J. DIFERENSIAL/E-ISSN:2775-9644
Vol.6(2),2024, 170–178
doi:https://doi.org/10.35508/jd.v6i2.16278

RESEARCH ARTICLE

Implementation of K-Nearest Neighbor Algorithm on Density-Based Spatial Clustering Application with Noise Method on Stunting Clustering

Friansyah Gani[1], Hasan S. Panigoro[2,*], Sri Lestari Mahmud[3], Emli Rahmi[4], Salmun K. Nasib[5], La Ode Nashar[6]

[1,2,3,4,5,6]*Department of Mathematics, Universitas Negeri Gorontalo, Bone Bolango 96554, Indonesia*
[*]*Corresponding author: hspanigoro@ung.ac.id*

Received: 23 May 2024; Revised: 7 August 2024; Accepted: 28 August 2024; Published:1 November 2024.

Abstract:
This paper studies the implementation of the K-Nearest Neighbor (KNN) algorithm on the Density-Based Spatial Clustering Application with Noise (DBSCAN) method on stunting Clustering in the eastern region of Indonesia in 2022. The DBSCAN method is used because it is more efficient in performing the Clustering process for irregular Clustering shapes. The main objective of this study is to apply the KNN algorithm to the DBSCAN Clustering technique in 161 Districts/Cities in 11 provinces in eastern Indonesia. A comparison of the performance evaluation of the DBSCAN Clustering technique is done by considering the value of the Silhouette, BetaCV, and Davies-Bouldin indexes indicating the quality of the Clusters formed with the lowest results scores of 0.67 and 1.84 with epsilon value = 3.4 and minimum point value = 2 resulting in 4 Clusters. The results of Clustering 161 Districts and Cities based on the factors that cause stunting formed 4 Clusters where Cluster 0 consists of 119 Districts and Cities with very high stunting characteristics, Cluster 1 consists of 3 Districts and Cities with high stunting characteristics, the results of Cluster 2 consist of 2 Districts and Cities with low stunting characteristics, then the results of Cluster 2 consist of 2 Districts and Cities with low stunting characteristics and Cluster 3 consists of 2 Cities with very low stunting characteristics.

Keywords: Silhouette, BetaCV, Davies-Bouldin, DBSCAN, KNN, Clustering

1. Introduction

Clustering represents the process of partitioning a dataset into several subsets, known as clusters, such that the intra-cluster similarity is maximized and the inter-cluster similarity is minimized [1]. This technique organizes data objects into clusters where objects within the same cluster exhibit high similarity to each other and significant dissimilarity to objects in other clusters. The partitioning is achieved through the application of clustering algorithms rather than manual grouping. Consequently, clustering is an effective method for identifying previously unknown groupings within a dataset. According to [2] unsupervised learning is the process of grouping unlabeled data into clusters, where the same sample belongs to the same cluster but different samples belong to different groups. Meanwhile, according to [3], unsupervised learning or what is often referred to as clustering techniques optimally partition data based on Euclidean distance to find better data patterns. In unsupervised learning, the data has no pattern at all and the goal of unsupervised learning is to find patterns in the data.

The way to partition a dataset into groups based on predefined criteria of similarity is the advantage of Clustering. A cluster refers to a collection of data objects that exhibit high similarity to each other within the same cluster and significant dissimilarity to objects in other clusters. The objective is to group data objects into clusters such that the intra-cluster similarity is maximized while the inter-cluster dissimilarity is also maximized. Object similarity is typically determined from the attribute values that characterize the data objects, which are often represented as points in a multidimensional feature space [4]. Clustering is frequently employed as an initial step in the data mining process. The resulting clusters serve as input for subsequent stages, such as training artificial neural networks. Given the vast scale of contemporary databases, applying clustering analysis at the outset is highly advantageous for optimizing the data mining process. [5]. K-Nearest Neighbors (KNN) retains all instances from the training dataset and begins constructing the classification model only when test data becomes available for prediction. KNN is considered a relatively simple classification technique. [6]. Other than that, this approach is the simplest of all machine learning approaches [7]. The following term is used for KKN.

$$\text{dist}(p, q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}, \tag{1.1}$$

To find out the distance in KNN of a data, it can be done by finding the farthest distance from the neighbors in a data. By looking for this farthest distance, the KNN distance is obtained which will later be used as a reference to find the value of the density parameter.

$$\text{KNN}(p) = \max \left\{ \text{dist}(p, q) \mid q \in k \text{ Neighbor(p)} \right\}, \tag{1.2}$$

Furthermore, DBSCAN is classified as a density-based clustering algorithm that is based on specific density criteria. A cluster is defined as a region of high density, with clusters being separated by regions of low density, which are regarded as noise. Data points are grouped based on their density, with dense regions being identified as clusters and sparse regions being classified as outliers. [8]. In DBSCAN, the concept of density refers to the quantity of data points within the MinPts radius (the minimum number of points within the Eps radius). Data points are classified based on their density, with core points being those that meet the minimum density requirement within the specified radius. This density-based classification leads to three types of data points: core points, border points, and noise [9]. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a method designed to identify clusters based on the density of data points and to detect noise or outliers within the dataset. In DBSCAN, density is defined as the number of data points (MinPts) within the Eps radius of each data point. This density-based approach results in the classification of data points into three categories: core points, border points, and noise [10]. Some terms in the DBSCAN method are given as follows.

1. Core: the center point in a Cluster is based on the density where there is the number of points that must be within Eps (radius or threshold value). MinPts (minimum points in the Cluster) which are user-defined.
2. Border: the point that becomes the boundary within the core.
3. Noise: a point that cannot be reached by the core and is not a border.

$$Noise = \{x \in X \mid \forall i : x \in C_i\} \tag{1.3}$$

Where X is the data Cluster, and Ci is the 1st Cluster.

4. Direct reachability density: a point is said to be a direct reachable point if it is directly connected in front of the center (core).

$$x \in NEps(y)|N_{eps}(y)| > MinPts \tag{1.4}$$

   where NEps (y) is the point around y within the eps radius and MinPts is the minimum points in the Cluster.
5. Reachable density: a point is said to be a reachable point if it is indirectly connected to the center point. Is indirectly connected to the center point (core).
6. Connected density: a point is said to be connected by other points. by another point. DBSCAN requires two input parameters, namely epsilon and minimum points. eps-around points are defined as:

$$N_{eps}(x) = (y \in D|dist(x, y) < eps) \tag{1.5}$$

   Where $N_{eps}(x)$ is the neighboring points of x in radius eps, D is the data Cluster, dist(x,y) is the Euclidean distance of object x and y and eps is the radius or threshold.

Silhouette index that has a value close to 1 means that the object fits better in the Cluster [11]. SI that has a negative value indicates that the object does not fit in the Cluster. SI that has a value of 0 or close to zero indicates that the object is located at the boundary between two Clusters [12].

$$SI = \frac{1}{k}\sum_{j=1}^{k} SI_j, \tag{1.6}$$

BetaCV is used to measure the goodness of a Cluster based on the relationship between the average distance in a Cluster and the average distance between Clusters. distance between Clusters. The lower BetaCV value means that the better the Clustering effect.

$$BetaCV = \frac{W_{in}/N_{in}}{W_{out}/N_{out}}, \tag{1.7}$$

DBI is a validity index that determines the average value of each data point in the collection. The value of each point is determined by adding the compactness value and dividing the result by the difference between the two Cluster centroids as separation [13]. A low DBI value means that the better the Clustering [14].

$$DBI = \frac{1}{c}\sum_{i=1}^{c} \max_{i \neq j} \left(\frac{d(x_i + d(x_j)}{d(c_i, c_j)} \right), \tag{1.8}$$

Based on the descriptions of the DBSCAN method and the K-Nearest Neighbors (KNN) algorithm, the researcher will apply these techniques to a case study on stunting. Stunting is a condition characterized by chronic malnutrition in toddlers, resulting in inadequate growth relative to age [15]. Undernutrition, particularly among children, represents a significant public health issue and is one of the leading global contributors to preventable child mortality. According to recent global estimates, by 2019, 144 million children under the age of 5 were projected to be stunted, 47 million wasted, and 14.3 million malnourished (DHS) [16]. Stunting remains a persistent nutritional challenge affecting toddlers worldwide.

Stunting is a sign of chronic malnutrition over a long period. Stunting is diagnosed by comparing height-for-age scores from growth charts that have been used worldwide. The World Health Organization growth charts, developed in 2005, are used in Indonesia to identify stunting [17]. According to [18], the main risk factors for stunting are genetic factors, economic status, birth spacing, history of LBW, anemia in the mother, hygiene, environmental sanitation, and environmental hygiene. mother, hygiene, environmental sanitation, and nutritional deficiencies.

Factors influencing the occurrence of stunting, as identified by [19], include the proportion of households with adequate sanitation, the proportion of households with access to safe drinking water, the proportion of infants who are exclusively breastfed for the first six months, the proportion of infants who initiate complementary feeding (MP-ASI) promptly, the proportion of children aged 12-59 months who have received all recommended vaccinations, average household expenditure per capita, and average daily caloric intake per person. Additionally, access to good sanitation, health insurance, family planning services, social assistance, healthy housing conditions, family food security, and the availability of diverse foods for children under five are considered key nutritional indicators [20].

According to the Indonesian Nutrition Status Survey (SSGI) by the Ministry of Health, the prevalence of stunting in Indonesia reached 21.6% in 2022. This represents a reduction of 2.8 percentage points from the previous year. Among the 13 provinces in eastern Indonesia, 11 provinces exhibit stunting prevalence rates above the national average, with North Sulawesi and Bali having lower prevalence rates of 20.5% each. Conversely, the province with the lowest stunting prevalence rate is Bali, at 8%. The provinces with stunting prevalence rates exceeding the national average include NTT, West Sulawesi, Papua, NTB, Central Sulawesi, Southeast Sulawesi, South Sulawesi, North Maluku, Maluku, and Gorontalo. In the next section, we present all mathematical results as well as their interpretation.

2. Results and Discussions

In this section, two parameters must be determined before running the DBSCAN algorithm. The minimum point and epsilon parameters, where the data used are 21 variables from 161 Districts and Cities in the eastern region of Indonesia in 2022.

The validation test of the DBSCAN algorithm was carried out using the Silhouette Index, BetaCV, and Davies-Bouldin Index (DBI). The Silhouette Index which has a value close to 1 means that the object fits better in the Cluster, while the BetaCV value, the lower the Clustering effect and the Davies-Bouldin Index (DBI). value, the lower the Clustering effect will be, and for a low Davies-Bouldin Index value is low can be interpreted that the better the Clusters are produced. To find the optimal number of Clusters, it is necessary to randomize the value of ε and Min Point to get the Silhouette, BetaCV, BetaCV, and Davies-Bouldin indexes. Below are the evaluation results of the DBSCAN algorithm presented in the form of a table.

From the evaluation results using the Silhouette Index, BetaCV, and Davies-Bouldin Index, the best Eps and Min Point values can be seen. Bouldin Index, the best Eps, and Min Point values can be seen at at Eps = 3.4 and Min Point = 2 based on the performance of the BetaCV and Davies-Bouldin Index where from these two parameters 4 Clusters are formed. Therefore, this value will be used to initialize the two parameters in the DBSCAN method.

The Cluster results can be used as mapping material and material for analyzing stunting handling plans based on their factors. The results of the four Clusters in this study show the existence of different characteristics between Clusters. The identification analysis is as follows:

Based on the table above, it is known that Cluster 0 consists of 119 Districts and Cities, Cluster 1 consists of 3 Districts and Cities, Cluster 2 consists of 2 Districts and Cities, Cluster 3 consists of 2 Districts and Cities, and 35 Districts and Cities including Noise. Where Cluster 0 has more members than Clusters 1, 2, and 3. The following is a mapping of Clustering results.

The figure above is a mapping of the results of Clustering 161 Districts and Cities based on the factors that cause stunting, with Green Districts and Cities are Districts and Cities included in Cluster 0, then yellow Districts and Cities are Districts and Cities included in Cluster 1 orange Districts and Cities are Districts and Cities included in Cluster 2, for Districts and Cities that are red are Districts and Cities included in Cluster 3 and Districts and Cities that are pink are Districts and Cities included

Table 2.1: Values of Cluster Validity Index Silhouette Index, BetaCV and Davies-Bouldin Index in DBSCAN Method

Eps	*Min Points*	*SI*	*BetaCV*	*DBI*
3	2	-0,0112	0,8114	2,2580
3	3	-0,0047	0,8339	3,1849
3	4	0,1188	0,8485	3,4177
3	5	0,0858	0,9119	3,6607
3	6	0,0693	0,9446	3,7574
3	7	0,0533	0,9782	3,9080
3,2	2	0,0082	0,7476	2,2225
3,2	3	0,0146	0,7576	2,5997
3,2	4	0,1612	0,7816	3,1754
3,2	5	0,1572	0,7893	3,1112
3,2	6	0,1406	0,8154	3,2067
3,2	7	0,1307	0,8330	3,2294
3,4	2	0,0724	**0,6776**	**1,8428**
3,4	3	0,1673	0,6893	2,4602
3,4	4	**0,2224**	0,6990	3,2170
3,4	5	0,2166	0,7057	3,2077
3,4	6	0,1995	0,7281	3,0999
3,4	7	0,1853	0,7467	3,1417

Table 2.2: Clustering Results Based on Cluster 0, 1, 2 and 3

Cluster	**Cities**
Cluster 0	Sumba Barat, Lembata, Morowali, Timor Tengah Selatan, Belu, Alor, Flores Timur, Sikka, Ende, Ngada, Manggarai Barat, Sumba Tengah, Sumba Barat Daya, Nagekeo, Manggarai Timur, Sabu Raijua, Malaka, Kota Kupang, Majene, Polewali Mandar, Mamuju, Pasangkayu, Mamuju Tengah, Merauke, Nabire, Biak Numfor, Mimika, Boven Digoel, Sarmi, Keerom, Waropen Supiori, Lombok Barat, Lombok Tengah, Lombok Timur, Sumbawa, Dompu, Bima, Sumbawa Barat, Lombok Utara, Kota Mataram, Kota Bima, Fakfak, Kaimana, Teluk Bintuni, Manokwari, Sorong Selatan, Sorong, Raja Ampat, Tambrauw, Maybrat, Manokwari Selatan, Banggai Kepulauan, Banggai, Poso, Donggala, Toli-Toli, Buol, Parigi Moutong, Tojo Una-Una, Sigi, Morowali Utara, Palu, Buton, Muna, Konawe, Kolaka, Konawe Selatan, Bombana, Wakatobi, Buton Utara, Konawe Utara, Kolaka Timur, Muna Barat, Buton Tengah, Buton Selatan, Kendari, Bau Bau, Kepulauan Selayar, Bulukumba, Bantaeng, Jeneponto, Takalar, Sinjai, Maros, Pangkajene dan Kepulauan, Barru, Bone, Soppeng, Wajo, Sidenreng Rappang, Pinrang, Enrekang, Luwu, Tana Toraja, Luwu Utara, Luwu Timur, Kota Parepare, Kota Palopo, Halmahera Barat, Halmahera Tengah, Kepulauan Sula, Halmahera Utara, Halmahera Timur, Pulau Morotai, Pulau Taliabu, Tidore Kepulauan, Kepulauan Tanimbar, Buru, Kepulauan Aru, Seram Bagian Barat, Seram Bagian Timur, Maluku Barat Daya, Boalemo, Gorontalo, Pohuwato, Bone Bolango, Gorontalo Utara, Kota Gorontalo
Cluster 1	Lembata, Kolaka Utara, Halmahera Selatan
Cluster 2	Morowali dan Tual
Cluster 3	Kota Makassar dan Ternate
Noise	Timor Tengah Utara, Manggarai, Rote Ndao, Mamasa, Jayawijaya, Jayapura, Kepulauan Yapen, Paniai, Puncak Jaya, Mappi, Asmat, Yahukimo, Pegunungan Bintang, Tolikara, Mamberamo Raya, Yalimo, Puncak, Dogiyai, Intan Jaya, Deiyai, Kota Jayapura, Teluk Wondama,Pegunungan Arfak, Kota Sorong, Banggai Laut, Konawe Kepulauan, Gowa, Toraja Utara, Kepulauan Kei, Maluku Tengah, Buru Selatan, Ambon

in noise/outliers. Noise detected on the map consists of 35 regencies and Cities, this is because the noise data is not successfully captured by eps which is determined based on the results of the comparison of the three indices.

Grouping 161 Districts and Cities in 11 provinces with DBSCAN based on factors affecting stunting and interpreting the characteristics of each Cluster formed can be seen in the average value of the variables in each Cluster as follows:

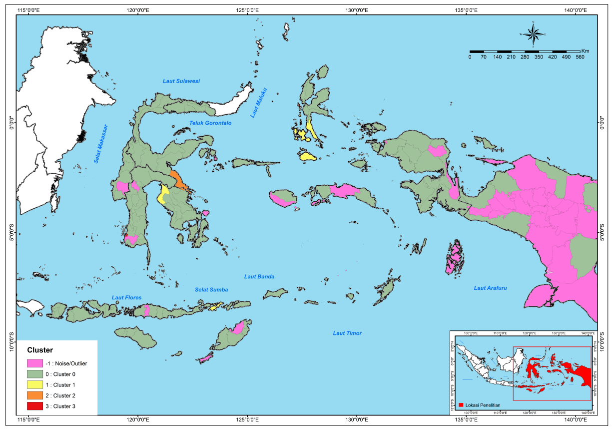

Figure 2.1: Map of Eastern Indonesia Stunting Clustering Results in 2022

Table 2.3: Average Value of Variables in Cluster

Variabel	Klaster			
	0	1	2	3
X_1	0.2014	0.5039	-0.7080	0.2625
X_2	0.0800	0.2372	0.2254	1.0511
X_3	0.0163	0.0856	3.0041	0.8215
X_4	0.1034	0.5018	0.8105	1.3763
X_5	-0.1564	0.2124	-0.9211	-1.0356
X_6	0.2583	0.4047	0.1147	0.6826
X_7	0.2893	-0.4473	-0.1236	0.2291
X_8	0.1451	0.3632	-0.2328	-0.2587
X_9	0.0095	-0.3106	0.3976	0.2043
X_{10}	-0.0670	-0.1598	0.7207	2.7812
X_{11}	-0.0380	-0.0730	-0.5030	2.5768
X_{12}	0.1891	-0.3009	-0.3711	0.0619
X_{13}	0.0196	-0.6115	-0.6841	2.5059
X_{14}	-0.0355	-0.6063	0.3176	-1.4191
X_{15}	-0.0602	4.9715	-0.2859	-0.5509
X_{16}	0.2632	0.1640	-0.6551	1.2068
X_{17}	-0.0411	0.0266	0.5869	-0.6864
X_{18}	0.3149	0.3942	0.1642	0.4524
X_{19}	0.1044	-0.3733	1.6108	1.3199
X_{20}	0.1925	0.3091	0.3642	0.7817
X_{21}	0.2484	-0.8265	-1.0655	-0.1806

Based on the average value of the variables in Table 4.9 Cluster zero consists of areas with the percentage of health insurance (X7), the percentage of midwives (X12), and the percentage of maternal education level (X21) with (X7), the percentage of midwives (X12) and the percentage of maternal education level (X21) is the highest compared to Cluster 1, Cluster 2 and Cluster 3. high values compared to Cluster 1, Cluster 2 and Cluster 3. A high value means that regions in Cluster 0 have sufficient midwives, health insurance and midwives, health insurance, and maternal education levels are better than Cluster 1, Cluster 2, and Cluster 3. Cluster 1, Cluster 2, and Cluster 3. The first Cluster consists of regions with a higher percentage of children under five years of age exclusively breastfed (X1), the percentage of households with decent drinking water (X5), and the percentage of households that (X1), percentage of households with safe drinking water (X5), percentage of under-five children fed with complementary foods (X8), and percentage of clinics (X8). (X8) and the percentage of clinics (X15) are the highest compared to Cluster 0, Cluster 2 and Cluster 3. Cluster 0, Cluster 2, and Cluster 3. The high value means that areas in Cluster 1 are exclusively breastfed, have access to proper drinking water decent drinking water, the clown is given complementary feeding and has a good clinic compared to Cluster 0, Cluster 2, and Cluster 3. better than Cluster 0, Cluster 2, and Cluster 3. The second Cluster consists of areas with a high percentage of children under five years of age with low birth weight (X3), the percentage of children under five not given MP-ASI (X9), the percentage of adequate Community Health Centers (X14), the percentage of women using injections (X17), and the percentage of women using injections (X17). using injections (X17) and the percentage of households using bottled water (X19) that are not using bottled water (X19) is the highest compared to Cluster 0, Cluster 1, and Cluster 2. Cluster 0, Cluster 1, and Cluster 3. The high value can be interpreted that regions in Cluster 2 have toddlers with LBW, under-five children who are not given MP-ASI, have a Community Health Center, women who use injections, and households using bottled water are more likely to have households that use bottled water are better than Cluster 0, Cluster 1 and Cluster 3. Cluster 3 consists of regions with a higher percentage of children under five years of age who received early breastfeeding initiation (X2), the percentage of children under five years of age who received complete basic immunization (X4), the percentage of households that have proper sanitation (X6), percentage of government hospital adequacy (X10), percentage of private hospital adequacy (X11), percentage of private hospital adequacy (X12) (X10), the percentage of private hospitals (X11), the percentage of doctors (X13), the percentage of children born alive adequacy of doctors (X13), percentage of children born alive (X16), percentage of households using gooseneck latrines (X18) and the percentage of households with a clean drinking water source (X18). who have a clean drinking water source (X20) is the highest compared to Cluster with Cluster 0, Cluster 1, and Cluster 2. The high value can be interpreted as that regions in Cluster 3 have the highest percentage of children under five years of age who received early breastfeeding initiation, complete basic immunization, households with proper sanitation, adequate households have proper sanitation, adequate government hospitals, adequate private hospitals, adequacy of doctors, children born alive, households using gooseneck latrines that use gooseneck latrines and have a better source of clean drinking water than Cluster 0. source of clean drinking water that is better than Cluster 0, Cluster 1 and Cluster 2.

3. Conclusion

In the 2022 clustering analysis of districts and cities in Eastern Indonesia, based on factors contributing to stunting, the DBSCAN algorithm was applied with parameters $\varepsilon = 3.4$ and MinPts = 2. This approach effectively identified four clusters and also detected noise and outliers. The validity of the clustering was confirmed by the BetaCV index, which achieved the lowest value of 0.677608963, and the Davies-Bouldin Index (DBI), which recorded the lowest value of 1.842872488. The clustering results revealed four distinct groups. Cluster 0 included 119 districts and cities with the most severe stunting rates, while Cluster 1 encompassed 3 districts and cities with high stunting rates. Cluster 2 comprised 2 districts and cities with lower stunting rates, and Cluster 3 contained 2 districts and cities with the lowest stunting rates. The analysis indicates that Cluster 3 represents areas with very low stunting rates, Cluster 2 includes areas with low stunting rates, Cluster 1 covers regions with high stunting rates, and Cluster 0 consists of regions with the highest stunting rates.

Acknowledgement

The authors would like to thank and appreciate the editors, reviewers, and all those who have supported us in improving this article. This research was funded by Lembaga Penelitian dan Pengabdian Kepada Masyarakat, Universitas Negeri Gorontalo via PNBP-UNG with DIPA-UNG No. SP DIPA-023.17.2.677521/2023/2023, under contract No. B/704/UN47.D1/PT.01.03/2023.

References

[1] Y. D. Darmi and A. Setiawan, “Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk,” *Jurnal Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017. View online.

[2] S. Soheily-Khah, P. F. Marteau, and N. Bechet, “Intrusion detection in network systems through hybrid supervised and unsupervised machine learning process: A case study on the iscx dataset,” *Proceedings - 2018 1st International Conference on Data Intelligence and Security, ICDIS 2018*, no. July, pp. 219–226, 2018. View online.

[3] Mustakim, U. Khairunnisa, A. Wenda, A. Ilham, F. E. Laumal, A. Daengsgs, D. S. Putra, I. B. A. I. Iswara, S. R. Fitriatien, and R. Rahim, “Unsupervised Learning As a Data Sharing Model in the FP-Growth Algorithm in Determining the Best Transaction Data Pattern,” *Journal of Theoretical and Applied Information Technology*, vol. 99, no. 11, pp. 2679–2689, 2021. View online.

[4] B. Probierz, J. Kozak, and A. Hrabia, “Clustering of scientific articles using natural language processing,” *Procedia Computer Science*, vol. 207, no. Kes, pp. 3443–3452, 2022. View online.

[5] I. P. S. Almantara, N. W. S. Ariyani, and I. B. A. Swamardika, “Spatial Data Analysis using DB-SCAN Method and KNN classification,” *International Journal of Engineering and Emerging Technology*, vol. 5, no. 2, pp. 5–8, 2020. View online.

[6] P. Putra, A. M. H. Pardede, and S. Syahputra, “Analisis Metode K-Nearest Neighbour (KNN) Dalam Klasifikasi Data Iris Bunga,” *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 6, no. 1, pp. 297–305, 2022. View online.

[7] X. Zhang, H. Xiao, R. Gao, H. Zhang, and Y. Wang, “K-nearest neighbors rule combining prototype selection and local feature weighting for classification,” *Knowledge-Based Systems*, vol. 243, p. 108451, 2022. View online.

[8] A. Arafa, N. El-Fishawy, M. Badawy, and M. Radad, “RN-SMOTE: Reduced Noise SMOTE based on DBSCAN for enhancing imbalanced data classification,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 8, pp. 5059–5074, 2022. View online.

[9] L. Qadrini, “Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017,” *Jurnal Statistika : Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 13, no. 2, pp. 5–11, 2020. View online.

[10] S. Budiman, D. Safitri, and D. Ispriyanti, “Perbandingan Metode K-Means dan Metode DB-SCAN Pada Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang Semarang,” *Jurnal Gaussian*, vol. 5, no. 4, pp. 757–762, 2016. View online.

[11] K. A. Abbas, A. Gharavi, N. A. Hindi, M. Hassan, H. Y. Alhosin, J. Gholinezhad, H. Ghoochaninejad, H. Barati, J. Buick, P. Yousefi, R. Alasmar, and S. Al-Saegh, “Unsupervised machine learning technique for classifying production zones in unconventional reservoirs,” *International Journal of Intelligent Networks*, vol. 4, no. December 2022, pp. 29–37, 2023. View online.

[12] A. F. Khairati, A. Adlina, G. Hertono, and B. Handari, “Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA,” *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 2, pp. 161–170, 2019. View online.

[13] M. T. Jatipaningrum, S. E. Azhari, and K. Suryowati, “Pengelompokan Kabupaten dan Kota di Provinsi Jawa Timur Berdasarkan Tingkat Kesejahteraan dengan Metode K-Means dan Density-Based Spatial Clustering Of Applications With Noise,” *Jurnal Derivat : Jurnal Matematika*, vol. 9, no. 1, pp. 70–81, 2022. View online.

[14] E. Rendón, I. Abundez, A. Arizmendi, and E. M. Quiroz, “Internal versus External cluster validation indexes,” *International Journal of Computers and Communications*, vol. 5, no. 1, pp. 27—-34, 2011. View online.

[15] N. Yulia, R. Saragih, and I. Ambarita, “Data Mining Pengelompokan Anak Stunting Berdasarkan Usia, Penyebab dan Pekerjaan Orang Tua Dengan Menggunakan Metode Clustering (Studi Ka-

sus : Dinas Kesehatan Kabupaten Langkat)," *Seminar Nasional Informatika (SENATIKA)Prosiding SENATIKA 2021*, p. 12, 2021. View online.

[16] F. Danso and M. A. Appiah, "Prevalence and associated factors influencing stunting and wasting among children aged 1 to 5 years in Nkwanta South Municipality, Ghana," *Nutrition*, vol. 110, p. 111996, 2023. View online.

[17] WHO and UNICEF, *Low Bithweight Country, Regional and Global Estimate*, vol. 9. 2004. View online.

[18] D. A. Candra, *Pencegahan dan Penanggulangan Stunting*. 2020. View online.

[19] D. Satriawan, "Pengelompokan Provinsi di Indonesia Berdasarkan Faktor Penyebab Balita Stunting," *Buletin Penelitian Sistem Kesehatan*, vol. 24, no. 4, pp. 308–317, 2021. View online.

[20] Kemenkes, "Hasil Survei Status Gizi Indonesia (SSGI) 2022," *Kementrian Kesehatan Indonesia*, pp. 1–7, 2023. View online.

Citation IEEE Format:
Gani et al., "*Implementation of K-Nearest Neighbor Algorithm on Density-Based Spatial Clustering Application with Noise Method on Stunting Clustering*", *Jurnal Diferensial*, vol. 6(2), pp. 170-178, 2024.