Data-Driven Telecommunication Infrastructure: AI Clustering and Geodesic Measurement for Strategic Tower Optimization

1st Sadam Al Rasyid
School of Electrical Engineering
Telkom University
Bandung, Indonesia
sadamalrasyid@student.telkomuniversity.ac.id

2nd Suryo Adhi Wibowo
School of Electrical Engineering
Telkom University
Bandung, Indonesia
suryoadhiwibowo@telkomuniversity.ac.id

***Abstract*—The optimization of Base Transceiver Station (BTS) location is a major challenge in current urban areas, owing to fast population increase and rising need for high-performance communications networks. This paper describes a revolutionary strategy to BTS deployment that employs advanced clustering algorithms to improve network performance and coverage in densely populated urban locations. Four clustering algorithms are assessed, including K-Means, DBSCAN, Hierarchical Clustering, and K-Medoids, while taking into account urban variables such as housing density, land use, and geographic distribution. The paper makes two major contributions: dynamic change of the K-Means algorithm's cluster count and efficient centroid initialization using real-world urban data. Geodesic distance measures are used to examine the spatial relationships between BTS locations, resulting in more accurate and efficient tower deployment. Experimental results show that the modified K-Means algorithm beats the other techniques, with a Calinski-Harabasz index of 1662.46 and a Davies-Bouldin index of 0.868, showing improved cluster cohesiveness and separation. This technique lowers deployment costs while improving network coverage, resulting in more precise BTS placement and better resource use. These findings fill a gap in the literature by providing vital insights into data-driven urban optimization methodologies. They also have substantial implications for the planning and development of smart city infrastructure, furthering the future of wireless network architecture in urban contexts.**

***Index Terms*—telecommunication optimization, base transceiver station (BTS), clustering algorithms, geodesic measurement.**

I. INTRODUCTION

In Indonesia, the rise of urban populations, expected to reach 68% by 2035, intensifies the need for smart cities to address urban challenges through advanced infrastructure and technology. Cities like Bandung aim to improve the quality of life by integrating ICTs for efficient resource management and responsive public services [1]. Key initiatives, such as Bandung's Command Center, exemplify steps toward real-time urban monitoring, aligning with national goals for sustainable, digitally connected cities [2].

To meet the growing demands for connectivity, establishing robust infrastructure, particularly strategically positioned Base Transceiver Station (BTS) towers, is essential. These towers are the backbone of telecommunication networks, providing critical coverage and data transmission capabilities that enable seamless connectivity for millions of users. However, traditional methods for tower placement often need help in complex urban environments such as Bandung. The city's diverse geography and demographics necessitate adaptive, data-driven strategies for optimizing BTS placement. Specifically, these systems must consider population density, building layouts, and geographic features, significantly influencing coverage and signal distribution [3].

Clustering algorithms present a powerful method for uncovering spatial patterns within large datasets, enabling telecom providers to make informed decisions regarding tower placements that maximize coverage, minimize interference, and optimize deployment costs. This paper explores four prominent clustering techniques: K-Means, DBSCAN, Hierarchical Clustering, and K-Medoids, assessing their scalability, efficiency, and applicability to the complex urban environment of Bandung. The goal is to determine which method is most suited for supporting strategic infrastructure deployment in such a setting.

The appropriate number of clusters may be automatically determined, and cluster centers can be dynamically initialized thanks to a new addition to the K-Means algorithm, which incorporates a noise algorithm. This advancement significantly improves clustering performance, particularly in identifying urban hotspots more effectively [4], [5]. On the other hand, multi-density clustering algorithms like DBSCAN and its variants perform well in urban environments with fluctuating population densities, as they are capable of detecting multiple density regions and nested clusters, offering notable advantages in complex city layouts [6], [7], [8]. Several methods have been proposed to optimize hierarchical clustering, including one that uses centroids to represent groups of adjacent points, which reduces computational costs without sacrificing performance [9]. Another technique presents a hierarchical clustering algorithm that is extremely effective and runs in linear time. This method can be thought of as a hierarchical grid-based strategy [10]. Additionally, A refined K-Medoids algorithm demonstrates enhanced performance compared to traditional methods. It achieves improved accuracy and computational efficiency by progressively fine-tuning the medoid selection process and optimizing the number of clusters. This approach leads to better cluster cohesion and separation, making it particularly well-suited for this investigation [11].

The study's objectives are to conduct a comparative analysis and determine the best clustering technique to maximize telecom tower placement within the unique urban environment of Bandung. The findings suggest potential for broader applications in urban planning, offering adaptable and scalable solutions to support the ongoing development of digital infrastructure in Indonesia and globally.

II. RELATED WORK

A major research topic is optimizing BTS locations in urban environments, particularly to optimize coverage while reducing interference in densely populated regions. Numerous clustering methods have been investigated to enhance this procedure. For instance, Li et al. added noise-handling capabilities to the classic K-Means algorithm, improving its stability in crowded conditions and enabling it to detect urban hotspots [4].

To overcome the drawbacks of the standard DBSCAN method, which is sensitive to factors like ϵ and minPts, Liu et al. developed Multi-Scaled DBSCAN (M-DBSCAN). M-DBSCAN, an enhanced version of DBSCAN, adjusts these parameters locally to better handle clusters of varying densities and sizes. This approach reduces uncertainty in cluster identification and minimizes noise. When the technique was applied to geotagged data from cities like Madison, Wisconsin, and Washington, D.C., it demonstrated an efficient identification of sparse clusters, improving clustering accuracy in urban environments [8].

The computing efficiency of hierarchical clustering algorithms has increased recently. Innovations such as grid-based and centroid-based hierarchical algorithms have reduced the computing load without compromising accuracy, even though classic hierarchical clustering can be resource-intensive. Bouguettaya et al. presented a centroid-based agglomerative hierarchical clustering method that effectively manages large datasets, making it highly suitable for applications in urban planning [9].

The K-Medoids technique also works well for BTS placement optimization. Yu et al. suggested an enhanced version of K-Medoids that gradually improves medoid selection to increase clustering efficiency while maintaining placement accuracy. This makes the technique especially useful in urban settings, where selecting accurate cluster centers is crucial for successful BTS deployment [11].

Although these methods have contributed to BTS location optimization, a clear gap exists in terms of dynamic cluster count determination and centroid initialization for urban environments. While methods like K-Means, DBSCAN, and K-Medoids have been extensively studied, they do not address the challenges of automatic determination of the optimal number of clusters or real-time adjustment of cluster centers, which are essential for urban areas with fluctuating population densities and complex spatial layouts. Furthermore, while M-DBSCAN and hierarchical clustering algorithms have been explored for handling different densities, none have fully integrated dynamic parameter adjustments specific to urban hotspot identification. This study aims to bridge these gaps by introducing an enhanced K-Means algorithm with dynamic cluster count determination and centroid optimization, incorporating geodesic measurements for more accurate spatial clustering in urban environments. This approach offers significant improvements in clustering quality and BTS placement accuracy, particularly in complex urban landscapes such as Bandung.

This research compares these clustering algorithms to evaluate their applicability to Bandung's urban landscape. The study aims to provide important data-driven insights to help optimize telecommunications infrastructure in accordance with Indonesia's smart city development goals.

III. MATERIAL AND RESEARCH METHOD

A. Datasets

This study uses two primary datasets from *Open Data Bandung* for optimizing telecommunication tower placement in Bandung's urban environment:

- **Telecommunication Tower Data**: This dataset includes detailed information about the locations, types, and technical specifications of telecommunication towers in Bandung. It is essential for spatial analysis, helping assess the distribution of BTS and their effectiveness in network coverage. This dataset is the foundation for optimizing tower placement to improve connectivity [12].
- **Residential Data**: Sourced from *Open Data Bandung*, this dataset contains information on the number and distribution of residential buildings across Bandung. It provides vital insights into population density, a critical factor for assessing the demand for telecommunication services. This dataset also includes geographical coordinates, building heights, and floor area data. It is used to model the distribution of people and their proximity to existing telecommunication towers, which is crucial for strategically positioning BTS towers. The data was cleaned and normalized to ensure consistency, and missing values were imputed based on statistical methods to ensure completeness [13].

The datasets were carefully processed, cleaned, and integrated to ensure alignment, establishing a solid foundation for the subsequent evaluation.

B. Data Collection and Preprocessing

The datasets used in this study were sourced from *Open Data Bandung*, which provides geospatial information related to telecommunication tower locations and residential areas within Bandung. The preprocessing of these datasets involved several key steps to ensure data quality and suitability for clustering analysis.

1) Data Cleaning: The raw datasets were subjected to a comprehensive cleaning process to eliminate any anomalies. Outliers and erroneous entries, such as inaccurate geographic coordinates or invalid technical specifications of telecommunication towers, were identified and rectified. Missing values in the residential dataset were addressed by suitable imputation methods, specifically mean imputation for continuous variables and mode imputation for categorical variables. This step ensured the integrity and consistency of the data for subsequent analysis.

2) Normalization: To address the issue of disparate data scales, particularly the geographic coordinates and technical specifications, normalization techniques were applied. The geographical data were normalized using min-max scaling to bring all values into a comparable range. For the technical specifications of the telecommunication towers, Z-score normalization was employed to standardize the data distribution. This preprocessing step was essential in averting characteristics with greater numerical ranges from significantly affecting the clustering outcomes.

3) Data Integration: Following the preprocessing steps, the telecommunication tower and residential data were integrated within a Geographic Information System (GIS). This integration facilitated spatial analysis and allowed for the visualization of the data in relation to geographic proximity and population density. By combining these datasets in a GIS environment, a comprehensive spatial understanding of the urban infrastructure was achieved, providing a robust foundation for the clustering analysis aimed at optimizing telecommunication tower placement.

C. Overview of Clustering Algorithms

This study employs four clustering techniques: K-Means, Hierarchical Clustering, DBSCAN, and K-Medoids, each utilizing a distinct approach for partitioning the data:

- **K-Means:** By minimizing the objective function, the K-Means algorithm seeks to divide data into K clusters, as shown in:

$$\min_{\boldsymbol{c_1},\ldots,\boldsymbol{c_K}} \sum_{i=1}^{K} \sum_{x_j \in S_i} \|x_j - \boldsymbol{c_i}\|^2 \tag{1}$$

  In this formulation, K denotes the total number of clusters, where S_i represents the i-th cluster, x_j refers to a data point in cluster S_i, and $\boldsymbol{c_i}$ stands for the centroid of S_i. The term $\|x_j - \boldsymbol{c_i}\|^2$ represents the squared Euclidean distance between the data point x_j and its associated centroid $\boldsymbol{c_i}$. The clustering process aims to minimize the objective function in (1).
- **DBSCAN:** DBSCAN uses the density of data points to find clusters. According to this method, a core point is comprised of at least MinPts points at a specific radius ϵ. The core distance p of a point is defined as follows:

$$\text{CoreDist}(p) = \min_{\{q \in N_\epsilon(p)\}} \|p - q\| \tag{2}$$

  where $N_\epsilon(p)$ represents the neighborhood of p within radius ϵ, and the Euclidean distance between points p and q is $\|p-q\|$. The core distance is calculated as shown in (2).
- **Hierarchical Clustering:** By repeatedly combining the nearest clusters according to a distance measure, hierarchical clustering creates a structure resembling a tree (dendrogram). The following formula determines the separation between two points x_i and x_j:

$$D(i,j) = \|x_i - x_j\|_2 \tag{3}$$

  where the Euclidean distance is indicated by $\|\cdot\|_2$. Every data point is first regarded as a separate cluster. Until the required number of clusters is achieved, the algorithm then gradually merges the closest clusters according to the smallest pairwise distance, as indicated in (3).
- **K-Medoids**: K-Medoids clustering, like K-Means clustering, uses real data points, or medoids, to indicate the cluster center instead of the mean. Reducing the overall dissimilarity within the clusters is the goal. The K-Medoids cost function has the following definition:

$$J' = \sum_{k=1}^{K} \sum_{\mathbf{x} \in S_k} \delta(\mathbf{x}, \mathbf{m}_k) \tag{4}$$

  In this formulation, K denotes the total number of clusters, and S_k denotes the set of data points within the k-th cluster. Each point $\mathbf{x}$ within S_k is evaluated against the cluster's medoid $\mathbf{m}_k$. The function $\delta(\mathbf{x}, \mathbf{m}_k)$ quantifies the dissimilarity between a data point $\mathbf{x}$ and its corresponding medoid $\mathbf{m}_k$. By minimizing J' in (4), this approach seeks to reduce the overall dissimilarity within clusters, leading to more compact and representative clusters.

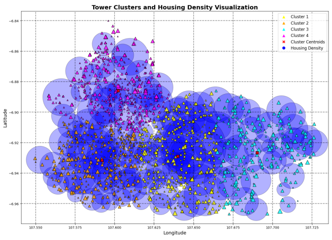
The applied clustering algorithms segment the data, facilitating its analysis to determine the optimal placement of Base Transceiver Stations (BTS) by leveraging both telecommunication tower and residential datasets. The outcomes of the four AI-based models are effectively illustrated in Fig. 1, demonstrating their performance in partitioning the data for enhanced network planning.

D. Geodesic Measurement Method

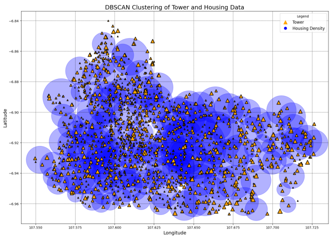
The Vincenty formula is used in this study to determine the geodesic distance between telecom towers [19]. This method accounts for the Earth's ellipsoidal shape, providing a more accurate measurement than traditional Euclidean distance calculations. The formula is implemented in Python using the `geopy` library.

The Vincenty formula calculates the separation distance d between two locations on the Earth's surface based on their latitudinal and longitudinal coordinates (ϕ_1, λ_1) and (ϕ_2, λ_2), as shown in (5).

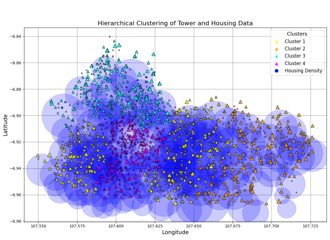
$$d = \text{geodesic}(\phi_1, \lambda_1, \phi_2, \lambda_2) \tag{5}$$



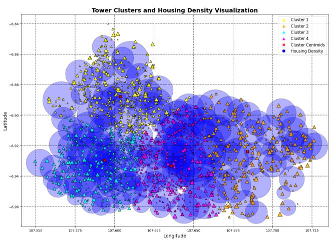
(a) K-Means clustering result



(b) DBSCAN clustering result



(c) Hierarchical clustering result



(d) K-Medoids clustering result

Fig. 1. Visualization of various clustering techniques: (a) K-Means, (b) DBSCAN, (c) Hierarchical, (d) K-Medoids.

Where ϕ_1, λ_1 represent the latitude and longitude of the first point, and ϕ_2, λ_2 correspond to the second point. The resulting distance is provided in kilometers.

Table I presents sample distances between Tower 1479 and several neighboring towers, all calculated using the Vincenty geodesic method.

TABLE I: Sample Distances Between Tower 1479 and Neighboring Towers

Tower ID 1	Tower ID 2	Distance (km)
1479	1530	8.08
1479	1531	11.07
1479	1532	6.29
1479	1533	2.52
1479	1534	0.38
1479	1535	4.78
1479	1536	10.54
1479	1537	10.88
1479	1538	5.98
1479	1539	8.93
1479	1540	9.54
1479	1541	4.81
1479	1542	8.09
1479	1543	8.50

The spatial relationships between the towers are further illustrated through a heatmap Fig. 2. This heatmap visually represents the distance matrix, with shorter distances in darker shades, allowing for easy identification of nearby towers.

E. Evaluation Metrics

The clustering models' performance was assessed using several widely used Evaluation measures, including Davies-Bouldin, Calinski-Harabasz, and Silhouette. These measures add to our understanding of the quality and effectiveness of clustering results. Fig. 3 summarizes the performance outcomes

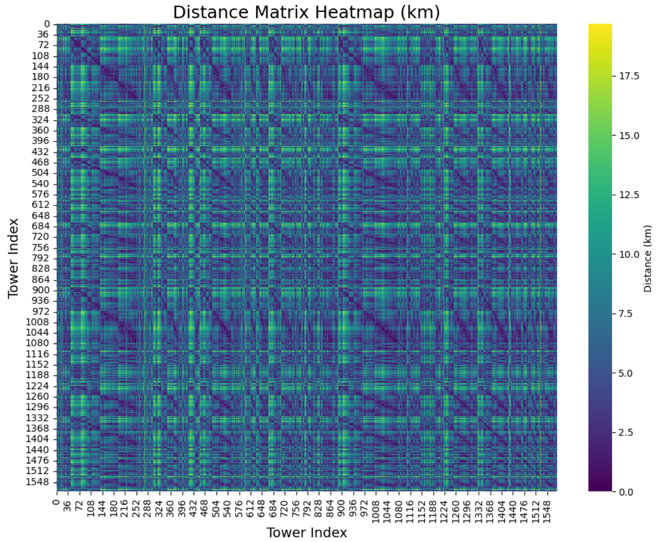


Fig. 2. Distance matrix heatmap for telecommunication towers.

- **Silhouette Score:** The silhouette is frequently used in k-means clustering to figure out how many groups are appropriate. A higher Silhouette Score indicates well-defined clusters, with the best value of k yielding the highest score, thereby ensuring optimal clustering performance [14], [15].
- **Calinski-Harabasz Score:** In order to evaluate the quality of clustering, this metric calculates the separation between clusters as well as the compactness inside them. Better performance is shown by higher values, which show that the clusters are closer together and more distinct from one another. [16], [17].
- **Davies-Bouldin Score:** By computing the average similarity between each cluster and its closest comparable counterpart, the Davies-Bouldin Score assesses how different a cluster is. Better separation between clusters is shown by a lower Davies-Bouldin Score, which shows the clusters are more distinct and less alike. [18].

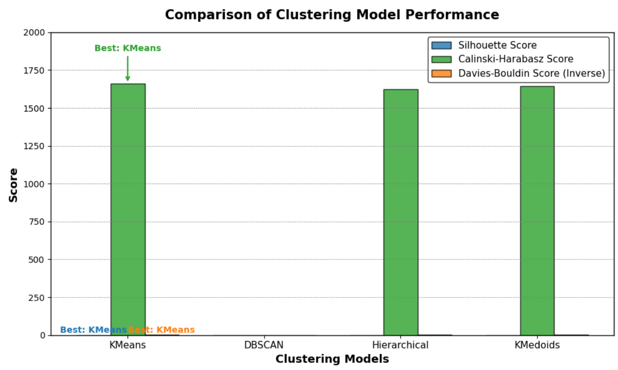


Fig. 3. Clustering performance comparison using Silhouette, Calinski-Harabasz, and Davies-Bouldin metrics, highlighting K-means as the best method.

Compared to other models, the K-means clustering method performs better across all evaluation criteria, as shown in Fig. 3. K-means, in particular, had the highest Calinski-Harabasz Score, suggesting improved cluster separation and cohesion. It also had the lowest Silhouette and Davies-Bouldin Scores, demonstrating its capacity to create clear, distinct clusters. These results demonstrate how well K-means clusters consistently outperform other models in clustering quality across all three parameters.

The comparative analysis decisively demonstrates that K-means provides this dataset's most effective clustering solution. It routinely beats other algorithms in terms of cluster cohesiveness & separation, making it the ideal candidate for this investigation.

IV. Results and Discussion

This section provides an analytical assessment of the clustering performance on the telecom tower dataset. Davies-Bouldin, Calinski-Harabasz, and Silhouette were the three evaluation metrics used. Every metric offers a different perspective on how cohesive, distinct, and separated the clusters are.

A. Clustering Model Evaluation

The efficiency of four cluster methods — K-means, Hierarchical, DBSCAN, and K-Medoids—was evaluated, as shown in Table II. With the lowest Davies-Bouldin Score and the highest Calinski-Harabasz and Silhouette Scores, K-means continuously outperformed the others, demonstrating its ability to create distinct clusters.

TABLE II: Comparison of Clustering Model Performance Across Evaluation Metrics

Model	Silhouette	Calinski-Harabasz	Davies-Bouldin
K-means	0.446	1662.46	0.868
DBSCAN	-1.000	-1.000	-1.000
Hierarchical	0.377	1622.11	0.885
K-Medoids	0.441	1642.85	0.877

B. Discussion

Fig. 4 illustrates the spatial distribution of telecommunication towers based on the results of K-means clustering applied to the dataset. The figure highlights the geographic distribution of the towers, plotted according to their latitude and longitude, with distinct clusters representing regions of high residential density. These clusters offer insight into the strategic placement of towers and how they are influenced by population distribution, ensuring optimal network coverage. The K-means algorithm partitions the towers into meaningful groups that align with urban residential patterns, allowing for more efficient network planning.

The K-means clustering algorithm effectively groups telecommunication towers based on proximity to residential areas and coverage potential. It identifies optimal tower locations in densely populated areas, with taller towers placed strategically to extend signal coverage, considering geographical factors like elevation.

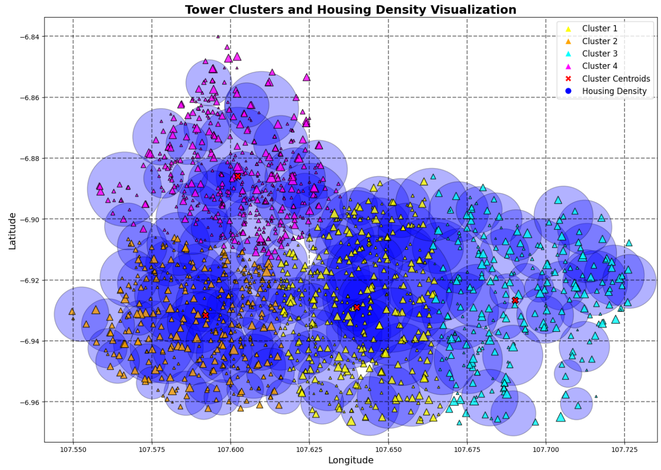


Fig. 4. Spatial distribution of telecommunication towers: results of K-means clustering.

The analysis of the x-axis (longitude) and y-axis (latitude) in the plot can guide the identification of potential sites for future tower placements. This approach underscores K-means' capability to integrate both geographic and demographic variables, offering actionable insights for informed decision-making in telecommunication infrastructure development.

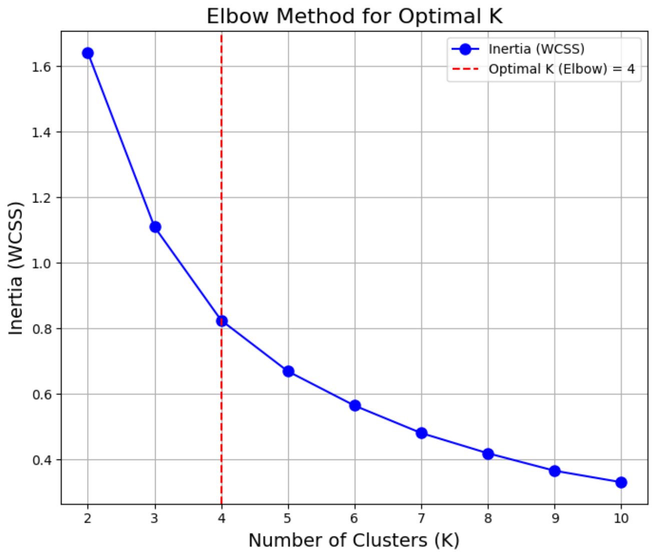


Fig. 5. Elbow method for determining the optimal number of clusters (K = 4).

The Elbow approach, a crucial strategy for figuring out the ideal number of clusters for K-means, as illustrated in Fig. 5. The curve shows an inflection point at $K = 4$, which indicates the ideal number of clusters to balance computing efficiency and clustering quality. This value of K was selected to refine the clustering process and improve the precision of tower placement recommendations, thereby enhancing the overall network planning and coverage optimization.

V. CONCLUSION

This study demonstrates that the enhanced K-Means clustering algorithm outperforms DBSCAN, Hierarchical Clustering, and K-Medoids in optimizing telecommunication tower placement in urban environments. The evaluation measures, including the Calinski-Harabasz index (1662.46) and the Davies-Bouldin index (0.868), reveal that K-Means has superior intra-cluster cohesiveness and inter-cluster separation. By considering factors such as population density, geographic features, and urban infrastructure, this approach improves network coverage and signal distribution. The findings suggest that K-Means is a robust method for strategic telecommunication planning and can play a significant role in smart city development.

This work highlights the potential of the K-Means algorithm for optimizing telecommunications infrastructure, which is crucial for enhancing network reliability and coverage in complex urban settings. Future research could focus on integrating real-time data analytics for dynamic adaptation to changing urban environments. Furthermore, combining K-Means with reinforcement learning techniques for long-term optimization and applying this method to other cities could expand its applicability. AI-driven predictive models for urban growth could also enhance the accuracy of tower placement, contributing to the effective development of telecommunication infrastructure globally.

ACKNOWLEDGMENT

The authors sincerely thank Telkom University's Image Processing and Vision Laboratory (IMV) for all of their help and support during this study. The success of this endeavor was facilitated by their resources and guidance.

REFERENCES

[1] A. Nuraeni, H. S. Firmansyah, G. S. Pribadi, A. Munandar, L. Herdiani, and N. Nurwathi, "Smart city evaluation model in Bandung, West Java, Indonesia," Oct. 2019, doi: https://doi.org/10.1109/tssa48701.2019.8985465.

[2] T. Ahmad, "Pusat Komando Bandung: Ruang kontrol ala film Star Trek," 2015. [Online]. Available: http://www.infobdg.com/v2/bandung-perintah-pusat-ruang-kontrol-Ala-film-bintang-perjalana. [Accessed: Nov. 8, 2024].

[3] V. Barrile, G. Armocida, and G. Bilotta, "GIS supporting the plan of BTS (base transceiver stations) for mobile network in urban context," *WSEAS Trans. on Communications*, vol. 8, no. 8, pp. 775–784, Aug. 2009.

[4] Y. Li, X. Zhou, J. Gu, K. Guo, and W. Deng, "A novel K-means clustering method for locating urban hotspots based on hybrid heuristic initialization," *Applied Sciences*, vol. 12, no. 16, p. 8047, Aug. 2022, doi: https://doi.org/10.3390/app12168047.

[5] X. Ran, X. Zhou, M. Lei, W. Tepsan, and W. Deng, "A novel K-means clustering algorithm with a noise algorithm for capturing urban hotspots," *Applied Sciences*, vol. 11, no. 23, p. 11202, Nov. 2021, doi: https://doi.org/10.3390/app112311202.

[6] E. Cesario, P. Lindia, and A. Vinci, "Detecting multi-density urban hotspots in a smart city: Approaches, challenges and applications," *Big Data and Cognitive Computing*, vol. 7, no. 1, p. 29, Feb. 2023, doi: https://doi.org/10.3390/bdcc7010029.

[7] X. Li, P. Zhang, and G. Zhu, "DBSCAN clustering algorithms for non-uniform density data and its application in urban rail passenger aggregation distribution," *Energies*, vol. 12, no. 19, p. 3722, Sep. 2019, doi: https://doi.org/10.3390/en12193722.

[8] X. Liu, Q. Huang, and S. Gao, "Exploring the uncertainty of activity zone detection using digital footprints with multi-scaled DBSCAN," *Int. J. of Geographical Information Science*, vol. 33, no. 6, pp. 1196–1223, Jan. 2019, doi: https://doi.org/10.1080/13658816.2018.1563301.

[9] A. Bouguettaya, Q. Yu, X. Liu, X. Zhou, and A. Song, "Efficient agglomerative hierarchical clustering," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2785–2797, Apr. 2015, doi: https://doi.org/10.1016/j.eswa.2014.09.054.

[10] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: An overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, Dec. 2011, doi: https://doi.org/10.1002/widm.53.

[11] D. Yu, G. Liu, M. Guo, and X. Liu, "An improved K-medoids algorithm based on step increasing and optimizing medoids," *Expert Systems with Applications*, vol. 92, pp. 464–473, Feb. 2018, doi: https://doi.org/10.1016/j.eswa.2017.09.052.

[12] Dinas Komunikasi dan Informatika Kota Bandung, "Open Data," bandung.go.id. https://opendata.bandung.go.id/dataset/lokasi-menara-telekomunikasi-di-kota-bandung. [Accessed: Nov. 8, 2024].

[13] Dinas Komunikasi dan Informatika Kota Bandung, "Open Data," bandung.go.id. https://opendata.bandung.go.id/dataset/jumlah-rumah-di-kota-bandung-2. [Accessed: Nov. 8, 2024].

[14] L. Sai, M. Shreya, A. Subudhi, B. Lakshmi, and K. Madhuri, "Optimal K-means clustering method using silhouette coefficient," *Int. J. of Applied Research on Information Technology and Computing*, vol. 8, pp. 335–344, 2017, doi: https://doi.org/10.5958/0975-8089.2017.00030.6.

[15] K. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *Proc. 2020 IEEE 7th Int. Conf. on Data Science and Advanced Analytics (DSAA)*, 2020, pp. 747–748, doi: https://doi.org/10.1109/DSAA49011.2020.00096.

[16] A. Rachwał, E. Popławska, I. Gorgol, T. Cieplak, D. Pliszczuk, Ł. Skowron, and T. Rymarczyk, "Determining the quality of a dataset in clustering terms," *Applied Sciences*, 2023, doi: https://doi.org/10.3390/app13052942.

[17] X. Wang and Y. Xu, "An improved index for clustering validation based on silhouette index and Calinski-Harabasz index," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 569, p. 052024, 2019, doi: https://doi.org/10.1088/1757-899X/569/5/052024.

[18] A. Rachwał, E. Popławska, I. Gorgol, T. Cieplak, D. Pliszczuk, Ł. Skowron, and T. Rymarczyk, "Determining the quality of a dataset in clustering terms," *Applied Sciences*, 2023, doi: https://doi.org/10.3390/app13052942.

[19] G. Bahati and E. Masabo, "Optimizing ambulance location based on road accident data in Rwanda using machine learning algorithms," *Research Square*, Nov. 2024, doi: https://doi.org/10.21203/rs.3.rs-5319700/v1.

[20] D. Aouladhadj, E. Kpre, V. Deniau, A. Kharchouf, C. Gransart, and C. Gaquière, "Drone detection and tracking using RF identification signals," *Sensors*, vol. 23, no. 17, p. 7650, Jan. 2023, doi: https://doi.org/10.3390/s23177650.