

MAPPING LITERACY DISPARITIES BETWEEN PROVINCES FOR FORMULATING NATIONAL LITERACY DEVELOPMENT POLICIES: A K- MEANS CLUSTERING APPROACH

Muhamad Rifqi Nurhidayat, Nayla Nur Syamila, Selvina Intan Budiman* & Hilma Mutiara Winata

UIN Sunan Gunung Djati Bandung, Indonesia

*Email: selvinaintanwork@gmail.com**

Article History

Received: 15 November 2025

Accepted: 15 December 2025

Published: 30 December 2025

Abstract

The disparity in literacy levels between provinces in Indonesia remains a serious problem that hinders equitable human resource development. Therefore, data-based mapping is needed to support more targeted literacy policies. This study aims to group provinces based on the Community Literacy Development Index, Reading Enthusiasm Level, and Library Service Equity, supported by the Engagement Model of Reading Motivation and Public Service Theory as theoretical foundations. The method used is a quantitative approach using the K-Means Clustering algorithm, complemented by validation analysis using the Elbow Method, Silhouette Coefficient, and Davies Bouldin Index. The results show the formation of two main clusters: a high-income cluster with better literacy infrastructure and a low-income cluster dominated by eastern Indonesia. These findings confirm the existence of structural disparities in access to literacy services and reading engagement, thus requiring differential design of national literacy policies based on regional characteristics.

Keywords: K-Means Clustering, Literacy, Regional Disparity.

A. INTRODUCTION

Literacy is a challenge in human resource development, a crucial factor for the advancement of the Indonesian nation (Mahmudi & Fatah, 2024). A love of reading is a fundamental indicator of literacy, making it a developed nation in terms of education and public knowledge (Hasanah D, 2024). However, the reality on the ground shows significant disparities in reading literacy levels, meaning that improvements in reading literacy across provinces are essential, particularly in terms of the quality and quantity of library infrastructure (Jelita, 2023).

Improving public literacy is a strategic agenda in Indonesia's human resource development (Jannah M, 2025). Literacy reflects not only the ability to read and write but also the ability to think critically, understand information, and actively participate in social life (Putri et al., 2024). The Indonesian government has strived to strengthen literacy culture through various policies and programs, such as the National Literacy Movement and strengthening the role of libraries and educational institutions. However, the gap in literacy rates between provinces remains quite wide, indicating that reading culture is not evenly distributed across Indonesia (Solihin et al., 2019).

According to research conducted by the Center for Education and Cultural Policy Research (Solihin et al., 2019), the reading literacy activity index in 34 Indonesian provinces shows significant differences. Provinces in Java and Bali tend to have higher literacy indexes

than regions in eastern Indonesia. This disparity is caused by differences in access to education, availability of reading materials, digital infrastructure, and local government support for literacy activities.

This geographic disparity underscores the need for a segmented and measured approach to formulating literacy policies. One-size-fits-all policies tend to be ineffective because they ignore the heterogeneity of social, economic, and infrastructure conditions across provinces. Therefore, a mapping that groups provinces with similar literacy characteristics into disparity clusters is needed so that resource allocation, including library services and government literacy programs, can be tailored to the specific needs of each cluster.

To achieve systematic mapping, the cluster analysis method has proven relevant. Research conducted by Sekar Kedaton et al. (2025) emphasized that mapping the level of public reading interest using the K-Means algorithm can help identify regions with similar reading habit patterns. Through this method, data on public reading interest is grouped into several clusters, thus identifying regions with high, medium, or low reading interest.

Based on these studies, it can be concluded that the K-Means clustering method has great potential for grouping regions based on literacy indicators and reading interest. Therefore, this study focuses on analyzing the grouping of provinces in Indonesia based on the public literacy development index, the level of library service equity, and the level of reading interest in 2024 using the K-Means clustering method.

This research is novel in its use of data on the public literacy development index, the level of library service equity, and the latest level of reading interest in 2024. This approach has not been widely used in previous research, which generally only focused on aspects of public reading interest. Thus, this research is expected to provide a more comprehensive and up-to-date picture of the state of public literacy in Indonesia. Furthermore, the results of this clustering will serve as an empirical database for formulating more targeted, effective, and equitable national literacy development policies, ensuring that central and regional government interventions are well-targeted in addressing inter-regional literacy disparities.

This research aims to map literacy disparities between provinces based on the community literacy development index, the level of library service equity, and the level of reading habit using the k-means clustering approach as a reference in formulating national literacy development policies.

B. LITERATURE REVIEW

This study uses the Engagement Model of Reading Motivation theory from Guthrie & Wigfield (2000) to examine how reading motivation and engagement influence literacy achievement. This model emphasizes that literacy is not only related to technical skills but also motivation, self-confidence, the environment, and reading opportunities. According to Harkina, Sandayanti, and Pradini (2023), the Reading Enthusiasm Level (TKM) variable represents aspects of reading motivation and engagement, while the Community Literacy Development Index (IPLM) is an indicator of final literacy outcomes.

Furthermore, this study is grounded in Public Service Theory to explain the role of the state in providing equitable access to literacy services through libraries (Maghfirah, 2025). Library services are viewed as a basic public service that must ensure equitable access to information and knowledge for all members of society (Jamridafrizal & Muhammad, 2024). Therefore, the Library Service Equity (PLP) variable is used to represent the level of equality of literacy services in various provinces, because the more equitable and quality the services are, the greater the opportunity for increasing public literacy (Muhaimin 2022). Library

services as public services and increasing public reading motivation can simultaneously encourage more optimal literacy improvements.

C. RESEARCH METHODOLOGY

This research employed a quantitative method with an exploratory data mining approach using the K-Means Clustering algorithm, which aimed to group provinces in Indonesia based on similar characteristics in the literacy level and reading interest variables. The quantitative approach was chosen because the research involved numerical analysis of literacy indices and reading interest indicators. As stated by Kasiram (2008) in Abdullah et al. (2021), quantitative research is the process of discovering knowledge through numerical data as an analytical tool. K-Means Clustering is used as a non-hierarchical clustering technique that separates data into several clusters based on the proximity of characteristics between observations (Amalina et al., 2022).

The data used in this study are secondary data sourced from the Community Literacy Development Index and the Community Reading Interest Level, along with their constituent indicators at the provincial level in 2024. All data were obtained from the official source, the Central Statistics Agency (BPS), via <http://www.bps.go.id/>, ensuring its quality and credibility. The use of official government data was chosen to ensure methodological suitability for empirical data-based research and to improve the accuracy of the clustering analysis.

The analysis process involved several key stages, starting with data preprocessing, which included checking for missing values, removing duplications, and handling inconsistencies to ensure dataset quality. Data normalization was then performed to ensure each variable had equal weight in cluster formation. In the clustering stage, the study employed three methods for determining the optimal number of clusters: the Elbow Method (Saji, 2025), the Silhouette Coefficient (Shahapure & Nicholas, 2020), and the Davies–Bouldin Index (Saji, 2025) to minimize subjectivity and strengthen the statistical basis of the clustering results. The final stage involved analysis and visualization, where patterns, characteristics, and differences between clusters were identified and then visualized to facilitate interpretation of relationships between variables and the validity of the obtained cluster structure.

D. RESULT AND DISCUSSION

The results of the K-Means analysis on three main variables: the Community Literacy Development Index (IPLM), Library Service Equity (PLP), and Reading Enthusiasm Level (TKM), indicate that there are quite distinct differences in literacy characteristics between provinces in Indonesia.

The number of clusters in this study was determined using the Elbow Method, which indicates an elbow point at $k = 2$, thus the optimal number of clusters for mapping literacy disparities between provinces. After the clustering process was completed, the quality of the clustering was validated using two evaluation metrics: the Silhouette Score and the Davies-Bouldin Index (DBI).

The evaluation results showed a Silhouette Score of 0.510, which is considered good, indicating a fairly clear cluster structure and strong similarities between members within each cluster. Furthermore, the DBI value of 0.687 is also considered good, indicating that the separation between clusters is relatively clear, although there is still slight overlap or closeness in characteristics between provinces at the boundaries of the two clusters. Thus, the configuration of the two clusters formed can be said to be able to adequately represent the segmentation of literacy disparities between provinces in Indonesia.



Figure 1. Interprovincial Disparity Map
Source: Diola Riset, 2025

The interprovincial literacy clustering map presented provides a visual depiction of literacy disparities in Indonesia based on three main indicators: the Community Literacy Development Index (IPLM), Library Service Equity (PLP), and Reading Interest Level (TKM). This visualization reinforces the K-Means results, which divide 38 provinces into two main clusters: the High Cluster and the Low Cluster.

Distribution of the High Cluster (marked in red on the map)

Cluster 0 (high) indicates that most provinces in this group have relatively high and more evenly distributed IPLM and TKM achievements, indicating a relatively good level of literacy development and reading interest. However, the fluctuating PLP scores between provinces indicate that high literacy rates have not been fully accompanied by equitable public participation in utilizing library services.

Low Cluster Distribution (marked in blue on the map)

Cluster 1 (low) shows literacy disparities in the provinces of eastern Indonesia, where IPLM and TKM scores tend to be in the moderate category, with better achievements in North Maluku and West Papua, but the lowest in Highlands Papua, South Papua, and Central Papua. The low PLP in most provinces in this cluster indicates minimal public participation in utilizing literacy services.

Table 1. Clustering Results

Province	IPLM	PLP	TKM	Cluster
DI Yogyakarta	86,39	0,7625	79,99	Tall
Bangka Belitung Islands	84,59	0,7356	77,47	Tall
Bali	66,05	0,6813	71,97	Tall
Riau Islands	74,24	0,6712	73,69	Tall
South Kalimantan	81,16	0,6251	74,63	Tall
Bengkulu	65,96	0,6097	68,83	Tall
West Sumatra	82,47	0,607	73,3	Tall
Jakarta Province	73,07	0,6047	72,19	Tall
East Kalimantan	78,34	0,5953	69,53	Tall
Central Java	70,57	0,5946	73,91	Tall
South Sulawesi	88,24	0,5404	74,46	Tall
West Kalimantan	75,15	0,5361	71,26	Tall
West Java	72,76	0,5236	75,07	Tall

ARTICLE

Gorontalo	77,46	0,4969	62,43	Tall
Southeast Sulawesi	72,31	0,4904	66,02	Tall
Jambi	65,43	0,4543	68,05	Tall
South Sumatra	72,24	0,4531	69,62	Tall
Central Sulawesi	71,7	0,4278	67,48	Tall
Lampung	64,81	0,426	67,67	Tall
East Nusa Tenggara	62,62	0,4189	70,34	Tall
North Sulawesi	62,54	0,4065	68,44	Tall
East Java	78,6	0,3809	77,15	Tall
West Nusa Tenggara	60,42	0,3673	65,67	Tall
Central Kalimantan	72,5	0,3637	68,34	Tall
North Sumatra	62,39	0,3606	68,57	Tall
Riau	69,24	0,3503	70,26	Tall
West Sulawesi	63,65	0,3344	67,06	Tall
Aceh	72,42	0,3212	69,93	Tall
Banten	61,88	0,3197	70,66	Tall
North Kalimantan	66,73	0,2972	72,8	Tall
Maluku	58,55	0,3501	62,58	Low
North Maluku	61,7	0,3139	60,52	Low
Papua	60,75	0,2049	50,86	Low
West Papua	65,86	0,1957	59,29	Low
Southwest Papua	54,14	0,1619	54,89	Low
Central Papua	48,93	0,1165	52,06	Low
South Papua	49,78	0,0843	54,82	Low
Mountainous Papua	35,25	0,0264	38,83	Low

Source: Processed by Researchers, 2025

Based on the Inter-Provincial Literacy Clustering Results Table, of the 38 provinces analyzed in this study, two main clusters emerged: a High Cluster with 30 provinces and a Low Cluster with 8 provinces. The High Cluster is dominated by provinces with IPLM, PLP, and TKM achievements relatively above the national average. Some of the provinces with the highest literacy performance include DI Yogyakarta, the Bangka Belitung Islands, and East Java, reflecting a strong level of literacy development and public interest in reading, along with relatively good participation in the use of literacy services.

Conversely, the Low Cluster is consistently dominated by provinces in eastern Indonesia, such as Papua, Central Papua, South Papua, Southwest Papua, and especially Highland Papua, which shows the lowest achievement. In general, provinces in this cluster have lower TKM scores and very minimal PLP, indicating low levels of public participation in literacy services.

The High Cluster is dominated by provinces with IPLM, PLP, and TKM achievements far above the national average. Among them, the Special Region of Yogyakarta demonstrated the highest performance (IPLM 86.39; PLP 0.7625; TKM 79.99), followed by the Bangka Belitung Islands (IPLM 84.59; PLP 0.7356; TKM 77.47), and East Java (IPLM 81.16; PLP 0.6251; TKM 77.15). These high literacy rates are accompanied by strong literacy infrastructure in these regions, as evidenced by the significant number of libraries. According to data obtained from the Ministry of Elementary and Secondary Education (2024), Central Java has approximately 14,514 libraries, East Java 12,975, and West Java 11,168, providing the public with broad access to reading materials and literacy services.

In contrast, the Low Cluster is consistently found in provinces in eastern Indonesia, such as Papua (IPLM 52.23; PLP 0.0801; TKM 50.86), Central Papua (IPLM 46.97; PLP 0.0471; TKM 52.06), South Papua (IPLM 48.33; PLP 0.0567; TKM 54.82), Southwest Papua (IPLM 55.94; PLP 0.0912; TKM 54.89), and especially in Highlands Papua, which recorded the lowest achievement (IPLM 35.25; PLP 0.0264; TKM 38.83). The low scores of these three indicators align with the small number of libraries in eastern Papua, with only around 1,343 libraries, Central Papua around 600, and Highlands Papua, the province with the lowest number of libraries nationally.

This disparity in literacy and infrastructure achievement reinforces Public Service Theory, which emphasizes that public services, including libraries, must be available and distributed equitably to ensure equal opportunities for the public to benefit from education and information. The striking difference between provinces with over 11,000 libraries and those with fewer than 1,500 demonstrates that access to literacy services is a structural factor determining the level of community engagement in reading.

This finding is also consistent with the Engagement Model of Reading Motivation (Guthrie & Wigfield, 2000), which states that reading engagement depends not only on internal motivation but also on a supportive external environment. Therefore, the low Reading Enthusiasm Level (TKM) in eastern provinces cannot be understood solely as a lack of reading interest, but rather as a reflection of the limited opportunities for the community to engage in literacy practices and the supporting infrastructure that will impact the Community Literacy Development Index (IPLM).

The results of this analysis indicate that Indonesia needs an adaptive, gradual, and evidence-based literacy policy. Differences in characteristics between provinces in the achievement of IPLM, PLP, TKM, and the availability of library facilities demonstrate that national policies cannot be applied uniformly. An evidence-based policymaking approach requires the government to formulate interventions based on empirical data demonstrating structural inequalities between regions (Lentari, 2024). Therefore, policy recommendations must be directed at identifying specific regional needs to optimize the effectiveness of the national literacy program.

Overall, national literacy policy needs to focus on three main strategic agendas: equitable distribution of literacy infrastructure, improving the quality of library services, and strengthening digital literacy. Equitable infrastructure includes the development of community-based libraries, increasing the distribution of appropriate reading materials, and expanding internet access to address inequality (Aurellia et al., 2025). Furthermore, improving the quality of literacy services needs to focus on digital library transformation, service modernization, and technology-based program innovation. Abdurokhim & Masruri (2025) found that digital libraries play a significant role in increasing community access and participation. These efforts must be accompanied by increased capacity of local governments to design and implement sustainable literacy programs.

In addition to infrastructure and service aspects, national policy must also consider factors such as community motivation and reading engagement. The Engagement Model of Reading Motivation suggests that reading interest and participation increase when individuals receive environmental support, access to reading resources, and relevant literacy activities (Achmad et al., 2023). Therefore, the national literacy program needs to incorporate a community-based approach, creative literacy activities, the provision of meaningful content, and the development of literacy activities appropriate to the regional socio-cultural context. The combination of equitable access, strengthening reading motivation, and improving

service quality is expected to gradually reduce the literacy gap and build an inclusive and sustainable national literacy ecosystem.

E. CONCLUSION

Based on the results of an analysis using the K-Means Clustering algorithm on three main variables: the Community Literacy Development Index (IPLM), Library Service Equity (PLP), and Reading Enthusiasm Level (TKM), it was found that there are clear literacy disparities among provinces in Indonesia. The clustering results indicate the formation of two main clusters: the High Cluster, consisting of 30 provinces, and the Low Cluster, consisting of 8 provinces.

The High Cluster is dominated by provinces with relatively adequate access to literacy and library infrastructure, and IPLM and TKM achievements above the national average. Conversely, the Low Cluster is primarily inhabited by provinces in eastern Indonesia, which have low levels of library access, minimal service equity, and lagging IPLM and TKM achievements.

Cluster quality validation using the Silhouette Score (0.510) and Davies-Bouldin Index (0.687) indicates that the cluster structure is of good quality and is able to represent the segmentation of literacy disparities representatively. The findings of this study confirm that literacy inequality is influenced not only by motivation and reading culture, but also by unequal access to library facilities and the literacy ecosystem.

To reduce literacy disparities between provinces, the government needs to prioritize increasing the distribution of library infrastructure, especially in low-income areas, while strengthening the quality of literacy services and expanding access to both physical and digital reading materials. Furthermore, literacy programs must be designed according to the socio-cultural context of each region, involving communities, schools, and local governments to increase public reading engagement. Further research is also needed, including other supporting variables, so that literacy mapping can form the basis for more comprehensive and targeted policies.

REFERENCE

- Abdullah, K., Jannah, M., Aiman, U., Hasda, S., Fadilla, Z., Taqwin, Masita, Ardiawan, K. N., & Sari, M. E. (2021). *Metodologi Penelitian Kuantitatif*. Yayasan Penerbit Muhammad Zaini.
- Abdurokhim, M., & Masruri, A. (2025). I Peran Perpustakaan Digital Dalam Melayani Masyarakat Informasi. *Kajian Kepustakawanan Dan Informas*, 6(1), 19–34.
- Achmad, C., Rizky, N., Amin, S., Studi, P., Ilmu, P., Sosial, P., Islam, U., Maulana, N., & Ibrahim, M. (2023). TERHADAP PRESTASI SISWA PADA PELAJARAN IPS. *Dinamika Sosial : Jurnal Pendidikan Ilmu Pengetahuan Sosial*, 2(1), 56–67.
- Amalina, T., Pramana, D. B. A., & Sari, B. N. (2022). Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food Tifani Amalina 1 , Danendra Bima Adhi Pramana 2 , Betha Nurina Sari 3 Teknik Informatika, Universitas Singaperbangsa Karawang. *Jurnal Ilmiah Wahana Pendidikan*, 8(September), 574–583.
- Aurellia, A., Athifa, C. N., & Amrozi, Y. (2025). Transformasi Digital Yang Adil: Peran Manajemen Risiko Dalam Mengurangi Ketimpangan Digital. *Nusantara Computer And Design Review*, 3(1), 54–63.
- Harkina, P., Sandayanti, V., & Pradini, S. (2023). Cognitive Strategy And Motivation In Reading English Literature On College Students. *PSYCHE: JURNAL PSIKOLOGI UNIVERSITAS MUHAMMADIYAH LAMPUNG*, 5(2), 234–253.

- Hasanah, D. F. (2024). Kesetaraan Wawasan Dunia Melalui Literasi: Evaluasi Ketercapaian Gerakan Literasi Nasional Melalui Data Pisa Dan Statistik Indonesia. *Journal Of Education For The Language And Literature Of Indonesia*, 2(2), 98-110.
- Jamridafrizal, Z., & Wajdi, M. F. (2024). Perpustakaan Sebagai Institusi. *Perpustakaan Sebagai Institusi Perspektif Organisasi Dan Regulasi*, 10-12.
- Jannah, M. (2025). Membangun Rumah Literasi Di Desa: Menumbuhkan Budaya Membaca Dan Meningkatkan Kualitas Pendidikan. *Jurnal Literasi Digital*, 5(1), 1-10.
- Jelita, M. (2023). *Analisis Clustering Menggunakan Metode K-Means Untuk Mengelompokkan Kabupaten/Kota Di Indonesia Berdasarkan Unsur-Unsur Pembangun Literasi Masyarakat (UPLM) Studi Kasus Data Tahun 2023*. <https://docs.rapidminer.com/latest/studio/guided/index.html>
- Kemendikdasmen. (2025). *Jumlah Perpustakaan Menurut Kondisi Tiap Provinsi Tahun 2024*. Portal Data Pendidikan. <https://data.kemendikdasmen.go.id/dataset/p/sarana-dan-prasarana/jumlah-perpustakaan-menurut-kondisi-tiap-propinsi-indonesia-sd-2024>
- Lentari, S. (2024). Evidence-Based Policy Making: Penggunaan Bukti Untuk Program Penanggulangan Kemiskinan Di Jawa Timur. *Publikauma: Jurnal Administrasi Publik Universitas Medan Area*, 12(1).
- MAGHFIRAH, A. (2025). Strategi Perpustakaan Umum Dalam Penguatan Literasi Keluarga Di Masyarakat Kota Payakumbuh.
- Mahmudi, & Fatah, Z. (2024). Analisis Pengaruh Jenis Buku Terhadap Minat Baca Mahasiswa Di Perpustakaan Ibrahimy Dengan Algoritma K-Means Clustering. *JURNAL ILMIAH SAINS TEKNOLOGI DAN INFORMASI*, 3(1), 53–65. <https://doi.org/10.59024/jiti.v3i1.1013>
- Putri, I. T. A., Agusdianita, N., & Desri, D. (2024, August). Literasi Dalam Meningkatkan Kemampuan Berpikir Kritis Peserta Didik Sekolah Dasar Era Digital. In *Social, Humanities, And Educational Studies (SHES): Conference Series* (Vol. 7, No. 3).
- Saji, B. (2025). *Elbow Method For Optimal Cluster Number In K-Means*. Analyticsvidhya. <https://www.analyticsvidhya.com/blog/2021/01/in-depth-intuition-of-k-means-clustering-algorithm-in-machine-learning/>
- Shahapure, K., & Nicholas, C. (2020). Cluster Quality Analysis Using Silhouette Score. *2020 IEEE 7th International Conference On Data Science And Advanced Analytics (DSAA)*, 747–748. <https://doi.org/10.1109/Dsaa49011.2020.00096>
- Sekar Kedaton, A., Alan Bukhari, B., Wahyu Febriyanto, R., & Arum Sari, A. (2025). Pemetaan Wilayah Berdasarkan Minat Baca Masyarakat Dengan Algoritma K-Means. *Seminar Nasional Teknologi Informasi Dan Bisnis (SENATIB)*, 2025.
- Solihin, L., Utama, B., Pratiwi, I., & Novirina, N. (2019). Indeks Aktivitas Literasi Membaca 34 Provinsi.