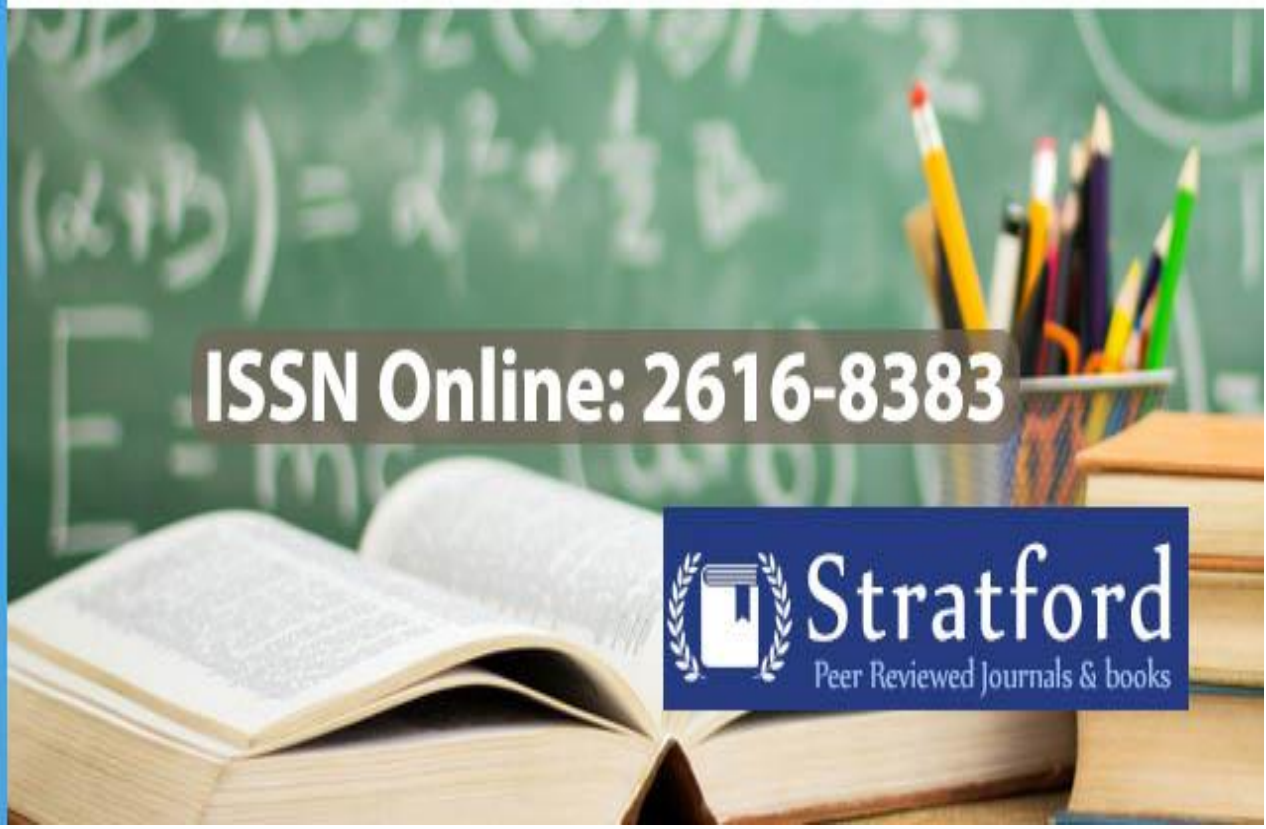
Journal of Education



Clustering of Primary and Secondary School in Indonesia Using The Fuzzy C-Means Method Based on School Self-Evaluation With Imputation Data

Agnes Tuti Rumiati, Muhammad Rif'an, Nur Achmey Selgi Harwanti & Haniza Annuril Chusna

ISSN: 2616-8383

Stratford Peer Reviewed Journals and Book Publishing
Journal of Education
Volume 4||Issue 8||Page 20-31 ||November||2021|
Email: info@stratfordjournals.org ISSN: 2616-8383

Clustering of Primary and Secondary School in Indonesia Using The Fuzzy C-Means Method Based on School Self-Evaluation With Imputation Data

[1*] Agnes Tuti Rumiati, [2] Muhammad Rif'an [3] Nur Achmey Selgi Harwanti, [4] Haniza Annuril Chusna

[1] Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota SBY, Jawa Timur 60111

**** agnes_tuti@statistika.its.ac.id***

[2] Faculty of Enginering, Universitas Negeri Jakarta
Jl. R. Mangun Muka Raya No.11, RT.11/RW.14, Rawamangun, Jakarta 13220
m.rifan@unj.ac.id

[3] Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota SBY, Jawa Timur 60111
nurachmeyselgi@gmail.com

[4]Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota SBY, Jawa Timur 60111
hanizachusna@gmail.com

***How to cite this article*: Rumiati A. T., Rif'an M., Harwanti N. A. S. & Chusna H. A (2021). Clustering of Primary and Secondary School in Indonesia Using The Fuzzy C-Means Method Based on School Self-Evaluation With Imputation Data. *Journal of Education. Vol* 4(8) pp. 20-31.**
https://doi.org/10.53819/81018102t2024

Abstract

The National Education Standard is one of the government's efforts to achieve equitable quality education. National Education Standards include eight outcomes, namely Graduation Competency Standards, Content Standards, Process Standards, Assessment Standards, Educators and Educators Standards, Facilities and Infrastructure Standards, Management Standards, and Financing Standards. This research was conducted to classify elementary and junior high schools in Indonesia based on SNP using the Fuzzy C-Means method. Prior to the cluster analysis, the missing value imputation was carried out using regression. The variables that have the lowest median and average value are the standard variables of educators and education personnel, while those with the highest value are the process standards. Based on the results of grouping using C-Means, the optimum number of clusters is four clusters with the most members being cluster 1 (the best cluster).

Keywords: *Education; Clustering; Fuzzy C-Means; imputation missing value*

Stratford Peer Reviewed Journals and Book Publishing
Journal of Education
Volume 4||Issue 8||Page 20-31 ||November||2021|
Email: info@stratfordjournals.org ISSN: 2616-8383

1.1 Introduction

Perception about measurement of the quality of a school depends on the site of whom to answer. Policy leaders tend to embrace standardized tests as the go to indicator. Many states, include other measures in their accountability systems measures like graduation rates or ones that show the narrowing of achievement gaps. Parents and community members may use test scores in their quest to understand a school's relative standing, but primarily, they rely on reputation, word-of-mouth, and what they perceive with their own eyes, looking at factors like facility conditions or student demographics (Scheneider *et al.*, 2017). Qualifications of teaching staff are found to be one of the most important factors affecting the perception of education quality (Alkareem & Hossain, 2014). Students sometimes suggest that learning is not correlated with the course design and instructor, what students actually learn does not always reflect in their grades, but recognized that student evaluation of teaching is treated as one of the widest research literature in applied psychology (Ginns, Prosser & Barrie, 2007).

An education indicator provides information about the health of the educational system. A statistic becomes an indicator when it is useful in a policy context. This statistic would qualify as an indicator when two conditions are met: 1) The statistic should measure something that relates to the health of the educational system and 2) An indicator must he placed in a particular context (Kaagan & Smith, 1985). Schools (internal) and the inspectorate (external) both influence the focus and indicators of school self-evaluation. These concepts determine the level of accountability and improvement of schools. Together, the above-mentioned elements provide input for a typology of the quality management of schools (Hofman, Djikstra &Hofman Smith, 2009). Education systems only function effectively if their strategies, approaches and funding are built on the solid foundations of data. The Digest aims to support countries as they strengthen these foundations to produce the data needed for international reporting, as well as for their own education priorities. The countries are under intense pressure to produce education data for a wide range of indicators: the 11 indicators used to monitor global progress towards SDG 4, plus a set of 32 thematic indicators to better support policymaking. Together, these indicators should deliver a full picture of progress and potential setbacks (UNESCO, 2019).

In Indonesia, the Ministry of Education and Culture has developed a National Education Standard (NES), which consists of 8 Standards as a reference for measuring the quality of education carried out by educational units. The 8 NESs consist of graduation competency standards, content, process, assessment, teachers and education staff, management, facilities and infrastructure and financing standards. By knowing the achievement of these 8 standards, the government can find out the map of the quality of education in school level, as well as regional and national level. The results of the quality mapping can be used by schools, local governments, and the central government as a reference in planning improvements and improving the quality of education according to their respective authorities (Republic of Indonesian Goverment, 2013).

The number of basic education are consist of 148.244 primary education and 38.960 Secondary Education under the management of the Ministry of Education and Culture, spread throughout of 34 Provinces and 514 Regencies / Cities. Education quality data is obtained from schools through self-assessment by the principal, teachers and educational staff. Data related standards of management and education process are measured through respondents' perceptions, while data related to infrastructure, human resources, finance are based on school records (not

perceptions) contained in the main education data (DAPODIK). School periodically completes the data through the PMP application. The 8 NSEs are described in 29 indicators and 163 sub-indicators.

The main issue in data collection is quality of data especially in completeness of information. Not all school give complete data, especially on facilities, infrastructure, and human resources data. As a result, the evaluation of sub indicators and indicators becomes not valid which also affects to the bias in NSE achievement measure. Therefore, in order to provide good and reliable analysis results, it is necessary to improve the quality of the data before analysis is carried out. In this research, data pre-processing was to overcome the missing value. Imputation of the missing data was using regression model. In order to make school quality classification we used cluster analysis for primary and secondary education level based on the 8 NSEs. The clustering methods used in this research is Fuzzy C-Means. Pseudo-F Statistics is used to evaluate the optimum number of cluster. The highest pseudo-F value indicates that the number of groups used to classify data has been optimal.

2.1 Research Methodology

2.1 Quality of School Measurement

Perception of the quality of school very depends on the site of whom to answer, it can be a standardized tests as the go to indicator, graduation rates, or the narrowing of achievement gaps, school reputation, word-of-mouth, and what they perceive with their own eyes, facility conditions or student demographics (Scheneider et al., 2017).

An education indicator provides information about the health of the educational system. A statistic becomes an indicator when it is useful in a policy context. This statistic would qualify as an indicator when two conditions are met: 1) The statistic should measure something that relates to the health of the educational system and 2) An indicator must be placed in a particular context (Kaagan & Smith, 1985).

How to measure the school quality indicator? The school self-evaluation is one process to measure quality of school by all school stakeholders. Self-evaluation is a process of reflection on practice, made systematic and transparent, to know the education process and used to improve school management, pupil, professional and organisational learning (Chapman & Sammons, 2013). School self-evaluation is a professional responsibility, self-knowledge assumes a greater sense of urgency when applied to teachers or other personnel as an accountable to their colleagues, to parents and their employers. School self-evaluation also a policy imperative, in a quality assurance context, ‘self’ is generally seen as applying to the school as an institution. A school is expected to know itself in all its complexity. In some cases, school self-evaluation is also regarded as a product, in the sense of the results of the process of school self-evaluation. The concept of school self-evaluation is a process may be narrowly defined as the “check” or “measurement” phase within a system of quality assurance (National Inspectorate of Education, 2006).

In Indonesia, quality of school measurement is developed based on The National Standard of Education (NSoE) build by The National Standard of Education Agency under responsibility by Ministry of Education and Culture. The NSoE is consist of 8 (eight) standard, i.e Graduation Competency Standards (GCS), Content, Process, Educational Assessment, Educators and Education Personnel, Facilities and Infrastructure, Management, and Financing Standards. Provisions regarding The NSoE are regulated in the Ministry of Education and Culture

Stratford Peer Reviewed Journals and Book Publishing
Journal of Education
Volume 4||Issue 8||Page 20-31 ||November||2021|
Email: info@stratfordjournals.org ISSN: 2616-8383

Regulation no 28, year 2016 (Ministry pf Education and Culture, 2016). The number of indicators is 8 NSoE are then elaborated into 29 indicators and 163 sub-indicators. Indicator performance is generally an arithmetic mean of the sub-indicators in the same indicator. Mathematically, the calculation of the composite value of indicator performance is written in equation.

$$T_{t,u} = \Sigma^{v} (\beta_{t,u,(v)} \bullet U_{t,u,(v)}) \quad (1)$$

Where:

$T_{t,u}$: achievement of the u-indicator at the t standard

$U_{t,u,v}$: composite sub indicator from measured variables

$\beta_{t,u,(v)}$: the weight of the calue of the v sub-indicator on the u- indicator and the t standard

The final standard performance score is calculated:

$$S_t = \Sigma^{u} (\gamma_{t,(u)} \bullet T_{t,(u)}) \quad (2)$$

Where:

S_t : performance of the t^{th} standard

$\gamma_{t,(u)}$: the weight of the u^{th} indicator value at the t^{th} standard with v:1,2,…,k, u:1,2,…,l and t: 1,2,…,8

All achievement figures, both sub-indicators to the standard, undergo a process of re-scaling the achievement figures to a scale of 0 to 7.Achievements to NSoE calculated using the composite method are categorized into 5 types, namely Towards SNP 1 (M1), Towards SNP 2 (M2), Towards SNP 3 (M3), Towards SNP 4 (M4), and SNP.

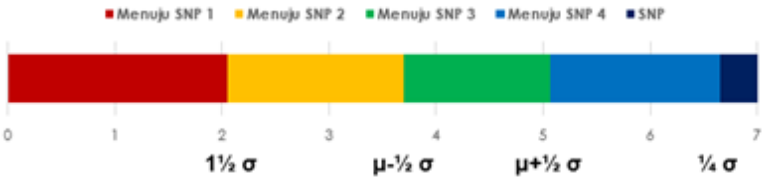


Figure 1. The range of achievements of the SNoE in the Ministry of Education and Culture grouping

The division of the SNoE performance category ranges is based on the standard deviation function. The determination of the upper and lower limits follows the pattern shown in Figure 1. There are five stages of achieving SNoE with an upper and lower limit at each stage are listed in Table 1.

Table 1: The stages of SNP achievement by the Ministry of Education and Culture

	M1	M2	M3	M4	SNP
Upper limit	0.00	2.05	3.71	5.07	6.67
Lower limit	2.04	3.70	5.06	6.66	7.00

https://doi.org/10.53819/81018102t2024

Stratford Peer Reviewed Journals and Book Publishing
Journal of Education
Volume 4||Issue 8||Page 20-31 ||November||2021|
Email: info@stratfordjournals.org ISSN: 2616-8383

2.2 Preprocessing Data

In this research, pre-processing is the process of preparing data in order to improve data quality. In this study, the pre-processing data carried out by deletion of schools with zero research standard scores, elimination of foreign schools, and overcoming missing grades.

Missing value is generally a state where there is a value of one or more variables that are missing / not available for analysis (Pyle, 1999). Missing values can cause the data become biased and allowing the invalid results of the analysis. Several treatments to overcome the missing value are:

1. If the missing value is less than <10%, it can be ignored or in other words can see the next analysis without imputation.
2. If the missing value is more than 15%, it can be a candidate for the elimination of the observation variable, but the 20% to 30% missing value criteria can still be resolved (depending on the researcher).
3. If the missing value is> 50%, data deletion can be performed

Several ways to input the missing values such as using the average value, median, estimation regression, and others (Santoso, 2010). The imputation lost values used in this study were performed using regression. Imputation of missing value with regression is an imputation method by predicting the value using information from complete data to fill in incomplete variables (Enders, 2010).

In the regression method, the response variable or also known as the dependent variable and explanatory variable or also known as the estimator variable or the independent variable (Nawari, 2010). In the imputation of missing value using the regression method, the Y variable is the variable that has missing, while the X variable is other variables to be used as explanatory variable from the complete data (without missing data). The multiple regression model is a model that studies the dependence of the response variable on two or more independent variables (Gujarati, 2006). The linear regression model in general is as shown in equation 3.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi} \quad (3)$$

2.3 Cluster Analysis to Classify the School Quality

The Fuzzy C-Means Method (FCM) is a data grouping technique, where each point in a cluster is determined by the degree of membership. Optimization partition conducted by selecting an object to be categorized into a cluster, then excluded again if the object is closed to other clusters. The basic concept of FCM is: 1). Determine the centre of the cluster, where the centre will mark the average location for each cluster and 2) Improve the cluster centre and degree of membership of each data point repeatedly so that the cluster centre moves to the right location (Bezdek, Ehrlich & Full, 1984). The loop is based on minimizing the objective function which describes the distance from the data point to the cluster centre weighted by the degree of membership of that data point.

The membership function is defined as how likely the data can become a member of a certain cluster. Each data has a membership value that is formed randomly as matrix $U^{(Y)}=[u_{ik}]$ (the membership matrix function in ith cluster. , $u_{ik}[0,1]$ is a membership value in the i^{th} cluster (i=1,2,...,c) and k^{th} object(k=1,2,...,n) with $\sum_{i=1}^{c} u_{ik} = 1, k = 1,2, \dots, n$ and $0< \sum_{i=1}^{c} u_{ik} <1$. The

Stratford Peer Reviewed Journals and Book Publishing
Journal of Education
Volume 4||Issue 8||Page 20-31 ||November||2021|
Email: info@stratfordjournals.org ISSN: 2616-8383

Cluster centre v_i is the average location for each cluster can be obtained by equation 6 (Bezdek, Ehrlich & Full, 1984).

$$\hat{v}_i = \frac{\sum_{k=1}^{n}(\hat{u}_{ik})^m y_k}{\sum_{k=1}^{n}(\hat{u}_{ik})^m} \quad (4)$$

Where:

v_i : the ith cluster centre with i=1,2,...,c

m : degree of the weight

Improvement of the cluster centre and membership value should be repeatedly give impact that the cluster will move to the right place, so the membership value is written as equation 7.

$$\hat{u}_{ik} = \left[\sum_{j=1}^{c}\left(\frac{\hat{d}_{ik}}{\hat{d}_{jk}}\right)^{2/(m-1)}\right]^{-1} \quad (5)$$

The objectives function J_m has to be improve after the cluster centres and it was done by using the following formula.

$$J_m(U,v) = \sum_{k=1}^{nN}\sum_{i=1}^{c}(u_{ik})^m \|y_k - v_i\|^2 \quad (6)$$

The iteration will stopp if $J_m < \varepsilon$, but if $J_m > \varepsilon$then re-calculate the new cluster centre.

The optimum number of clusters can be determined based on the criteria for the Pseudo F value. The highest pseudo-F value indicates that the number of groups used to classify the data has been optimal (Hair et al., 2010). The equation used to calculate the Pseudo F value is in equation bellow:

$$Pseudo\ F = \frac{\left(\frac{R^2}{c-1}\right)}{\left(\frac{1-R^2}{n-c}\right)} \quad (7)$$

Where:

$$R^2 = \frac{SST-SSW}{SST} \quad (8)$$

SST : the sum of the squares of the sample distances to the overall mean

SSW : the sum of the squares of the sample distances to the group mean

n : number of sample and c : number of cluster

Determining the best cluster method in this study is by looking at the ICD (internal cluster dispersion) rate. The smaller the ICD rate, the better the grouping results. The ICD rate is the level of disperse in the cluster, this value can be written as an equation 9 (National Inspectorate of Education, 2006).

$$\text{ICD rate} = 1 - \frac{\text{SST}-\text{SSW}}{\text{SST}} = 1 - \frac{\text{SSB}}{\text{SST}} = 1 - \text{R}^2 \quad (9)$$

where SSB (Sum Square Between) is

$$SSB = \sum_{j=1}^{c}\sum_{k=1}^{p}(x_{ijk} - \bar{x}_j)^2 \quad (10)$$

Stratford Peer Reviewed Journals and Book Publishing
Journal of Education
Volume 4||Issue 8||Page 20-31 ||November||2021|
Email: info@stratfordjournals.org ISSN: 2616-8383

C : number of variables,

P : number of cluster,

x_{ijk} : the i[th] sample of j[th] variable at k[th] cluster and

$\bar{x}_j$: sample mean in j[th] variable

3.1 Research Findings and Discussions

3.1 Data Collection

3.1.1 Participants

Nine hundred eighty-two high school students (female 534, male 443, gender not given 5, with a mean age of 17 years) from 23 randomly chosen high schools participated in the quantitative part of the study. The schools (gymnasiums) were located in three small cities (< 20,000 inhabitants), seven medium-sized cities (20,000–100,000 inhabitants) and six large cities (>100,000 inhabitants) in northwestern Germany. For detailed information on the sample grades, see online resource 1, Table 1. In the qualitative part of the study, six students (female 3, male 3; 15–16 years old; grade 10) volunteered to participate in interviews 1 week after they had filled in the questionnaire. Interviews took approximately 30 min each.

Before data collection, none of the students had engaged in the topic–context and topic–activity combinations described in the questionnaire in the classroom. As evolution is covered mainly at the end of high school before graduation, participants had little prior knowledge of evolution.

Table 2. Differences of the Average Score of Infrastructure Facilities Standard before and After Imputation

	Before imputation	After Imputation
Primary School	3,8463	5,1220
Secondary School	3,9114	5,3000

Further analysis was carried out to determine the characteristics of the 8 National Standard od Primary and Secondary school Education in Indonesia as a whole. Table 4 shows that the variable has the highest average score at both Primary and Secondary School is the standard process with an average score at primary school level of 6.5424 and secondary school level of 6.4089. Meanwhile, the lowest average score is the standard variable for educators and human resource with an average value for primary school level of 3.8947 and junior high school level of 3.3564.

Stratford Peer Reviewed Journals and Book Publishing
Journal of Education
Volume 4||Issue 8||Page 20-31 ||November||2021|
Email: info@stratfordjournals.org ISSN: 2616-8383

Table 3. The Average Score of Each National Standard of Education at Primary and Secondary School Level

Code	Standard	Primary School level	Secondary School Level
1	Standard of Graduate Competence	6,0712	6,1828
2	Standard of Content	5,7470	5,7853
3	Standard of Process	6,4524	6,4089
4	Standard of Assessment	6,0312	5,9256
5	Standard of Quality of Personnel	3,8947	3,3564
6	Standard of Infrastructure Facilities	5,1220	5,3000
7	Standard of School Management	5,7837	5,7517
8	Standard of School Funding	5,8180	5,7651

As explained in sub-chapter 2.1, it is mentioned that the indicators for setting standards for human resource quality are the availability and competence of teachers, school principals, teachers, laboratory assistants, and librarians. The quality of human resources in education quite low compared to other standards. The primary schools in Indonesia are still lacking in terms of the quality of human resources for both teachers and staffs.

3.1.2 Primary and Secondary School Level Clustering Based on National Standard of Education

The School clustering is done by using the fuzzy C-Means method. In this analysis, three clusters were determined, i.e 4, 5 and 6 clusters. Table 4 shows the results of the school clustering of Primary and Secondary level in Indonesia in 4 to 6 clusters using the fuzzy C means method. The optimum number of clusters will be selected by looking at the Pseudo F-Statistic. The following is the pseudo F value of each cluster generated.

Table 4. Pseudo F of possible number of cluster

Number of Cluster	Pseudo F Statistics	
	Primary school level	Secondary School level
4	58204,01	13265,54
5	47019,81	12160,56
6	39515,31	9001,06

Table 4 shows that the largest Pseudo F value at both the Primary and Secondary school level is for cluster 4 which indicates that the optimum number of clusters is 4 clusters. The next analysis was to look at the differences in characteristics between clusters.

Table 5. The Rating Cluster Status

Cluster	Rating Status	Number of School	
		Primary	Secondary
Cluster 1	Best	63.477 (45%)	16.551 (45%)
Cluster 2	Medium	44.680 (31%)	9.666 (26%)
Cluster 3	Upper lower	31.487 (22%)	10.091 (27%)
Cluster 4	Lowest	2.650 (2%)	532 (1%)

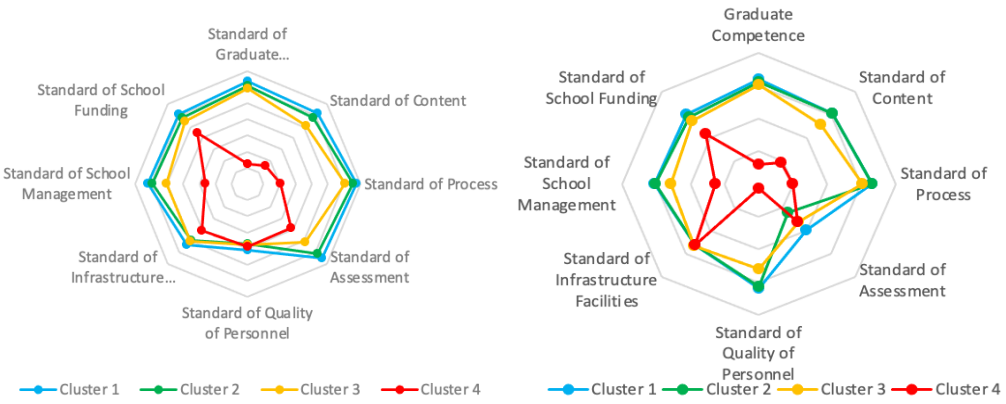
https://doi.org/10.53819/81018102t2024

Stratford Peer Reviewed Journals and Book Publishing
Journal of Education
Volume 4||Issue 8||Page 20-31 ||November||2021|
Email: info@stratfordjournals.org ISSN: 2616-8383

Table 5 shows that the highest number of members is in cluster 1 or schools with the best status, fulfilled of 45% Primary school and 16,551 Secondary school (45%). The medium status of primary school about 31% and for Secondary school are about 26% and so on.

The performance of each cluster is shown in Figure 2, whereas for Primary School level, the main problem is on quality of human resources. Meanwhile for secondary school level the main problem is standard of assessment. It means that the teachers knowledge and skill of teacher to assess student ability still in problem. Member of schools in cluster 1 shown in blue line appear to have the highest mean across all standards, while those in cluster 4 which are shown in red line have the lowest mean. Table 6 shows the characteristics of each cluster, so the rating status for each school group. It can be seen that in clusters 1, 2, and 3, both at the primary and secondary school levels, the highest average score is on standard process, while in cluster 4 the highest average score is in the assessment standard. At the primary school level, all average values for each standard in cluster 1 are higher than in cluster 2, especially in standard facilities and infrastructure which have the highest difference, which is a difference of 0,4034. A slightly different thing happened at the secondary school level where the average value of standard facilities and infrastructure in cluster 2 was slightly higher than cluster 1, which is a difference of 0.0038, but the overall value in cluster 1 was still higher than in cluster 2.

Figure 2. Radar Chart of the Average score for each Cluster



(a). Primary School Level **(b). Secondary School Level**

The lowest cluster for both primary and secondary school level have problem almost in all standards, standard of content, process and assessment and a result the graduate competence also very low. Besides, the quality of school management and human resources also quite low, but the infrastructure and funding is moderate. This condition is happened in 2650 primary schools and 532 secondary school level. In general the main problem of school quality in Indonesia is human resources and management capability. In learning process, capability of the teacher in assessment still an issue.

https://doi.org/10.53819/81018102t2024

Stratford Peer Reviewed Journals and Book Publishing
Journal of Education
Volume 4||Issue 8||Page 20-31 ||November||2021|
Email: info@stratfordjournals.org ISSN: 2616-8383

Table 6. The Average Score of Each National Standard of Education and Each Cluster at Primary and Secondary School Level

	Cluster	Standard of Graduate Competence	Standard of Content	Standard of Process	Standard of Assessment	Standard of Quality of Personnel	Standard of Infrastructure Facilities	Standard of School Management	Standard of School Funding
Primary School	Cluster 1	6,3493	6,1906	6,7588	6,5178	4,1046	5,3537	6,1939	6,0733
	Cluster 2	6,0735	5,8063	6,5569	6,1309	3,7059	4,9503	5,9144	5,7593
	Cluster 3	5,9140	5,1199	6,0569	5,0944	3,7419	4,9889	5,0367	5,5026
	Cluster 4	1,2417	1,5724	2,0502	3,8261	3,8649	4,0504	2,6321	4,4399
Secondary School	Cluster 1	6,3821	6,1274	6,6488	3,9520	6,3531	5,2954	6,0894	5,9815
	Cluster 2	6,2631	6,1024	6,6178	2,4261	6,2735	5,2992	6,0202	5,7575
	Cluster 3	6,0421	5,1292	6,0480	3,2770	5,1905	5,3097	5,1116	5,4920
	Cluster 4	1,1934	1,8256	2,0021	3,2379	0,2530	5,2788	2,5099	4,3542

The next analysis was carried out to comparing the two methods, clustering with Fuzzy C-Means and classification being carried out by the Ministry of Education and Culture (MoEC). It is shown in Table 7 below.

Table 7. Comparation of Number of member using Fuzzy Means method and MoEC Classification

School Level	*Fuzzy-C Means*	MoEC Classification			
		M4	M3	M2	M1
Primary School	*Cluster 1*	63.477	0	0	0
	Cluster 2	44.602	78	0	0
	Cluster 3	23.006	8362	119	0
	Cluster 4	0	3	2643	4
Secondary School	*Cluster 1*	16.551	0	0	0
	Cluster 2	9.484	182	0	0
	Cluster 3	7.335	2.744	12	0
	Cluster 4	0	20	416	96

Based on Table 7, it can be seen that all members of Cluster 1 (the best status) using Fuzzy C-Means also been classified in the best group (M4). The majority of schools in cluster 2 and cluster 3 using the Fuzzy C-Means are classified in the M4 (the best group of MoEC classification).

Not much different from the comparison at the Primary School level, all members of Cluster 1 have also been classified as the best group in the MoEC classification (M4), but cluster 2 and cluster 3 in the Fuzzy C-Means grouping, the majority are still classified as the best group (M4) in MoEC classification.

https://doi.org/10.53819/81018102t2024

4.1 Conclusion and Recommendation

Based on the pre-processing data, imputation process on the sub-indicators of standard infrastructure facilities with the regression method can increase the average standard of infrastructure for both primary and secondary high school levels so that the standard value of infrastructure is better.

The basic problem of school quality in Indonesia is quality of human resources, indicator that have the lowest average score is the quality of teachers and staffs, and also the capability of teacher to make student assessment in learning process. Otherwise, the highest quality is the learning process. The score average of learning process mosly more than 5.

Cluster Analysis using Fuzzy C-Means gives a results that the optimum number of clusters in school clustering in Indonesia using the Fuzzy C-Means method are 4 clusters with the order that has the best quality, cluster 1 with 63,477 primary school and 16,551 secondary school members, cluster 2 with 44,680 primary school and 9,666 Secondary school members. Cluster 3 with 31,487 primary school and 10,091 secondary members, and cluster 4 with 2,650 primary school and 532 secondary members. There are 45% of primary school and secondary school in Indonesia, which are classified in cluster 1 (likely can fulfilled the national standard). Otherwise Indonesia still have 2% primary school and 1% of secondary school level having very low quality.

Table 8. Four Clusters with the Order That Has the Best Quality

Cluster	Number of School	
	Primary	Secondary
Cluster 1	63.477 (45%)	16.551 (45%)
Cluster 2	44.680 (31%)	9.666 (26%)
Cluster 3	31.487 (22%)	10.091 (27%)
Cluster 4	2.650 (2%)	532 (1%)

The classification using Fuzzy C-Means Cluster analysis seem giving consistence classification for cluster 1 and cluster 4 with M1 and M4 (based on MoEC classification). This is because the variability of school quality in this cluster is very low. For the cluster 2 and 3 are quite different with M2 and M3 of MoEC classification in term of number of school's member.

Indonesia government has to improve the school quality for about 55% school in Indonesia. Especially for schools in cluster 4 need to pay more attention to the quality of teacher and staff is still very low. Teacher exchange program can encourage improving the lowest quality besides training program for learning process improvement. The school in cluster 4 also need to improve management skill for the principle and management team and improvement in quality of infrastructure. Recommendation for further measurement is to improve the data quality especially to fulfil all sub indicators in all standards.

References

Alkareem, H. S., & Hossain, S. S. (2014). *Determinants of education quality: what makes student's perception different.* 52-67. https://doi.org/10.1080/23265507.2016.1155167

Bezdek, J., Ehrlich, R., & Full, W. (1984). Fuzzy C-Means Clustering Algorithm. *Compters & Geosciences Vol 10, no. 2-3*, 191-203. https://doi.org/10.1016/0098-3004 (84)90020-7

Chapman, C., & Sammons, P. (2013). *School Self-Evaluation for School Improvement: What Works and Why?*

Enders, C. (2010). *Applied Missing Data Analysis.* New York: The Guilford Press.

Ginns, P., Prosser, M., & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. *Studies in Higher Education, 32*(5), 603–615. https://doi.org/10.1080/03075070701573773

Gujarati, D. (2006). *Ekonometrika Dasar Diterjemahkan oleh Sumarto Zain.* Jakarta: Erlangga.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate Data Analysis. In *Vectors*. https://doi.org/10.1016/j.ijpharm.2011.02.019

Hofman, R. H., Dijkstra, N. J., & Hofman, W.H.A., (2009, March 10). *School self-evaluation and student achievement.* Retrieved from tandfonline.com: https://www.tandfonline.com/doi/full/10.1080/09243450802664115

Kaagan, S., & Smith, M. S. (1985). Indicators of Educational Quality. *Educational Leadership*, 21-24.

Ministry of Education and Culture. (2016). *Ministry of Education and Culture Regulation no 28 regarding the Quality Assurance System.* Jakarta.

National Inspectorate of Education. (2006). *Proportional supervision and school improvement from an international perspective. A study into the side effects of utilising school selfevaluations for inspection purposes in Europe*, Utrecht, The Netherlands: Inspectie van het onderwijs. Retrieved from tandfonline.com : https://www.tandfonline.com/doi/full/10.1080/09243450802664115

Nawari. (2010). *Analisis Regresi dengan MS Excel 2007 dan SPSS.* Jakarta: PT. Elex Media Komputindo.

Pyle, D. (1999). *Data Preparation for Data Mining.* USA: Academic Press.

Republic of Indonesia Government. (2013). *Government Regulation of the Republic of Indonesia No 32 of 2013 concerning Amendments to Government Regulation No 19 of 2005 concerning National Education Standards.* Jakarta.

Santoso, S. (2010). *Statistik Parametrik, Konsep dan Aplikasi dengan SPSS.* Jakarta: PT. Gramedia.

Schneider, J., Jacobsen, R., White, R., & Gehlbach, H. (2017). Building a better measure of school quality. In *Phi Delta Kappan* (Vol. 98, Issue 7). https://doi.org/10.1177/0031721717702631

UNESCO. (2019). *SDG 4 Data Digest: How to Produce and Use the Global and Thematic Education Indicators.* Canada: Institute for Statistics.