

## K-MEANS CLUSTERING FOR CLASSIFYING THE QUALITY MANAGEMENT OF SECONDARY EDUCATION IN INDONESIA

Zaenal Mustakim \*, Rahmat Kamal

Institut Agama Islam Negeri Pekalongan, Pekalongan, Indonesia

\*e-mail: zaenalmustakim@iainpekalongan.ac.id

**Abstract:** Mapping educational management is critical for improving national education quality. This study aims to classify Indonesian provinces in terms of secondary education management within quality standard parameters. Framed in a quantitative approach with the K-means cluster analysis, this study applied eleven parameters to classify the provinces into three main clusters. These parameters were schools, new entrants, students, repeaters, drop-outs, graduates, teachers, classrooms, laboratories, libraries, and school health services. Secondary data related to secondary education in Indonesia published by the Ministry of Education and Culture were analyzed. With the R software, the clusters were formed based on the shared characteristics of the provinces. This study found that Cluster 1 had twenty-two provinces, Cluster 2 two provinces, and Cluster 3 ten provinces. Cluster 2 was found to be the best cluster, while West Java and East Java shared similar characteristics, hence in the same cluster. Further, 5.88% of the provinces were eligible to be pilot models for the standard quality management of education in Indonesia. There is, therefore, a pressing need for the improvement of education infrastructure to support a better-quality education.

**Keywords:** educational management, education quality, K-means, clustering, senior secondary school

### KLUSTER *K-MEANS* UNTUK MENGLASIFIKASIKAN MANAJEMEN PENDIDIKAN MENENGAH BERKUALITAS DI INDONESIA

**Abstrak:** Pemetaan manajemen pendidikan sangat penting guna meningkatkan kualitas pendidikan nasional. Penelitian ini bertujuan untuk mengklasifikasikan provinsi-provinsi di Indonesia dalam hal manajemen pendidikan menengah berdasarkan parameter standar kualitas atau mutu. Dalam bingkai pendekatan kuantitatif dengan analisis kluster *K-means*, penelitian ini melibatkan sebelas parameter untuk mengklasifikasikan provinsi-provinsi tersebut menjadi tiga kluster utama. Parameter yang dimaksud adalah sekolah, pendatang baru, siswa, repeater, putus sekolah, lulusan, guru, ruang kelas, laboratorium, perpustakaan, dan pusat kesehatan sekolah. Data sekunder terkait dengan pendidikan menengah di Indonesia yang dipublikasikan oleh kementerian pendidikan dan kebudayaan dianalisis. Dengan bantuan *R software*, kluster-kluster dibentuk berdasarkan kemiripan karakteristik masing-masing provinsi. Hasil penelitian ini menunjukkan bahwa kluster 1 mencakup dua puluh dua provinsi, kluster 2 dua provinsi, dan kluster 3 sepuluh provinsi. Kluster 2 didapati sebagai cluster terbaik, sedangkan Jawa Barat dan Jawa Timur memiliki kemiripan karakteristik, sehingga berada di kluster sama. Selanjutnya, 5,88% provinsi memenuhi syarat untuk menjadi model percontohan dalam hal manajemen standar pendidikan yang berkualitas di Indonesia. Oleh karena itu, perbaikan infrastruktur pendidikan sangat dibutuhkan guna mendukung kualitas pendidikan yang lebih baik.

**Kata Kunci:** manajemen pendidikan, kualitas pendidikan, K-means, clustering, sekolah menengah atas

#### INTRODUCTION

The management of senior secondary school requires thorough attention from numerous sectors. Education at this school level, along with its well-established decisions, is crucial because it can prepare meaningful human resources for the community, prospective university students, or competition in the labor market (Hanushek et al., 2017; Xu & Trimble, 2016;

Yuguda Kotirde et al., 2015). Therefore, completing senior secondary school is a pivotal investment to compete in the job market. A study by Mahmudah, Suhartono, and Fatimah (2017) suggested that it is vital to investigate the quality of the senior secondary school in a closer look to foster educational management in Indonesia (Mahmudah et al., 2017). The concept of quality management becomes a critical factor in the

success and survival of organizations. A wide array of methods has also been carried out to escalate the quality management (Sadeghi Moghadam et al., 2018).

This quality management constitutes a fundamental aspect in the education field (Anttila & Jussila, 2017; Mahmudah et al., 2018) and has become a big concern for educational organizations (Mahmudah et al., 2018; Stukalina, 2010). Managing education quality is influenced by various factors, i.e. the increase in teachers and students' knowledge, skills, and abilities in quality management (Abdinazar, 2020; Bendermacher et al., 2017; Rahayuningsih et al., 2018; Töremen et al., 2009; Wong, 1999). The management of the education system and the enhancement of tools and technology are deemed urgent tasks in modern education (Anttila & Jussila, 2017). Quality management practices are demanded to maximize pedagogical potential and professionalism in the future (Omigie et al., 2019). Various models and methods of quality management have been documented in a study that employed systematic literature and text-mining analysis approaches (Sadeghi Moghadam et al., 2018).

One way to portray the quality of education is to classify it into several categories (Yamauchi, 2011). The categories are frequently referred to as clusters to view the profile and shared characteristics of the determined objects (Everitt & Hothorn, 2011; Härdle & Simar, 2013). Cluster profiling of senior secondary school management conducted by each province allows the government to better understand their performance in managing education based on quality indicators. Such profiling illustrates that the government can employ effective education management strategies to foster the nationally mandated quality of education. Given these demands, the present study provides a foundation based on empirical data to serve as a reference for constructing education policies so that the quality of Indonesian education can compete globally. It also has a direct implication for society, i.e. gaining a better quality education since both the provincial and the central government can manage education management in better further directions.

In creating the groups, cluster analysis is employed to serve as a powerful tool in articulating groups of objects that possess similar characteristics (Everitt & Hothorn, 2011; Härdle & Simar, 2013; Thrun, 2018). With this approach, the most similar objects tend to be put

in the same cluster. Further, the determination of the groups in cluster analysis adopts pairwise similarities (Härdle & Simar, 2013; Vörös & Snijders, 2017). In response to the cluster analysis, one of the prominent partitioning methods is K-means approach (Thrun, 2018). Here, the main concept in the clustering process is looking for a cluster center iteratively, which is determined based on the minimum distance of each data center on the cluster (Butarbutar et al., 2017; Nurzahputra et al., 2017; Windarto, 2017). Unlike other multivariate analyses estimating a set of variables, the purpose of cluster analysis is to classify objects based on those variables (Adachi, 2016; Everitt & Hothorn, 2011). The cluster formed tends to enact high internal homogeneity and high external heterogeneity (Everitt & Hothorn, 2011). Cluster analysis, likewise, can be utilized as a tool in organizing specific profiles (Hartanto et al., 2017; Tkaczynski, 2017).

Studies on the frame of cluster analysis for creating groups, classification, and particular profiles in the realm of education are burgeoning (Bojesen & Rayce, 2020; Kabók et al., 2017; Perrotta & Williamson, 2018). Prior research by (Darcan & Badur, 2012), for instance, employed cluster analysis to obtain student profiles based on different backgrounds in the management information systems department at Bogazici University, Turkey (Darcan & Badur, 2012). Cluster analysis was also applied to a group of European countries concerning the competitiveness of higher education (Kabók et al., 2017). Another research conducted by Dumuid, Olds, Martín-Fernández, Lewis, Cassidy, & Maher (2017) reported that the taxonomy of school children in Australia based on their lifestyle and behavior was also carried out by using cluster analysis. The indicators determined in the study are categorized into light, moderate, and strong physical activities; sedentary and sleep behavior derived from 24-hour accelerometry; self-reported screen time and diet.

In a further vein, the cluster analysis was adopted to classify senior secondary schools by previous researchers (Aoyama et al., 2011; Chabrol et al., 2015; Desmet et al., 2015; Ng et al., 2016). Papi & Teimouri's (2014) study applied cluster analysis to classify the motivational configuration of senior high school students in Iran when learning a second language. Cluster analysis was outlined by performing the K-means approach in distributing senior secondary school teachers in Indonesia (Widiyaningtyas et

al., 2017). Other studies also exposed the application of cluster analysis in mapping the quality of education in Indonesian school settings (Ananda, 2019; Nugraha & Hairani, 2018; Oktavianty et al., 2019; Prayoga & Zain, 2016; Wijayanto, 2016).

Unfortunately, there has not been much research that focuses on mapping and classifying Indonesian provinces in secondary school education management based on education quality standards. Given the critical role of this mapping for improving the quality of education on the national scale, this study primarily aims to classify all provinces in Indonesia in the forms of clusters based on similar characteristics of educational quality. It employs the K-means approach, one of the most widely used clustering methods, to obtain these clusters.

## METHODS

This study employed a quantitative approach and relied on secondary data, i.e. senior secondary school statistical data in the 2019-2020 academic period. The data were collected from documents released by the Ministry of Education and Culture, the Republic of Indonesia. The Secretariat of the Directorate General of Primary and Secondary Education and the Directorate of Senior High School Development collected the data through the Ministry of Education's data center, which had verified and validated the information before they were tabulated. This database is an integrated national scale data collection system and plays a role as the primary data source for national education in Indonesia (Ministry of Education and Culture, 2020). The data-informed from the Basic Education Data (*Data Pokok Pendidikan*) covered 34 provinces of Indonesia to unravel the implementation of secondary school education. Provincial classifications are enacted regarding national education quality standards. Eight indicators of the quality of Indonesian education were reported, i.e. standards of management, graduate competency, content, process, assessment, teacher and student, facilities and infrastructure, and cost. This research employed eleven variables referring to these standard rules: the number of schools, new entrants, pupils, repeaters, drop-outs, graduates, teachers, classrooms, laboratories, libraries, and school health centers (Ministry of Education and Culture, 2020).

This research employed a cluster analysis method by using K-Means approach, which is usually used to group a set of objects based on

similarity characteristics. The process of administering a provincial cluster was demonstrated by grouping data possessing the same parameters into one or more clusters. Conversely, data with distinct characteristics were grouped with other groups. This clustering method functions to detect observation groups equipped with homogeneous features and differentiate them from other groups (Everitt & Hothorn, 2011). Additionally, this study also employed the three-cluster solution. The similarities between objects were carried out by comparing pairs of observations  $(x_i, x_j)$ , where  $x_i^T = (x_{i1}, \dots, x_{ip})$ ,  $x_j^T = (x_{j1}, \dots, x_{jp})$ , and  $x_{ip}, x_{jp} \in \{0,1\}$  (Härdle & Simar, 2013).

To hone in cluster analysis, the following steps were taken (Everitt & Hothorn, 2011; Thrun, 2018):

*Step 1.* Selecting a set of variables used to group objects into clusters ( $C_k$ ).

*Step 2.* Initializing a random cluster center, called a centroid. This process was performed in the input room.

*Step 3.* Determining the data point nearest to each centroid to its point.

*Step 4.* Mapping the data points. The centroid was moved such that its distance from the setpoint to the corresponding centroid was minimized. The measure of the similarity matrix used the Euclidean distance (Härdle & Simar, 2013):

$$d_{ij} = \left\{ \sum_{k=1}^p |x_{ik} - x_{jk}| \right\}^{1/r}$$

$d_{ij}$  demonstrates the Euclidean distance for both  $i$  and  $j$  with the range of  $r \geq 1$ .  $x_{ik}$  shows the value of variable  $k^{th}$  on the object  $i$  while  $x_{jk}$  depicts the value of variable  $k^{th}$  on the object  $j$ . This equation also reveals the class of distance for the variation of  $r$  in the inequality measurement with unequal weights (Härdle & Simar, 2013).

*Step 5.* Finding out the closest cluster pairs of  $C_1, C_2, \dots, C_k$  performing the K-means method. For instance, the nearest pairs were  $C_i$  and  $C_j$ , then combined these clusters, removed  $C_j$  and reduced the number of clusters one by one.

*Step 6.* The process stopped when the number of clusters was equal to one. If it did not fulfill this condition, step 5 was repeated continuously until the number of clusters was equal to one.

**Table 1. KMO and Bartlett's Test**

Tests		Values
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.86
Bartlett's test of sphericity	Approx. Chi-square	1235.68
	df	45
	Sig.	.00

This study adopted these stages to cluster subgroups from all Indonesian provinces in managing senior secondary education practices concerning the quality of national education. The R program was used to analyze all of the results.

Furthermore, measuring the sample representation was critical before using a clustering analysis to cluster a group of provinces with similar characteristics. This study used the Kaiser-Meyer-Olkin (KMO) and Bartlett test to perform the test. Table 1 shows the results of the Kaiser-Meyer-Olkin (KMO) and Bartlett test.

As shown in Table 1, it can be noticed that the data obtained in this study met the assumptions required in the cluster analysis, i.e. the population representation test. Table 1 highlights that the KMO and Bartlett's test yielded a value higher than 0.5, namely 0.86. In addition to this attainment, Table 1 exemplifies that the test gained a significance value equal to 0.00, which was less than 0.05. It indicates that the clustering process using the K-means approach in these determined provinces could be further carried out.

## FINDINGS AND DISCUSSION

### Findings

This study establishes eleven indicators representing the quality of senior high schools in the 2019-2020 school year in the context of Indonesia. Informed by obtained secondary data, the following table portrays the descriptive indicators regarding the quality of Indonesian senior secondary schools.

Inspired by the statistical calculation, Table 2 demonstrates that the classification of 34 samples was derived from the total number of provinces in Indonesia. It is essential to notice that Table 2 employed eleven indicators administered in clustering the subgroups. Schools (X1) reveal the number of schools in the province stored in the database of the Ministry of Education and Culture, the Republic of Indonesia, both public and private senior secondary schools in each province. New entrants (X2) constitute the number of students who have just entered senior secondary schools, both public and

private. Pupils (X3) depict the number of students currently studying at school. Repeaters (X4) refer to the number of students who repeat in the province. Drop-outs (X5) portray the number of students who quit before graduating from senior high school. Teachers (X7) illustrate the total number of teachers in the province, including school principals. Classrooms (X8) showcase the total number of classrooms used in the teaching and learning process. Laboratories (X9) frame the total number of laboratories. Libraries (X10) inform the number of school libraries in each province. School health centers (X11) denote the number of health facilities in senior secondary schools located in the province.

Furthermore, Table 2 captures that the average number of senior high schools in the overall provinces of Indonesia which was 410 senior high schools. It was obtained that the minimum number of the schools was 61, whereas the maximum number was 1,663 schools from a total of 13,939 senior high schools throughout the country. North Kalimantan was the province with the lowest number of public and private senior high schools. Conversely, West Java was the province with the highest number of senior high schools (Ministry of Education and Culture, 2020). The statistical calculation also illustrates that the average number of students per year who had entered senior secondary school was 50,059. North Kalimantan with 5,902 new entrants reflected the province with the fewest new students, while West Java with 236,215 new entrants was considered as the province with the highest number of new students. It is important to note that the number of new entrants in Indonesian senior high schools in the 2019-2020 school year was 1,702,019 (Ministry of Education and Culture, 2020). The average number of pupils (X3) in Indonesia was 146,357, in which the smallest number of students was 17,386, and the biggest number of students was 693,450. Respectively, North Kalimantan and West Java represented provinces with the smallest and biggest senior high school students (Ministry of Education and Culture, 2020).

Table 2 also showcases that the average number of repeaters was 500 each year. Mean-

while, the fewest and biggest numbers of repeaters were 48 and 1,563 respectively. Bali and East Java were identified as the provinces with the fewest and most repeaters, respectively. The total number of repeaters in Indonesian senior secondary school was 17,017 students in the 2019-2020 school year (Ministry of Education and Culture, 2020). Table 2, furthermore, shows that the average number of senior high school drop-outs during the examined school year was 790. Bangka Belitung and South Sulawesi represented the provinces with the lowest and highest number of drop-outs from senior secondary school; there were 31 and 2,595 drop-outs, respectively. In this school year, 26,864 Indonesian students quit senior secondary education. As depicted in Table 2, in the same school year, it is also obtained that the average number of senior high school graduates were 45,464, where the lowest and highest numbers of high school graduates were 5,132 and 223,434, respectively. North Kalimantan and West Java were the provinces with the lowest and highest number of graduates of senior high schools. It was also reported that there were 1,545,784 senior high school graduates in Indonesia during the 2019-2020 school year (Ministry of Education and Culture, 2020).

Table 2, likewise, provides a summary that the average number of Indonesian senior secondary school teachers, including principals, was 9,468 during the 2019-2020 school year. North Kalimantan with 1,334 teachers and West Java with 37,314 teachers respectively came to the surface as the provinces with the lowest and highest numbers of teachers. The calculation also shows that the total number of teachers and school principals in Indonesia in the school year comprised 321,914 (Ministry of Education and Culture, 2020). As shown in table 2, it is portrayed that the average number of classrooms in senior secondary schools in Indonesia was 5,215. The minimum and maximum numbers of the classrooms were 623 and 23,999, respectively. North Kalimantan and West Java were reported as the provinces with the smallest and the biggest numbers of classrooms for each. The validated total number of senior high school classrooms in Indonesia in the 2019-2020 academic year was 177,323 (Ministry of Education and Culture, 2020).

Table 2 informs that the average number of laboratories situated in Indonesian senior secondary school was 1,200. North Kalimantan and West Java were also found out to be the

lowest and highest numbers of laboratories; they had 126 and 5,048 laboratories respectively. Meanwhile, the average number of libraries in senior secondary school was 398. North Kalimantan and West Java were also measured as the provinces with the smallest and biggest numbers of libraries. They respectively had 53 and 1,611 libraries. Lastly, the average number of health centers in Indonesian senior secondary schools was 232. It is revealed that the North Kalimantan with 29 school health centers was classified as the province with the lowest number of school health centers. In contrast, the East Java province with 1,033 school health centers was considered as the province with the highest number of school health centers (Ministry of Education and Culture, 2020). Inspired by these statistical calculations, the study finding reported that West Java and North Kalimantan provinces demonstrated distinct characteristics to be grouped in different clusters.

Table 3 reveals the results of provincial clustering in Indonesia regarding education quality indicators using the K-means approach and a 3-cluster solution. Table 3 shows the history of the iteration process that occurred to get the right cluster.

Table 3 illustrates that there were three iteration processes required to obtain a convergent cluster solution. Then, it can also be viewed that the minimum distance between flashlights was 5.75. Further, as revealed in Table 4, the Analysis of Variance (ANOVA) results of the clustering process performed variables mainly contributing to the gained cluster solution.

Next, Table 4 yields that the Classrooms (X8) variable has the most significant contribution to the obtained 3-cluster solutions, where the *F* value results in the greatest one; it was 95.84. Then, the smallest *F* value is in the Repeaters variable (X4). Meanwhile, Table 5 highlights the final results of the clustering process employing a 3-cluster solution. The last cluster center was calculated as the mean for each variable in every final cluster. This cluster reflects typical case characteristics for the respective cluster.

As reported in Table 5, Cluster 1 consisted of the provinces with a low number of overall indicators as all values demonstrated negative numbers. A negative value denotes that the data are below the total average. Conversely, a positive value means that the data are above the total average. Table 5 also portrayed that Cluster 2 comprises the provinces that had a high number of indicators. Meanwhile, Cluster 3 contains

provinces with relatively moderate overall indicators. The cluster membership of each province was spotted in Table 6. Table 6 also illuminates the Euclidean distances indicating similar characteristics of each province.

Table 6 shows that three clusters were established in this study, i.e. Cluster 1 with 22 provinces, Cluster 2 consisting of two provinces, and Cluster 3 comprising 10 provinces.

**Table 2. Descriptive Statistics of Indicators (N=34)**

Indicators	N	Minimum	Maximum	Mean	Std. Deviation
Schools (X <sub>1</sub> )	34	61	1,663	410	381
New Entrants (X <sub>2</sub> )	34	5,902	236,215	50,059	52,113
Pupils (X <sub>3</sub> )	34	17,386	693,450	146,357	152,208
Repeaters (X <sub>4</sub> )	34	48	1,563	500	406
Drop-outs (X <sub>5</sub> )	34	31	2,595	790	716
Graduates (X <sub>6</sub> )	34	5,132	223,434	45,464	49,135
Teachers (X <sub>7</sub> )	34	1,334	37,314	9,468	8,475
Classrooms (X <sub>8</sub> )	34	623	23,999	5,215	5,267
Laboratories (X <sub>9</sub> )	34	126	5,048	1,200	1,231
Libraries (X <sub>10</sub> )	34	53	1,611	398	373
School Health Centers (X <sub>11</sub> )	34	29	1,033	232	257

**Table 3. Iteration History**

Iteration	Change in Cluster Centers		
	Cluster 1	Cluster 2	Cluster 3
1	1.37	1.48	1.61
2	0.14	0.00	0.31
3	0.00	0.00	0.00

Note: Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 3. The minimum distance between initial centers is 5.749.

**Table 4. ANOVA Statistical Calculations**

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Schools (X <sub>1</sub> )	14.10	2	0.15	31	91.13	0.00
New Entrants (X <sub>2</sub> )	13.88	2	0.17	31	82.12	0.00
Pupils (X <sub>3</sub> )	13.93	2	0.17	31	84.00	0.00
Repeaters (X <sub>4</sub> )	6.28	2	0.66	31	9.52	0.00
Drop-outs (X <sub>5</sub> )	7.46	2	0.58	31	12.80	0.00
Graduates (X <sub>6</sub> )	13.99	2	0.16	31	86.31	0.00
Teachers (X <sub>7</sub> )	14.10	2	0.15	31	91.22	0.00
Classrooms (X <sub>8</sub> )	14.20	2	0.15	31	95.84	0.00
Laboratories (X <sub>9</sub> )	13.44	2	0.20	31	67.95	0.00
Libraries (X <sub>10</sub> )	14.15	2	0.15	31	93.46	0.00
School Health Centers (X <sub>11</sub> )	13.49	2	0.19	31	69.39	0.00

**Table 5. Final Cluster Centers**

ZScore	Cluster		
	Cluster 1	Cluster 2	Cluster 3
ZScore: Schools (X <sub>1</sub> )	-0.52	3.13	0.52
ZScore: New entrants (X <sub>2</sub> )	-0.52	3.09	0.52
ZScore: Pupils (X <sub>3</sub> )	-0.52	3.10	0.52
ZScore: Repeaters (X <sub>4</sub> )	-0.41	1.67	0.57
ZScore: Drop-outs (X <sub>5</sub> )	-0.47	1.54	0.73
ZScore: Graduates (X <sub>6</sub> )	-0.52	3.13	0.51
ZScore: Teachers (X <sub>7</sub> )	-0.55	2.96	0.63
ZScore: Classrooms (X <sub>8</sub> )	-0.53	3.12	0.54
ZScore: Laboratories (X <sub>9</sub> )	-0.52	2.98	0.56
ZScore: Libraries (X <sub>10</sub> )	-0.53	3.09	0.55
ZScore: School Health Centers (X <sub>11</sub> )	-0.49	3.12	0.47

**Table 6. Cluster Membership of Indonesian Senior Secondary Schools**

Case Number	Provinces	Cluster	Distance
1	Jakarta	3	2.23
2	West Java	2	1.48
3	Banten	3	1.58
4	Central Java	3	3.75
5	Yogyakarta	1	0.71
6	East Java	2	1.48
7	Aceh	3	2.16
8	North Sumatera	3	3.36
9	Sumatera Barat	3	2.18
10	Riau	3	1.51
11	Riau Islands	1	0.81
12	Jambi	1	0.36
13	South Sumatera	3	0.86
14	Bangka Belitung	1	1.22
15	Bengkulu	1	0.72
16	Lampung	1	2.02
17	West Kalimantan	1	2.04
18	Central Kalimantan	1	0.54
19	South Kalimantan	1	0.16
20	East Kalimantan	1	0.27
21	North Kalimantan	1	1.35
22	North Sulawesi	1	0.58
23	Gorontalo	1	0.93
24	Central Sulawesi	1	0.21
25	South Sulawesi	3	1.85
26	West Sulawesi	1	0.98
27	Southeast Sulawesi	1	0.80
28	Maluku	1	0.81
29	North Maluku	1	0.68
30	Bali	1	0.97
31	West Nusa Tenggara	1	2.00
32	East Nusa Tenggara	3	1.40
33	Papua	1	2.58
34	West Papua	1	0.81

### Discussion

This study adopted cluster analysis with the K-means approach to classifying 34 provinces in managing the education quality of Indonesian senior secondary schools. Framed in the 3-cluster solution, the study determined three categories, i.e. high, medium, and low (Butarbutar et al., 2017; Windarto, 2017). Nevertheless, this method of analysis requires certain data assumptions to provide precise and reliable clusters. Inspired by Table 1, the study enacted the KMO and Bartlett test in examining the statistical assumptions required in cluster analysis. The results of the assumption test indicate that the data used were able to represent the population, so it was deemed effective to consider the data in this analysis. It was attained by the KMO and Bartlett values of 0.86 and a significance value of 0.00. Overall, Table 1 spotlights that the variables and data could be used in the cluster analysis process performing

the K-means approach. Consequently, the next stage was further analyzed, e.g. investigating provinces that possessed similar characteristics regarding the predetermined indicators (Thrun, 2018). It then was continued to determine the cluster membership of each province. Therefore, each cluster formed consisted of provinces showing adjacent characteristics determined by the Euclidean distance (Adachi, 2016; Everitt & Hothorn, 2011; Thrun, 2018).

There were changes in each step in the clustering process with a 3-cluster solution, as depicted in Table 3. The changes in the cluster centers suggest the progress of the clustering process at every phase (Everitt & Hothorn, 2011; Windarto, 2017). Table 3 also elucidates that there was a significant shift in the center of the cluster in the early stages of the iteration process. Likewise, in the third iteration, it was obvious to note that this stage fulfilled the convergence criteria. The criteria are reached when there is no

change or minor change in the iteration step (Adachi, 2016). For this reason, the analysis findings demonstrate that it required three iteration processes to create a convergent cluster. This result also documents that the iteration processes continued until the data in the last iteration are the same as those in the previous process (Windarto, 2017). In response to this, the iteration processes were conducted on 34 objects, all provinces in Indonesia, which were classified into small clusters capturing similar characteristics.

In a further vein, it is essential to note that cluster analysis explores how much the contribution of each variable is employed in the formation of the clusters. Table 4 provides detailed information concerning the contribution of these variables. The amount of contribution is usually viewed from the F value and the significance of each variable. Informed by Table 4, the analysis results illuminate that the variables of the number of classrooms and the number of Repeaters were reported to have the highest and lowest contribution in forming clusters. This can be noticed from the F values of the two variables, i.e. 95.84 and 9.52, respectively. The attainment was also highlighted to be the biggest and the smallest value among the F values of other variables. Table 4 also depicts that the higher the F value in a variable, the more significant the difference in the variable situated in the formed cluster. Thus, the F value in the ANOVA table in the cluster analysis indicates the magnitude of the contribution of the variable to the cluster solution, which is directly gearing proportional to the mean square value.

The statistical calculation may speak to that the number of classrooms facilitated by Indonesian senior secondary schools in a province constituted the variable that most indicates the difference between the provinces in the three formed clusters. In contrast, repeaters of senior high schools were reported to reveal the smallest differences. In this respect, the provinces with the highest number of classrooms were projected not to be situated in the same cluster as the provinces with the lowest number of classrooms. This notion implies that these provinces did not have similar characteristics forming the basis for cluster membership establishment (Everitt & Hothorn, 2011; Härdle & Simar, 2013; Martinez et al., 2004; Thrun, 2018). In other words, the classification for taxonomies was acknowledged in relation to the Euclidean distance-based measure (Butarbutar et al., 2017; Windarto, 2017). This notion is evidenced by the

cluster membership in Table 6 demonstrating that North Kalimantan and West Java were not categorized in the same cluster. It was documented that both provinces represented the lowest and highest number of classrooms, respectively.

Table 6 also suggests that Bali and East Java provinces did not emerge from the same cluster since they performed different characteristics regarding the number of repeaters. Table 1 portrays that the two provinces exposed the lowest and the highest number of repeaters. Table 4 reveals that the second-biggest contribution was the number of libraries, whereas the second-lowest was the number of drop-outs, in which the F values of each variable were 93.46 and 12.80 respectively. The results can also be expected that the provinces possessing the most number of libraries were not categorized in the same cluster as those with the most number of drop-outs. Table 6, additionally, confirms that West Java and Bangka Belitung were not in the same cluster. The two provinces were also reported to have the largest number of libraries and the highest number of drop-outs, respectively.

Documented in Table 6, the analysis results succeeded in forming three clearly identified clusters, i.e. Cluster 1 consisting of 22 provinces, Cluster 2 with two provinces, and Cluster 3 comprising 10 provinces. The taxonomy of clusters is based on the similarity of each province's characteristics, noticed from the minimum distance (Butarbutar et al., 2017; Everitt & Hothorn, 2011; Windarto, 2017). The results elucidate that the provinces in one cluster were documented to have better similarities than provinces in other clusters when all indicators were considered and have the fewest minimum distance.

Informed by the result of cluster analysis, 22 provinces were classified in Cluster 1, i.e. Special Region of Yogyakarta, Riau Islands, Jambi, Bangka Belitung, Bengkulu, Lampung, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Gorontalo, Central Sulawesi, West Sulawesi, Southeast Sulawesi, Maluku, North Maluku, Bali, West Nusa Tenggara, Papua, dan West Papua. Two provinces were yielded in Cluster 2, i.e. West Java and East Java. Meanwhile, 10 provinces were categorized in Cluster 3, i.e. Banten, Central Java, Aceh, North Sumatera, West Sumatera, Riau, South Sumatera, South Sulawesi, and West Nusa Tenggara.

Table 6, furthermore, reported that West Java and North Kalimantan were shown to have



the largest diversity in characteristics in terms of the determining variables (Ministry of Education and Culture, 2020). Thus, it is no wonder that the two provinces were situated in different clusters as demonstrated in Table 6. West Java was classified in Cluster 2 while North Kalimantan in Cluster 1.

Captured in Table 5, distance illuminates the similarities in the characteristics of the objects in the cluster (Erdoğan & Esen, 2016; Everitt & Hothorn, 2011; Thrun, 2018; Windarto, 2017). Table 5 also indicates that the distance from West Java and East Java provinces was the same, namely 1.48. This attainment shows the similar characteristics of the two provinces when all parameters of national education quality standards are taken into account. Therefore, the two provinces should be classified into the same group. This result is consistent with the previous studies pointing out that objects that had similar characteristics needed to be categorized in the same cluster (Dumuid et al., 2017; Kabók et al., 2017; Nurzahputra et al., 2017; Sadeghi Moghadam et al., 2018).

Table 5 elucidates that the provinces in Cluster 1 performed an average number fewer than the total value for all parameters. This can be noted from all indicator values that have a negative sign. Meanwhile, the provinces classified in Cluster 3 yielded moderate scores. Cluster 1 comprises the provinces with a high average number of all indicators because all values showed high numbers compared to other clusters, then, all values were also considered positive. It implies that the provinces in Cluster 2 represented the best high school quality standards compared to the provinces in both Cluster 1 and Cluster 3. West Java and East Java provinces were identified to have quite a lot of schools compared to other provinces, accounting for 1,663 and 1,542, respectively. Besides, the two provinces also had a large number of new students, with a total of 236,215 and 185,897 students respectively. Regarding the number of students, the two provinces also had large numbers, namely 693,450 and 544,274, for each. Lastly, the number of libraries and the number of laboratories also showed the same characteristics (Ministry of Education and Culture, 2020).

In a comprehensive direction, the clustering process enables us to map the quality of education (Ananda, 2019; Nugraha & Hairani, 2018; Oktavianty & Handayani, 2019; Subanar, 2011). In response to this, every province's management of senior secondary schools is

pivotal for in-depth analysis. The classification of the clusters functions to create a profile in mapping provinces involved in the main cluster, provinces documented in the medium cluster, and those categorized as the last clusters. The existence of this map allows education offices or educational institutions in Indonesia to compare the quality of education in each province or district (Nugraha & Hairani, 2018). In addition to the classification, the main cluster can serve as a pilot model for other provinces in managing senior secondary education to foster the betterment of education quality. It, for this reason, is crucial to make improvements in the provision of educational facilities to support the quality of senior high school education in Indonesia (Prayoga & Zain, 2016).

Promoting the quality of senior high school education may also directly impact the input of qualified prospective students in tertiary institutions and human resources participating in society. Besides, the business world has also enjoyed positive impacts, namely improving the quality of job seekers with high school graduates. It suggests that the provinces classified in the main cluster could become motors in escalating the quality of education in Indonesia thoroughly. Thus, to perform better education quality standards, provinces that were not in Cluster 2 can provide several indicators used as the basis for the quality of Indonesia's education. Thus, these provinces are worthy of being employed as a model in increasing the quality of education standards for other provinces in other clusters, especially Cluster 3. In short, the government can compare the quality of education of each province to make appropriate policies because of the outlined profile map.

## CONCLUSION

This present study was designed to classify Indonesian provinces in terms of secondary education management based on quality standards parameters by employing the K-means clustering approach. The analysis results of this study show that, in general, Indonesian provinces were classified into three clusters concerning the similarities of each province. Cluster 1 included twenty-two provinces, Cluster 2 two provinces, and Cluster 3 ten provinces. The analysis also confirmed that Cluster 2 was in the high category, Cluster 3 in the medium category, and Cluster 1 in the low one in terms of educational management. The two provinces in Cluster 2, namely West Java and East Java, should serve as role

models in the management of senior secondary education quality in Indonesia because they are capable of managing the available resources to provide the highest possible quality of education. Other provinces, therefore, should learn from and reflect on the two said provinces' approaches to managing educational quality to achieve optimal outcomes. This way, the quality of national education can be improved in a systematic and structured manner.

More in-depth studies are further recommended to examine the inclusion of other indicators of senior secondary education quality. In addition, diverse taxonomy methods can also be applied to compare the final results of the cluster analysis. Likewise, other educational parameters, including those in higher education, also need to be classified deeming the quality standards of national education.

#### ACKNOWLEDGMENTS

The author would like to thank the Ministry of Education and Culture, the Republic of Indonesia, for allowing us to obtain valid and reliable data for this research.

#### REFERENCES

- Abdinazar, N. (2020). Managing the quality of training of pedagogical personnel's on the basis of TQM - (Total Quality Management). *International Journal of Psychosocial Rehabilitation*, 24(5), 3759–3770. <https://doi.org/10.37200/IJPR/V24I5/PR202085>
- Adachi, K. (2016). *Matrix-based introduction to multivariate data analysis*. Springer.
- Ananda, R. (2019). Analisis mutu pendidikan sekolah menengah atas program ilmu alam di Jawa Tengah dengan algoritme K-means terorganisir. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 2(1), 65–72. <https://doi.org/10.20895/inista.v2i1.97>
- Anttila, J., & Jussila, K. (2017). Understanding quality – conceptualization of the fundamental concepts of quality. *International Journal of Quality and Service Sciences*, 9(3–4), 251–268. <https://doi.org/10.1108/IJQSS-03-2017-0020>
- Aoyama, I., Barnard-Brak, L., & Talbert, T. L. (2011). Cyberbullying among high school students: Cluster analysis of sex and age differences and the level of parental monitoring. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)*, 1(1), 25–35. <https://doi.org/10.4018/ijcbpl.2011010103>
- Bendermacher, G. W. G., oude Egbrink, M. G. A., Wolfhagen, I. H. A. P., & Dolmans, D. H. J. M. (2017). Unravelling quality culture in higher education: a realist review. *Higher Education*, 73(1), 39–60. <https://doi.org/10.1007/s10734-015-9979-2>
- Bojesen, A. B., & Rayce, S. B. (2020). Effectiveness of a school-based road safety educational program for lower secondary school students in Denmark: A cluster-randomized controlled trial. *Accident Analysis and Prevention*, 147, 105773. <https://doi.org/10.1016/j.aap.2020.105773>
- Butarbutar, N., Windarto, A. P., Hartama, D., & Solikhun, S. (2017). Komparasi kinerja algoritma fuzzy c-means dan k-means dalam pengelompokan data siswa berdasarkan prestasi nilai akademik siswa. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 1(1), 46. <https://doi.org/10.30645/jurasik.v1i1.8>
- Chabrol, H., Melioli, T., Van Leeuwen, N., Rodgers, R., & Goutaudier, N. (2015). The Dark Tetrad: Identifying personality profiles in high-school students. *Personality and Individual Differences*, 83, 97–101. <https://doi.org/10.1016/j.paid.2015.03.051>
- Darcan, O. N., & Badur, B. Y. (2012). Student profiling on academic performance using cluster analysis. *Journal of E-Learning & Higher Education*, e1-8. <https://doi.org/10.5171/2012.622480>
- Desmet, A., Aelterman, N., Bastiaensens, S., Van Cleemput, K., Poels, K., Vandebosch, H., Cardon, G., & De Bourdeaudhuij, I. (2015). Secondary school educators' perceptions and practices in handling cyberbullying among adolescents: A cluster analysis. *Computers and Education*, 88, 192–201. <https://doi.org/10.1016/j.compedu.2015.05.006>
- Dumuid, D., Olds, T., Martín-Fernández, J.-A., Lewis, L. K., Cassidy, L., & Maher, C.

- (2017). Academic performance and lifestyle behaviors in Australian School Children: A cluster analysis. *Health Education & Behavior*, 44(6), 918–927. <https://doi.org/10.1177/1090198117699508>
- Erdoğan, N., & Esen, M. (2016). Classifying universities in Turkey by hierarchical cluster analysis. *Education & Science/Eğitim ve Bilim*, 41(184). <https://doi.org/10.15390/EB.2016.6232>
- Everitt, B., & Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. Springer Science & Business Media.
- Hanushek, E. A., Schwerdt, G., Woessmann, L., & Zhang, L. (2017). General education, vocational education, and labor-market outcomes over the lifecycle. *Journal of Human Resources*, 52(1), 48–87. <https://doi.org/10.3368/jhr.52.1.0415-7074R>
- Härdle, W. K., & Simar, L. (2013). Applied multivariate statistical analysis. In *Applied Multivariate Statistical Analysis*. <https://doi.org/10.1007/978-3-642-17229-8>
- Hartanto, Y. I., Rusgiyono, A., & Wuryandari, T. (2017). Penerapan analisis kluster metode ward terhadap kabupaten/kota di Jawa Tengah berdasarkan pengguna alat kontrasepsi. *Jurnal Gaussian*, 6(4), 528–537. <https://doi.org/10.14710/j.gauss.v6i4.30387>
- Kabók, J., Radišić, S., & Kuzmanović, B. (2017). Cluster analysis of higher-education competitiveness in selected European countries. *Economic Research-Ekonomska Istrazivanja*, 30(1), 845–857. <https://doi.org/10.1080/1331677X.2017.1305783>
- Mahmudah, U., Suhartono, & Fatimah, S. (2017). A robust approach to analyzing the factors influencing quality education in Indonesia. *New Educational Review*, 49(3), 77–90. <https://doi.org/10.15804/ner.2017.49.3.06>
- Mahmudah, U., Suhartono, S., & Rohayana, A. D. (2018). A robust data envelopment analysis for evaluating technical efficiency of Indonesian high schools. *Jurnal Pendidikan IPA Indonesia*, 7(1), 114–121. <https://doi.org/10.15294/jpii.v7i1.9883>
- Martinez, P. R., Molnar, J., Trejos, E., Meyer, D., Meyer, S. T., & Tollner, W. (2004). Cluster membership as a competitive advantage in aquacultural development: Case study of tilapia producers in Olancho, Honduras. *Aquaculture Economics & Management*, 8(5–6), 281–294. <https://doi.org/10.1080/13657300409380370>
- Ministry of Education and Culture, I. (2020). Senior high school in statistics. In *Ministry of Education and culture* (Vol. 53, Issue 9).
- Ng, B. L. L., Liu, W. C., & Wang, J. C. K. (2016). Student motivation and learning in mathematics and science: A cluster analysis. *International Journal of Science and Mathematics Education*, 14(7), 1359–1376. <https://doi.org/10.1007/s10763-015-9654-1>
- Nugraha, G. S., & Hairani, H. (2018). Aplikasi pemetaan kualitas pendidikan di Indonesia menggunakan metode k-means. *Jurnal MATRIK*, 17(2), 13–23. <https://doi.org/10.30812/matrik.v17i2.84>
- Nurzahputra, A., Muslim, M. A., & Khusniati, M. (2017). Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa. *Techno. Com*, 16(1), 17–24. <https://doi.org/10.33633/tc.v16i1.1284>
- Oktavianty, E., Junaidi, & Handayani, L. (2019). Pengelompokan kabupaten/kota di Sulawesi berdasarkan indikator pendidikan menggunakan analisis kluster average linkage dan median linkage. *Natural Science: Journal of Science and Technology*, 8(3), 191–197. <https://doi.org/10.22487/25411969.2019.v8.i3.14960>
- Omigie, C. A., Ikenwe, I. J., & Idhalama, O. U. (2019). The role of knowledge management for education in Nigeria. *International Multidisciplinary Research Journal*, 20–23. <https://doi.org/10.25081/imrj.2019.v9.5496>
- Papi, M., & Teimouri, Y. (2014). Language

- learner motivational types: A cluster analysis study. *Language Learning*, 64(3), 493–525.  
<https://doi.org/10.1111/lang.12065>
- Perrotta, C., & Williamson, B. (2018). The social life of Learning Analytics: cluster analysis and the ‘performance’ of algorithmic education. *Learning, Media and Technology*, 43(1), 3–16.  
<https://doi.org/10.1080/17439884.2016.1182927>
- Prayoga, A., & Zain, I. (2016). Analisis faktor dan pengelompokan kecamatan berdasarkan indikator mutu pendidikan jenjang pendidikan dasar di Kabupaten Sidoarjo. *Jurnal Sains Dan Seni ITS*, 4(2).  
<https://doi.org/10.12962/j23373520.v4i2.10920>
- Rahayuningsih, R. S., Fajaruddin, S., & Manggalasari, L. C. (2018). The implementation of total quality management in vocational high schools. *Psychology, Evaluation, and Technology in Educational Research*, 1(1), 31–40.  
<https://doi.org/10.33292/petier.v1i1.20>
- Sadeghi Moghadam, M. R., Safari, H., & Yousefi, N. (2018). Clustering quality management models and methods: systematic literature review and text-mining analysis approach. *Total Quality Management and Business Excellence*.  
<https://doi.org/10.1080/14783363.2018.1540927>
- Stukalina, Y. (2010). Using quality management procedures in education: Managing the learner-centered educational environment. *Technological and Economic Development of Economy*, 16(1), 75–93.  
<https://doi.org/10.3846/tede.2010.05>
- Subanar, H. (2011). *Analisis kluster untuk pemetaan mutu pendidikan di Aceh*. [Yogyakarta]: Universitas Gadjah Mada.
- Thrun, M. C. (2018). Projection-based clustering through self-organization and swarm intelligence. In *Projection-Based Clustering through Self-Organization and Swarm Intelligence*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-20540-9>
- Tkaczynski, A. (2017). Segmentation using two-step cluster analysis. In *Segmentation in social marketing* (pp. 109–125). Springer. [https://doi.org/https://doi.org/10.1007/978-981-10-1835-0\\_8](https://doi.org/https://doi.org/10.1007/978-981-10-1835-0_8)
- Töremen, F., Karakuş, M., & Yasan, T. (2009). Total quality management practices in Turkish primary schools. *Quality Assurance in Education*, 17(1), 30–44.  
<https://doi.org/10.1108/09684880910929917>
- Vörös, A., & Snijders, T. A. B. (2017). Cluster analysis of multiplex networks: Defining composite network measures. *Social Networks*, 49, 93–112.  
<https://doi.org/10.1016/j.socnet.2017.01.002>
- Widiyaningtyas, T., Prabowo, M. I. W., & Pratama, M. A. M. (2017). Implementation of K-means clustering method to distribution of high school teachers. *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 1–6.
- Wijayanto, F. (2016). Clustering analysis on Indonesian education quality performance using input–output model. *Advanced Science Letters*, 22(10), 2799–2803.  
<https://doi.org/10.1166/asl.2016.7127>
- Windarto, A. P. (2017). Penerapan datamining pada ekspor buah-buahan menurut negara tujuan menggunakan k-means clustering method. *Techno.Com*, 16(4), 348–357.  
<https://doi.org/10.33633/tc.v16i4.1447>
- Wong, A. (1999). Total quality management in the construction industry in Hong Kong: A supply chain management perspective. *Total Quality Management*, 10(2), 199–208.  
<https://doi.org/10.1080/0954412997956>
- Xu, D., & Trimble, M. (2016). What about certificates? Evidence on the labor market returns to nondegree community college awards in two states. *Educational Evaluation and Policy Analysis*, 38(2), 272–292.  
<https://doi.org/10.3102/0162373715617827>
- Yamauchi, F. (2011). School quality, clustering and government subsidy in post-apartheid South Africa. *Economics of Education Review*, 30(1), 146–156.  
<https://doi.org/10.1016/j.econedurev.2010.08.002>

Yuguda Kotirde, I., Bin, J., & Yunos, M.  
(2015). ScienceDirect-NC-ND license  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>). Peer-review under  
responsibility of Faculty of Technical and  
Vocational Education The Processes of

Supervisions in Secondary Schools  
Educational System in Nigeria. *Procedia -  
Social and Behavioral Sciences*, 204,  
259–264.  
<https://doi.org/10.1016/j.sbspro.2015.08.149>