Determination of characteristics of stunting risk families in Subulussalam, Aceh in 2022 using k-modes cluster analysis

Muhammad Iqbal[1], *Hizir* Sofyan[2], *Tarmizi* Usman[3], and *Marzuki* Marzuki[2*]

[1]The Regional Development Planning Agency, Subulussalam City, 24781 Subulussalam, Indonesia
[2]Statistics Department, Universitas Syiah Kuala, 23111 Banda Aceh, Indonesia
[3]Mathematics Department, Universitas Syiah Kuala, 23111 Banda Aceh, Indonesia

Abstract. Stunting is a disruption in the growth and development of children caused by nutritional deficiencies and recurrent infections. Stunting can lead to a decline in intelligence levels and lower educational achievements. The aim of this research is to determine the characteristics of families at risk of stunting in the Subulussalam City. This study utilizes data from the Family Census conducted by the National Family Planning Coordination Board (BKKBN) in 2021. The data consists of 8,954 households at risk of stunting, using 10 variables for stunting risk calculation, such as households with children under two years old, under five years old, households with a fertile-aged wife, households without access to safe drinking water, and others. This study employs the k-modes cluster analysis method, which involves grouping a dataset based on specific variables into k clusters. The k-modes algorithm randomly selects k objects as cluster centers and calculates the distance of each object to the cluster modes. Households at risk of stunting in Subulussalam City mostly exhibit characteristics of having a fertile-aged wife and having too many children. Most households at risk of stunting already have access to safe drinking water and proper toilets. Clustering using k-modes resulted in an optimal k model with $k = 2$ (SW/SB ratio) and $k = 5$ (R-Squared). The model with $k = 2$ revealed significant differences in the variables of households with children under two years old and ownership of proper toilets.

1 Background

Stunting is a condition of impaired growth and development in children caused by recurring malnutrition and infections. Stunting is characterized by height or length below the established standards set by the government health authority.

The decline in intelligence levels, developmental delays, lower educational achievements, cognitive development, motor skills development, language development, and socio-emotional development are some consequences of stunting. If left untreated, stunting can lead to obesity, increased cardiovascular risks, and metabolic syndromes [1,2, 3, 4 5].

*Corresponding author: marzuki@usk.ac.id

According to the 2022 Indonesia Nutrition Status Survey (SSGI) by the Ministry of Health [6], Aceh is the fifth-highest province in terms of stunting prevalence in Indonesia, reaching 31.2 percent. At the provincial level, Subulussalam City in Aceh ranks first with a staggering 47.9 percent. This alarming rate necessitates research with the aim of determining the characteristics of families at risk of stunting in Subulussalam City. The findings are expected to provide valuable insights to the government in efforts to reduce the prevalence of stunting, particularly in the Aceh province.

With the high prevalence of stunting, interventions are needed for families at risk of stunting, obtained through the 2021 Family Census (PK21) conducted by the National Family Planning Coordination Board (BKKBN). These interventions are carried out by categorizing families at risk of stunting, allowing for more effective and efficient policy decisions. The suitable statistical method for categorical data from PK21 is the K-Mode Cluster method. This method is appropriate for identifying cluster centers using the modes of a dataset.

2 Research Methodology

The data used in this study are derived from the Family Census conducted by the National Family Planning Coordinating Board (BKKBN) in 2021 (PK21). The dataset comprises 8,954 households at risk of stunting, along with 10 variables for calculating the risk of stunting. The ten variables are as follows:

- Households with Children Under Two Years Old;
- Households with Children Under Five Years Old;
- Households with Wives of Reproductive Age (15-49 Years);
- Households with Wives of Reproductive Age and Pregnant;
- Households without Adequate Sanitation Facilities/Proper Toilets;
- Households without Access to Safe Drinking Water;
- Households with Young Wives (Wife's Age <20 Years);
- Households with Older Wives (Wife's Age 35-40 Years);
- Households with Children Born Too Close Together (Inter-child spacing <2 Years); and
- Households with Too Many Children (Total children 3 or more).

The method used in this study is k-modes cluster analysis. K-modes analysis is a clustering analysis of a dataset based on specific variables into k groups [7]. K-modes is an extension of k-means cluster analysis with changes in:

- Differences in distance measurement calculations;
- Calculation of cluster centers based on k-means becomes k-modes;
- The use of modes as a measure of calculation [8].

The k-modes algorithm is as follows [9]:

1. Randomly selecting k objects as cluster centroids;
2. Calculating the distance between each object and the cluster modes [10];
   The distance measurement calculation for k-modes uses the Hamming equation. The equation for the Hamming distance function between two objects (X_i) and (X_j) is as follows:
   $$d(X_i, X_j) = \sum_{i=1,j=1}^{m} \delta(x_i, x_j) \quad (1)$$
   where
   $$\delta(x_i, x_j) = \begin{cases} 0, x_i = x_j \\ 1, x_i \neq x_j \end{cases} \quad (2)$$
   with m being the number of variables.
3. Grouping objects into clusters that have the shortest distance to the cluster centroids chosen in point 1;
4. Selecting new modes for each cluster and comparing them with the previous modes. If there is a difference, points 2 and 3 are repeated. If there is no difference, then stop.

The measurement of the optimal number of clusters is conducted using two methods, namely the ratio of the within-cluster standard deviation (SW) to the between-cluster standard deviation (SB) or the SW/SB ratio, and R-Squared [9]. A smaller SW/SB ratio indicates a more optimal cluster model [11]. As for R-Squared, the closer it is to the value of "1", the more optimal the cluster model generated [12].

All analyzes in this study used R-software and Ms. Excel. Meanwhile, the R package used is “klaR” with a set seed of 123.

3 Results and Discussion

3.1 Descriptive Analysis

The ten research variables are calculated as percentages for each category of "yes" and "no." The total percentages are presented in the following table.

Table 1. Percentage of households based on research variables.

Number	Variables	Yes		No	
		n	%	n	%
1	Households with Children Under Two Years Old	1796	18.23	8058	81.77
2	Households with Children Under Five Years Old	3587	36.40	6267	63.60
3	Households with Wives of Reproductive Age (15-49 Years)	9796	99.41	58	0.59
4	Households with Wives of Reproductive Age and Pregnant	523	5.31	9331	94.69
5	Households without Adequate Sanitation Facilities/Proper Toilets	1823	18.50	8031	81.50
6	Households without Access to Safe Drinking Water	1878	19.06	7976	80.94
7	Households with Young Wives (Wife's Age <20 Years)	157	1.59	9697	98.41
8	Households with Older Wives (Wife's Age 35-40 Years)	3684	37.39	6170	62.61
9	Households with Children Born Too Close Together (Inter-child spacing <2 Years)	262	2.66	9592	97.34
10	Households with Too Many Children (Total children 3 or more)	7300	74.08	2554	25.92

Most households at risk of stunting fall into the category of wives of reproductive age with too many children. This is in line with the research conducted by Karundeng et al. [13]. Additionally, the majority of households at risk of stunting already have access to safe drinking water and proper sanitation facilities. This indicates that despite the risk of stunting,

conditions related to access to safe drinking water and ownership of proper sanitation facilities are generally met. However, there is still 8.69 percent of the population without access to adequate drinking water and proper sanitation facilities. This aligns with the percentage of households in Subulussalam City in 2021, where only 85.39 percent have private toilets, while the rest use communal toilets (2.89 percent), public bath-toilet facilities (2.31 percent), do not use toilets (0.77 percent), and have no toilets (8.54 percent). The population still engages in open defecation practices, posing the potential risk of stunting in households (Aceh Province in Figures, 2021, 2022).

An interesting point is that around 74.08 percent of households at risk of stunting in Subulussalam City have 3 or more children. This figure indicates that the larger the household size, the higher the risk of a household being stunted. The number of children in a family affects food security within the family. Insufficient food intake due to a large family size is a contributing factor to the nutritional status.

Growth and development issues are more likely to be experienced by later-born children, as parents' burden increases with the growing number of children. The first child's needs are more likely to be met because the parents' burden is still light, allowing more attention and fulfillment of all the child's needs.

Parents' age at the time of having their first child is relatively young, so their stamina is still in good condition. However, for the third child and beyond, parents are relatively older, and their stamina decreases. Parents' age and physical stamina also affect their parenting style. Rufaida et al. [14] also revealed that having more than 2 children indirectly becomes a risk factor for stunting. This is related to the family's food availability [15].

3.2 Cluster Analysis

3.2.1 Determination of optimal cluster number

The determination of the optimal number of clusters is conducted using two metrics: the SW/SB ratio and R-Squared. The SW/SB ratio and R-Squared for each cluster group (k=1 to k=10) are as follows:

Table 2. The SW/SB ratio and R-Squared for each cluster group.

Number	Number of Clusters (k)	SW/SB Ratio	R-Square
1	2	0.0238	0.5498
2	3	0.0293	0.6157
3	4	0.0285	0.7185
4	5	0.0310	0.7419
5	6	0.0422	0.6609
6	7	0.0502	0.6232
7	8	0.0623	0.5560
8	9	0.0711	0.5233
9	10	0.1400	0.2421

The optimal number of clusters according to the SW/SB ratio is the model with k=2, with a SW/SB ratio value of 0.0238. Meanwhile, the optimal number of clusters according to R-Squared is the model with k=5, with an R-Squared value of 0.7419 or 74.19 percent.

3.2.2 Clustering families at risk of stunting with k=2

Most households at risk of stunting are in cluster 2 (81.05 percent). The differences in characteristics between each cluster are related to the variable of households with children under two years old and households without proper toilets. In cluster 1, households have children under two years old and do not have proper toilets. On the other hand, in cluster 2, households do not have children under two years old and already have proper toilets.

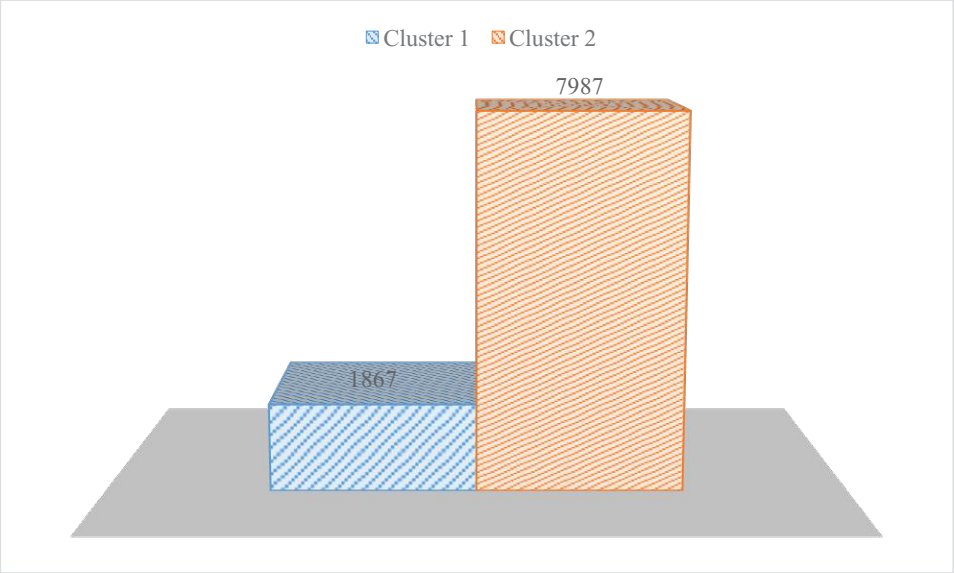

Fig. 1. The number of households in each cluster (k=2).

Table 3. The modes in each cluster.

Number	Variable	Cluster 1	Cluster 2
1	Households with Children Under Two Years Old	Yes	No
2	Households with Children Under Five Years Old	No	No
3	Households with Wives of Reproductive Age (15-49 Years)	Yes	Yes
4	Households with Wives of Reproductive Age and Pregnant	No	No
5	Households without Adequate Sanitation Facilities/Proper Toilets	Yes	No
6	Households without Access to Safe Drinking Water	No	No
7	Households with Young Wives (Wife's Age <20 Years)	No	No

8	Households with Older Wives (Wife's Age 35-40 Years)	No	No
9	Households with Children Born Too Close Together (Inter-child spacing <2 Years)	No	No
10	Households with Too Many Children (Total children 3 or more)	Yes	Yes

Serious attention is needed for cluster 1, especially regarding the variable of owning proper toilets. As a majority of the population lacks proper toilets, 64.27 percent of households in cluster 1 do not have proper toilets, which equals 1,200 households. Additionally, households in cluster 1 generally have children under two years old and an excessive number of children. This indicates that the risk of stunting in cluster 1 will persist for 1 to 2 years, contributing to the number of children under five years old with stunting status. Nutrition and dietary fulfillment for children under two years old in this cluster are necessary in preventing stunting during the under-five years old stage.

Research conducted by Sugianti et al. [16] found that the most influential nutritional intervention in preventing stunting is the provision of healthy toilet facilities. Furthermore, the age of children between 12-24 months is another variable that affects the occurrence of stunting. These findings reinforce the need to focus stunting reduction efforts on households with children under two years old. To reduce stunting in rural and food-insecure areas, programs to provide healthy toilets should be enhanced.

Several interventions for stunting in households in cluster 1 with children under two years old include increasing maternal awareness of stunting prevention behaviors, such as avoiding open defecation, handwashing with soap, managing household water and food, securing household waste, managing household liquid waste, maternal nutrition, infant and child feeding, and monitoring child growth [17]. In addition to maternal behavior, specific interventions by local governments can be implemented, such as improving the nutritional intake of chronically energy-deficient pregnant women, increasing the number of pregnant women consuming a minimum of 90 iron supplement tablets during pregnancy, adolescent girls consuming iron supplement tablets, infants under 6 months receiving exclusive breastfeeding, children aged 6-23 months receiving complementary feeding alongside breastfeeding, malnourished children under five receiving malnutrition management services, children under five being monitored for growth and development, malnourished children under five receiving additional nutritional intake, and children under five obtaining complete basic immunizations (Presidential Regulation Number 72 of 2021).

A study by Efendi et al. [18] revealed that children under two years old who do not receive exclusive breastfeeding, are not given appropriate complementary feeding, are not given vitamin A capsules, and have incomplete immunizations are more at risk of stunting. Factors affecting stunting in children under two years old include the provision of exclusive breastfeeding, complementary feeding, and the status of completing basic immunizations [19]. Meanwhile, a study by Talarima et al. [20] suggested that specific interventions for stunting should include exclusive breastfeeding, early initiation of breastfeeding, proper provision of complementary feeding alongside breastfeeding, and the management of infectious diseases in children under five years old.

3.2.3 Clustering families at risk of stunting with k=5

Most households at risk of stunting are in cluster 3 (48.45 percent). The differences in characteristics between each cluster are related to the variables of households with children

under two years old, households with children under five years old, lacking access to safe drinking water, households without proper toilets, and having too many children.

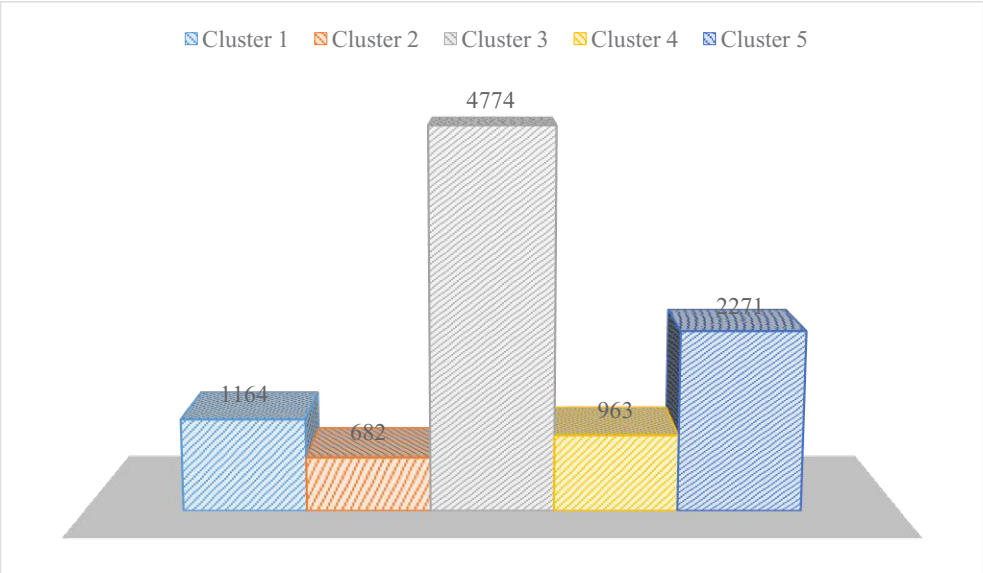

Fig. 2. The number of households in each cluster (k=5).

Table 4. The modes in each cluster.

Number	Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
1	Households with Children Under Two Years Old	No	No	No	Yes	No
2	Households with Children Under Five Years Old	No	Yes	No	Yes	No
3	Households with Wives of Reproductive Age (15-49 Years)	Yes	Yes	Yes	Yes	Yes
4	Households with Wives of Reproductive Age and Pregnant	No	No	No	No	No
5	Households without Adequate Sanitation Facilities/Proper Toilets	Yes	Yes	No	No	No
6	Households without Access to Safe Drinking Water	Yes	Yes	No	No	No
7	Households with Young Wives (Wife's Age <20 Years)	No	No	No	No	No
8	Households with Older Wives (Wife's Age 35-40 Years)	No	No	No	No	No
9	Households with Children Born Too Close Together (Inter-child spacing <2 Years)	No	No	No	No	No

10	Households with Too Many Children (Total children 3 or more)	Yes	No	No	No	No

The table above explains that only cluster 4 generally has children under two years old, while the other clusters generally do not have children under two years old. Clusters 2 and 4 generally have children under five years old, while the other clusters do not have children under five years old. Cluster 4 generally has children under two years old and children under five years old. Clusters 1 and 2 do not have access to safe drinking water and proper toilets, while the other clusters have proper toilets. Cluster 1 also has too many children.

The two cluster model results generated have their respective strengths and weaknesses. The Subulussalam City Government can use either cluster model as the basis for policy-making to reduce stunting rates in accordance with existing regulations. Both models can serve as alternative policy stages in development, such as focusing on specific time frames. Additionally, special attention needs to be given to the two main indicators in identifying households at risk of stunting, namely households with access to safe drinking water and proper sanitation facilities. This is crucial as over 20 percent of households in Subulussalam City still lack access to safe drinking water. Furthermore, adequate toilets are essential for proper sanitation, and more than 20 percent of households in Subulussalam City still lack proper sanitation facilities.

4 Conclusion

Households at risk of stunting in Subulussalam City mostly exhibit characteristics of wives of reproductive age with too many children. Additionally, the majority of households at risk of stunting already have access to safe drinking water and proper toilets.

The clustering using k-modes resulted in an optimal model with k=2 (SW/SB ratio) and k=5 (R-Squared). The model with k=2 showed significant differences in the variables of households with children under two years old and ownership of proper toilets. Cluster 1 presented serious issues in both variables and requires attention to reduce the risk of stunting. In cluster 1, 64.27 percent of households do not have proper toilets, totaling 1,200 households. Meanwhile, the model with k=5 produced differences in all five clusters, including the variables of households with children under two years old and under five years old, ownership of safe drinking water and proper toilets, and having too many children (3 or more).

The Subulussalam City Government can formulate policies by considering the planned development priorities, making the clustering of families at risk of stunting one of the alternatives. Dividing families at risk of stunting into several groups can assist the government in implementing development gradually in areas with stunting issues, taking into account resource and budget constraints. Additionally, the cluster models with k=2 and k=5 serve as considerations for policy stages, paying attention to the indicators of concern within each formed cluster.

References

1. M.A. Alam, S.A. Richard, S.M. Fahim, M. Mahfuz, B. Nahar, et al, PLoS One. **15**, 2 (2020)
2. D. Coetzee, Plessis, D. Wilmarie, D. Staden, J. Child. Educ. **10**, 1 (2020)
3. L.S. Adair, D.B. Carba, N.R. Lee, J.B. Borja, Econ. Hum. Biol. **42**, 100999 (2021)

4. J. Beckmann, C. Lang, R.d. Randt, A. Gresse, K.Z. Long, S. Ludyga, I. Müller, S. Nqweniso, U. Pühse, J. Utzinger, C. Walter, M. Gerber, Int. J. Environ. Res. Public Health **18**, 8 (2021)
5. A. Deshpande, R. Ramachandran, Econ. Hum. Biol. **44**, 101099 (2022)
6. Kemenkes R. I., Buku Saku Hasil Survei Status Gizi Indonesia (SSGI). **2022**, (2022)
7. S. C. Satapathy, P.S. Avadhani, S.K. Udgata, S. Lakshminarayana, *Advances in Intelligent Systems and Computing*, in ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India **248**, 137-144 (2014)
8. N.K. Visalakshi, K.A. Prabha, IJERD, **11,** 03 (2015)
9. H. Sofyan, M. Iqbal, M. Marzuki, M. Muhammad, *The comparison of k-modes clustering and ROCK clustering to the poverty indicator in Samadua Subdistrict, South Aceh*, in Proceedings The 10th Annual International Conference on Science and Engineering, IOP Conf. Series: Materials Science and Engineering, 15-16 October 2020, Banda Aceh, Indonesia (2020)
10. J.Z. Huang, Encyclopedia of Data Warehousing and Mining (IGI Global, Pennsylvania (2009)
11. A. Dewi, Sutikno, H. Kuswanto, Metode *Cluster* untuk Pengelompokan Desa Pedesaan di Provinsi Riau **4**, 1-16 (2012)
12. Q. Zhao, Cluster Validity in Vlustering Methods. (P. Pasanen & P. Kilpelainen, Eds.), Publications of the University of Eastern Finland, Dissertations in Forestry and Natural Sciences. Joensuu, Finland: University of Eastern Finland (2012)
13. L.R. Karundeng, A.Y. Ismayanto, R. Kundre, e-Kep **3**, 1 (2015)
14. F.D. Rufaida, A.M. Raharjo, A. Handoko, AMS, **6**, 1 (2020)
15. A. Candra, JNH, **1**, 1 (2013)
16. E. Sugianti, B.D. Putri, Amerta Nutrition **6**, 1 (2022)
17. F.K. Fildzah, A. Yamin, S. Hendrawati, J. Keperawatan Muhammadiyah **5**, 2 (2020)
18. N.F. Efendi, L. Sitoayu, R. Nuzriana, L.P. Dewanti, Y. Wahyuni, Ilmu Gizi Indonesia **5**, 1 (2021)
19. I. Hannum, A.J. Hadi, O. Wisudawan, H. Ahmad, Z. Nasution, Media Publikasi Promosi Kesehatan Indonesia **6**, 6 (2023)
20. B. Talarima, I.V. Lawalata, J. Endurance **8**, 2 (2023)