Classification of Foods Based on Nutritional Content Using K-Means and DBSCAN Clustering Methods

Fitria Nurulhikmah[1*], Deden Nur Eka Abdi[2]

[1,2]Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, East Java, Indonesia
Email: [1*]6025232060@student.its.ac.id, [2]6025232025@student.its.ac.id

(Received: 8 Sep 2024, revised: 9 Oct 2024, 21 Oct 2024, accepted: 22 Oct 2024)

Abstract

This study classifies foods based on their nutritional content using K-Means and DBSCAN clustering methods. The clustering quality was evaluated using the Davies-Bouldin Index (DBI) and Silhouette Score. K-Means was tested with different k values, while DBSCAN was analyzed with varying min_samples parameters. Additionally, a function was developed to group foods into three categories: Weight Gain, Obesity Prevention, and Weight Loss, based on calories, protein, fat, and carbohydrate content. The results show that K-Means is more effective than DBSCAN in clustering foods by nutritional content, yielding lower DBI values and higher Silhouette Scores. For example, K-Means with k = 3 achieved a DBI of 0.694930 and a Silhouette Score of 0.538921, while DBSCAN with eps = 0.75 and min_samples = 4 produced a DBI of 0.34546577 and a Silhouette Score of 0.492830814. This study concludes that K-Means provides superior clustering performance, enabling more specific dietary recommendations tailored to individual nutritional needs.

Keywords: Clustering, K-Means, DBSCAN, Nutritional

I. INTRODUCTION

Nutritional data plays a vital role in many aspects of life, from individual health and wellness to dietary programs and the food industry. Proper nutrition not only impacts physical health but also mental well-being, thus having a comprehensive understanding of food nutrition is essential to meet societal needs and promote healthy living. Nutritional knowledge enables better decision-making regarding food consumption, which can help prevent diet-related diseases and improve overall well-being [1].

However, one of the major challenges in nutritional data analysis is the fragmentation of information. Nutritional content data is often scattered across various sources, such as food product labels, research studies, and databases, making it difficult to gather and analyze comprehensively [2]. This fragmented data complicates efforts for nutritionists, researchers, and consumers to identify clear patterns in nutritional content and make informed dietary recommendations.

To address this, clustering techniques offer a practical approach by grouping foods with similar nutritional profiles. Clustering is a machine learning method that identifies groups, or "clusters," of data points with similar characteristics [3]. In the context of food nutrition, clustering can group foods based on attributes such as protein, fat, carbohydrates, vitamins, and minerals. Such techniques can simplify the food selection process, aligning it with individual nutritional needs, thus helping to prevent diet-related health issues.

This research focuses on classifying foods based on nutritional content using two clustering methods: K-Means and DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Both methods are applied to group foods based on their nutrient composition, which is critical for recommending appropriate dietary choices that cater to weight management, obesity prevention, and healthy living. The K-Means algorithm, suitable for numerical data, faces challenges in determining the optimal number of clusters. To resolve this, evaluation metrics such as the Silhouette Coefficient and Davies-Bouldin Index (DBI) are used to assess clustering quality [4].

On the other hand, DBSCAN is selected for its ability to handle data with varying shapes and densities and its effectiveness in detecting outliers. By comparing the results of these two methods, this research aims to provide a comprehensive analysis of clustering techniques in food classification based on nutritional data [5].

II. RELATED WORK

Previous studies have applied clustering methods in the context of health and nutrition with various approaches aimed at improving understanding of food consumption patterns. For example, a study by Lin (2021) applied the hierarchical clustering method to analyze nutritional data at the country level. The aim was to develop a fair and sustainable food system. The results of this study showed that although the existing food system is quite efficient, there are still inequalities in consumption patterns and nutrient distribution across regions [1].

In the context of health data analysis in Indonesia, a study by Septiani (2022) used the K-Means algorithm to cluster foods based on their nutritional content. The aim of this study was to provide more precise dietary recommendations based on clustered nutritional data. The results showed that K-Means was effective in clustering foods based on nutritional value, with evaluation using the Davies-Bouldin Index (DBI) and Silhouette Score to assess the quality of clustering [6].

Another study by Harjanto et al. (2021) compared K-Means with DBSCAN in the context of clustering patient data based on health conditions. The aim of this study was to evaluate the effectiveness of both algorithms in identifying relevant patterns from data that often contains noise. The results showed that DBSCAN was superior in dealing with data with irregular cluster shapes and in detecting outliers, while K-Means was better for data with a clearer distribution [7].

This study aims to compare the effectiveness of the K-Means and DBSCAN algorithms in clustering food data based on their nutritional content, such as calories, protein, fat, and carbohydrates. Using the same evaluation metrics, namely DBI and Silhouette Score, the results showed that K-Means provided better clustering quality than DBSCAN, with lower DBI values and higher Silhouette scores at k = 3. In addition, this study found that the formed clusters showed significant differences in nutritional profiles between categories, reflecting the potential use of clustering methods to provide more accurate and targeted dietary recommendations.

III. DATASET AND FEATURES

In this study, we describe the dataset used for food nutrition analysis, including the number of examples used for training, validation, and testing, the preprocessing steps applied, and the features utilized in this research.

1. Dataset Description

The dataset used in this study was sourced from [https://kaggle.com/datasets/anasfikrihanif/indonesian-food-and-drink-nutrition-dataset]. This dataset contains nutritional information for different types of foods in Indonesia, including calories, protein, fat, and carbohydrate content. The dataset contains 1,346 food examples, which were categorized as shown in Figure 1.

id	calories	proteins	fat	carbohydrate	name	image
1	280	9.2	28.4	0	Abon	https://img-cdn.medkomtek.com/PbrY9X3ignQ8sVuj_LU9UXljyew=/0x0
2	513	23.7	37	21.3	Abon haruwan	https://img-global.cpcdn.com/recipes/cbf330fbd1ba6316/1200x630cq
3	0	0	0.2	0	Agar-agar	https://res.cloudinary.com/dk0z4ums3/image/upload/v1644389489/at
4	45	1.1	0.4	10.8	Akar tonjong segar	https://images.tokopedia.net/img/cache/200-square/product-1/2018/
5	37	4.4	0.5	3.8	Aletoge segar	https://nilaigizi.com/assets/images/produk/produk_1577340236.jpg
6	85	0.9	6.5	7.7	Alpukat segar	https://katakabar.com/assets/images/upload/news/medium_news_1
7	96	3.7	0.6	19.1	Ampas kacang hijau	https://images.tokopedia.net/img/cache/215-square/shops-1/2016/10
8	414	26.6	18.3	41.3	Ampas Tahu	https://palpres.disway.id/upload/9e9c1ba592cac7270de5e48889a3352
9	75	4.1	2.1	10.7	Ampas tahu kukus	https://cdns.diadona.id/diadona.id/resized/640x320/news/2020/02/26
10	67	5	2.1	8.1	Ampas tahu mentah	https://cdn-image.hipwee.com/wp-content/uploads/2021/03/hipwee-
11	184	18.8	14	0	Anak sapi daging gemuk segar	https://png.pngtree.com/png-clipart/20220124/ourmid/pngtree-a-plat
12	174	19.6	10	0	Anak sapi daging kurus segar	https://asset.kompas.com/crops/BzqLdMzV0nYuuT_-vqxhGoDU0ag=/0
13	190	19.1	12	0	Anak sapi daging sedang segar	https://koran-jakarta.com/images/article/tips-memilih-daging-sapi-s
14	99	4.6	1	18	Andaliman segar	https://cf.shopee.co.id/file/3c033b2480e951e2634909817e417218
15	25	1.6	0.2	5.3	Andewi	https://www.satuharapan.com/uploads/cache/news_77440_15181851
16	30	0.5	0.2	6.8	Anggur hutan segar	https://cf.shopee.co.id/file/ed7f26c6de5f13ec39e9931263ade6b9
17	354	16.4	31.5	0	Angsa	https://cdn.idntimes.com/content-images/community/2020/06/image
18	126	3.4	7.9	10.3	Anyang sayur	https://static.republika.co.id/uploads/images/inpicture_slide/anyang
19	58	0.3	0.4	14.9	Apel	https://asset.kompas.com/crops/smfd25xgXRE3HpMLb2aamPeulYM=/
20	57	0.5	0.4	12.8	Apel malang segar	https://pasarsegar.co.id/wp-content/uploads/2020/09/name-63-1.jpg

Figure 1. Dataset from Kaggle Website

Before clustering analysis, we performed several preprocessing steps to ensure data quality and maximize the performance of the considered algorithms. The preprocessing steps included the following steps as shown in Figure 2.

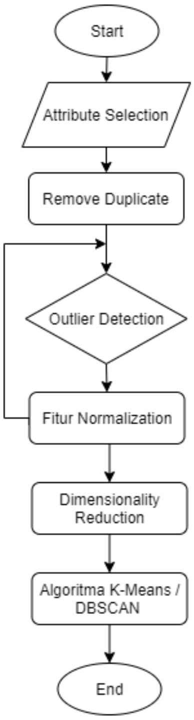

Figure 2. Preprocessing Flow

2. Attribute Selection

The process of selecting a subset of the most relevant features for use in analysis or machine learning models aims to enhance model performance by reducing data dimensionality and eliminating irrelevant or redundant features.

3. Remove Duplicates

Removing duplicates refers to eliminating repeated entries in a dataset to ensure data uniqueness. This is a crucial step in data preprocessing that can enhance data quality and analysis accuracy.

4. Outlier Detection

Cleaning the dataset of outliers before applying clustering algorithms, such as K-Means and DBSCAN, is crucial to ensure that the analysis is based on representative data and is not influenced by extreme values. This process typically

DOI: 10.34148/teknika.v13i3.1067

begins with outlier detection using the Interquartile Range (IQR) method, where values outside the range defined by the first quartile (Q1) and third quartile (Q3) are identified as outliers. Once the outliers were identified, the next step removing them from the dataset to ensure that the clustering results were not distorted by uncommon values. This process is essential for ensuring the validity of clustering results and improving the analysis accuracy [4][8].

5. Fitur Normalization

Feature normalization plays a crucial role in the performance of the K-Means and DBSCAN clustering algorithms. For K-Means, feature normalization is necessary because this algorithm is sensitive to the scale of features in distance calculations between data points, which can affect the formation of optimal clusters. DBSCAN also requires feature normalization to ensure consistency in distance measurements within the feature space, which is crucial for identifying density-based clusters. By standardizing or normalizing features using methods like *StandardScaler* before applying these algorithms, both K-Means and DBSCAN can yield more accurate and consistent clustering results. This facilitates better interpretation of the data structure [9].

IV. METHODS

In this section, we describe the clustering algorithms used in this research: K-Means and DBSCAN. The proposed algorithms were applied to a food nutrition dataset to cluster data based on similar nutritional characteristics.

1. K-Means

K-Means clustering method is capable of dividing data into two or more groups [10]. This method group data such these data with similar characteristics are grouped together, while data with different characteristics are placed in different groups [11]. K-means clustering algorithm is one of the simplest and most commonly used clustering algorithms.

To calculate the distance between the i data point (xi) and the k cluster center (ck), which is called (dik), the following Euclidean Formula 1 is used:

$$d_{ik} = \sqrt{\sum_{j=1}^{m} \left(c_{kj} - x_{ij}\right)^2} \tag{1}$$

Here, $c_{kj}$is the coordinate of the k cluster center in dimension j, and x_{ij} is the coordinate of the i data in dimension j. This equation is used to determine the proximity between data points and cluster centers, and it optimizes data clustering based on their characteristics.

2. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

A Density-Based Spatial Clustering Algorithm with Noise (DBSCAN) can identify high-density core samples and expand clusters from these samples [10]. The proposed algorithm has two main parameters that determine cluster formation: minimum number of samples (MinPts) and epsilon distance (ε). The first parameter determines the minimum number of points that can be considered as a core sample and determines the algorithm's noise tolerance level [3][5].

The steps involved in the DBSCAN algorithm are as follows:

1. MinPts and Eps parameters.
2. Randomly select a starting point (p).
3. Calculate the epsilon distance (Eps) or all reachable point distances from p using the following Euclidean distance Formula (2):

$$d_{ij} = \sqrt{\sum_{a=1}^{p} \left(x_{ia} - x_{ja}\right)^2} \tag{2}$$

Where x_{ia} is the a variable of the i object (i=1, ..., n; a=1, ..., p) and $d_{ij}$is the value of the Euclidean distance.

4. A cluster is formed if the number of points within the epsilon distance (Eps) is greater than MinPts, and point p becomes a core point.
5. Repeat steps 3-4 until all points have been processed. If p is a border point and no points are reachable from p, the process continues to the next border point.

By applying these steps, DBSCAN can recognize high-density clusters and outliers in the dataset, which makes it an effective method for data analysis with varying shapes and densities.

3. Silhouette Coefficient

The silhouette coefficient is a useful tool for measuring confidence in the clustering process of observations. The silhouette coefficient values range from -1 to 1, where clusters with a silhouette coefficient closer to 1 are considered good, while those closer to -1 are considered poor [12]. Calculating the Silhouette Value for Each Object ($s(i)$) is as Formula 3 follows:

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

$a(i)$ is the average distance between object iii and all other objects in the same cluster. $b(i)$ is the average distance between object i and all objects in the nearest different cluster. The silhouette value ranges between $-1 \leq s(i) \leq 1$ [7].

4. Davies-Bouldin Index (DBI)

The Davies-Bouldin Index (DBI) is a metric used to evaluate the quality of clustering by considering the separation and cohesion values within clusters. Separation refers to the distance between cluster centroids, and cohesion represents the similarity of data points to the centroid of their respective clusters. A smaller DBI indicates better clustering results as shown in Formula 4.

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{i \neq \mathrm{j}} \left(Ri, \mathrm{j}\right) \tag{4}$$

k is the number of clusters, and Ri,j is the ratio between clusters i and j.

V. EXPERIMENTS

Implementing the DBSCAN and K-Means algorithm with several preprocessing steps, such as data selection, duplicate removal, outlier detection, data normalization, and dimensionality reduction, based on nutritional criteria yields different clustering results for each food category. Subsequently, an experimental process was conducted to apply the DBSCAN algorithm and the K-Means algorithm for comparison, as illustrated in Figure 3.

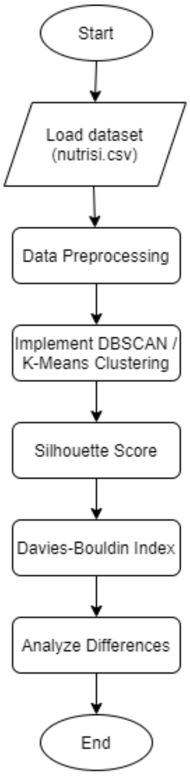

Figure 3. The Experimental Execution Flow

In this study, we explored determining the optimal epsilon (eps) value in the DBSCAN clustering algorithm and the optimal k value in the K-Means algorithm. The Davies-Bouldin Index (DBI) was used as the evaluation metric to compare the clustering quality of both methods.

The implementation involved DBSCAN with various optimal eps values and min_samples = 4, 5, 6, 7, and 8. The determination of the optimal epsilon (eps) value in the DBSCAN clustering algorithm is critical for obtaining good clustering results. One approach to determining the optimal epsilon value is using the K-Nearest Neighbor (KNN) method, as illustrated in Figure 4. The results obtained by applying DBSCAN are presented in Table 1.

The DBSCAN algorithm was implemented using various optimal epsilon (**eps**) values alongside different **min_samples** parameters set to 4, 5, 6, 7, and 8. Determining the optimal epsilon value in the DBSCAN clustering algorithm is crucial for achieving effective clustering results. One method employed to find the optimal epsilon value is the K-Nearest Neighbor (KNN) approach, where the distance to the k-th nearest neighbor is calculated to create a distance graph. This helps in visualizing the point at which the density of points significantly drops, indicating the appropriate value for **eps**.

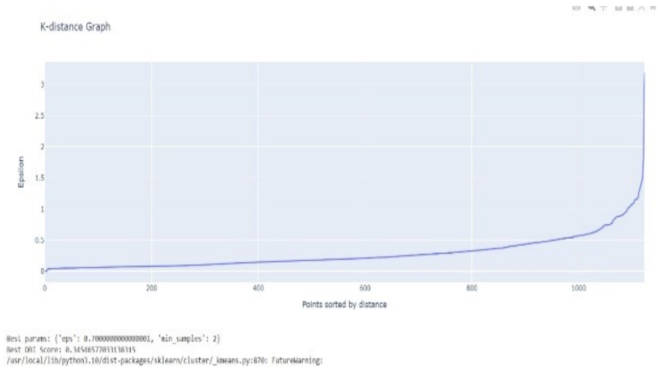

Figure 4. Optimal epsilon values for DBSCAN

Table 1. Comparison DBI and Silhouette Scores of DBSCAN

Eps	Min_Sample	DBI Score DBSCAN	Silhouette Score DBSCAN
0.7	4	0.4096385287	0.478855636
0.7	5	0.4096385287	0.478855636
0.7	6	0.839437517	0.459351336
0.7	7	0.839437517	0.459351336
0.7	8	0.862154841	0.470780629
0.75	4	0.34546577	0.492830814
0.75	5	0.369333203	0.487545501
0.75	6	0.369333203	0.487545501
0.75	7	0.369333203	0.487545501
0.75	8	0.737241311	0.471212884
0.8	4	0.34546577	0.492830814
0.8	5	0.34546577	0.492830814
0.8	6	0.34546577	0.492830814
0.8	7	0.369333203	0.487545501
0.8	8	0.788498651	0.459580366

Table 1. Comparison of DBI and Silhouette scores across different DBSCAN configurations, illustrating the performance at various Eps and min_sample values.

Implementation was performed using K-Means with the optimal k value. The k value obtained from Elbow Method. This resulted in determining the optimal k value for the K-Means clustering method. The results of applying DBSCAN and K-Means are presented in Table 2.

Table 2. DBI and Silhouette Scores in K-Means

K	DBI Score K-Means	Silhouette Score K-Means
2	0.859863	0.528081
3	0.694930	0.538921
4	0.741324	0.491049

DOI: 10.34148/teknika.v13i3.1067
ISSN 2549-8037, EISSN 2549-8045

5	0.746615	0.473861
6	0.744451	0.476737
7	0.737927	0.475092

Figures 5 and 6 present the results of clustering the nutritional data of foods using the DBSCAN and K-Means algorithms. In these graphs, the horizontal axis represents the number of Calories, and the vertical axis represents the amount of Fat in various types of foods. Three clusters are identified, each marked with a different color:

- Cluster 0 (indicated by blue dots): Foods that tend to have low-calorie and fat content.
- Cluster 1 (indicated by red dots): foods with moderate-calorie content and varying fat content.
- Cluster 2 (indicated by green dots): Foods with high caloric content and varying fat content

Figure 5. DBSCAN Clustering Results

Figure 6. K-Means Clustering Results

In this study, we developed the `categorize_cluster` function, which classifies foods into three categories based on their nutritional clustering. These categories are determined by key nutritional indicators such as calories, protein, fat, and carbohydrates. This function aids in specific dietary recommendations aligned with the goal of either weight gain, obesity prevention, or weight loss.

Weight Gain Enhancer foods are categorized as those with high caloric content, typically ranging from 250 to 400 calories, which are suitable for individuals aiming to increase body weight. Obesity Prevention foods are characterized by moderate caloric content, also within the 250 to 400 calorie range. These foods provide adequate protein intake but should be consumed in moderation to prevent excessive weight gain. Weight-loss foods, with caloric content ranging from 0 to 250 calories, are suitable for diets aimed at reducing weight. However, unbalanced consumption of these low-calorie foods may lead to nutritional deficiencies. These calorie ranges were inferred from the clustering results based on the nutritional content analyzed in this study.

In this approach, we provide more specific recommendations in diet management and food selection tailored to individual nutritional needs. The results of the food clustering are presented in Tables 3 and 4, with 10 samples taken from each cluster. However, to clarify, the categorization presented in Table 3 (Obesity Prevention, Weight Loss, Weight Gain Enhancer) is directly derived from the nutritional characteristics of the clusters, based on key nutritional metrics such as calorie, protein, fat, and carbohydrate content. These categories were determined after applying the K-Means and DBSCAN algorithms, as well as evaluating their performance using the Davies-Bouldin Index and Silhouette Score, which ensures the clustering's validity. The calculations and justifications for these categories are aligned with the evaluation of clustering quality presented earlier in this study.

Table 3. List of Foods from DBSCAN Clustering Results

Cluster	Category	Food List
-1	Undefined (Noise)	Coconut meal, fried oncom (fermented soybean), fried eel, fresh soybeans, tofu dregs, fresh green soybeans, cowpea seeds, country beans, fried fresh eel, dried starfruit, etc.
0	Obesity Prevention	Fried rice, macaroni, honey, young mango, pastry, crab, purple potatoes, cream puffs, pineapple cakes, tofu, beef intestines, etc.

Table 4. List of Foods from K-Means Clustering Results

Cluster	Category	Food List
0	Low-calorie weight loss	Agar-agar, pumpkin, soy sauce, radish, turmeric, kool kembang, mangosteen, young mango, lime, young snow peas, etc.
1	Obesity prevention	Fried chicken thigh, fresh shrimp, shredded beef, beef, tofu egg, pomfret fish, ice cream, crab, fried carp fish, cream puffs, etc.
2	Weight gain enhancer	Tepung terigu, Wheat flour, wingko babat (coconut cake), jam, tiwul (fermented cassava), rice flour, emping rice crackers, fried banana, onde-onde (fried glutinous rice ball), kue satu (layered cake), corn rice, etc.

VI. CONCLUSION

This study successfully applied the DBSCAN and K-Means clustering algorithms to classify food data based on nutritional content, utilizing the Davies-Bouldin Index (DBI) and Silhouette Score as evaluation metrics. The results revealed that the optimal parameters for DBSCAN were an epsilon (eps) value of approximately 0.75 with a minimum sample size of 4, achieving a DBI score of 0.34546577 and a Silhouette Score of 0.492830814. Conversely, K-Means produced the best clustering outcomes with an optimal cluster count of 3, resulting in a DBI score of 0.694930 and a Silhouette Score of 0.538921. Furthermore, the categorization of food items into three distinct groups—Weight Gain Enhancers, Obesity Prevention, and Weight Loss—was facilitated through the development of a categorize_cluster function, providing specific dietary recommendations tailored to individual nutritional needs. This research underscores the significance of parameter optimization in clustering algorithms, demonstrating their potential in offering more accurate and actionable insights into food selection based on nutritional content. Overall, the findings highlight the applicability of these clustering methods in enhancing dietary decision-making and addressing the complexities of food nutrition.

REFERENCES

[1] Lin, Q. (2021). Equity-ecological food system based on hierarchical clustering. *2021 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*.

[2] Hargreaves, D., Mates, E., Menon, P., Alderman, H., Devakumar, D., Fawzi, W., Greenfield, G., Hammoudeh, W., He, S., Lahiri, A., Liu, Z., Nguyen, P. H., Sethi, V., Wang, H., Neufeld, L. M., & Patton, G. C. (2022). Strategies and interventions for healthy adolescent growth, nutrition, and development. *The Lancet, 399*(10320), 198–210.

[3] D. Deng, “Research on Anomaly Detection Method Based on DBSCAN Clustering Algorithm,” in Proceedings - 2020 5th International Conference on Information Science, Computer Technology and Transportation, ISCTT 2020, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 439– 442. doi: 10.1109/ISCTT51595.2020.00083.

[4] Vredizon, P. A., Firmansyah, H., Salsabila, N. S., & Nugroho, W. E. (2024). Penerapan algoritma K-Means untuk mengelompokkan makanan berdasarkan nilai nutrisi. *Journal of Technology and Informatics (JoTI)*, *5*(2), April.

[5] Z. Zhou, L. Luo, Z. Li, M. Chao, W. Tang, and Y. Wang, “Harmonic pollution zoning method based on improved DBSCAN clustering,” in China International Conference on Electricity Distribution, CICED, IEEE Computer Society, 2022, pp. 1099–1103. doi:

[6] I. W. Septiani, Abd. C. Fauzan, Dan M. M. Huda, “Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker ParuParu,” Jurnal Sistem Komputer Dan Informatika (Json), Vol. 3, No. 4, Hlm. 556, Jul 2022, Doi: 10.30865/Json.V3i4.4055

[7] Muhammad Sholeh, Dina Andayati, Rr. Yuliana Rachmawati. (2022). "Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor dengan Normalisasi untuk Prediksi Penyakit Diabetes." Jurnal TeIKa, Volume 12, Nomor 2, Oktober 2022.

[8] A. Ambarwari, Q. Jafar Adrian, dan Y. Herdiyeni. (2020). "Analysis Of The Effect Of Data Scaling On The Performance Of The Machine Learning Algorithm For Plant Identification." Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), Vol. 4(No. 1), hlm. 117-122. DOI: 10.29207/resti.v4i1.1517.

[9] Hadikurniawati, W., Hartomo, K. D., & Sembiring, I. (2023). Spatial clustering of child malnutrition in Central Java: A comparative analysis using K-Means and DBSCAN. *Proceedings of the 2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA)*.

[10] Rimelda Adha, Nana Nurhaliza, Ummi Soleha, Mustakim. (2021). Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia. Jurnal Sains, Teknologi dan Industri, Vol. 18(No. 2), 206-211. ISSN 1693-2390 (print)/ISSN 2407-0939 (online).

[11] Rahmawati, & Faisal, M. (2019). Analisis Cluster untuk Pengelompokan Desa Berdasarkan Indikator Penyakit Diare. Saintifik, Vol. 5(No. 1), 75-80.

[12] Harjanto, T. D., Vatresia, A., & Faurina, R. (2021). Analisis penetapan skala prioritas penanganan balita stunting menggunakan metode DBSCAN clustering. *Jurnal Rekursif*, *9*(1), 1-XX. ISSN 2303-0755.