Data Collection for Nearest Public Facility Using Ball Tree Algorithm and Google Maps API

H Ramadhan[1]*

[1] BPS-Statistics Indonesia Kalimantan Barat Province, West Kalimantan, Indonesia

*Corresponding author's email: handika.ramadhan@bps.go.id

Abstract. Accessibility to public facilities is a crucial factor in regional development, including at the village level as the smallest administrative unit. The Central Bureau of Statistics (BPS) currently collects data on public facilities and their distances to village offices through interviews, making the results dependent on respondents' perceptions. This research aims to measure the nearest distance from village offices to public schools by utilizing the BallTree algorithm and the Google Maps API. The dataset consists of 128 village offices and a list of public schools classified into four categories. BallTree was used to filter the nearest school candidates within a given radius, after which the route distance of the ten nearest candidates was calculated using the Google Maps Distance Matrix API to identify the school with the nearest route distance based on the road network. The findings show that straight-line distance often aligns with route distance, although not at all, highlighting the importance of Google Maps route calculation. This research concludes that combining BallTree and the Google Maps API improves computational efficiency while providing objective and reliable information.
Keyword: Accessibility, Ball Tree, Distance Matrix, Google Maps API.

1. Introduction

Public facilities are fundamental elements in regional development. The availability of basic infrastructure such as roads, schools, markets, healthcare centers, and other public services not only serves a physical function but also reflects the overall quality of life of the community [1]. The proximity of public facilities can enhance access to economic opportunities, accelerate regional growth, and reduce disparities between areas. Conversely, regions that are distant from public facilities often face significant challenges in improving the well-being of their residents. Therefore, it is essential for every region to possess, or at least be located near, adequate public facilities.

Villages, as the smallest administrative units, hold a strategic role in regional development as they serve as the frontline for delivering services directly to the community. It is therefore important for each village to have adequate public facilities, as their availability determines the quality of life at the local level. The Central Bureau of Statistics (BPS) collects data on the availability of public facilities in each village through the Village Potential Survey (PODES) [2]. In this survey, BPS records the presence of public facilities within a village, and if certain facilities are absent, BPS records the distance to thenearest facility in another village. This approach highlights that accessibility to public facilities, whether within the village itself or in surrounding areas, is a key indicator of a village's level of development. However, the measurement of the nearest distance is obtained solely through direct questioning of village officials, making the results prone to subjectivity.

Nevertheless, the method used by BPS to collect data on the distance to the nearest public facility still has limitations, as the measurement is primarily obtained through direct questioning of village officials or other respondents [3]. While this approach is relatively fast and practical, it is prone to bias because the reported distances are often approximate and may not accurately reflect actual conditions in the field. In fact, with current technological advancements, both distance and travel time can be measured more precisely, one of which is through Google Maps, enabling a more accurate calculation of distances [4].

One of the key strengths of Google Maps is the availability of Application Programming Interfaces (APIs), which allow researchers and practitioners to calculate distances between an origin point and multiple destination points by utilizing geographic coordinates. Through these APIs, accessibility to public facilities can be measured in a more objective and standardized manner compared to relying on subjective reports [4]. However, the use of Google Maps APIs also presents certain challenges. If calculations are carried out for all destinations simultaneously, the computational process becomes lengthy and complex. Moreover, the Google Maps API imposes usage limits within specific time frames, which means that attempts to calculate distances to all facilities at once may exceed the service's capacity [5]. This condition requires a more efficient approach to obtain accurate results while staying within the API limits.

To address these efficiency limitations in the use of Google Maps API, one applicable approach is the BallTree algorithm [6]. This algorithm was developed to accelerate nearest-neighbor searches by partitioning a set of points into a tree structure based on hyperspheres, allowing the search process to run faster than direct distance calculations to all points [7]. In the context of spatial data on villages and public facilities, BallTree can be used to filter facilities located within a certain radius from the origin village. In other words, BallTree functions as an initial filter to determine geographically relevant candidate facilities without requiring distance calculations for all destinations.

Once the list of nearest candidates is obtained through BallTree, the next step is to calculate their distances more precisely using the Google Maps API. With this strategy, only the preselected points are sent to the API, significantly reducing the number of requests. This not only conserves the API usage quota but also shortens computation time, particularly when analyzing a large number of villages and public facilities. The integration of BallTree as a spatial filter and Google Maps API as a precise distance calculator provides an effective combination for producing accessibility data that is more accurate, efficient, and compliant with the API's usage limits.

Based on this discussion, there is a clear gap between the conventional method of collecting public facility data through surveys and interviews and the potential use of modern spatial technologies that are more accurate and efficient. While BPS has made efforts to document the availability and distance of public facilities, its current method still relies on subjective reports from village officials. Meanwhile, although the Google Maps API offers precision, it faces challenges of efficiency and usage limitations. Therefore, an alternative approach is needed that combines the efficiency of algorithms such as BallTree with the accuracy of the Google Maps API.

Considering this gap, the present research is directed at offering an alternative approach to collecting data on public facilities. The main objective of this research is to utilize the BallTree algorithm as an initial filter for identifying candidate nearest facilities, followed by integration with the Google Maps

API to calculate route distances with greater precision. Through this strategy, it is expected to establish a method that is not only efficient in terms of computational resources and API quota but also capable of producing more reliable accessibility data to support analysis and regional development planning at the village level.

2. Research Method

2.1. Scope of Research

This research is subject to several defined scopes in order to ensure a more focused analysis. First, the type of public facility examined is limited to public schools. The selection of public schools is based on the consideration that their data are relatively more accessible, including official school registries as well as spatial coordinates of their locations. In addition, the names of public schools are generally unique within the same regency, which facilitates the process of spatial data identification and matching.

Second, the research focuses on Kubu Raya Regency as the research area. The choice of this location is motivated by its diverse characteristics, encompassing urban zones, rural areas [8], and regions with geographical barriers such as rivers that cannot be entirely traversed by land routes. With these conditions, Kubu Raya Regency is considered representative for testing methods of public facility search in a spatially heterogeneous context. Therefore, the findings of this research are expected to provide a more comprehensive understanding of the challenges and opportunities in applying the proposed method.

Another limitation of this research is that the route distance calculation relies solely on Google Maps as the data source, even though alternative platforms such as OpenStreetMap (OSM) or other GIS-based routing services are available. The selection of Google Maps was made because it provides a consistent and easily verifiable reference, as the calculated routes can be directly validated using the publicly accessible Google Maps application. This approach ensures transparency and reproducibility of the results, although it may exclude certain rural areas that are not fully covered by Google's road network data.

2.2. Data Input Collection

In this research, two primary datasets were used: village office data and public school data. The list of villages was obtained from the rural–urban classification released by Statistics Indonesia (BPS), resulting in 128 villages [9]. Meanwhile, public school data were collected from the Ministry of Education's Dapodik website, comprising 450 records across four categories: elementary schools (SD), junior high schools (SMP), senior high schools (SMA), and vocational high schools (SMK). The geographic coordinates of both datasets were retrieved using the Google Maps Places API.

For village office coordinates, the query was constructed by adding the keyword "Kantor Desa" (village office) before the village name and ", Kabupaten Kubu Raya" after it. For public schools, the query was generated by appending ", Kabupaten Kubu Raya" after the school name. Following the coordinate search with the Places API, the data were validated through several steps: checking for missing location results, identifying duplicate entries, and mapping to ensure that all points were located within Kubu Raya Regency. If such issues were detected, corrections were made through manual searches.

2.3. Utilization of Google Maps API

In this research, the Google Maps API was utilized for two main purposes: obtaining the geographic coordinates of public facilities and calculating travel distances between locations. Two modules were employed: the Places API and the Distance Matrix API.

First, the Places API was used to collect the coordinates of village offices as origin points and public schools as destination points. The Places API enables place searches based on keywords, categories, and specific search radii. By providing appropriate keywords, such as “Kantor Desa Sungai Ambangah” or “SMP Negeri 1 Sungai Kakap, Kabupaten Kubu Raya”, the API returns information including the place name, address, and latitude–longitude coordinates. The choice of public schools as research objects was supported by the uniqueness of their names within the regency, which minimizes the risk of duplication and facilitates automated searches through the API.

Second, the Distance Matrix API was applied to compute travel distances and times from each village office to the selected public schools. Unlike Euclidean methods that measure straight-line distance, the Distance Matrix API relies on the actual road network to generate realistic routes. This feature is crucial because accessibility is not only determined by geographic proximity but also by road connectivity. The advantages of using Google Maps’ road network can be summarized as follows:1. Spatial realism: Google Maps incorporates existing road networks, producing distance measurements that better reflect real accessibility experienced by communities; 2. Geographic barrier adaptation: In areas such as Kubu Raya Regency, rivers may obstruct overland travel. Straight-line distance disregards such barriers, whereas Google Maps API adjusts routes according to the available road network; 3. Scalability: The Distance Matrix API can calculate distances from a single origin to multiple destinations in one request, supporting large-scale spatial analysis.

Thus, the combination of the Places API and the Distance Matrix API provides a critical foundation for this research. The Places API ensures that the coordinates of villages and schools are obtained accurately and consistently, while the Distance Matrix API enables the calculation of route distances that are more realistic compared to straight-line distance methods. The results serve as the primary input for spatial data–based analyses of public facility accessibility.

2.4. Using the Ball Tree Algorithm

Ball Tree is a data structure designed to accelerate the process of nearest neighbor search. In general, a Ball Tree partitions a dataset into a hierarchical tree, where each node represents a circle (ball) that encloses a subset of data points [6]. The partitioning is performed recursively until smaller nodes are formed, making the search process more efficient. With this approach, the nearest neighbor search does not require direct comparison with all data points (brute force), but only with nodes deemed relevant.

In the context of spatial data, each point stored in the Ball Tree structure represents geographic coordinates, such as latitude and longitude of villages or public facilities. The Ball Tree algorithm can calculate distances between coordinates using various metrics, such as Euclidean distance or haversine distance, which is more suitable for Earth’s curvature. By leveraging Ball Tree, the search for nearest points within spatial datasets containing thousands or even millions of records can be performed much faster than conventional methods.

One way to understand how Ball Tree operates is through the visual representation of space partitioned into hierarchical circles (balls). Each circle represents a cluster of points, and the partitioning is performed recursively until smaller circles are formed at the leaf level. With this approach, nearest neighbor search or radius-based queries can be executed more efficiently, as the algorithm only traverses relevant circles. Figure 1 illustrates how a set of spatial points is divided into nested circles to construct the hierarchical structure of a Ball Tree.

ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025

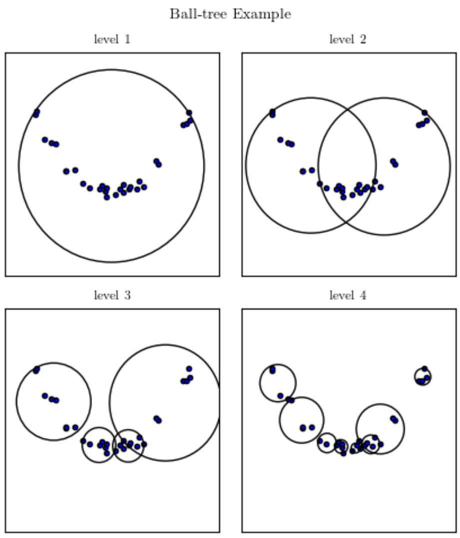

Figure 1. Illustration of spatial partitioning using Ball Tree (AstroML)

From Figure 1, it can be seen that searching within a certain radius is performed by utilizing the hierarchical partitioning of circles. Circles that are completely outside the search radius can be ignored, while circles that are inside or intersecting with the radius are further explored. In this way, the algorithm accelerates the process of finding points within the radius without the need to check all data directly.

In this research, Ball Tree is employed as an initial stage to filter public facility points located within a certain radius from the village office. The candidate points obtained from the Ball Tree selection then serve as inputs for the Distance Matrix API to calculate more precise route distances. This strategy reduces the number of requests sent to the Google API, which is limited by quota and cost, while still maintaining accuracy in distance calculation. Thus, Ball Tree functions as an efficient spatial filter before proceeding to the actual route distance computation using the Google Maps API.

2.5. Research Workflow

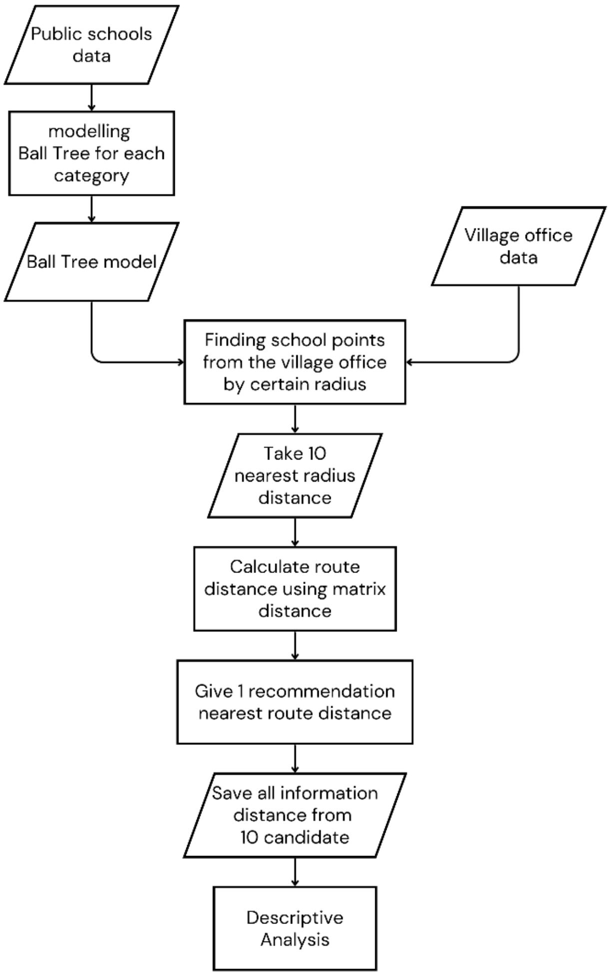

Figure 2. Research Workflow

Figure 2 shows the research workflow of this research. The initial stage begins with the collection of input data, namely public school data and village office data, each of which has been supplemented with geographic coordinate information as described in the data input collection section.
The next step is the construction of Ball Tree models based on public school data categorized into four different levels, resulting in four Ball Tree models according to their respective categories. These models serve to accelerate spatial searches, allowing the identification of sets of schools located within a certain radius from each village office.

Subsequently, for each Ball Tree model, a search is performed to identify schools within a given radius from the village office. This process produces 512 search rows, derived from 128 village offices multiplied by the four school categories. Each search row returns the ten nearest schools based on straight-line distance. Thus, ten candidate schools that are geographically closest to each village office are obtained.

The next stage involves calculating route distances using the Google Maps Distance Matrix API. The ten candidate schools from each village office are used as inputs for the API, and this process returns the actual route distances based on the road network. This calculation enables the identification of the school with the nearest route distance, which provides a more realistic measure than straight-line distance alone. However, if no valid route is found among the candidates (e.g., due to rivers or limited road infrastructure), the nearest school is determined using straight-line distance from the Ball Tree output.

The output of this process is stored as complete information on the nearest school from each village office. The stored data includes not only the selected school but also all ten nearest candidates, including school names, straight-line distances, and Google Maps route distances. This comprehensive information is important as it allows users or decision-makers to manually validate the results if necessary.

The final step of this research is a descriptive analysis of the collected information. The analysis aims to derive insights into the accessibility of educational facilities from each village office and its variation across different geographical contexts, particularly contrasting the outcomes between urban and rural villages.

3. Result and Discussion

For each village office, the nearest public schools were searched across all categories and levels. As a result, the dataset of nearest distance searches from village offices to public schools consists of 512 rows. In each row, ten candidate facilities were identified based on the nearest straight-line distance. Subsequently, route distances from the village office to these ten candidates were calculated using Google Maps. The nearest facility was then determined based on the route distance. Information including the school name, straight-line distance, and Google Maps route distance was recorded as a recommendation of the nearest facility, which can also be used for manual verification if necessary.

This study reduces the exhaustive search approach of distance computation using the Google Maps API by introducing a filtering mechanism based on the Ball Tree algorithm. Instead of querying all possible facility pairs, the process is limited to the nearest distance candidates identified through spatial indexing. This reduction significantly minimizes the number of API requests while maintaining the accuracy of the nearest-route identification, particularly in large-scale spatial datasets such as village-level accessibility analysis.

Not all distance searchings can be measured using Google Maps

In the search for route distances using Google Maps, there were rows in which all ten nearest school candidates returned null values, indicating that no measurable road network was available in Google Maps. As a result, 64 search rows yielded no measurable route distance. Figure 3 shows that when classified by urban–rural categories, only one case came from an urban area, while the remaining 63 were from rural areas. Furthermore, when the searches were broken down by school level, the results are presented in Figure 4. These findings indicate that the higher the level of education, the greater the number of public schools that cannot be measured using Google Maps. This suggests that such facilities are accessible only through smaller roads or, in some cases, not accessible by land routes at all.

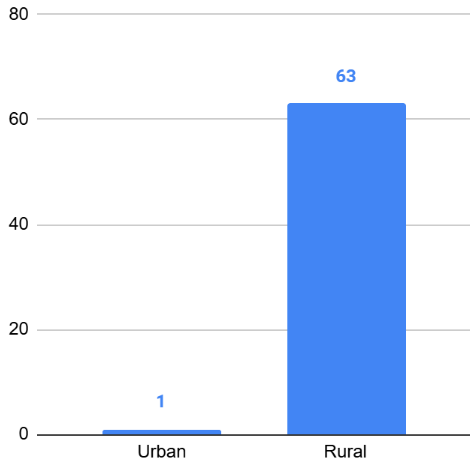

Figure 3. Number of nearest-distance searchings from village offices to public schools that could not be measured using Google Maps by urban–rural classification

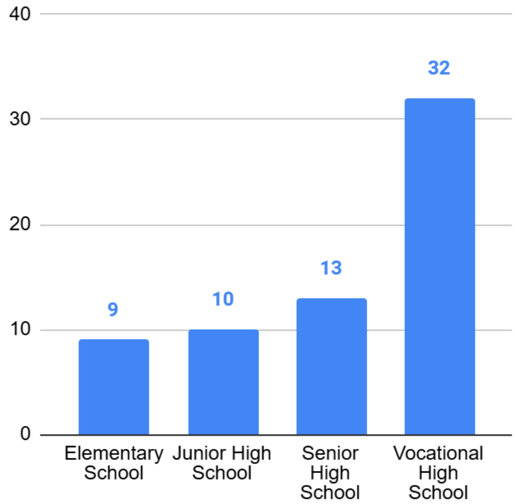

Figure 4. Number of nearest-distance searchings from village offices to public schools that could not be measured using Google Maps by school level

3.1 Among the 10 measured candidates, how many have a route available in Google Maps?

In this research, efficiency was achieved by not measuring the route distance from each origin to all destination points. Instead, the process generated 10 nearest candidates based on straight-line distance, which were then measured using the Google Maps Distance Matrix. However, not all of the 10 candidates produced measurable Google Maps route distances. As shown in Table 1, only 47.12% of the searches in urban villages resulted in route distances for all candidates, while in rural villages the proportion was only 21.32%.

Table 1. Total and percentage of searches producing Google Maps route distances by total measurable candidates and urban-rural classification

Total of measurable distance	Urban	Rural
0	1 (0.96)	63 (15.44)
1	5 (4.81)	49 (12.01)
2	0 (0.00)	22 (5.39)
3	2 (1.92)	17 (4.17)
4	2 (1.92)	11 (2.70)

5	2 (1.92)	13 (3.19)
6	2 (1.92)	19 (4.66)
7	26 (25.00)	69 (16.91)
8	2 (1.92)	28 (6.86)
9	13 (12.50)	30 (7.35)
10	49 (47.12)	87 (21.32)

When examined further in Figure 3, in urban villages more than 80% of the searches produced Google Maps route distances for at least 7 candidates, whereas in rural villages the searches producing Google Maps route distances for at least 7 candidates did not exceed 60%. Moreover, in rural villages, a quarter of the searches resulted in only 0 or 1 measurable candidate.

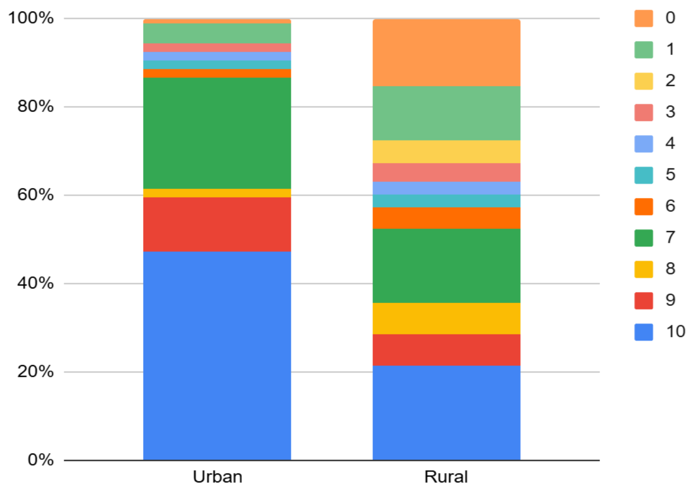

Figure 5. Percentage of searchings that produced Google Maps route distances by total measurable candidates and urban–rural classification

A non-parametric Mann–Whitney U test was applied to compare the distributions of measurable distances obtained from Google Maps between urban and rural villages. The test was chosen because it is robust to non-normal data and suitable for groups with unequal sample sizes. The urban group consisted of 104 measurable distance observations, while the rural group included 408 observations, representing the number of distance searches successfully returned by Google Maps in each area.

The test result yielded a p-value of 2.09×10^{-11}, indicating a statistically significant difference between the two groups. The median measurable distance in urban villages (9.0) was higher than in rural villages (7.0), suggesting that urban areas generally have a greater proportion of successful Google Maps distance measurements. This finding indicates that public facilities are more available in urban villages, supported by better road network connectivity compared to rural areas, where limited infrastructure often prevents measurable route detection.

3.2 Nearest distance using Google Maps based on the order of proximity by straight-line distance radius

In this research, the nearest distance searching was conducted by evaluating 10 candidates (the first candidate being the closest by straight-line distance, the second candidate being the second closest by straight-line distance, and so forth). Each candidate's distance was then calculated using Google Maps, and the shortest route distance was selected as the nearest distance. Table 2 shows that the nearest distance was predominantly produced by the first candidate, both in urban and rural villages, at 77.67% and 81.16% respectively. Meanwhile, the farthest candidate by radius that became the nearest distance was the 8th candidate, while the 9th and 10th candidates never resulted as the nearest distance.

Table 2. Total and percentage of nearest distance searching by candidate order and urban–rural classification

Order of Candidates	Urban	Rural
1	80 (77.67)	280 (81,16)
2	18 (17.48)	39 (11,3)
3	2 (1.94)	8 (2,32)
4	2 (1.94)	10 (2,90)
5	1 (0.97)	5 (1,45)
6	0 (0.00)	0 (0.00)
7	0 (0.00)	2 (0,58)
8	0 (0.00)	1 (0,29)
9	0 (0.00)	0 (0.00)
10	0 (0.00)	0 (0.00)

Figure 6 shows the comparison of the second candidate that became the nearest distance by urban–rural classification and whether the first candidate had a measurable Google Maps distance or not. The results indicate that in most cases the first candidate had a measurable route, but the second candidate turned out to be closer than the first. This demonstrates that the nearest distance based on straight-line radius is not always the nearest distance when measured using actual road networks.

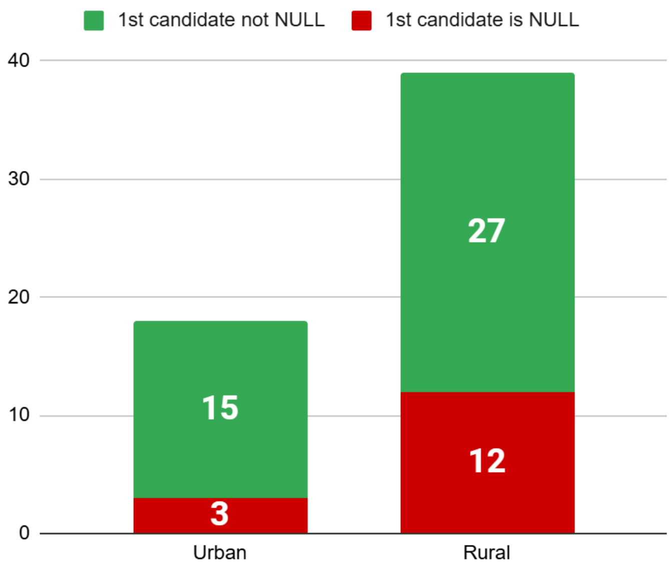

Figure 6. Total searches where the nearest distance was the second candidate by urban–rural classification and the first candidate returned a null value or not

Figure 7 shows the comparison of the third to fifth candidates becoming the nearest distance by urban–rural classification and whether the previous candidates could be measured by Google Maps or not. The results indicate that in urban villages, the same pattern still holds, where the nearest distance mostly occurs when the previous candidates had a route distance but were not shorter. However, in rural villages, a shift begins to appear, showing that the nearest distance is obtained because the previous candidates returned no route distance values.

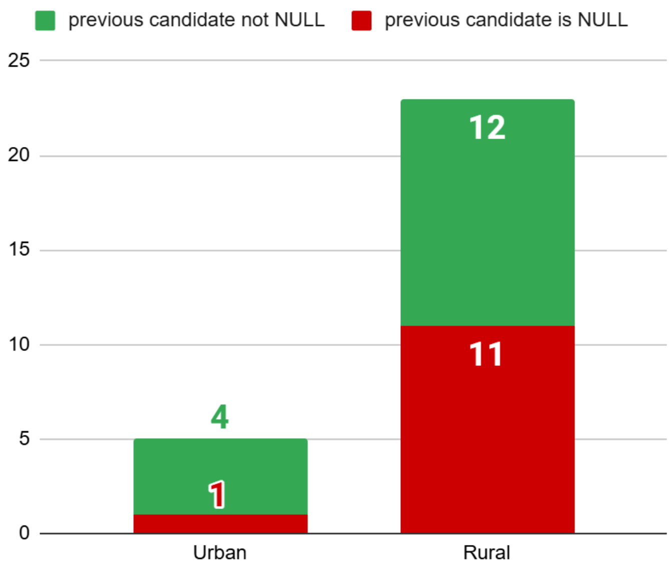

Figure 7. Total searchings where the nearest distance was the third, fourth, or fifth candidate by urban–rural classification and whether the previous candidates returned null or not

Figure 8 presents a comparison of the seventh and eighth candidates that became the nearest distance by urban–rural classification. The results indicate that no urban villages had their nearest distance determined at the seventh or eighth candidate, while in rural villages there were still three searchings. Meanwhile, as shown in Table 1, the sixth, ninth, and tenth candidates never became the nearest distance. These findings suggest that the searching process can be made more efficient without extending to all ten candidates; even limiting it to the fifth candidate already provides sufficiently accurate results.

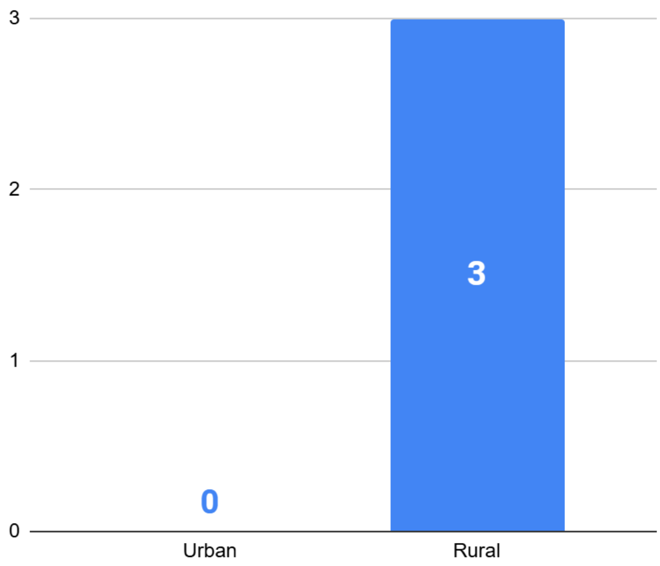

Figure 8. Total searchings in which the nearest distance was the seventh and eighth candidates by urban–rural classification

4. Conclusion

This research was successfully conducted, producing information on the nearest distance between village offices and public schools using route distance calculations based on the Google Maps Distance Matrix API. Efficiency in the searching process was achieved through the application of the Ball Tree algorithm, which enabled an initial filtering of candidate schools within a certain radius before route distance calculations using the road network were performed. This approach minimized the use of the API without reducing the accuracy of the results.

In addition, the research also presented comprehensive information on the ten nearest candidate schools based on straight-line distance for each village office. This information not only provides a more complete spatial overview but can also be used for manual validation or as a basis for decision-making. The results demonstrate that the combination of Ball Tree and Google Maps API can be effectively applied to support accessibility analysis of public facilities, particularly educational facilities at the village level.

The findings show that the nearest route distance could be found as far as the eighth candidate. This indicates that the number of candidates can actually be reduced, with even five candidates being sufficient to produce reliable results. Therefore, processing efficiency can be further improved in future applications.

The descriptive analysis also revealed that the nearest straight-line distance often coincides with the nearest route distance. Nevertheless, this is not the case in all situations, highlighting the importance of route distance calculations using the Google Maps API to obtain more accurate results. Another finding indicates that searchings in urban areas performed better due to the availability of more road networks recognized by Google Maps, enabling a more comprehensive comparison to identify the nearest location.

In contrast, in rural areas, particularly regions not fully covered by Google Maps due to limited road networks or geographical barriers such as rivers, the nearest route distance results require further examination. Nevertheless, this research still provides value through recommendations based on complete information from the ten nearest straight-line distance candidates. This information can assist users in selecting the target facility while also serving as a reference for further manual calculations.

In addition, this research offers a methodological innovation in the context of data collection for accessibility analysis — offering an alternative to the conventional survey and interview-based approach used in the Village Potential Survey (PODES). By leveraging spatial computation and real road network data, the proposed framework provides a more objective and reproducible method for measuring accessibility to public facilities. Furthermore, this approach can be adopted by BPS-Statistics Indonesia in future PODES implementations to enhance the accuracy, efficiency, and objectivity of accessibility data collection in line with actual road conditions.

References

[1] Opris, A.; Necsulescu (Popa), "*A. Public Facilities Assessment Methodology*". Postmod. Open. 2020, 11, 64–78.

[2] Badan Pusat Statistik, "*Buku 2: Pedoman Pencacahan Pemutakhiran Data Perkembangan Desa 2025 (PODES 2025)*". Jakarta. 2025

[3] Badan Pusat Statistik, "*Buku 3: Pedoman Konsep dan Definisi Pemutakhiran Data Perkembangan Desa 2025 (PODES 2025)*". Jakarta. 2025

[4] Suhaedi, D., Harahap, E. "*Traffic Data Mining With Distance Matrix Api by Google Maps*". In: AIP Conference Proceedings, vol. 3065, 2024.

[5] Alsobky, A., Mousa, R., 2020. "*Estimating Free Flow Speed Using Google Maps API: Accuracy, Limitations, and Applications*." Adv. Transport. Stud. 50, 49–64

[6] S. M. Omohundro. "*Five Balltree Construction Algorithms*". Berkeley: International Computer Science Institute. 1989

[7] M. Dolatshah, "*Ball*-tree: Efficient spatial indexing for constrained nearest-neighbor search in metric spaces*," , arXiv:1511.00628, Nov 2015.

[8] Sivakumar I, Manimekalai K, Ranjithkumar A, "*Accessing Public Health Facilities: Rural and Urban Disparities*". J Crit Rev 7(3):382–388. 2020

[9] Badan Pusat Statistik, "*Peraturan Kepala Badan Pusat Statistik Nomor 120 Tahun 2020 Tentang Klasifikasi Desa Perkotaan Dan Perdesaan Di Indonesia Tahun 2020*". Jakarta. 2020