Information System for Evaluating Specific Interventions of Stunting Case Using K-means Clustering

Widya Sartika[1][*], *Suryono* Suryono[2], *Adi* Wibowo[3]

[1]Magister Program of Information System, School of Postgraduate Studies, Diponegoro University, Semarang- Indonesia
[2]Department of Physics, Faculty of Science and Mathematics, Diponegoro University, Semarang- Indonesia
[3]Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Semarang- Indonesia

Abstract. One of the causes of death of children under five is chronic malnutrition or stunting. The government has made a policy as an effort to reduce stunting. Specific intervention nutrition programs are activities that directly address the occurrence of stunting. Evaluation of edit interventions that have currently been carried out but have not separated the process based on intervention indicators so that it results in a long processing time and large exit costs so we need a system that can accelerate the process of evaluating specific interventions. Information systems for evaluating specific interventions can be a solution to this problem. In this study, an information system for evaluating specific interventions of stunting cases was developed web-based to facilitate real-time operation. The evaluation information system is designed to collect national data and carry out a clustering process using the k-means method and evaluation stucture cluster using silhouette method. In this study an evaluation value of 34 provinces was obtained which was divided into several clusters as a result of evaluation process in the information system.

1 Introduction

One form of chronic malnutrition that lasts for a long time due to failure to meet nutrition in infants is called stunting [1]. Stunting not only interferes with physical growth but more than that, stunting also interferes with brain development which can affect cognitive abilities, productivity and creativity [2]. Regional Health Research (RISKESDAS) in 2018 noted that Indonesia has a stunting prevalence rate of 30.8%, where the WHO annual target is no more than 20% [3]. In 2017 the National Team for the Acceleration of Poverty Reduction (TNP2K) by the Ministry of National Development Planning / Indonesian National Development Planning Agency has established priority areas for stunting intervention action plans. Starting from 34 provinces with 100 priority districts / cities based on specific and sensitive

---

[*] Corresponding author: widyakoto78@gmail.com

intervention indicators. Growth in age / height for the global average is surprisingly from the National Center for Health Statistics and the Cambridge reference, has experienced resistance after birth and continues thereafter and persisted into the third year. Conclusion. These findings highlight the need for prenatal and early life interventions to prevent growth failure [4].

Research on stunting has been carried out related to clustering analysis in the case of stunting in the study, the variables used are the length and age of the baby. The results of the study stated that 48% of infants experienced severe stunting, 22% of infants experienced stunting, 28% of normal infants and 2.5% of infants experienced abnormal. With 70% value of the accuracy of k-means clustering [2]. Further research with complex health data has been carried out to determine the right algorithm in getting the optimal group. The algorithm used is k-means and DBSCAN with silhouette as the evaluation method. The dataset used is a movement activity dataset from H Myhealtavatar. Analysis of the study shows that there is strong intra-cluster cohesion and inter-cluster separation where the k-means algorithm has better performance than DBSCAN in cluster accuracy and execution time [5].

Adaptation of the k-means clustering method to the evaluation information system has been carried out on the academic performance of students at a private university in Nigeria where the research aims to help academics make effective decisions for future academic planning. The results of these studies found 9 study programs that have the value of academic performance of students with good interpretation of a total of 79 students [6]. A research in Ethiopia has been conducted to look for interventions that reduce linear growth retardation (stunting) in children aged 6-36 months over a 5 year period in a food insecure population. This research was conducted with a cross-sectional survey using quantitative and qualitative tools to evaluate the impact of the interventions carried out. the results of study are that hygiene improvement practices have a significant impact 12%-1% on stunting levels in the area. in this study took 5 years to evaluate the interventions in the area [7].

The weakness of previous research is the absence of evaluations that focus on case-specific stunting interventions. where specific interventions are nutrition program activities which are directly aimed at stunting cases. besides that the length of time for evaluation and the amount of costs incurred for evaluation can be a limiting factor in efforts to reduce stunting rates. Based on this problem, we provide a solution with an evaluation information system that can accelerate the evaluation of the performance of nutrition programs for specific stunting case interventions that can assist government efforts in suppressing the prevalence of national stunting. Evaluation Information System was developed with a web-based system to facilitate the operation of the system and can be used in realtime. Furthermore, the evaluation process is automatically carried out by grouping of available datasets. the method used by the information system is the k-means clustering method. based on previous research the k-means algorithm is a partial algorithm that is widely used for pattern recognition with the nature of available data, ease of implementation, efficiency and empirical success [8]. Thereafter cluster structure evaluation will be carried out using the silhouette method to get the best cluster results.

2 Materials and Methods

2.1 Materials

In this study, we used problem data and nutrition program performance results that contained specific intervention indicators for stunting cases. We use data collected through from Status Monitoring National Nutrition book and Indonesia's Health Profile 2017. The

research data consisted of 16 variables obtained from problem data and nutrition program performance results for stunting. These variables consist of initiation of breastfeeding (IMD)> 1 hour and <1 hour, breast milk in the last 24 hours, exclusive breastfeeding, toddlers having a healthy going card (KMS), vitamin A for toddlers 6-59 months, thin toddlers getting food additional (PMT), weigh>=4 times, Pregnant women risk of chronic energy deficiency (KEK), women of childbearing age (WUS) risk of chronic energy deficiency (KEK), Pregnant women KEK can PMT, pregnant women can add blood tablets (TTD) >= 90 and <= 90, puerperal mothers get vitamin A>=2, young women get TTD, consume iodized salt. All of variable used enter the nutritional stunting intervention program indicators that have been designed by the government based on the manual for implementing integrated stunting reduction interventions in the district / city [9].

2.2 Methods

An evaluation technology includes research, improvement and application of testing methods and testing tools, the construction of a basic database and the preparation of related standard [10]. In this research, evaluation stage is carried out by conducting a clustering process on the performance data using k-means algorithm. The k-means algorithm is a partial algorithm that is widely used for pattern recognition with the nature of available data, ease of implementation, efficiency and empirical success [8]. Before entering into the k-means algorithm, the preprocessing data is performed first. Preprocessing data with Kaiser-Mayer Olkin test and Multikolinearity test using for clustering assumption. analysis of the sample to be used whether it can represent the population with a representative sample test [11]. the data used can represent the population if the results of the KMO (Kaiser-Mayer Olkin) table>0.50. The KMO table in this study is shown in table 1.

Table 1. KMO test.

Kaiser Meyer Olkin Test		
Size Of Sample Adequacy With Kmo		0,57
Barlett	Approx Chi-Square	510,638
	Df	91
	Sig.	0

From the KMO test results it can be seen that the KMO value in the dataset is 0.570 then the dataset has KMO assumption > 0.5. Furthermore, the non-multicollinearity test is performed to ensure that there are no independent variables that influence each other and seen from the value of Variance Inflating Factor (VIF) <10 [11]. The following VIF value can be seen in table 2 multicollinearity output.

Table 2. Multikolinearity

ID	Collinearity Statistic	
	Variable	VIF
X1	Initiation of breastfeeding (IMD)> 1 hour	3,034
X2	Initiation of breastfeeding (IMD)< 1 hour	3,233
X3	Breast milk in the last 24 hours	6,45

X4	Exclusive breastfeeding	5,923
X5	Toddlers having a healthy going card (KMS)	5,295
X6	Vitamin A for toddlers 6-59 months	4,492
X7	Thin toddlers getting food additional (PMT)	3,45
X8	Weigh> = 4 times	9,213
X10	Pregnant women KEK given PMT	3,899
X11	Women of childbearing age (WUS) risk of chronic energy deficiency (KEK)	3,758
X12	Pregnant women given blood tablets (TTD) < = 90	5,4
X13	pregnant women given blood tablets (TTD) < = 90	4,276
X14	Puerperal mothers get vitamin A> = 2	4,569
X15	Young women get TTD	1,526
X16	Consume iodized salt	2,905

Dependent Variabel : Pregnant Women Risk Of Chronic Energy Deficiency (Kek) (X9)

From Multikolineary test result it can be seen that value of VIF from all of variabel have <10 so the variables used in this study have met the assumptions of cluster analysis because there is no linear relationship between the independent variables. After preprocessing data. We can enter to k-means clustering algorithm as following below:

Step 1: Determine the number of clusters needed and the value of initial centroid using equation 1

$$v = \frac{\sum_{i}^{n} x_i}{n}; i = 1,2,3,..., \qquad (1)$$

In equation 1 where v is the centroid of the cluster, x_i is the i object and n is the number of objects / number of objects that are members of the cluster.

Step 2: Calculate the distance of data with center of the cluste with euclidean distance formula with the equation 2 below

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}; i = 1,2,3,...,n \qquad (2)$$

In equation (2) it is known that $d(x,y)$ is the distance of objecct x with each centroid y, where x is an object, y is a centroid, x_i is the i attribute value of object x, y_i is the attribute value to - i of the objects y and n are the number of attributes. Then the object is inserted into the cluster by looking at the shortest distance to the centroid. If there is still an object movement then iterate again by determining the new centroid position and doing the euclidean distance calculation using equation 2.

Step 3: After we get the results of the cluster analysis is performed to get the best cluster structure using the silhouette method like below.

$$SI(i) = \frac{b(i) - a(i)}{\max(a(i) - b(i))} \qquad (3)$$

The following SI (i) is a silhouette, a (i) is the average inequality i with all objects in clusters a, b (i) is the minimum value. The value of the silhouette coefficient results lies in the range of values -1 to 1. The more the value of the silhouette coefficient approaches the

value of 1, the better the grouping of data in one cluster. Conversely, if the silhouette coefficient approaches -1, the worse the grouping of data in one cluster.

Step 4: the last step is to profiling analysis of cluster characteristics. In this analysis we use the spearman correlation and the average value to get the characteristics and values of the clusters. Spearman correlation formula can be seen in the equation below.

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N^3 - N} \qquad (4)$$

in equation (4), d_i is the difference between the two rankings and N is the number of observations (4).

3 Implementation

Evaluation information systems are designed to collect datasets and run the k-means clustering algorithm automatically. In addition, the evaluation information system will also display spatial maps based on clusters and evaluation values based on cluster criteria analysis. The following are some interface views of the stunting specific intervention evaluation information system which can be seen in the figure below.
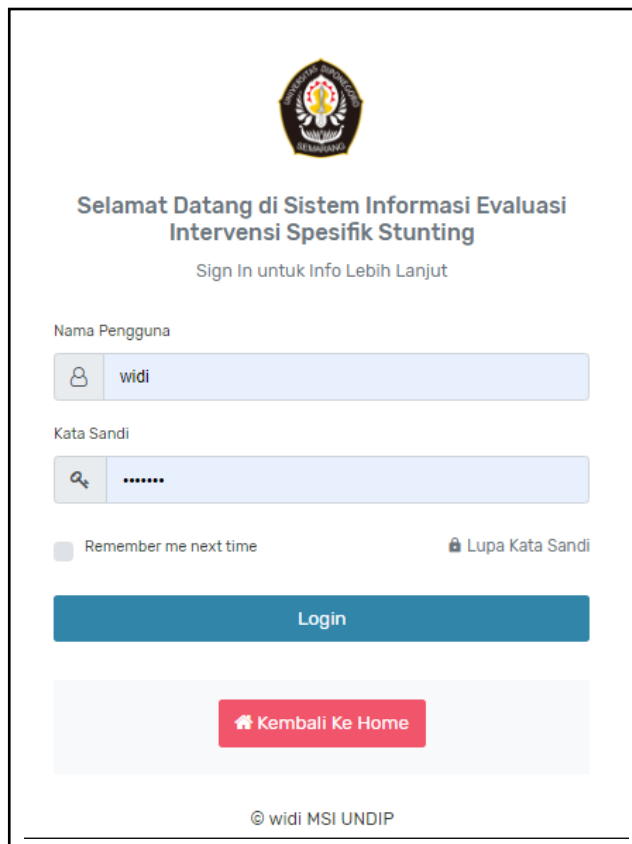

Fig. 1. Login interface

the process of k-means clustering analysis in the evaluation information system starts by entering data into the master data menu and then opening import data. Display the master data menu can be seen below.

Fig. 2. Master Data

In *klaster* menu the data will be processed with k-means. Fisrt step is desired number of clusters, in this research using k=2, k=3, k=4, k=5 and after that initial centroid will automatically come out from system. After that system will automatically calculate the shortest distance with Euclidean distance formula until iteration stop when member of cluster not moving anymore and the results of cluster will come out in *perhitungan* page. The following below displays *Perhitungan* page for the k-means process.

Fig. 3. *Perhitungan* page

4 Result and Discussion

The best cluster of this research can find with conducted several experiments using the silhouette method. Silhouette method is an analysis used to get the best cluster structure. then do a silhouette score search with the orange tools application. The results of the silhouette analysis can be seen in the table below.

Table 3. Silhouette score

	Silhouette score cluster
k=2	0,315
k=3	0,252
k=4	0,21
k=5	0,206

In table 3 the results of the analysis of the silhouette method show that cluster k = 2 with a score of 0.315 has the best cluster structure compared to the cluster 3 of the other cluster types.

After that based on table 2 the dataset in this research consist of independent variables is the reason for this study choosing cluster analysis using distance measurements so that the results of the cluster are not based on similarity patterns but based on the closest distance to the center of the cluster [12]. then the assessment of the characteristics of cluster members is done by looking at the level of similarity in the value of the weight of the cluster members. In cluster 1 the average value of the weight of cluster members is 0.64917 and the average value of the weight of the members of cluster 2 is 0.45343. the average value of cluster weights 1 and 2 can be seen in the table below.

Table 4. Results of analysis cluster 1

No.	Province	Cluster	Weight
1	DKI Jakarta	1	0,62625
2	Jawa Barat	1	0,603125
3	Kalimantan Selatan	1	0,610625
4	Nusa Tenggara Barat	1	0,606875
5	DI Yogyakarta	1	0,791875
6	Jawa Timur	1	0,6725
7	Banten	1	0,586875
8	Jawa Tengah	1	0,743125
9	Nusa Tenggara Timur	1	0,60125
	Average of weight cluster		**0,649167**
	Evaluation value		**Good**

Table 5. Results of analysis cluster 2

No.	Province	Cluster	Weight
1	Sulawesi Utara	2	0,41375
2	Sumatra Utara	2	0,306875
3	Bali	2	0,404375
4	Aceh	2	0,47
5	Jambi	2	0,495625

6	Bengkulu	2	0,31625
7	Kep Bangka Belitung	2	0,45625
8	Kepulauan Riau	2	0,536875
9	Kalimantan Timur	2	0,510625
10	Maluku	2	0,33625
11	Kalimantan Utara	2	0,535625
12	Gorontalo	2	0,538125
13	Maluku Utara	2	0,4375
14	Riau	2	0,3825
15	Kalimantan Tengah	2	0,426875
16	Sumatera Selatan	2	0,506875
17	Kalimantan Barat	2	0,47625
18	Sulawesi Barat	2	0,55
19	Sumatra Barat	2	0,519375
20	Sulawesi Tengah	2	0,4675
21	Sulawesi Tenggara	2	0,410625
22	Sulawesi Selatan	2	0,53
23	Papua Barat	2	0,4375
24	Lampung	2	0,526875
25	Papua	2	0,343125
	Average of weight cluster		0,453425
	Evaluation value		Less

Then spearman correlation analysis is performed to determine whether the research variables have a strong relationship with stunting cases so that they can be a reference in determining which clusters have better nutrition program performance. Spearman analysis results can be seen in Table 6 below.

Table 6. Analysis Spearman Correlation

No.	Variabel		value
1	Pregnant women KEK given PMT		+0,321
2	Women of childbearing age (WUS) risk of chronic energy deficiency (KEK)	**Stunting**	+0,316
3	Thin toddlers getting food additional (PMT)		-0,306
4	Toddlers having a healthy going card (KMS)		-0,236

5	Initiation of breastfeeding (IMD)< 1 hour	-0,226
6	Vitamin A for toddlers 6-59 months	+0,208
7	Pregnant women given blood tablets (TTD) < = 90	-0,191
8	Weigh> = 4 times	-0,168
9	pregnant women given blood tablets (TTD) > = 90	+0,149
10	Exclusive breastfeeding	-0,125
11	Pregnant women KEK given PMT	-0,121
12	Breast milk in the last 24 hours	-0,112
13	Young women get TTD	+0,090
14	Initiation of breastfeeding (IMD)> 1 hour	-0,075
15	Puerperal mothers get vitamin A> =2	+0,029
16	Consume iodized salt	+0,029

In Table 3 Spearman correlation analysis, it can be seen that the correlation value of specific nutrition program intervention variables with stunting cases is 0.321 which means that between the variables and stunting has a weak relationship. Therefore it is not fair to evaluate the performance of specific intervention nutrition programs based only on stunting cases that occur in cluster members. then to get the evaluation value sought factors that state the superiority of the two clusters and the results are based on the analysis of the average value of each variable. cluster 1 has a high value on 12 variables while cluster 2 only excels on 4 variables. then it can be said that cluster 1 is good and cluster 2 is less.

5 Conclusion

Based on this research it was concluded Information system for evaluating specific interventions of stunting cases was developed web-based with facilitate real-time operation. Information system for evaluating specific interventions of stunting cases help the evaluation of specific intervention programs to save time and costs because the process carried out with information systems is proven to be faster and easier. Evaluation value is not given based on the number of stunting cases because in this research all of the variables used has a low correlation value with stunting cases which is 0.321. The results of performance evaluation intervention specific stunting have good value for cluster 1 consisting of 9 provinces and less for cluster 2 consisting of 25 provinces.

References

[1] S. Kastro Dake, B. Solomon, T. M. Bobe, H. A. Tekle, and E. G. Tufa, "Predictors of stunting among children 6-59 months of age in Sodo Zuria District, South Ethiopia: a community based cross-sectional study," *BMC Nutr.*, (2019).

[2] P. W. Sudarmadji, C. Elena, and B. Bire, "Implementation Of K-Means Clustering Algoritm To Determine Stunted Status In Children Under Two Years Old," (2019).

[3] Ministry of Health Research and development Health, "RISKESDAS 2018," Republic of Indonesia, (2018).

[4] R. Shrimpton *et al.*, "Worldwide Timing of Growth Faltering: Implications for Nutritional Interventions," *Pediatrics*, vol. **107**, May (2001).

[5] G. Ogbuabor and F. . Ugwoke, "Clustering Algorithm for a Healthcare Dataset Using Silhouette Score Value," *Int. J. Comput. Sci. Inf. Technol.*, vol. **10**, no. 2, pp. 27–37, (2018).

[6] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, "Application of k-Means Clustering algorithm for prediction of Students' Academic Performance," *IJCSIS) Int. J. Comput. Sci. Inf. Secur.*, vol. **7**, no. 1, (2010).

[7] B. Fenn, A. T. Bulti, T. Nduna, A. Duffield, and F. Watson, "An evaluation of an operations research project to reduce childhood stunting in a food-insecure area in Ethiopia.," *Public Health Nutr.*, vol. **15**, no. 9, pp. 1746–1754, (2012).

[8] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. **31**, no. 8, pp. 651–666, (2010).

[9] Kementerian PPN / Bappenas, *Pedoman Pelaksanaan Intervensi Penurunan Stunting Terintegrasi Di Kabupaten/ Kota*. (2018).

[10] hao hao song, "Testing and Evaluation System for Cloud Computing Information Security Products," *Procedia Comput. Sci.*, pp. 84–87, (2020).

[11] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, "Multivariate Data Analysis," in *Pearson Prentice Hall*, 7th ed., PEARSON, (2010), pp. 455–458.

[12] C. A. Utomo, "Analisis Multivariat Clustering K-Means Untuk Apbd Kecamatan Di Kota Semarang Tahun 2015," Universitas Negeri Semarang, Semarang, (2017).