Clustering School Operational Aid Using Fuzzy Cluster Means Method

Rochdi . Wasono[1], Dwi Haryo. Ismunarti[2]
rochdi@unimus.ac.id, dwiharyois@gmail.com

[1]Muhammadiyah University of Semarang, Indonesia
[2]Diponegoro University of Semarang, Indonesia

Abstract. Clustering is needed to classify the amount of School Operational Aid (BOS)funds with other influential variables with the aim as the consideration in making policy on the distribution and amount of School Operational Aid funds. The method of Fuzzy Cluster Means was used because it has the advantage of always converging or being able to cluster to the degree of quadratic convergence. The variables used in this study were the School Operational Aid (BOS) funds, total coaching costs, and total regency/ city management costs in Central Java Province. The best result of the clustering process was to use cluster 4 with a value of Root Mean Square Error (RMSE) of 0.335.

Keywords: Clustering, Fuzzy C Means,School Operational Aid (BOS) funds

1 Introduction

School Operational Aid (BOS) funds is an Indonesian government program which is basically to provide operational and non-personnel costs for education units as a compulsory education program. One of the factors influencing the success of the BOS program is the management of funds and all resources available in the BOS program [1][10][16]. The problem is determining the amount of BOS funds because each regency and city in Central Java has a different student population in primary and junior secondary schools [15].In addition, the different geographical conditions in Central Java Province causes different funding to build educational facilities, so clustering is necessary to classify the amount of BOS funds with other influential variables with the purpose as the consideration of making distribution policy and the amount of BOS funds [11][12]. The use of appropriateCluster method will support the need for information dissemination in the form of groups or clusters to improve coordination and integration of the distribution of School Operational Aid (BOS) fund. The most important thing is the data on the receipt of aid funds that had been received by the schools in the Regencies and Cities in Central Java Province [2][7][8].

Fuzzy C-Means (FCM) is amethod of clustering to minimize objective functions set in theprocess of clustering, while other clustering methods generally try to minimize variations within a cluster and to maximize variations between clusters [9][20]. The advantages of using Fuzzy C Means algorithm are:it is always convergent or able to perform clustering with quadratic convergence; it does not require complicated calculation operations; and light computing loads.

ICE 2019, September 28-30, Purwokerto, Indonesia

DOI 10.4108/eai.28-9-2019.2291052

Therefore, convergence can be achieved relatively depending on the number of data and cluster to be achieved.

Previous studies on Fuzzy Cluster Means had been conducted, including Hidayatullah[5] who conducted a research on the selection ofclusters optimumon Fuzzy C-Means using case studies on the clustering of Regencies/ Cities in Central Java based on Human Development Index Indicators. Concludes that schools in Semarang city accept BOS, but the government and the community are still weak in supervision [16]. Conducted a study on the spatial effects of BOS using a spatial analysis which concluded that there is no association in the distribution of BOS funds in the regencies and cities in Central Java [14].

Based on the background above, this research will present the clustering of the School Operational Aid (BOS) funds, the total cost of coaching, and the total cost of the regency/ city management in Central Java Province using themethod of. Fuzzy C-Means

2 Method

The data used in this research was the data obtained from the Ministry of Education and Culture for the period of 2018. In this study, the observation units were the regencies and cities in Central Java Province. The variable used was based on the Constitutional Court Minutes No. 13 / PUU-VI / 2008[15]. The full explanation can be seen below:

Table 1. Operational definitions of variables

Variables	Indicators	Description
X1	BOS value per regency and city	Million rupiah
X2	Total supervision costs per regency and city	Million rupiah
X3	Total management costs per regency and the city	Million rupiah

2.1. Fuzzy Cluster Means

The first FCM basic concept is to determine the cluster center which will mark the average location for each cluster. In the initial condition, the cluster center is still inaccurate. Each data has a degree of membership for each cluster. By improving the cluster center and membership value of each data repeatedly, it can be seen that the cluster center will move to the right location[3][4].

The FCM algorithm is as follows :

1. Determining:
   a. Matrix X sized n x m, with n = the number of data to be clustered; and m = the number of variables.
   b. Number of clusters to be formed = C (≥ 2).
   c. Rank (weighting) = W (> 1)
   d. Maximum iteration
   e. Termination criteria = ξ (very small positive value)
   f. Initial iteration, t = 1 and Δ = 1.
2. The initial partition matrix form of U^{0} is as follows:

$$U = \begin{bmatrix} \mu_{11}(x_1)\mu_{12}(x_2)...\mu_{1n}(x_n) \\ \mu_{21}(x_1)\mu_{22}(x_2)...\mu_{2n}(x_n) \\ \vdots \qquad\qquad \vdots \\ \mu_{c1}(x_1)\mu_{c2}(x_2)...\mu_{cn}(x_n) \end{bmatrix}$$

(initial partition matrix is usually chosen randomly)

3. Calculating the cluster center V for each cluster:

$$V_{ij} = \frac{\sum_{k=1}^{n} (\mu_{ik})^{w} x_{kj}}{\sum_{k=1}^{n} (\mu_{ik})}$$

4. Fixing the degree of membership of each data in each cluster (fixing the partition matrix), as follows:

$$\mu_{ik} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(w-1)} \right]^{-1}$$

with:

$$d_{ik} = d(x_k - V_i) = \left[\sum_{j=1}^{m} (x_{kj} - V_{ij}) \right]^{1/2}$$

5. Determining the termination criteria which is the change of the partition matrix in the current iteration with the previous iteration as follows:

$$\Lambda = \left\| U^{t} - U^{t-1} \right\|$$

If $\Lambda \leq \xi$, the iteration is stopped; however, if $\Lambda > \xi$, then it increases the iteration (t = t + 1) and returns to step 3.

3 Results and Discussion

The Distribution Pattern of the School Operational Aid (BOS)in Central Java Province is mapped in the following figure:

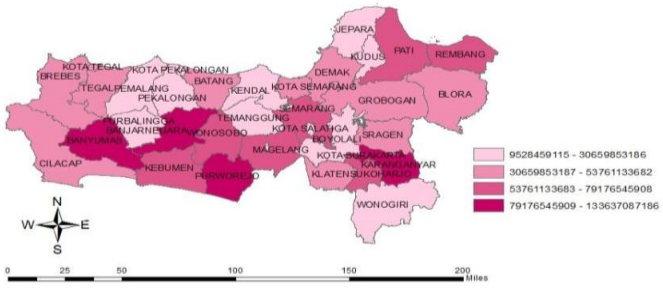

Fig 1. BOS Funding in Central Java in 2018

Based on the BOS data, it can be seen that the regencies/ cities in the BOS distribution are divided into four colors; the darker the color of the location, the higher the BOS. It can be seen that the regencies/ cities with the darkest color BOS are Banyumas, Banjarnegara, Purworejo, Karanganyar, and Surakarta City that got the biggest BOS funds among other cities and regencies. The regions that received the smallest BOS funds are Pemalang, Pekalongan, Purbalingga, Kendal, Temanggung, Boyolali, Jepara, Kudus, Pekalongan City, and Salatiga City.

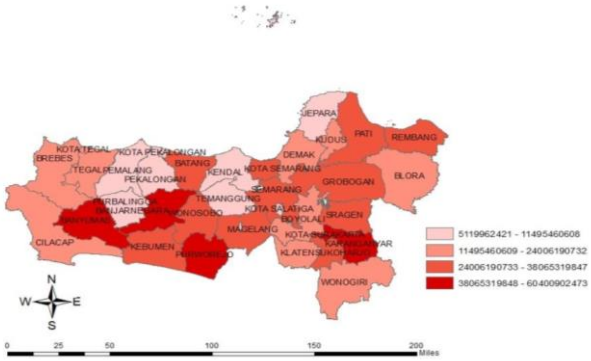

Fig 2. Funding of Total Coaching Costs in Central Java in 2018

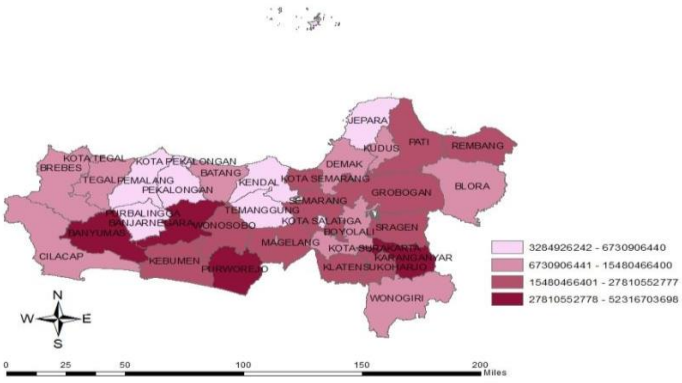

Fig 3. Funding of Total Management Costs in Central Java in 2018

3.1 Algorithm of*Fuzzy C-Means*

The process of clustering using the FCM algorithm performed early was to determine the following values: the number of clusters formed (c): 3,4,5 and 6; Rank (Weight/ w): 2; Maximum iteration: 500; The smallest expected error: 0.000001; Initial objective function (Po): 0 and initial iteration: 1.

Next, the initial random figures were generated and obtained the following results

Table 2. Initial Random Figures

Data	Cl.1	Cl.2	Cl.3	Cl.4
1	0.198068	0.130435	0.207729	0.463768
2	0.071429	0.090909	0.415584	0.422078
3	0.371542	0.391304	0.189723	0.047431
⋮	⋮	⋮	⋮	⋮
35	0.383966	0.012658	0.21097	0.392405

Then, the cluster center was calculated to obtain cluster center as follows:

Table 3. ClusterCenter

Cluster	X1	X2	X3

Cl.1	-1.29682	-1.33872	-1.19931
Cl.2	-0.36288	-0.34226	-0.2445
Cl.3	0.50369	0.79277	0.42432
Cl.4	-0.99501	-1.03925	-0.93302

Next, the objective function value for the first iteration was calculated. When the objective function value did not meet the established criteria, the next iteration was performed with the new membership matrix.When the epsilon or error value reached the expected error, the iteration process was stopped. In this study, the iteration was made 131 times to meet the epsilon value with an objective function of 7.490. Thus, the results of the cluster method using *Fuzzy C-Means* are presented in the following table:

Table 4. Membership cluster

Data	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster Results
1	0.068156	0.213287	0.025788	0.692769	4
2	0.063847	0.180974	0.659986	0.095193	3
3	0.954734	0.008204	0.003158	0.033904	1
⋮	⋮	⋮	⋮	⋮	⋮
35	0.089759	0.024504	0.006089	0.879648	4

From the table above, the details of each cluster are as follows: Cluster 1 is 5, Cluster 2 is 13 Cluster 3 is 9, and cluster 4 is 9.

The clustering process using FCM algorithm was made by testing various multiple clusters with the following results:

Table 5. Clustering Results of grouping using FCM

Number of Cluster	Iteration	Objective Function	RMSE	MAD	Within Cluster
3	46	12.036	1.318	5.286	14.814
4	131	7.490	0.335	1.436	10.529

5	34	5.114	1.744	4.916	7.452
6	147	4.243	1.513	4.546	6.965

From the table above it can be seen that the index of *Root Mean Square Error* (RMSE) isminimum at the number of cluster of 4 with the RMSE value of 0.335. the smaller the RMSE, the greater the success rate of the clustering process. Then, the best result from the clustering process in the data is by using cluster 4.

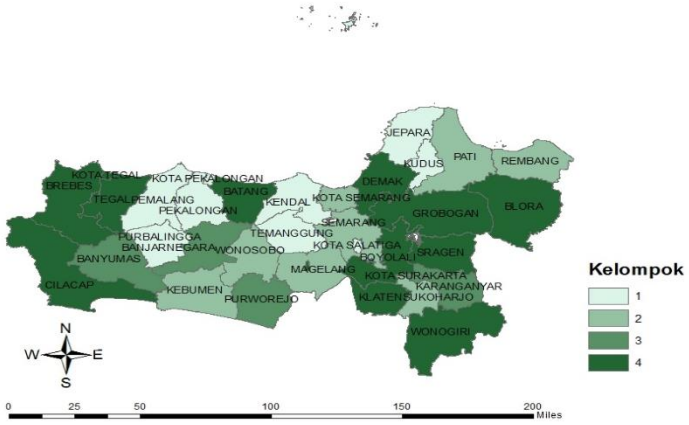

Fig. 4. Clustering Based on FCM

4 Conclusion

The clustering usng the method of Fuzzy C Means in the School Operational Aid (BOS) funds, total coaching costs, and total regency/ city management costs in Central Java Province shows the best results when using cluster 4 with the value of Root Mean Square Error (RMSE) of 0.335.

REFERENCES

[1]A. Fitri. "Pengelolaan Dana BantuanOperasionalSekolah (BOS) SekolahDasarNegeriKecamatanMandianginKotoSelayan Kota

Bukittinggi". *JurnalBahanaManajemenPendidikan*, *2*(1), 33-39,2014.

[2] A.Karim,"KajianEfekSpasialBantuanOperasionalSekolah (Bos) MenggunakanAnalisisSpasial",*JurnalStatistika*, *2*(1), 2014.

[3] H.L. Sari & D. Suranti, " PerbandinganAlgoritma Fuzzy C-Means (FCM) Dan Algoritma Mixture DalamPenclusteran Data Curah Hujan Kota Bengkulu", *JurnalFakultasHukum UII*, 2016.

[4] S. R. Kannan, S. Ramathilagam & P. C. Chung, " Effective fuzzy c-means clustering algorithms for data clustering problems", *Expert Systems with Applications*, *39*(7), 6292–6300, 2012. https://doi.org/10.1016/j.eswa.2011.11.063.

[5] K. H. Hidayatullah, "AnalisisKlasterUntukPengelompokanKabupaten/Kota di ProvinsiJawa Tengah BerdasarkanIndikatorKesejahteraan Rakyat", *JurnalStatistikaUniversitasMuhammadiyah Semarang*, *2*(1), 2014.

[6] C. Ciccarelli& S. Fachin, " Regional growth with spatial dependence: A case study on early Italian industrialization. *Papers in Regional Science*, 2016.

[7] C. Q. Ho & D. A. Hensher, " A workplace choice model accounting for spatial competition and agglomeration effects". *Journal of Transport Geography, 51*, pp.193-203, 2016.

[8] D. J. Lacombe & S. G. McIntyre, " Local and global spatial effects in hierarchical models", *Applied Economics Letters*, pp.1-5, 2016.

[9] H. L. Sari&D. Suranti. "PerbandinganAlgoritma Fuzzy C-Means (FCM) Dan Algoritma Mixture DalamPenclusteran Data Curah Hujan Kota Bengkulu". *JurnalFakultasHukum UII*, 2016.

[10] Karding, AK, "EvaluasiPelaksanaan Program BantuanOperasionalSekolah (BOS) SekolahMenengahPertamaNegeri di Kota Semarang ",Tesis Program PascasarjanaAkuntanPublikUniversitasDiponegoro. Semarang, 2008.

[11] Kangjuan, Tao, danSiyi. " Study on the Spatial Effect of Provincial Education Investment based on Spatial Statistics". International Journal of Information and Education Technology, Vol. 2, 2012.

[12] LeSage& Pace, "*Introduction to spatialeconometrics*", Boca Raton, London and New York, NY. CRC Press, 2009.

[13] PeraturanPemerintah RI. No. 48 Tahun 2008 tentang*Pendanaanpendidikan*.Jakarta: DepartemenPendidikanNasional, 2008.

[14] R. Wasono, A. Karim, M. Y. Darsyah&S, Suwardi, "Perencanaan Program BantuanOperasionalSekolah (BOS) di ProvinsiJawa Tengah Berbasiskan Model Spatial Autoregressive (SAR) Dan Spatial Error Model (SEM)". In *Prosiding Seminar Nasional &Internasional* (Vol. 1, No. 1), 2018.

[15] RisalahSidangPerkaraNomor 13/PUU-VI/2008. Jakarta. MahkamahKonstitusi, 2008.

[16] S.A.Hogantara, &J.M. Kodoatie. "EvaluasiBantuanOperasionalSekolah di Kota Semarang (Benefit Incidence Analysis)" (Doctoral dissertation, UniversitasDiponegoro), 2011

[17] S.Rahayu, U. Ludigdo, and G. Irianto, " Budgeting of School Operational Assistance Fund Based on The Value of Gotong Royong". *Procedia-Social and Behavioral Sciences*, *211*, pp.364-369, 2015.

[18] Setiawan, I. S. Ahmad, A. Karim, A."Study of Spatial Weight Matrices of SDM and SDEM for Modelling GDP Main Sector in JawaTimurIndonesia ". Research paper presented at

International Conference on Statistics and Mathematics Institut Teknologi Sepuluh Nopember, 2015.

[19] UU RI No. 20 Tahun 2003 tentang *Sistem pendidikan nasional*.Jakarta: KementrianPendidikanNasional.

[20] Z. Ji, Y. Xia, Q. Sun&G. Cao." Interval-valued possibilistic fuzzy C-means clustering algorithm. *Fuzzy Sets and Systems*",*253*(1), 138–156, 2014. https://doi.org/10.1016/j.fss.2013.12.011