Vol.1 No.1 October 2024: 19-24
International Journal of Computer Science and Humanitarian AI (IJCSHAI)
e-ISSN: 3064-4372
DOI: 10.21512./ijcshaijournal.v1i1.12149

Implementation of Spatial Constraints in Clustering Algorithms: A Study on Basic Infant Immunization in Lamongan District During the COVID-19 Pandemic

Mohammad Dian Purnama [1*], Moh Jainor Miko Wahyudi [2]

[1]Department of Mathematics, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Surabaya,
Surabaya, Indonesia 6023

[2]Departement of Management, Faculty of Economics and Business,
University of Madura,
Pamekasan, Indonesia 693271

mohammaddian.20053@mhs.unesa.ac; jainurmiko089@gmail.com

***Abstract*— Algorithms for clustering data are important for data analysis, especially when finding patterns in big datasets. Nevertheless, the spatial limitations that are important in real-world contexts are generally ignored by the classic clustering approaches. Spatial variables have become more significant in the health industry, particularly during the COVID-19 pandemic, in terms of assessing population requirements and the allocation of healthcare resources. The purpose of this work is to investigate the use of spatial restrictions in clustering algorithms and to apply this method to COVID-19 immunization data from Lamongan District. The data analysis includes the 27 subdistricts of Lamongan District for the year 2021. Based on the peak of the COVID-19 pandemic, which had a major effect on baby immunization coverage, 2021 was chosen. The four basic baby immunization coverages—DPT-HB-Hib3, Polio 4, Measles, and BCG—are the variables that are used. Two methods are used: a neighborhood-like hierarchical clustering algorithm and spatial limitations. Distance-based spatial weights are better than proximity-based spatial weights when it comes to spatial constraints. This method is employed because an infant's coverage of all essential vaccines may be impacted by the spatial structure. We discovered that the fundamental baby immunization variable formed five clusters. It was discovered that cluster five had the highest immunization coverage among all the clusters. The three sub-districts that make up this cluster are Mantup, Kembangbahu, Tikung, Sarirejo, Deket, Glagah, and Karangbinangun.**

Keywords—Spatial Constraint, Spatial Clustering, Clustering, Basic Infant Immunization

Received: Aug. 21, 2024; received in revised form: Oct. 10,2024; accepted: Oct. 17,2024; available online: Oct. 17,2024.

**Corresponding: mohammaddian.20053@mhs.unesa.ac*

I. INTRODUCTION

Clustering algorithms play a significant role in data analysis, particularly in identifying patterns within large datasets [1]. However, the traditional clustering methods often overlook the spatial constraints that are crucial in real scenarios [2]. Incorporating spatial constraints into clustering algorithms allows for more accurate results in geographical contexts, making the data-driven decision-making process more robust.

In the health sector, especially during the COVID-19 pandemic, spatial factors became increasingly important in understanding the distribution of healthcare services and population needs [3]. One critical area that was impacted by the pandemic is the administration of basic infant immunizations [4]. The pandemic disrupted healthcare systems worldwide, including immunization programs, leading to potential public health risks such as outbreaks of vaccine-preventable diseases.

Lamongan district is one of the districts in the province of East Java, Indonesia. It is located in the north of Java Island, 46.3 km from Surabaya, the capital of East Java province. With 27 sub-districts, Lamongan district has the most sub-districts in East Java, second only to Bojonegoro district [5]. Understanding the spatial distribution of immunization coverage during pandemic periods is critical for public health planning and resource allocation.

With disruptions to health services and changes in population movement patterns due to COVID-19, understanding and optimizing the spatial distribution of immunization efforts is critical [6]. This study aims to evaluate how spatial constraints can improve clustering algorithms to provide actionable insights for public health officials and policy makers, ultimately contributing to more effective vaccination strategies and better health outcomes for infants in Lamongan district.

By applying spatially constrained clustering algorithms, health authorities can identify areas with lower immunization rates, which can help focus efforts and resources more effectively. Research related to clustering of basic infant immunization coverage has been conducted by Saputra and Chusyairi [7] with several clustering method. Meanwhile, research related to Clustering with spatial constraints has been conducted by Jaya, et.al [8] on the case of diarrhea in Bandung city.

This study aims to explore the application of spatial constraints in clustering algorithms and apply this approach to Lamongan District immunization data during the COVID-19 pandemic. The goal is to improve understanding of how spatial distribution affects immunization coverage and provide insights that can improve future public health strategies in the face of similar challenges.

II. PROPOSED METHODS

This study examines basic infant immunization coverage in Lamongan District during the COVID-19 pandemic by using spatial constraints in clustering techniques. Through the incorporation of geographical links into the clustering process, the research aims to detect spatial patterns and trends in the distribution of vaccinations. The first step in the technique is data gathering, where information is obtained from different regions on baby immunization rates and pertinent determinants. Descriptive statistical analysis is then carried out to give a summary of the vaccination coverage by region. The next stage is to capture the spatial links between regions by constructing the spatial weight matrix based on either contiguity or distance. This matrix takes into account the influence of surrounding areas and provides a basis for spatial modelling. The spatial dependency is then incorporated into the clustering method by estimating the mixed parameter (α). During clustering, this parameter aids in balancing the impact of spatial proximity with other factors. Lastly, the geographical limitations are taken into account while applying Ward's agglomerative clustering. Taking into account the spatial proximity of the regions, this hierarchical clustering algorithm classifies regions based on similarities in immunization coverage. The steps of the experiment, including the incorporation of spatial limitations, are depicted in Figure 1.

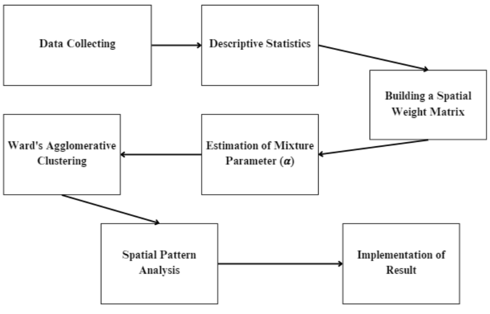

Figure 1. Steps of experiment

A. Data Acquisition

Secondary data from the Lamongan District Health Office release page were used in this study. The 27 subdistricts in Lamongan District for the year 2021 are included in the data analysis [9]. The selection of 2021 is predicated on the peak of the COVID-19 pandemic, which had a significant impact on infant immunization coverage. The data for variables from Lamongan health profile was provided by the Lamongan District Health Office and represents the most relevant information available for that period. The variables used are 4 coverage of basic infant immunization which includes DPT-HB-Hib3, Polio 4, Measles, and BCG. The selection of these variables is based on basic immunization in infants which involves the administration of the four types of vaccines [10].

B. Descriptive Statistic

Descriptive statistical analysis, which entails interpreting the collected data, is one method of data examination. With the use of variables like mean, standard deviation, lowest and greatest values, and so on, the goal is to provide a comprehensive overview of the data. According to Anugrahayu and Azmi [11], descriptive statistics aid in the transformation of data into more easily understood information and provide an understanding of between the study's variables.

C. Spatial Clustering

There are various methods to classify areas into specific clusters especially on immunization coverage which is not only determined by traditional mapping techniques. Clustering methods are needed to identify which areas have low immunization coverage, especially considering spatial dependency due to the nature of infectious diseases such as COVID-19. To incorporate this geographical dependency, two inequality matrices were used.

First, D_0 is the Euclidean distance matrix based on four basic infant immunization coverage includes DPT-HB-Hib3, Polio 4, Measles, and BCG, across 27 subdistricts in Lamongan. Second, D_1 is a dissimilarity matrix that accounts for the geographic proximity between these districts, ensuring that spatial relationships are considered in the clustering process [12].

D. Ward-linkage method

Pseudo Inertia, assume that K clusters, $P_K = (C_1, \dots, C_K)$, may be formed by partitioning the regions $A = \{A_1, \dots, A_n\}$. In the case of dissimilarity data, the pseudo inertia of a C_k cluster can be generalized as follows [12]:

$$I(C_k) = \sum_{i \in C_k} \sum_{j \in C_k} \frac{w_i w_j}{2\mu_k} d_{ij}^2$$

where the weight of C_k is given by $\mu_k = \Sigma_{i \in C_k} w_i$. When the pseudo-inertia $I(C_k)$ is minimal, it means that the observations are more homogenous within cluster-k [12]. Here is the partition P_K's pseudo within-cluster inertia.

$$W(P_K) = \sum_{k=1}^{K} I(C_K)$$

For K clusters, the variance between clusters is described by the total pseudo inertia $I(C_k) = W(P_K)$. Variation between the clusters is more homogeneous when $W(P_K)$ is minimal.

E. Spirit of Ward Hierarchical Clustering

The objective is to combine the two clusters, A and B, of P_{K+1} so that the new partition has the least amount of within-cluster inertia possible. This yields a new partition, P_K in K clusters from a given partition, P_{K+1} in $K+1$ cluster.

$$\arg \min_{A,B \in P_{k+1}} W(P_K)$$

where $W(P_K) = W(P_{K+1}) - I(A) - I(B) + I(A \cup B)$. In this case, two distinct dissimilarity matrices are used: one is based on the spatial weight matrix (i.e., distance and contiguity), while the other is based on various basic infant immunization coverage. The size of these two distinct matrices must be taken into account during the partitioning process [12]. Here, we employ the cluster's mixed pseudo-inertia.Descriptive statistical analysis,

$$I_\alpha(C_k^\alpha) = (1-\alpha) \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} \frac{w_i w_j}{2\mu_k^\alpha} d_{0,ij}^2 + \alpha \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} \frac{w_i w_j}{2\mu_k^\alpha} d_{0,ij}^2$$

where the weight of $C_k^\alpha$is represented by $\mu_k^\alpha = \Sigma_{i \epsilon C_k^\alpha} w_i$, and the normalized dissimilarity between observations i and j in D_0 and D_1, respectively, is represented by $d_{0,ij}^2$ and $d_{1,ij}^2$.

III. Experimental Result

Based on the results of the classical assumption test carried out, the following results were obtained.

A. Descriptive Statistics

There were four variables in the Basic Infant Immunization data used in this research. Table 1 provides a detailed summary for every variable.

TABLE I. TABLE TYPE STYLES

Variable	*Mean*	*Standar Deviation*	*Min*	*Max*
DPT-HB-Hib3	78.71	16.58	40.1	114.2
Polio 4	85.41	17.95	35.6	131.5
Measles	97.54	14.66	56.6	130.5
BCG	104.04	17.51	48.2	146.6

B. Spatial Clustering Without Spatial Constraint

Firstly, we examine geographical clustering of basic infant immunization diseases in the absence of spatial limitations. The dendrogram is used to determine the proper number of clusters K. Figure 2's dendrogram recommends keeping K=5 clusters.

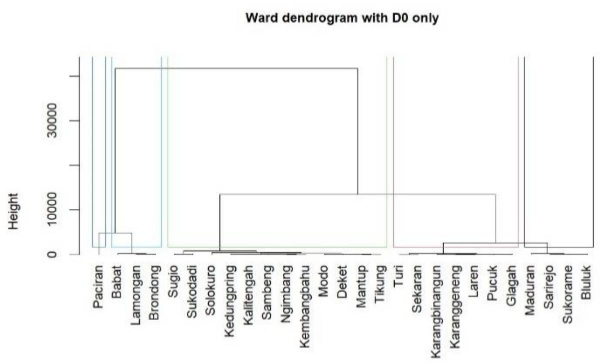

Figure 2. Spatial Clustering Dendogram without Constraint

The appropriate partition of five groups is shown using the choropleth map in Fig. 3.

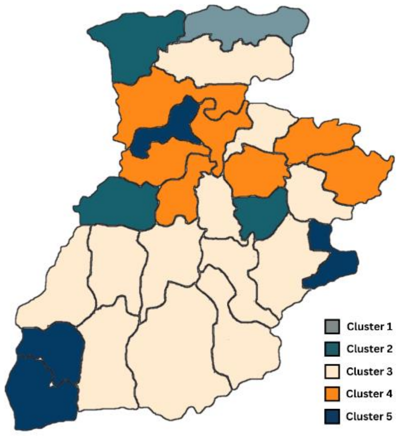

Figure 3. Spatial Clustering Map without Constraint

The subdistrict clustering of Lamongan's basic infant immunization coverage is depicted in Figure 2. But there is no indication of regional clustering on the map. Perhaps the dissimilarity matrix isn't taking into account how close 27 subdistricts are to one another. Cluster 5 is situated in Lamongan's east and southwest.. Moreover, Cluster 2 is dispersed. Cluster 4 is separating it. It makes it challenging to identify the subdistricts that make up the high-risk cluster.

C. Spatial Clustering With Spatial Constraint

The spatial constraint is incorporated into the clustering method in order to resolve the preceding difficulty. We maintain cluster size $K = 5$, which we derive from the dendrogram in Figure 4.

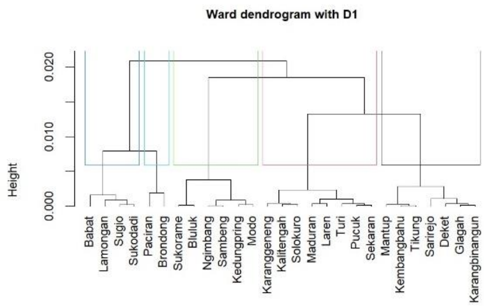

Figure 4. Spatial Clustering Dendogram with Constraint

By incorporating the geographical distance matrix D1, the spatial cluster in Figure 3 can be enhanced to produce more compact clusters. The dendrogram with spatial constraints is displayed on Figure 4. Five built clusters are what we observed. To enhance the geographical cohesion of the five clusters without negatively impacting coverage of basic infant immunization cohesion, a mixing value of α is necessary. The relevance of D0 and D1 in the clustering process is defined by the mixing parameter $\alpha \in [0, 1]$. When $\alpha = 0$ is defined, geographical dissimilarities are not considered. When $\alpha = 1$, on the other hand, the distance coverage of basic infant immunization are not included, and the clusters only include geographical distances. By displaying the quality criterion Q0 and Q1 plot of the partitions P_K^α in Figure 5 and Figure 6, we demonstrate the trade-off between geographical and coverage of basic infant immunization variable distances.

Cite this article as: *M. D. Purnama, and M. J. M. Wahyudi, "Implementation of Spatial Constraints in Clustering Algorithms: A Study on Basic Infant Immunization in Lamongan District During the COVID-19 Pandemic ", International Journal of Computer Science and Humanitarian Artificial Intelligence (IJCSHAI), vol. 1, no. 1, pp. 19-24, Oct. 2025. DOI: https://doi.org/10.21512/ijcshai.v1i1.12149*

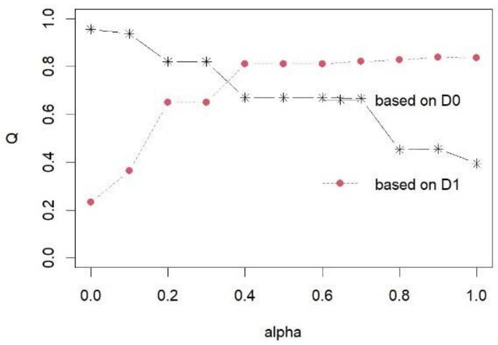

Figure 5. Proportion of pseudo-inertias that are explained, $Q_0(P_K^\alpha)$ versus α (shown as a black solid line) and $Q_1(P_K^\alpha)$versus α (shown as a dashed line).

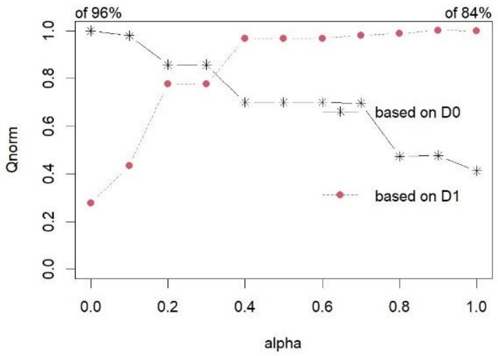

Figure 6. Proportion of pseudo-inertias that are explained, $Q_0^*(P_K^\alpha)$ versus α (black solid line) and $Q_1^*(P_K^\alpha)$ versus α (dashed line)

TABLE II. PROPORTION OF EXPLAINED PSEUDO-INERTIA

Alpha	Q(Q0)	Q(Q1)	QNorm(Q0)	QNorm(Q1)
0.0	0.96	0.23	1.00	0.28
0.1	0.94	0.36	0.98	0.43
0.2	0.82	0.65	0.86	0.78
0.3	0.82	0.65	0.86	0.78
0.4	0.67	0.81	0.70	0.97
0.5	0.67	0.81	0.70	0.97
0.6	0.67	0.81	0.70	0.97

Based on the table 2, we found that 0.40 is the optimal alpha value, with the highest Q0 and high Q1 values. It suggests that the construction of the clusters is significantly influenced by the geographic and basic infant immunization characteristics. Put otherwise, the distances associated with the geographical and coverage of basic infant immunization variables exhibit a good quality level, with values exceeding 0.70. The spatial clustering maps for the provided distance spatial weight matrix are shown in Figure 7.

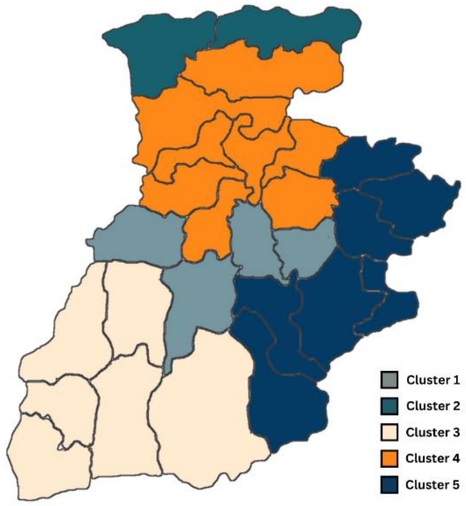

Figure 7. Spatial Clustering Map with Constraint into K = 5 clusters according to the "neighborhood" distances of the districts D1 with α = 0.4 and basic infant immunization coverage D0.

Given that the clustering process takes place in part because of the geographical distance, Figure 8 and Figure 9 demonstrates that clusters with greater geographic distance are more geographically compact. Based on spatial contiguity, we also take geographical dissimilarity into account. Figure 8 and Figure 9 presents the trade-off geographical dissimilarity based on variables related to coverage of basic infant immunization and queen contiguity.

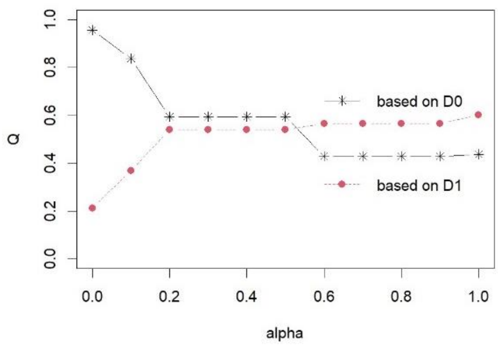

Figure 8. Proportion of pseudo-inertias that are explained, $Q_0(P_K^\alpha)$ versus α (shown as a black solid line) and $Q_1(P_K^\alpha)$versus α (shown as a dashed line).

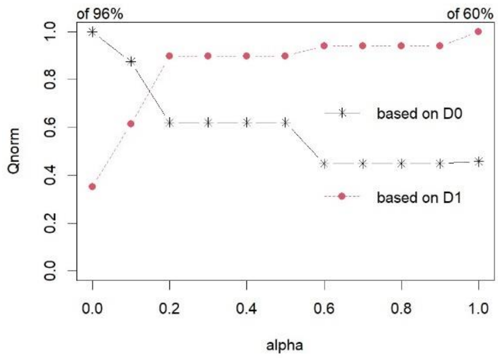

Figure 9. Proportion of pseudo-inertias that are explained, $Q_0^*(P_K^\alpha)$ versus α (black solid line) and $Q_1^*(P_K^\alpha)$ versus α (dashed line)

TABLE III. PROPORTION OF EXPLAINED PSEUDO-INERTIA GEOGRAPHICAL CONTIGUITY

Alpha	Q(Q0)	Q(Q1)	QNorm(Q0)	QNorm(Q1)
0.0	0.96	0.21	1.00	0.35
0.1	0.84	0.37	0.88	0.61
0.2	0.60	0.54	0.62	0.90
0.3	0.60	0.54	0.62	0.90
0.4	0.60	0.54	0.62	0.90
0.5	0.60	0.54	0.62	0,90

With the trade-off between regional dissimilarity and coverage of basic infant immunization factors taken into account, 0.20 is the optimal value for alpha. Based on normalization, Q0 and Q1 have quality levels of 0.93 and 0.61, respectively. Based on criteria Q 0 and Q 1, the quality of the partition spatial clusters of basic infant immunization coverage with "neighborhood" contiguity is marginally lower than that of the other partition spatial clusters of basic baby immunization coverage with "neighborhood" contiguity (62% compared to 70%, 90% compared to 97%).

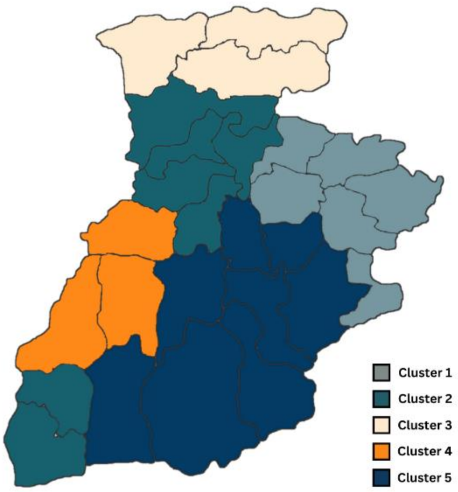

Figure 10. Spatial Clustering Map with Constraint into K = 5 clusters using the "neighborhood" continuity D1 and the basic infant immunization coverage variables D0 with $\alpha = 0.2$.

Spatial clusters of geographically adjacent basic infant immunization coverage are displayed in Figure 7. In terms of appearance, it differs from a spatial cluster determined by distance. Geographic proximity is marginally worse than geographic distance, though. It suggests that in the event of basic infant immunization coverage, geographic distance is more suitable. This could be the result of the basic infant immunization coverage spreading not just in nearby areas but potentially even far away.

TABLE IV. AVERAGE VALUES OF VARIABLES FOR EACH CLUSTER

Cluster	DPT-HB-Hib3	Polio 4	Measles	BCG
1	85.40	92.47	100.53	106.75
2	82.34	90.16	101.83	114.46
3	56.27	63.83	73.87	78.77
4	88.95	97.62	109.92	118.25
5	75.09	79.47	95.79	97.04

Cluster 1 ranks fourth with good immunization coverage, though not as high as some other clusters. The coverage for DPT-HB-Hib3 is 78.93, while Polio 4 reaches 80.83. Measles coverage stands at 97.60, close to the standard target. BCG shows the highest coverage in this cluster with a value of 104.58. Overall, the performance is quite good, but there is room for improvement in DPT-HB-Hib3 and Polio 4. The subdistrict includes Babat, Lamongan, Sugio, and Sukodadi.

Cluster 2 is ranked fifth with the lowest immunization coverage. The coverage for DPT-HB-Hib3 is only 64.35, the lowest among all clusters. Polio 4 reaches 77.95, while Measles coverage is 82.50. BCG coverage in this cluster is 94.05. The low coverage across all immunizations indicates that this cluster needs special attention to improve access and implementation of immunization. The subdistrict includes Paciran and Brondong.

Cluster 3 ranks second with relatively high coverage across various types of immunizations. DPT-HB-Hib3 is at 79.21, while Polio 4 is at 85.33. Measles coverage is 97.78, and BCG reaches 100.61. This cluster shows consistently good performance in all categories, though there is some room for slight improvement in DPT-HB-Hib3. The subdistrict includes Sukorame, Bluluk, Ngimbang, Sambeng, Kedungpiring, and Modo.

Cluster 4 ranks third with also good coverage. DPT-HB-Hib3 in this cluster is 78.76, while Polio 4 reaches 86.08. Measles coverage is 97.35, and BCG has the highest coverage with a value of 108.05. This cluster is close to Cluster 3 in terms of coverage but is slightly lower in DPT-HB-Hib3 and Measles. The subdistrict includes Karanggeneng, Kalitengah, Solokuro, Maduran, Laren, Turi, Pucuk, and Sekaran.

Cluster 5 ranks first with the highest immunization coverage among all clusters. DPT-HB-Hib3 coverage is 82.20, while Polio 4 is at 89.47. Measles coverage is the highest in this cluster at 101.83, and BCG stands at 104.94. This cluster shows the best results in immunization coverage, particularly in DPT-HB-Hib3, Polio 4, and Measles. The subdistrict includes Mantup, Kembangbahu, Tikung, Sarirejo, Deket, Glagah, and Karangbinangun.

IV. CONCLUSSION

Immunization is the process of intentionally boosting a person's immunity against a disease, increasing the likelihood that they will either never contract the illness or just suffer from its mild symptoms if they are ever exposed to it. A spatial phenomena is also seen in the equitable coverage of basic newborn immunization. Determining the equality of vaccination coverage requires the identification of spatial groupings. Spatial constraints and a neighborhood-like hierarchical clustering technique are used. For spatial limitations, distance-based spatial weights are preferable than proximity-based spatial weights. When clustering the distribution of newborns' basic immunization coverage, spatial restrictions must be taken into account to identify the locations with high coverage. This approach is used because the spatial structure may affect an infant's coverage of all core immunizations. We found that there were five clusters produced by the basic infant immunization variable. Out of all the clusters, cluster five was found to have the highest immunization coverage. This cluster consists of Mantup, Kembangbahu, Tikung, Sarirejo, Deket, Glagah, and Karangbinangun as its three sub-districts.

REFERENCES

[1] M. Allaoui, M. L. Kherfi, and A. Cheriet, "Considerably improving clustering algorithms using UMAP dimensionality reduction technique: a comparative study," in *Int. Conf. Image Signal Process.*, Cham: Springer Int. Publishing, June 2020, pp. 317–325.

[2] H. Usman and M. Lizam, "Exploring some spatially constrained delineation methods in segmenting the Malaysian commercial property market," *Int. J. Strategic Property Manag.*, vol. 27, no. 6, pp. 379–390, 2023.

[3] I. Franch-Pardo, B. M. Napoletano, F. Rosete-Verges, and L. Billa, "Spatial analysis and GIS in the study of COVID-19. A review," *Sci. Total Environ.*, vol. 739, p. 140033, 2020.

[4] S. Basu et al., "Impact of the COVID-19 pandemic on routine vaccine landscape: A global perspective," *Hum. Vaccines Immunotherapeut.*, vol. 19, no. 1, p. 2199656, 2023.

[5] B. M. Rizaldi, F. Setiawan, W. A. Violando, and A. D. Muttaqin, "Analisis Perubahan Stok Karbon Berdasarkan Perubahan Penutupan Lahan dari Tahun 1990–Tahun 2020 di Wilayah Pesisir Kabupaten Lamongan, Provinsi Jawa Timur," in *Prosiding Forum Ilmiah Tahunan (FIT)-Ikatan Surveyor Indonesia (ISI)*, vol. 1, pp. 81–88, 2021.

[6] P. K. Jithesh, "A model based on cellular automata for investigating the impact of lockdown, migration and vaccination on COVID-19 dynamics," *Comput. Methods Programs Biomed.*, vol. 211, p. 106402, 2021.

[7] P. R. N. Saputra and A. Chusyairi, "Perbandingan metode clustering dalam pengelompokan data puskesmas pada cakupan imunisasi dasar lengkap," *J. RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 6, pp. 1077–1084, 2020.

[8] I. G. Jaya, B. N. Ruchjana, Y. Andriyana, and R. Agata, "Clustering with spatial constraints: The case of diarrhea in Bandung city, Indonesia," *J. Phys.: Conf. Ser.*, vol. 1397, 2019.

[9] Dinas Kesehatan Kabupaten Lamongan, "Profil Kesehatan Kabupaten Lamongan 2021," [PDF]. Available: https://lamongankab.go.id/dinkes/post/1872.

[10] M. D. Purnama, C. D. Farrel, S. F. Wahyuniarti, I. P. Nuraini, and A. Oktaviarina, "Analisis Perbedaan Data Komponen Kesehatan Bayi Provinsi Jawa Timur Tahun 2022 Menggunakan MANOVA," *J. Sains dan Seni ITS*, vol. 12, no. 6, 2023.

[11] M. Anugrahayu and U. Azmi, "Stock portfolio optimization using mean-variance and mean absolute deviation model based on K-medoids clustering by dynamic time warping," *J. Matematika, Statistika dan Komputasi*, vol. 20, no. 1, pp. 164–183, 2023.

[12] M. Chavent, V. Kuentz-Simonet, A. Labenne, and J. Saracco, "ClustGeo: an R package for hierarchical clustering with spatial constraints," *Comput. Stat.*, vol. 33, no. 4, pp. 1799–1822, 2018.