PROC. INTERNAT. CONF. SCI. ENGIN. ISSN 2597-5250
Volume 3, April 2020 | Pages: 353-357 E-ISSN 2598-232X

Compare of Clustering School Operational Aid Using Fuzzy Cluster Means and Fuzzy Geographically Weighted Clustering Method

Rochdi Wasono[1,*], Dwi Haryo Ismunarti[2,]**
[1]Statistics Departement, Faculty Mathematics and Natural Science, Muhammadiyah University of Semarang,
Jl. Kedung Mundu Raya no 18 Semarang, Indonesia.
[2]Department of Oceanography, Fisheries and Marine Sciense Faculty, Diponegoro University of Semarang,
Jl. Prof. Sudarto Tembalang, Semarang, Indonesia.
Email: rochdi@unimus.ac.id*, dwiharyois@gmail.com**

Abstract. The use of appropriate Cluster method will support the distribution of School Operational Aid (BOS) fund. 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. Compare of method Fuzzy Cluster Means and Fuzzy Geographically Weighted Clustering were used. 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 Fuzzy Geographically Weighted Clustering use cluster 3.

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

INTRODUCTION

School Operational Aid (BOS) funds are government programs that are used to provide non-personnel operating costs for education units as a compulsory education program. One of the factors that influenced the success of the BOS program was the management of funds and all the resources in the program. Good management of BOS funds is a school's success through a systematic collaborative process from planning, implementation, to evaluation (Fitri, 2014). Each Regency and City Ministry of Education and Culture in Central Java has a different student population in primary, junior and senior high school, in addition to the different geographical conditions in Central Java Province causing different funding to build educational facilities.

Fuzzy C-Means (FCM) is a clustering method to minimize objective functions in the clustering process, while other clustering methods generally try to minimize variations within a cluster and maximize variation between clusters (Sari & Suranti, 2016). The advantage of using Fuzzy C Means algorithm is that it is always convergent or capable of clustering (with a quadratic convergence level), does not require complicated calculation operations, and the computational burden of light training so that convergence can be achieved more quickly depending on the amount of data and cluster to be achieved. The Fuzzy C-Means method has several weaknesses, including requiring the number of groups and the matrix of group membership predetermined (Ramadhan et al., 2015). The initial group membership matrix was randomly initialized which caused the Fuzzy C Means method has inconsistency problems Fuzzy C Means is also relatively sensitive to initialization, without good initialization this can produce fewer cluster values than previously determined (Ji et al., 2014). Fuzzy C-Means has several limitations that are very sensitive to initial solutions (initialization), constrained by local minimums and noise-sensitive, automatic central initialization methods to reduce the computational complexity of Fuzzy C Means by improper centers of the actual dataset class (Kannan et al., 2012).

One solution to overcome the weaknesses of the Fuzzy C-Means method, can be done by using the analysis of Fuzzy Geographically Weighted Clustering (FGWC) which was first introduced by (Mason & Jacobson, 2006). Fuzzy Geographically Weighted Clustering is an integration of Classical Fuzzy Clustering methods use geographically elements. Fuzzy Geographically Weighted Clustering includes geographical elements in its analysis so that the clusters formed will be sensitive to environmental effects and affect the central cluster values to create a cluster that is "geographically aware" (Sara, 2018). Fuzzy C-Means algorithm still has weaknesses in the initialization stage. To overcome weaknesses and limitations in the FCM algorithm, the FGWC algorithm is used to determine clusters that have a geographical effect therein. The use of an appropriate cluster method supports 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) funds. The main data used is data on revenue or aid funds that have

been received by schools in the Regency/City of Central Java Province.

Previous studies regarding Fuzzy Cluster Means research have been carried out, including conducted research on the selection of optimum clusters on Fuzzy C-Means with a case study on the grouping of Regencies/Cities in Central Java based on Human Development Index Indicators. Hogantara (2011) concludes that, schools in the city of Semarang accept BOS, but the government and the community are still weak in supervision. (Wasono,R. et al.,2018), conducted a study of the spatial effects of BOS using a spatial analysis which concluded that there was no linkage of the distribution of BOS funds for districts and cities in Central Java. Research on Fuzzy Geographically Weighted Clustering has been carried out by Sara (2018) regarding the grouping of People's Welfare Indicators in Central Java Province with the results of the study forming 3 optimum clusters with each different characteristic where FGWC analysis produces more significant values and fulfills assumptions compared to classic fuzzy clustering.

Based on the above background, this research will present the grouping of School Operational Aid (BOS) funds, total coaching costs, and total district/city management costs in Central Java Province using Fuzzy C-Means and Fuzzy Geographically Weighted Clustering methods.

METHOD

The data used in this study are data obtained from the Ministry of Education and Culture for the 2018 period. In this study, the observation units are regencies and cities in Central Java Province. The variable used is based on the Constitutional Court Minutes No. 13 / PUU-VI / 2008. 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

Algoritma:
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:

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

   (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.

RESULTS AND DISCUSSION

The distribution pattern of BOS (School Operational Aid) in Central Java Province is explained in the figure as follows:

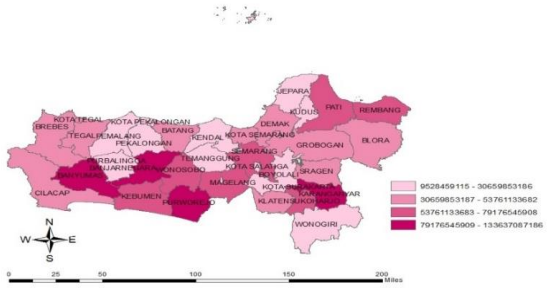

Figure 1. BOS Funding in Central Java in 2018.

Based on the BOS data, it can be seen that the districts/cities in the BOS distribution are divided into 4 colors, the color of the location is getting darker, the BOS is getting higher. It can be seen that districts / cities that have the darkest color BOS are Banyumas, Banjarnegara, Purworejo, Karanganyar, and Surakarta City that get the biggest BOS funds among other cities and regencies, regions that receive the smallest BOS funds are Pemalang, Pekalongan, Purbalingga, Kendal,

Temanggung, Boyolali, Jepara, Kudus, Pekalongan City, and Salatiga City.

Result-1

The grouping process using the FCM algorithm is done by testing a variety of many clusters, the results are as follows:

Table 2. Clustering Results of grouping using FCM.

Number of Cluster	Iterasi	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 **Table** 2 above it can be seen that the minimum Root Mean Squre Error (RMSE) index in many clusters = 4. The smaller the RMSE, the greater the success rate of the grouping process. So that the best results from the grouping process on the data is to use cluster 4.

Algorithm:

1. The first thing to do is determine the following:
   a) Number of clusters formed (c): 4.
   b) Rank (Weight / w): 2.
   c) Maximum iteration: 500.
   d) The smallest expected error: 0.000001.
   e) Initial objective function (Po): 0.
   f) Initial iteration: 1.
2. Generating initial random numbers and the following results are obtained. (For complete results on Excel)

Table 3. 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

3. Calculate the center of the cluster so that the center of the cluster is obtained as follows:

Table 4. Center Cluster.

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

4. Calculates the value of the objective function for the first iteration, and when the objective function value does not meet the specified criteria, the next iteration is performed with a new membership matrix.
5. Calculates the change in the u_ik membership matrix
6. When the epsilon or error value has reached the expected error, the iteration process is stopped. In this study, iteration was carried out 131 times to get the epsilon value fulfilled with an objective function of 7,490. So that the results obtained from the cluster using Fuzzy C-Means are presented in the following **Table** (more complete in the excel file).

Table 5. Membership cluster.

Data	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster
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

With the details of each cluster

Cluster 1: 5 Cluster 2: 13
Cluster 3: 8 Cluster 4: 9

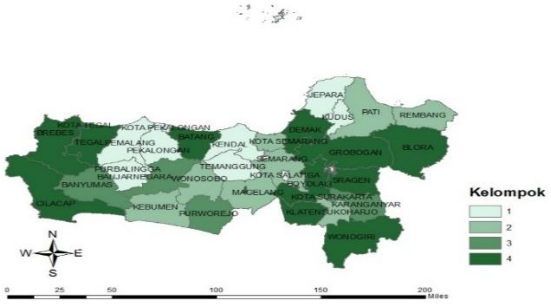

Figure 2. Clustering Based on FCM.

Result 2

The grouping process using the FGWC algorithm is done by testing a variety of many clusters, the results are as follows:

Table 6. Clustering Results of grouping using FGWC.

Number of Cluster	Objective Function	CE Index	Separation Index	PC Index
3	**26.085**	**0.948**	**0.541**	**0.439**
4	20.511	1.203	6.970	0.356
5	13.912	1.374	1.149	0.313
6	11.868	1.548	8.754	0.269

In the concept of fuzzy clustering, a member can be a member of several clusters at once according to their degree of membership. In the clustering process always look for the best solution for the defined parameters. To determine the optimal number of clusters it is necessary to have a validity index measurement. Partition Coefficient (PC) is a method that measures the number of overlapping clusters. In measuring the validity index using the PC index, the most optimal cluster is determined based on the greatest PC value (Sara, 2018). Classification Entropy (CE) is a method that measures fuzziness and cluster partitioning. The most optimal

cluster is determined based on the smallest CE value. According to (Sormin, et al, 2015) the CE index evaluates the randomness of data in clusters whose values are in the range [0.1] so that if the value gets smaller it approaches 0 then the cluster quality becomes better. The Separation Index uses the minimum distance separator for partition validity. The optimum number of clusters is indicated by the minimum S index value (Sara, 2018).

So that the best results from the grouping process on the data is to use cluster 3. And the results obtained from the cluster using FGWC are presented in the following **Table** (more complete in the excel file).

Table 7. Membership cluster FGWC.

Data	Cluster 1	Cluster 2	Cluster 3	Cluster
1	0.268667	0.147046	0.584287	3
2	0.325131	0.361276	0.313593	2
3	0.582537	0.179541	0.237923	1
⋮	⋮	⋮	⋮	⋮
35	0.77077	0.090261	0.13897	1

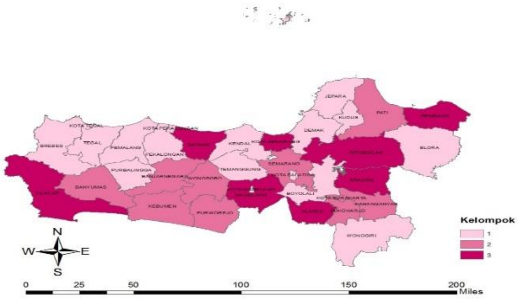

Figure 4. Clustering Based on FCM.

Result 3

Comparison of clustering results using the FCM method with FGWC

Testing multivariate normal distribution is done by looking at the correlation between mahalanobis distance and chi square values in the data. The results of multivariate normal distribution testing in the FWGC and FCM analysis are as follows:

Table 8. Pengujian Distribusi Normal Multivariate.

Analysis	Sig. (2-tailed)	Keputusan
FCM	0.151	Ho diterima
FGWC	0.000	Ho ditolak

The above **Table** is the result of a multivariate normal distribution test, with a significance level of 95%. In the FCM analysis test there is no relationship between mahalanobis distance and chi square values in the data. In this case, the data on FCM is not multivariate normally distributed and on FGWC data, it is normally multivariate. Homogeneity testing of multivariate data is done by Lavene's Test, with the following results:

Table 9. Homogeneity Testing Matriks Varians-Kovarians.

Jenis Analisis	P-Value
FCM	0.000
FGWC	0.000

Based on the **Table** above, each p-value in the FCM and FGWC analysis is the same, that is 0.000 which is less than a = 95% so the testing decision rejects Ho. Because Ho is rejected, there are differences in the variance-covariance matrix in the data, which means that the data are heterogeneous in nature, so the second assumption is not fulfilled.

One Way Manova Test and Comparison between FCM and FGWC

The last step is to conduct a one-way manova test, which is obtained to determine differences in the characteristics of each cluster. In this one way manova test uses a significance level of 95%, and the following results are obtained:

Table 10. Testing One Way Manova.

Criterion	FCM	FGWC
Wilks'	0.000	0.000
Lawley-Hotelling	0.000	0.000
Pillai's	0.000	0.000
Roy's	0.000	0.000

The above **Table** is the result of one way manova test, because the second assumption is that the homogeneity of the variance-covariance matrix is not fulfilled, so the one way manova analysis test uses Pillai's Trace. FCM and FGWC analyzes have the same value of 0.000 less than 0.05, and both show significant results. The characteristics of a good cluster according to (Hidayatullah, 2014) are multivariate normally distributed data and have different characteristics from one another (heterogeneity).

Discussion

The clustering using the method of Fuzzy C Means shows the best results when using cluster 4 with the value of Root Mean Square Error (RMSE) of 0.335 and the best results FGWC method , the grouping process on the data is to use cluster 3. This can be interpreted that the FGWC analysis is able to fulfill the characteristics of a good cluster compare FCM.

CONCLUSION

The data used in this study are data obtained from the Ministry of Education and Culture for the 2018 period.

In this study, the observation units are regencies and cities in Central Java Province. The variable used is based on the Constitutional Court Minutes No. 13 / PUU-VI / 2008. The clustering using the method of Fuzzy C Means shows the best results when using cluster 4 with the value of Root Mean Square Error (RMSE) of 0.335 and the best results FGWC method , the grouping process on the data is to use cluster 3. This can be interpreted that the FGWC analysis is able to fulfill the characteristics of a good cluster compare FCM.

REFERENCES

Fitri, A. 2014. Pengelolaan Dana Bantuan Operasional Sekolah (BOS) Sekolah Dasar Negeri Kecamatan Mandiangin Koto Selayan Kota Bukittinggi. *Jurnal Bahana Manajemen Pendidikan*, *2*(1), 33-39.

Hidayatullah, K. H. 2014. Analisis Klaster Untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Statistika Universitas Muhammadiyah Semarang*, *2*(1).

Hogantara, S. A., & Kodoatie, J. M. 2011. *Evaluasi Bantuan Operasional Sekolah di Kota Semarang (Benefit Incidence Analysis)* (Doctoral dissertation, Universitas Diponegoro).

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

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

Mason, G. A., & Jacobson, R. D. (2006). Fuzzy Geographically Weighted Clustering. In Proceedings of the 9th International Conference on Geocomputation (pp. 1– 7).

Ramadhan, R., Djauhari, Z., Kurniawandy, A., Jurusan, M., Sipil, T., Jurusan, D., & Sipil, T. (2015). Analisis perbandingan metode, *2*(2), 1–15.

Sara, D. S. 2018. *Fuzzy Geographically Weighted Clustering Untuk Pengelompokan Indikator Kesejahteraan Rakyat di Provinsi Jawa Tengah* (Universitas Muhammadiyah Semarang).

Sari, H. L., & Suranti, D. (2016). Perbandingan Algoritma Fuzzy C-Means (FCM) Dan Algoritma Mixture Dalam Penclusteran Data Curah Hujan Kota Bengkulu. *Jurnal Fakultas Hukum UII*.

Wasono, R., Karim, A., Darsyah, M. Y., & Suwardi, S. 2018. Perencanaan Program Bantuan Operasional Sekolah (BOS) di Provinsi Jawa Tengah Berbasiskan Model Spatial Autoregressive (SAR) Dan Spatial Error Model (Sem). In *Prosiding Seminar Nasional & Internasional* (Vol. 1, No. 1).

THIS PAGE INTENTIONALLY LEFT BLANK