		Kalimantan Timur	
x_{24}	0.9072	0.0663	0.0265	0.9072	C_1	Kalimantan Utara		
x_{25}	0.0090	0.9630	0.0280	0.9630	C_2		Sulawesi Utara	
x_{26}	0.0004	0.9987	0.0008	0.9987	C_2		Sulawesi Tengah	
x_{27}	0.0349	0.5567	0.4084	0.5567	C_2		Sulawesi Selatan	
x_{28}	0.0049	0.9874	0.0076	0.9874	C_2		Sulawesi Tenggara	
x_{29}	0.0338	0.4173	0.5489	0.5489	C_3			Gorontalo
x_{30}	0.0034	0.9871	0.0094	0.9871	C_2		Sulawesi Barat	
x_{31}	0.0455	0.5797	0.3747	0.5797	C_2		Maluku	
x_{32}	0.9878	0.0092	0.0030	0.9878	C_1	Maluku Utara		
x_{33}	0.3837	0.5328	0.0836	0.5328	C_2		Papua Barat	
x_{34}	0.1383	0.7929	0.0688	0.7929	C_2		Papua	

4.4. Comparison

Comparison of member clustering in K-Means and Fuzzy C-Means shown on Table 15.

Table 15. Comparison of Membership Clustering

Membership K-Means Clustering			Membership Fuzzy C-Means Clustering		
C_1	C_2	C_3	C_1	C_2	C_3
x_4	x_1	x_{11}	x_4	x_1	x_{11}
x_9	x_2	x_{12}	x_9	x_2	x_{12}
x_{10}	x_3	x_{13}	x_{10}	x_3	x_{13}
x_{19}	x_5	x_{14}	x_{19}	x_5	x_{14}
x_{20}	x_6	x_{15}	x_{20}	x_6	x_{15}
x_{21}	x_7	x_{16}	x_{21}	x_7	x_{16}
x_{22}	x_8	x_{17}	x_{24}	x_8	x_{17}
x_{23}	x_{18}		x_{32}	x_{18}	x_{29}
x_{24}	x_{25}			x_{22}	
x_{32}	x_{26}			x_{23}	
x_{33}	x_{27}			x_{25}	
	x_{28}			x_{26}	
	x_{29}			x_{27}	
	x_{30}			x_{28}	
	x_{31}			x_{30}	
	x_{34}			x_{31}	
				x_{33}	
				x_{34}	
N=11	N=16	N=7	N=8	N=18	N=8

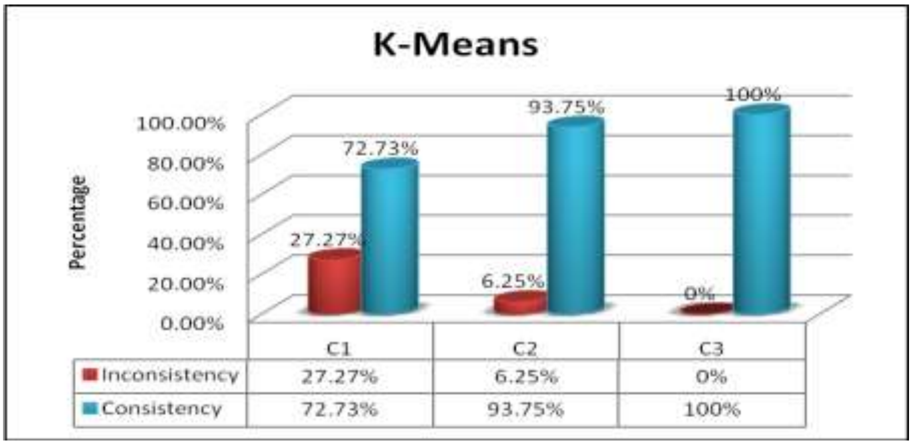

Figure 4. Comparison of Consistency K-Means

Figure 4 shown comparison of membership in K-Means, on Cluster 1 (C_1), data point (N=11), 3 data point (x_{22}, x_{23}, x_{33}) inconsistency → 3/11*100% = 27.27% and consistency → 8/11*100% = 72.73%. Cluster 2 (C_2), point data (N=16), 1 data point (x_{29}) inconsistency →1/16*100% = 6.25% and consistency →15/16*100% = 93.75%. Cluster 3 (C_3), data point (N=7), 0% inconsistency and 100% consistency.

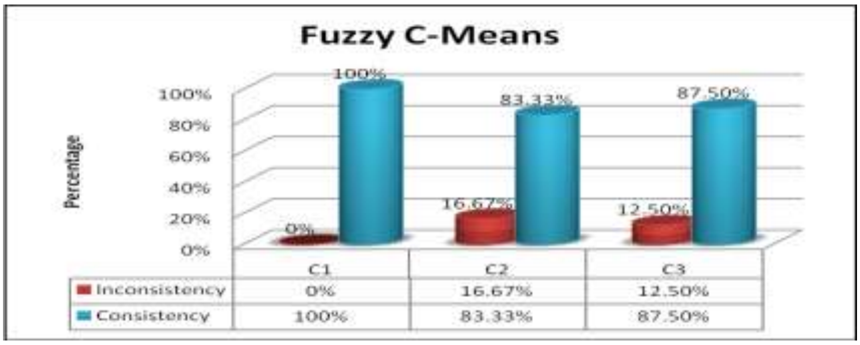

Figure 5. Comparison of Consistency Fuzzy C-Means

Figure 5 shown comparison of membership in Fuzzy C-Means, on Cluster 1 (C_1), data point (N=8), 0% inconsistency and 100% consistency. Cluster 2 (C_2), data point (N=18), 3 data point (x_{22}, x_{23}, x_{33}) inconsistency → 3/18*100% = 16.67%, and consistency → 15/18*100% = 83.33%. Cluster 3 (C_3), 1 data point (x_{29}) inconsistency → 1/8*100% = 12.50%, and consistency → 7/8*100% = 87.50%.

Table 16. Comparison of Rice Productivity for K-Means Clustering

x_i	C_1	2014 (Ku/Ha)	2015 (Ku/Ha)	x_i	C_2	2014 (Ku/Ha)	2015 (Ku/Ha)	x_i	C_3	2014 (Ku/Ha)	2015 (Ku/Ha)
x_4	Riau	36,35	36,63	x_1	Aceh	48,39	50,56	x_{11}	DKI Jakarta	53,86	55,95
x_9	BangkaBelitung	23,62	22,85	x_2	SumateraUtara	50,62	51,74	x_{12}	Jawa Barat	58,82	61,22
x_{10}	Kepulauan Riau	36,44	36,46	x_3	SumateraBarat	50,06	50,25	x_{13}	Jawa Tengah	53,57	60,25
x_{19}	NTT	33,46	35,61	x_5	Jambi	45,53	44,31	x_{14}	DIYogyakarta	57,87	60,65
x_{20}	KalimantanBarat	30,35	29,40	x_6	SumateraSelatan	45,26	48,67	x_{15}	Jawa Timur	59,81	61,13
x_{21}	KalimantanTengah	34,57	35,07	x_7	Bengkulu	40,20	44,92	x_{16}	Banten	52,95	56,61
x_{22}	KalimantanSelatan	42,05	41,87	x_8	Lampung	51,18	51,49	x_{17}	Bali	60,12	62,14
x_{23}	Kalimantan Timur	42,55	41,20	x_{18}	NTB	48,80	51,71	Rice Productivity in C_3 > Indonesia = increase			
x_{24}	Kalimantan Utara	36,05	27,27	x_{25}	Sulawesi Utara	48,91	49,05				
x_{32}	Maluku Utara	34,01	35,11	x_{26}	SulawesiTengah	46,54	48,57				
x_{33}	Papua Barat	40,21	42,12	x_{27}	SulawesiSelatan	52,17	52,41				
Rice Productivity in C_1 < Indonesia = decrease				x_{28}	SulawesiTenggara	46,84	47,07				
				x_{29}	Gorontalo	50,20	55,51				
				x_{30}	Sulawesi Barat	47,65	49,41				
				x_{31}	Maluku	47,52	55,72				
				x_{34}	Papua	43,09	43,95				
				Rice Productivity in C_2 < Indonesia = decrease							
Indonesia		51,35	53,41	Indonesia		51,35	53,41	Indonesia		51,35	53,41

Table 17. Comparison of Rice Productivity for Fuzzy C-Means Clustering

x_i	C_1	2014 (Ku/Ha)	2015 (Ku/Ha)	x_i	C_2	2014 (Ku/Ha)	2015 (Ku/Ha)	x_i	C_3	2014 (Ku/Ha)	2015 (Ku/Ha)
x_4	Riau	36,35	36,63	x_1	Aceh	48,39	50,56	x_{11}	DKI Jakarta	53,86	55,95
x_9	BangkaBelitung	23,62	22,85	x_2	SumateraUtara	50,62	51,74	x_{12}	Jawa Barat	58,82	61,22
x_{10}	Kepulauan Riau	36,44	36,46	x_3	SumateraBarat	50,06	50,25	x_{13}	Jawa Tengah	53,57	60,25
x_{19}	NTT	33,46	35,61	x_5	Jambi	45,53	44,31	x_{14}	DIYogyakarta	57,87	60,65
x_{20}	KalimantanBarat	30,35	29,40	x_6	SumateraSelatan	45,26	48,67	x_{15}	Jawa Timur	59,81	61,13
x_{21}	KalimantanTengah	34,57	35,07	x_7	Bengkulu	40,20	44,92	x_{16}	Banten	52,95	56,61
x_{24}	Kalimantan Utara	36,05	27,27	x_8	Lampung	51,18	51,49	x_{17}	Bali	60,12	62,14
x_{32}	Maluku Utara	34,01	35,11	x_{18}	NTB	48,80	51,71	x_{29}	Gorontalo	50,20	55,51
Rice Productivity in C_1 < Indonesia = decrease				x_{22}	KalimantanSelatan	42,05	41,87	Rice Productivity in C_3 > Indonesia = increase			
				x_{23}	Kalimantan Timur	42,55	41,20				
				x_{25}	Sulawesi Utara	48,91	49,05				
				x_{26}	Sulawesi Tengah	46,54	48,57				
				x_{27}	Sulawesi Selatan	52,17	52,41				
				x_{28}	SulawesiTenggara	46,84	47,07				
				x_{30}	Sulawesi Barat	47,65	49,41				
				x_{31}	Maluku	47,52	55,72				
				x_{33}	Papua Barat	40,21	42,12				
				x_{34}	Papua	43,09	43,95				
				Rice Productivity in C_2 < Indonesia = decrease							
Indonesia		51,35	53,41	Indonesia		51,35	53,41	Indonesia		51,35	53,41

Table 16 shown Comparison of Rice Productivity C_1 and Productivity of National Food (Indonesia) is decrease. Comparison of Rice Productivity C_2 and Productivity of National Food (Indonesia) is decrease. Comparison of Rice Productivity C_3 and Productivity of National Food (Indonesia) is increase. Cluster 2 (C_2), there is 3 data point data (x_{27}, x_{29}, x_{31}): 3/16 *100% = 18.75% → anomaly.

Table 17 shown Comparison of Rice Productivity C_1 and Productivity of National Food (Indonesia) is decrease. Comparison of Rice Productivity C_2 and Productivity of National Food (Indonesia) is decrease. Comparison of Rice Productivity C_3 and Productivity of National Food (Indonesia) is increase. Cluster 2 (C_2), there is 2 data point data (x_{27}, x_{31}): 2/18 *100% = 11.11% → anomaly. Cluster 2 (C_3), there is 1 data point data (x_{29}): 1/8 *100% = 12.50% → anomaly.

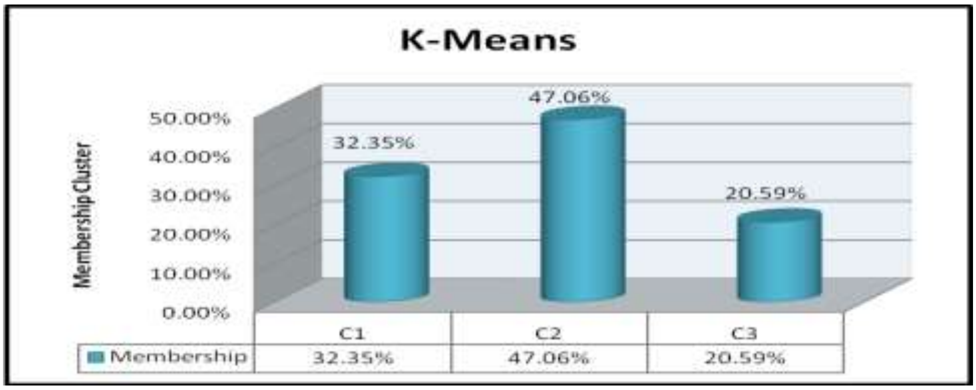

Figure 6. Membership Cluster of K-Means

Figure 6 shown membership cluster in K-Means Clustering. Percentage membership data point in Cluster 1 is 11 from 34 data point using = (11/34)*100% = 32.35%. Percentage membership data point in Cluster 2 is 16 from 34 data point using = (16/34)*100% = 47.06%. Percentage membership data point in Cluster 3 is 7 from 34 data point using = (7/34)*100% = 20.59%.

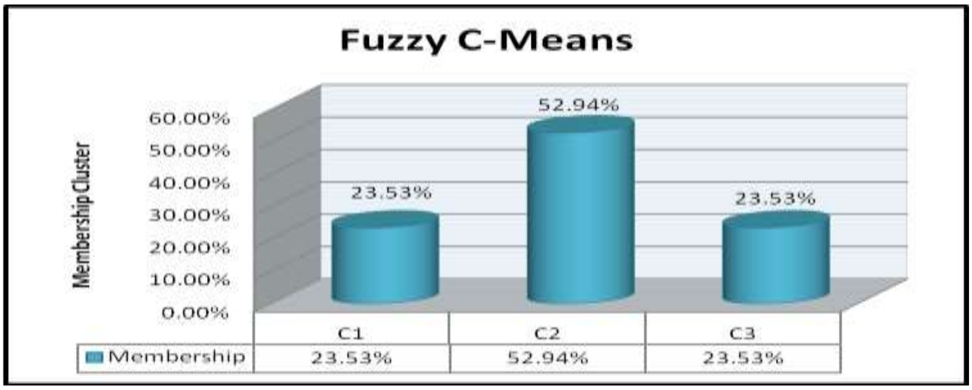

Figure 7. Membership Cluster of Fuzzy C-Means

Figure 7 shown membership cluster in Fuzzy C-Means Clustering. Percentage membership data point in Cluster 1 is 8 from 34 data point using = (8/34)*100% = 23.53%. Percentage membership data point in Cluster 2 is 18 from 34 data point using = (18/34)*100% = 52.94%. Percentage membership data point in Cluster 3 is 8 from 34 data point using = (8/34)*100% = 23.53%.

5. Conclusion

Experimental result of clustering using K-Means and Fuzzy C-Means produce three clusters. Cluster 1, Cluster 2, and Cluster 3 is convergence. Anomaly data in Membership Cluster 1 (x_{22}, x_{23}, x_{33}). Anomaly data in Membership Cluster 2 (x_{22}, x_{23}, x_{29}, x_{33}). Anomaly data in Membership Cluster 3 (x_{29}). Consistency data on K-Means (C_1 = 72.73%, C_2 = 93.75%, C_3 = 100%). Consistency data on Fuzzy C-Means (C_1 = 100%, C_2 = 88.33%, C_3 = 87.50%). Rice Productivity in Cluster 1 (decrease), Cluster 2 (decrease, except 3 provinces), and Cluster 3 (increase, except 1 province). It is meaning that Rice Productivity on Cluster 1 and Cluster 2 is low productivity. From 34 provinces, there is 10 provinces = high productivity (10/34*100% = 29.41%) and 24 provinces = low productivity (24/34*100% = 70.59%). Majority membership cluster in K-Means is C_2 (47.06%). Majority membership cluster in Fuzzy C-Means is C_2 (52.94%). Concluded that is majority of rice productivity per province in Indonesia is low. Future work, clustering by using Heuristic Fuzzy Clustering and Hierarchical Clustering.

References

[1] T. Scaria, G. Stephen and J. Mathew, "Gene expression data analysis using fuzzy c-means clustering technique", International Journal of Computer Applications, vol. 135 , no. 8, (**2016**), pp. 33-36.

[2] E. Mehdizadeh and A. Golabzaei, "Electrical fuzzy c-means: a new heuristic fuzzy clustering algorithm", Cogent Engineering, (**2016**), pp. 1-15.

[3] Z. Cebeci and F. Yildiz, "Comparison of k-means and fuzzy c-means algorithms on different cluster structures", Journal of Agricultural Informatics, (**2015**), pp. 13-23.

[4] A. Gong, Y. Gao, X. Ma, W. Gong, H. Li and Z. Gao, "An optimized artificial bee colony algorithm for clustering", International Journal of Control and Automation, vol. 9, no. 4, (**2016**), pp. 107-116.

[5] A. X. Ye and Y. X. Jin, "A fuzzy c-means clustering algorithm based on improved quantum genetic algorithm", International Journal of Database Theory and Application, vol. 9, no.1, (**2016**), pp. 227-236.

[6] E. A. S. Brar and K. Singh, "Potato defect detection using fuzzy c-mean clustering based segmentation", Indian Journal of Science and Technology, vol. 9, no. 32, (**2016**), pp. 1-6.

[7] S. Ghosh and S. K. Dubey, "Comparative analysis of k-means and fuzzy c-means algorithms", International Journal of Advanced Computer Science and Applications, vol. 4, no. 4, (**2013**), pp. 35-39.

[8] S. Singla and B. S. Khehra, "A comparative study of k-means, fuzzy c-means and possibilistic fuzzy c-means algorithm on noisy grayscale images", International Journal of Computer Applications, (**2016**), pp. 1-5.

[9] S. Majumder, S. A. Anand, K. A. Javid and V. Katarwar, "Brain tumor segmentation mechanism by using k mean and fuzzy c means", International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, iss. 4, (**2016**), pp. 7445-7451.

[10] A.Sheshasayee and P. Sharmila, "Comparative study of fuzzy c means and k means algorithm for requirements clustering", Indian Journal of Science and Technology, vol. 7, no. 6, (**2014**), pp. 853-857.

[11] D. Bora and A. K. Gupta, "A comparative study between fuzzy clustering algorithm and hard clustering algorithm", International Journal of Computer Trends and Technology, vol.10, no. 2,(**2014**), pp.08-113.

[12] Widiarina, "Data clustering using k-means dynamic algorithm", Journal of Computer Engineering AMIK BSI, vol. 1, no. 2, (**2015**), pp. 260-265.

[13] B. N. Prasad, M. Rathore, G. Gupta and T. Singh, "Performance measure of hard c-means, fuzzy c-means and alternative c-means algorithms", International Journal of Computer Science and Information Technologies, vol. 7 no. 2, (**2016**), pp. 878-883.

[14] Nurjanah, A. Farmadi and F. Indriani, "Implementation fuzzy c-means on clustering system of paddy variety", Journal Collection of Computer Science, vol. 1, no. 1, (**2014**), pp. 23-32.

