Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14311
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Analysis of Stunting Risk Factors Using K-Means Clustering and PCA in Sambas Regency, Indonesia

M. Hilma Minanur Rohman[1)], Farrikh Al Zami[2)], Heru Pramono Hadi[3)*], Zaenal Arifin[4)], Titien Suhartini Sukamto[5)], Ayu Ashari[6)], Moh. Yusuf[7)]
[1,2,3,4,5,6)] Universitas Dian Nuswantoro, Indonesia
7) Faculty of Dentistry, Universitas Islam Sultan Agung, Indonesia
[1)]112202106708@mhs.dinus.ac.id, [2)]alzami@dsn.dinus.ac.id, [3)]heru.pramono.hadi@dsn.dinus.ac.id,
[4)]zaenal@dosen.dinus.ac.id, [5)]titien.suhartini@dsn.dinus.ac.id, [6)]ayu.ashari@dsn.dinus.ac.id,
[7)]mohyusuf@unissula.ac.id

Submitted :Dec 8, 2024 | **Accepted** : Dec 31, 2024 | **Published** : Jan 8, 2025

Abstract: Stunting, characterized by impaired growth and development in children, is one of the most serious public health problems often caused by chronic malnutrition. This study aims to identify patterns among stunting cases through clustering analysis of child health data. The algorithm used in this research uses K-Means. The dataset used in this study uses health data from 599 children in the Sambas Regency area of East Kalimantan Province. This dataset has several features that are quite diverse such as height, weight, age, nutritional intake, socioeconomic status, and others. This research process begins with cleaning the data, as well as looking at the correlation between features. One of the methods used is to conduct a data analysis process using Principal Component Analysis (PCA) which aims to reduce the dimensions of the data. After that, the process of finding the number of clusters using the Elbow method is carried out to determine the optimal number of clusters. This research uses 4 clusters in the process. The clustering results revealed that family structure (main family vs extended family) and parental income levels significantly influence stunting prevalence in the region.

Keywords: clustering; elbow method; k-means; principal component analysis (PCA); stunting;

INTRODUCTION

Stunting is one of the most serious public health problems, characterized by impaired growth and development in children due to chronic malnutrition. It is calculated based on measurements of height-for-age that are more than two standard deviations below World Health Organization (WHO) growth standards. addressing stunting is critical, as it applies not only to health, but also to human resource development and economic growth; in addition, stunting can lead to long-term cognitive and physical impairment, which affects future education levels and productivity (Paul et al., 2021; K. Takele et al., 2019). In the field of computer science, stunting can be addressed through data analysis and machine learning techniques to identify affected populations and maximize resource allocation for nutritional needs; for example, predictive analytics can assist in understanding the socioeconomic factors that play a role in stunting, enabling targeted programs that leverage technology for better health outcomes (Aguilera Vasquez & Daher, 2019; Nemerimana et al., 2023). In addition, integrating health data with computational tools can improve monitoring and evaluation of nutrition programs, ultimately contributing to stunting reduction (Aheto, 2020). Thus, the intersection between stunting and computer science presents an important opportunity for innovative solutions to pressing global challenges.

Further investigation of stunting is crucial as this phenomenon impacts three key aspects of national stability, namely health, human resource development, and long-term economy. Stunting can be identified from the imbalance between height and age of the sufferer, which is specifically caused by chronic malnutrition, and is considered a cause of increased morbidity and mortality in children, existing research shows that children affected by stunting have limitations in terms of cognition; This limitation can certainly hamper various things, namely in terms of education and productivity; Both of these things will certainly affect, and have a negative impact on their future (Tola et al., 2023). The high rate of stunting is very significant, especially for developing countries; this is

e-ISSN : 2541-2019
p-ISSN : 2541-044X

specifically the impact of low and middle income; this problem is the forerunner of food security problems, and is exacerbated by difficult access to health services (Gebreayohanes & Dessie, 2022; S. M. J. Rahman et al., 2021). Research is needed to address the multiple causes of stunting, including maternal health, dietary practices, and environmental conditions (Gansaonré et al., 2022; B. A. Takele et al., 2022). By understanding these, more effective strategies can be implemented to reduce stunting and improve overall child health outcomes.

Clustering, an essential unsupervised machine learning technique, offers a systematic approach to identify patterns in stunting data by grouping similar characteristics without predefined labels (Chen et al., 2023). This method is particularly valuable in public health research as it can reveal natural groupings of risk factors and affected populations. By automatically organizing complex health datasets into meaningful clusters, it enables researchers to uncover hidden patterns that might be overlooked using traditional analytical methods (S. Rahman et al., 2022). Among various clustering algorithms, K-means has proven effective in health-related studies due to its ability to handle large datasets and identify distinct patient groups, while hierarchical clustering provides detailed relationship structures between variables (Chen et al., 2023; Yu et al., 2024). In the context of stunting, clustering can help identify groups of children with similar risk factors, enabling more targeted and effective interventions.

The application of clustering techniques in healthcare has emerged as a powerful tool for public health analysis and policy-making. In the context of health research, clustering effectively identifies significant patterns within health-related data, enabling evidence-based resource allocation and targeted interventions (Whitaker et al., 2021). This analytical approach transforms complex health data into actionable insights, allowing stakeholders to make informed decisions about community health initiatives (Ariza Colpas et al., 2020). Beyond traditional health metrics, clustering analysis can reveal important relationships between health outcomes and socioeconomic determinants, providing a more comprehensive understanding of health disparities (Satoh, 2022). While the Indonesian Code of Medical Ethics presents certain limitations in data collection, careful methodology and ethical considerations can help overcome these challenges (Janssen et al., 2019). The versatility of clustering in healthcare extends beyond identifying high-risk populations; it can effectively segment populations based on socioeconomic factors, enabling policymakers to develop more equitable and targeted health interventions (Liao et al., 2022). This multifaceted approach is particularly relevant for addressing complex public health challenges like stunting, where both health and socioeconomic factors play crucial roles.

Principal Component Analysis (PCA) serves as a crucial analytical tool in stunting research by effectively identifying key patterns in child growth data. Its primary strength lies in dimensionality reduction, enabling researchers to focus on the most influential features affecting stunting outcomes. A comprehensive study of child malnutrition in India demonstrated PCA's effectiveness in handling multicollinearity among 21 independent variables, revealing significant spatial patterns of stunting across districts (Vennam et al., 2020). Furthermore, advanced applications like Functional Principal Component Analysis (FPCA) have enhanced our understanding of child growth trajectories by identifying correlations between growth patterns and various factors, including economic conditions and breastfeeding practices (Karuppusami et al., 2022). The versatility of PCA extends to predictive modeling of growth patterns, offering valuable insights into the multifaceted nature of stunting (Massara et al., 2023). This analytical approach is particularly relevant for our study, as it helps identify the most significant variables among the numerous potential factors affecting stunting in the Sambas region.

The present study builds upon previous research conducted in Sambas Regency, West Kalimantan Province, covering an area of 6,395.70 km2 (4.36% of the province). This predominantly rural region, bordering Malaysia, encompasses 19 sub-districts, with research activities conducted in 17 sub-districts across 30 villages. The data originated from two comprehensive cross-sectional studies conducted in 2016 and 2017 (Sartika et al., 2021). The initial study in 2016 followed a systematic four-phase approach: preparation (April 11-24), data collection (April 25-May 11), data processing and analysis (June), and report compilation (through August). The follow-up study in 2017 maintained a similar structured approach: preparation (April 3-14), data collection (May 17-June 2), data processing and analysis (May), and report finalization (May-June 8). This methodical approach across both studies ensured data consistency and reliability, providing a robust foundation for our current analysis.

Previous applications of K-means clustering in stunting analysis have demonstrated its effectiveness in uncovering patterns and relationships within health data. Mokalla and Mendu's research (Mokalla & Rao Mendu, 2022) established foundational insights into stunting patterns across specific populations, emphasizing the interplay between socio-economic factors and growth outcomes. This analytical approach was further validated by Ahmed et al. (Ahmed et al., 2022a), who revealed significant correlations between stunting and dietary diversity through cluster analysis. Regional variations in stunting prevalence were effectively mapped by Takele et al. (B. A. Takele et al., 2022), who employed K-means clustering to identify high-risk areas requiring targeted interventions. Additionally, Namirembe et al. (Namirembe et al., 2022) utilized this technique to uncover crucial relationships between maternal education levels and child nutritional outcomes. These studies collectively demonstrate K-means clustering's versatility in analyzing various aspects of stunting, from nutritional factors to socio-economic determinants, despite methodological variations across studies. This body of research provides a

e-ISSN : 2541-2019
p-ISSN : 2541-044X

strong methodological foundation for our current study, while highlighting opportunities for novel applications in the Indonesian context.

LITERATURE REVIEW

Clustering analysis, particularly through the application of K-Means and Principal Component Analysis (PCA), has emerged as a vital approach in understanding the risk factors associated with stunting in children. Stunting, defined as low height-for-age, is a significant public health issue affecting millions of children worldwide, particularly in low- and middle-income countries. This literature review synthesizes recent studies that utilize K-Means and PCA to analyze the multifaceted risk factors contributing to stunting.

K-Means clustering is frequently employed to categorize children based on various risk factors associated with stunting. For instance, Takele et al. (B. A. Takele et al., 2022) conducted a Bayesian multilevel analysis to determine the pooled prevalence of stunting and its associated factors among children aged 6–59 months in Sub-Saharan Africa. Their findings revealed that children with a history of fever had higher odds of stunting, highlighting the importance of health-related factors in clustering analyses. Similarly, Aheto (Aheto, 2020) identified multiple socio-economic and maternal factors that increase the risk of severe chronic malnutrition in Ghana, demonstrating how K-Means can effectively group children based on shared risk characteristics.

The integration of PCA in clustering analysis serves to reduce dimensionality and enhance the interpretability of the data. For example, Gebreayohanes and Dessie (Gebreayohanes & Dessie, 2022) utilized PCA to analyze the prevalence of stunting in a pastoralist community in Northeast Ethiopia. Their study underscored the critical role of nutritional deficiencies and recurrent infections in stunting, which were effectively captured through PCA, allowing for a clearer understanding of the underlying factors. Furthermore, the work of Nemerimana (Nemerimana et al., 2023) on recovery from stunting in Rwanda illustrates how PCA can identify key determinants that influence recovery trajectories, thus providing insights into effective interventions.

The combination of K-Means and PCA offers a robust framework for exploring the complex interactions between various risk factors. For instance, Ahmed et al. (Ahmed et al., 2022b) examined the relationship between maternal employment and stunting, employing K-Means to cluster children based on socio-economic variables. Their analysis revealed significant disparities in stunting rates, which were further elucidated through PCA, highlighting the multifactorial nature of malnutrition. Additionally, Rahman et al (S. M. J. Rahman et al., 2021) applied machine learning techniques alongside K-Means to predict stunting risk factors in Bangladeshi children, demonstrating the adaptability of these methods in contemporary research.

Moreover, the systematic review by Vasquez and Daher (Aguilera Vasquez & Daher, 2019) emphasizes the impact of nutrition and cash-based interventions on stunting reduction, indicating that clustering analyses can inform policy decisions aimed at economic development in low-and-middle-income countries. This aligns with the findings of Tola (Tola et al., 2023), who explored stunting prevalence among neonates in Ethiopia, underscoring the necessity of targeted interventions based on clustered risk factors.

METHOD

The research methodology employs clustering, a robust unsupervised learning technique for grouping unlabeled data based on inherent similarities while distinguishing between dissimilar groups. Figure 1 illustrates our data processing workflow, which consists of three main stages: preprocessing, model training, and results analysis.

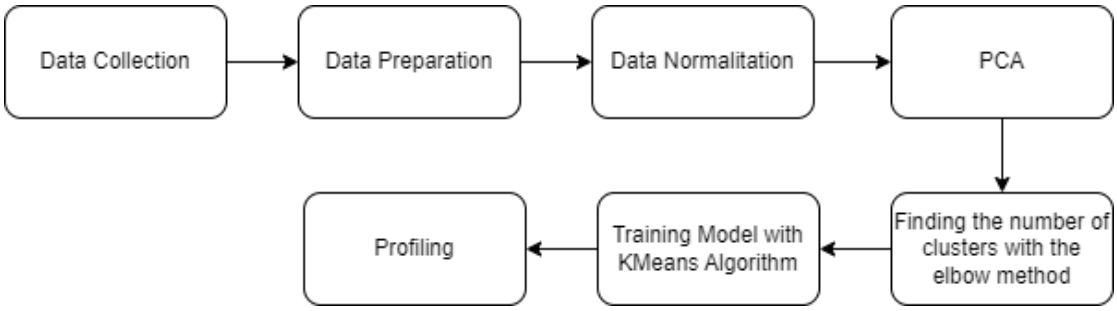

Fig. 1 The main stages in the clustering process include preprocessing, model training, and final results

Data Collection

This study utilizes a comprehensive dataset from a prospective, repeated cross-sectional study conducted in Sambas Regency. The dataset comprises 559 infants aged 0-11 months, incorporating data from two surveys: the 2016 maternal survey and the 2017 mother and child survey. The data structure includes 20 potential stunting

e-ISSN : 2541-2019
p-ISSN : 2541-044X

predictors, categorized into four main domains: household characteristics, maternal characteristics, antenatal care services, and child characteristics. These variables were initially analyzed using logistic regression to establish baseline associations with stunting outcomes.

Data Preparation

Data preparation constitutes a critical phase in ensuring data quality and analytical reliability. The initial dataset consisted of 559 observations across 660 variables, comprising 168 numerical features, 471 categorical variables, and 13 temporal features. Eight variables were excluded due to their high cardinality (unique values), which could potentially introduce noise into the analysis. The prepared dataset underwent three main preprocessing steps: missing value handling, duplicate removal, and data transformation.

1) Missing Value

Missing value treatment followed a systematic two-step approach:

a. Initial Assessment:
   1. Features with minimal research relevance were identified and removed to reduce unnecessary variance
   2. Missing values were categorized based on their nature:
      a) Intentional non-responses (indicating non-applicable criteria)
      b) Absence-related missing data (subject unavailability)

b. Treatment Strategy:
   1. Following established protocols from referenced literature:
      a) Categorical missing values were encoded as 66
      b) Numerical missing values were replaced with 0
   2. Features with >90% missing values were excluded
   3. Eight features (n5, jnskeln, jnspeny, jnskec, rsnlain, udtyp, supbr, and supfr) were removed due to excessive missing values and high uniqueness

This approach maintains data integrity while ensuring meaningful representation of missing information in the analysis.

2) Cleaning Duplicate Data

Duplicate analysis was conducted at both record and feature levels. While no duplicate records were identified, feature-level examination revealed 121 redundant variables. For instance, overlapping information was found between 'income_percapita' and 'income_percapitaNEW' columns, necessitating the removal of one variant. Additionally, we screened for constant values across features, finding none that would warrant removal.

3) Data Transformation

The final preprocessing step involved data transformation, consisting of three key operations:

a. Categorical Encoding:
   1. Conversion of categorical variables to numerical format following established protocols from reference studies
   2. Implementation of binary encoding for specific variables as illustrated in Table 1
b. Feature Engineering:
   1. Created new binary feature 'HaveHP' to indicate mobile phone ownership
   2. Value 1 assigned for non-empty 'nohp' entries, representing phone ownership
c. Data Type Standardization:
   1. Corrected data type inconsistencies, particularly for floating-point values
   2. Ensured proper numerical representation for subsequent modeling

Table 1 demonstrates the encoding methodology using the 'chmsex' variable as an example.

Table 1 Examples of transformed features

Conversion feature chmsex	
Before converted	After converted
boy	0
girl	1

PCA

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14311
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Principal Component Analysis (PCA) is implemented as a dimensionality reduction technique to address the high-dimensional nature of our dataset while preserving essential information. This method enables model simplification, accelerates training processes, and mitigates potential overfitting issues. The PCA implementation follows a systematic mathematical framework:

a. The first step that needs to be done when using PCA is to standardize the data so that each feature has a mean of 0 and a variance of 1. This process is done by subtracting the mean and dividing by the standard deviation for each feature :

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \quad (1)$$

Information :
X_{ij}= value feature j from row i
μ_j= average of feature j
σ_j= standard deviation from feature j

b. After the data is normalized, the next step is to measure the relationship of each feature in the dataset.

$$C = \frac{1}{n-1} Z^T Z \quad (2)$$

Information :
Z= Data matrix after normalization
n= number of existing rows

c. We need to calculate the eigenvalue and eigenvector of the covariance matrix C, to get the principal components. The eigenvalue (λ) indicates the amount of variance explained by each principal component, and the eigenvector (v) is the direction of the principal component.

$$Cv = \lambda v \quad (3)$$

Information :
λ= the magnitude of the variance in the eigenfactor (eigenvalue)
v= calculation direction from component main
5. Matrix Projection (Transformation)

d. After getting the eigenvectors (principal components), we can use them to transform the original data into the principal component space. If we want to reduce the data dimension to k dimensions, we select the k largest eigenvectors (with the highest eigenvalue), then we multiply it with the original data Z to get the data in the principal component space:

$$Z' = ZW_k \quad (4)$$

Information :
W_k= matrix $p \times k$ consisting of from k eigenvector the biggest.
Z'= result transformation from data to in room new dimension k.

e. Sorting eigenvalues from largest to smallest eigenvector associated with the largest eigenvalue will become the main component, followed by the second largest eigenvalue to the smallest. These principal components will be used to reduce the dimensionality of the data.

Elbow Method

Before the modeling process, the data will be normalized first to prevent the data range from being too far. Then we will find the number of clusters, using the Elbow Method on the normalized data. Elbow method offers a distortion score approach that drastically reduces distortion in determining the number of clusters.

$$Distortion\ (Inertia) = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (5)$$

Information :
k = Number of clusters
C_i= Cluster i
$x \in C_i$= Data points that are within the cluster C_i
μ_i = Centroid from cluster C_i
$\|x - \mu_i\|^2$= Euclidean distance squared between data $x$points and centroids μ_i

K-Means

K-Means is an unsupervised machine learning algorithm that aims to recognize patterns and cluster data. The way of clustering is by determining the starting point (k) randomly. After that, the point will be moved until it finds the most ideal group. The advantage of the K-Means algorithm is that it can handle relatively large data. In addition, K-Means can adapt when there are new points available. The disadvantage of K-Means is the manual

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14311
e-ISSN : 2541-2019
p-ISSN : 2541-044X

selection of the number of clusters (k). So in this case we need another algorithm to get the optimal k to be applied to the K Means algorithm. One such algorithm is the Elbow Method that we have described earlier.

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^{n} (x_{ij} - c_{kj})^2} \qquad (6)$$

Information :
x_i = vector feature from *i*- th data point.
c_k = centroid of the kth cluster.
n = number feature.

RESULT

In the methods section we have done data collection, as well as data preparation. In this section we implement the analysis following a two-stage approach that combines dimensionality reduction and clustering techniques. At first, Principal Component Analysis (PCA) is applied to reduce the dimensionality of the data while retaining important information patterns. Next, cluster analysis is performed using the K-means algorithm. All process is done using Python, and scikit-learn library

The optimal number of clusters was determined using the Elbow Method, which analyzed distortion scores across a range of K values (1 to 10) . As seen in figure 2 the analysis revealed an optimal point at K=4, with a distortion score of 7062.78, indicating the most efficient cluster configuration for our dataset. This finding guided our decision to implement K-means clustering with four distinct clusters for the final analysis.

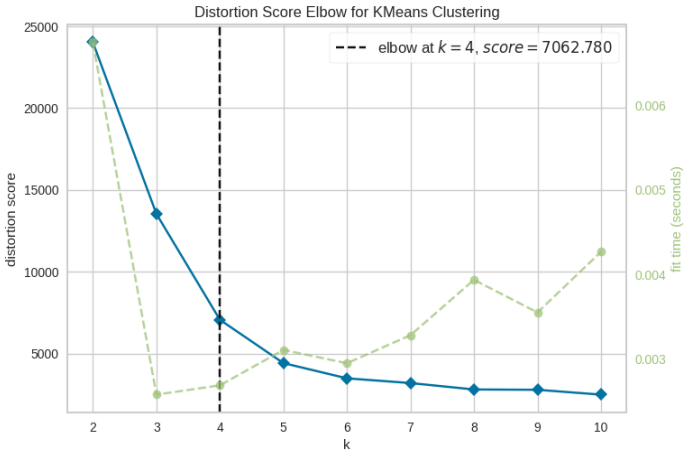

Fig. 2 Determining Amount Cluster Using the Elbow Method

We have determined that we apply 4 clusters to the K-Means algorithm used. Next we train the data, and get the data distribution as shown in the following figure 3,

Fig. 3 Data Distribution from All Generated Clusters

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14311
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Because we previously used PCA, we will get a cluster division of the data reduced by PCA. This can be seen in Figure 4,

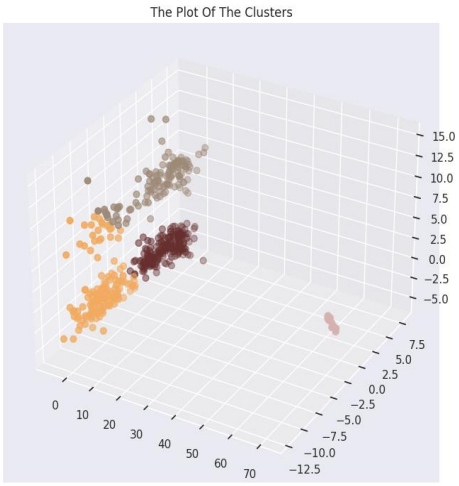

Fig. 4 Visualization Clusters in Principal Component Analysis (PCA)

DISCUSSIONS

Cluster Analysis

From the clustering process that has been processed previously, 4 clusters were obtained which will be explained as follows:

Child's Life Status

In Figure 5, The cluster analysis revealed distinct patterns in child mortality distribution across the four identified clusters. Cluster 0, 2, and 3 predominantly contained cases of surviving children, while Cluster 1 uniquely captured infant mortality cases, suggesting potential risk factors specific to this group.

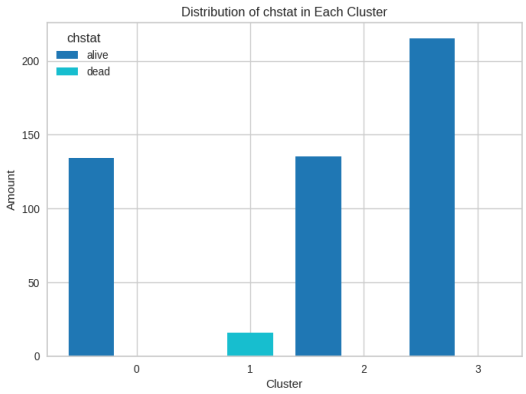

Fig. 5 Visualization of chstat Features Based on Cluster

Mortality Conditions Analysis (Conditions When a Baby Dies)

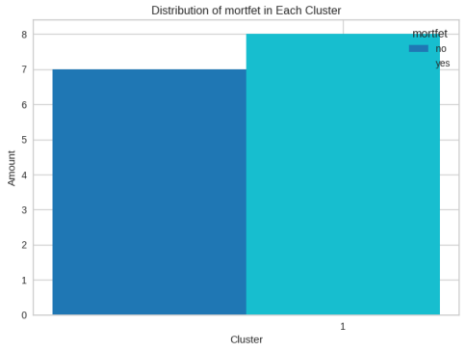

Fig. 6 mortfet Feature Visualization Based on Cluster

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14311
e-ISSN : 2541-2019
p-ISSN : 2541-044X

We get information from figure 6 about the condition of the babies who died. Further examination of Cluster 1 indicated that the majority of infant deaths occurred during the prenatal period, potentially highlighting the critical importance of maternal healthcare during pregnancy (experienced malnutrition, or other factors while in the womb)

Family Structure Impact (Family Type)

From Figure 7, we can see the difference in family types in each cluster. Cluster 0 is a collection of children who are raised in main families. While cluster 3 is a collection of children who are raised in extended families. However, cluster 2 is a balanced collection of children who are raised in main families and extended families.

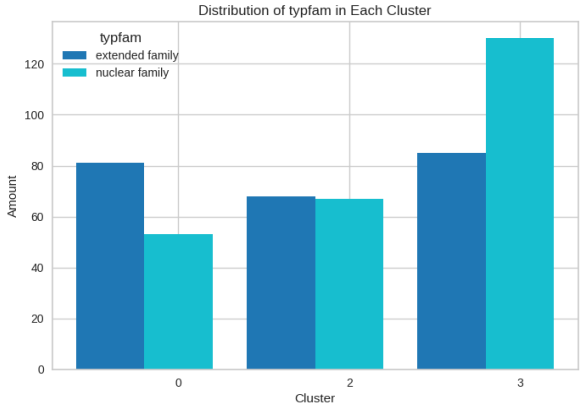

Fig. 7 Typefam Feature Visualization Based on Cluster

Socioeconomic Analysis (Family Income)

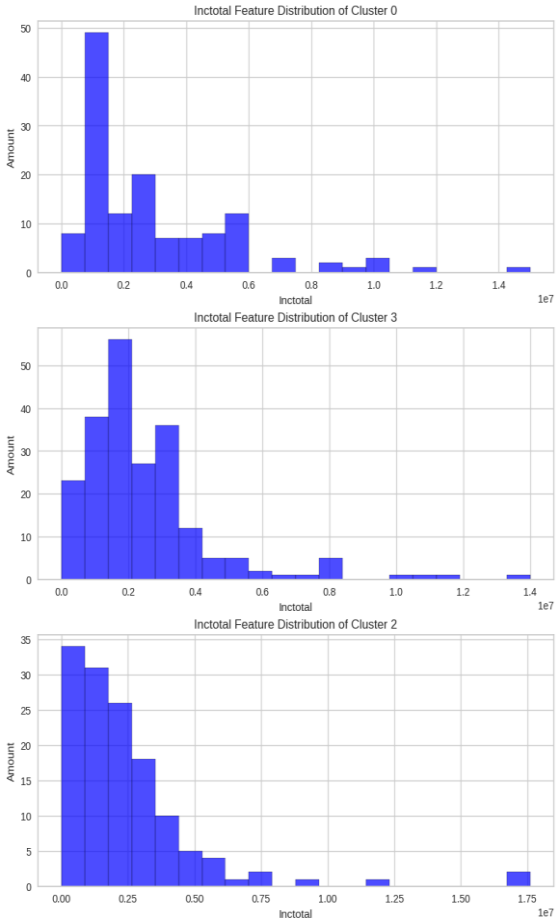

Fig. 8 Inctotal Feature Visualization Based on Cluster

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14311
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figure 8 explains that the economic stratification across clusters revealed distinct patterns: Cluster 0: Higher socioeconomic status with above-average family income; Cluster 2: Lower-middle income segment with economic constraints; Cluster 3: Middle-income group with stable financial conditions.

This economic distribution correlates significantly with child health outcomes, particularly in relation to stunting prevalence. The analysis suggests that economic factors may serve as key determinants in accessing adequate nutrition and healthcare services.

Healthcare Utilization Patterns (Child Health Checks at Posyandu)

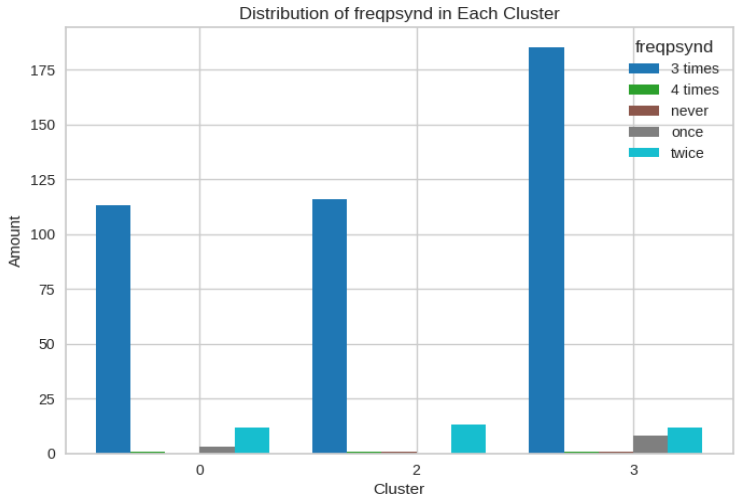

Fig. 9 Visualization of freqpsynd Features Based on Cluster

In Figure 9, it can be seen that most parents bring their children to the integrated health post 3 times. The figure also indicates parents' concern about the importance of integrated health posts. We can see a positive trend, namely that very few parents have never brought their children to the integrated health post. Thus, Analysis of integrated health post (Posyandu) visits revealed consistent healthcare-seeking behavior across clusters:

- The majority of families maintained regular visits (three times on average)
- Low prevalence of families with zero visits
- Consistent attendance patterns regardless of economic status

This finding indicates positive community engagement with primary healthcare services, suggesting effective healthcare accessibility despite socioeconomic variations.

Stunting Status

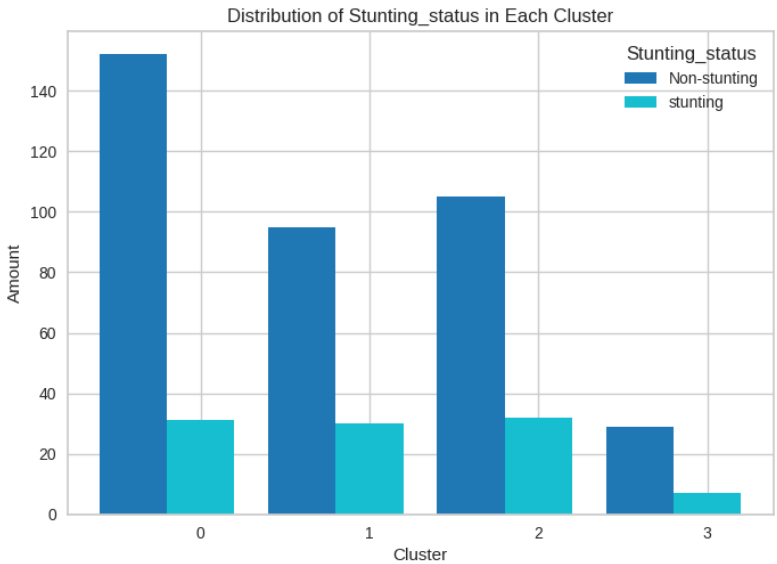

Fig. 10 Visualization of Stunting_status Feature Based on Cluster

e-ISSN : 2541-2019
p-ISSN : 2541-044X

When viewed in figure 10, it can be seen that Cluster 2 and Cluster 3 have higher stunting rates than Cluster 0. If we look at previous research, we know that stunting can be caused by a person's economic status. Citizens with lower middle economy can be given special attention to the dangers of stunting. By any means, The cluster analysis revealed notable patterns in stunting prevalence:

- Clusters 2 and 3 demonstrated higher stunting rates compared to Cluster 0
- Economic status showed strong correlation with stunting occurrence
- Lower-middle income groups (Cluster 2) exhibited increased vulnerability to stunting

These findings align with existing literature suggesting socioeconomic status as a critical factor in stunting prevention. Particularly noteworthy is the lower stunting prevalence in Cluster 0, characterized by:

a. Higher income levels
b. main family structure
c. Regular healthcare utilization

Regional Context and Cultural Considerations

The unique characteristics of Sambas Regency play a crucial role in shaping the context of our findings and their implications for stunting prevention. As a border region in West Kalimantan Province, Sambas presents distinct geographic challenges that influence various aspects of public health intervention. The region's location affects access to resources and services, potentially creating disparities in healthcare delivery systems across different areas. Additionally, its position as a border region may impact economic opportunities for families, which our clustering analysis has shown to be significantly correlated with stunting outcomes.

The cultural landscape of Sambas Regency adds another layer of complexity to our findings. Local customs and traditions significantly influence dietary habits, healthcare-seeking behavior, and family structure decisions. These cultural factors may explain some of the variations observed in our cluster analysis, particularly regarding family structures and their association with stunting rates. Traditional dietary practices, for instance, may affect nutritional intake patterns regardless of economic status, while cultural norms around family composition could influence resource distribution and childcare practices. Understanding these cultural nuances is crucial for developing effective interventions that will be well-received and sustainable within the community.

This regional and cultural context must be carefully considered when designing and implementing stunting prevention programs. While our clustering analysis provides valuable insights into risk factors and potential intervention points, the success of any prevention strategy will largely depend on its alignment with local cultural values and practical considerations of the border region setting. Future interventions should therefore strike a balance between evidence-based practices and cultural sensitivity, ensuring that programs are both effective and culturally appropriate for the unique context of Sambas Regency.

Policy Implications and Intervention Strategies

Our clustering analysis results provide valuable insights that suggest several key areas for policy intervention. The strong correlation between economic status and stunting prevalence, particularly evident in Clusters 2 and 3, indicates a need for targeted financial assistance programs for lower-middle income families. These could include nutritional subsidy programs for at-risk households and economic support initiatives aimed at improving household food security. Such interventions align with Paul et al.'s (Paul et al., 2021) findings on the effectiveness of economic interventions in reducing stunting rates among vulnerable populations.

The distinct patterns observed in family structures across clusters suggest the need for customized support programs based on family composition. Our findings indicate that extended families, particularly in Cluster 3, may benefit from specific resource optimization strategies and targeted nutritional education programs. These family-specific interventions should be designed to address the unique challenges faced by different household structures while maintaining cultural sensitivity. This approach is supported by Tola et al.'s (Tola et al., 2023) research, which emphasizes the importance of family-centered interventions in stunting prevention.

Despite the encouraging findings regarding Posyandu attendance across all clusters, our analysis suggests the need for enhancement of existing healthcare services. While basic healthcare accessibility appears strong, there is room for improvement in service quality and program scope. This could include strengthening nutritional monitoring programs and increasing focus on prenatal care services, particularly given the mortality patterns observed in Cluster 1. The implementation of these enhanced healthcare services should be coordinated with existing community health initiatives to ensure maximum effectiveness and resource utilization. This comprehensive approach to healthcare service enhancement aligns with Gansaonré et al.'s (Gansaonré et al., 2022) recommendations for integrated healthcare interventions in stunting prevention.

The regional context of Sambas Regency, particularly its position as a border region, necessitates careful consideration in policy implementation. Geographic factors may affect access to resources and services, while local cultural practices influence dietary habits and healthcare-seeking behavior. Therefore, any intervention

e-ISSN : 2541-2019
p-ISSN : 2541-044X

strategies must be adapted to account for these regional specificities while maintaining their fundamental effectiveness in addressing stunting risk factors.

Limitations and Future Works

Our study encountered several methodological limitations that should be considered when interpreting the results. The cross-sectional nature of our data collection limits our ability to establish causal relationships and track changes over time. Additionally, while our clustering analysis revealed significant patterns, there may be unmeasured confounding variables that influence stunting outcomes in ways not captured by our current dataset. The complexity of socio-cultural factors in Sambas Regency also presents challenges in generalizing our findings to other regions.

These limitations open up several promising avenues for future research. Longitudinal studies would be particularly valuable in tracking the effectiveness of interventions and understanding how family structure impacts child development over time. There is also a need for more detailed investigation of specific nutrition intervention outcomes, especially in the context of different family structures and economic conditions. Furthermore, the integration of cultural factors in stunting prevention strategies requires more in-depth study, particularly in border regions like Sambas Regency where cultural practices may significantly influence health outcomes.

From an analytical perspective, future studies could benefit from incorporating advanced machine learning techniques for predictive modeling of stunting risk factors. The development of risk assessment tools based on our clustering findings could aid healthcare providers in early identification of at-risk children. Additionally, integrating qualitative research methods with our quantitative approach could provide deeper insights into the cultural and behavioral aspects of stunting prevention. Such mixed-method approaches would be particularly valuable in understanding the complex interplay between socioeconomic factors, family dynamics, and child health outcomes.

CONCLUSION

This study employed Principal Component Analysis and K-means clustering to analyze stunting patterns in Sambas Regency, West Kalimantan Province. Through the Elbow method, four distinct clusters were identified, each revealing unique characteristics and risk factors associated with stunting. The analysis uncovered significant patterns in infant mortality, predominantly in Cluster 1, where prenatal deaths were most common, indicating the crucial role of maternal nutrition during pregnancy. The remaining clusters (0, 2, and 3) exhibited varying family structures and socioeconomic conditions that correlated strongly with stunting prevalence. A notable finding emerged from Cluster 0, characterized by nuclear family structures and higher socioeconomic status, which demonstrated significantly lower stunting rates. This suggests that the combination of family structure and economic stability plays a crucial role in reducing stunting risk. While healthcare utilization, particularly through regular Posyandu visits, remained consistent across all clusters, the economic disparities between clusters appeared to be a determining factor in stunting outcomes. This research faced limitations primarily in the comprehensive analysis of the extensive feature set, which complicated the profiling process. Future research opportunities lie in exploring additional stunting determinants, conducting more detailed analyses of specific risk factors, and examining the longitudinal effectiveness of interventions. The findings emphasize the importance of developing integrated intervention programs that address both economic factors and family structure in stunting prevention efforts. Such interventions should particularly focus on supporting lower-income families and providing targeted nutritional support during pregnancy to reduce stunting prevalence in the region.

ACKNOWLEDGMENT

This research has been conducted in collaboration with IDSS Research Center Faculty of Computer Science Universitas Dian Nuswantoro.

REFERENCES

Aguilera Vasquez, N., & Daher, J. (2019). Do nutrition and cash-based interventions and policies aimed at reducing stunting have an impact on economic development of low-and-middle-income countries? A systematic review. *BMC Public Health*, *19*(1), 1419. https://doi.org/10.1186/s12889-019-7677-1

Aheto, J. M. K. (2020). Simultaneous quantile regression and determinants of under-five severe chronic malnutrition in Ghana. *BMC Public Health*, *20*(1), 644. https://doi.org/10.1186/s12889-020-08782-7

Ahmed, M., Zepre, K., Lentero, K., Gebremariam, T., Jemal, Z., Wondimu, A., Bedewi, J., Melis, T., & Gebremeskel, A. (2022a). The relationship between maternal employment and stunting among 6–59 months old children in Gurage Zone Southern Nation Nationality People's region, Ethiopia: A

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14311
e-ISSN : 2541-2019
p-ISSN : 2541-044X

comparative cross-sectional study. *Frontiers in Nutrition*, *9*, 964124. https://doi.org/10.3389/fnut.2022.964124

Ahmed, M., Zepre, K., Lentero, K., Gebremariam, T., Jemal, Z., Wondimu, A., Bedewi, J., Melis, T., & Gebremeskel, A. (2022b). The relationship between maternal employment and stunting among 6–59 months old children in Gurage Zone Southern Nation Nationality People's region, Ethiopia: A comparative cross-sectional study. *Frontiers in Nutrition*, *9*, 964124. https://doi.org/10.3389/fnut.2022.964124

Ariza Colpas, P., Vicario, E., De-La-Hoz-Franco, E., Pineres-Melo, M., Oviedo-Carrascal, A., & Patara, F. (2020). Unsupervised Human Activity Recognition Using the Clustering Approach: A Review. *Sensors*, *20*(9), 2702. https://doi.org/10.3390/s20092702

Chen, L., Zhong, C., & Zhang, Z. (2023). Explanation of clustering result based on multi-objective optimization. *PLOS ONE*, *18*(10), e0292960. https://doi.org/10.1371/journal.pone.0292960

Gansaonré, R. J., Moore, L., Bleau, L., Kobiané, J., & Haddad, S. (2022). Stunting, age at school entry and academic performance in developing countries: A systematic review and meta-analysis. *Acta Paediatrica*, *111*(10), 1853–1861. https://doi.org/10.1111/apa.16449

Gebreayohanes, M., & Dessie, A. (2022). Prevalence of stunting and its associated factors among children 6–59 months of age in pastoralist community, Northeast Ethiopia: A community-based cross-sectional study. *PLOS ONE*, *17*(2), e0256722. https://doi.org/10.1371/journal.pone.0256722

Janssen, D., Rechberger, S., Wouters, E., Schols, J., Johnson, M., Currow, D., Curtis, J., & Spruit, M. (2019). Clustering of 27,525,663 Death Records from the United States Based on Health Conditions Associated with Death: An Example of Big Health Data Exploration. *Journal of Clinical Medicine*, *8*(7), 922. https://doi.org/10.3390/jcm8070922

Karuppusami, R., Antonisamy, B., & Premkumar, P. S. (2022). Functional principal component analysis for identifying the child growth pattern using longitudinal birth cohort data. *BMC Medical Research Methodology*, *22*(1), 76. https://doi.org/10.1186/s12874-022-01566-0

Liao, J., Scholes, S., Mawditt, C., Mejía, S. T., & Lu, W. (2022). Comparing relationships between health-related behaviour clustering and episodic memory trajectories in the United States of America and England: A longitudinal study. *BMC Public Health*, *22*(1), 1367. https://doi.org/10.1186/s12889-022-13785-7

Massara, P., Lopez-Dominguez, L., Bourdon, C., Bassani, D. G., Keown-Stoneman, C. D. G., Birken, C. S., Maguire, J. L., Santos, I. S., Matijasevich, A., Bandsma, R. H. J., & Comelli, E. M. (2023). A novel systematic pipeline for increased predictability and explainability of growth patterns in children using trajectory features. *International Journal of Medical Informatics*, *177*, 105143. https://doi.org/10.1016/j.ijmedinf.2023.105143

Mokalla, T. R., & Rao Mendu, V. V. (2022). Application of quantile regression to examine changes in the distribution of Height for Age (HAZ) of Indian children aged 0–36 months using four rounds of NFHS data. *PLOS ONE*, *17*(5), e0265877. https://doi.org/10.1371/journal.pone.0265877

Namirembe, G., Ghosh, S., Ausman, L. M., Shrestha, R., Zaharia, S., Bashaasha, B., Kabunga, N., Agaba, E., Mezzano, J., & Webb, P. (2022). Child stunting starts in utero: Growth trajectories and determinants in Ugandan infants. *Maternal & Child Nutrition*, *18*(3), e13359. https://doi.org/10.1111/mcn.13359

Nemerimana, M., Havugarurema, S., Nshimyiryo, A., Karambizi, A. C., Kirk, C. M., Beck, K., Gégout, C., Anderson, T., Bigirumwami, O., Ubarijoro, J. M., Ngamije, P. K., & Miller, A. C. (2023). Factors associated with recovery from stunting at 24 months of age among infants and young children enrolled in the Pediatric Development Clinic (PDC): A retrospective cohort study in rural Rwanda. *PLOS ONE*, *18*(7), e0283504. https://doi.org/10.1371/journal.pone.0283504

Paul, P., Arra, B., Hakobyan, M., Hovhannisyan, M. G., & Kauhanen, J. (2021). The determinants of under-5 age children malnutrition and the differences in the distribution of stunting–A study from Armenia. *PLOS ONE*, *16*(5), e0249776. https://doi.org/10.1371/journal.pone.0249776

Rahman, S., Johnson, V. E., & Rao, S. S. (2022). A Hyperparameter-Free, Fast and Efficient Framework to Detect Clusters From Limited Samples Based on Ultra High-Dimensional Features. *IEEE Access*, *10*, 116844–116857. https://doi.org/10.1109/ACCESS.2022.3218800

Rahman, S. M. J., Ahmed, N. A. M. F., Abedin, Md. M., Ahammed, B., Ali, M., Rahman, Md. J., & Maniruzzaman, Md. (2021). Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. *PLOS ONE*, *16*(6), e0253172. https://doi.org/10.1371/journal.pone.0253172

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Sartika, A. N., Khoirunnisa, M., Meiyetriani, E., Ermayani, E., Pramesthi, I. L., & Nur Ananda, A. J. (2021). Prenatal and postnatal determinants of stunting at age 0–11 months: A cross-sectional study in Indonesia. *PLOS ONE*, *16*(7), e0254662. https://doi.org/10.1371/journal.pone.0254662

Satoh, M. (2022). Clustering of health behaviors among Japanese adults and their association with socio-demographics and happiness. *PLOS ONE*, *17*(4), e0266009. https://doi.org/10.1371/journal.pone.0266009

Takele, B. A., Gezie, L. D., & Alamneh, T. S. (2022). Pooled prevalence of stunting and associated factors among children aged 6–59 months in Sub-Saharan Africa countries: A Bayesian multilevel approach. *PLOS ONE*, *17*(10), e0275889. https://doi.org/10.1371/journal.pone.0275889

Takele, K., Zewotir, T., & Ndanguza, D. (2019). Understanding correlates of child stunting in Ethiopia using generalized linear mixed models. *BMC Public Health*, *19*(1), 626. https://doi.org/10.1186/s12889-019-6984-x

Tola, G., Kassa, A., Getu, M., Dibaba, B., & Neggesse, S. (2023). Prevalence of stunting and associated factors among neonates in Shebadino woreda, Sidama region South Ethiopia; a community-based cross-sectional study 2022. *BMC Pediatrics*, *23*(1), 276. https://doi.org/10.1186/s12887-023-04080-4

Vennam, T., Agnihotri, S., & Chinnasamy, P. (2020). Examining Spatial Dependency in Child Malnutrition Across 640 Districts in India for Context Specific Planning and Intervention. *Current Developments in Nutrition*, *4*, nzaa053_124. https://doi.org/10.1093/cdn/nzaa053_124

Whitaker, V., Oldham, M., Boyd, J., Fairbrother, H., Curtis, P., Meier, P., & Holmes, J. (2021). Clustering of health-related behaviours within children aged 11–16: A systematic review. *BMC Public Health*, *21*(1), 137. https://doi.org/10.1186/s12889-020-10140-6

Yu, B., Xu, R., Cai, M., & Ding, W. (2024). A clustering method based on multi-positive–negative granularity and attenuation-diffusion pattern. *Information Fusion*, *103*, 102137. https://doi.org/10.1016/j.inffus.2023.102137