ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 14 (2024) No. 7, pp. 51-71
DOI:10.33168/JSMS.2024.0703

Data Mining and Visualization for Toddler Nutrition Monitoring in Community Health Centers

Melissa Indah Fianty[1], Monika Evelin Johan[1], Fahmy Rinanda Saputri[2], Fidelia Rahmah[1]

[1]Information System, Faculty of Engineering and Informatics, Multimedia Nusantara University, Tangerang Indonesia

[2]Engineering Physics, Faculty of Engineering and Informatics, Multimedia Nusantara University, Tangerang Indonesia

melissa.indah@umn.ac.id, monika.evelin@umn.ac.id, fahmy.rinanda@umn.ac.id

Abstract. This research employed data mining methodologies to evaluate the nutritional well-being of children aged under 5 by utilizing standard anthropometric indices. Employing a knowledge discovery approach, data sourced from an Indonesian community health center was analyzed. The study focused on four key anthropometric indices—weight/age, height/age, weight/height, and BMI/age—as defining attributes. Through the application of k-means clustering, the toddlers were segmented into distinct groups exhibiting similar nutritional traits. The analysis of these clusters delineated three categories, highlighting groups with satisfactory nutritional indicators, those with low weight/height ratios, and others displaying inadequate height/age metrics. While a significant proportion of toddlers showcased acceptable nutritional profiles across these clusters, a substantial portion demonstrated irregular height or weight metrics. The utilization of Tableau visualizations yielded actionable insights, showcasing the practical utility of data mining in assessing nutritional status, potentially guiding health center interventions. However, to consolidate these findings, future research endeavors with larger sample sizes are imperative.

Keywords: clustering, data mining, k-means, nutritional status of toddlers

1. Introduction

In Indonesia, the issue of nutrition, particularly concerning toddlers, remains a top priority to be addressed (Meher et al., 2023). The field of nutrition categorizes toddlers, including Toddlers up to five years old, as a group vulnerable to malnutrition (Novyriana et al., 2022). Based on the findings from the Ministry of Health's Integrated Nutritional Status Study of Toddlers in 2022, it is revealed that 21.6% of toddlers in Indonesia suffer from stunting (short stature), 17.1% experience underweight, and 7.7% of Indonesian toddlers are in a wasting or thin condition (Kebijakan et al., 2022). Nutrition in toddlers, especially those up to five years old, plays a crucial role as it significantly influences overall health and development. Nutrition issues in toddlers take priority due to their profound and lasting impact on health and quality of life. Children within this age range are highly vulnerable to problems such as stunting, underweight, and other forms of malnutrition. These nutritional issues not only affect physical growth but also impact brain development, cognitive abilities, and academic performance in school. Moreover, the correlation between malnutrition in toddlers and high mortality rates underscores the urgency of addressing this issue seriously. Therefore, toddler nutrition is a priority due to its extensive impact on children's health and development, as well as the potential long-term effects on future life.

The extensive impact of stunting and poor nutrition in toddlers is concerning. Children affected by stunting face heightened risks, including infections, brain developmental disorders, and limitations in cognitive abilities, impacting their learning in school. These limitations in physical and mental aspects also extend to adulthood, affecting productivity. Moreover, the close link between poor nutrition and mortality in children under five emphasizes the vulnerability of undernourished children to diseases leading to premature death. In Indonesia, data on poor nutrition in toddlers not only highlights its direct impact on children's health but also underscores its significant long-term implications on growth, development, and future productivity. To ensure the development of high-quality human resources, monitoring and assessing children's nutritional status and growth trends according to established standards become essential (Yanhui et al., 2019).

The Indonesian government, through its Community Health Center, has been collecting data on the nutritional status of toddlers (Fauziah et al., 2021). However, the data within these health centers still do not determine the nutritional status of toddlers in accordance with the standards set by the government, leading to a manual data mapping process (Pratiwi et al., 2022). This, of course, can make it challenging for Community Health Center to make quick and accurate decisions in addressing toddler nutrition issues (M. Masturina et al., 2023). The concept of data mining using the k-means clustering method can be employed to classify data into several categories of poor nutritional status (Amelia et al., 2022). Data mining involves the exploration and analysis of large volumes of data to extract valuable patterns. The primary objective of data mining is to categorize and predict data (Chaudhry et al., 2023). Data mining is not specific to any particular type of media or data and can be applied to all types of information storage, such as scientific and medical data (Tran et al., 2023). Clustering techniques have been employed to extract useful patterns from medical data, typically aiming to identify patients with similar attributes and therefore categorize them into the same risk group (Dol & Jawandhiya, 2023). Clustering using the k-means algorithm can be utilized to group the nutritional status of toddlers (Iuliano et al., 2023).

This research utilizes four standard anthropometric indices based on weight and height parameters, namely Body/Age, Height/Age, Body Weight/Height, and Body Mass Index/Age, as attributes for the data mining process using the k-means clustering algorithm (Rae et al., 2023). The use of anthropometric indices in assessing children's nutritional status is crucial as it provides a globally recognized standard framework. Through indices such as BMI, Body Weight/Height, and Height/Age, this research offers a clear picture of children's growth and nutritional status. These relatively easy measurements also offer critical information for healthcare professionals to identify nutritional issues, enabling appropriate intervention adjustments and monitoring responses to nutritional changes. In research, the use of anthropometric indices provides a strong foundation for analyzing data related to children's nutritional status, aiding in identifying patterns and risk factors associated with nutritional issues, and supporting recommendations or policy changes to improve children's health.

The categorization of nutritional status in toddlers based on these four attributes refers to the Table of Anthropometric Standards for Assessing Child Nutritional Status from the Ministry of Health Regulation No. 2 of 2020 regarding Child Anthropometric Standards (Menteri Kesehatan Republik Indonesia, 2020). These attributes are derived from the child's Date of Birth (to determine their age), Gender, Height, and Weight data available at the Community Health Center (Perumal et al., 2020). Child anthropometric standards constitute a collection of data concerning body measurements, proportions, and composition, serving as a reference for assessing a child's nutritional status (Nagari & Inayati, 2020). These standards are based on weight and height parameters, comprising four indices: Body/Age, Height/Age, Body Weight/Height, and Body Mass Index/Age (Limauro et al., 2023). The classification of nutritional status based on anthropometric indices aligns with the nutritional status categories outlined in the WHO Child Growth Standards for children aged 0 to 5 years (Virgia & Widiyawati, 2023).

The nutritional status categories for the Weight/Age index are differentiated into severely underweight, underweight, normal, and overweight (S. Masturina et al., 2023). The Height/Age index is categorized as severely stunted, stunted, normal, and tall (Dwi Wulandari & Laksono, 2023). Both the Weight/Height and Body Mass Index/Age indices share the same nutritional status categories, which include malnutrition, undernutrition, good nutrition, overnutrition, and obesity (Rahmawati & Retnaningrum, 2022). Monitoring all four standard anthropometric indices simultaneously is essential for determining a child's nutritional status, enabling the identification of growth-related issues that require appropriate intervention (Bhattacharya et al., 2014).

Monitoring and mapping the nutritional status of toddlers using Tableau tools has the primary goal of enhancing understanding, making better decisions, and enabling more effective interventions concerning toddler nutrition (Sharma & Deshmukh, 2020). Data monitoring can also promptly identify nutritional issues in toddlers, such as malnutrition or obesity, allowing for timely and appropriate actions to be taken (Zikri et al., 2017) (Dubey & Chakrawarti, 2023). Through data visualization with Tableau, health centers can discern trends and patterns that might not be readily apparent in raw data, thereby facilitating improved decision-making regarding health interventions and programs (Purich et al., 2023). Community Health Centers are crucial in gathering, managing, and storing essential health information concerning individuals, particularly children, encompassing data relevant to toddler nutrition. This includes gathering anthropometric measurements like height, weight, and age, alongside demographic particulars such as gender and address. This comprehensive data collection assists in evaluating and overseeing toddlers' nutritional well-being, serving as a cornerstone for strategies or policies aimed at enhancing child nutrition. The collected data aligns entirely with governmental standards, ensuring accurate and efficient utilization in addressing toddler nutrition issues. However, the collected data presents several issues that need addressing. There are empty rows and columns separating values and attributes in the dataset, unnecessary sentences and labels that disrupt the dataset's attribute structure for data mining analysis, and three instances of data concerning toddlers above 60 months old. This necessitates data processing methods following the Knowledge Discovery in Databases (KDD) framework. This also enhances efficiency and accuracy in assessing the nutritional status of toddlers (Kruger et al., 2023). Data mapping aids in planning more precise nutritional programs, including the provision of dietary supplements, nutritional education initiatives, or suitable medical interventions (Nurzaman & Jayadi, 2022). Furthermore, ongoing monitoring and mapping help measure the long-term impact of nutritional programs and foster effective communication with all stakeholders (Sholahuddin et al., 2022) (Lin et al., 2023).

The research problem centers on utilizing the k-means clustering algorithm to categorize toddler data by nutritional status, utilizing four anthropometric indices. It raises key inquiries: whether clusters indicating good nutritional categories in Weight/Height or BMI/Age indices correspond consistently to normal height categories in the Height/Age index and normal weight categories in the Weight/Age index. Moreover, it explores if clusters demonstrating good nutritional status and normal weight and height percentages surpass other clusters. The primary research objectives are to cluster toddler data from Community Health Centers and analyze cluster characteristics derived from the k-means method, generating visualizations and conclusions. The research outcome aims to identify toddler clusters based on their nutritional status at Community Health Centers, providing deeper insights into each cluster's characteristics, especially regarding nutrition. Ultimately, it seeks to help these centers understand clustering outcomes to make more informed decisions about addressing toddlers' nutritional statuses that fall short of good and normal standards.

The hypotheses propose that the k-means method can categorize toddler data using anthropometric indices for nutritional status. However, clusters with good nutrition may not always have normal height and weight. Additionally, it's expected that over 60% of data in clusters display normal weight, height, and good nutrition. The data collection involves interviews with Community Health Centers staff to acquire necessary toddler datasets, seeking specific parameters and requirements. These interviews provide demographic details like age, gender, height, weight, and relevant indices. Data processing follows the Knowledge Discovery in Databases framework, encompassing data cleaning, transformation, and analysis to prepare for the k-means algorithm. The research outcome categorizes toddler data into clusters via k-means clustering, offering nuanced insights into various nutritional attributes among toddlers. Analyzing cluster characteristics provides a comprehensive view of toddler nutrition based on anthropometric indices. Tableau-generated visualizations aid in interpreting complex patterns and formulating tailored interventions. The conclusions drawn provide health centers actionable insights to implement targeted strategies for improving toddler nutrition and health.

2. Related Work

In the first research study, titled "Identification of Toddlers’ Nutritional Status using Data Mining Approach," the focus was on utilizing the k-means clustering method to classify the nutritional status of toddlers (Winiari et al., 2018). However, the approach was based on variables such as age, height, and weight without taking into consideration standard anthropometric indices like Body/Age, Height/Age, Body Weight/Height, and Body Mass Index/Age. This indicates that the research was primarily interested in non-anthropometric factors when assessing the nutritional status of toddlers. In the second study, "Implementation of Clustering Using K-Means Method to Determine Nutritional Status," the k-means clustering method was also employed. However, the difference lies in the fact that this research incorporated attributes of weight and age as factors used to analyze the nutritional status of toddlers. This represents a more comprehensive approach that includes anthropometric indices and other factors in the nutritional status analysis.

On the other hand, the third research, "Identifying Body Size Group Clusters from Anthropometric Body Composition Indicators," centered around the use of the k-means clustering method and referred to standard anthropometric indices like Body/Age, Height/Age, Body Weight/Height, and Body Mass Index/Age to classify the nutritional status of toddlers (Majumder & Sharma, 2015). However, the primary aim of this research was to identify human body size groups based on body anthropometrics. Therefore, this study placed more emphasis on the physical characteristics and growth of toddlers. In the fourth research, "Application of Combination of K-Means and Agglomerative Hierarchical Clustering Method to Determine Nutritional Status in Toddlers," combined the k-means clustering method with agglomerative hierarchical clustering to determine the nutritional status of toddlers (Alpiana & Anifah, 2019). The attributes used included Body/Age, Height/Age, Body Weight/Height, and Body Mass Index/Age. This demonstrated a more complex and diverse approach to nutritional status analysis by considering more than one clustering technique. The fifth study, "Clustering as Data Mining Technique in Risk Factors Analysis of Diabetes, Hypertension, and Obesity," utilized the k-means clustering method to analyze the nutritional status of toddlers (Ahamad et al., 2018). However, the difference lies in the fact that this research was more focused on the analysis of risk factors that could contribute to diseases like diabetes, hypertension, and obesity, thus being more oriented toward further health issues that can be influenced by nutritional status. Finally, the sixth research, "Different Pathophysiology and Outcomes of Heart Failure with Preserved Ejection Fraction Stratified by K-Means Clustering, Frontiers in Cardiovascular Medicine," also employed the k-means clustering method. However, this research had a stronger focus on heart failure with preserved ejection fraction and its associated pathophysiology, rather than the nutritional status of toddlers (Harada et al., 2020).

Clustering techniques, particularly k-means clustering, serve as valuable tools for analyzing various aspects of health sciences, ranging from the nutritional status of toddlers to disease risk and other advanced health issues (Bholowalia & Kumar, 2014). This reflects the flexibility and diversity of data mining methods in the field of health (Flores et al., 2023). Data mining methods, particularly k-means clustering, are employed in determining the nutritional status of toddlers, categorizing their nutritional status (Krishnamoorthy & Karthikeyan, 2022). This underscores the significance of analytical and computational approaches in assessing and understanding the nutritional status of children (Julianto et al., 2022). The use of different attributes indicates variations in the factors considered for analysis. Some studies use age and weight attributes, while others incorporate anthropometric indices such as Body/Age, Height/Age, Body Weight/Height, and Body Mass Index/Age (Macuácua et al., 2023). This demonstrates the diversity of approaches in determining nutritional status, ranging from identifying body size groups to analyzing specific disease risk factors. It highlights the flexibility of data mining within the healthcare context. The importance of data integration and statistics in health emphasizes the value of data mining and statistical analysis in gaining deeper insights into health issues, including the nutritional status of toddlers (Han et al., 2023). This aids health practitioners in uncovering insights that may not be evident through traditional methods. The importance of health-focused data mining is evident in the distinct focus of each research, both in terms of the attributes considered and the objectives of the analysis, underscoring the importance of a focused and contextual approach in the field of health.

The primary hypothesis is rooted in the existing problem, suggesting that employing clustering methods, specifically the k-means algorithm, can effectively categorize toddler data based on the four standard anthropometric indices linked to nutritional status. This hypothesis emerges from the recognition of the algorithm's potential in revealing significant patterns within the complex dataset of toddler nutrition. It's based on the understanding that these anthropometric indices provide a robust framework for delineating nutritional categories, allowing for a systematic and structured assessment of toddlers' nutritional status. The second hypothesis stems from the realization that clusters representing good nutritional status don't always guarantee normal weight and height among toddlers. This hypothesis delves into the intricate relationship between nutritional status and physical measurements, emphasizing the complexity and multifaceted nature of this correlation. It underscores the limitations of solely relying on weight and height as individual indicators of a toddler's nutritional well-being. This hypothesis aims to challenge conventional assumptions, suggesting that factors beyond weight and height might contribute significantly to determining a toddler's nutritional status. Lastly, the hypothesis aiming to validate that more than 60% of the data will align with clusters characterized by normal weight, normal height, and good nutritional status is driven by the intention to identify distinct and healthy groups within the toddler dataset. By setting this threshold, the hypothesis seeks to affirm or contradict the assumption that a considerable portion of the data would fit into clusters signifying optimal nutritional health. Confirming this hypothesis could offer pivotal insights into managing toddler nutritional status by defining normative criteria and refining current assessment practices.

The current research demonstrates a meticulous understanding of toddler nutrition. The primary hypotheses suggest the effectiveness of employing k-means in clustering toddlers based on standard anthropometric indices, hinting at the method's potential in uncovering crucial patterns within nutritional data. Subsequent hypotheses challenge assumptions, highlighting the intricate relationship between nutritional status and body measurements. Specifically, these hypotheses emphasize the limitations of using weight and height as sole indicators of nutritional well-being, advocating for a more comprehensive assessment.

The formulated approach and hypotheses reflect a contextual and nuanced perspective crucial within the healthcare domain. It underscores the dynamism and necessity for comprehensive, multidimensional analyses to effectively comprehend and manage toddler nutritional status. For the dashboard development, a monitoring tool will be utilized to collect data from the server database, followed by data extraction into the monitoring tool for continuous 24-hour tracking. After the monitoring process, the data will undergo the Knowledge Discovery in Databases process, involving data cleaning, integration, selection, transformation, mining, and pattern evaluation. Subsequently, this dataset will undergo further analysis to draw conclusions based on the mining results and will be prepared for visualization using visualization tools, presenting the data through graphs, tables, and other formats for easier interpretation. Once visualized, the data will be ready for analysis to achieve long-term strategic goals effectively in the future.

3. Method

In this research, the chosen method of data collection involves conducting interviews to gather information by asking questions to the respondents. The questions provided to the respondents are crucial to capture understanding, opinions, and knowledge about a particular event or fact. These interviews are conducted through discussions with nutrition experts from the Community Health Center. The purpose of these interviews is to gather necessary information required for the research. During these interviews, the Community Health Center also provided data on toddlers from Community Health Center, which will be used for subsequent data mining. The toddler data from Community Health Center covers the period from February 2021 to February 2023, with an average of 2911 toddler data entries per year, stored in Microsoft Excel format. These datasets encompass attributes such as gender, date of birth, village/urban sub-district, weight, and height.

The sample size of toddler data from Community Health Center during this period averages around 2911 entries per year. This information is crucial as it provides an overview of the sample size utilized in this research, reflecting the relevant data coverage for analysis. The demographic information within this dataset includes fundamental attributes like gender, date of birth, and location of the village/urban sub-district where the toddlers reside. These details are key to interpreting the research findings, depicting the data source, time frame, and attributes considered in the analysis. With a substantial sample size and a significant time span, this research has the potential to offer a robust representation in analyzing the nutritional status of toddlers.

The process of data cleaning was executed using Microsoft Excel on the dataset procured from the Community Health Center. This choice of software was primarily dictated by the dataset's file format, .xlsx, ensuring ease of access and manipulation. The user-friendly interface of Microsoft Excel facilitated the cleansing process, aligning the dataset for seamless importation into Google Colab for subsequent data mining activities. The dataset from the Community Health Center exhibited several issues that required meticulous handling. These included the presence of empty rows and columns, disrupting the dataset's structural coherence. The resolution involved a systematic removal of these redundant rows and columns, preserving the dataset's integrity. Additionally, non-essential descriptions beyond attribute-related content were eliminated to streamline the dataset for focused analysis. Furthermore, to align the dataset with the study's scope, data related to children aged over 60 months was excluded, ensuring the dataset remained relevant to the research objectives. This systematic approach aimed to refine the dataset for subsequent analysis, enabling a more focused exploration of the pertinent attributes. The phase of data integration emerged as vital due to the dataset's fragmented nature, divided into multiple tables. Combining these disparate tables into a unified format was imperative to ensure a coherent dataset for comprehensive analysis. For data selection, a focused approach was adopted, emphasizing attributes such as ID, Date of Birth, Gender, Address, Height, and Weight. Extraneous attributes were pruned to streamline the dataset, retaining only the relevant information pertinent to the research objectives.

In this research, both quantitative and qualitative analyses are integrated to offer a comprehensive overview of the nutritional status of toddlers. Quantitative analysis involves the utilization of numerical data, such as weight and height, which are statistically analyzed to classify the nutritional status of children into categories. This is achieved through methods like clustering, grouping toddler data based on growth patterns or nutritional status. Conversely, qualitative analysis involves interviews or discussions with nutrition experts, healthcare workers, or parents of toddlers. This aims to acquire a deeper understanding of social, cultural, or dietary factors that might influence the nutritional status of toddlers. Insights gained from these interviews provide essential context to comprehend the causes or factors associated with the toddlers' nutritional status. By amalgamating these two types of analyses, this research aims to provide a more comprehensive understanding of toddlers' nutritional status. While quantitative analysis offers a robust framework of data, qualitative analysis provides context necessary to comprehend the factors that may influence the quantitative results.

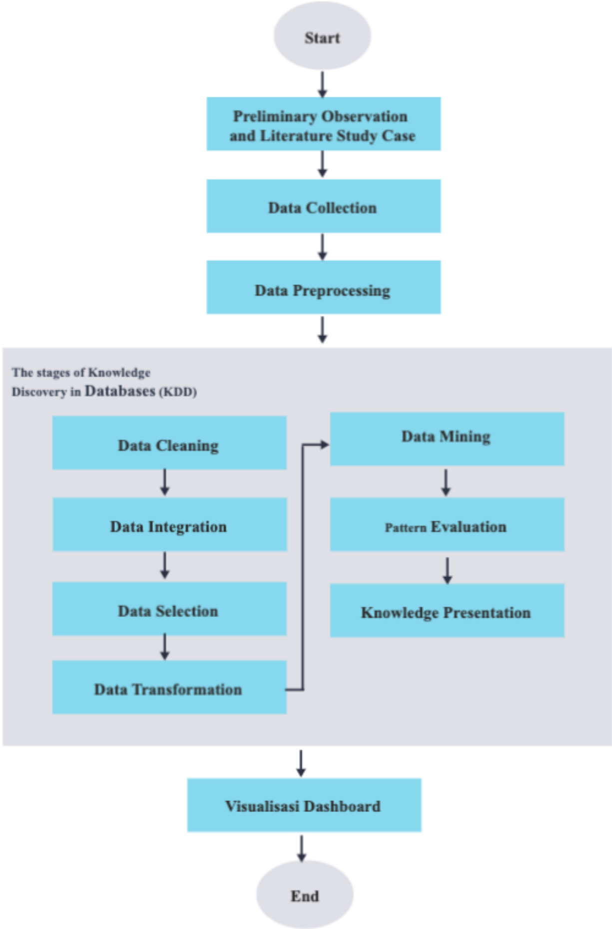

Fig. 1: Research Methodology

The data processing method follows the stages of Knowledge Discovery in Databases, which consist of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge presentation (Nwagu et al., n.d.). Data Cleaning is the initial stage where the database is cleaned by eliminating irrelevant data, rectifying invalid or mistyped entries, and removing unnecessary components. The dataset provided by the Community Health Center contains labels and information that are not required for the data mining process and thus must be eliminated. Furthermore, the row and column structure in the health center's dataset is disorganized and needs to be organized. The purpose of data cleaning is to ensure that the dataset used contains data that aligns with the research needs and to prevent a decrease in the quality of the data mining results. Data Integration is the subsequent stage involving the merging of separate data provided by the health center into a unified table. Data Selection is the phase where attribute selection is conducted to choose the necessary attributes for data mining and the subsequent analysis of data mining results. In this data selection process, irrelevant attributes are identified and removed from the dataset before applying it to the data mining algorithm to avoid affecting patterns and outcomes. The list of attributes in the dataset includes No, Date of Birth, Gender, Address, Height, and Weight.

The Data Transformation stage is a critical step in the Knowledge Discovery in Databases process (Duque et al., 2023). This process is essential for converting data into a suitable format for data mining analysis, such as converting text data into numerical data. Additionally, in the Data Transformation stage, data normalization is performed to ensure that all attributes have a balanced weight in the analysis process.

The selection of specific data mining techniques, such as the k-means clustering algorithm and min-max normalization, is typically based on their alignment with research goals, effectiveness in handling dataset attributes, and proven reliability in similar studies within the domain of toddler nutritional status assessment. K-means, known for segmenting data into clusters based on similarities, is chosen for grouping toddlers by nutritional status using anthropometric indices. Min-max normalization, scaling attribute values within a specified range while preserving data distribution, was likely selected for its demonstrated lower classification errors in similar studies. These techniques were likely chosen for their compatibility and effectiveness in achieving the study's objectives. Data Transformation strategies, including attribute construction, normalization, and concept hierarchy generation for nominal data, customized to the data's characteristics. The attribute construction strategy allows for modifying attributes to align with the requirements of data mining analysis, including adding attributes that may be needed in the data mining process. Data normalization, the next strategy, is crucial because algorithms relying on distance measurements, like k-means, can introduce bias if attribute value ranges vary significantly. K-means is selected due to its effectiveness in partitioning data into clusters based on similarities. It's suitable when the number of clusters is unspecified and aims to minimize variance within clusters. K-means' simplicity and efficiency are advantageous, particularly with large datasets. However, it is sensitive to the scales of attributes. Therefore, ensuring fair attribute influence becomes crucial, leading to the next step. Data normalization is vital as it equalizes the scales of attributes, preventing certain attributes from disproportionately influencing clustering results. Algorithms like k-means, reliant on distance calculations, can introduce bias if attribute values have widely varying scales. By normalizing the data, each attribute is transformed to fit within a consistent range, ensuring fair and unbiased clustering outcomes.

This research employs the min-max normalization method, which has been shown to produce fewer classification errors compared to other methods like z-score normalization (Pambudi et al., 2023). The choice of min-max normalization is specifically made due to its ability to scale attribute values within a predetermined range, often between 0 and 1. This technique preserves the data's distribution while standardizing attribute scales. In this research, it's adopted because it has demonstrated a lower incidence of classification errors compared to alternative normalization methods, such as z-score normalization. This lower susceptibility to errors is crucial, especially when attributes have significantly varying ranges, as it ensures more accurate clustering results. These strategies showcase a meticulous approach to handling data characteristics. The attribute construction aligns data with mining analysis needs, normalization mitigates bias from varying attribute scales, and min-max normalization specifically balances attribute influences while minimizing errors in clustering, enhancing the reliability and accuracy of the overall analysis.

Furthermore, the strategy of concept hierarchy generation for nominal data is utilized to transform specific nominal data into more general data, enabling improved data clustering outcomes. For example, initially detailed attributes like street addresses and house numbers are transformed into more generalized information, such as the neighborhood. This helps make the address attributes more uniform. Following the Data Transformation process, the Data Mining stage is conducted to discover patterns and new knowledge within the data using clustering methods with the k-means algorithm. The Data Mining process is executed within a web application called Google Colab, which supports the Python programming language. The results of the Data Mining process go through the Pattern Evaluation stage, where patterns and findings from the data are analyzed, leading to conclusions. This stage yields valuable insights from the data analysis, which is the primary goal of this research. Finally, the Knowledge Presentation stage involves presenting the analysis results in descriptive and visual forms, including diagrams and tables. This facilitates better communication and more effective decision-making in toddler nutrition management. Tools such as Tableau are employed in this stage for visualizing toddler nutrition data, enhancing comprehension and sharing the analysis results.

Through the entire Knowledge Discovery in Databases process, this research aims to test the hypotheses that have been put forward, particularly concerning the use of the k-means clustering method in grouping toddler nutrition data based on standard anthropometric indices. The anthropometric indices in this study are calculated using standard formulas related to toddler growth and nutritional status. Some of the anthropometric indices used include:

1. Body/Age: Measures a toddler's weight concerning their age. In this case, the toddler's weight is measured against their age, which is then compared to growth standards based on their age.
2. Height/Age: Measures a toddler's height concerning their age. It is obtained by comparing a toddler's height to the normal range of height at a specific age.
3. Body Weight/Height: Indicates the ratio between a toddler's weight and height. It helps identify the distribution of a toddler's weight relative to their height.
4. Body Mass Index/Age: Body Mass Index is calculated by dividing weight in kilograms by the square of height in meters. Here, BMI is used as a significant indicator in evaluating overweight or undernutrition in toddlers.

All these anthropometric indices are computed based on the data obtained from each toddler in the dataset. Once these values are calculated, they are compared to growth standards and nutritional status categories to determine whether a toddler falls into a category of normal growth, malnutrition, or undernutrition. These hypotheses also encompass the consideration that clusters with good nutritional status do not always have normal weight and height, and that a majority of the data will be included in clusters characterized by normal weight, normal height, and good nutritional status.

Understanding the potential constraints or factors impacting the study's outcomes is crucial when discussing limitations or assumptions in methodology. These factors include sample size affecting generalizability, data quality influencing accuracy, algorithm limitations in real-world scenarios, assumptions about variable relationships, incomplete temporal coverage, and biases in data collection or analysis. Acknowledging these constraints is essential to interpret the study's conclusions within their specified boundaries.

4. Results and Discussion

Microsoft Excel was used to clean the Community Health Center. The dataset supplied by Community Health Center was in the.xlsx file format, which makes it simple to open and work with in Microsoft Excel, which is why that program was selected. Additionally, Microsoft Excel was used for data cleaning due to its user-friendly interface, and the cleaned dataset could be directly imported into Google Colab for data mining. The dataset from Community Health Center was divided into several tables, which required the combining of these tables into a single, unified table. This is why data integration procedures were performed.

The required attributes for this research are as follows: No (Number), Date of Birth, Gender, Address, Height, and Weight. Unnecessary attributes will be removed from the dataset. The Body Mass Index (BMI) attribute is calculated using the BMI formula, which is the weight divided by the square of the height in units of kg/m2. Meanwhile, attributes such as Body/Age, Height/Age, Body Weight/Height, and Body Mass Index/Age are obtained by matching the age, gender, height, weight, and BMI values from each data point in the dataset with the nutritional status categories referring to the Anthropometric Standards for Assessing Child Nutritional Status and the Category and Thresholds for Child Nutritional Status from the Ministry of Health Regulation No. 2 of 2020, as detailed in the appendix. Data normalization is performed on the attributes that will be used for data mining, which are Body/Age, Height/Age, Body Weight/Height, and Body Mass Index/Age using min-max normalization. Data normalization is carried out during data processing in Google Colab using the Python programming language. The hierarchy of the "Address" attribute, which initially contains detailed information such as street, village, district, and so on, is transformed and standardized to only include the village address information of the toddlers.

4.1. Data Mining

The dataset is adjusted according to the established requirements. The K-means algorithm is employed using Google Colab and the Python programming language. The process begins by opening Google

Colab via the link "colab.research.google.com" using a Gmail account. This is required because the project will be saved in Google Drive associated with the Gmail account. After uploading the dataset, the next steps involve importing the necessary libraries for data analysis with clustering methods, specifically the K-means algorithm. Once the libraries are imported, the uploaded dataset is converted into a data frame. Information about the dataset is displayed using the "info ()" command to ensure data consistency and to identify attribute indices. After checking the data frame information and ensuring there are no issues, the next step is to create a new data frame containing the attributes required for clustering, such as Weight/Age, Height/Age, Weight/Height, and Body Mass Index/Age. The data frame containing the required attributes is then converted into an array. This array is then normalized using the min-max normalization method (min-max scaler) to ensure uniform weighting of data values, avoiding the dominance of attributes with larger weights that could influence clustering results. After normalization, the prepared data array is used to determine the optimal number of clusters.

```
clusters = []
for i in range (1, 11):
    km = KMeans(n_clusters=i).fit(mtbs_x)
    clusters.append(km.inertia_)

fig, ax = plt.subplots(figsize=(8, 4))
sns.lineplot(x=list(range(1, 11)), y=clusters, ax=ax)
ax.set_title('Mencari Nilai K dengaqn Elbow Method')
ax.set_xlabel('Clusters')
ax.set_ylabel('Inertia')
```

Fig. 2: Optimal Number of Clusters with Elbow Method

Fig 2. The Elbow method is used to find the cluster value located at the elbow of the curve, indicating a significant drop in the previous cluster value and the stability of the next cluster value. This is important for obtaining quality clustering results.

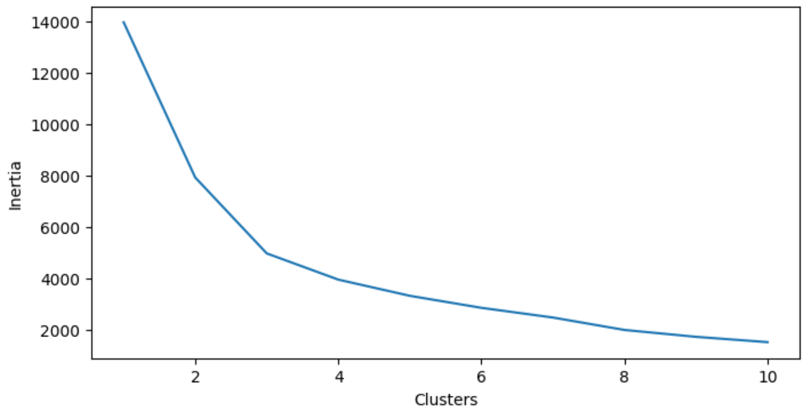

Fig. 3: Elbow Method Curve

The curve in Figure 3 is seen to be at a value of 3 on the Clusters axis, so it can be concluded that the optimal number of clusters for clustering this data is 3. The next step is to perform the data clustering process on the array that has been previously normalized. The data clustering process at this stage uses the optimal number of clusters determined in the previous step, which is 3. The next step is to add the clustering results to the initial data frame. After clustering the data, the results are added to the initial data frame as a column or attribute labeled "Cluster." The content of the Cluster attribute is the cluster code for the data, indicating whether the data belongs to Cluster 0, Cluster 1, or Cluster 2. After performing the clustering process, the next step is to display the centroid values for each cluster. These centroid values can be used to understand the characteristics of each cluster based on the centroid values, allowing for further analysis. The next step is to add the clustering results to the initial data frame. After clustering the data, the results are added to the initial data frame as a column or attribute labeled "Cluster." The content of the Cluster attribute is the cluster code for the data, indicating whether the data belongs to Cluster 0, Cluster 1, or Cluster 2.

4.2. K-means clustering

The centroid values of each cluster are one of the results of the k-means clustering process. The centroid values of each cluster in Table 1. are used to determine the characteristics of each cluster by transforming them into a descriptive form. In Table 2, the characteristics of each cluster are based on the centroid values.

Table 1: Centroid Value 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
BMI/Age	0.49	0.38	0.45

Table 1. displays the characteristics of each cluster in clustering analysis. Centroid values represent the center of each cluster and are utilized to identify the features of these clusters. Data is grouped into clusters based on the observed attributes, which can assist in decision-making and a deeper understanding of patterns within the data.

Table 2: Characteristics of Each Cluster Based on Centroid Value

Attribute	Cluster 0	Cluster 1	Cluster 2
Weight/Age	Normal Weight	Very Underweight	Normal Weight
Height/Age	Very Short	Very Short	Normal
Weight/Height	Good Nutrition	Good Nutrition	Good Nutrition
BMI/Age	Good Nutrition	Good Nutrition	Good Nutrition

Based on the characteristics of the centroid values in Table 2, it can be concluded that Cluster 0 is a cluster of toddlers with suboptimal growth conditions because, despite having normal weight and good nutrition, their height falls into the very short category, as referenced by the Weight for Height and Body Mass Index for Age. Cluster 1 represents a cluster of toddlers with a rather serious condition as they have very low weight and very short height for their age. Nevertheless, the values for Weight for Height and Body Mass Index for Age, indicating Good Nutrition, suggest that even though their weight and height are significantly low, the nutritional status of toddlers in Cluster 1 is still considered good. Cluster 2 comprises toddlers with well-balanced and healthy body conditions. The weight and height of these toddlers fall within the normal range for their age, and their nutritional status is categorized as Good, both based on Weight for Height and Body Mass Index for Age. Cluster 2 can be considered a group of toddlers experiencing optimal growth and nutritional conditions because their centroid values indicate that their weight and height are appropriate for their age, and they have good nutritional status.

4.3. Total data from each cluster

The distribution of data within the clusters resulting from clustering analysis is organized based on their quantities, and this explanation provides details about the clusters with the largest, second largest, and smallest data quantities. Furthermore, the percentage of data included in each cluster is mentioned to provide a more comprehensive picture of the data proportion in the clustering analysis.

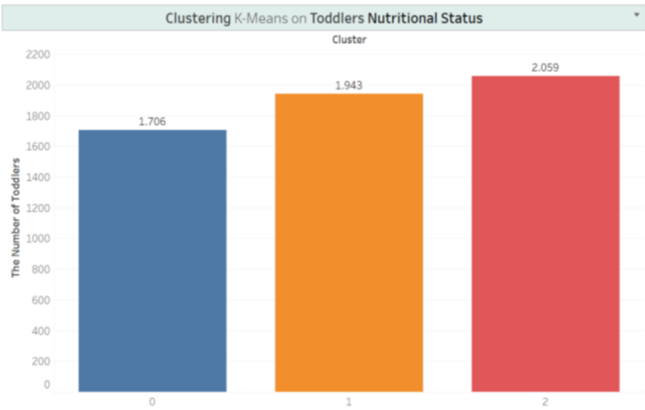

Fig. 4: Visualization of total data for each cluster

Based on Figure 4, the data distribution in each cluster resulting from the clustering analysis can be observed. Cluster 2 appears to be the largest, with a total of 2059 data points, indicating that the majority of the data in the clustering analysis belong to this cluster. It is followed by Cluster 1, which is the second largest, comprising 1943 data points, a significant number though lower than Cluster 2. Meanwhile, Cluster 0 is the cluster with the lowest number of data, encompassing only 1706 data points. Out of the total 5708 data points analyzed, the percentage of data in each cluster can be observed, with Cluster 0 contributing 29.88% of the total data, Cluster 1 around 34.03%, and Cluster 2 around 36.07%. The data distribution within clusters can be used to support decision-making and gain a better understanding of the characteristics of each cluster.

4.4. Analysis of Data Characteristics in Each Cluster

The clusters obtained from the data mining process will then be analyzed in detail for their characteristics as follows.

Table 3: Distribution of Data Count by Nutritional Status Categories in Clusters 0, 1, and 2

Attribute	Nutritional Status Category	Percentage of Data (%)		
		Cluster0	Cluster1	Cluster 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

Cluster 0 in clustering analysis predominantly consists of toddlers with normal weight for their age (86.8%), but approximately 9.6% are at risk of being overweight. Regarding height, about 72.4% of toddlers have shorter stature than the average, and 27.6% have very short height, indicating growth issues. In terms of weight and height, the majority of toddlers (around 77.9%) have good nutritional status, but about 15.4% are at risk of being overweight. When looking at body mass index and age, approximately 69.9% have a good nutritional status, while 18.2% are at risk of being overweight. There are no toddlers experiencing poor nutrition or undernutrition. Overall, the nutritional status of toddlers in Cluster 0 can be considered good and healthy, although some are at risk of being overweight or facing nutritional issues, with no cases of poor nutrition or undernutrition in this group.

Cluster 1 represents the characteristics of toddlers with nutrition problems. Based on the attributes of weight (W) and age (A), approximately 60.7% of toddlers are severely underweight, and 39.3% are underweight. There are no toddlers with normal weight or at risk of being overweight in this cluster. Based on height (H) and age (A), around 97.1% of toddlers have very short stature, with only 2.6% having short stature, and 0.3% having normal height. In terms of weight and height, about 88.1% of toddlers have good nutritional status, 7.8% are undernourished, and 2.3% are at risk of overnutrition. For the body mass index

(BMI) attribute and age, approximately 80.8% of toddlers have good nutritional status, 6.5% are at risk of overnutrition, 6.3% are undernourished, and 5.7% have poor nutrition. A small percentage of toddlers (0.7%) have excess nutrition, and only a few (0.1%) are classified as obese. Overall, the characteristics of toddlers in Cluster 1 are having very short stature and being underweight to severely underweight for their age, but predominantly fall into the category of good nutritional status in the Weight/Height and BMI/Age attributes.

Cluster 2 represents characteristics of toddlers with generally good nutritional status. Based on the attributes of weight and age, approximately 77.2% of toddlers have normal weight, 22.2% are underweight, and only 0.6% are at risk of being overweight. When comparing height and age, almost all toddlers (99.8%) have normal height. In terms of weight and height, around 82.3% of toddlers have good nutritional status. Regarding the body mass index and age attribute, approximately 82.3% of toddlers have good nutritional status, 8.1% are at risk of being overweight, 3.9% are overweight, 3.1% are obese, and only 0.1% have poor nutrition. Therefore, the characteristics of toddlers in Cluster 2 are predominantly having normal weight and underweight, as well as a majority having normal height. Referring to Weight/Height and Body Mass Index/Age, the nutritional status of toddlers in Cluster 2 is mainly dominated by normal nutrition.

4.5. Dashboard Visualization

Every visualization has different purposes and functions. The first visualization provides an understanding of the distribution of the number of toddlers in two neighborhoods, which can serve as a basis for making decisions related to toddler healthcare services. The second visualization compares the physical characteristics of male and female toddlers. Meanwhile, the third visualization offers insights into the distribution of toddler ages, which may be relevant in the context of data analysis. The main goal of these visualizations is to make data more easily understood, facilitate the comprehension of patterns, and present information in a way that is more digestible to dashboard users. This aids in making better decisions and provides deeper insights into the analyzed data.

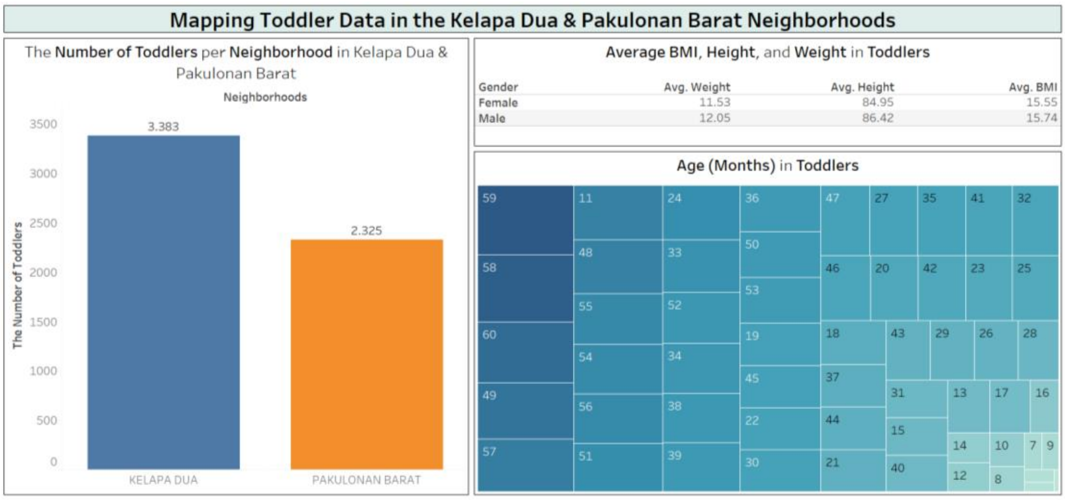

Fig. 5: Visualization of Toddler Data Mapping in Kelapa Dua & Pakulonan Barat Subdistricts

Fig. 5 toddler Data Mapping in the Kelapa Dua and Pakulonan Barat Subdistricts provides three visualizations that offer insights into:

1. Comparison of the Number of Toddlers per Sub District in Kelapa Dua & Pakulonan Barat: This visualization takes the form of a bar chart that illustrates a significant difference in the number of toddlers between two subdistricts, namely Kelapa Dua and Pakulonan Barat. Kelapa Dua has a total of 3,883 toddlers, while Pakulonan Barat has only 2,325 toddlers. Consequently, Kelapa Dua has a larger toddler population than Pakulonan Barat.

2. Average Body Mass Index, Height, and Weight of Toddlers: This visualization is presented in the form of a table displaying the average body mass index (BMI), weight, and height of toddlers, with a breakdown between males and females. The results show that the average weight of male toddlers is

11.68, which is greater than that of female toddlers with an average weight of 11.10. Additionally, the average BMI for males is 15.59, while female toddlers have an average BMI of 15.47. Furthermore, the average height of male toddlers is 85.41, which is taller than female toddlers' average height of 83.55. These data indicate that the average physical attributes of male toddlers tend to be larger than those of female toddlers.

3. Age (months) of Toddlers: This visualization takes the form of a Tree Map, providing an overview of the distribution of toddler age. In this visualization, toddlers at 11 months of age have the darkest color and the largest box, indicating that in the data, toddlers at 11 months of age constitute the largest group compared to other age groups. Toddlers at 59 months of age rank second, with a significant number of toddlers. This visualization offers an understanding of the distribution of toddler ages within the analyzed population.

With this visualization, Community Health Center can easily identify the differences in the number of toddlers between two neighborhoods, understand the comparison of the average physical attributes of toddlers based on gender, and gain insights into the age distribution of toddlers within the population. This information can be used for making better decisions based on child health-related data in those two neighborhoods. Visualization simplifies the understanding of complex data and aids in devising appropriate strategies to improve the well-being of toddlers in the area.

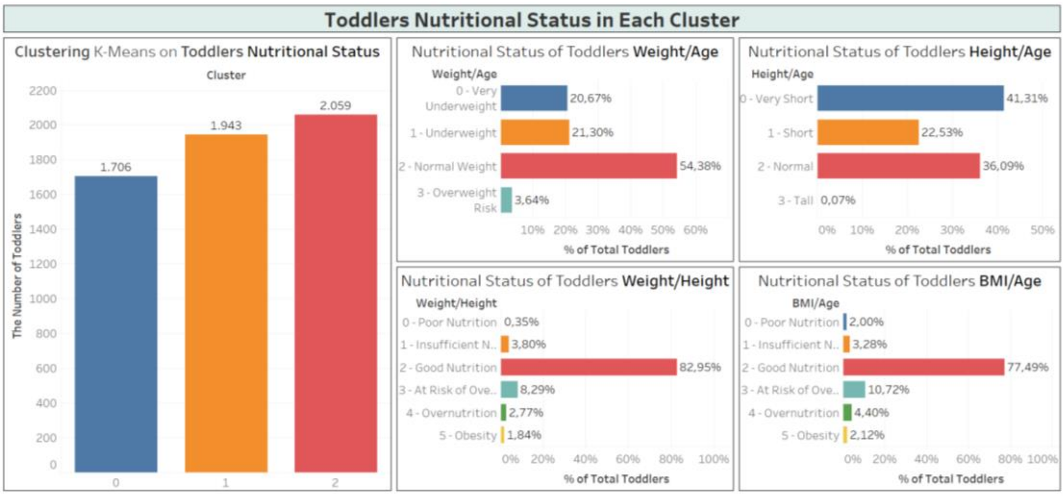

Fig. 6: Visualization of Mapping the Nutritional Status of Toddlers in Each Cluster

In the Toddler Nutrition Status Cluster Dashboard, there are five visualizations that provide an overview of the toddler birth situation within a specific cluster.

1. Comparison of the Number of K-Means Clusters in Toddler Nutrition Status: This visualization uses a bar chart to divide toddler nutrition status into three categories: cluster 0, 1, and 2. From this visualization, it is evident that the data mapping across the three clusters is relatively even. Cluster 0 has 1706 data points, cluster 1 has 1943 data points, and cluster 2 has 2059 data points.

2. Comparison of Toddler Nutrition Status by Weight/Age: This visualization uses a bar chart to compare toddler nutrition status based on weight and age. The majority of the data indicates normal weight (54.38%). Some toddlers have low weight (21.30%) or very low weight (20.67%), while a small portion have an overweight risk (3.64%).

3. Comparison of Toddler Nutrition Status by Height/Age: In this bar chart visualization, toddler nutrition status is compared based on height and age. The majority of the data indicates very short height (41.31%). Some others have normal height (36.09%) or short height (22.53%), and only a small fraction have tall height (0.07%).

4. Comparison of Toddler Nutrition Status by Weight/Length: In this visualization, toddler nutrition status is compared based on weight and length. The majority of the data indicates good

nutrition status (82.95%). Some others are at risk of overnutrition (8.29%), undernutrition (3.80%), overnutrition (2.77%), obesity (1.84%), or even poor nutrition (0.35%).

5. Comparison of Toddler Nutrition Status by Body Mass Index/Age: This visualization compares toddler nutrition status based on body mass index and age using a bar chart. The majority of the data indicates good nutrition status (77.49%). Some toddlers are at risk of overnutrition (10.72%), overnutrition (4.40%), undernutrition (3.28%), obesity (2.12%), or even poor nutrition (2.00%).

With this visualization, one can see how the nutritional status of toddlers differs within each cluster and based on various attributes, such as weight, height, weight/length ratio, and body mass index. This visualization aims to provide valuable information to Community Health Center in understanding and making decisions related to the nutritional status of toddlers in various clusters. The benefits include a better understanding of the distribution of nutritional status, making informed decisions, monitoring progress, providing guidance, and evaluating programs. With this visualization, Community Health Center can improve toddler nutrition care, inform families, and ensure the effectiveness of existing programs.

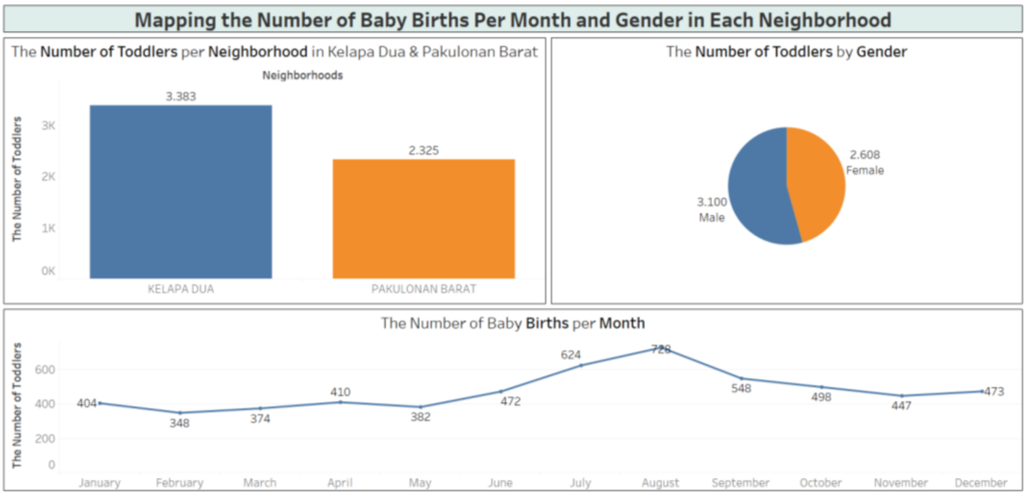

Fig. 7: Visualization of mapping the number of births each month and gender in each sub district

Fig. 7 displays three visualizations that provide information about the situation of toddler births in a specific area:

1. Comparison of the Number of Toddlers in Kelapa Dua and Pakulonan Barat: In a bar chart visualization, a significant difference in the number of toddlers in two neighborhoods, Kelapa Dua and Pakulonan Barat, is evident. Kelapa Dua has a toddler population of 3,883, while Pakulonan Barat has 2,325 toddlers. This indicates that Kelapa Dua has a higher number of toddlers compared to Pakulonan Barat.

2. Comparison of Toddler Gender: Through a pie chart visualization, the distribution of toddler genders in the Kelapa Dua and Pakulonan Barat areas can be seen. There is a significant difference in the number of male and female toddlers. The number of male toddlers is 3,100, while the number of female toddlers is 2,608. This data shows that there are more male toddlers than female toddlers in that region.

3. Number of Toddler Births per Month: Through a line chart, we can observe the pattern of the distribution of the number of toddler births throughout the year. In this visualization, the month of July shows the highest number of toddler births, with 624 toddlers, followed by August with 728 toddler births. In contrast, February records the lowest number of toddler births, with only 348 toddlers. This pattern reflects a significant fluctuation in the number of births during the year, with a noticeable increase from February to July and then a decrease.

This data illustrates the fluctuations in the number of toddler births over the course of a year. Understanding this pattern helps the health center manage resources and facilities more effectively in accordance with the fluctuations in the number of toddler births. With a deeper understanding of the characteristics of toddler births in these two neighborhoods, the health center can plan and organize services that are more effective and appropriate to support the well-being of toddlers and mothers in the area.

4.6. The Implications of Research Findings

In the first statement, the hypothesis focuses on the ability of the k-means method to group toddler data based on nutritional status categories using the four standard anthropometric indices. The research findings indicate that the k-means method is capable of clustering toddler data based on the nutritional status attribute measured from anthropometric indices. This aligns with the hypothesis's expectation that the k-means algorithm can segregate data based on nutritional status categories. For the second hypothesis, there was an expectation that clusters with good nutritional status might not always have normal height and weight. The research findings confirm this hypothesis by showing that there are clusters categorized with good nutritional status but do not always possess height and weight conforming to standard norms. This suggests that having a good nutritional status does not necessarily imply normal height and weight within each cluster.

Meanwhile, the third hypothesis assumes that the percentage of data within clusters characterized by normal weight, normal height, and good nutrition would be over 60%. The research results indicate that the majority of the data, consistent with the hypothesis, belong to clusters with these attributes. This reinforces the assumption that most data would have normal weight and height as well as good nutritional status. Overall, the research findings support the assumptions stated in the hypotheses. The k-means method effectively categorizes data based on nutritional status, while clusters with good nutritional status do not always have normal height and weight. Moreover, the majority of the data fall within clusters that align with the characteristics outlined in the hypotheses—having normal weight, normal height, and good nutritional status.

Every The results of the conducted data analysis have significant practical implications for the Community Health Center and parents. These implications are particularly related to a deeper understanding of toddlers' nutritional status and aid in making more accurate and effective decisions regarding childcare and health.

Implications for Health Centers:

1. Targeted Health Program Development: The Community Health Center can utilize these findings to design more focused health programs. They can identify toddler groups based on the generated nutritional clusters, allowing the provision of more specific and targeted health services.
2. Resource Prioritization: With a deeper understanding of different toddler groups, health centers can allocate resources more effectively. They can provide more attention to groups requiring more intensive health interventions.
3. Evaluation of Existing Health Programs: Cluster analysis enables the evaluation of existing health programs. Health centers can review the effectiveness of current programs and adjust strategies based on the findings from this data analysis.

Implications for Parents:

1. Deeper Understanding of Child Health: Parents can gain more detailed information about their child's nutritional status. This helps them understand their child's health and growth conditions better.
2. Child Health Monitoring: With a better understanding of attributes such as weight, height, and body mass index of their child, parents can regularly monitor their child's growth. This allows them to take prompt action if there are significant changes in their child's health condition.
3. Selection of More Appropriate Nutritional Policies: By understanding the nutritional clustering of their child, parents can make smarter decisions regarding dietary patterns and nutrition. They can adapt food policies more suited to their child's specific nutritional needs.

Therefore, these data analysis results provide a more detailed insight into toddlers' health conditions. This enables health centers and parents to take more targeted and specific steps in the care and nutritional planning for children.

4.7. Discussion

The results of this data analysis provide a clear picture of the nutritional status of toddlers and the influencing factors. By utilizing clustering methods and analyzing various attributes such as weight, height, and body mass index, the findings address the research question regarding the nutritional status of toddlers in three distinct clusters. The hypothesis underlying this analysis is that there is a connection between growth patterns and nutritional status among toddlers. The results support this hypothesis by identifying three groups of toddlers with different growth characteristics and nutritional statuses. This illustrates a correlation between factors like weight, height, and body mass index with toddler nutrition status, confirming the proposed hypothesis.

This study relates its findings to previous similar studies concerning the understanding of toddler nutritional status. Previous studies also conducted cluster analyses of children's nutritional statuses using similar methods. The findings in those studies broadly support the results obtained in this research, showing similar patterns in clustering toddler nutritional status. However, other previous studies showed striking differences in clustering children's nutritional statuses. This was due to variations in the attributes used or diversity within the studied population. These differing results highlight the complexity in interpreting children's nutrition clustering data. The correlation between the findings of this research and previous studies provides a broader understanding of potential patterns in toddler nutritional statuses. It also offers a better understanding of the reliability and consistency of findings in this field of research.

This study offers significant insights into toddler nutritional status, guiding the development of nutritional programs and child health-related policies. The analysis provides a profound understanding of children's growth patterns and their nutritional statuses. Concerning toddler nutritional programs, these findings offer insight into various nutritional conditions experienced by children. Understanding these clusters allows health centers to develop more targeted and focused nutritional programs, allocate resources effectively based on the needs of specific groups requiring special attention, and evaluate the effectiveness of existing programs. In terms of related policies, these results can assist in developing more detailed and adaptive nutritional guidelines or policies. For instance, by understanding the distribution of children's groups based on nutritional status, nutritional policies can be better directed to meet the specific needs of each group. This helps ensure that every child receives nutritional support suitable for their condition. Overall, the findings of this research provide a strong foundation for the development of more effective toddler nutritional programs and more directed policies, ensuring that each child receives nutritional support tailored to their needs.

5. Conclusion

The application of the k-means clustering algorithm is used to group toddler data based on their nutritional status, which is analyzed through four anthropometric indices: Body/Age, Height/Age, Body Weight/Height, and Body Mass Index/Age as attributes. In this study, toddler data is grouped into three clusters: Cluster 0, Cluster 1, and Cluster 2. The clustering process utilizes the similarity of anthropometric attributes to group toddlers with similar characteristics into the same cluster. By understanding the unique characteristics of each cluster, more specific and measurable intervention steps can be taken to address toddler nutrition issues. Analyzing the data similarities within each cluster allows conclusions to be drawn about the unique characteristics of each cluster.

A cluster with data characteristics indicating good nutritional status according to the anthropometric indices Weight/Height and Body Mass Index/Age contains a group of toddlers with abnormal height or weight based on the Height/Age and Weight/Age anthropometric indices. Toddlers in Cluster 1 have data characteristics indicating good nutrition based on the Weight/Height and Body Mass Index/Age indices. However, when referring to the Weight/Age and Height/Age anthropometric indices, it is evident that they have very low weight and short stature for their age.

The data analysis of toddlers at the Community Health Center reveals that the majority of toddlers in all clusters have adequate nutritional conditions and fall within the normal range for weight and height, with a percentage of 62.7%. However, there are 37.3% of toddlers who fall into the group with abnormal height or weight and poor nutritional status. This percentage should be a matter of serious concern for the Community Health Center and parents, considering the importance of their role in routinely monitoring the health of toddlers. Through consistent monitoring using a visualization dashboard, the Community Health Center can be more effective in providing appropriate interventions and supporting the optimal development of toddlers.

This research is crucial as it provides a deeper understanding of toddler nutrition. Through the k-means clustering approach, the study successfully categorized toddler data into three clusters, opening opportunities to comprehend the unique characteristics of each cluster. The ability to precisely identify each group of toddlers lays a strong foundation for designing more specific interventions to address nutritional challenges in toddlers more effectively. In-depth analysis of data similarity within each cluster also provides the opportunity to draw conclusions about the distinctive characteristics of each group, enriching our understanding of the variations in nutritional conditions among toddlers.

As for future plans to expand this study, several strategic steps are considered. Firstly, leveraging alternative data mining methods like classification will enrich the analysis and enhance the accuracy of findings. Secondly, collecting datasets with broader coverage and longer durations will yield stronger and sustainable findings. This step will deepen our understanding of the dynamics of toddler nutrition. Furthermore, the development of an application prototype utilizing the attributes and outcomes of this research is expected to educate the community and encourage routine check-ups for toddlers. Thus, there will be broader and more targeted efforts to expand knowledge about toddler nutrition and support their overall health and well-being.

References

Ahamad, M. G., Ahmed, M. F., & Uddin, M. Y. (2018). Clustering as Data Mining Technique in Risk Factors Analysis of Diabetes, Hypertension and Obesity. *European Journal of Engineering and Technology Research*, *1*(6), 88–93. https://doi.org/10.24018/ejeng.2016.1.6.202

Alpiana, I., & Anifah, L. (2019). Penerapan Metode KnA (Kombinasi K-Means dan Agglomerative Hierarchical Clustering) dengan Pendekatan Single Linkage untuk Menentukan Status Gizi pada Balita. In *Indonesian Journal of Engineering and Technology (INAJET)* (Vol. 1, Issue 2). https://journal.unesa.ac.id/index.php/inajet

Amelia, N., Yaddarabullah, Y., Arifitama, B., & Jusman, Y. (2022). Classification of toddler nutrition status using data mining with algorithm C4.5 (Case study: Integrated service centers in Ciasihan Village). In *AIP Conference Proceedings* (Vol. 2499). https://doi.org/10.1063/5.0105013

Bhattacharya, M., Ehrenthal, D., & Shatkay, H. (2014). Identifying Growth-Patterns in Children by Applying Cluster analysis to Electronic Medical Records. In *Proceedings - 2014 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2014*. https://doi.org/10.1109/BIBM.2014.6999183

Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. In *International Journal of Computer Applications* (Vol. 105, Issue 9).

Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D., Thompson, E., & Ashraf, I. (2023). A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. *Symmetry*, *15*, 1679. https://doi.org/10.3390/sym15091679

Dol, S. M., & Jawandhiya, P. M. (2023). Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining — A survey. *Engineering Applications of Artificial Intelligence*, *122*, 106071. https://doi.org/https://doi.org/10.1016/j.engappai.2023.106071

Dubey, P., & Chakrawarti, R. (2023). *Analysis of Clustering-Algorithms for Efficient Data Mining*.

Duque, J., Silva, F., & Godinho, A. (2023). Data Mining applied to Knowledge Management. *Procedia Computer Science*, *219*, 455–461. https://doi.org/https://doi.org/10.1016/j.procs.2023.01.312

Dwi Wulandari, R., & Laksono, A. (2023). *Factors Related to Stunted Toddlers in Eastern Indonesia*. *3*, 1–10. https://doi.org/10.32695/JKIT.V3I1.451

Fauziah, M., Permatasari, T., Herdiansyah, D., Latifah, N., Murod, M., & Defriyana, D. (2021). Factors Associated with the Nutritional Status of Toddlers at the Limo Community Health Center, Depok City, Indonesia. *International Journal of Early Childhood Special Education*, *13*, 701–706. https://doi.org/10.9756/INT-JECSE/V13I2.211110

Flores, K. R., de Carvalho, L. V. F. M., Reading, B. J., Fahrenholz, A., Ferket, P. R., & Grimes, J. L. (2023). Machine learning and data mining methodology to predict nominal and numeric performance body weight values using Large White male turkey datasets. *Journal of Applied Poultry Research*, *32*(4), 100366. https://doi.org/https://doi.org/10.1016/j.japr.2023.100366

Han, J., Pei, J., & Tong, H. (2023). *Data Mining: Concepts and Techniques*.

Harada, D., Asanoi, H., Noto, T., & Takagawa, J. (2020). Different Pathophysiology and Outcomes of Heart Failure With Preserved Ejection Fraction Stratified by K-Means Clustering. *Frontiers in Cardiovascular Medicine*, *7*. https://doi.org/10.3389/fcvm.2020.607760

Iuliano, S., Poon, S., Robbins, J., Wang, X., Bui, M., & Seeman, E. (2023). Provision of High Protein Foods Slows the Age-Related Decline in Nutritional Status in Aged Care Residents: A Cluster-Randomised Controlled Trial. *The Journal of Nutrition, Health & Aging*, *27*, 1–6. https://doi.org/10.1007/s12603-022-1868-7

Julianto, I., Kurniadi, D., Nashrulloh, M., & Mulyani, A. (2022). Data Mining Clustering Food Expenditure in Indonesia. *Jurnal Teknik Informatika (Jutif)*, *3*, 1491–1500. https://doi.org/10.20884/1.jutif.2022.3.6.331

Kebijakan, B., Kesehatan, P., & Ri, K. K. (2022). *Buku Saku Hasil Survei Status Gizi Indonesia (SSGI) 2022*.

Krishnamoorthy, M., & Karthikeyan, R. (2022). Pattern mining algorithms for data streams using itemset. *Measurement: Sensors*, *24*, 100421. https://doi.org/https://doi.org/10.1016/j.measen.2022.100421

Kruger, H., Visser, M., Malan, L., Zandberg, L., Wicks, M., Ricci, C., & Faber, M. (2023). Anthropometric nutritional status of children (0–18 years) in South Africa 1997–2022: a systematic review and meta-analysis. *Public Health Nutrition*, 1–17. https://doi.org/10.1017/S1368980023001994

Limauro, R., Cioffi, L., Bianco, V., Caruso, V., Casani, A., Gaizo, D., Esposito, A., Farris, E., Gallo, P., Giuliano, M., Iannone, A., Izzo, A., Vecchia, M., Raineri, L., Sabetti, P., Sassi, R., Servodidio, C., Sorice, N., Cioffi, V., & Valerio, G. (2023). *Nutritional counselling and risk factors for obesity: an observational study in toddlers.* https://doi.org/10.21203/rs.3.rs-3178335/v1

Lin, Y., Li, H., Wu, A., Wang, Y., & Qu, H. (2023). Dashboard Design Mining and Recommendation. *IEEE Transactions on Visualization and Computer Graphics*, *PP*, 1–15. https://doi.org/10.1109/TVCG.2023.3251344

Macuácua, J. C., Centeno, J. A. S., & Amisse, C. (2023). Data mining approach for dry bean seeds classification. *Smart Agricultural Technology*, *5*, 100240. https://doi.org/https://doi.org/10.1016/j.atech.2023.100240

Majumder, J., & Sharma, L. (2015). Identifying Body Size Group Clusters from Anthropometric Body Composition Indicators. *Journal of Ecophysiology and Occupational Health*, *15*, 1–8. https://doi.org/10.15512/joeoh/2015/v15i3-4/121588

Masturina, M., Salam, A., Indriasari, R., Thaha, Abd. R., & Jafar, N. (2023). Description of family characteristics and nutritional status in toddlers. *Community Research of Epidemiology (CORE)*, 101–114. https://doi.org/10.24252/corejournal.vi.37731

Masturina, S., Hardjito, K., & Rahayu, D. (2023). The relationship between feeding patterns and nutritional status of toddlers. *Science Midwifery*, *11*, 153–164. https://doi.org/10.35335/midwifery.v11i1.1248

Meher, C., Zaluchu, F., & Eyanoer, P. (2023). Local approaches and ineffectivity in reducing stunting in children: A case study of policy in Indonesia. *F1000Research*, *12*, 217. https://doi.org/10.12688/f1000research.130902.1

Menteri Kesehatan Republik Indonesia. (2020). *Peraturan Menteri Kesehatan Republik Indonesia*.

Nagari, S., & Inayati, L. (2020). Implementation of Clustering Using K-Means Method to Determine Nutritional Status. *Jurnal Biometrika Dan Kependudukan*, *9*, 62. https://doi.org/10.20473/jbk.v9i1.2020.62-68

Novyriana, E., Irtanti, P., Sumaji, V., Susilo, N., & Rahmadhani, W. (2022). Factors Affecting Balanced Nutrition In Babies And Toddlers. *Journal of Sexual and Reproductive Health Sciences*, *1*, 70. https://doi.org/10.26753/jsrhs.v1i2.707

Nurzaman, A., & Jayadi, R. (2022). Dashboard System for Predicting Student Practicum Performance Using the Data Mining Method. *CESS (Journal of Computer Engineering, System and Science)*, *7*, 394. https://doi.org/10.24114/cess.v7i2.35394

Nwagu, C., Kenneth, O., & Chinecherem, I. (n.d.). *Knowledge Discovery in Databases (KDD): An Overview*. https://sites.google.com/site/ijcsis/

Pambudi, E., Fahrezi, A., & Badharudin, A. (2023). Forecasting Brown Sugar Production Using k-NN Minkowski Distance and Z-Score Normalization. *Journal of Information Systems and Informatics*, *5*, 580–589. https://doi.org/10.51519/journalisi.v5i2.485

Perumal, N., Namaste, S., Qamar, H., Aimone, A., Bassani, D., & Roth, D. (2020). Anthropometric data quality assessment in multisurvey studies of child growth. *The American Journal of Clinical Nutrition*, *112*. https://doi.org/10.1093/ajcn/nqaa162

Pratiwi, N., Atika, P., & Herlawati, H. (2022). Nutritional Status Classification of Toddlers Using K-Nearest Neighbor Algorithm. *Piksel : Penelitian Ilmu Komputer Sistem Embedded and Logic*, *10*, 117–132. https://doi.org/10.33558/piksel.v10i2.5604

Purich, J., Srinivasan, A., Correll, M., Battle, L., Setlur, V., & Crisan, A. (2023). *Toward a Scalable Census of Dashboard Designs in the Wild: A Case Study with Tableau Public*.

Rae, S., Ong, F., Dennis, C.-L., Hamilton, J., Pullenayegum, E., Maguire, J., & Birken, C. (2023). Parent involvement in child anthropometric measurement. *BMC Primary Care*, *24*, 89. https://doi.org/10.1186/s12875-023-02028-2

Rahmawati, W., & Retnaningrum, D. (2022). The Role of Mothers Knowledge Regarding Nutritional Needs of Toddlers Nutritional Status. *Jurnal MID-Z (Midwivery Zigot) Jurnal Ilmiah Kebidanan*, *5*, 139–143. https://doi.org/10.56013/jurnalmidz.v5i2.1638

Sharma, B., & Deshmukh, S. (2020). Data Visualization for Accelerated Business Intelligence in the Indian Health Care Sector using Tableau. *International Journal of Recent Technology and Engineering (IJRTE)*, *9*, 304–308. https://doi.org/10.35940/ijrte.D4979.119420

Sholahuddin, M., Atqiya, F., Faridah, H., & Nurianti, N. (2022). Google Data Studio Implementation for Visualizing West Java Province Toddler Stunting Data. *The IJICS (International Journal of Informatics and Computer Science)*, *6*, 81. https://doi.org/10.30865/ijics.v6i2.4696

Tran, T., Ruppert, T., Eigner, G., & Abonyi, J. (2023). Assessing human worker performance by pattern mining of Kinect sensor skeleton data. *Journal of Manufacturing Systems*, *70*, 538–556. https://doi.org/https://doi.org/10.1016/j.jmsy.2023.08.010

Virgia, V., & Widiyawati, R. (2023). Fortification of Moringa Leaves to Fulfill the Nutritional Needs of Stunting Toddler. *International Journal of Scientific Research and Management (IJSRM)*, *11*, 87–96. https://doi.org/10.18535/ijsrm/v11i10.nd01

Winiari, S., Yuliansyah, H., & Andi, A. (2018). Identification of Toddlers' Nutritional Status using Data Mining Approach. *International Journal of Advanced Computer Science and Applications*, *9*. https://doi.org/10.14569/IJACSA.2018.090122

Yanhui, D., Lau, W. C., Dong, B., Zou, Z., Yide, Y., Wen, B., Ma, Y., Hu, P., Song, Y., Ma, J., Sawyer, S., & Patton, G. (2019). Trends in physical fitness, growth, and nutritional status of Chinese children and adolescents: a retrospective analysis of 1·5 million students from six successive national surveys between 1985 and 2014. *The Lancet Child & Adolescent Health*, *3*. https://doi.org/10.1016/S2352-4642(19)30302-5

Zikri, A., Adrian, J., Soniawan, A., Azim, R., Dinur, R., & Akbar, R. (2017). Implementasi Business Intelligence untuk Menganalisis Data Persalinan Anak di Klinik Ani Padang dengan Menggunakan Aplikasi Tableau Public. *Jurnal Online Informatika*, *2*, 20. https://doi.org/10.15575/join.v2i1.70