



Mapping of Education Quality and E-Learning Readiness to Enhance Economic Growth in Indonesia

Setia PRAMANA¹, Erni Tri ASTUTI²

Received: August 20, 2021. Revised: November 21, 2021. Accepted: December 30, 2021.

Abstract

Purpose: This study is aimed to map the provinces in Indonesia based on the education and ICT indicators using several unsupervised learning algorithms. **Research design, data, and methodology:** The education and ICT indicators such as student-teacher ratio, illiteracy rate, net enrolment ratio, internet access, computer ownership, are used. Several approaches to get deeper understanding on provincial strength and weakness based on these indicators are implemented. The approaches are Ensemble K-Mean and Fuzzy C Means clustering. **Results:** There are at least three clusters observed in Indonesia the education quality, participation, facilities and ICT Access. Cluster with high education quality and ICT access are consist of DKI Jakarta, Yogyakarta, Riau Islands, East Kalimantan and Bali. These provinces show rapid economic growth. Meanwhile the other cluster consisting of six provinces (NTT, West Kalimantan, Central Sulawesi, West Sulawesi, North Maluku, and Papua) are the cluster with lower education quality and ICT development which impact their economic growth. **Conclusions:** The provinces in Indonesia are clustered into three group based on the education attainment and ICT indicators. Some provinces can directly implement e-learning; however, more provinces need to improve the education quality and facilities as well as the ICT infrastructure before implementing the e-learning.

Keywords : clustering, education, ICT, economic growth in Indonesia.

JEL Classification Code: H52, I25, D83

1. Introduction

Ensuring inclusive and equitable quality education and promoting lifelong learning opportunities for all is goal number 4 of the Sustainable Development Goals defined by the United Nations to be achieved by 2030. The goal has seven targets, such as primary and secondary education and Gender equality and inclusion. Achieving quality education is also one of the main priorities in the national development plan of Indonesia. As one of the major economies in Southeast Asia, Indonesia, through its government, focuses on its social and economic development. Hence, the development of human capital is very crucial. Moreover, the Indonesian government has recently shifted the

development focus from infrastructure to improving the quality of human resources. However, Program for International Student Assessment (PISA) of the Organization for Economic Co-operation and Development's (OECD) shows that the Indonesian students performs three years behind as compared to the OECD average (OECD, 2019). Furthermore, United Nations Development Program (UNDP) reports that Indonesia's Human Development Index (HDI) for 2020 was ranked 107th out of 189 countries (UNDP, 2021).

The Indonesian government has placed the improvement of the quality of basic education as the top development priority. It is implemented by spending 20% of the government expenditure for the educational development.

1. First and Corresponding Author, Associate Professor, Computational Statistics Department, Politeknik Statistika STIS, Indonesia. Email: setia.pramana@stis.ac.id
 2. Associate Professor, Statistics Department, Politeknik Statistika STIS, Indonesia. Email: erni@stis.ac.id

© Copyright: The Author(s)
 This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

In addition, the OECD has suggested that educational development in Indonesia can be focused on quality, participation, and efficiency. As the regions in Indonesia are so many and diverse, understanding the strengths and weaknesses of every region are important for efficiency of the education development program for providing the right policy for the right district/region.

Rapid technology development has led the education system to adopt the ICT technology to improve the quality of education attainments. The e-learning system nowadays has become the main requirement for educational institutions to support their learning activities. However, to adopt e-learning, a large strategy from all stake holders (governments, industries, private sectors, and communities) and resources (e.g., Internet connection, electricity, etc.) for optimal application are required. Moreover, a study aimed to evaluate e-learning readiness of the schools and regions needs to be performed to obtain the best strategy for the development. As Indonesia is very diverse in terms of geographics and infrastructure, it is needed to map the provinces based on the education development and for e-learning adoption readiness.

In general, the higher the education level of residents in a country or region, the higher the level of economic progress of the country/region. Improving the education system can reduce the poverty level of a region (Nguyen & Nguyen 2019). Furthermore, Reza and Widodo (2013) have shown that education has a positive impact on economic growth in Indonesia. Although the impact is relatively different among the provinces. The technology adoption in different sectors is also expected to boost the country's economic growth (Parente, 1994). ICT development has been shown to have a huge impact on regional economic growth (Agustina & Pramana, 2019).

This study is aimed to map the province in Indonesia based on the education and ICT indicators using several unsupervised learning algorithms.

2. Research Methods and Materials

2.1. Data

For the study, 27 indicators obtained from BPS' Statistics Indonesia and Ministry of Education were used (see Table 1). These indicators measure four dimensions: Education Quality, Participation, Facilities, and ICT Access.

2.2. Statistical Approach

To cluster the provinces based on these indicators, the following unsupervised learning approaches are implemented and compared:

Table 1: Education Quality and ICT Access Indicators

Dimension	Indicators
Quality	Adult (15-59) Literacy Rate Average Years of Schooling Percentage of Qualified Teachers in Primary School (PS) Percentage of Qualified Teachers in Junior High School (JHS) Percentage of Qualified Teachers in Senior High School (SHS) Graduated rate in PS Graduated rate in JHS Graduated rate in SHS
Participation	Net Enrolment Ratio in higher education Net Enrollment Ratio in JHS Net Enrollment Ratio in SHS Dropout Rate in PS Dropout Rate in JHS Dropout Rate in SHS
Facilities	Number of students per class PS Number of students per class JHS Number of students per class SHS Ratio of Students per Schools PS Ratio of Students per Schools JHS Ratio of Students per Teachers PS Ratio of Students per Teachers JHS Ratio of Students per Teachers SHS
ICT Access	ICT Development Percentage of Individuals with access to computer Percentage of Individuals with access to mobile cellular Percentage of students with access to computer Proportion of adults (15-59) with ICT skills

2.2.1 K-Means

This method is one of the simplest and popular unsupervised machine learning algorithms to find k clusters. The k number of centroids is calculated using the mean-and then allocating every data point to the nearest centroid. This procedure is repeated to optimize the position of the centroids until the centroids have stabilized or the defined number of iterations has been achieved (Pramana et al., 2018).

2.2.2. Fuzzy C-Means

This method is an improved method of K-Means algorithm. This clustering technique permits a data item to belong to some clusters with a defined fuzzy membership grade. It is robust to extreme observation. The FCM algorithm minimizes the following equation:

$$J_m(U, V) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \|x_j - v_i\|^2,$$

where u_{ij} is cluster membership and

$$v_i = \frac{\sum_{j=1}^n (u_{ij})^m x_j}{\sum_{j=1}^n (u_{ij})^m}, \quad 1 \leq i \leq c,$$

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{\|x_j - v_i\|^2}{\|x_j - v_k\|^2} \right)^{1/(m-1)} \right]^{-1}, \quad 1 \leq i \leq c, 1 \leq j \leq n.$$

(Bezdek, J. C., Robert, E., & William, F. (1984).

2.2.3. Cluster Validation Index

The biggest challenge in clustering is to check how good the clustering result is. Since the number of clusters need to specify in advance, obtaining the right number of clusters is crucial. Several measurements for cluster validations are available to find the right/optimal number of clusters (Pramana et al., 2018). In this study, the following clustering indices are carried out:

- a. **Within Class Variation.** It is the simplest way to measure to find the optimal number of clusters. Its defined as:

$$W_k = \sum_{r=1}^k \frac{1}{2n_r} D_r$$

where

$$D_r = \sum_{i \in C_r} \sum_{j \in C_r} \|x_i - x_j\|^2$$

$$= 2n_r \sum_{i \in C_r} \|x_i - \bar{x}\|^2$$

- b. **Silhouette Index.** Developed by Rousseeuw in 1987, it measures how well an observation is clustered and estimates the average distance between clusters. The values ranges from -1 to +1. For each observation, we can calculate the silhouette value using the following formula:

$$s(i) = \begin{cases} 1 - \frac{a(i)}{b(i)} & \text{if } a(i) < b(i), \\ 0 & \text{if } a(i) = b(i), \\ \frac{b(i)}{a(i)} - 1 & \text{if } a(i) > b(i) \end{cases}$$

where $a(i)$ is the mean distance between the i -th observation and all other data points in the same cluster, and $b(i)$ is the smallest mean distance of the i -th observation to all points in any other clusters. It measures how similar an object is to its own cluster (cohesion) as compared to other clusters (separation). High silhouette value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters.

- c. **The gap statistic** published by Tibshirani et al. (2001) compares the total within intra-cluster variation for different values of k with the expected values under null reference distribution of the data. The large value of gap statistic means that the clustering structure is far away from the random uniform distribution of points. Hence, the higher the value, the better the classification.

- d. **Xie Beni Index.** Wang and Zhang (2007) proposed a Xie Beni Index, which is a ratio of compactness and separation as defined below:

$$V_{XB} = \frac{J_m(u, v)/n}{Sep(v)} = \frac{\sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2}{n \min_{i,j} \|v_i - v_j\|^2}.$$

The number of optimal clusters can be seen from the smallest value of Xie Beni index.

2.2.4. Ensemble clustering

One of the limitations of partitioning clustering with such as K-means is that different initial partitions can result in different final clusters. One popular approach to mitigate this problem is the ensemble approach. Ensemble approach generates different clustering results, using different subsets of the data, algorithms, or different number of clusters, and then combine the results into a single consensus solution. It is shown that Ensemble methods give a more robust clustering results (Chiu & Talhouk, 2018; Firmansyah & Pramana, 2017).

In this study, K-means cluster ensemble with majority voting proposed by Ayad HG and Kamel MS (2010) is implemented by using the package *diceR* (Chiu & Talhouk, 2018). For fuzzy clustering, the ensemble based Fuzzy C-means (Firmansyah & Pramana, 2017) is carried out by using the R package *advclust*.

Once the clusters are identified, next is to profile the clusters based on their characteristics using Biplot and radar plot. All the analysis is conducted using the R software (R. Core team, 2017; Pramana et al., 2017).

3. Results and Discussion

First step is to find the optimal number of clusters using several cluster indices. Figure 1 shows that based on all four measurements, (a) within class variation, (b) Gap Statistics, (c) Silhouette score, and (d) Xie Beni Index, the optimal number of clusters is three. The Ensemble K-Means and Ensemble Fuzzy C Means are implemented and find similar results.

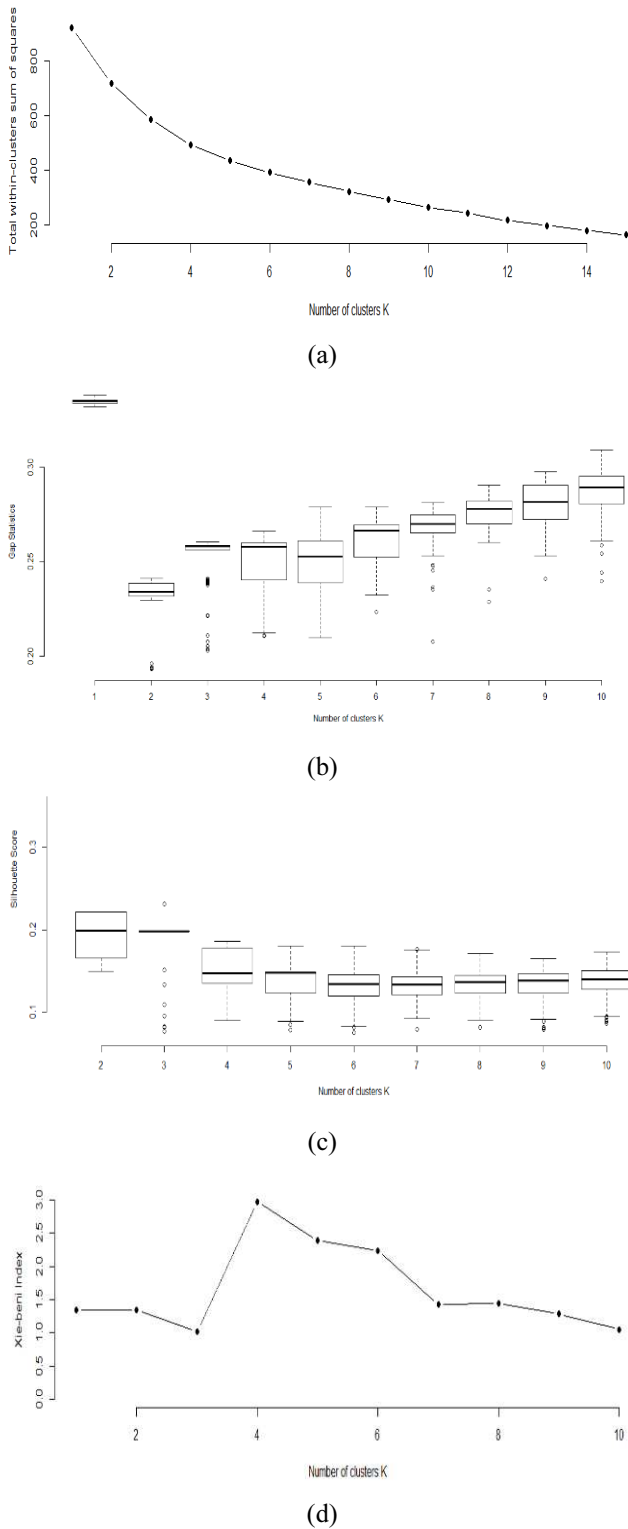


Figure 1: Clustering indices for different number of clusters (a) within class variation, (b) Gap Statistics, (c) Silhouette score, (d) Xie beni Index

Radar plot shows that Cluster 3 is the cluster with high education attainment and quality, as well as the ICT development. Whereas cluster 1 is the cluster with low education quality and ICT development. The other 23 provinces in cluster 2 obtain medium education and ICT quality.

From the Biplot shown in Figure 2, all members of cluster 3, except Yogyakarta, have highest number ICT Access (e.g., ICT Development index, Percentage Individual access mobile cellular). DKI Jakarta has the best on ICT infrastructure. Yogyakarta is best in education quality (high literacy rate, qualified teachers, and graduate rate) and participation (high Net Enrolment ratio, and low dropout rate).

In cluster 2, the drop out is very high, as well as the number of students per class and ratio of students per school. The other variables, such as literacy rate, ICT development, and number of qualified teachers, are very low in this cluster.

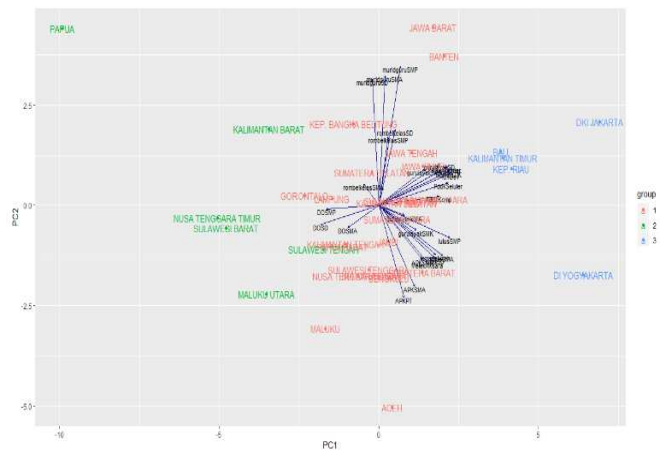


Figure 2: Biplot of the clustering results. Cluster 1, 2 and 3 are represented by color red, green, and blue, respectively.

Cluster 1 is the largest group that has high literacy rate, however the other variables are on average. Banten, Bangka Belitung, and West Java have high number of students per class and ratio of students per teachers. It shows that these provinces need to be improved in terms of the facilities such as more number of classrooms. Central Java, East Java and South Sumatera, are provinces with the number of qualified teachers and ICT development above other provinces in this cluster. The teachers of these provinces not only come from better education background but also have a good access on quality improvement and higher education. Moreover, the dropout rate is low in these provinces as well. The good quality teacher can have better education attainment then improving the quality of students.

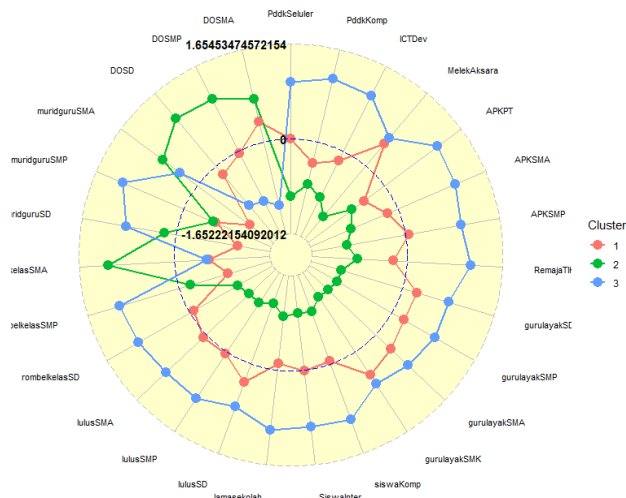


Figure 3: Radar plot of the Clustering Results

Figure 4 shows the mapping the group of provinces based on the education quality and ICT development. Most of the provinces in Sumatera and Java are included in cluster 2 where the education and ICT development is average.

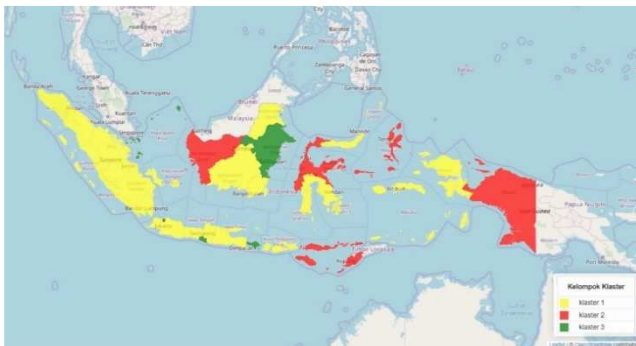


Figure 4: Mapping the Education Quality and ICT development.

Furthermore, it is interesting that the education attainment and ICT development of the provinces in Kalimantan Island are quite diverse. The East Kalimantan province has better education attainment and ICT infrastructure and access as compared to the other provinces from the same island. The West Kalimantan province show the lowest performance.

Moreover, the result shows that in cluster 3, DKI Jakarta, Yogyakarta, Riau Islands, East Kalimantan, and Bali provinces are ready for implementing e-learning system not only inside school, but also in a broader scope. The ICT infrastructure, ICT experts and digital literacy coupled with good education quality, facilities, and participations make it ready to support the implementation of e-learning. These provinces, such as DKI Jakarta, have large Gross Regional

Domestic Product (GRDP) and show positive economic growth during 2019.

The other provinces, such as Central Java, East Java, and South Sumatera with quite a good ratio of qualified teachers and ICT development, can also be included in e-learning system planning by improving their facilities to improve the participation and quality. These provinces have relatively medium GRDP and slightly high economic growth.

4. Conclusions

The study has revealed that the provinces in Indonesia can be clustered into three clusters based on the education quality, participation, facilities and ICT Access. The cluster with high education quality and ICT access consists of DKI Jakarta, Yogyakarta, Riau Islands, East Kalimantan, and Bali. These provinces shows high economic growth and can directly implement the e-learning system. Meanwhile, the other cluster consisting six provinces (East Nusa Tenggara, West Kalimantan, Central Sulawesi, West Sulawesi, North Maluku, and Papua) is the cluster with lower education quality and ICT development. The improvement of education facilities and ICT infrastructure is the most crucial in this cluster to enhance the economic growth. The study provides better insight on the cluster obtained and can be used for focusing on the education improvement program.

References

- Agustina, N., & Pramana, S. (2019). The Impact of Development and Government Expenditure for Information and Communication Technology on Indonesian Economic Growth. *The Journal of Business Economics and Environmental Studies*, 9(4), 5-13.
- Ayad, H., & Kamel, M. (2010). On voting-based consensus of cluster ensembles. *Pattern Recognition*, 43, 1943-1953. doi:doi: 10.1016/j.patcog.2009.11.012
- Bezdek, J. C., Robert, E., & William, F. (1984). FCM: The Fuzzy C-means Clustering. *Computers & Geosciences*, 10, 191-203.
- Chiu, D., & Talhouk, A. (2018). diceR: An R Package for Class Discovery using an Ensemble Driven Approach. *BMC Bioinformatics*. doi: 10.1186/s12859-017-1996-y
- Firmansyah, A., & Pramana, S. (2017). Ensemble Based Gustafson Kessel Fuzzy Clustering. *Journal of Data Science and Its Application*, 1(1), 1-9.
- Nguyen, H. H., & Nguyen, N. V. (2019). Factor Affecting Poverty and Policy Implication of Poverty Reduction: A Case Study for the Khmer Ethnic People in Tra Vinh Province. *Viet Nam Journal of Asian Finance*, 6 (Economics and Business), 315-319.
- OECD. (2018). *PISA 2018 Results Combined Executive Summaries*. Turkey: OECD. Retrieved from https://www.oecd.org/pisa/Combined_Executive_Summary

- es_PISA_2018.pdf
- Parente, S. L. (1994). Technology Adoption, Learning-by-Doing, and Economic Growth. *Journal of Economic Theory*, 63(2), 346-369.
- Pramana, S., Yordani, R., Kurniawan, R., & Yuniarto, B. (2017). *Dasar-dasar Statistika dengan Software R Konsep dan Aplikasi* (2nd ed.). Jakarta: InMedia.
- Pramana, S., Yuniarto, B., Mariyah, S., Santoso, I., & Nooraeni, R. (2018). *Data Mining dengan R, Konsep serta Implementasi*. Jakarta: InMedia.
- Reza, F., & Widodo, T. (2013). The Impact of Education on Economic Growth in Indonesia. *Journal of Indonesian Economy and Business*, 1, 28.
- Rousseeuw, P. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics*, 20, 53-65. doi:10.1016/0377-0427(87)90125-7
- Team, R. C. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from URL <https://www.R-project.org/>
- Tibshirani, R., Walther, G., & Hastie, T. (2000). Estimating the Number of Clusters in a Dataset Via The Gap Statistic. *Journal Royal Statistical Society*, 63(2), 411-423.
- UNDP. (2021). *Human Development Report 2020 The Next Frontier: Human Development and the Anthropocene*. UNDP. Retrieved from <http://hdr.undp.org/sites/all/themes/hdrtheme/country-notes/IDN.pdf>
- Wang, W., & Zhang, Y. (2007). On Fuzzy Cluster Validity Indices. In *Fuzzy Sets and Systems* (pp. 2095-2117).