IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 5, October 2025, pp. 3771~3780
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i5.pp3771-3780

Data-driven clustering and prediction of high school graduation rates in Indonesia (2015-2023) using machine learning

Muhammad Salman Arrosyid, Marzuki, Widihastuti, Haryanto, Maria Angelina Fransiska Mbari
Department of Educational Research and Evaluation, Graduate School, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

Article Info

Article history:

Received Jan 7, 2025
Revised Jun 24, 2025
Accepted Aug 6, 2025

Keywords:

Education
Education disparity
Graduation rate
K-means clustering
Machine learning
Neural network

ABSTRACT

This study aims to analyze the graduation rate of senior high school education in 34 Indonesian provinces during the period 2015-2023 and identify patterns of educational disparities between regions. To achieve the objectives, this study applies a neural network to predict education completion patterns based on historical data, then the prediction results are analyzed using K-means clustering technique utilizing the elbow method to select the ideal number of clusters. The clustering results show three categories of provinces based on education completion rates: high, medium, and low. The provinces with high completion rates, generally, supported with good education infrastructure and effective policies, while the medium category faces challenges in resource distribution, but still potentially improve. In contrast, the low category suffers from limited access, geographical constraints, and socio-economic disparities. This research contributes to education policy-making by offering a machine learning-based approach to understanding education disparities between regions. The new insight offered by this study lies in the integration of neural network and K-means clustering in mapping education completion rates to support strategies for improving access and quality of education in Indonesia.

Corresponding Author:

Muhammad Salman Arrosyid
Department of Educational Research and Evaluation, Graduate School, Universitas Negeri Yogyakarta
St. Colombo No.1 Karangmalang Yogyakarta 55281, Indonesia
Email: muhammad0039pasca.2023@student.uny.ac.id

1. INTRODUCTION

Education is the main foundation of the Indonesian country in creating a smart and civilized society [1]. However, this education sector is still plagued by complex problems, especially the inequitable distribution of education, which hinders access to quality education for all levels of society [2], [3]. Based on Law Number 21 of 2003. According to the National Education System, education refers to a purposeful and systematically designed effort to build an environment and learning process that encourages learners to be active and interact in developing their potential. The purpose of this education is to form students to develop spiritual strength, self-regulation, intelligence, commendable character, and competencies essential for oneself, the community, the nation, and the country [4], [5].

Increasing complexity of global challenges [6], such as climate change, poverty, and social inequality [7]. Therefore education has a strategic role as the main instrument in shaping and increasing the capacity of human resources and empowering the community to contribute to building and realizing social welfare without inequality [8]. Through education, humans not only acquire knowledge, but also skills, critical thinking and awareness to adaptively face the global dynamics.

The senior high school education level is an important and decisive phase in the educational journey of students [9]. Senior high school is a crucial step for students towards higher education and employment

[10], [11], Therefore, students at the senior high school level must be equipped with basic knowledge, critical thinking skills, and cognitive and affective skills to face various challenges after they complete the senior high school level [12]. Contextually, it is important to ensure that senior secondary education is equally accessible to all people in Indonesian provinces with no disparities in facilities, quality of teaching and learning, or learning opportunities [13]. The data from the Central Bureau of Statistics or Badan Pusat Statistik (BPS) shows that the completion rate of senior high school education in 34 provinces during 2015-2023 experienced significant disparities. Geographical and socio-economic factors are the main obstacles, especially in remote, border and underdeveloped areas. This disparity affects the quality and quantity of high school graduates, and hampers equitable human resource development in Indonesia [14].

Education equity is not only served as an effort to ensure justice for all learners, but also a fundamental element in creating balanced human resource development throughout Indonesia. This inequality in educational, social and economic access has a significant negative impact on the quality and quantity of high school graduates [15]. By ensuring sustainable education equity, every province in Indonesia has the potential to produce high school graduates who not only excel in quality, but also increase in quantity. This equity will contribute to the strengthening the local economy in each province and support national economic growth holistically. In addition, the graduate students will have adequate competencies to face the dynamics and challenges of global competition, thus being able to contribute significantly to nation building.

Along with the rapid enhancement of technology, the application of machine learning is increasingly widespread in various sectors and applications [16]. Machine learning is a branch of artificial intelligence (AI) that enables information systems to automatically learn patterns, relationships, and characteristics in data without requiring explicit programming instructions [17]. With these capabilities, machine learning is able to generate new knowledge, train algorithms, identify relationships, and recognize hidden patterns that have not been previously detected [18]. The patterns and relationships discovered through this process can be used to analyze new and unknown data, enabling more accurate predictions and supporting efficient and adaptive process optimization [19].

This study aims to analyze the graduation rates of high school students in various provinces in Indonesia through a machine learning approach, by integrating the neural network method to predict graduation rates and the K-means clustering algorithm to group provinces based on completion patterns during the period 2015-2023. K-means is known as one of the oldest and most widely used partitioning methods [20]. This algorithm has been the object of extensive study with various developments in the literature and applied in various substantive fields [21]. This method allows the process of grouping data based on the similarity of certain characteristics [22]. This study provides a more systematic picture of the pattern of education equity and inequality in each province. The findings are expected to serve as a scientific basis in formulating strategic policies to support equitable access to education and improve the quality of senior secondary school graduates. In addition, the results of this study also contribute to strengthening national competitiveness through the development of human resources that are more adaptive and competitive in the midst of global dynamics.

2. METHOD

This study analyzes the completion rate of senior high school education in 34 provinces in Indonesia in the period 2015-2023 using data from BPS. This study is conducted in two phases, the are data preparation stage and the clustering stage, as shown in Figure 1. The data preparation stage is carried out in several stages, namely collecting raw data, cleaning data to eliminate errors or irrelevant data, filtering data, combining all data, and converting data to suit the needs of the analysis. Further data analysis was conducted by combining machine learning neural network method and K-means clustering, to achieve a thorough comprehension of the pattern of graduation or completion of education at the high school level. This integrated approach provides a robust framework for understanding regional differences and patterns of educational outcomes, especially high school completion rates in Indonesia. Figure 1 shows the stages of data analysis.

2.1. Data

This research utilizes a dataset containing data on the percentage of completion or graduation rates of senior high school students in 34 provinces in Indonesia, presented in Table 1. This data is sourced from the BPS and covers the period 2015-2023. The initial stage was carried out by processing and preparing the data using RStudio software before further analysis was carried out by performing several stages, namely collecting raw data, cleaning data to eliminate errors or irrelevant data, filtering data, combining all data, and converting data to suit the needs of the analysis. The analysis was conducted to identify graduation patterns and educational trends in each province, which contributes to a detailed analysis of the determinants affecting students success rate in completing senior secondary education. The following is the data on the percentage (%) of senior high school completion rates from provinces in Indonesia from 2015 to 2023 in Table 1.

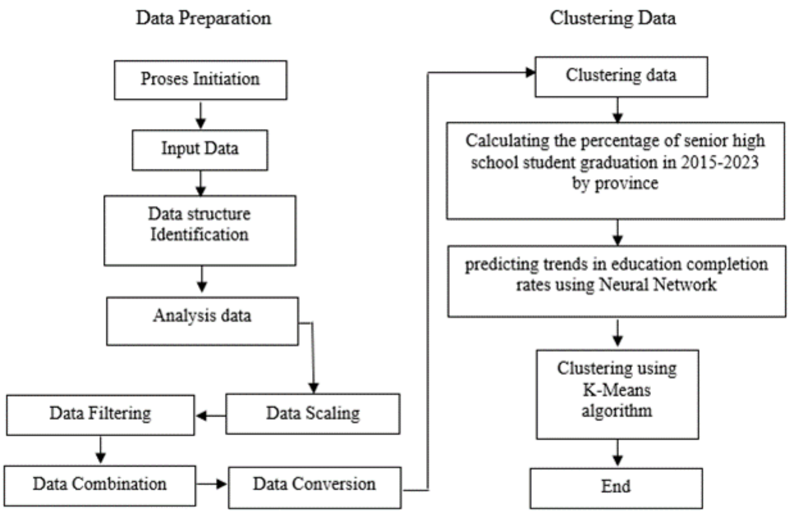

Figure 1. Flowchart of data analysis

Table 1. Student pass percentage data

Province	2015	2016	2017	2018	2019	2020	2021	2022	2023
Aceh	68.16	74.46	70.64	70.68	69.96	70.07	74.36	70.67	74.46
Sumatera Utara	59.54	69.69	67.16	68.34	65.21	70.39	72.81	77.16	74.43
Sumatera Barat	58.04	64.97	60.8	65.34	60.32	67.11	70.06	65.96	68.64
Riau	57.28	62.12	61.9	63.71	58.78	66.62	68.94	66.91	67.79
Jambi	49.05	60.5	58.27	66.06	56.87	63.66	64.51	65.85	66.62
Sumatera Selatan	48.9	55.37	54.15	63.94	58.23	65.42	67.2	67.07	64.81
Bengkulu	55.94	64.31	62.57	58.86	61.47	62.73	62.46	64.88	63.41
Lampung	40.6	47.62	48.75	54.89	54.87	57.59	60.09	62.42	64.54
Kep. Bangka Belitung	43.46	53.84	51.55	55.01	53.84	56.74	63.98	66.87	68.96
Kep. Riau	65.28	75.93	83.55	82.86	78.14	78.65	81.07	73.93	78.97
Jakarta	74.1	74.74	78.25	83.48	84.35	85.67	84.98	87.71	88.1
Jawa Barat	48.53	55.03	48.32	61.04	57.46	63.56	64.89	67.05	66.47
Jawa Tengah	43.86	44.59	51.52	55.62	49.79	55.82	59.9	58.75	58.35
DI Yogyakarta	80.77	79.95	85.53	81.96	84.54	87.99	90.12	87.92	89.69
Jawa Timur	52.04	55.13	59.9	62.48	57.74	63.53	66.33	66.87	68.65
Banten	52.95	60.83	59.87	67.54	56.94	64.24	66.9	66.02	70.07
Bali	69.08	73.65	74.62	78.67	64.52	74.88	75.86	76.59	76.51
Nusa Tenggara Barat	51.83	55.01	59.1	52.6	57.6	64.66	65.71	61	63.66
Nusa Tenggara Timur	37.78	48.95	41.44	43.41	43.85	50.65	44.88	38.47	43.46
Kalimantan Barat	35.69	35.69	42.7	47.66	49.29	55.23	54.27	58.4	55.58
Kalimantan Tengah	47.28	52.42	56.48	53.47	50.01	60.77	61.04	61.88	63.93
Kalimantan Selatan	44.85	52.91	56.75	61.09	59.52	63.05	63.59	67.81	68.35
Kalimantan Timur	67.56	66.76	67.72	68.73	64.74	71.63	74.26	74	73.63
Kalimantan Utara	47.64	58.6	57.43	58.22	61.1	67.77	62.3	54.8	59.5
Sulawesi Utara	55.5	72.33	67.46	70.02	67.58	73.79	68.56	66.66	67.57
Sulawesi Tengah	45.84	61.79	62.73	53.84	52	57.68	61.16	53.73	55.69
Sulawesi Selatan	50.85	59.56	63.82	56.86	60.97	66.22	69.43	68.32	67.41
Sulawesi Tenggara	61.52	67.12	67.75	67.67	64.26	68.28	70.65	65.97	68.28
Gorontalo	44.67	50.79	55.3	52.39	50.87	55.35	53.73	45.12	46.19
Sulawesi Barat	39.29	53.45	56.17	37.65	48.2	56.6	56.22	55.18	54.79
Maluku	58.59	72.87	73.58	66.42	67.82	70.55	68.12	72.08	75.01
Maluku Utara	57.12	64.87	65.14	60.07	59.13	66.52	66.95	67.1	64.61
Papua Barat	55.24	56.12	62.81	60.47	50.95	61.49	59.08	57.07	59.99
Papua	28.23	35.69	33.82	29.56	27.44	30.92	32.95	39.01	39.5

2.2. Neural network

Neural network is used to predict trends in education completion rates based on historical data. This method was chosen because it has the ability to capture non-linear patterns and complex relationships between variables. The neural network model is trained using a dataset that includes education indicators, such as graduation rates in each province. The model used is a multi-layer perceptron (MLP) with several hidden layers optimized to capture complex relationships between variables. The prediction results from this neural network provides an estimation of the education graduation or completion rate, which used as the basis for the clustering process using K-means. The output of this model allows for more specific analysis in clustering provinces based on the educational patterns identified through the predictions generated. To assess the accuracy and effectiveness of the neural network model, validation was conducted using four main metrics, namely mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2) [23].

2.2.1. Mean squared error

MSE measures the average of the squared difference between the actual value (*Yi*) and the predicted value (*yi*). The smaller the MSE value, the better model predicts the data with minimal error.

$$\boldsymbol{MSE = \frac{1}{n} \sum_{i=1}^{n} Y_i - y_i} \tag{1}$$

2.2.2. Root mean squared error

RMSE is an error measure that quantifies the variation between forecasted and true values. RMSE is determined by applying the square root to the MSE. A smaller RMSE value indicates that the model has a low prediction error rate.

$$\boldsymbol{RMSE = \sqrt{MSE}} \tag{2}$$

2.2.3. Mean absolute error

MAE is a measure of the average absolute error of the actual value (*Yi*) and the predicted value (*yi*). MAE measures how large the average difference is between the predicted value and the actual value, regardless of the direction of the error (positive or negative). The smaller the MAE value, the better the model is at making predictions.

$$\boldsymbol{MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i - y_i|} \tag{3}$$

2.2.4. R-squared

R^2 is a statistical metric used to assess the extent to which the model can explain variability in the actual data (*Yi*). R^2 values fall within the range of 0 to 1, where a value approaching 1 indicates that the model is getting better at explaining data variation.

$$\boldsymbol{R^2 = 1 - \frac{\sum(Y_i - y_i)^2}{\sum(Y_i - y_i)^2}} \tag{4}$$

2.3. K-means clustering

K-means clustering serves to classify provinces based on similar educational characteristics. The K-means clustering method is used because of its ability to efficiently group provinces based on the similarity of initial centroid values of educational characteristics [24], [25], because it is able to evaluate changes in the total within-cluster sum of squares (WSS) value against various numbers of clusters. The elbow method is used to determine the optimal number of clusters by balancing intra-cluster variability and optimal model complexity, when the decrease in WSS starts to slow down significantly. The selection of the number of clusters aims to achieve a balance point between clustering accuracy and model complexity, so that the analysis results are more representative and the data can be interpreted properly into fixed groups [26].

2.4. Relevance of neural network and K-means clustering

This data analysis combines neural network and K-means clustering as they have significant relevance in education analysis. Neural network is used to model the complex relationships between various educational factors, with its ability to capture non-linear patterns allowing for more accurate prediction of educational completion rates. The model applied is a MLP with optimized hidden layers to improve prediction accuracy. After obtaining prediction results from the neural network. K-means clustering method is applied to group provinces based on similar educational characteristics [27]. This approach allows the identification of provincial clusters with homogeneous patterns of education completion [28]. So that it can be the basis in formulating educational policies that are more targeted [29]. Neural network acts as a prediction tool that captures complex relationships in the data, while K-means clustering groups provinces based on the prediction results, allowing for more systematic and in-depth analysis [30]. The combination of these two methods not only improves the accuracy in projecting education trends, but also provides a clearer mapping of areas based on education characteristics, thus supporting data-driven decision-making more optimally.

3. RESULTS AND DISCUSSION

This study applies the neural network and K-means clustering methods to analyze the graduation rate of senior high school education in 34 provinces of Indonesia during the period 2015-2023. The data used is sourced from the BPS and includes the main variables in the form of the percentage of education completion each year. These variables are used as a representation of the level of educational success in various provinces throughout the study period.

At the initial stage of the analysis, a data normalization process was carried out using the Z-score scaling method. This step aims to equalize the scale between variables so that no variable has a dominant influence on clustering results. The scaling process is used in the neural network to improve the stability of model learning and ensure the activation function works optimally [31]. In addition, scaling is also applied to K-means clustering so that the clustering of provinces based on the pattern of education completion rates can be done objectively, taking into account the equal distribution of the data [32].

The results of the neural network analysis are visualized through a model structure that describes the relationship between the input data in the form of annual graduation scores (2015-2023) and the output in the form of graduation. The weights between neurons that represent the impact of individual variables on the predictive outcome results. The following are the results of the neural network visualization using RStudio in Figure 2.

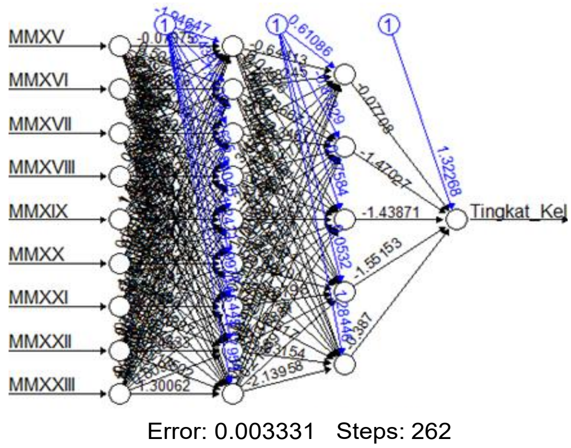

Figure 2. Neural network results

Based on the neural network results above, the model consists of two hidden layers that progressively process input data in the form of annual graduation rates (2015-2023) into output in the form of graduation rates. The weights between neurons, as shown in Figure 2, reflect the level of contribution of each input variable to the prediction result [33]. The visualized neuron activations show how the model learns complex patterns of relationships, including non-linear patterns that cannot be captured by conventional analysis methods. The implemented neural network model shows a high level of accuracy in predicting education completion patterns. The model validation results were conducted using four main evaluation metrics, namely the MSE of 0.0001936, which indicates a very small prediction error rate after 262 iterations. In addition, the RMSE of 0.0139 and the result of the MAE of 0.0100 indicate that the model has a good performance in reducing the prediction error. Furthermore, the R^2 value of 0.662 provides information that the model is able to explain 66.2% of data variability, while the rest which influenced by other factors not included in the model. Additionally, the overall error value of 0.003331 listed in the analysis results confirms the stability of the model during the training process. Thus, the neural network validation results confirm that the model can effectively identify the relationship between historical variables and education completion rates, thus supporting the validity of the model in making predictions.

The next step in this research aims to identify the optimal cluster count based on the neural network prediction results used in performing K-means clustering. Cluster analysis is a method in unsupervised learning that seeks to divide data into specific groups, ensuring that elements within a cluster have homogeneous properties, while differences between clusters are made as clear as possible [34]. To achieve this goal, the elbow method is applied, a visual approach that illustrates the relationship between the number of clusters and the total WSS value. The graphs generated by the elbow method provide guidance in identifying the most statistically and interpretatively optimal number of clusters. Figure 3 shows the number of clusters generated.

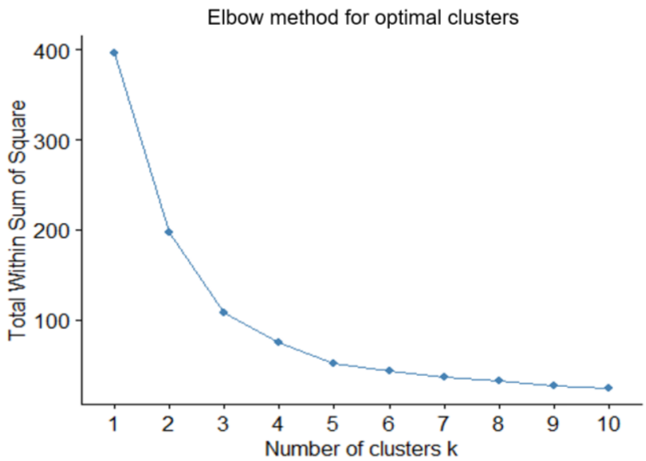

Figure 3. Optimal number of clusters

Based on the resulting elbow method graph, it can be seen that the total WSS value decreases sharply from k=1k, k=2k, and k=3k. After k=3k, the decrease in WSS value becomes slower and insignificant. This pattern indicates that the use of three clusters is the optimal choice, as the number of additional clusters after 3k does not provide a significant improvement in reducing within-cluster variation. Therefore, the optimal number of clusters chosen in K-means clustering analysis is three clusters, which provides a balance between the model's performance in representing data variation and the complexity of the model formed.

K-means algorithm is an algorithm with partitioning [35], because K-means method requires determining the number of clusters at the beginning of the process, the algorithm starts by setting an initial centroid value that will serve as the center point for each group [36]. The K-means algorithm requires a definite number of clusters to be determined before the clustering process can be performed [37], because the initial positions of the cluster centers can vary, this may lead to inconsistent clustering outcomes for the data. This clustering aims to present a nuanced perspective of the differences and similarities between provinces related to high school graduation rates, which can be used to understand regional dynamics more comprehensively. The following is the result of clustering provinces using the K-means algorithm shown in Figure 4.

Based on Figure 4, which presents the outcomes of the clustering analysis performed with the K-means algorithm clustering method, the three clusters formed are visualized with different symbols and colors to facilitate interpretation. This visualization shows the distribution of provinces based on the level of high school completion in each province. Dimension 1 (90.5%) and dimension 2 (4.5%) represent the principal components resulting from dimension reduction using the principal component analysis (PCA) method, which covers a total of 95% of the data variability. Each point on the graph represents the position of each province in the dimension space. The following is an interpretation of the data based on the clusters formed.

Cluster one (red), which consists of most provinces such as Central Java, East Java, and North Sumatra, has moderate levels of education completion. Provinces in this cluster face a number of challenges, including variations in education accessibility, inequality in infrastructure, limited resources and the need to improve teaching quality. Nevertheless, this cluster has a great opportunity to improve education performance through more targeted policies and infrastructure development that supports equitable access to education.

The second cluster (green) includes provinces with low education completion rates, such as Papua, East Nusa Tenggara (NTT), and West Sulawesi, which face significant challenges such as difficult geographical access, lack of education facilities and stark socio-economic disparities. These challenges require strategic interventions such as improving education accessibility in remote areas, developing supportive infrastructure and reducing disparities between regions. The cluster findings provide important insights into the variations in education performance in Indonesia, which can inform the formulation of data-driven education policies to improve the quality and equity of education nationwide.

The third cluster (blue) includes provinces with high education completion rates, such as DKI Jakarta and DI Yogyakarta. Provinces in this cluster have excellent access to education, adequate infrastructure and effective implementation of education policies. These factors enable the provinces in this cluster to consistently achieve education performance above the national average.

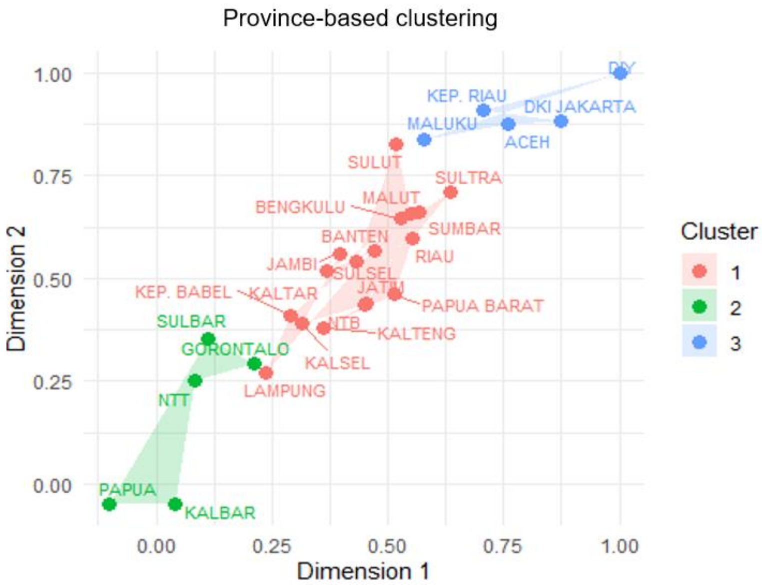

Figure 4. Visualization of K-means algorithm results

4. CONCLUSION

This study examines the graduation rate of senior high school education in 34 provinces in Indonesia from 2015-2023 by utilizing a combination of neural network and K-means clustering methods. The neural network model showed superior ability in predicting education completion rates with a MSE of 0.0001936, reflecting its ability to capture non-linear relationships from historical data. The resulting predictions were then used in a K-means clustering analysis to group the provinces into three main clusters: high, medium, and low completion rates. The combined approach of neural network and K-means clustering provides a comprehensive picture of education disparities in Indonesia. The findings of this study offer a strong basis for data-driven policy making to improve education quality and reduce disparities among provinces in Indonesia. Future research by including additional variables is expected to broaden the analysis and strengthen policy recommendations that are more targeted.

ACKNOWLEDGEMENTS

This research would not have been possible without the support of various parties. Our appreciation goes to the Central Bureau of Statistics or Badan Pusat Statistik (BPS) for providing the education data on which this research is based, and to our institution for the facilities and opportunities provided. We also thank our fellow researchers for their valuable inputs, which enriched the quality of this study. Special thanks go to Prof. Erna Andriyanti, S.S., M.Hum., Ph.D., for her insightful guidance, suggestions, and continuous support, which greatly helped in the completion of this article.

FUNDING INFORMATION

This research was funded by the Center for Financing and Assessment of Higher Education (PPAPT), Ministry of Education, Culture, Research and Technology (Kemdikbudristek), through the decision of the Head of the Center for Education Financing, Kemdikbudristek, No. 00571/BPPT/BPI.06/9/2023. Additional support was provided by the Education Fund Management Institution (LPDP), Ministry of Finance of the Republic of Indonesia, in the form of a scholarship awarded to the fifth author. The authors also acknowledge Nusa Nipa University Maumere for supporting the fifth author's academic career.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Muhammad Salman Arrosyid	✓	✓	✓		✓		✓		✓	✓	✓		✓	
Marzuki	✓			✓		✓		✓		✓		✓		
Widihastuti		✓		✓				✓		✓	✓	✓	✓	
Haryanto			✓			✓		✓	✓		✓	✓		
Maria Angelina Fransiska Mbari	✓	✓			✓		✓		✓	✓			✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

DATA AVAILABILITY

The data that support the findings of this study are openly available in *Badan Pusat Statistik* or Central Bureau of Statistics at https://www.bps.go.id/id/statistics-table/2/MTk4MSMy/tingkat-penyelesaian-pendidikan-menurut-jenjang-pendidikan-dan-wilayah.html.

REFERENCES

[1] M. C. R.-Jiménez, R. M.-Jiménez, A. L.-Gutiérrez, and E. G-Martí, "Students' attitude: key to understanding the improvement of their academic results in a flipped classroom environment," *International Journal of Management Education*, vol. 20, no. 2, pp. 1–16, 2022, doi: 10.1016/j.ijme.2022.100635.

[2] C. Chabbott and M. Sinclair, "SDG 4 and the COVID-19 emergency: textbooks, tutoring, and teachers," *Prospects*, vol. 49, no. 1–2, pp. 51–57, 2020, doi: 10.1007/s11125-020-09485-y.

[3] E. F. S. Rini, F. T. Aldila, and R. P. Wirayuda, "A study of student learning discipline in senior high school," *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, vol. 7, no. 1, pp. 33–37, 2023, doi: 10.22437/jiituj.v7i1.26698.

[4] M. S. Arrosyid, F. Fakhruddin, and Z. Elmubarok, "Developing an attitude assessments instrument and Pancasila student profiles in islamic religious education subject," *Journal of Research and Educational Research Evaluation*, vol. 11, no. 2, 2022.

[5] M. Huda, A. Sudrajat, R. Muhamat, K. S. M. Teh, and B. Jalal, "Strengthening divine values for self-regulation in religiosity: insights from Tawakkul (trust in God)," *International Journal of Ethics and Systems*, vol. 35, no. 3, pp. 323–344, 2019, doi: 10.1108/ijoes-02-2018-0025.

[6] A. K.-Dinya, "Managing challenges of increasing complexity in sustainability," *Ecocycles*, vol. 6, no. 2, pp. 49–53, 2020, doi: 10.19040/ecocycles.v6i2.159.
[7] Heriberta, Zulfanetti, and R. Setiawati, "Exploring the impact of age and motivation on self-determination: insights from informal sector mothers," *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, vol. 8, no. 1, pp. 306–321, 2024, doi: 10.22437/jiituj.v8i1.32880.
[8] J. Jeong, D. Hong, and J. Chang, "Analysis in Republic of Korea : long-term research conducted," *System*, vol. 11, pp. 1–20, 2023.
[9] J. L. Fua, I. S. Wekke, Z. Sabara, and R. U. Nurlila, "Development of environmental care attitude of students through religion education approach in Indonesia," *IOP Conference Series: Earth and Environmental Science*, vol. 175, no. 1, pp. 1–7, 2018, doi: 10.1088/1755-1315/175/1/012229.
[10] D. Purohit, M. Jayswal, and A. Muduli, "Factors influencing graduate job choice – a systematic literature review," *European Journal of Training and Development*, vol. 45, no. 4–5, pp. 381–401, 2021, doi: 10.1108/EJTD-06-2020-0101.
[11] N. C. Hai, N. H. Thanh, T. M. Chau, T. V. Sang, and V. H. Dong, "Factors affecting the decision to choose a university of high school students: a study in an Giang Province, Vietnam," *International Journal of Evaluation and Research in Education*, vol. 12, no. 1, pp. 535–545, 2023, doi: 10.11591/ijere.v12i1.22971.
[12] C. Carrico, H. M. Matusovich, and M. C. Paretti, "A qualitative analysis of career choice pathways of college-oriented rural central appalachian high school students," *Journal of Career Development*, vol. 46, no. 2, pp. 94–111, 2019, doi: 10.1177/0894845317725603.
[13] A. Poncini, "Standards setting in religious education: addressing the quality of teaching and assessment practices," *Religions*, vol. 14, no. 3, pp. 1–14, 2023, doi: 10.3390/rel14030315.
[14] Z. Ma, F. O. Agyeman, A. K. Sampene, M. Li, and S. Botchway, "Senior high school teachers' preference for professional development: a case study in the central region of Ghana under the current free senior high school policy," *SAGE Open*, vol. 13, no. 3, pp. 1–14, 2023, doi: 10.1177/21582440231185116.
[15] S. Irianti and P. Prasetyoputra, "Rural–urban disparities in access to improved sanitation in Indonesia: a decomposition approach," *SAGE Open*, vol. 11, no. 3, 2021, doi: 10.1177/21582440211029920.
[16] N. Kühl, M. Goutier, R. Hirt, and G. Satzger, "Machine learning in artificial intelligence: towards a common understanding," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2019, pp. 5236–5245, doi: 10.24251/HICSS.2019.630.
[17] I. H. Sarker, "Machine learning: algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, 2021, doi: 10.1007/s42979-021-00592-x.
[18] R. C. Fong, W. J. Scheirer, and D. D. Cox, "Using human brain activity to guide machine learning," *Scientific Reports*, vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-23618-6.
[19] M. L. Hoang and N. Delmonte, "K-centroid convergence clustering identification in one-label per type for disease prediction," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 1, pp. 1149–1159, 2024, doi: 10.11591/ijai.v13.i1.pp1149-1159.
[20] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14, pp. 281–297, 1967.
[21] M. Alhawarat and M. Hegazi, "Revisiting K-means and topic modeling, a comparison study to cluster Arabic documents," *IEEE Access*, vol. 6, pp. 42740–42749, 2018, doi: 10.1109/ACCESS.2018.2852648.
[22] A. Tsitsulin, J. Palowitch, B. Perozzi, and E. Müller, "Graph clustering with graph neural networks," *Journal of Machine Learning Research*, vol. 24, pp. 1–21, 2023.
[23] C. Stylianoudaki, I. Trichakis, and G. P. Karatzas, "Modeling groundwater nitrate contamination using artificial neural networks," *Water*, vol. 14, no. 7, pp. 1–15, 2022, doi: 10.3390/w14071173.
[24] Y. Duan, Q. Liu, and S. Xia, "An improved initialization center k-means clustering algorithm based on distance and density," *AIP Conference Proceedings*, vol. 1955, no. August, 2018, doi: 10.1063/1.5033710.
[25] Y. Meng, J. Liang, F. Cao, and Y. He, "A new distance with derivative information for functional K-means clustering algorithm," *Information Sciences*, vol. 463–464, pp. 166–185, 2018, doi: 10.1016/j.ins.2018.06.035.
[26] W. Yanbo, L. Li, P. Xinfu, and F. Enpeng, "Load forecasting based on improved K-means clustering algorithm," *China International Conference on Electricity Distribution, CICED*, pp. 2751–2755, 2018, doi: 10.1109/CICED.2018.8592023.
[27] T. S. Xu, H. D. Chiang, G. Y. Liu, and C. W. Tan, "Hierarchical K-means method for clustering large-scale advanced metering infrastructure data," *IEEE Transactions on Power Delivery*, vol. 32, no. 2, pp. 609–616, 2017, doi: 10.1109/TPWRD.2015.2479941.
[28] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
[29] J. F. Lu, J. B. Tang, Z. M. Tang, and J. Y. Yang, "Hierarchical initialization approach for K-Means clustering," *Pattern Recognition Letters*, vol. 29, no. 6, pp. 787–795, 2008, doi: 10.1016/j.patrec.2007.12.009.
[30] Arnita, F. Marpaung, F. Ramadhani, and D. Dinata, "Machine learning for mapping and forecasting poverty in north sumatera: a data-driven approach," *Sains Malaysiana*, vol. 53, no. 7, pp. 1715–1728, 2024, doi: 10.17576/jsm-2024-5307-18.
[31] K. L. Du, "Clustering: a neural network approach," *Neural Networks*, vol. 23, no. 1, pp. 89–107, 2010, doi: 10.1016/j.neunet.2009.08.007.
[32] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, pp. 1–17, 2021, doi: 10.3390/e23060759.
[33] M. Cho, K. A.-Vahid, S. Adya, and M. Rastegari, "Differentiable K-means clustering layer for neural network compression," in *ICLR 2022 - 10th International Conference on Learning Representations*, 2022, pp. 1–19.
[34] V. Melnykov and S. Michael, "Clustering large datasets by merging k -means solutions," *Journal of Classification*, vol. 37, no. 1, pp. 97–123, 2020, doi: 10.1007/s00357-019-09314-8.
[35] P. Garikapati, K. Balamurugan, T. P. Latchoumi, and R. Malkapuram, "A cluster-profile comparative study on machining AlSi7/63% of SiC hybrid composite using agglomerative hierarchical clustering and K-means," *Silicon*, vol. 13, no. 4, pp. 961–972, 2021, doi: 10.1007/s12633-020-00447-9.
[36] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-means clustering method and elbow method for identification of the best customer profile cluster," *IOP Conference Series: Materials Science and Engineering*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.
[37] J. Kauffmann, M. Esders, L. Ruff, G. Montavon, W. Samek, and K. R. Muller, "From clustering to cluster explanations via neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 2, pp. 1926–1940, 2024, doi: 10.1109/TNNLS.2022.3185901.

BIOGRAPHIES OF AUTHORS

Muhammad Salman Arrosyid, S.Pd., M.Pd. obtained a master's degree in Educational Research and Evaluation at Universitas Negeri Semarang, Indonesia and a bachelor's degree from UIN Raden Intan Lampung, Lampung, Indonesia. Now a student in the Educational Research and Evaluation Doctoral Study Program, Universitas Negeri Yogyakarta Postgraduate School. He can be contacted at email: muhammad0039pasca.2023@student.uny.ac.id.

Prof. Dr. Marzuki, M.Ag. completed his undergraduate studies at the Faculty of Tarbiyah, UIN Sunan Kalijaga, Yogyakarta, in 1990. He earned his master's degree in Islamic Studies from the Graduate School of UIN Syarif Hidayatullah, Jakarta, in 1997, and completed his doctoral degree at the same institution in 2007. In 2021, he was appointed as a professor in the field of Islamic Religious Education at the Faculty of Social Sciences, Universitas Negeri Yogyakarta. He can be contacted at email: marzuki@uny.ac.id.

Dr. Widihastuti, S.Pd., M.Pd. obtained a bachelor's degree in Education from UNY in 1998. Graduated with a master's degree from UNY Postgraduate Program in 2007. Obtained a doctorate in Education Research and Evaluation from the UNY Postgraduate Program in 2014. She can be contacted at email: widihastuti@uny.ac.id.

Dr. Drs. Ir. Haryanto, M.Pd., M.T. is a lecturer at the Faculty of Engineering and Research and Educational Evaluation, Universitas Negeri Yogyakarta. His research interests focus on artificial intelligence control, education research, and technical and vocational education. He can be contacted at email: haryanto@uny.ac.id.

Maria Angelina Fransiska Mbari, M.Pd. became a doctoral student in the Education Research and Evaluation Study Program at the Postgraduate School of Universitas Negeri Yogyakarta. She currently works as a permanent lecturer in the Primary School Teacher Education Study Program at Universitas Nusa Nipa, Maumere. She can be contacted at email: anjelinaanselmus@gmail.com.