Journal of Information Systems and Informatics
Vol. 6, No. 4, December 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i4.932 Published By DRPM-UBD

Clustering Sugar Content in Children's Snacks for Diabetes Prevention Using Unsupervised Learning

Irma Darmayanti[1]*, Dhanar Intan Surya Saputra[2], Inka Saputri[3], Nurul Hidayati[4], Nandang Hermanto[5]

[1,3,4]Information Technology Department, Amikom Purwokerto University, Indonesia
[2,5]Informatics Department, Amikom Purwokerto University, Indonesia
Email: [1]irmada@amikompurwokerto.ac.id, , [2]dhanarsaputra@amikompurwokerto.ac.id
[3]inkasaputri1@gmail.com, [4]nurulhid371@gmail.com,
[5]nandanghermanto@amikompurwokerto.ac.id

Abstract

Diabetes is a chronic health problem with increasing prevalence, especially among children, due to the consumption of sugary foods/beverages. This study aims to cluster children's snack products based on sugar content using unsupervised learning by combining Hierarchical Clustering and K-Means algorithms optimized using Silhouette Score. This combined approach utilizes Hierarchical Clustering to determine the optimal value (k) of K-Means, ensuring the efficiency and accuracy of data clustering. A total of 157 sample data were effectively clustered with K-means. The test results with Silhouette Score yielded the highest value of 0.380 for 2 clusters, while 3 clusters scored 0.350 and 0.277 for 4 clusters. For this reason, 2 clusters better represent the homogeneity of the data in the cluster, although it has not reached the ideal condition. Further analysis showed that high sugar and calorie content in sugary drinks, including milk, could increase blood glucose levels. These findings can be the basis for the development of consumer-friendly nutrition labels. However, support is needed from the government to create a labelling policy to ensure the sustainability of implementation in the community as an educational effort to prevent the risk of diabetes in children.

Keywords: Diabetes, Snacks, Sugar, Clustering, Unsupervised Learning

1. INTRODUCTION

Diabetes is a global health problem that is increasing in prevalence and is the main focus of efforts to prevent and manage chronic diseases[1]. Diabetes is a chronic disease that can lead to major side effects such renal failure, heart disease, blindness, and amputation of limbs[2]. It is estimated that more than 400 million people worldwide have diabetes, and it is projected that number to rise to 700 million by 2045[3]. Worryingly, diabetes is now commonly found in children and adolescents[4].

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Findings from an observational study conducted in six regions in the United States show that, between 2001 and 2017, the number of young persons with type 1 and type 2 diabetes cases has significantly increased[5]. This phenomenon also occurs in Indonesia, where the prevalence of diabetes in children is reported by IDAI to reach two cases per 100,000 children, with most sufferers aged between 10-14 years old[6]. The problem is compounded by unhealthy diets and lack of physical activity in the digital age, where children spend more time on activities such as playing games or watching videos[7], [8].

Parents tend to prioritise calmness and comfort by buying instant food or snacks that children like, and they neglect adequate physical activity[9][10]. Lack of parental knowledge about the nutritive value of food and drink itemsmakes many children accustomed to sweet snacks from an early age[11][12]. This condition is exacerbated by parents' limited time, knowledge, and awareness of the importance of maintaining nutritional balance in their daily diet[13]. In fact, the habit of consuming unhealthy products high in energy, sugar and salt can trigger blood pressure and the chance of developing diabetes and obesity in later life [14].

Based on data from Indonesia's Basic Health Research (Riskesda) in 2018 shows that the level of consumption of sugary foods and drinks in Indonesia is very high, namely 87.9% consumption of sugary foods and 91.49% consumption of sugary drinks[15], as attached in Table 1.

Table 1. Percentage of Risky Food and Beverage Consumption in Ages 3 Above

ID	Type	Total (%)
1	Sugary Drinks	91,49
2	Foods with MSG	88,4
3	Instant Noodles	87,9
4	Sweet Foods	87,9
5	Fatty Foods	86,7
6	Salty Food	72,7
7	Burnt Food	39
8	Processed Food	27,7
9	Carbonated Food	13,2

This Table shows the prevalence of excessive sugar consumption in society, including among children. This high consumption of sugary foods and beverages is one of the primary risk factors for diabetes, especially diabetes type 2, which is now increasingly found in young people[16]. Therefore, it is essential to identify product categories for food and drink favored by children based on their nutritional content that could trigger diabetes to help the public, especially parents, make healthier food choices for their children.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Previous research on public health and diabetes risk factors has shown that unhealthy food consumption patterns, including foods high in fat and sugar, contribute significantly to the occurrence of diabetes mellitus[17]. A study at RSUD Dr. Fauziah Bireuen Aceh Province also revealed a strong correlation between the prevalence of diabetes mellitus and a diet high in fatty, sugary, and salty foods[18]. Other research points to the fact that children with longer screen time are less likely to consume vegetables and fruits, while snacks and sugar-sweetened beverages (SSBs) tend to be higher[19].

These findings emphasise the importance of a broader strategic approach to encourage healthy eating to prevent and control diabetes early on. One approach that can be applied is to cluster children's snacks based on their nutritional content. Unsupervised learning techniques can cluster snacks based on sugar content, making it easier for parents and children to make healthier food choices. With precise clustering results, it is hoped to increase nutritional awareness from an early age, which is the primary goal of this research. It also encourages changes in healthier eating behavior, supporting efforts to prevent diabetes at a young age.
In this regard, the K-Means and Fuzzy C-Means algorithms have proven effective in mapping products' nutritional content to identify health risks[20]. In addition, hierarchical clustering can divide products into two main categories that provide insight into the health profile of each product group[21]. The K-Means method also successfully identified 40 packaged products that should be avoided in diabetic patients[22].

Considering the results of applying data clustering using algorithms from unsupervised learning techniques, we applied a novel approach focusing on snacks for children in this study. This less researched product group often contains high sugar content, It greatly raises the chance of developing diabetes at a young age, by combining Hierarchical and K-Means Clustering algorithms in the Unsupervised Learning method to cluster children's snacks based on sugar content, which is optimized using silhouette score to improve clustering accuracy. One strategy to increase awareness of the value of a healthy diet and lower the risk of diabetes at an early age is to group children's snacks according to their sugar content. This will help to promote early diabetes prevention.

2. METHODS

This section describes the systematic approach applied in analyzing children's snack products based on sugar content to classify these products using unsupervised learning techniques. The process steps include data collection, pre-processing, algorithm application, evaluation, and interpretation of results. The process is designed to ensure that the analyses are transparently reproducible. A visual understanding of the methodology is seen in Figure 1.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

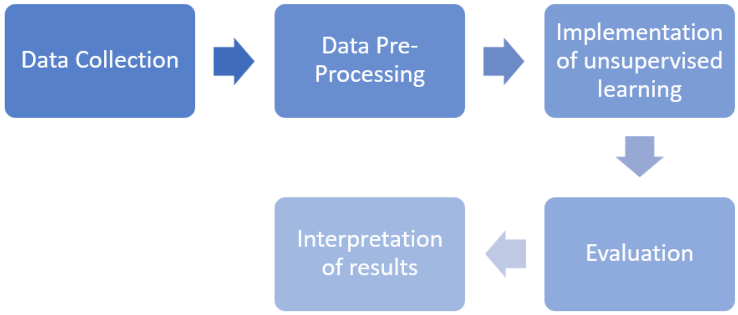


Figure 1. Method

2.1. Data Collection

Data collection is an essential phase in providing the basis for analysis using unsupervised learning techniques, namely Hierarchical Clustering and K-Means Clustering. the data collection process was carried out by taking samples of children's snack products from various well-known brands often consumed by Indonesians. The information collected includes key nutritional content, such as sugar, calories, total fat, protein, carbohydrates, and salt, obtained from product packaging labels. These components are recognized as factors that can increase the risk of diabetes if consumed in excess[23][24]. A whopping 162 product data points were collected, covering various brands and variants of snacks that are popular with children. Below are the top 5 and bottom 5 data points from the collected data.

Table 2. Sample Data

No	Product	Sugar	Calorie	Fat	Carbohydrate	Protein	Sodium
1	Ultra Milk Chocolate	19	160	4	24	6	0.045
2	Yakult	11	50	0	12	1	0.03
3	Cimory	31	220	7	33	6	0.1
4	Yoforia	20	140	2	24	4	0.068
…	…	…	…	…	…	…	…
158	Lemonilo Chimi	1	110	7	12	0	0.08
159	Chitato Sapi Bumbu Bakar	0	100	6	11	2	0.105
160	Chitato Lite	1	110	6	11	1	0.085
161	Pota Bee Black Truffle	1	160	10	15	3	0.16
162	Jet Z Sweet Stick	5	100	4,5	14	1	0.075

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2.2. Data Pre-Processing

The pre-processing phase is essential in ensuring the data quality is met in the analysis. At this stage, missing values and anomaly detection are checked to ensure that the data used is complete and does not contain any discrepancies that could affect the results[25][26]. One of the main steps in pre-processing is data standardisation, which aims to convert variables into a consistent scale. This step is crucial because, in clustering techniques, scale differences between variables can give greater weight to variables with larger values, leading to data bias[27]. With Z-score normalisation, each variable is treated equally, which allows the clustering algorithm to work more optimally without being affected by different variable scales. Z-score is formulated as shown in Equation 1.

$$Z= \frac{x-\bar{x}}{\sigma} \quad (1)$$

2.3. Implementation of Unsupervised Learning

This research uses a combination of Hierarchical Clustering and K-means algorithms, which provide multiple advantages. Hierarchical clustering gives the data structure a preliminary overview, identifies the number of existing clusters, and provides insight into the similarities between products. The Hierarchical Clustering method effectively builds a hierarchy of clusters based on the proximity or similarity of data to each other [28]. On the other hand, K-Means is more efficient in performing clustering based on a predetermined number of clusters, thus providing more precise results in grouping products based on sugar content. In this research, the K-Means process is performed after Hierarchical Clustering to optimise the clustering results. In the application of Hierarchical Clustering, the average linkage technique is used to determine how to measure the distance between the clusters formed by calculating the average distance between all pairs of data points, where each pair consists of one point from each cluster.

$$d_{GA}(G,H)= \frac{1}{N_G N_H} \sum i \in G \sum \acute{i} \in H \, d_{i\acute{i}} \quad (2)$$

The application of K-Means uses the results of hierarchical clustering as the initial initialisation to ensure a more targeted distribution of initial centroids. Thus, this technique improves the accuracy of clustering data based on sugar content. It allows for clearer identification of clusters of children's snack products that are high in sugar or otherwise. This combination allows us to use both strengths: Hierarchical Clustering provides an overall picture, while K-Means provides a more focused clustering.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2.4. Evaluation

Evaluation of Hierarchical Clustering and K-Means results using metrics such as the Silhouette Score aims to ensure the quality and accuracy of clustering. Silhouette Score measures how healthy objects in a cluster neighbour each other, with objects in the same cluster and objects in different clusters[29]. Higher values indicate that the data is more well-clustered[30]. Silhouette Score is formulated as Shown in Equation 3 to 8.

1) Calculate the two-component SI values of the i-th data:

$$d_i^j = \frac{1}{m_j - 1} \sum_{\substack{r=1 \\ r \neq 1}}^{m_j} d(x_i^j, x_r^j) \tag{3}$$

$$b_i^j = \min_{\substack{n=1,\ldots k \\ n \neq j}} \left\{ \frac{1}{m_n} \sum_{r=1, r \neq 1}^{m_n} d(x_i^j, x_r^n) \right\} \tag{4}$$

2) Calculate the Silhouette Index of the i-th data in a cluster:

$$SI_i^j = \frac{b_i^j}{\max\{d_i^j, b_i^j\}} \tag{5}$$

3) Average Silhouette Index of cluster j:

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{m_j} SI_i^j \tag{6}$$

4) Calculate the global SI values:

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{m_j} SI_i^j \tag{7}$$

5) Calculate the average Silhouette Index of the dataset

$$SI = \frac{1}{k} \sum_{j=1}^{k} SI_j \tag{8}$$

2.5. Interpretation of result

Interpretation of results is the final stage of the data analysis process, which aims to label the clusters that have been formed. The labelling of clusters is based on the dominant sugar content in the group. The labels are created to provide information easily understood by the public, especially parents, when choosing healthier snacks for their children.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3. RESULTS AND DISCUSSION

From the total 162 sample data collected, 157 met the criteria and were ready for analysis after cleaning, including removing incomplete or inconsistent data. Processed data is then normalised to ensure that each variable has a uniform scale, so that clustering analysis can be carried out effectively. Below are the top 5 punti dati and bottom 5 punti dati of the dataset that have gone through the pre-processing phase.

Table 2. Dataset

No	Product	Sugar	Calorie	Fat total	Carbohydrate	Protein	Sodium
1	Ultra Milk Chocolate	1,1205	1,5880	0,2049	1,1138	2,3129	-0,0822
2	Anchor Boneeto UHT Creamy Vanilla	0,4894	-0,3803	-0,5082	-0,2192	0,7574	-0,0816
3	Yakult	0,2790	-1,1183	-1,2212	-0,3404	-0,2796	-0,0830
4	Cimory	2,3828	3,0641	1,2745	2,2045	2,3129	-0,0793
5	Yoforia	1,2257	1,0959	-0,5082	1,1138	1,2759	-0,0810
…	…	…	…	…	…	…	…
153	Lemonilo Chimi	-0,7729	0,3578	1,2745	-0,3404	-0,7981	-0,0803
154	Chitato Sapi Bumbu Bakar	-0,8781	0,1118	0,9179	-0,4616	0,2389	-0,0790
155	Chitato Lite	-0,7729	0,3578	0,9179	-0,4616	-0,2796	-0,0801
156	Pota Bee Black Truffle	-0,7729	1,5880	2,3440	0,0232	0,7574	-0,0761
157	Jet Z Sweet Stick	-0,3522	0,1118	0,3831	-0,0980	-0,2796	-0,0806

The first analysis was done with Hierarchical Clustering using average linkage to determine the similarity based on the nutritional content of sugar, calories, total fat, protein, carbohydrates, and sodium with similar values. Hierarchical clustering analysis is best visualised with a dendrogram[31]. Visualisation of Hierarchical Clustering using average linkage shows the formation of a hierarchy of clusters that allows identification of patterns in data, as shown in Figure 2.

Reading the dendrogram is done from right to left, with clusters with high similarity being merged early into the same group. The horizontal lines in the dendrogram show the distance or the extent of dissimilarity among clusters and connect all data that belongs to one group. This line becomes essential for determining the final number of clusters, especially after the decision to stop the merging process. Dendrogram visualisation helps reveal the data's hierarchical structure, enabling the identification of groups of products with similar characteristics[32]. This information is then used as a guide to determine the optimal number of clusters to be applied in the K-Means analysis, ensuring more representative and accurate clustering results.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

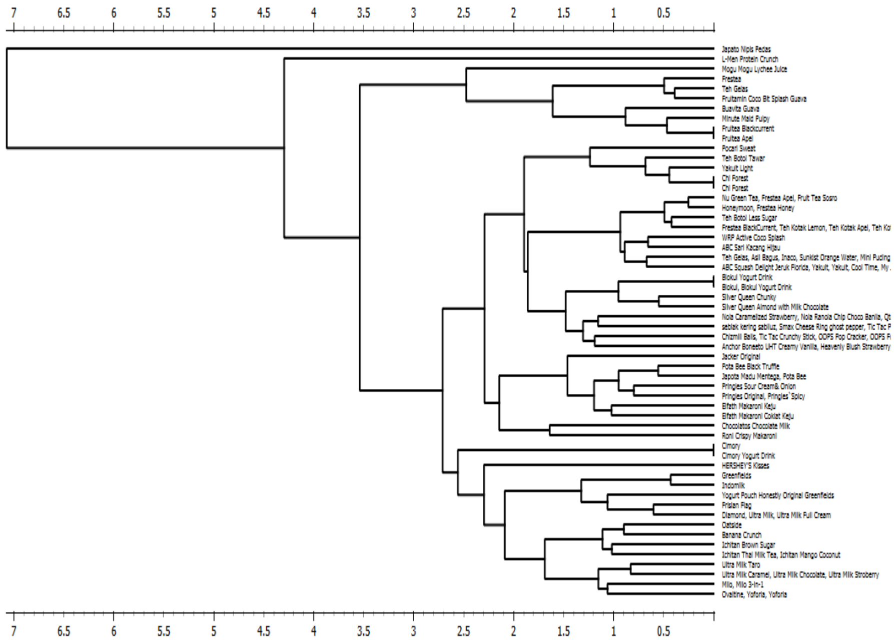
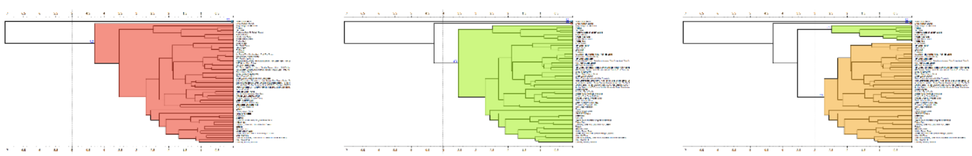


Figure 2. Dendrogram

The approach taken to determine the optimal number of clusters is to observe the 'cut-off' in the dendrogram, which is the point at which the cluster merging distance starts to increase significantly. At this stage, the dendrogram divides the data into clusters that are considered to represent the actual structure of the dataset. The result of the 'cut-off' approach on the dendrogram is shown in Figure 3.



(a) (b) (c)

Figure 3. Visualising dendrogram with added line cut-off

Suppose the vertical and horizontal lines in the dendrogram are close. In that case, this indicates that the level of uniformity between the clusters combined at that stage is relatively consistent. Therefore, the final clustering boundary is set where these lines are no longer adjacent. This ensures that the resulting clusters are more representative and conform to the data structure.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Clustering using K-Means begins with determining the number of clusters (k), which is the initial parameter for the clustering process. This study determined the number of clusters based on dendrogram analysis with cut-off lines, as shown in Figure 3. The outcomes of this visualization identified the ideal quantity of clusters as 2, 3, and 4 clusters. The clustering process begins with the initialization of a randomly selected initial centroid. The k-means results are visualized using a scatter plot. The scatter plot uses variables from sugar content and calories in this study. It is known that sugar content is identified as affecting the total calories in products that can potentially increase obesity and blood glucose levels.

Figure 4. K-Means Analysis Scatter Plot Visualisation

The clustering results were evaluated using the Silhouette Score, which is a measure of cluster validity to ensure the data distribution is logically reflected in the number of clusters formed. Based on Table 3, Cluster_2 shows the most optimal result with a Silhouette score of 0.382. This score indicates that the data in the cluster has a pretty good level of homogeneity, although not ideal. The relatively low score (0.382) suggests that there are limitations in cluster reliability. This could be due to overlapping data in different clusters or a need for more apparent separation between clusters.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Table 3. Evaluation Results

No	Cluster	Silhouette Score
1	Cluster_2	0.382
2	Cluster_3	0.350
3	Cluster_4	0.277

The results of the two clusters provide an essential insight that products such as sweetened beverages, including packaged milk, tend to fall into the high-risk category. Products in the high-risk cluster tend to have high sugar and calorie content that could potentially contribute to significantly elevated blood sugar levels, which could lead to future diabetes risk, especially if consumed in excess and over the long term. These results serve as a basis for developing consumer-friendly nutrition labelling. One approach is to apply visual indicators, such as color indicators (red for high risk, green for low risk), to provide information easily understood by the public, especially parents, when choosing healthier snacks for their children.

Table 4. Clustering Result Data

Id	Product	Sugar	Calorie	Risk
1	Ultra Milk Chocolate	1,1205	1,5880	High
4	Cimory	2,3828	3,0641	High
5	Yoforia	1,2257	1,0959	High
10	Teh Gelas	2,3828	3,0641	High
156	Pota Bee Black Truffle	-0,7729	1,5880	High
…	…	…	…	
61	Cimory Yogurt Squeeze	-0.0415	-0.6236	Low
62	Yakult	-0,2790	-1,1183	Low
76	Kinder Joy	0.1686	0.3574	Low
155	Chitato Lite	-0,7729	0,3578	Low
157	Jet Z Sweet Stick	-0,3522	0,1118	Low

However, implementing this label development presents challenges. One is the limited data diversity, as the current research dataset only covers children's snack products. In addition, implementing these labelling results requires government policy support to ensure its successful implementation and sustainability. Another challenge is the complexity of larger and more varied datasets, which may decrease the accuracy of clustering algorithms, especially if there are many interrelated variables. Therefore, the dataset needs to be expanded so that the clustering can reflect a wider diversity of globally relevant products. Support from the government's food/beverage product labelling policy is required to accelerate public understanding of the sugar content in products and create a positive long-term impact in suppressing diabetes cases, especially in children. In addition, applying modern algorithms such as DBSCAN or Gaussian Mixture Models

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

(GMM) can be a solution to overcome low silhouette scores and make the clustering results more accurate.

This merging method has been proven to improve the quality of clustering compared to using a single algorithm, as has often been applied in previous studies. This is due to the more systematic and data-driven process of selecting the number of clusters (k) using the dendrogram in Hierarchical Clustering. This approach results in an optimal number of clusters, ensuring the clustering better reflects the data structure. In contrast, in previous studies, the number of clusters was often determined randomly or based on initial assumptions without validating whether the clusters formed were optimal.

4. CONCLUSION

This study shows that clustering products based on sugar content using a combined Hierarchical Clustering and K-means method can provide significant insight into the health risks of consuming certain products. The selection of two clusters of high-risk and low-risk resulted in a balanced and informative classification, especially in identifying products that could potentially increase the risk of diabetes in children. However, the results still need improvement. The results of this study serve as a basis for developing consumer-friendly nutrition labelling. One approach can be used to apply visual indicators, such as color indicators (red for high risk, green for low risk), to increase public awareness of the sugar content in food and beverage products. In this study, the government's support of food/beverage product labelling policy is needed to create a positive long-term impact in suppressing diabetes cases, especially in children. In addition, the limited scale of the dataset is a significant challenge because it limits the generalizability of the results. For this reason, it is necessary to expand the dataset so that the clustering performed can reflect a wider diversity of products and be globally relevant. The application of modern algorithms such as DBSCAN or Gaussian Mixture Models (GMM) can be a solution to overcome the low Silhouette score and make the clustering results more accurate.

REFERENCES

[1] S.-K. dr. Mohammad Robikhul Ikhsan, MKes., "Sardjito Menyapa Acces To Diabetes Care," *RSUP Dr. Sardjito Yogyakarta*, vol. 1, pp. 1–15, 2021.
[2] E. F. Dungga and Y. Indiarti, "Risk Factors For Type 2 Diabetes Mellitus Patients At The Monano Health Center, North Gorontalo District," *Jambura Nurs. J.*, vol. 6, no. 1, pp. 40–56, 2024.
[3] S. Ellahham, "Artificial Intelligence: The Future for Diabetes Care," *Am. J. Med.*, vol. 133, no. 8, pp. 895–900, 2020, doi: 10.1016/j.amjmed.2020.03.033.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[4] S. Nurvita, "Diabetes Mellitus Tipe 1 Pada Anak Di Indonesia," *PREPOTIF J. Kesehat. Masy.*, vol. 7, no. April, pp. 635–639, 2023.
[5] J. M. Lawrence et al., "Trends in Prevalence of Type 1 and Type 2 Diabetes in Children and Adolescents in the US, 2001-2017," *JAMA*, vol. 326, no. 8, pp. 717–727, 2021, doi: 10.1001/jama.2021.11165.
[6] Y. Hasanah, "Diabetes Pada Anak," *Conf. Med. Sci. Dies Natalis Fac. Med. Univ. Sriwij.*, vol. 1, no. 1, pp. 19–27, 2019, doi: 10.32539/confmednatalisunsri.v1i1.3.
[7] A. Khoiriyah, D. M. Sholikhah, and D. N. Supriatiningrum, "Correlation Between Snack Consumption Habits with Nutritional Status of Children at Muhammadiyah Elementary School in the Gresik District," *Ghidza Media J.*, vol. 1, no. 2, pp. 107–116, 2023.
[8] A. R. Hidayat, H. Hanipah, A. Nurjanah, and R. Farizki, "Upaya untuk Mencegah Penyakit Diabetes pada Usia Dini," *J. Forum Kesehat. Media Publ. Kesehat. Ilm.*, vol. 11, no. 2, pp. 63–69, 2022, doi: 10.52263/jfk.v11i2.229.
[9] T. Ragelienė, "Do children favor snacks and dislike vegetables? Exploring children's food preferences using drawing as a projective technique. A cross-cultural study," *Appetite*, vol. 165, p. 105276, 2021, doi: 10.1016/j.appet.2021.105276.
[10] I. Darmayanti et al., "Menjadi Orang Tua Digital: Panduan Praktis Untuk Mengawasi Penggunaan Internet Anak," vol. 5, no. 4, pp. 1090–1099, 2024.
[11] A. A. Fitriana, "Pemahaman Orang Tua Mengenai Gizi Anak," *J. Pendidik. Mod.*, vol. 5, no. 3, pp. 96–101, 2020, doi: 10.37471/jpm.v5i3.92.
[12] L. Rahmawati, "Hubungan Antara Pengetahuan Gizi Dan Kebiasaan Konsumsi Jajanan Tinggi Energi Dengan Status Gizi Pada Anak Usia Sekolah Dasar Di Kecamatan Kanor Kabupaten Bojonegoro," Universitas Islam Negeri Walisongo Semarang, 2022.
[13] R. D. Rizkina et al., "Type 1 Diabetes Mellitus in Children: Diagnosis and Management," *J. Biol. Trop.*, vol. 23, no. 1, pp. 104–111, 2023, doi: 10.29303/jbt.v23i4b.5820.
[14] N. M. Almoraie, R. Saqaan, R. Alharthi, A. Alamoudi, L. Badh, and I. M. Shatwan, "Snacking patterns throughout the life span: potential implications on health," *Nutr. Res.*, vol. 91, pp. 81–94, 2021, doi: 10.1016/j.nutres.2021.05.001.
[15] R. I. Kemenkes, "Laporan Nasional Riskesdas 2018," 2019.
[16] S. D. . Pasaribu, O. Komalasari, Suheti, and R. A. . Putri, "Hubungan Peran Orang Tua Dengan Perilaku Jajan Tidak Aman Pada Anak Usia Sekolah Dasar Di RW 006 Parigi Lama Pondok Aren Tangerang Selatan," *J. Kesehat. STIKes IMC Bintaro*, vol. VI, pp. 1–8, 2023.
[17] R. Vena and M. C. Yuantari, "Kajian Literatur: Hubungan Antara Pola Makan Dengan Kejadian Diabetes Melitus," *J. Kesehat. Masy. STIKES Cendekia Utama Kudus P-ISSN*, vol. 9, pp. 255–266, 2022.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[18] A. Nur, E. Fitria, A. Zulhaida, and S. Hanum, "Hubungan Pola Konsumsi dengan Diabetes Melitus Tipe 2 pada Pasien Rawat Jalan di RSUD Dr. Fauziah Bireuen Provinsi Aceh," *Media Penelit. dan Pengemb. Kesehat.*, vol. 26, no. 3, pp. 145–150, 2017, doi: 10.22435/mpk.v26i3.4607.145-150.

[19] J. Huo et al., "Screen Time and Its Association with Vegetables, Fruits, Snacks and Sugary Sweetened Beverages Intake among Chinese Preschool Children in Changsha, Hunan Province: A Cross-Sectional Study," *Nutrients*, vol. 14, no. 19, 2022, doi: 10.3390/nu14194086.

[20] I. Darmayanti, D. Mustofa, N. Hidayati, and I. Saputri, "K - Means and Fuzzy C - Means Cluster Food Nutrients for Innovative Diabetes Risk Assessment," vol. 13, pp. 2175–2182, 2024.

[21] A. Nurhayati, R. Mardaweni, and R. M. Widiastuti, "Noodle Grouping Based on Nutritional Similarity with Hierarchical Cluster Analysis Method," *Sainteks J. Sain dan Tek.*, vol. 2, no. 2, pp. 112–125, 2023, doi: 10.37577/sainteks.v.

[22] N. Husna, F. Hanum, and M. F. Azrial, "Pengelompokkan Produk Kemasan yang Harus Dihindari Penderita Diabetes Menggunakan Algoritma K-Means Clustering," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 4, no. 1, pp. 167–174, 2019, doi: 10.30743/infotekjar.v4i1.1484.

[23] D. Selvia and A. Wahyuni, "Hubungan Pola Makan dan Obesitas Dengan Kejadian Diabetes Melitus," *J. Kesehat. Saintika Meditory*, vol. 4, no. 4657, pp. 78–84, 2022.

[24] D. Suprapti, "Hubungan Pola Makan Karbohidrat, Protein , Lemak, Dengan Diabetes Mellitus pada Lansia," *J. Borneo Cendekia*, vol. 11, no. 1, pp. 8–20, 2017.

[25] Baiq Nikum Yulisasih, H. Herman, and S. Sunardi, "K-Means Clustering Method For Customer Segmentation Based On Potential Purchases," *J. ELTIKOM*, vol. 8, no. 1, pp. 83–90, 2024, doi: 10.31961/eltikom.v8i1.1137.

[26] A. Karim, F. Nurhadi, I. K. O. Setiawan, I. A. Rizky, and R. Br. Manurung, "Pengaruh Normalisasi Data pada Klasifikasi Harga Ponsel Berdasarkan Spesifikasi Menggunakan Klasifikasi Naive Bayes dan Multinomial Logistic Regression," *J. Rekayasa Elektro Sriwij.*, vol. 4, no. 1, pp. 8–16, 2023, doi: 10.36706/jres.v4i1.59.

[27] R. G. Whendasmoro and J. Joseph, "Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 872, 2022, doi: 10.30865/jurikom.v9i4.4526.

[28] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura J. Math.*, vol. 1, no. 2, pp. 64–78, 2019, doi: 10.34312/jjom.v1i2.2354.

[29] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, pp. 1–17, 2021, doi: 10.3390/e23060759.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[30] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," in 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 2020, pp. 747–748. doi: 10.1109/DSAA49011.2020.00096.

[31] E. Yakın et al., "'In-between orthorexia' profile: the co-occurrence of pathological and healthy orthorexia among male and female non-clinical adolescents," *J. Eat. Disord.*, vol. 10, no. 1, pp. 1–14, 2022, doi: 10.1186/s40337-022-00673-z.

[32] O. Yim and K. T. Ramdeen, "Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data," 2015.