ISSN xxxx-xxxx

K-Means Analysis of Indonesia's 2025 Education Participation Disparities

Egi Dio Bagus Sudewo[1], Dimas Ade Irwanda[2], Syahul Arifin[3], Diantoro Putra Wiguna[4], Sigit Nugraha[5]

Sistem Informasi/Universitas Royal, Asahan, Indonesia[1,2,3,4,5]
Jl. Imam Bonjol No.179, Kisaran Kota, Kec. Kota Kisaran Barat, Kabupaten Asahan, Sumatera Utara 21211 Indonesia

e-mail : egidio8gali@gmail.com[1], irwandadimasade@gmail.com[2], sa0247096@gmail.com[3], toroputra07@gmail.com[4], Sigitnugraha2704@gmail.com[5]

Abstract

The Pure Participation Rate (PPR), also known in Indonesia as Angka Partisipasi Murni (APM), represents an important indicator for assessing the level of public participation in Indonesia's education system across different educational stages. This study aims to classify the 38 provinces in Indonesia into clusters based on PPR values at three educational levels, namely elementary/equivalent, junior high/equivalent, and senior high/equivalent, by employing the K-Means clustering method. The dataset utilized in this research was obtained from the Central Bureau of Statistics (BPS) for the year 2025. The research process involved several stages, including data preprocessing, normalization through the StandardScaler technique, identification of the optimal number of clusters using the Elbow Method, and evaluation of clustering performance through the Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. The findings indicate that the provinces in Indonesia are optimally categorized into two primary clusters. Cluster 0 comprises 32 provinces characterized by relatively high PPR values (Elementary: 96.53%, Junior High: 79.16%, Senior High 66.49%), whereas Cluster 1 consists of six provinces in the Papua region that demonstrate comparatively lower participation rates (Elementary: 78.35%, Junior High: 56.07%, Senior High: 40.31%). Furthermore, the clustering model achieved strong evaluation results, reflected by a Silhouette Score of 0.8463, a Davies-Bouldin Index of 0.2478, and a Calinski-Harabasz Index of 324.8115, indicating high cluster quality and separation. Overall, the outcomes of this study are expected to provide valuable insights for policymakers in designing more effective and targeted educational strategies tailored to regional characteristics across Indonesia.

Keywords: K-Means, Clustering, Pure Participation Rate, Education, Indonesia.

Abstrak

Angka Partisipasi Murni (APM) merupakan indikator penting untuk menilai tingkat partisipasi masyarakat dalam sistem pendidikan di Indonesia pada berbagai jenjang pendidikan. Penelitian ini bertujuan untuk mengelompokkan 38 provinsi di Indonesia ke dalam beberapa klaster berdasarkan nilai APM pada tiga jenjang pendidikan, yaitu sekolah dasar/sederajat, sekolah menengah pertama/sederajat, dan sekolah menengah atas/sederajat, dengan menggunakan metode klasterisasi K-Means. Dataset yang digunakan dalam penelitian ini diperoleh dari Statistics Indonesia tahun 2025. Proses penelitian melibatkan beberapa tahapan, meliputi prapemrosesan data, normalisasi menggunakan teknik StandardScaler, identifikasi jumlah klaster optimal melalui Elbow Method, serta evaluasi performa klasterisasi menggunakan Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index. Hasil penelitian menunjukkan bahwa provinsi-provinsi di Indonesia secara optimal terbagi ke dalam dua klaster utama. Klaster 0 mencakup 32 provinsi yang ditandai dengan nilai APM relatif tinggi (SD: 96.53%, SMP: 79.16%, SMA: 66.49%), sedangkan Klaster 1 terdiri atas enam provinsi di wilayah Papua yang menunjukkan tingkat partisipasi pendidikan relatif lebih rendah (SD: 78.35%, SMP: 56.07%, SMA: 40.31%). Selain itu, model klasterisasi menghasilkan performa evaluasi yang sangat baik, ditunjukkan oleh nilai Silhouette Score sebesar 0,8463, Davies-Bouldin Index sebesar 0,2478, dan Calinski-Harabasz Index sebesar 324,8115, yang mengindikasikan kualitas klaster dan tingkat pemisahan antarkelompok yang tinggi. Secara keseluruhan, hasil penelitian ini diharapkan dapat memberikan wawasan yang bermanfaat bagi para pembuat kebijakan dalam merancang strategi pendidikan yang lebih efektif, tepat sasaran, dan sesuai dengan karakteristik wilayah di Indonesia.

Kata Kunci : K-Means, Clustering, Angka Partisipasi Murni, Pendidikan, Indonesia.

INTRODUCTION

Education is one of the main pillars in the development of a nation. The Pure Participation Rate (PPR), also known in Indonesia as Angka Partisipasi Murni (APM), functions as an important parameter to indicate the percentage of students in a specific age group participating in educational programs appropriate to their age level (Statistik, 2025). Not all regions have equality in their education sector, therefore special attention is needed from the government for areas considered to have unequal education (Manik & Ariesta, 2023). Classification and categorization of education levels in a region will make it easier for the government to determine which areas need more attention in the implementation of education (Manik & Ariesta, 2023).

Based on data from the Central Bureau of Statistics for 2025, there are significant variations in PPR between provinces in Indonesia at various education levels. The disparity in education quality in Indonesia occurs due to inequalities in access to educational facilities, differences in regional economic situations, and the not yet optimal distribution and quality of teaching staff (Ardianti et al., 2025). The inequality in education quality is clearly visible in isolated areas that face various obstacles ranging from minimal accessibility and educational infrastructure, challenging geographical locations, to socio-cultural factors that hinder progress (Abduh et al., 2022).

The K-Means clustering approach is a data mining method that has proven efficient in classifying data according to the similarity of attributes possessed. The working principle of the K-Means algorithm is by calculating the Euclidean distance from each data point to the cluster center, then allocating each data point to the cluster with the nearest center (Hardiani, 2022). Previous research shows that K-Means has been successfully applied to classify the quality of education management in Indonesia (Mustakim & Kamal, 2021), Education institution performance indicator (Fatmawaty et al., 2024), and cluster higher education data (Marthasari, 2021).

Several previous studies have successfully applied the K-Means method to analyze secondary education management parameters in Indonesia, map national educational primary data, and perform clustering in higher education institutions. However, most of these studies still rely on outdated datasets collected prior to the administrative expansion of provinces in Indonesia. The research gap addressed by this study lies in the utilization of the most recent 2025 spatial education data released by the Statistics Indonesia, which incorporates the latest administrative division of 38 provinces. This approach is crucial for capturing the actual educational disparities following regional expansion, particularly in the Papua region, thereby enabling this study to provide a more precise and contextually relevant foundation for policymaking.

Despite their contributions, prior studies exhibit notable limitations that constrain the comprehensiveness of their findings. Studies such as (Mustakim & Kamal, 2021) and (Manik & Ariesta, 2023) were conducted using provincial configurations predating the 2022–2023 administrative expansions, thereby failing to reflect the educational reality of the newly established Papua provinces. Furthermore, these studies generally adopted a single-metric evaluation approach, which limits the reliability of the clustering validity assessment. The absence of multi-level comparative analysis across educational stages also restricts the depth of policy-relevant insights, particularly regarding the progression of educational disparities from elementary to senior secondary levels.

The K-Means algorithm was selected for this study for several methodological reasons. First, K-Means is computationally efficient and well-suited for numerical datasets with continuous variables, such as PPR values across multiple educational levels. Second, the algorithm's iterative centroid-based mechanism allows for clear and interpretable cluster boundaries, which is particularly valuable for policy communication purposes. Third, K-Means has been extensively validated in educational data mining contexts within Indonesia and internationally, providing a reliable benchmark for comparative analysis. While hierarchical or density-based alternatives exist, K-Means offers superior scalability and reproducibility for datasets of the size and structure employed in this study.

This research aims to classify provinces in Indonesia based on The Pure Participation Rate (PPR) using the K-Means clustering method and to analyze the characteristics of each resulting cluster. The findings are expected to provide valuable insights for policymakers in formulating strategies to improve and equalize the quality of education across Indonesia.

METHODS

Figure 1 represents the research workflow involving key stages, beginning with the collection of The Pure Participation Rate (PPR) data, followed by data cleaning and normalization, determination of the optimal number of clusters, implementation of the K-Means algorithm, and evaluation of clustering results alongside final outcome analysis. The data clustering process was implemented using scientific Python libraries, with the K-Means algorithm parameters rigidly configured to ensure experimental reproducibility, namely $(k = 2)$, cluster centroid initialization using init='k-means++', a maximum iteration limit of max_iter=300, and a fixed random seed random_state=42. The quality of the clustering output was assessed using three formal mathematical metrics: the Silhouette Score (to measure inter-cluster separation), the Davies–Bouldin Index (to evaluate the ratio of intra-cluster distance to inter-cluster distance), and the Calinski–Harabasz Index (to measure the ratio of between-cluster variance to within-cluster variance).

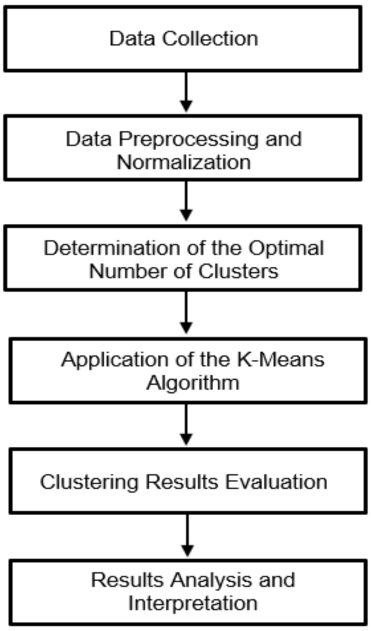

Figure 1. Research Flow
Source: Research Documentation

The research began with the collection of Pure Participation Rate (PPR) data from 38 provinces in Indonesia for elementary/equivalent, junior high/equivalent, and senior high/equivalent education levels in 2025 through the official portal of the Central Bureau of Statistics (BPS). After the data were obtained, a cleaning process was conducted to ensure the absence of missing or invalid values, thereby improving data accuracy and consistency. The StandardScaler preprocessing technique was selected because PPR values across different educational levels, while measured on similar scales, may exhibit varying degrees of variance; standardization ensures that no single variable disproportionately dominates the distance calculations during clustering. Subsequently, the dataset underwent normalization using the StandardScaler method to standardize the scale of all variables and prevent any single variable from exerting excessive influence during the clustering process. The determination of the optimal number of clusters was then carried out using the Elbow Method by calculating the Within-Cluster Sum of Squares (WCSS) for several k values, where the elbow point on the graph served as an indicator of the most appropriate cluster number. Following this stage, the K-Means algorithm was applied to the normalized data to classify provinces according to the similarity of their PPR values. To evaluate the quality and validity of the clustering results, three evaluation metrics—Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index—were employed to assess cluster cohesion and separation. Finally, the characteristics of each generated cluster were analyzed, and the clustering results were interpreted within the context of Indonesia's education sector to provide insights and formulate relevant policy recommendations aimed at improving educational participation across provinces.

Data Source

The data used in this research is The Pure Participation Rate (PPR) data for provinces in Indonesia for 2025 sourced from the Central Bureau of Statistics (BPS) (Statistik, 2025). The data covers 38 provinces with three PPR variables, namely elementary/equivalent, junior high/equivalent, and senior high/equivalent levels. The dataset consists of 38 records with 4 attributes: province name, Elementary PPR, Junior High PPR, and Senior High PPR.

Table 1. Pure Participation Rate (PPR) Data by Province and Education Level

No	Province	Elementary PPR	Junior High PPR	Senior High PPR
1	Aceh	96.32	81.37	70.36
2	North Sumatra	95.99	77.41	70.40
3	West Sumatra	97.06	82.70	71.57
4	Riau	97.54	81.40	65.22
5	Jambi	97.02	78.47	63.05
6	South Sumatra	95.98	74.27	63.08
7	Bengkulu	96.12	78.83	69.32
8	Lampung	97.64	78.47	67.22
9	Bangka Belitung Islands	97.00	74.06	65.94
10	Riau Islands	96.60	78.52	67.94
11	Jakarta Special Capital Region	97.60	82.65	71.07
12	West Java	96.74	78.80	66.13
13	Central Java	97.34	80.01	67.24
14	Special Region of Yogyakarta	97.48	83.46	72.34
15	East Java	97.12	79.92	66.89
16	Banten	96.59	77.66	66.34
17	Bali	97.52	82.75	72.39
18	West Nusa Tenggara	96.16	75.48	62.19
19	East Nusa Tenggara	94.76	71.83	57.54
20	West Kalimantan	96.00	76.10	60.82
21	Central Kalimantan	96.46	77.80	65.59
22	South Kalimantan	96.97	81.11	68.01
23	East Kalimantan	96.88	79.61	68.52
24	North Kalimantan	96.50	77.57	64.65
25	North Sulawesi	96.74	81.76	72.77
26	Central Sulawesi	95.22	76.67	62.22

No	Province	Elementary PPR	Junior High PPR	Senior High PPR
27	South Sulawesi	96.54	79.82	68.67
28	Southeast Sulawesi	95.76	78.09	64.86
29	Gorontalo	96.26	78.43	66.89
30	West Sulawesi	95.40	75.47	60.08
31	Maluku	95.44	77.81	63.46
32	North Maluku	95.53	76.47	61.14
33	West Papua	81.81	61.18	45.08
34	Papua	75.95	52.88	37.02
35	Central Papua	76.09	53.40	38.42
36	Highland Papua	76.09	53.49	39.06
37	South Papua	82.94	61.51	44.40
38	Southwest Papua	80.61	58.98	41.55

Source: Central Bureau of Statistics, Pure Participation Rate (PPR) by Province and Education Level 2025

Kmeans Clustering

K-Means Clustering is one of the unsupervised machine learning algorithms designed to classify data into several clusters according to the level of similarity among data characteristics. The main purpose of this method is to divide a dataset into relatively homogeneous groups, in which data points within the same cluster possess greater similarity, while maintaining clear distinctions between different clusters (Wala & Umar, 2024; Yulisasih et al., 2024). In its implementation, the K-Means algorithm works by determining the centroid or center point of each cluster and subsequently assigning every data point to the nearest centroid based on a distance calculation, with Euclidean Distance being the most commonly applied metric for measuring similarity between observations (Das et al., 2023; Qi et al., 2023). Through this mechanism, K-Means enables data grouping to be performed systematically and efficiently, making it widely applicable in various fields for pattern recognition and data segmentation purposes.

The mathematical formulas applied in this study are defined as follows:

a) Euclidean Distance is used to measure the distance between each data point and the cluster centroid (Black, 2025). The formula is:

$$d(x_i, \mu_j) = \sqrt{\sum_{k=1}^{n} \left(x_{ik} - \mu_{jk}\right)^2}$$

where x_i is a data point, μ_j is the centroid of cluster j, and x_{ik} and μ_{jk} are their respective k-th feature values.

b) Silhouette Score measures how similar an object is to its own cluster compared to other clusters (Choi et al., 2019). The formula is:

$$s(i) = \frac{b(i) - a(i)}{\max\left(a(i), b(i)\right)}$$

where $a(i)$ is the mean intra-cluster distance and $b(i)$ is the mean nearest-cluster distance. The score ranges from -1 to 1, with higher values indicating better-defined clusters.

c) Davies-Bouldin Index (DBI) evaluates the average similarity ratio between clusters. A lower DBI indicates better clustering (Alom et al., 2022). The formula is:

$$DBI = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d\left(c_i, c_j\right)} \right)$$

where σ_i and σ_j denote the average distances within clusters i and j respectively, and $d(c_i, c_j)$ is the distance between their centroids.

d) Calinski-Harabasz Index (CHI) measures the ratio of between-cluster dispersion to within-cluster dispersion (Hilmani et al., 2025). Higher values indicate better-separated clusters. The formula is:

$$CHI = \frac{\frac{SSB}{k-1}}{\frac{SSW}{n-k}}$$

where SSB is the between-cluster sum of squares, SSW is the within-cluster sum of squares, k is the number of clusters, and n is the total number of data points.

RESULTS AND DISCUSSION

This section presents the results of the K-Means clustering analysis on provincial Pure Participation Rate (PPR) data in Indonesia, followed by an in-depth discussion of the findings. The analysis successfully identified distinct patterns of educational participation across Indonesian provinces. Initial testing using the Elbow Method revealed a decline in the inertia value **(**Within-Cluster Sum of Squares/WCSS**)** that began to level off significantly at $(k = 2)$. This finding was further validated by the Silhouette Score graph, which demonstrated the peak coefficient value at the same number of clusters. The clustering performance evaluation, based on the interaction of three evaluation metrics, is transparently presented in Table 2 below.

Table 2. Evaluation Metrics of K-Means Model Quality (k = 2)

Evaluation Metric	**Exact Value**	**Academic Interpretation**
Silhouette Score	0.8463	Close to +1.00, indicating a highly robust cluster structure and very clear separation between groups.
Davies–Bouldin Index (DBI)	0.2478	A low value (close to 0), indicating high internal cohesion among data points and well-separated clusters.
Calinski–Harabasz Index (CHI)	324.8115	A high value, demonstrating that between-cluster variance is substantially greater than within-cluster variance.

Source: Research Documentation

After the model was optimally executed, the dataset of 38 provinces in Indonesia was successfully divided into distinct educational performance characteristics. The interpretative structure of the clusters is presented in Table 3 below.

Table 3. Characteristics of Average PPR Values and Policy Implications by Cluster

Cluster	**Number of Members**	**Average Primary School PPR**	**Average Junior High School PPR**	**Average Senior High School PPR**	**Policy Implications**
Cluster 0 (High PPR)	32 Provinces	96.53%	79.16%	66.49%	Maintenance of instructional quality, digitalization of learning materials, and enhancement of school laboratory capacity.

Cluster	Number of Members	Average Primary School PPR	Average Junior High School PPR	Average Senior High School PPR	Policy Implications
Cluster 1 (Low PPR)	6 Provinces	78.35%	56.07%	40.31%	High-priority affirmative interventions, expansion of scholarship programs, improvement of transportation access, and physical infrastructure development in remote, frontier, and underdeveloped (3T) areas.

Source: Research Documentation

Table 3 interprets the characteristics of each cluster based on PPR values. The high PPR group indicates the need for maintaining and developing education quality, while the low PPR group requires priority intervention to address education disparities at Junior High and Senior High levels.

In-Depth Discussion of Regional Disparities

Based on the spatial mapping results, Cluster 1 exclusively grouped six newly established provinces in the Papua region, namely West Papua, Papua, Central Papua, Highland Papua, South Papua, and Southwest Papua. The average Senior High School Pure Participation Rate (PPR) in Cluster 1, which reached only 40.31%**,** reflects a striking portrait of educational inequality compared to Cluster 0, which encompasses the remaining 32 provinces and achieved an average senior high school PPR of 66.49%**.**

This statistical deviation phenomenon confirms the theoretical foundation proposed by Ardianti et al. (2025), in which disparities in educational quality in Indonesia are driven by limited physical accessibility to educational institutions and the macroeconomic conditions of local regions. These barriers are further exacerbated in the Papua cluster due to geographical challenges, including extreme mountainous terrain, limited availability of transportation infrastructure, and the unequal distribution of qualified teachers in isolated areas, consistent with the empirical findings of Abduh et al. (2022). As a consequence, student retention declines substantially as educational levels increase, dropping from 78.35% at the primary school level to only 40.31% at the senior high school level. These findings emphasize that national educational equity policies cannot be uniformly implemented through generalized regulations; instead, they require an affirmative, cluster-based regional approach tailored to the specific characteristics and constraints of each region.

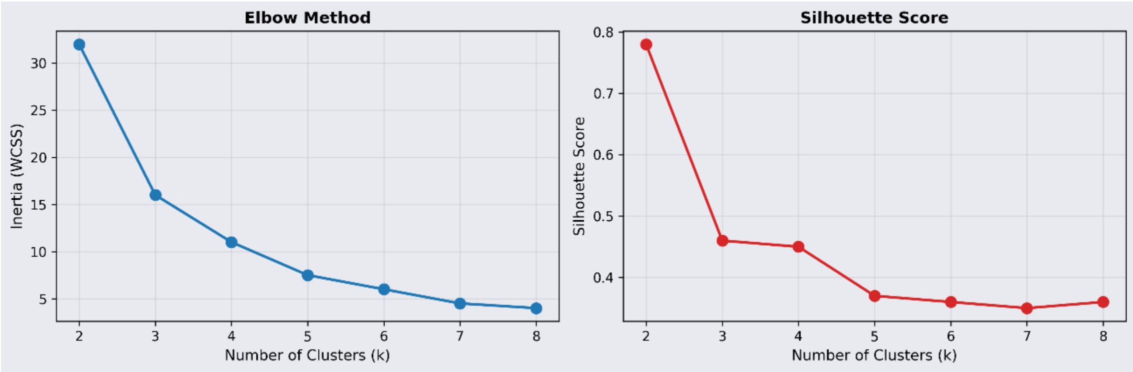

Figure 2. Elbow Method Graph for Optimal Cluster Determination (k = 2)
Source: Research Documentation

Figure 2 shows the use of the Elbow Method to determine the optimal number of clusters. The elbow point location on the graph indicates that the optimal number of clusters is 2, which demonstrates efficient separation between provinces with high and low PPR.

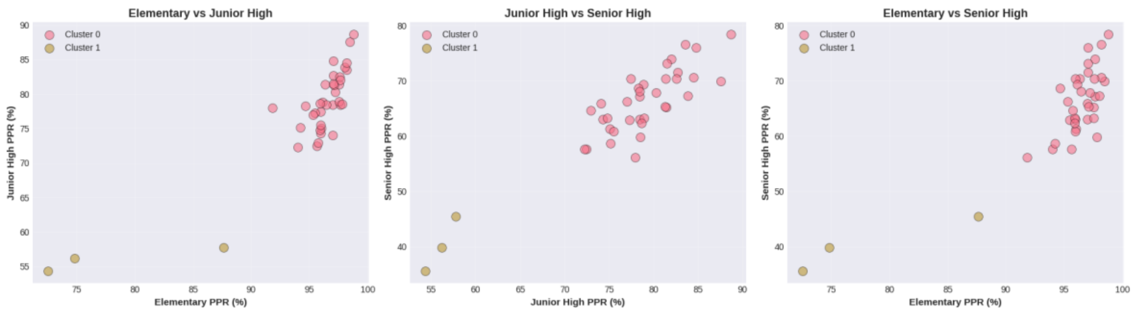

Figure 3. PPR Clustering Results Scatter Plot
Source: Research Documentation

Figure 3 shows the clustering results of provinces in Indonesia based on their PPR values. This visual representation demonstrates clear separation between the two groups, with high PPR and low PPR provinces well classified.

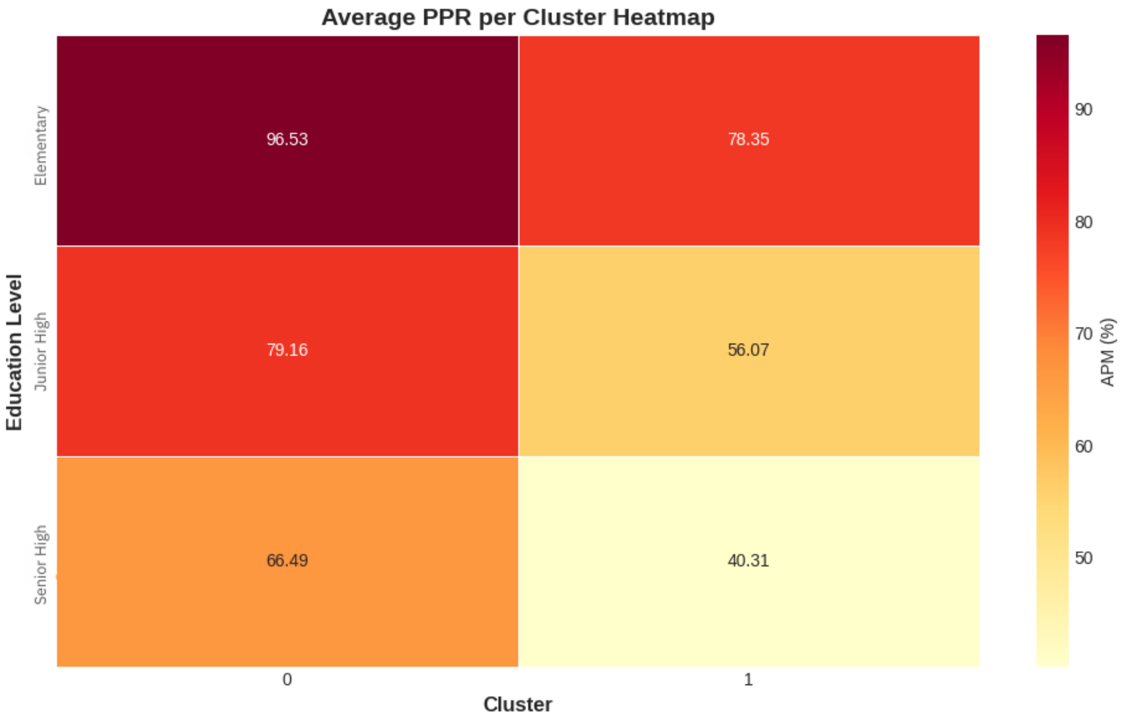

Figure 4. Cluster Characteristics Heatmap
Source: Research Documentation

Figure 4 displays PPR distribution across cluster characteristics. Each cluster reveals a distinct and interpretable pattern. In Cluster 0, PPR values remain relatively stable across educational levels, with a moderate decline from the elementary level (96.53%) to the junior high level (79.16%) and a further reduction at the senior high level (66.49%), suggesting that while some dropout pressure exists, educational retention remains broadly manageable. In contrast, Cluster 1 exhibits a pronounced and accelerating decline in PPR as educational level increases, falling from 78.35% at elementary to 56.07% at junior high and dropping sharply to 40.31% at senior high school level. This pattern indicates that educational retention in the Papua cluster deteriorates not just in absolute terms, but at an increasing rate with each successive educational stage. The widening gap between Cluster 0 and Cluster 1 across educational levels, from approximately 17.6 percentage points at elementary to 25.5 percentage points at senior high, underscores that access and retention challenges become progressively more severe in remote regions as students advance in their education. From a policy standpoint, this implies that interventions in the Papua cluster must not only address entry-level access barriers but also the structural impediments to educational continuation, including transportation infrastructure, teacher availability, and economic support for families.

CONCLUSIONS

This study successfully implemented the K-Means clustering algorithm (k = 2) to map educational participation disparities across 38 Indonesian provinces using the 2025 Pure Participation Rate (PPR) dataset. Validated by high-quality metrics—including a Silhouette Score of 0.8463, a Davies-Bouldin Index of 0.2478, and a Calinski-Harabasz Index of 324.8115 the model clearly separated 32 provinces into a high-performing cluster (Cluster 0) and 6 newly expanded provinces in the Papua region into a low-participation cluster (Cluster 1). Cluster 1 exhibits a severe, accelerating drop in educational retention, plunging from a 78.35% elementary school PPR to a striking 40.31% at the senior high school level, which underscores how the 2022–2023 administrative expansions have exposed critical, localized vulnerabilities previously hidden in aggregate data. These findings provide an empirical, data-driven foundation for policymakers, highlighting that addressing Indonesia's educational divide requires shifting away from generalized frameworks toward urgent, cluster-based affirmative interventions such as expanding secondary school infrastructure, deploying specialized teachers, and enhancing economic support specifically tailored to the geographic and structural realities of the Papua region.

Acknowledgment
The authors would like to thank Universitas Royal for supporting this research, and the Central Bureau of Statistics (BPS) for providing the open-access dataset.

Conflict of Interest
The authors declare no conflict of interest regarding the publication of this manuscript.

Author Contribution
E.D.B.S. conceived the research idea and coordinated the project; D.A.I. and S.A. performed data preprocessing and model implementation; D.P.W. and S.N. analyzed the evaluation metrics and drafted the manuscript. All authors reviewed and approved the final manuscript.

REFERENCES

Abduh, M., Basiru, A. A., Narayana, M. W., Safitri, N., & Fauzi, R. (2022). Potret Pendidikan di Daerah Terpencil Kampung Manceri Cigudeg Kabupaten Bogor. *Jurnal Citizenship Virtues*, *2*(1), 291–300.

Alom, R., Mazumdar, A., Prasad, R. K., Basumatary, G., & Baruah, B. (2022). Analysis of Seismic Data Using Partition-Based Clustering Techniques. *2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT)*, 1–6. https://doi.org/10.1109/GlobConPT57482.2022.9938362

Ardianti, R., Memi, & Lestari, A. (2025). KETIMPANGAN PENDIDIKAN DI INDONESIA: KAJIAN LITERATUR DAN WAWANCARA. *Global Research and Innovation Journal*, *01*(02), 65–70.

Black, D. (2025). Accelerating Euclidean Distance Transforms : A Fast and Flexible Approach With Multi-Vendor Multi-Language Support. *IEEE Access*, *13*(February), 44636–44649. https://doi.org/10.1109/ACCESS.2025.3548563

Choi, H. W., Qureshi, N. M. F., & Shin, D. R. (2019). Comparative Analysis of Electricity Consumption at Home through a Silhouette-score prospective. *2019 21st International Conference on Advanced Communication Technology (ICACT)*, 589–591. https://doi.org/10.23919/ICACT.2019.8701923

Das, D., Kayal, P., & Maiti, M. (2023). A K-means clustering model for analyzing the Bitcoin extreme value returns. *Decision Analytics Journal*, *6*, 100152. https://doi.org/10.1016/j.dajour.2022.100152

Fatmawaty, V. S., Riadi, I., & Herman, H. (2024). Higher Education Institution Clustering Based on Key Performance Indicators using Quartile Binning Method. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *24*(1), 141–154. https://doi.org/10.30812/matrik.v24i1.4244

Hardiani, T. (2022). ANALISIS CLUSTERING KASUS COVID 19 DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS. *Jurnal Nasional Pendidikan Teknik Informatika*, *11*(2), 156–165.

Hilmani, A., Fouad, A., Sabri, Y., Maizate, A., & Aouad, S. (2025). Adaptive K-Means Clustering for Energy Optimization in Wireless Sensor Networks Using Calinski-Harabasz Index. *2025 8th International Conference on Advanced Communication Technologies and Networking*

(CommNet), 1–5. https://doi.org/10.1109/CommNet68224.2025.11288894

Manik, R. A., & Ariesta, A. (2023). Data Mining Untuk Klasterisasi Provinsi Di Indonesia Berdasarkan Data Pokok Pendidikan Nasional. *Jurnal TICOM: Technology of Information and Communication*, *11*(3), 159–164.

Marthasari, G. I. (2021). ANALISIS DATA PENDIDIKAN TINGGI MENGGUNAKAN PENDEKATAN DATA MINING. *Jurnal Simantec*, *5*(3), 165–172.

Mustakim, Z., & Kamal, R. (2021). K-MEANS CLUSTERING FOR CLASSIFYING THE QUALITY MANAGEMENT OF SECONDARY EDUCATION IN INDONESIA. *Jurnal Cakrawala Pendidikan*, *40*(3), 725–737. https://doi.org/10.21831/cp.v40i3.40150

Qi, K.-T., Zhang, H.-S., Zheng, Y.-G., Zhang, Y., & Ding, L.-Y. (2023). Stripe segmentation of oceanic internal waves in SAR images based on Gabor transform and K-means clustering. *Oceanologia*, *65*(4), 548–555. https://doi.org/10.1016/j.oceano.2023.06.006

Statistik, B. P. (2025). *Angka Partisipasi Murni (APM) Menurut Provinsi dan Jenjang Pendidikan 2025*.

Wala, J., & Umar, R. (2024). *Implementasi K-Means Clustering pada Pengelompokan Pasien Penyakit Jantung*. *9*(3), 205–216.

Yulisasih, B. N., Herman, H., Sunardi, S., & Yuliansyah, H. (2024). Evaluation of K-Means Clustering Using Silhouette Score Method on Customer Segmentation. *ILKOM Jurnal Ilmiah; Vol 16, No 3 (2024)DO - 10.33096/Ilkom.V16i3.2325.330-342* . https://jurnal.fikom.umi.ac.id/index.php/ILKOM/article/view/2325