Explainable GIS-Based Decision Support System for Internet Infrastructure Planning Using Spatial K-Means Clustering

Irena Santi Widjaja
Doctoral Program of Information Systems, Postgraduate School, Universitas Diponegoro, Semarang, Indonesia
irenasantiwidjajaja@students.undip.ac.id (corresponding author)

R. Rizal Isnanto
Department of Computer Engineering, Faculty of Engineering, Universitas Diponegoro, Semarang, Indonesia
rizal@ce.undip.ac.id

Maman Somantri
Department of Electrical Engineering, Faculty of Engineering, Universitas Diponegoro, Semarang, Indonesia
mmsomantri@live.undip.ac.id

Received: 4 February 2026 | Revised: 24 February 2026, 2 March 2026, and 14 March 2026 | Accepted: 15 March 2026

ABSTRACT

This study aimed to develop an explainable GIS-based decision-support system for planning Internet infrastructure tailored to regional spatial features. Spatial K-Means clustering is applied to demographic, socioeconomic, and network infrastructure data to identify geographic clustering patterns. Experiments using various clusters ($k = 3$ and $k = 5$) were used to evaluate the model's ability to distinguish regional characteristics. The results show that the proposed method reveals significant regional differences, with population densities from 445 to 3515 people/km^2, Internet penetration rates between 52.4% and 82.7%, and Base Transceiver Stations (BTSs) ranging from 2.9 to 12.3 units per cluster. Integrating interactive GIS visualization with explainable analytics enhances the transparency of results and helps identify priority development regions, transition regions, and regions with mature infrastructure. These findings offer practical implications for formulating internet infrastructure development policies that are more targeted and tailored to regional needs.

Keywords-spatial clustering; explainable; infrastructure planning; regional analysis; policy-oriented analytics

I. INTRODUCTION

Access to reliable internet infrastructure is now essential for socioeconomic growth and improving the quality of public services. The internet is no longer just an auxiliary facility, but a strategic infrastructure that determines how the public participates in the digital economy, education, and modern governance [1, 2]. Obtaining equitable internet access remains challenging, especially in regions beyond urban centers. This gap is often due to technological limitations and infrastructure planning that does not adequately consider various spatial conditions and local needs [3, 4].

Internet infrastructure planning strongly depends on spatial dimensions. Factors such as population density, settlement patterns, geographic conditions, and availability of existing infrastructure interact, creating complexities that conventional planning methods struggle to encompass [5]. In many cases, decision-making continues to rely on administrative boundaries or broad aggregate indicators that often do not accurately reflect real-world conditions. This can lead to inefficient decisions, increasing both investment costs and inadequate service coverage.

Geographic Information Systems (GIS) are extensively used to support location-based analysis and planning. They enable the integration of diverse spatial and non-spatial datasets and provide visualization tools that enhance the understanding of regional characteristics [6-8]. However, as infrastructure datasets become increasingly multidimensional and analytically demanding, visualization alone is insufficient to guide structured policy decisions. Analytical outputs must be systematically processed, interpreted, and translated into actionable planning priorities. In this context, a Decision Support System (DSS) provides a structured computational framework that integrates three essential components: data management, model management, and user-oriented presentation. The data layer organizes and harmonizes heterogeneous spatial datasets, the model layer performs analytical procedures such as clustering and spatial evaluation, and the presentation layer communicates interpretable outputs to decision-makers. Embedding GIS analysis within a well-defined DSS architecture enhances transparency, consistency, and accountability in infrastructure planning processes [9].

Within the model management component of a spatial DSS, clustering techniques are frequently employed to structure and analyze regional patterns. In GIS-based analytical environments, clustering enables spatial regionalization by grouping areas with similar demographic, infrastructure, and geographic characteristics. For instance, hierarchical clustering integrated with geospatial visualization has been applied to support decision-oriented analytical systems [10]. Applied infrastructure studies have also utilized clustering approaches, including K-Means, to support regional segmentation and infrastructure-based analysis [11]. The K-Means algorithm remains widely adopted due to its computational efficiency, scalability for large datasets, and centroid-based interpretability [12]. Furthermore, spatial clustering has been shown to capture regional heterogeneity more effectively than rigid administrative segmentation approaches [13]. However, the success of each approach depends largely on the analysis goals, the application context, and the spatial characteristics of the region. The K-Means algorithm partitions spatial units by minimizing intra-cluster variance, typically computed using Euclidean distance or related similarity measures [14].

The lack of transparency in analytical outputs is a significant issue, particularly when such systems guide public infrastructure decisions with long-term socio-economic implications. Policymakers require not only segmentation results, but structured explanations of how regional characteristics influence cluster formation. Therefore, the concept of explainable decision-support systems has emerged to enhance interpretability, accountability, and transparency in data-driven planning environments [15, 16].

Despite the growing application of spatial clustering in geospatial and infrastructure-related studies, several methodological limitations remain. First, many existing works emphasize descriptive segmentation without incorporating formal cluster validation metrics to assess separation quality and robustness. Second, comparative evaluations between clustering approaches in large-scale infrastructure contexts are often limited, particularly in countrywide spatial planning scenarios. Third, clustering outputs are rarely embedded within a structured DSS architecture that translates analytical results into transparent and policy-oriented recommendations. Consequently, there is a need for a validated and explainable spatial clustering framework that integrates methodological rigor with decision-support functionality.

This study proposes an explainable GIS-based decision-support framework for internet infrastructure planning using spatial K-Means clustering, integrating GIS-based spatial analysis with structured cluster interpretation mechanisms to enhance transparency and policy relevance. By embedding explainability as a core component, the framework promotes a more evidence-driven and context-aware planning process across heterogeneous regional settings.

II. METHODOLOGY

This study uses an Explainable Spatial Decision Support System (SDSS) approach that combines GIS-based analysis with spatial K-Means clustering to support transparent and data-driven planning of internet infrastructure. The method aims to generate location suggestions and regional priorities, as well as to offer clear explanations of the decision-making process. Figure 1 illustrates the research stages undertaken.

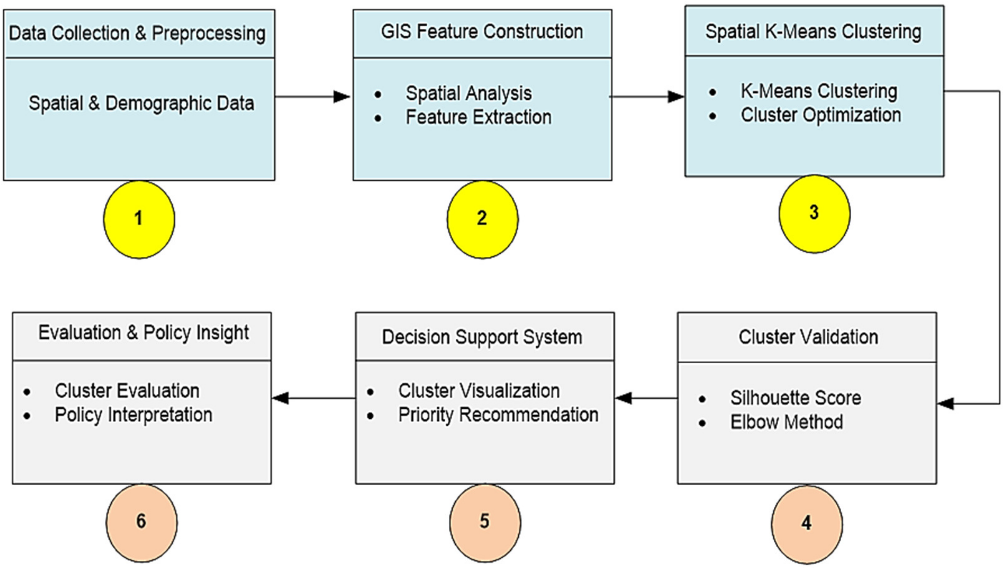


Fig. 1. Methodological framework of the explainable GIS-based DSS.

A. Data Collection and Preprocessing

This study utilizes both spatial and non-spatial data related to internet infrastructure planning across regional units throughout Indonesia. The dataset covers multiple provinces throughout the country, representing diverse demographic, socio-economic, and geographic conditions. The dataset was compiled from official data sources, including the Ministry of Communications and Digital (KOMDIGI) and Statistics Indonesia (BPS). Infrastructure-related indicators such as BTS distribution and network coverage were obtained from national telecommunication records, while demographic and socio-economic variables were collected from publicly available statistical databases. All data correspond to the most recent reporting period available at the time of the analysis.

Clustering analysis is conducted on regional units across Indonesia, and no direct comparative analysis is performed between specific provinces. The analytical framework incorporates demographic, socio-economic, and infrastructure-related indicators, including population density, settlement patterns, availability, and distribution of existing network infrastructure, regional accessibility, and geographic characteristics such as topography and spatial conditions. These variables collectively represent demand intensity, infrastructure capacity, and geographic constraints relevant to regional internet planning. All spatial data were standardized to a common coordinate reference system to ensure spatial consistency. Data cleaning procedures were conducted to eliminate missing values, attribute inconsistencies, and duplicate entries. Furthermore, all numerical variables were normalized before clustering to prevent scale dominance and to ensure balanced influence in the distance-based segmentation process. Table I presents an illustrative sample of the processed dataset used in this study (selected regions from Java and Sulawesi), provided solely for illustrative purposes without limiting the national scope of the analysis. Each row signifies a single analysis unit (region), identified by Region ID and Region name, and categorized by Province and Area type (urban or rural). This data helps to examine spatial variation and regional traits relevant to infrastructure planning.

Demographic details, estimate the potential demand for internet services in each region. Typically, areas with higher population and density require more infrastructure, particularly in urban zones. In addition, Average income (USD) serves as a marker of community economic capacity, which can influence internet service adoption levels. Network infrastructure conditions are indicated by the Number of BTSs and Fiber network, reflecting the existing facilities in each region. The Internet penetration (%) and Number of complaints offer insights into the perceived quality and accessibility of internet services by the community, helping to identify areas with notable service deficiencies. Along with non-spatial attributes, this table includes geographic coordinates (Longitude and Latitude) to facilitate integration with a GIS environment. This spatial data serves as the foundation for spatial analysis and spatial K-Means clustering, enabling regions to be grouped based on attribute similarities and geographic proximity.

B. GIS-Based Feature Construction

The preprocessed data is incorporated into a GIS environment to build spatial features that thoroughly depict each region's characteristics. Each regional unit is treated as a spatial entity, integrating demographic, socioeconomic, and internet infrastructure data with geographic location details. The longitude and latitude coordinates define a point layer in GIS. Non-spatial attributes such as population density, average income, BTS count, fiber network presence, internet penetration, and complaint numbers are associated via a spatial join. This approach represents each region as a multidimensional feature vector suitable for advanced spatial analysis. Additionally, basic spatial analyses, such as overlays and distance calculations, are used to maintain consistency in spatial relationships between regions and to assess how geographic proximity influences infrastructure planning. This stage produces a structured spatial dataset that combines attributes and locations, ready for use in the next phase of spatial K-Means clustering.

Figure 2 illustrates an example of a GIS-based point layer utilized for internet infrastructure area analysis. Each point represents a single territorial unit and is located based on geographic coordinates (longitude and latitude). These points include non-spatial attributes, such as demographic data (population size and density), socioeconomic data (average income), and internet infrastructure details (number of base transceiver stations, fiber network availability, internet penetration rate, and complaint count). This visualization provides a regional overview of the spatial distribution and highlights regional patterns and concentrations before further analysis. In this stage, the map does not display clustering results; all points are shown uniformly, depicting the initial spatial data used as input to the subsequent spatial K-Means clustering process.

TABLE I. SAMPLE DATASET FOR INTERNET INFRASTRUCTURE PLANNING

Region ID	Region name	Province	Area type	Population	Population density (people/km^2)	Average income (USD)	Number of BTS	Fiber network	Internet penetration (%)	Number of complaints	Longitude	Latitude
ID_1001	Village_A0	West Java	Urban	10,227	4,360	421	10	Yes	82.32	17	106.73	-6.42
ID_1002	Village_B1	South Sulawesi	Urban	17,784	1,130	355	8	No	70.40	32	119.74	-5.01
ID_1003	Village_C2	South Sulawesi	Rural	3,112	591	251	2	No	44.59	4	119.82	-5.18
ID_1004	Village_D3	South Sulawesi	Rural	3,358	637	166	4	No	51.64	9	120.14	-3.02
ID_1005	Village_E4	South Sulawesi	Rural	6,558	231	101	3	No	51.85	9	120.45	-2.87

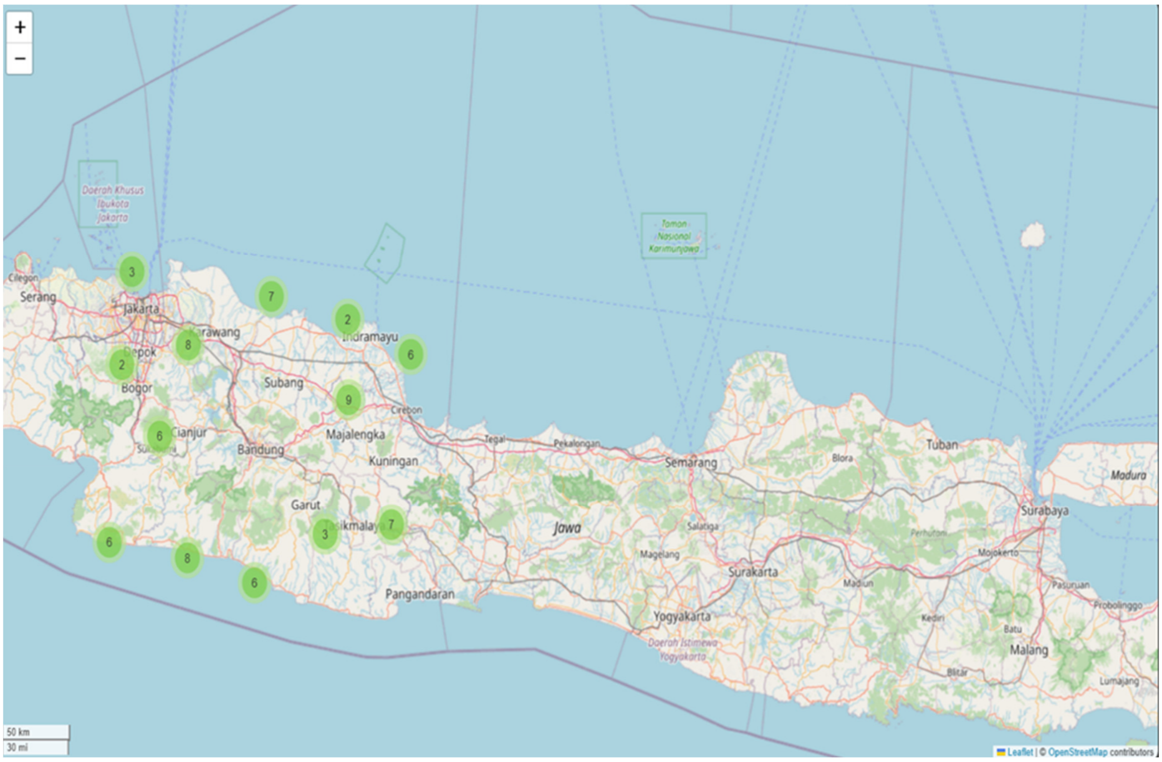


Fig. 2. GIS-based point layer of internet infrastructure analysis regions.

C. Spatial K-Means Clustering

The structured spatial dataset from the previous stage serves as input for the Spatial K-Means clustering process. Each region is represented by a multidimensional feature vector that combines demographic, socio-economic, internet infrastructure, and geographic coordinate (longitude and latitude) attributes. Integrating both non-spatial and spatial attributes enables clustering to consider regional similarities alongside geographic proximity.

Before clustering, all numerical variables were normalized using min-max scaling to ensure balanced contribution in the distance computation. The Spatial K-Means algorithm employed the Euclidean distance to measure similarity between multidimensional feature vectors. The Euclidean distance between two regions x_i and x_j is defined as:

$$(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2} \tag{1}$$

where n represents the number of attributes, including demographic, socio-economic, infrastructure, and projected spatial coordinates. Geographic coordinates were transformed into a projected coordinate reference system before clustering to maintain spatial consistency.

Two clustering scenarios (k = 3 and k = 5) were evaluated to analyze different regional grouping configurations. The resulting clusters depict spatial patterns of internet infrastructure planning, including regions with high penetration and adequate infrastructure, transition zones requiring network improvements, and priority areas with limited access.

Figure 3 presents a sample of Spatial K-Means clustering results, dividing the analysis areas into three groups based on demographic, socioeconomic, and internet infrastructure factors, while also accounting for geographic proximity. Each point on the map corresponds to a territorial unit, with colors indicating cluster membership. These clusters tend to be compact and geographically localized, highlighting the importance of spatial proximity in the clustering. Clusters with specific dominant colors are mainly found in regions with higher population density, internet penetration, and better infrastructure. In contrast, other clusters cover areas with less developed infrastructure, fewer BTS, and moderate to low internet usage. This pattern reflects regional differences in internet infrastructure development needs and priorities.

D. Cluster Validation and Robustness Analysis

To ensure that the spatial clustering results are statistically reliable and not merely visually interpretable, cluster validation and robustness analysis were conducted. This stage aimed to evaluate the internal quality of the clustering structure and to justify the selection of the number of clusters used in the spatial segmentation process.

The Elbow Method was applied to determine the appropriate number of clusters (k) by analyzing the Within-Cluster Sum of Squares (WCSS). WCSS measures cluster compactness by calculating the total squared distance between each data point and its corresponding cluster centroid. Formally, WCSS is defined as:

$$WCSS = \sum_{i=1}^{K} \sum_{x \epsilon C_i} ||x - \mu_i||^2 \tag{2}$$

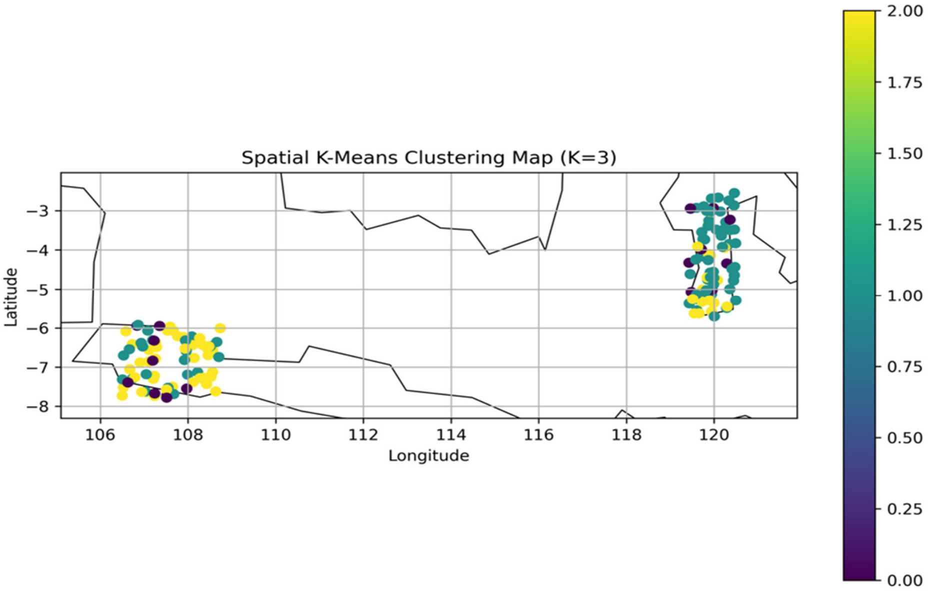


Fig. 3. Spatial K-Means clustering map of Internet infrastructure planning regions.

where C_i denotes cluster i, μ_i represents the centroid of cluster i, and $||x - \mu_i||^2$ corresponds to the squared Euclidean distance between a data point and its assigned centroid. The optimal number of clusters is identified at the inflection point of the WCSS curve, where additional clusters result in diminishing improvements in intra-cluster variance reduction. In addition to the Elbow Method, clustering quality was evaluated using the Silhouette coefficient, which measures both intra-cluster cohesion and inter-cluster separation. For each observation i, the Silhouette score is computed as:

$$s(i) = \frac{b(i) - a(i)}{max\ (a(i), b(i))} \quad (3)$$

where $a(i)$ is the average distance between observation i and other observations within the same cluster, and $b(i)$ is the average distance between observation i and observations in the nearest neighboring cluster. The Silhouette score ranges from −1 to 1, with values closer to 1 indicating well-separated and cohesive clusters, values near 0 suggesting overlapping clusters, and negative values implying potential misclassification.

To enhance methodological robustness, all numerical variables were normalized before clustering to prevent scale dominance in distance computation. The clustering process was evaluated across multiple k values, including $k = 3$ and $k = 5$, to examine the trade-off between segmentation granularity and cluster interpretability. This validation ensures that the spatial grouping structure is statistically sound before being integrated into the decision support and policy evaluation stages.

E. Explainable Decision Support and Policy Interpretation

Spatial K-Means clustering results are used to create an explainability mechanism that supports transparent, understandable decision-making for stakeholders. The primary goal is to clarify why a specific region is assigned to a certain cluster and what this means for internet infrastructure planning. Explainability is achieved by analyzing cluster profiles (centroids), which reflect the averages of key variables such as population density, income levels, the number of BTS, fiber network availability, internet penetration rates, and complaint counts. Comparing these centroids helps identify the main factors that distinguish each cluster. This allows each cluster to be described semantically, such as an area with mature infrastructure, a transition zone, or a development priority. Cluster interpretations are then integrated into a GIS-based DSS. Thematic maps combined with attribute data generate operational policy recommendations, such as building new base stations, expanding fiber networks, or optimizing services in areas with high internet use but many complaints. This ensures that recommendations are based on clear numerical patterns and logical, contextual explanations.

F. Decision Support Output and Policy-Oriented Evaluation

This is the final phase of the proposed framework, in which all prior analyses spanning spatial modeling and clustering to cluster interpretation are combined to generate operational and policy-oriented DSS outputs. The primary objective is to present the results in a clear, actionable, and decision-maker-friendly format. The DSS output is a regional priority summary that categorizes regions into high, medium, or low priority for internet infrastructure development. These categories stem directly from the cluster interpretation results in the previous stage, taking into account key factors such as infrastructure constraints, demand potential, and service quality perceptions. This approach ensures that each policy recommendation is grounded in a transparent, well-supported analysis. Additionally, this stage features scenario-based policy evaluation, allowing the cluster to simulate various development strategies, such as installing base stations in priority zones, expanding fiber networks, or enhancing capacity in areas with existing infrastructure.

III. RESULTS AND DISCUSSION

This section outlines the key findings of a GIS-based DSS for internet infrastructure planning. The system combines spatial and non-spatial data to generate outputs that are both analytical and directly applicable to policy development. The results are presented in the following order: first, the system output; then, its interpretation and policy implications.

A. *Experimental Results*

The experimental process involved applying the Spatial K-Means clustering method with different cluster counts (k = 1, 3, and 5) to assess the stability of spatial patterns and the level of detail in the generated information. The k = 1 setup served as a baseline to represent overall conditions across all regions without segmentation, providing a general view of average internet infrastructure characteristics. Next, the k = 3 setup was used to identify key spatial patterns that distinguish regions based on infrastructure needs and readiness. Meanwhile, the k = 5 configuration provided a more detailed segmentation, enabling the identification of local differences and potential infrastructure gaps across regions. The primary outputs include spatial regional groupings, cluster profiles reflecting the average traits of each group, and thematic maps illustrating the geographical distribution of the clusters. Table II presents a comparative summary of centroid characteristics and cluster sizes across different k values.

Table II offers a comparative overview of spatial clustering outcomes using the K-Means method with different cluster counts (k = 1, 3, and 5). It highlights the attributes of the cluster centers (centroids), representing the average conditions within each regional group, including population density, average income, number of BTS, fiber network availability, internet penetration, service complaints, and the number of regions per cluster. This comparison illustrates how changing the number of clusters affects the interpretation of results and their application in internet infrastructure planning. When k = 1, all regions form a single cluster, providing a broad overview of the internet infrastructure. However, this approach overlooks regional differences and variations, limiting its usefulness for tailoring region-specific policies.

TABLE II. COMPARATIVE SUMMARY OF SPATIAL K-MEANS CLUSTERING RESULTS

k	**Cluster**	**Population density**	**Average income (USD)**	**Total BTS**	**Fiber network**	**Internet penetration (%)**	**Total complaints**	**Total region**
1	C0	1453.51	238.57	6.53	0.41	63.94	15.73	150
3	C0	699.65	184.10	3.59	1.00	60.47	9.94	17
	C1	455.75	177.76	3.05	0.00	52.80	8.08	75
	C2	2964.69	333.18	11.90	0.76	79.37	27.33	58
5	C0	546.27	164.87	2.87	1.00	59.86	10.60	15
	C1	2482.38	288.66	12.33	1.00	76.31	32.75	24
	C2	445.49	177.33	2.96	0.00	52.39	8.00	74
	C3	3514.68	381.93	11.27	1.00	82.70	18.68	22
	C4	2664.47	324.02	11.60	0.00	77.74	27.47	15

Increasing the number of clusters to k = 3 makes the regional structure more clearly defined. One cluster indicates an area with relatively low population density and internet access, along with few BTS, making it a priority for development. Another cluster shows characteristics of a transitional zone with moderate infrastructure needs and availability. Meanwhile, the cluster with high population density and income has more BTSs and higher internet penetration, but also experiences more complaints, indicating usage pressure and service quality issues. This setup balances detail with simplicity, making it most useful for strategic planning.

In the k = 5 setup, clustering yields more detailed results by dividing areas into finer subclusters. This enables better identification of variations in infrastructure conditions and service demand, useful for operational analysis or micro-level planning. However, more clusters also mean increased interpretative complexity, which may diminish their usefulness for macro or strategic policy decisions. To objectively evaluate the clustering structures obtained for different k values, quantitative validation metrics were computed. The Silhouette score was used to assess inter-cluster separation and intra-cluster cohesion.

As shown in Figure 4, k = 3 achieves a higher Silhouette score (0.375) compared to k = 5 (0.333), indicating stronger structural separation and more compact cluster formation at k = 3. Although k = 5 provides finer segmentation and reduced within-cluster variance, the lower Silhouette score suggests weaker inter-cluster distinction. This reveals a trade-off between segmentation granularity and statistical stability. Therefore, k = 3 offers a more coherent configuration for strategic-level infrastructure planning, while k = 5 remains useful for operational-level and localized analysis where finer differentiation is required. Based on this evaluation, the k = 5 configuration was selected for dashboard visualization to demonstrate detailed spatial differentiation.

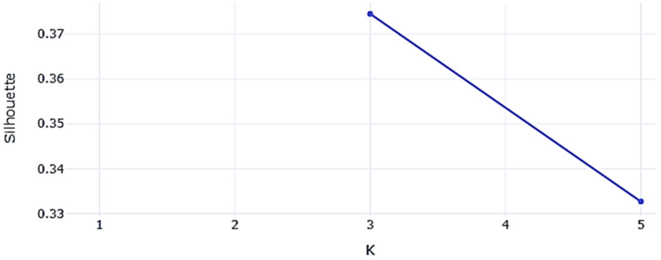
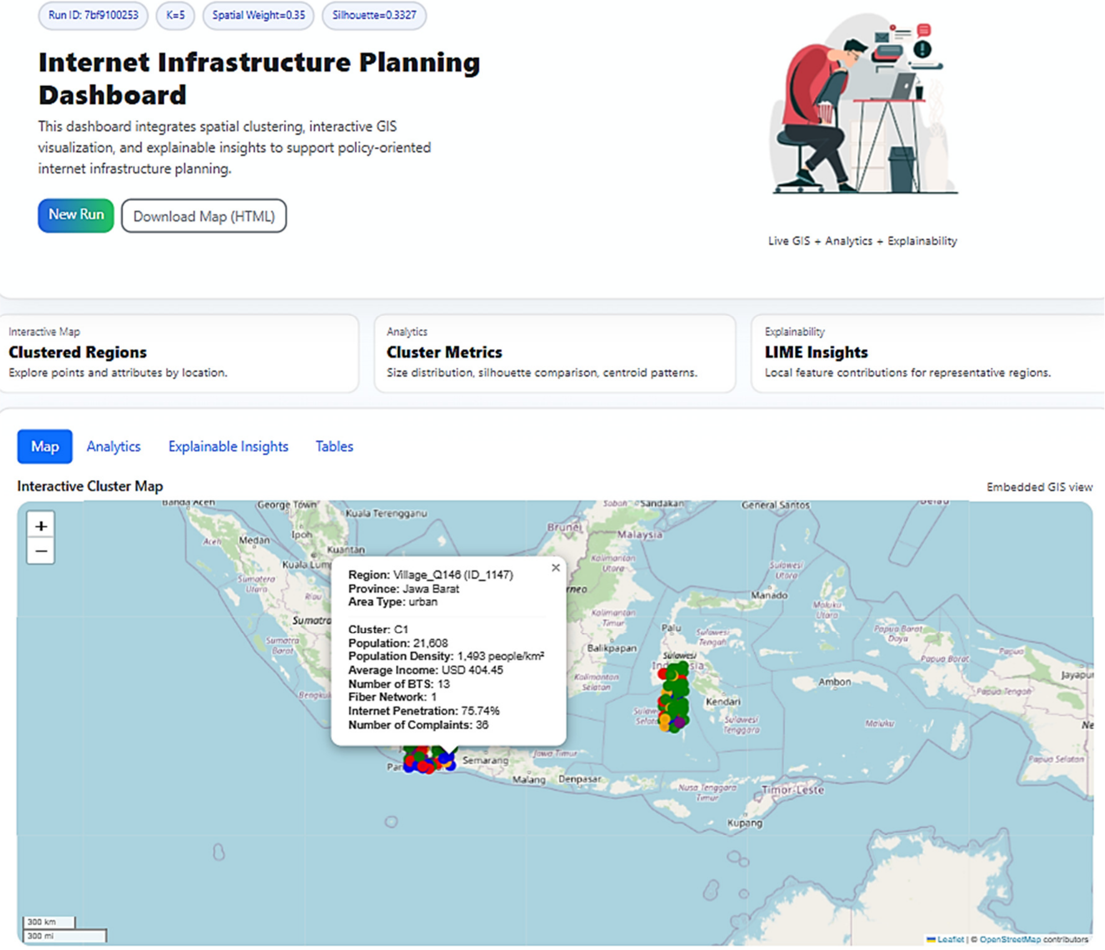


Fig. 4. Silhouette score as a function of the Number of Clusters (k).

Figure 5 shows the GIS-based decision-support dashboard configured with k = 5 for internet infrastructure planning. The interactive map illustrates the spatial distribution of regions grouped into five clusters based on demographic, socioeconomic, and network infrastructure features. Each cluster has a distinct color to facilitate the identification of spatial patterns and service gaps across regions. Figure 6 shows the distribution of regions within each cluster.



Internet Infrastructure Planning Dashboard

This dashboard integrates spatial clustering, interactive GIS visualization, and explainable insights to support policy-oriented internet infrastructure planning.

Region: Village_Q146 (ID_1147)
Province: Jawa Barat
Area Type: urban

Cluster: C1
Population: 21,608
Population Density: 1,493 people/km²
Average Income: USD 404.45
Number of BTS: 13
Fiber Network: 1
Internet Penetration: 75.74%
Number of Complaints: 36

Fig. 5. Interactive GIS-based decision support dashboard for Internet infrastructure planning (k = 5).

Fig. 6. Distribution of regions across clusters (k =5).

Figure 7 offers a heatmap that compares the characteristics of the cluster centers (centroids) for k = 5. These panels together provide a quantitative overview that complements the spatial visualization and aids in interpreting differences between clusters. Each row corresponds to a cluster, and each column highlights key indicators of internet infrastructure: population density, average income, number of BTS, fiber network coverage, internet penetration rate, and service complaints. The heatmap's color intensity indicates each indicator's value, facilitating the identification of clusters with varying levels of infrastructure demand and capacity. This visualization reveals that some clusters have high population density, high internet penetration, and many BTS, suggesting areas with well-developed infrastructure but potential service quality challenges. In contrast, other clusters display lower indicator values, indicating locations with limited infrastructure and scope for network growth.

Based on the spatial clustering and quantitative analysis shown in Figures 4 and 5, each cluster corresponds to a region with unique internet infrastructure features. These variations provide a solid foundation for regional analysis and the development of more specific, context-aware policy recommendations. Table IV outlines the interpretation of the clusters and the policy recommendations, connecting each cluster's main features to their implications for internet infrastructure development. Each cluster exhibits different internet infrastructure conditions, from regions with limited access and capacity to those with more developed infrastructure but service quality issues. Accordingly, tailored policy recommendations are proposed for each cluster, including expanding basic infrastructure, enhancing existing network capacity, and improving service quality.

TABLE III. CLUSTER INTERPRETATION AND POLICY RECOMMENDATIONS

Cluster	Dominant characteristic	Cluster interpretation	Planning implication	Policy recommendation
C0	Density≈546.3, Income≈USD 164.87, BTS≈2.9, Fiber≈1.0, Penetration≈59.9, Complaints≈10.6	Development Priority Region	Limited access and capacity indicate a strong need for infrastructure expansion.	Deploy new BTS sites, extend coverage, and introduce infrastructure incentives.
C1	Density≈2482.4, Income≈USD 288.66, BTS≈12.3, Fiber≈1.0, Penetration≈76.3, Complaints≈32.8	Transition Region	Basic infrastructure exists but remains suboptimal for demand stability.	Strengthen existing BTS capacity and improve service quality.
C2	Density≈445.5, Income≈USD 177.33, BTS≈3.0, Fiber≈0.0, Penetration≈52.4, Complaints≈8.0	Development Priority Region	Limited access and capacity indicate a strong need for infrastructure expansion.	Deploy new BTS sites, extend coverage, and introduce infrastructure incentives.
C3	Density≈3514.7, Income≈USD 381.93, BTS≈11.3, Fiber≈1.0, Penetration≈82.7, Complaints≈18.7	Transition Region	Basic infrastructure exists but remains suboptimal for demand stability.	Strengthen existing BTS capacity and improve service quality.
C4	Density≈2664.5, Income≈USD 324.02, BTS≈11.6, Fiber≈0.0, Penetration≈77.7, Complaints≈27.5	Mature Infrastructure Region	High usage intensity results in quality-of-service challenges.	Optimize capacity, upgrade backhaul/fiber, and apply demand-aware QoS management.

B. Discussion

Based on the experimental results, it was determined that combining GIS-based spatial analysis with spatial clustering methods provides a more comprehensive view of the heterogeneity in internet infrastructure conditions across regions than traditional planning methods. The spatial visualization and quantitative analysis shown in Figures 4 and 5 demonstrate that regions with similar demographic and socioeconomic features tend to form consistent spatial clusters, highlighting the significance of the geographic aspect in network infrastructure planning [17].

The key results from the $k = 5$ setup reveal distinct structural variations among clusters, particularly in population density, network infrastructure, and internet service coverage and quality. Clusters characterized by low population density and fewer BTS tend to have lower internet adoption rates. This supports earlier research indicating that physical infrastructure availability is crucial for narrowing the digital divide in developing areas [18, 19].

Using a cluster-based method can more objectively and data-drivenly identify key areas for development. On the other hand, clusters characterized by high population density and income tend to have higher internet penetration but also experience more service complaints. This pattern shows that greater access does not necessarily mean better service quality, particularly in high-usage areas. The results align with existing research, which indicates that urban or semi-urban regions often encounter issues such as capacity saturation and declining service quality despite having basic infrastructure [20, 21].

Therefore, planning strategies in these areas should focus more on capacity optimization and service quality management than on expanding the network. From the perspective of the cluster size distribution, it is evident that increasing the number of clusters does not necessarily yield a better cluster configuration. Although the configuration with $k = 5$ offers more detailed segmentation, it also makes interpretation more complex. This observation aligns with existing research on the trade-off between model accuracy and interpretability, indicating that more complex models often achieve higher accuracy but are less interpretable. Conversely, simpler, more interpretable models tend to have only slight performance reductions, making them more suitable for DSSs that demand transparency and ease of understanding for users [22, 23].

IV. CONCLUSION

This study developed and evaluated an explainable GIS-based DSS for planning internet infrastructure, incorporating spatial analysis, the spatial K-Means clustering method, and interpretability features. The clustering results show that areas with similar demographic, socioeconomic, and infrastructure characteristics tend to cluster spatially, highlighting the importance of geographic factors in network infrastructure planning. Experiments with different numbers of clusters showed that increasing the analysis's granularity results in more detailed regional segmentation but also makes interpretation more complex. These results confirm the trade-off between analytical depth and the clarity of the results in DSSs. In this context, a balanced cluster setup provides enough detailed information while remaining easy for decision-makers to understand, making it more useful for strategic policy decisions. Integrating interactive GIS visualization with quantitative analysis helps identify priority development areas, transition zones, and regions with relatively mature infrastructure but facing service quality issues.

Additionally, applying an explainable analytics approach improves system transparency by offering cluster-based explanations that clarify how key indicators contribute to each region's characteristics. This approach bridges the gap between technical analysis and practical decision-making needs. Moreover, this study shows that a GIS-based DSS combining spatial clustering and interpretability can support more adaptable, regionally focused, and understandable Internet infrastructure planning. The proposed method is not only useful for telecommunications network planning, but also has potential for other types of spatial infrastructure planning that require balancing analytical complexity with ease of interpretation.

Future studies might examine different spatial clustering techniques, such as density-based or hierarchical clustering, to better identify regional patterns in areas with uneven spatial distributions and more intricate cluster structures than K-Means.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the Ministry of Communications and Digital (KOMDIGI) and Statistics Indonesia (BPS) for providing access to the geospatial, infrastructure, and demographic datasets that formed the foundation of this research. They also extend their appreciation to industry experts and telecommunication providers whose valuable insights and feedback significantly enhanced the practical relevance of the proposed spatial recommendation model. Special thanks are conveyed to Diponegoro University for its academic support.

DATA AVAILABILITY STATEMENT

The dataset used in this study was compiled from official national sources, including the Ministry of Communication and Digital (KOMDIGI) and Statistics Indonesia (BPS). Due to institutional data-sharing agreements and regulatory constraints, the raw dataset is not publicly available. However, aggregated indicators and methodological details necessary to reproduce the analytical framework are available from the corresponding author upon reasonable request.

REFERENCES

[1] F. Xu and G. Peng, "Internet infrastructure, digital development and urban energy efficiency," *Journal of Digital Economy*, vol. 3, pp. 62–74, Dec. 2024, https://doi.org/10.1016/j.jdec.2024.11.001.

[2] R. Purnamasari, A. I. Hasanudin, R. Zulfikar, and H. Yazid, "Technological infrastructure and financial resource availability in enhancing public services and government performance: The role of digital innovation adoption in Indonesia," *Social Sciences & Humanities Open*, vol. 11, 2025, Art. no. 101621, https://doi.org/10.1016/j.ssaho.2025.101621.

[3] J. M. Graves, D. A. Abshire, S. Amiri, and J. L. Mackelprang, "Disparities in Technology and Broadband Internet Access Across Rurality: Implications for Health and Education," *Family & Community Health*, vol. 44, no. 4, pp. 257–265, Oct. 2021, https://doi.org/10.1097/FCH.0000000000000306.

[4] N. Aleisa, "Key factors influencing the e-government adoption: a systematic literature review," *Journal of Innovative Digital Transformation*, vol. 1, no. 1, pp. 14–31, Aug. 2024, https://doi.org/10.1108/JIDT-09-2023-0016.

[5] J. C. Kostelnick, J. B. Thayn, and K. Sinha, "Mapping and Spatial Analysis to Expand Rural Broadband Access," *Papers in Applied Geography*, vol. 10, no. 2, pp. 154–175, Apr. 2024, https://doi.org/10.1080/23754931.2024.2332238.

[6] X. Li *et al.*, "Development of Geographic Information System Architecture Feature Analysis and Evolution Trend Research," *Sustainability*, vol. 16, no. 1, Dec. 2023, Art. no. 137, https://doi.org/10.3390/su16010137.

[7] C. Zhou, "Exploring future GIS visions in the era of the scientific and technological revolution," *Information Geography*, vol. 1, no. 1, June 2025, Art. no. 100007, https://doi.org/10.1016/j.infgeo.2025.100007.

[8] P. B. Keenan and P. Jankowski, "Spatial Decision Support Systems: Three decades on," *Decision Support Systems*, vol. 116, pp. 64–76, Jan. 2019, https://doi.org/10.1016/j.dss.2018.10.010.

[9] I. Marzuki *et al.*, "Hierarchical Clustering-Based Geospatial Analysis for a Personalized Tourism Destination Recommender System," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 24794–24799, Aug. 2025, https://doi.org/10.48084/etasr.11055.

[10] D. E. Agustia, N. Fadhly, and C. Z. Oktaviani, "Clustering of Districts Based on Infrastructure Indicators Using K-Means And Average Linkage Methods," *International Journal of Science, Technology & Management*, vol. 6, no. 1, pp. 54–60, Jan. 2025, https://doi.org/10.46729/ijstm.v6i1.1193.

[11] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics*, vol. 9, no. 8, Aug. 2020, Art. no. 1295, https://doi.org/10.3390/electronics9081295.

[12] F. Mendler, B. Koch, B. Meißner, C. Voglstätter, and T. Smolinka, "Evaluation of spatial clustering methods for regionalisation of hydrogen ecosystems," *Energy Strategy Reviews*, vol. 57, Jan. 2025, Art. no. 101627, https://doi.org/10.1016/j.esr.2024.101627.

[13] E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, "Comprehensive Review of K-Means Clustering Algorithms," *International Journal of Advances in Scientific Research and Engineering*, vol. 07, no. 08, pp. 64–69, 2021, https://doi.org/10.31695/IJASRE.2021.34050.

[14] S. Ali *et al.*, "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence," *Information Fusion*, vol. 99, Nov. 2023, Art. no. 101805, https://doi.org/10.1016/j.inffus.2023.101805.

[15] A. Almtrf, "Integrating Explainable AI (XAI) Into Decision Support Systems: A Framework for Enhancing Transparency and Trust in Managerial Decision-Making," *International Journal of Managerial Studies and Research*, vol. 13, no. 9, pp. 9–22, 2025, https://doi.org/10.20431/2349-0349.1309002.

[16] L. Duan, Z. Gu, Y. Zhang, and Y. Chen, "From Clustered to Networked: Multi-Dimensional and Multi-Scale Performance Evaluation of Polycentric Urban Structure Evolution in Shenzhen, China," *Land*, vol. 14, no. 9, Sept. 2025, Art. no. 1899, https://doi.org/10.3390/land14091899.

[17] Y. Feng, J. Dai, and L. Zhang, "Digital infrastructure and income disparities: A quasi-natural experiment based on the 'Broadband China' strategy," *International Review of Economics & Finance*, vol. 102, Sept. 2025, Art. no. 104350, https://doi.org/10.1016/j.iref.2025.104350.

[18] Y. Hao, W. Liu, and M. Liu, "The Effect of Internet Infrastructure's Impact on Foreign Investment Inequality: Based on Digital Divide Perspective," *Sage Open*, vol. 15, no. 2, Apr. 2025, Art. no. 21582440251333183, https://doi.org/10.1177/21582440251333183.

[19] I. Maket, I. S. Kanó, and Z. Vas, "Quality of urban infrastructural service accessibility and human well-being in Sub-Saharan Africa," *World Development Sustainability*, vol. 4, June 2024, Art. no. 100155, https://doi.org/10.1016/j.wds.2024.100155.

[20] E. Syaodih, "The Challenges of Urban Management in Indonesia," presented at the Social and Humaniora Research Symposium (SoRes 2018), Mar. 2019, pp. 485–488, https://doi.org/10.2991/sores-18.2019.111.

[21] Z. Chen, F. Xiao, F. Guo, and J. Yan, "Interpretable machine learning for building energy management: A state-of-the-art review," *Advances in Applied Energy*, vol. 9, Feb. 2023, Art. no. 100123, https://doi.org/10.1016/j.adapen.2023.100123.

[22] F. Van Der Sluis and E. L. Van Den Broek, "Model interpretability enhances domain generalization in the case of textual complexity modeling," *Patterns*, vol. 6, no. 2, Feb. 2025, Art. no. 101177, https://doi.org/10.1016/j.patter.2025.101177.