Jurnal Administrasi Negara
Volume 32 No. 1, Edition April 2026
p-ISSN: 1410-8399, e-ISSN: 2615-3424

SPATIAL ANALYSIS AND K-MEANS CLUSTERING IN MAPPING EDUCATIONAL PARTICIPATION INEQUALITY ACROSS PROVINCES IN INDONESIA

Misbahuddin [1] and Sutrisman Basir [2*]

[1, 2*] Public Administration, Makassar State University
email: misbha.misbahuddin5@gmail.com
email: sutrisman.basir@gmail.com

Article History
Submited: 14 February 2026
Review: 04 May 2026
Publish: 03 June 2026

Keywords:
Educational Participation; Spatial Analysis; K-Means Clustering; Education Policy; Educational Equity;

ABSTRACT

Equal access to education is an important indicator in human resource development and reducing regional inequality in Indonesia. Differences in geographical conditions, socio-economic factors, and availability of educational infrastructure may create disparities in educational participation across provinces. This study aims to analyze the pattern of educational participation inequality across provinces in Indonesia and to identify regional groupings based on education participation levels.This research employs a quantitative approach using provincial educational participation data by education level in 2025. The analysis was conducted through descriptive statistical analysis, spatial mapping, and regional clustering using the K-Means clustering method.The results show that participation at the basic education level is relatively evenly distributed, with an average participation rate of 98.06 percent at the primary school level and 94.66 percent at the junior secondary school level. However, participation declines at the senior secondary level to 76.75 percent and decreases more significantly at the higher education level to 29.48 percent. The clustering results identify three regional groups consisting of a high participation group with two provinces, a medium participation group with twenty-four provinces, and a relatively low participation group with twelve provinces. This study contributes theoretically by integrating spatial analysis and unsupervised machine learning in examining educational inequality, thereby advancing the application of computational approaches in public policy analysis. From a policy perspective, the findings provide an empirical basis for data-driven governance and support the formulation of place-based education policies aimed at reducing regional disparities in educational participation.

INTRODUCTION

Ensuring equitable educational access is a strategic instrument for enhancing the quality of human resources and strengthening Indonesia's national competitiveness. Within the long-term development framework toward the Golden Indonesia Vision 2045, the education sector is a primary pillar in creating a knowledgeable, productive, and globally competitive society. The success of educational development is measured not only by the availability of facilities, but also by the level of public participation in accessing educational services equitably across all regions. The School

* Correspondence Author
Email: sutrisman.basir@gmail.com

Enrollment Rate or SER (*Angka Partisipasi Sekolah*/APS) serves as a key indicator as it represents the affordability, accessibility, and inclusiveness of the national education system (Hidayat & Nuraini, 2022; Safira & Wibowo, 2021).

Nationally, the SER shows an upward trend across all school-age groups. However, this aggregate achievement does not fully reflect the real conditions across regions. Variations among provinces are influenced by differences in geographical characteristics, regional fiscal capacity, the quality of educational governance, and the availability of socioeconomic infrastructure. This disparity has the potential to widen the gap in human resource quality and hinder the integration of long-term national development (Ananda et al., 2025; Lestari & Kurniawan, 2023; Ningsih et al., 2022). These conditions underscore the importance of a region-based analytical approach to support the formulation of evidence-based educational policies (Hidayat & Sari, 2021).

The analytical approaches used thus far remain dominated by aggregative descriptive statistics and have been unable to capture the complexity of regional disparities. The use of national averages tends to mask extreme underdevelopment in specific areas. Furthermore, conducting analysis without considering spatial dimensions potentially yields biased conclusions, as it overlooks the influence of geographical proximity between regions (Putri, 2023; Rahmawati & Siregar, 2024). This limitation highlights the need for an analytical approach that simultaneously integrates both statistical and spatial dimensions.

Empirical studies that integrate geospatial approaches with data mining methods in educational policy analysis remain relatively limited. Most previous research has focused on trend analysis or simple regional comparisons without exploring multidimensional patterns (Ardianti et al., 2025; Wahyuni & Setiawan, 2024). This condition indicates a research gap in developing a more comprehensive region-based analytical model for educational disparity.

This gap underscores the need to develop an analytical approach capable of identifying patterns of disparity more objectively based on data. This study aims to analyze the disparity patterns of the School Enrollment Rate across provinces in Indonesia and identify regional clusterings based on educational participation levels using a spatial analysis approach and the K-Means clustering method. This approach enables the identification of geographical disparity distributions while simultaneously generating a regional typology based on educational participation characteristics.

This study provides a scientific contribution to the development of data-driven public administration studies through the integration of spatial analysis and machine learning methods in educational policy research. The resulting findings also offer an empirical foundation for the formulation of region-based policies (place-based policies) and support the implementation of data-driven governance to reduce disparities in educational participation across regions in Indonesia.

LITERATURE REVIEW

Strengthening the Public Administration Perspective on Region-Based Educational Equity

Contemporary public administration studies indicate that educational equity is understood not merely as the distribution of public services, but also as a strategic instrument for creating social justice and inclusive development. The state holds a constitutional obligation to guarantee educational access for all citizens, free from discrimination based on region, socioeconomic status, or demographic characteristics. This context positions education as a component of basic public services that possesses simultaneous social, economic, and political dimensions. Educational disparities across regions have the potential to weaken the legitimacy of public policy, as they can create inequities in the distribution of social welfare (Denhardt & Denhardt, 2021).

The New Public Service paradigm emphasizes that the primary orientation of public administration is not merely service efficiency, but also the creation of public value through the equitable fulfillment of societal needs. In the context of education, this positions the government not only as a provider of educational facilities, but also as an actor that ensures the sustainability of educational access across levels and regions. The equalization of educational participation is integral to long-term social development strategies, as education plays a vital role in enhancing social mobility and reducing economic disparities among the public (Denhardt & Denhardt, 2021).

The Good Governance framework emphasizes that the provision of educational services must meet the principles of transparency, accountability, effectiveness, and responsiveness to community needs. Educational equity is determined not only by the number of educational facilities but also by the effectiveness of the distribution of those services. Regions with limited transportation access, extreme geographical conditions, and low regional fiscal capacity often face constraints in providing advanced educational services. These conditions demonstrate that educational equity requires policy synergy across sectors within regional development (Osborne, 2023).

The Public Value concept positions education as a long-term social investment that has a direct impact on improving the community's quality of life. Educational equity produces not only outputs in the form of increased school enrollment rates, but also outcomes such as enhanced human resource quality, increased economic productivity, and strengthened social stability. This perspective emphasizes that education policy must be designed to focus not only on achieving statistical indicators, but also on creating sustainable social value for society (Moore, 2024).

Government decentralization context defines the distribution of educational services as a shared responsibility between the central government and regional governments. Regional governments hold the authority to manage educational resources in accordance with the unique characteristics of their respective territories. Disparities in regional fiscal capacity can influence the quality of educational service provision. Regions with high fiscal capacity tend to have a greater ability to provide quality educational facilities compared to regions with limited fiscal capacity. These conditions indicate that educational equity policies require adaptive fiscal intervention mechanisms based on regional needs (Pollitt & Bouckaert, 2022).

The digital government transformation era shows that the concept of data-driven governance has become an essential approach to improving the quality of public policy planning. The utilization of educational statistics, socioeconomic data, and spatial data enables the government to identify patterns of educational inequality more accurately. Integrating spatial data with statistical analysis allows for a more objective, evidence-based identification of priority areas for policy intervention. This approach reduces the risk of policy bias that relies solely on national averages without considering regional heterogeneity (Janssen & Van den Hoven, 2024).

The evidence-based policy approach emphasizes that public policy must be grounded in valid and reliable empirical data analysis. In the educational equity context, the utilization of spatial analysis and clustering methods enables a more accurate identification of educational regional typologies. This approach allows the government to develop differential policies based on regional characteristics, thereby making policy interventions more effective and efficient in reducing educational disparities (Head, 2023).

The development administration perspective positions education as a strategic sector for accelerating the socioeconomic transformation of society. Regional disparities in education can reinforce economic development inequality, as the quality of human resources is a primary factor in increasing regional economic productivity. Regions with low educational participation rates have the potential to experience long-term economic development delays. These conditions indicate that educational equity policies must become an integral part of national development strategies (Todaro & Smith, 2021).

Institutional capacity theory in public administration explains that the quality of public services is highly influenced by the capacity of government institutions to design and implement policies. In the educational context, institutional capacity encompasses the ability to plan policies, manage education budgets, and oversee the quality of educational services. Regions with strong institutional capacity tend to have better educational service quality compared to regions with limited institutional capacity (Grindle, 2022).

The concept of collaborative governance emphasizes the importance of stakeholder engagement in the provision of educational services. The government, private sector, civil society, and educational institutions play crucial roles in improving access to and quality of education. A collaborative approach enables the optimization of educational development resources, particularly in regions with limited fiscal capacity and educational infrastructure (Ansell & Gash, 2023).

The context of region-based public policy indicates that a place-based policy approach has become an important strategy for addressing development disparities. Region-based education policies allow for the design of policy interventions tailored to the geographical, socioeconomic, and demographic characteristics of a region. This approach is more effective than uniform national policies that do not consider differences in regional characteristics (Barca, 2024).

Strengthening the integration of spatial analysis into education policy planning also supports the concept of smart governance, where policy decision-making is backed by multidimensional data analysis. The use of analytical technology in evaluating education policy allows for real-time, adaptive monitoring of educational disparities in response to changes in the community's socioeconomic conditions (Kettl, 2023).

The sustainable development perspective positions educational equity as part of global development goals that emphasize the importance of inclusive and quality education for all people. Education serves as the foundation for supporting sustainable economic development, poverty reduction, and the improvement of community welfare. Regional educational disparities have the potential to hinder the achievement of sustainable development at the national level (UNDP, 2022).

Strengthening educational equity policies requires the integration of modern public administration approaches, spatial data analysis, and region-based policy approaches. The integration of these approaches enables the formulation of education policies that are more adaptive, responsive, and tailored to specific regional needs. This approach also supports the transformation of government governance toward a data-driven and evidence-based governance system (World Bank, 2025).

Therefore, this study offers novelty through the integration of spatial analysis and the K-Means algorithm in mapping disparities in School Enrollment Rates (SER) across provinces as a foundation for formulating region-based educational equity policies (World Bank, 2025).

Methodological Approach in Spatial Data-Based Educational Inequality Analysis

The development of research methodologies in public administration indicates a shift from conventional analytical approaches toward multidimensional data-driven analytical approaches. The complexity of public policy issues, particularly in the field of educational equity, demands the use of analytical methods capable of simultaneously integrating statistical, geographical, and socioeconomic dimensions. A quantitative approach based on statistical data becomes essential in supporting more objective and evidence-based public policy decision-making (Kim & Lee, 2020; Hughes, Giest, & Tozer, 2020).

In the study of educational inequality, descriptive statistical analysis is frequently used as an initial stage to outline the general condition of educational participation across regions. However, this approach has limitations in identifying spatial relationship patterns between regions. National mean values often fail to represent the heterogeneity of regional conditions accurately. Therefore, an analytical approach capable of integrating spatial dimensions into education policy analysis is required (Wang & vom Hofe, 2020; Kovacs-Györi et al., 2020).

Spatial analysis has become one of the methodological approaches widely utilized in region-based public policy studies. This approach enables the identification of geographical distribution patterns of social phenomena as well as the proximity relationships between regions. In the educational context, spatial analysis can be used to identify regions with high or low levels of educational participation, as well as to identify concentration patterns of educational inequality across administrative regions (Wegmann, Schwalb-Willmann, & Dech, 2020; Lambert & Zanin, 2020).

The spatial analysis approach also possesses advantages in supporting the formulation of region-based policies. The visualization of thematic maps allows policymakers to understand the distribution of educational inequality more intuitively compared to using only statistical tables. This approach supports a more effective public policy communication process because it can present policy information in a visual format that is easily understood by stakeholders (Lambert & Zanin, 2020; Wang & vom Hofe, 2020).

In addition to spatial analysis, advancements in data mining and machine learning methods have contributed significantly to data-driven public policy analysis. Data mining methods enable the exploration of hidden patterns within multidimensional data that cannot be identified using conventional statistical methods. In the context of education policy, data mining methods can be used to identify regional typologies based on educational participation characteristics (Yusuf & Razi, 2025; Sari, Adzima, Sahila, & Khotimah, 2025).

One of the machine learning methods widely applied in public policy analysis is the K-Means Clustering algorithm. This method is an unsupervised learning technique that groups data based on the degree of similarity in characteristics among objects. In the analysis of educational inequality, the K-Means method enables the grouping of regions based on educational participation levels across multiple educational levels simultaneously (MacQueen, 1967; Saputro & Ulkhaq, 2026).

The advantage of the K-Means method lies in its capability to efficiently group large volumes of data and produce relatively stable cluster typologies. This method is widely used in public policy studies because it can objectively generate development region segmentations based on statistical data. In the context of education policy, the clustering results can be used as a foundation for determining priorities for policy interventions based on regional characteristics (Sari et al., 2025; Arfan & Pekei, 2025).

In its implementation, the K-Means method requires a data standardization process to avoid variable scale bias. Variables with larger scales have the potential to dominate the grouping results if a data normalization process is not conducted. Therefore, the data preprocessing stage becomes a vital part of ensuring the accuracy of the clustering analysis results (Han, Kamber, & Pei, 2022; Arfan & Pekei, 2025).

Determining the optimal number of clusters is a crucial step in K-Means analysis. The Elbow method is used to identify the optimal number of clusters based on changes in the within-cluster sum of squares value. The Silhouette method is used to evaluate cluster quality based on the level of similarity among members within the same cluster and the differences from other clusters. The combination of these two methods can enhance the validity of the clustering results (Rousseeuw, 1987; Schubert, 2022).

The integration of spatial analysis and K-Means clustering yields a more comprehensive region-based policy analysis approach. Spatial analysis provides an overview of the geographical distribution of educational phenomena, while clustering provides a classification of regional typologies based on data characteristics. The integration of these two approaches enables the development of region-based education policy models that are more accurate and adaptive to local conditions (Kovacs-Györi et al., 2020; Haryowidyatna, Ichikawa, & Fujita, 2025).

In the context of modern public administration, the use of multidimensional data-driven methodological approaches aligns with the concept of smart policy analysis. This approach emphasizes the importance of utilizing analytical technology to support the public policy decision-making process. The utilization of spatial analysis and machine learning in education policy supports the transformation of governance toward data-driven government systems (Kim & Lee, 2020; Hughes et al., 2020; Kovacs-Györi et al., 2020).

The use of a data-driven methodological approach also supports the enhancement of public policy accountability. Policies formulated based on empirical data analysis possess a higher level of legitimacy because they are grounded in objective evidence. In the context of educational equity, this approach enables the government to design more targeted policies in reducing educational participation disparities between regions (Hughes et al., 2020; Yusuf & Razi, 2025).

Furthermore, data-driven methodological approaches also support the efficiency of educational development resource allocation. The government can identify priority areas for policy intervention more accurately so that education budget allocation can be executed more effectively. This approach supports the principle of efficiency in modern public policy management (Wang & vom Hofe, 2020; Haryowidyatna et al., 2025).

Thus, the integration of spatial analysis and K-Means clustering serves as a relevant methodological approach in region-based educational inequality analysis. This approach can generate policy information that is more comprehensive, objective, and adaptive to the dynamics of regional development. Strengthening data-driven methodological approaches is an essential part of supporting the transformation of education policy toward public policies that are more grounded in empirical evidence and specific regional needs (Sari et al., 2025; Saputro & Ulkhaq, 2026; Arfan & Pekei, 2025).

In the development of contemporary public administration studies, a need has emerged to integrate data-driven analytical approaches with conceptual frameworks of public policy. The Computational Social Science approach offers opportunities to process large-scale social data and identify complex patterns that are not easily detected through conventional methods. Nonetheless, the integration between computational approaches and public administration studies, particularly in the context of education policy, remains relatively limited. This study positions itself as a bridge between the discipline of Public Administration and Computational Social Science approaches by integrating spatial analysis and machine learning-based clustering methods into education policy studies. This approach not only enriches the analytical methods within public administration studies but also expands the utilization of computational techniques in generating an evidence-based policy that is more adaptive to regional complexities. Consequently, this study contributes to driving the transformation of public policy analysis toward a more data-driven, integrative, and multidimensionally based approach (Lazer et al., 2020; Salganik, 2018; Kovacs-Györi et al., 2020).

RESEARCH METHODS

This study utilizes descriptive statistical analysis, exploratory spatial analysis, and K-Means clustering analysis as the primary methods for data processing. Descriptive statistical analysis is employed to outline the general condition of School Enrollment Rates across provinces through the measurement of mean, minimum, and maximum values, as well as standard deviation. This analysis aims to provide an initial overview of the variations in educational participation levels across regions, thereby enabling the identification of general trends in the national distribution of educational participation (Montgomery & Runger, 2020).

Exploratory spatial analysis is used to identify the geographical distribution patterns of educational participation across provinces. This method enables the identification of regional concentrations with both high and low educational participation rates, as well as the identification of potential geographical clustering patterns. Spatial analysis was chosen because of its ability to integrate geographical dimensions into public policy analysis, thereby generating more context-specific, region-based policy information. This approach is considered more comprehensive than conventional statistical analysis, which only displays numerical relationships without considering the aspect of geographical proximity (Anselin, 2019).

The K-Means clustering method is used to group provinces based on similarities in educational participation characteristics across multiple educational levels simultaneously. The K-Means method is an unsupervised learning technique that functions by grouping data based on the proximity distance between objects, thereby producing data groups that are homogeneous within clusters and heterogeneous between clusters. This method was chosen because of its ability to efficiently process large volumes of multidimensional data and objectively generate regional typologies based on data characteristics (Han et al., 2021).

The selection of the K-Means method is also based on its ability to produce development region segmentations that can be utilized as a foundation for formulating region-based public policies. In the context of education policy, the clustering results can be used to identify groups of regions with high, medium, and low levels of educational participation, thereby facilitating the government in determining priorities for education policy interventions in a more targeted manner (Jain, 2020).

The implementation of clustering analysis begins with a data standardization process to avoid scale bias among variables. Standardization is performed so that all variables carry equal weight in influencing the cluster formation process. This stage is crucial because variables with larger value ranges have the potential to dominate the clustering results if data normalization is not conducted (Hair et al., 2019).

The determination of the optimal number of clusters is carried out using the Elbow and Silhouette methods. The Elbow method is used to identify the optimal number of clusters based on changes in the within-cluster sum of squares value. The Silhouette method is used to evaluate the quality of the cluster results based on the level of similarity of objects within a single cluster and their differences from other clusters. The combination of these two methods enhances the validity of the clustering results, thereby producing a more representative regional grouping (Rousseeuw, 1987).

The integration of descriptive statistical analysis, spatial analysis, and K-Means clustering was chosen because it is capable of producing a comprehensive, region-based analysis of educational inequality. Statistical analysis provides a general overview of educational participation conditions. Spatial analysis illustrates the geographical distribution of educational participation. Clustering provides regional typologies based on educational participation characteristics. The integration of these three methods is capable of supporting data-driven and region-based education policy decision-making more effectively (Osborne, 2023). The use of a multidimensional data-driven analytical approach aligns with developments in modern public administration, which emphasize the use of evidence-based policy in public policy formulation. This approach enables the government to design policies that are more objective, adaptive, and tailored to specific regional needs, thereby enhancing the effectiveness of national educational equity policies (Head, 2022).

RESEARCH RESULTS

Patterns of Educational Participation Across Educational Levels

The results of the analysis indicate that School Enrollment Rates (SER) in Indonesia exhibit significantly different patterns across educational levels. At the primary education level, the SER is remarkably high and relatively evenly distributed across almost all regions. The mean SER reaches 98.06 percent at the Elementary School level and 94.66 percent at the Junior High School level. This condition reflects the success of the compulsory education policy in expanding access to primary education nationwide, accompanied by a relatively low level of spatial disparity.

A shift in patterns begins to emerge at the senior high school and higher education levels. The mean SER at the Senior High School level drops to 76.75 percent, while the Higher Education level experiences a more substantial decline down to 29.48 percent. This decline demonstrates that as the educational level increases, the influence of socioeconomic factors, regional accessibility, and the availability of educational facilities becomes more pronounced in determining the continuity of educational participation.

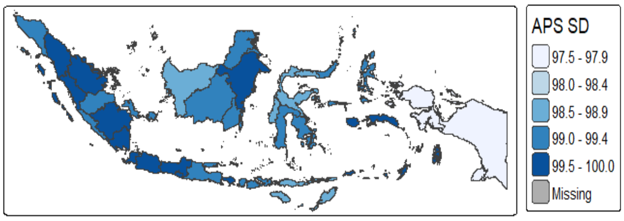

Figure 1. School Enrollment Rates at the Elementary School Level

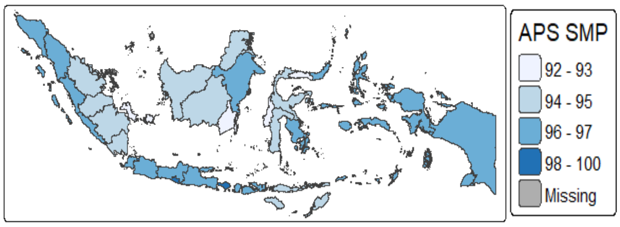

Figure 2. School Enrollment Rates at the Junior High School Level

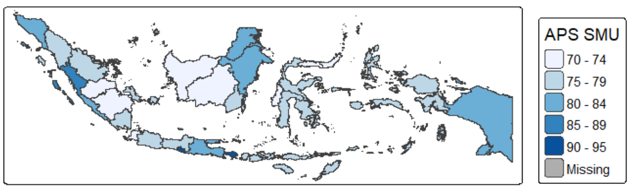

Figure 3. School Enrollment Rates at the Senior High School Level

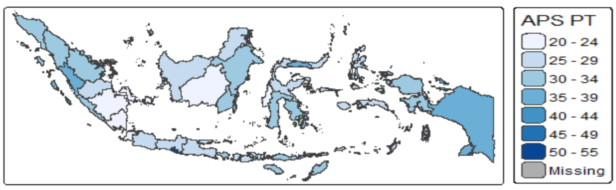

Figure 4. School Enrollment Rates at the Higher Education Level

Source: Data Processing Results using R-Studio

This spatial visualization not only illustrates the geographical distribution of educational participation but also reveals concentration patterns indicating that areas with high educational participation tend to accumulate in regions with advanced urbanization. These regions are supported by relatively comprehensive educational infrastructure, better accessibility to public services, and a socioeconomic ecosystem conducive to educational continuity. This condition reflects a close interrelation between regional development and educational participation outcomes, where infrastructural progress and local economic dynamics serve as leverage factors in enhancing access to and the continuity of education across various levels.

Regions facing limited transportation access, complex geographical conditions, such as archipelagic and remote areas, and relatively low community economic capacity tend to exhibit lower educational participation rates, particularly at advanced educational levels. These limitations not only impact physical access to educational facilities but also affect the households' ability to bear indirect educational costs, thereby cumulatively creating structural barriers to educational continuity.

This finding confirms that the critical juncture of educational inequality in Indonesia does not reside in access to primary education, but rather in the transition phase toward advanced education, where structural factors begin to play a more dominant role in determining the continuity of educational participation. Primary education has achieved a relatively well-distributed level of equity as a result of universal national policy interventions. However, the continuity toward senior high school and higher education levels is still confronted with multidimensional constraints, including limited geographical access, economic capacity disparities, and the uneven distribution of educational facilities.

The implication of this finding suggests that the future direction of education policy can no longer merely focus on expanding basic access. It must shift toward strengthening the educational transition system to guarantee the continuity of participation across levels. The policy focus needs to be directed at increasing access to advanced education through educational infrastructure development in underdeveloped regions, the provision of more inclusive educational financing schemes, and the enhancement of regional connectivity capable of mitigating geographical barriers. This approach is crucial, particularly for regions characterized by limited access and economic capacity, ensuring that education policy is not only formally equitable, but also substantively fair in addressing the specific needs of each region.

Determination of the Optimal Number of Clusters

The determination of the optimal number of clusters was conducted by combining the Elbow and Silhouette methods to obtain statistically more robust results. The evaluation results using the Elbow method indicate that a significant decline in the Within-Cluster Sum of Squares (WSS) value occurs up to the cluster count of k = 3. Past this point, the decline in the WSS value tends to flatten, indicating that the primary structure of the data has been optimally represented in three clusters.

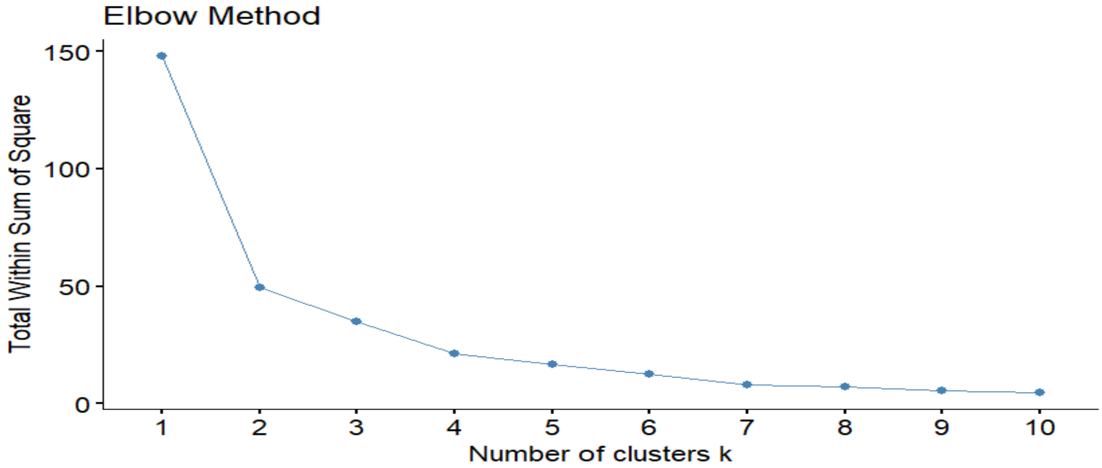

Figure 5. Elbow Method
Source: Data Processing Results using R-Studio

The visualization of the Elbow Method illustrates the relationship between the number of clusters (k) and the Within-Cluster Sum of Squares (WSS) value as an indicator of the degree of data homogeneity within each cluster. The curve pattern exhibits a sharp drop in WSS during the initial phase, particularly from k = 1 to k = 2, and continues significantly up to k = 3, indicating that the segmentation process at this stage is capable of substantially reducing within-cluster variation. This steep decline reflects a high degree of heterogeneity in the initial data structure. Therefore, partitioning it into the first few clusters yields a highly significant improvement in grouping quality. The model's capacity to capture the underlying structure of the data appears optimal as the number of clusters increases up to three, as most of the major variations have been successfully represented within this division.

The shift in the curve pattern, which begins to flatten after k = 3, indicates a phenomenon of diminishing returns, where adding more clusters no longer yields a substantially meaningful reduction in WSS. The WSS values from k = 4 to k = 10 experience relatively small and stable decreases, thereby demonstrating that an increase in model complexity is not accompanied by a proportional improvement in clustering quality. This condition signifies that the primary structure of the data has been successfully captured by the previous number of clusters, meaning that additional clusters merely contribute to splitting relatively homogeneous groups into smaller subgroups without providing significant analytical value. The most prominent change in the curve's gradient, or the elbow point, is identified at k = 3, which methodologically represents the optimal balance between the efficiency of variance reduction and the number of clusters utilized.

The existence of the optimal point at k = 3 suggests that the natural distribution structure of the data tends to divide into three primary groups capable of representing variations in educational participation rates more proportionally. Forming three clusters allows for a more comprehensive identification of regional categories with high, medium, and low characteristics compared to classifications that are either too simplistic or overly complex. The implication of this finding shows that the selection of the cluster count is grounded not only in mathematical considerations but also in the model's ability to generate meaningful and easily interpretable segmentations. This approach is highly relevant to public policy analysis, as a more representative regional segmentation enables the formulation of policies that are more adaptive, contextual, and based on empirical variations that reflect the actual conditions across regions.

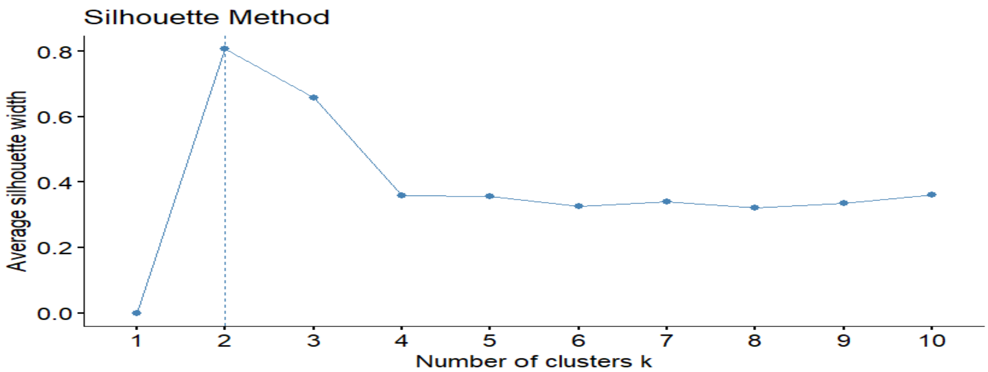

Figure 6. Silhouette Method
Source: Data Processing Results using R-Studio

The visualization of the Silhouette Method illustrates the relationship between the number of clusters (k) and the average silhouette width value as a measure of cluster separation quality, which considers the balance between intra-cluster cohesion and inter-cluster separation. The graph pattern demonstrates that the silhouette value increases sharply from k = 1 until it reaches its peak at k = 2 with a value approaching 0.8, indicating an excellent level of cluster separation. This high value shows that most observations lie very close to their assigned clusters and are significantly distant from other clusters, meaning the grouping structure at k = 2 can be considered the most distinct and mathematically optimal in separating the data.

A significant shift occurs after k = 2, where the silhouette value experiences a relatively sharp decline at k = 3 and continues to decrease, remaining within the range of approximately 0.3–0.4 for the subsequent values of k. This decline indicates that increasing the number of clusters begins to diminish the separation quality between groups, even though the cluster structure remains within a reasonable category (moderate structure). Values within this range imply that some observations begin to exhibit relative proximity to more than one cluster, causing the boundaries between clusters to become less defined compared to the two-cluster configuration. The stability of the silhouette values at k > 3, which tend to flatten, demonstrates that increasing the number of clusters no longer provides any meaningful improvement to the quality of the grouping structure.

The overall interpretation indicates that k = 2 is the statistically optimal number of clusters based on the silhouette criterion, as it yields the most distinct and consistent separation. However, the value at k = 3 still indicates an acceptable quality, making it viable for use within the context of complex and multidimensional social data analysis. The selection of k = 3 clusters in this study remains justified when considering the need for a richer interpretative capacity, specifically in identifying a middle/medium group that cannot be captured in a two-cluster division. This approach demonstrates that methodological decisions do not rely solely on numerical optimality but also consider the model's ability to generate segmentations that are more informative and relevant for public policy analysis.

Cluster Quality Evaluation

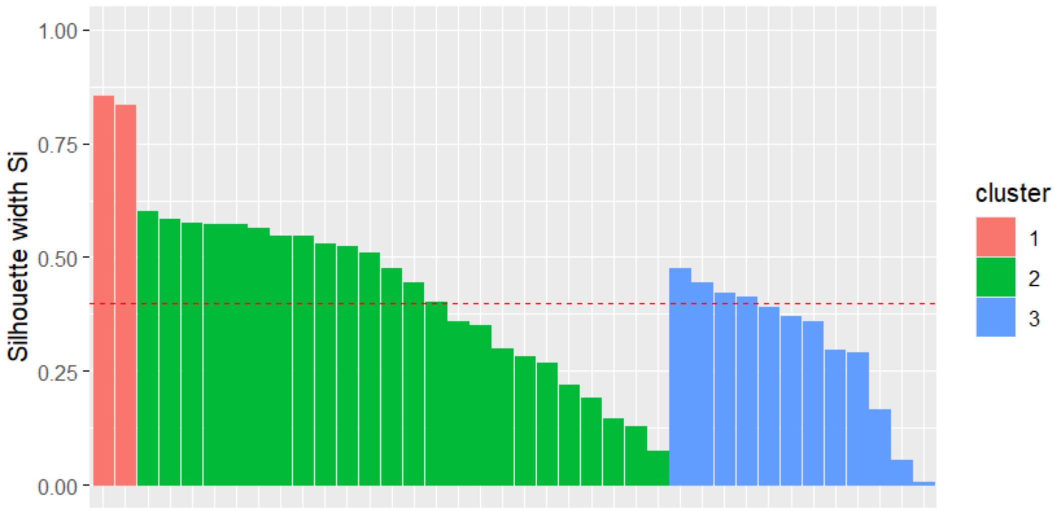

Figure 7. Cluster Quality Evaluation
Source: Data Processing Results using R-Studio

The cluster quality evaluation results show that the average Silhouette Score obtained is 0.398658, or approximately 0.40. This value falls within the range of 0.25 to 0.50, which is statistically categorized as a moderate or reasonably good cluster structure (moderate structure). This value indicates that the average proximity distance of objects within a single cluster is smaller than the distance of objects to other clusters, demonstrating that the separation process between clusters has been adequately formed.

A Silhouette Score approaching 0.40 indicates that the level of within-cluster homogeneity is in the moderate category, while the level of between-cluster heterogeneity is also at a reasonably distinct level. Numerically, this value implies that most objects have been grouped into their appropriate clusters, although there are still some objects that exhibit relative proximity to other clusters.

The obtained range of silhouette values demonstrates that the data grouping results have met the minimum clustering validity criteria for exploratory social data analysis. A value approaching 0.50 indicates a trend toward an increasingly better cluster structure, whereas a value approaching 0.25 represents the lower bound of a cluster structure that remains statistically acceptable.

Spatial Clustering Results of School Enrollment Rates (SER)

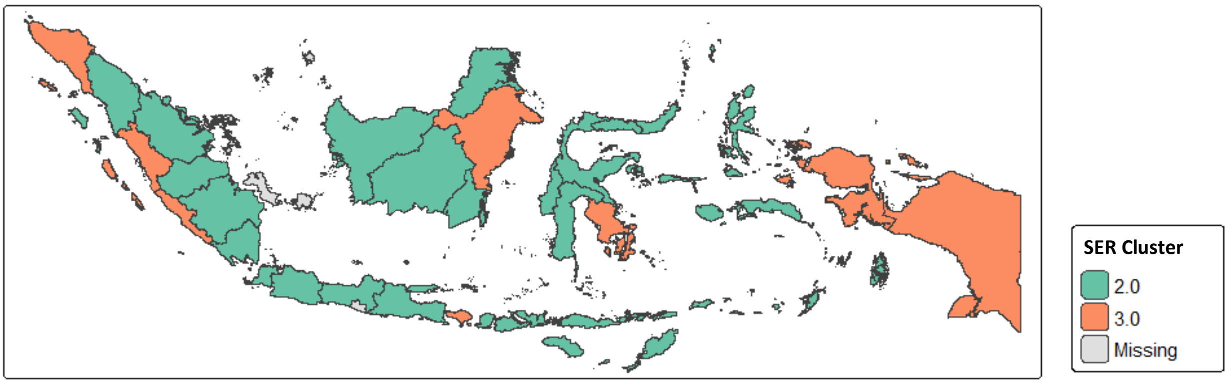

Figure 8. Spatial Clustering of School Enrollment Rates
Source: Data Processing Results using R-Studio

The spatial clustering results show that the distribution of provinces in Indonesia is significantly dominated by Cluster 2, which comprises 24 provinces or approximately 63.16 percent of the total 38 provinces. It is empirically representing the largest concentration in the structure of the national SER data grouping. This dominance reflects not only a quantitative proportion but also indicates that the majority of Indonesian regions share relatively similar educational participation characteristics and fall into the medium category within the SER value distribution. The dominant position of this cluster indicates a centralization trend in data distribution, where most provinces are concentrated within the middle-value range. This is implying that variation across regions is not entirely extreme but predominantly occurs within a moderate spectrum. This condition demonstrates that the senior high school and higher education levels serve as the primary factors driving SER value differentiation across provinces, while also acting as the main determinants in shaping the overall clustering structure.

Cluster 3 consists of 12 provinces or approximately 31.58 percent of the total provinces, proportionally indicating the extensive presence of a regional group with relatively lower educational participation rates on a fairly broad scale. This composition indicates that nearly one-third of Indonesia's regions still face constraints regarding educational access and continuity, particularly at advanced educational levels, thereby reflecting pockets of structural inequality that are geographically uneven. The characteristics of this cluster show that SER values fall below the national average, with the most pronounced declines occurring at the senior high school and higher education levels. These conceptually represent critical phases in educational continuity. The existence of this cluster confirms that disparities in educational participation are not confined to isolated regions but have formed a fairly systemic distribution pattern within the national educational development structure.

Cluster 1 constitutes the smallest group, with only 2 provinces or approximately 5.26 percent of the total provinces, demonstrating a very high level of exclusivity in the context of educational achievement distribution. This highly restricted proportion indicates that only a small fraction of regions is capable of consistently achieving very high educational participation rates across all educational levels. It is reflecting a concentration of optimal achievement in specific areas generally backed by robust educational infrastructure and more established socioeconomic conditions. Statistically, this cluster sits within the upper quartile of the national SER distribution. It shows that this group is not only relatively superior but also maintains a significant gap compared to other groups in terms of educational participation outcomes. The existence of this cluster further accentuates a vertical disparity in educational distribution, wherein a tiny fraction of regions can attain exceptionally high performance while the vast majority remain at medium or even low levels.

The distribution of cluster members reveals that the data grouping structure is asymmetrical, with a powerful dominance in the medium group, thereby reflecting an imbalance in the distribution of educational achievements across regions. The dominance of Cluster 2 demonstrates that the majority of provinces possess relatively homogeneous educational participation characteristics within the medium category. Groups with extreme characteristics, either exceptionally high or exceptionally low, comprise a far smaller proportion. This distribution pattern shows that variation in SER values occurs mostly within a moderate range, indicating that differences between regions tend to be gradual rather than sharply polarized. Such a structure also implies that the dynamics of educational inequality are determined not only by differences between extreme groups but also by internal variations within the medium group, which holds a substantial potential for stagnation or limited advancement.

The overall statistical distribution of the clusters indicates that the national SER data structure forms a centralized pattern around the medium group, with value dispersion moving progressively toward the high and low groups. This pattern indicates that educational participation inequality in Indonesia is not solely characterized by an extreme chasm between regions, but is more frequently manifested as widespread medium-level variations that form a complex distribution structure. These conditions suggest that the primary challenge in education policy lies not only in intervening within the lowest-performing regions but also in efforts to push capacity building in medium-group regions so they can transition toward higher categories. A policy approach sensitive to this distribution becomes crucial to ensure that educational development does not merely reduce extreme disparities but also enhances the quality and sustainability of educational participation as a whole.

DISCUSSION

The research findings demonstrate that the level of educational participation in Indonesia forms a multi-tiered and systemic pattern of inequality across educational levels. While participation at the primary education level is exceptionally high and relatively evenly distributed across almost all regions, the senior high school and higher education levels experience a significant decline accompanied by increasingly conspicuous spatial variations among provinces. This pattern of differentiation not only reflects the dynamics of educational access but also indicates a complex transformation in the determinants of educational participation as the educational level advances. Therefore, there is a robust empirical foundation to evaluate and recontextualize various theoretical frameworks within public administration and regional development studies (Hidayat & Nuraini, 2022; Ningsih et al., 2022).

The perspective of spatial inequality explains that the distribution of public services, including education, is never entirely geographically neutral. It is, however, influenced by the interaction between regional structural factors, regional fiscal capacity, and the development level of socioeconomic infrastructure. The high equity achieved in primary education, which fails to persist consistently into advanced educational levels, indicates that universal policy interventions possess limited effectiveness during the early stages of the educational cycle, whereas advanced stages demand a policy approach that is more contextual and adaptive to regional characteristics. This condition aligns with previous empirical findings demonstrating that the uneven distribution of public services in Indonesia remains heavily influenced by development heterogeneity across regions (Abdullah & Firmansyah, 2021; Rahmawati & Siregar, 2024). These findings explicitly reinforce the spatial inequality perspective within the context of educational service distribution.

The framework of core–periphery theory gains empirical support from the concentration patterns of higher educational participation in regions with advanced urbanization and economic activity. Regions functioning as economic growth centers tend to have broader access to educational infrastructure, including the availability of advanced schools and higher education institutions, thereby enabling them to maintain higher educational participation rates at the senior high school and higher education levels. Conversely, peripheral regions confront various structural constraints, such as geographical isolation, limited educational facilities, and low local economic opportunities, which cumulatively impact the low continuity of educational participation. This finding strengthens the argument that educational inequality cannot be decoupled from asymmetrical regional development dynamics (Iskandar & Putri, 2024; Santoso & Hadi, 2022).

The research findings offer a critical contribution by challenging the normative assumptions in education policy, which tend to assume that equitable access to education can be achieved through a uniform policy approach across all educational levels. Success at the primary education level does not automatically guarantee participation continuity at the next level, which is instead increasingly influenced by more complex factors, such as household socioeconomic conditions, indirect educational costs, geographical accessibility, and the local economic structure. Those factors determines individual incentives to pursue higher education. This finding underscores the necessity of shifting the education policy paradigm from an access-expansion approach toward strengthening the dimension of sustainability across educational levels (Lestari & Kurniawan, 2023; Pugu, 2025).

From a public administration standpoint, a data-driven governance approach is positioned as an essential instrument to support evidence-based and contextual policy formulation. Utilizing spatial analysis combined with clustering methods enables a more precise identification of regional typologies, thereby overcoming the limitations of aggregative approaches that frequently oversimplify the complexity of social realities. The integration of this approach reinforces the position of computational social science in public policy analysis, particularly in understanding social phenomena that are inherently multidimensional and non-linear (Handayani & Prabowo, 2023; Ningsih et al., 2022).

The utilization of machine learning methods, specifically clustering techniques, expands the analytical approach in public policy studies through its capacity to identify latent structures within the distribution of educational participation that cannot be captured by conventional statistical methods. This approach indicates that data mining techniques possess great potential in supporting more adaptive and evidence-based policy decision-making, particularly within the context of complex and heterogeneous regional development planning. This finding is in line with literature emphasizing the effectiveness of computational approaches in regional development and public policy analysis (Fadilah & Hartono, 2022; Firmansyah et al., 2023; Wahyuni & Setiawan, 2024).

The clustering results reveal that the majority of regions fall into the medium participation category, with significant variations at advanced educational levels. This distribution structure extends the understanding of educational inequality by demonstrating that disparities occur not only within extreme regional groups but also within the medium group, which holds the potential to experience stagnation or a decline in educational participation. This finding supports literature that emphasizes the importance of distributional analysis in understanding development inequality more comprehensively (Darmawan & Laksmi, 2023; Yuliana et al., 2023).

Policy implications point to the need for a paradigm shift in educational development from a basic access equity-based approach to an approach that emphasizes the sustainability of educational participation across levels. Policy interventions need to be focused on strengthening the educational transition from the primary level to senior high school and higher education by considering the spatial and socioeconomic characteristics of the region. A place-based policy approach becomes highly relevant because it can accommodate differences in regional characteristics to formulate more adaptive and contextual policies (Utami et al., 2023).

Ultimately, these research findings do not merely confirm the relevance of spatial inequality and regional development theories, but also expand the conceptual framework by demonstrating that educational inequality is dynamic, multi-tiered, and increasingly complex at higher educational levels. The integration of spatial analysis and machine learning methods indicates that data-driven approaches possess significant potential to enrich public policy analysis, specifically in generating policies that are more adaptive, evidence-based, and responsive to the complexities of regional development.

CONCLUSION

This study aims to analyze the disparities in School Enrollment Rates (SER) across provinces in Indonesia through spatial analysis and K-Means clustering approaches as part of strengthening data-driven public policy. The results of the analysis indicate that educational participation at the primary education level is relatively equitable, with a mean SER of 98.06 percent for Elementary Schools and 94.66 percent for Junior High Schools. A decline in educational participation occurs at the Senior High School level, with a mean of 76.75 percent, and drops more significantly at the Higher Education level, with a mean of 29.48 percent. This condition demonstrates that regional educational disparities are more dominantly prevalent at the senior high school and higher education levels. From a public administration perspective, this condition indicates that the equitable distribution of public services in advanced education is not yet optimal as part of the state's efforts to realize social justice and equitable human development.

The spatial analysis results show that regions in the medium educational participation category dominate most provinces in Indonesia, whereas regions with relatively low educational participation remain concentrated in specific areas. The clustering results reveal the formation of three regional typologies of educational participation: a high-participation cluster comprising 2 provinces, a medium-participation cluster with 24 provinces, and a relatively low-participation cluster with 12 provinces. An average Silhouette Score of 0.398658 demonstrates that the quality of the regional grouping falls within a reasonably good category. This finding confirms that a data-driven analytical approach can provide an objective overview of the typologies of regional educational inequality.

The public administration implications suggest that data-driven regional typology mapping can serve as a foundation for applying region-based education policies and supporting the implementation of evidence-based public policies in national educational development planning. This approach aligns with governance principles that emphasize the effectiveness, efficiency, accountability, and distributive justice of educational public services across regions.

Recommendations for future research are directed toward incorporating socioeconomic variables, utilizing advanced spatial analysis, and leveraging time-series data to analyze the dynamics of educational inequality shifts more comprehensively. Developing studies based on multidimensional data is expected to strengthen the formulation of more integrative educational policies, as well as support the reinforcement of sustainable and inclusive educational public service systems within the framework of national development. The objective is to analyze inequalities in School Enrollment Rates across provinces in Indonesia through a spatial analysis approach and K-Means clustering as part of strengthening data-driven public policy.

REFERENSI

Abdullah, M., & Firmansyah, D. (2021). Regional development inequality and public service distribution in Indonesia. *Jurnal Administrasi Publik Indonesia*, *7*(2), 145–158.

Ananda, R., Yuliani, C. E., Misnati, K., Azzikra, R., & Rahmadiah, P. (2025). Analisis kesenjangan layanan pendidikan sekolah dasar antara sekolah perkotaan dan daerah 3T di Indonesia [Analysis of primary school education service gaps between urban schools and 3T regions in Indonesia]. *Jurnal Pendas*, *10*(1), 87–98.

Anselin, L. (2019). *Spatial econometrics: Methods and models*. Springer.

Ansell, C., & Gash, A. (2023). *Collaborative governance and public policy implementation*. Oxford University Press.

Ardianti, R., Memi, M., & Lestari, A. (2025). Ketimpangan pendidikan di Indonesia: Kajian literatur dan wawancara [Educational inequality in Indonesia: A literature review and interviews]. *Global Research and Innovation Journal*, *1*(2), 65–70. *(Note: Assumed "Memi" is a first initial/name; adjust if it is a double surname).*

Barca, F. (2024). *Place-based development policies and regional inequality*. European Commission Press.

Darmawan, R., & Laksmi, N. (2023). Public service accessibility and human development disparity in Indonesian provinces. *Jurnal Ilmu Administrasi Negara*, *19*(1), 33–47.

Denhardt, R. B., & Denhardt, J. V. (2021). *The new public service: Serving, not steering* (Updated ed.). Routledge.

Fadilah, N., & Hartono, B. (2022). Machine learning approach in regional development clustering analysis. *Jurnal Statistika Terapan Indonesia*, *6*(1), 55–67.

Firmansyah, A., Nugroho, T., & Sari, M. (2023). Spatial data mining for regional policy planning in Indonesia. *Jurnal Geografi Dan Pembangunan Wilayah*, *8*(2), 102–116.

Grindle, M. S. (2022). *Good governance and institutional capacity in public sector reform*. Oxford University Press.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Han, J., Kamber, M., & Pei, J. (2021). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
Handayani, T., & Prabowo, H. (2023). Data-driven governance in public sector decision making. *Jurnal Kebijakan Dan Administrasi Publik*, *27*(1), 1–15.
Haryowidyatna, E., Ichikawa, M., & Fujita, K. (2025). Region-based recommendations for upper secondary school types in Indonesia using clustering analysis. *Proceedings of International Conference on Management in Emerging Markets*.
Head, B. W. (2022). Evidence-based policymaking: Principles and requirements. *Australian Journal of Public Administration*, *81*(3), 345–362.
Head, B. W. (2023). *Evidence-based public policy: Principles and practice*. Policy Press.
Hidayat, R., & Nuraini, S. (2022). Educational inequality and regional disparity in Indonesia. *Jurnal Pendidikan Dan Kebijakan Publik*, *14*(2), 120–134.
Hidayat, R., & Sari, N. (2021). The effectiveness of social assistance programs in reducing poverty in Indonesia. *Journal of Public Policy and Governance*, *11*(2), 89–102.
Hughes, S., Giest, S., & Tozer, L. (2020). Accountability and data-driven urban climate governance. *Nature Climate Change, 10*, 1085–1090.
Iskandar, D., & Putri, F. (2024). Infrastructure development and education accessibility in developing regions. *Jurnal Pembangunan Wilayah Dan Kota*, *20*(1), 44–56.
Jain, A. K. (2020). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666.
Janssen, M., & Van den Hoven, J. (2024). *Data-driven governance and public policy: Ethical and social implications*. Springer.
Kettl, D. F. (2023). *The transformation of governance: Public administration for the twenty-first century*. Johns Hopkins University Press.
Kim, S., & Lee, J. (2020). Do the right thing right! Understanding the hopes and hypes of data-based policy. *Government Information Quarterly, 37*(3), 101491.
Kovacs-Györi, A., Ristea, A., Havas, C., Mehaffy, M., Hochmair, H. H., Resch, B., Juhasz, L., Lehner, A., Ramasubramanian, L., & Blaschke, T. (2020). Opportunities and challenges of geospatial analysis for promoting urban livability in the era of big data and machine learning. *ISPRS International Journal of Geo-Information, 9*(12), 752.
Lambert, N., & Zanin, C. (2020). *Practical handbook of thematic cartography: Principles, methods, and applications*. CRC Press.
Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science, 369*(6507), 1060–1062.
Lestari, W., & Kurniawan, A. (2023). Regional fiscal capacity and education quality disparity in Indonesia. *Jurnal Ekonomi Pembangunan Indonesia*, *24*(1), 67–82.
MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.
Montgomery, D. C., & Runger, G. C. (2020). *Applied statistics and probability for engineers* (7th ed.). Wiley.
Moore, M. H. (2024). *Creating public value: Strategic management in government* (Revised ed.). Harvard University Press.
Ningsih, S., Prasetyo, A., & Ramadhan, R. (2022). Public policy data integration in digital governance era. *Jurnal Administrasi Publik*, *12*(3), 201–214.
Osborne, S. P. (2023). *Public service logic: Creating value for public service users, citizens, and society*. Routledge.
Pollitt, C., & Bouckaert, G. (2022). *Public management reform: A comparative analysis—Into the age of austerity* (4th ed.). Oxford University Press.

Pugu, M. R. (2025). Ketimpangan pendidikan dan mobilitas sosial di Indonesia [Educational inequality and social mobility in Indonesia]. *Jurnal Sosiologi Pendidikan*, *14*(3), 210–223.

Putri, N. A. (2023). Analisis minat belajar siswa terhadap pembelajaran matematika di sekolah menengah pertama [Analysis of student learning interest in mathematics learning in junior high school]. *Jurnal Pendidikan Matematika Dan Sains*, *11*(1), 21–30.

Rahmawati, D., & Siregar, H. (2024). Spatial inequality and public service distribution analysis. *Jurnal Perencanaan Wilayah Indonesia*, *9*(1), 15–29.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Safira, N., & Wibowo, Y. H. (2021). Angka partisipasi kasar dan angka partisipasi murni sebagai indikator keberhasilan pendidikan Indonesia [Gross enrollment rate and net enrollment rate as indicators of educational success in Indonesia]. *Jurnal PAKAR Pendidikan*, *19*(2), 101–115.

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton University Press.

Santoso, B., & Hadi, P. (2022). Urbanization and education participation gap in developing countries. *Jurnal Ilmu Sosial Dan Humaniora Indonesia*, *11*(2), 210–224.

Saputro, H. B., & Ulkhaq, M. M. (2026). Clustering regional educational performance in Indonesia using K-Means. *Journal of Computing and Smart Ecosystems*.

Sari, N. N., Adzima, K. R., Sahila, S., & Khotimah, T. H. (2025). K-Means clustering to classify Indonesian provinces based on school participation and socio-economic indicators. *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam*.

Schubert, E. (2022). Stop using the elbow criterion for K-Means and how to choose the number of clusters instead. *arXiv*.

Todaro, M. P., & Smith, S. C. (2021). *Economic development* (13th ed.). Pearson.

United Nations Development Programme (UNDP). (2022). *Human development report 2021/2022: Uncertain times, unsettled lives: Shaping our future in a transforming world*. United Nations Development Programme.

Utami, R., Wahyuni, L., & Setiawan, D. (2023). Spatial statistical approach in regional development inequality study. *Jurnal Statistika Dan Aplikasi*, *7*(1), 1–12.

Wahyuni, S., & Setiawan, B. (2024). Machine learning implementation in public policy analytics. *Jurnal Sistem Informasi Pemerintahan*, *6*(1), 45–58.

Wang, X., & vom Hofe, R. (2020). *Selected methods of planning analysis*. Springer.

Wegmann, M., Schwalb-Willmann, J., & Dech, S. (2020). *Introduction to spatial data analysis*. Pelagic Publishing.

World Bank. (2025). *Data-driven development and public sector transformation*. World Bank.

Yuliana, N., Hidayah, R., & Kurnia, D. (2023). Regional disparity and public service delivery in Indonesia. *Jurnal Administrasi Dan Kebijakan Publik*, *5*(2), 89–104.

Yusuf, D., & Razi, F. (2025). Analisis faktor sosial ekonomi yang mempengaruhi rendahnya capaian pendidikan di Indonesia menggunakan kombinasi metode data mining. *Jurnal Sistem Informasi (JUSIN)*.