Journal of Information Systems and Informatics

Vol. 5, No. 4, December 2023 e-ISSN: **2656-4882** p-ISSN: **2656-5935**

DOI: 10.51519/journalisi.v5i4.586 Published By DRPM-UBD

Application of Clustering-Based Data Mining for the Assessment of Nutritional Status in Toddlers at Community Health Centers

Melissa Indah Fianty[1], Monika Evelin Johan[2], Azka Aulia[3], Mella Margareta Veronica[4]

Information System, Universitas Multimedia Nusantara, Tangerang, Indonesia
Email: [1]melissa.indah@umn.ac.id, [2]monika.evelin@umn.ac.id, [3]azka.aulia@student.umn.ac.id, [4]mella.margareta@student.umn.ac.id

Abstract

Nutritional status is a crucial foundation for human health and development. Global facts indicate serious challenges in ensuring adequate nutrition, and the situation is no different in Indonesia. This research collected data from the Kelapa Dua Tangerang community health center and utilized data mining techniques with the k-means clustering algorithm to delve deeper into the nutritional status of toddlers. The research findings revealed that nearly 37.3% of toddlers experience issues with abnormal height or weight, as well as poor nutritional conditions, highlighting the importance of careful and timely intervention. With regular health monitoring by community health centers and active parental involvement, actions can be taken to support the optimal growth and development of these children. The results of this research provide a strong understanding to address malnutrition issues, which will ultimately support the formation of a healthier and more promising future generation in Indonesia.

Keywords: Data mining, K-means Clustering, Malnutrition, Nutritional Status, Toddlers

1. INTRODUCTION

Malnutrition is a nutritional disorder that can have fatal consequences for toddler health [1]. Malnutrition can lead to growth impediments in children. Toddlers go through a growth and development cycle that requires a greater amount of nutrients compared to other age groups, making them more susceptible to nutritional disorders [2]. Poor nutritional status in toddlers can have a significant impact on physical, mental, and cognitive growth [3]. Toddlers suffering from malnutrition may experience a decrease in intelligence, such as a decrease in Intelligence Quotient (IQ) by up to ten percent. In Indonesia, malnutrition issues in toddlers remain a major concern that needs to be addressed. Nutrition science categorizes toddlers, including infants up to five years old, as a vulnerable group to nutritional deficiencies [4]. According to the Integrated Nutritional Status of Toddlers in 2019, it was found that in Indonesia, 27.67% of toddlers experienced stunting (short stature), 16.29% were underweight, and 7.44% of toddlers in

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

Indonesia were in a wasting condition. Creating a high-quality human resource requires optimal child growth, and to achieve optimal development in every child, monitoring and assessing nutritional status and growth trends according to standards are necessary [5].

The Indonesian government, through Community Health Centers, has been collecting data on the nutritional status of toddlers. However, the data available at community Health Center has not yet determined the nutritional status of toddlers according to government standards, making the data mapping process manual. This can undoubtedly hinder Community Health Centers in taking quick and accurate actions to address toddler nutrition problems. The concept of applying data mining with the k-means clustering method can be used to classify data into several categories of malnutrition status [6]. Data mining involves the exploration and analysis of large amounts of data to obtain useful patterns [7]. The main goal of data mining is to categorize and predict data [8]. Data mining is not specific to one type of media or data and can be applied to all types of information storage, such as scientific and medical data. Clustering techniques have been applied to extract useful patterns from medical data, usually aiming to identify patients with similar attributes and therefore belonging to the same risk group. Clustering using the k-means algorithm can be used to classify the nutritional status of toddlers [9].

This research uses four standard anthropometric indices based on weight and height parameters: Weight/Age, Height/Age, Weight/Height, and Body Mass Index/Age, as attributes for data mining using the clustering algorithm k-means [10]. The categorization of toddler nutritional status based on these four attributes refers to the Anthropometric Standards Assessment of Child Nutrition by the Minister of Health Regulation No. 2 of 2020 regarding Child Anthropometric Standards based on attributes such as Date of Birth (to determine the age of toddlers), Gender, Height, and Weight of toddlers in the data from the Kelapa Dua Community Health Center.

Child anthropometric standards are a collection of data on the size, proportions, and composition of the body as a reference for assessing children's nutritional status. Child anthropometric standards are based on weight and height parameters and consist of four indices: Weight for Age, Height for Age, Weight for Height, and Body Mass Index for Age. The classification of nutritional status based on anthropometric indices corresponds to nutritional status categories in the WHO Child Growth Standards for children aged 0 – 5 years [11].

The nutritional status categories for the Weight for Age index are divided into severely underweight, underweight, normal, and at risk of overweight [12]. The Height for Age index is categorized as severely stunted, stunted, normal, and tall. The Weight/Height and Body Mass Index/Age indices have the same nutritional status categories: malnutrition, undernutrition, good nutrition, overnutrition, and

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

obesity. All four standard anthropometric indices must be considered together to determine a child's nutritional status to identify growth problems and take appropriate actions [13].

The results of this research involve the categorization of toddler data based on their nutritional status into several clusters. The characteristics of toddler nutritional status categories within each cluster are then analyzed, and conclusions are drawn. These conclusions can be used to determine appropriate and efficient actions in addressing toddler nutritional status based on the characteristics of each cluster.

2. METHODS

The method employed to collect and process data in order to test hypotheses based on the model to be used and obtain research answers, as well as the data processing steps using the Knowledge Discovery in Databases (KDD) methodology [14].

2.1. Research Methods

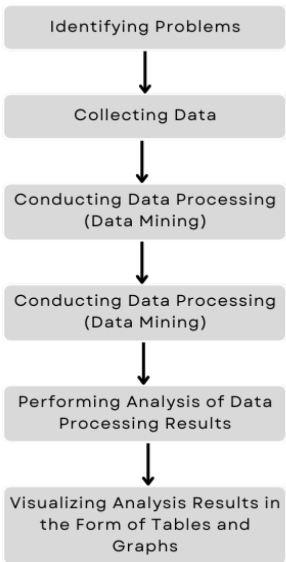

Figure 1. Research Flow

Knowledge Discovery in Databases (KDD) is a process of extracting valuable, useful, and potentially hidden information from large and complex datasets. The KDD process encompasses a series of steps or stages designed to identify patterns, relationships, or knowledge that can be used for decision-making. It begins with the selection and collection of data from various sources, followed by data preprocessing to clean and prepare the data for analysis. Data transformation may

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

be required to make it suitable for analysis. The heart of the process involves developing models using various data analysis algorithms and techniques such as clustering, classification, regression, or association to extract patterns and insights. These models are then evaluated and validated to ensure their reliability. The results of data analysis provide valuable insights and a deeper understanding of the data, which can be used to support decision-making across various.

KDD is a complex process often requiring specialized software and techniques, aimed at transforming raw data into actionable knowledge. It is highly relevant in the business and research world, enabling better decision-making and predictions based on data [15]. The explanation steps in Figure 1, in the first stage, involve identifying issues related to toddler malnutrition. The second stage involves searching for and gathering reference sources from journals, books, or other media related to the research topic and collecting data from toddlers at the Tangerang Community Health Center Clinic and gathering information related to this data through interviews with representatives from the health center staff [16].

The fourth stage is determining the appropriate research method drawn from various references so that research questions can be answered, and research objectives can be achieved using the stages of knowledge discovery in databases (KDD) for this study [17]. The data mining process for this research uses the k-means clustering algorithm and processes the data obtained using the method specified in the previous stages [18]. The fifth stage involves analyzing the data resulting from data mining, where the data will be studied and analyzed to obtain answers to research questions [19]. The sixth stage involves drawing conclusions based on the research results and data processing conducted and providing recommendations related to the research findings [20].

2.2. Data Processing Method

The research process follows the stages of Knowledge Discovery in Databases (KDD), which are as follows:

1) Data Cleaning, the first stage in data processing is data cleaning. In the data cleaning stage, the dataset's structure is improved. Afterward, unnecessary labels or information that could disrupt the data mining process are removed [21].
2) Data Integration, the next stage involves data integration, where data that is separated into multiple tables is merged into one [22].
3) Data Selection, the subsequent stage is data selection, where data or attributes needed for analysis and data mining are chosen.
4) Data Transformation, data transformation is necessary to convert data into the required format for data mining and analysis. Activities in the data transformation stage include converting data from text to numeric form.

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

Additionally, data normalization is performed in the data transformation stage to standardize the data used in data mining to have equal weights [23].

5) Data Mining, data mining is the stage where data is processed using specific methods and algorithms to uncover valuable patterns and information. In this stage, the selection of data mining methods that are appropriate for the data and research objectives is carried out [24].
6) Pattern Evaluation, the analysis, identification, and conclusion-making regarding the information and patterns discovered during data mining are conducted in the pattern evaluation stage. The conclusions drawn during this pattern evaluation stage will serve as the research's results [25].
7) Knowledge Presentation, the knowledge presentation stage is dedicated to creating visualizations and explanations of the information obtained from the pattern evaluation stage for easy comprehension [26].

3. RESULTS AND DISCUSSION

3.1 Dara Processing Process

The data processing in the stages using the Knowledge Discovery in Databases (KDD) method includes:

1) Data Cleaning, Data cleaning is performed on the Kelapa Dua Tangerang Community Health Center dataset using Google Sheets because the dataset provided by the Kelapa Dua Health Center is in .xlsx format. Additionally, Google Sheets is used for data cleaning due to its ease of use and facilitation of collaboration. Once the dataset is cleaned, it will be imported into Google Colab for data mining.
2) Data Integration, the data integration stage is conducted because the dataset obtained from the Kelapa Dua Community Health Center is separated into multiple tables, necessitating the merging of these tables into a single comprehensive table.
3) Data Selection, the attributes needed for this research are Gender, Date of Birth, Village/Subdistrict, Weight, and Height. Unnecessary attributes will be removed from the dataset.
4) Data Transformation, the data is transformed and consolidated into the format required by the data mining process. In the data transformation process, there are several things that need to be done, some of which are:
    a) Attribute Construction present in the dataset: Body Mass Index (BMI), Weight/Age, Height/Age, Weight/Height, and Body Mass Index/Age.
    b) Data normalization is applied to the attributes used for data mining, namely Weight/Age, Height/Age, Weight/Height, and Body Mass Index/Age, using min-max normalization. Data normalization is carried out while processing data in Google Colab using the Python programming language.

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

5) Data Mining, in this step, the previously prepared dataset is processed according to specified requirements. The K-means algorithm is employed as the method for data processing using Google Colab and the Python programming language. The initial step involves opening the Google Colab application by accessing the link "colab.research.google.com" through a web browser and signing in using a Gmail account. Access to Google Colab can only be achieved using a Gmail account because the project to be worked on will be saved in a Google Drive folder associated with that Gmail account. Upon successful login, the next step is to create a notebook to initiate the data mining project by clicking on the "New Notebook". After successfully uploading the dataset, the next step is to import the necessary libraries for data mining using the clustering method, namely the K-means algorithm.

```
[3] df = pd.read_excel("penimbangan_balita.xlsx", sheet_name="SATUIN")
    df.head()
```

	JK	Tgl Lahir	Umur (Bulan)	Desa/Kel	Berat	Tinggi	IMT	BB/U	TB/U	>+3SD	BB/PB	IMT/U
0	L	2023-01-31	5	PAKULONAN BARAT	4.90	56.0	15.6	1	0	-2SD s.d +1 SD	2	2
1	L	2023-01-16	6	PAKULONAN BARAT	3.70	52.0	13.7	0	0	-2SD s.d +1 SD	2	1
2	L	2023-01-23	6	KELAPA DUA	3.80	51.3	14.4	0	0	-2SD s.d +1 SD	2	1
3	L	2023-01-11	6	PAKULONAN BARAT	4.50	55.0	14.9	0	0	-2SD s.d +1 SD	2	2
4	L	2023-01-24	6	PAKULONAN BARAT	4.28	53.0	15.2	0	0	-2SD s.d +1 SD	2	2

Figure 2. Performing Input of the Dataset and Transforming It into a Data Frame

In figure 2, after importing the libraries, the dataset that was previously imported is input and converted into a data frame. Then, the 'info()' code is used to display information about the dataset to ensure data correctness and to understand the index numbers of each attribute.

```
[6] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5708 entries, 0 to 5707
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   JK            5708 non-null   object
 1   Tgl Lahir     5708 non-null   datetime64[ns]
 2   Umur (Bulan)  5708 non-null   int64
 3   Desa/Kel      5708 non-null   object
 4   Berat         5708 non-null   float64
 5   Tinggi        5708 non-null   float64
 6   IMT           5708 non-null   float64
 7   BB/U          5708 non-null   int64
 8   TB/U          5708 non-null   int64
 9   BB/PB         5708 non-null   int64
 10  IMT/U         5708 non-null   int64
dtypes: datetime64[ns](1), float64(3), int64(5), object(2)
memory usage: 490.7+ KB
```

Figure 3. Displaying Information from the Data Frame

In Figure 3, displaying information from the data frame is a crucial step in data analysis that aids analysts in understanding the structure of the dataset being

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

processed. This process involves using commands such as 'info()' in programming languages, which generate a concise summary of the dataset, including the number of rows and columns, attribute data types, and whether there are any missing values. This information is highly valuable for identifying potential issues within the dataset, determining the types of analyses that can be performed, and ensuring that the data is prepared for the subsequent stages of data analysis. In this way, displaying info from the data frame helps analysts commence their data analysis process on a solid foundation.

```
mtbs_x = df.iloc[:, 7:11]
mtbs_x.head()
```

	BB/U	TB/U	BB/PB	IMT/U
0	1	0	2	2
1	0	0	2	1
2	0	0	2	1
3	0	0	2	2
4	0	0	2	2

Figure 4. Selecting Attributes for Clustering

In figure 4, after reviewing the data frame information and confirming the absence of any issues, the next stage involves creating a new data frame that includes the attributes intended for clustering. These attributes comprise Weight/Age, Height/Age, Weight/Height, and Body Mass Index/Age.

In this step, the dataset prepared earlier is processed according to the established requirements. The K-means algorithm is used as the method for processing the data using Google Colab and the Python programming language. The elbow point on Figure 5 is observed to be at a value of 3 on the Clusters axis, thus, it can be concluded that the optimal number of clusters for clustering this data is 3. The next step is to carry out the data clustering process on the array that has been normalized in the previous stage. The data clustering process in this stage utilizes the previously determined optimal number of clusters, which is 3.

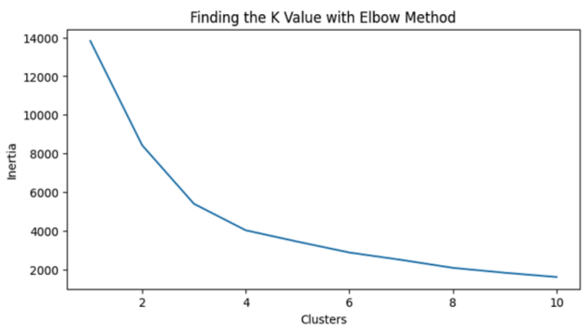

Figure 5. Elbow Method Curve Display

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

3.2 The centroid values of each cluster

The centroid values of each cluster are one of the results of the k-means clustering process.

Table 1. The centroid values of each cluster

Attribute	Cluster 0	Cluster 1	Cluster 2
Weight/Age	0,59	0,13	0,69
Height/Age	0,24	0,01	0,67
Weight/Height	0, 46	0,39	0,44
Body Mass Index (BMI)	0,49	0,38	0,45

The centroid values of each cluster can be seen in Table 1, where the data can be used to understand the characteristics of each cluster by converting them into descriptive form. The characteristics of each cluster based on centroid values can be observed in Table 2.

Table 2. Cluster Characteristics Based on Centroid Values

Attribute	Cluster 0	Cluster 1	Cluster 2
Weight/Age	Normal Weight	Severely Underweight	Normal Weight
Height/Age	Very Short	Very Short	Normal
Weight/Height	Good Nutrition	Good Nutrition	Good Nutrition
Body Mass Index (BMI)	Good Nutrition	Good Nutrition	Good Nutrition

Based on the characteristics described in Table 2, it can be concluded that Cluster 0 represents toddlers with suboptimal growth conditions because, despite having normal weight and good nutrition, their height falls into the category of very short, as indicated by Weight/Height and Body Mass Index/Age.

Cluster 1 represents a cluster of toddlers with a relatively serious condition due to severely low weight and very short height for their age. However, the values of Weight/Height and Body Mass Index/Age, indicating Good Nutrition, suggest that even though their weight is severely low and height is very short, the nutritional status of toddlers in Cluster 1 still falls into the good category.

Cluster 2 represents toddlers with well-balanced and healthy body conditions. The weight and height of toddlers in this cluster fall within the normal range for their age, and their nutritional status is categorized as Good, both based on Weight/Height and Body Mass Index/Age. Cluster 2 can be considered a group of toddlers experiencing optimal growth and nutrition conditions because their

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

centroid values indicate weight and height that are appropriate for their age, as well as good nutritional.

3.3 Analysis of Data Characteristics in Each Cluster

The clusters obtained from the data mining process will be analyzed in detail as follows:

Table 3. Percentage of Data Count Based on Nutritional Status Categories in Clusters 0, 1, and 2

Atribut	Kategori Status Gizi	Persentase Jumlah Data (%)		
		Klaster 0	Klaster 1	Klaster 2
Weight/Age	Very Underweight	0	60,7	0
	Underweight	3,6	39,3	22,2
	Normal Weight	86,8	0	77,2
	Overweight Risk	9,6	0	0,6
Height/Age	Very Short	27,6	97,1	0
	Short	72,4	2,6	0
	Normal	0	0,3	99,8
	Tall	0	0	0,2
Weight/Height	Poor Nutrition	0	0,9	0,1
	Insufficient Nutrition	0.7	7,8	2,6
	Good Nutrition	77,9	88,1	82,3
	At Risk of Overnutrition	15,4	2,3	8,1
	Overnutrition	4	0,5	3,9
	Obesity	1,9	0,5	3,1
Body Mass Index (BMI)	Poor Nutrition	0	5,7	0,1
	Insufficient Nutrition	0	6,3	3,2
	Good Nutrition	69,9	80,8	80,7
	At Risk of Overnutrition	18,2	6,5	8,5
	Overnutrition	8,4	0,7	4,5
	Obesity	3,5	0,1	3

Based on Table 3, toddlers in Cluster 0 mostly have normal weight for their age but shorter height, indicating a growth problem in height for most toddlers in this group. However, most toddlers have a good body mass index, indicating that their nutritional status is good. Additionally, some toddlers are at risk of being overweight and experiencing overnutrition. Overall, the nutritional status of toddlers in Cluster 0 can be considered good and healthy, with the majority falling

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

into the good nutritional category, and no toddlers experiencing malnutrition or undernutrition in this group.

Cluster 1, from the breakdown of nutritional status category percentages based on the four indices or attributes, consists of toddlers with very short height and significantly low weight for their age. However, based on the processed data, the nutritional status categories for toddlers in Cluster 1, as indicated by both Weight/Height and Body Mass Index/Age, are predominantly in the good nutrition category.

The data characteristics of Cluster 2, based on the breakdown of nutritional status category percentages according to the four attributes, reveal that toddlers in Cluster 2 have normal weight, but there is still a significant percentage of underweight toddlers. However, in terms of height, most toddlers have normal height, and when considering Body Mass Index/Age, the nutritional status of toddlers in Cluster 2 is predominantly categorized as normal nutrition.

4. CONCLUSION

This research employs the k-means clustering algorithm to group toddler data based on their nutritional status, utilizing four anthropometric indices (Weight/Age, Height/Age, Weight/Height, and Body Mass Index/Age) as attributes. The result is three clusters with different characteristics. In this study, the analysis involves clustering toddler data into three clusters: Cluster 0, Cluster 1, and Cluster 2. Clusters characterized by good nutritional status based on the anthropometric indices of Weight/Height or Body Mass Index/Age may not necessarily have normal height categories based on the Height/Age anthropometric index or normal weight categories based on the Weight/Age anthropometric index. This fact can be observed in the anthropometric indices of Weight/Height and Body Mass Index/Age for toddlers in Cluster 1, which exhibit characteristics of good nutritional status. However, when referring to the Weight/Age and Height/Age anthropometric indices, it becomes evident that toddlers in Cluster 1 have characteristics of very low weight and significantly short stature for their age.

The majority of toddlers in this study have adequate nutritional conditions and fall within the categories of normal weight and height. However, approximately 37.3% of toddlers belong to groups with abnormal height or weight and poor nutritional status. These findings underscore the importance of regular health monitoring for toddlers by the Kelapa Dua Tangerang Community Health Center and parents. With consistent monitoring, they can provide appropriate interventions and support the optimal development of these children.

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

REFERENCES

[1] R. K. Dewi and I. N. Budiantara, "Faktor-Faktor yang Mempengaruhi Angka Gizi Buruk di Jawa Timur dengan Pendekatan Regresi Nonparametrik Spline," *Jurnal Sains dan Seni ITS*, vol. 1, no. 1, 2012, doi: 10.12962/j23373520.v1i1.1968.

[2] S. Winiari, H. Yuliansyah, and A. Andi, "Identification of Toddlers' Nutritional Status using Data Mining Approach," *International Journal of Advanced Computer Science and Applications*, vol. 9, Jan. 2018, doi: 10.14569/IJACSA.2018.090122.

[3] W. Aprilius, "Big Data dan Perawatan Kesehatan Studi Awal Menuju Perawatan Kesehatan Masa Depan," *Jurnal ULTIMA InfoSys*, vol. 6, pp. 64–70, Jun. 2015, doi: 10.31937/si.v6i1.281.

[4] M. Bhattacharya, D. Ehrenthal, and H. Shatkay, *Identifying Growth-Patterns in Children by Applying Cluster analysis to Electronic Medical Records*. 2014. doi: 10.1109/BIBM.2014.6999183.

[5] S. Nagari and L. Inayati, "Implementation of Clustering Using K-Means Method to Determine Nutritional Status," *Jurnal Biometrika dan Kependudukan*, vol. 9, p. 62, Jun. 2020, doi: 10.20473/jbk.v9i1.2020.62-68.

[6] J. Majumder and L. Sharma, "Identifying Body Size Group Clusters from Anthropometric Body Composition Indicators," *Journal of Ecophysiology and Occupational Health*, vol. 15, pp. 1–8, Dec. 2015, doi: 10.15512/joeoh/2015/v15i3-4/121588.

[7] R. Rahim *et al.*, "Unsupervised Data Mining Technique for Clustering Library in Unsupervised Data Mining Technique for Clustering Library in Indonesia Indonesia CORE View metadata, citation and similar papers at core.ac.uk provided by DigitalCommons@University of Nebraska Unsupervised Data Mining Technique for Clustering Library in Indonesia," 2021. [Online]. Available: https://digitalcommons.unl.edu/libphilprac

[8] Frentzen, J. Wiratama, and R. S. Oetama, "KNN And Naïve Bayes Algorithms for Improving Prediction of Indonesian Film Ratings using Feature Selection Techniques," in *2023 4th International Conference on Big Data Analytics and Practices (IBDAP)*, 2023, pp. 1–6. doi: 10.1109/IBDAP58581.2023.10271977.

[9] I. Alpiana and L. Anifah, "Penerapan Metode KnA (Kombinasi K-Means dan Agglomerative Hierarchical Clustering) dengan Pendekatan Single Linkage untuk Menentukan Status Gizi pada Balita," 2019. [Online]. Available: https://journal.unesa.ac.id/index.php/inajet

[10] D. Harada, H. Asanoi, T. Noto, and J. Takagawa, "Different Pathophysiology and Outcomes of Heart Failure with Preserved Ejection Fraction Stratified by K-Means Clustering," *Front Cardiovasc Med*, vol. 7, Nov. 2020, doi: 10.3389/fcvm.2020.607760.

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

[11] I. Bombelli, I. Manipur, M. R. Guarracino, and M. B. Ferraro, "Representing ensembles of networks for fuzzy cluster analysis: a case study," *Data Min Knowl Discov*, 2023, doi: 10.1007/s10618-023-00977-x.
[12] M. A. Andila, R. Sunardi, O. 2*, and R. Artikel, "Evaluation And Solution For SAP Implementation Using Technology Acceptance Model: A Case Study In An Indonesian Food Trading Company Informasi Artikel ABSTRAK," *Jurnal Teknologi Terapan*, vol. 7, no. 4, pp. 1153–1162, 2023, doi: 10.33379/gtech.v7i4.2488.
[13] I. U. Sikder, "A variable precision rough set approach to knowledge discovery in land cover classification," *Int J Digit Earth*, vol. 9, no. 12, pp. 1206–1223, Dec. 2016, doi: 10.1080/17538947.2016.1194489.
[14] C. Nwagu, O. Kenneth, and I. Chinecherem, "Knowledge Discovery in Databases (KDD): An Overview." [Online]. Available: https://sites.google.com/site/ijcsis/
[15] M. Guarascio, G. Manco, and E. Ritacco, "Knowledge Discovery in Databases," in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds., Oxford: Academic Press, 2019, pp. 336–341. doi: https://doi.org/10.1016/B978-0-12-809633-8.20456-1.
[16] M. A. Andila, R. Sunardi, O. 2*, and R. Artikel, "Evaluation And Solution For SAP Implementation Using Technology Acceptance Model: A Case Study In An Indonesian Food Trading Company Informasi Artikel ABSTRAK," *Jurnal Teknologi Terapan*, vol. 7, no. 4, pp. 1153–1162, 2023, doi: 10.33379/gtech.v7i4.2488.
[17] P. Rai and S. Shubha, "A Survey of Clustering Techniques," *Int J Comput Appl*, vol. 7, Oct. 2010, doi: 10.5120/1326-1808.
[18] C. C. Aggarwal, "Cluster Analysis," in *Data Mining: The Textbook*, C. C. Aggarwal, Ed., Cham: Springer International Publishing, 2015, pp. 153–204. doi: 10.1007/978-3-319-14142-8_6.
[19] M. Kaur and S. Garg, "Survey on Clustering Techniques in Data Mining for Software Engineering," *International Journal of Advanced and Innovative Research*, vol. 3, pp. 238–243, May 2014.
[20] P. Bholowalia and A. Kumar, "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN," 2014.
[21] C. Saranya and G. Manikandan, "A study on normalization techniques for privacy preserving data mining," vol. 5, pp. 2701–2704, Jun. 2013.
[22] M. Chaudhry, I. Shafi, M. Mahnoor, D. Vargas, E. Thompson, and I. Ashraf, "A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective," *Symmetry (Basel)*, vol. 15, p. 1679, Aug. 2023, doi: 10.3390/sym15091679.
[23] I. Julianto, D. Kurniadi, M. Nashrulloh, and A. Mulyani, "Data Mining Clustering Food Expenditure in Indonesia," *Jurnal Teknik Informatika (Jutif)*, vol. 3, pp. 1491–1500, Dec. 2022, doi: 10.20884/1.jutif.2022.3.6.331.

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

[24] P. Dubey and R. Chakrawarti, *Analysis of Clustering-Algorithms for Efficient Data Mining*. 2023.

[25] A. Firdaus and W. I. Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi: (Sebuah Ulasan)," 2021.

[26] Institute of Electrical and Electronics Engineers., *Proceedings of National Conference on Innovations in Emerging Technology: (NCOIET '11) : Feb. 17 & 18, 2011, technology, ideas & our future*. IEEE, 2011.