Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases

Indah Purnama Sari [1]*, Al-Khowarizmi[1], Oris Krianto Sulaiman[2], Dicky Apdilah[3]
[1]Department of Information Technology, Universitas Muhammadiyah Sumatera Utara, Indonesia
[2]Department of Informatics Engineering, Universitas Islam Sumatera Utara, Indonesia
[3]Department of Informatics Engineering, Universitas Asahan, Indonesia

ABSTRACT

Stunting is still a serious public health problem in Indonesia, where the prevalence of this condition is 37.2%, up from 35.6% in 2019 and 36.8% in 2020. The length or height of a child who is short (dwarf) is below average for his age. Stunting has a negative impact on IQ deficiencies, infectious diseases, mental health problems, and child development. Toddlers with stunting cases are detected when their growth and development does not match their age, but currently there is no data grouping based on these criteria that is of concern to parents and posyandu cadres. Data can be grouped using the K-Means data mining technique. The K-Means algorithm is often used by researchers as a grouping procedure to ascertain whether children are stunted or not. 395 datasets are used in this research data. The Knowledge Discovery In Databases (KDD) approach, a comprehensive nontrivial procedure for detecting and recognizing patterns in data, underlies this research. Based on the variables of age, weight and height, this study aims to identify groups or clusters of stunting status in children under five. The best number of clusters with K = 2 was determined by the findings of this investigation. There are 392 children in cluster 0-Shanum, Rizka, Nurjanah, and others-and three toddlers in cluster 1-Ezra, M. Abidza, and Abd Mahmud. With a total of 287 stunted toddlers and 108 toddlers with normal development, the most ideal DBI value is 0.007 which is close to 0, this shows that the clusters under review provide quality clusters.

Keyword : K-Means Algorithm; Data Mining; KDD; Clusterization; Stunting.

Corresponding Author:
Indah Purnama Sari,
Department of Information Technology
Universitas Muhammadiyah Sumatera Utara
Jl. Kapten Mukhtar Basri No 3 Medan, 20238, Indonesia.
Email : indahpurnama@umsu.ac.id

Article history:
Received Jul 15th 2023
Revised Aug 1st 2023
Accepted Aug 28th 2023

1. INTRODUCTION

In Indonesia with a stunting prevalence of 37.2%, up from 35.6% in 2019 and 36.8% in 2020, the majority of the population is still affected. The Indonesian Ministry of Health estimates that the stunting prevalence rate will reach 38.9% in 2020 (Indraputra, 2020). Stunted toddlers (kerdil) are shorter or taller than average for their age. The World Health Organization (WHO) average growth standards for children are used to establish whether a person's height or length is abnormally high or low (Indraputra, 2020). Stunting, a physical growth problem characterized by slow growth rates, is caused by unbalanced nutrition (Jiwandono, et al., 2021). Children can become stunted during pregnancy, childbirth, breastfeeding, or the postpartum period. such as supplementary feeding, which does not provide adequate nutrition to toddlers (Nabila, et al., 2021). A measurement technique known as "Anthropometry" can be used to assess the stunting status of children under five. Age (U), weight (BW), and height (TB) are anthropometric categories (Rahmawati, 2023).

The K-Medoids algorithm can be used to classify stunted toddlers in Indonesia into high and low clusters, with 28 provinces having the highest cluster and 6 provinces having the lowest cluster, according to research (Pascalina, et al., 2022) on the application of this algorithm to categorize stunted toddlers in Indonesia. Using the K-means Clustering technique based on an additional study from the article (Isnain, et al., 2022), the nutritional values of under-fives were divided into groups based on the factors of height and weight of under-fives, including obesity, over-nutrition, good nutrition, poor nutrition, and malnutrition. Using the height and weight parameters, the K-means clustering algorithm was used to divide the nutritional values of children under five into five categories: obesity, overnutrition, good nutrition, undernutrition, and malnutrition. As can be seen from some of the descriptions above, the two studies are relevant to this final project because they both discuss stunting cases in toddlers. However, this research is different from the others because it uses the K-means algorithm and selects the best cluster based on the smallest DBI value closest to 0.

The purpose of this research is to find out how to group the status of stunted children based on age, weight, and height variables which are then divided into two categories, namely Normal and Stunting. Therefore, the K-Means Clustering approach is used in this final project. Due to its simplicity and efficiency in grouping very large data based on processing speed by organizing items into classes that have similarities, clustering is a very popular and most commonly used method in data processing. K-Means clustering approach to determine the stunting status of under-fives. Due to its effectiveness and simplicity, the K-Means clustering method has become one of the most popular clustering algorithms (Dwi, et al., 2019). 395 datasets were used in this research data. The Knowledge Discovery in Databases (KDD) approach, a comprehensive non-trivial process for discovering and identifying patterns in data, was used to assist this final project (Sari, et al., 2021). The patterns discovered in this project are valid, novel, valuable, and understandable.

2. RESEARCH METHOD/MATERIAL AND METHOD/LETERATURE REVIEW

A. Research Stages

The stages in this research are described in the flowchart below:

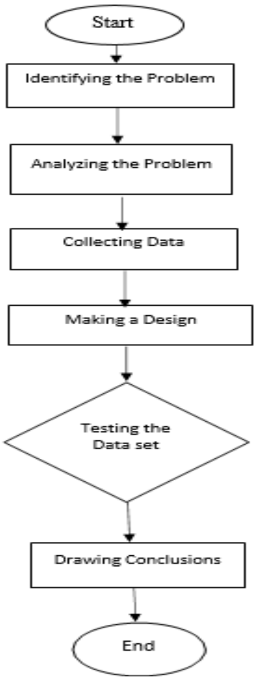


Figure 1. Flowchart of Research Stages

The description of the research stages carried out based on the stages in the figure above is:

a. Problem Identification
Problem identification is the first step in research that aims to classify data using the K-Means algorithm.

b. Problem Analysis
At this stage, an analysis of the needs of the research object is carried out, as well as analyzing the elements needed by the research object.

c. Collecting Data
The data collection technique used in this study, namely the observation technique used in this study is a passive observation technique, where the researcher comes directly to the Posyandu implementation site to observe the activities carried out but is not directly involved in these activities. Then the results of weighing records from each Posyandu are taken as research material.

d. Making a Design
The research stages cover the research steps from start to finish, in this study the authors used the Knowledge Discovery in Databases data mining stage in data processing. The phases start from raw data and end withknowledge or information that has been processed, which is obtained as a result of the following stages in Figure 2:

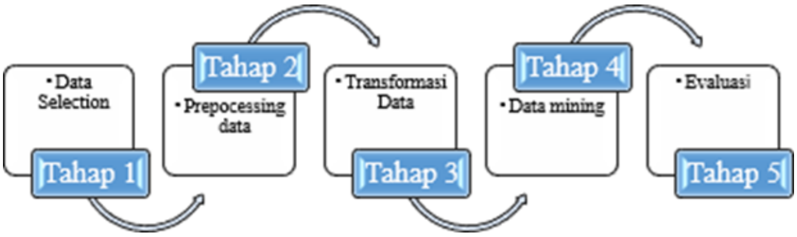
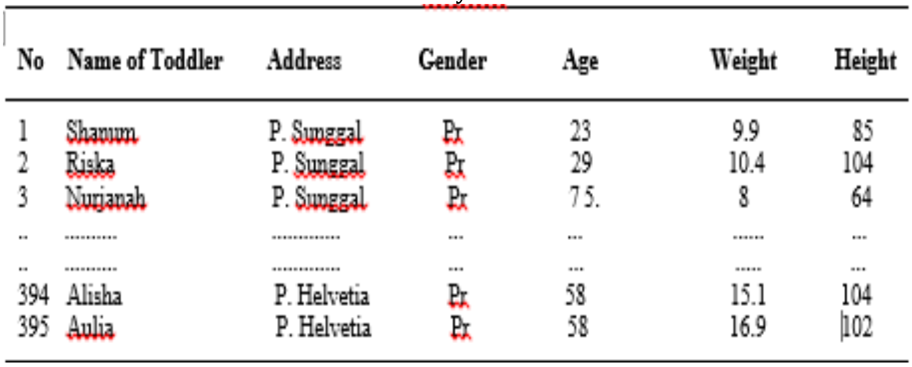


Figure 2. Data Mining Stages Using KDD

e. Testing Data Set
The data retrieval process used in the data clustering process, which seeks to select or select the collected data, is the initial stage. The dataset used is the Posyandu weighing data observed in January 2023 with the help of Microsoft Excel software. A total of 395 records with 7 attributes make up the collection. The following attributes are listed in Table 1: No, Toddler Name, Address, Gender, Age, Weight, and Height.

Table 1. Posyandu Dataset



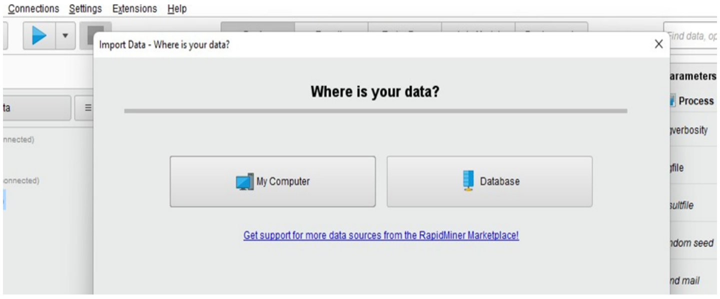
No	Name of Toddler	Address	Gender	Age	Weight	Height
1	Shanum	P. Sunggal	Pr	23	9.9	85
2	Riska	P. Sunggal	Pr	29	10.4	104
3	Nurjanah	P. Sunggal	Pr	7.5.	8	64
..	....	....	...	...	......	...
..	....	....	...	...	......	...
394	Alisha	P. Helvetia	Pr	58	15.1	104
395	Aulia	P. Helvetia	Pr	58	16.9	102

The pick Attribute operator in the RapidMiner program was used to select data for this investigation. The organized data set was imported into RapidMiner before the data selection procedure. Creating a process sheet that includes applying the algorithm and testing using the test methods offered by the RapidMiner operators is another thing to do. The explanation below shows how to perform these actions.

f. Conclusion
This stage is the stage of making conclusions from all the steps taken in designing a good system.

B. Import Data

Data in *.xls format is first imported into RapidMiner by clicking "Add Data" and selecting the storage location of the dataset to be used, before the K-Means method is used. Select the dataset location from "My Computer" as the dataset is a file with *.xls extension and not derived from a database such as SQL. The steps involved in data import are illustrated in Figures 3 and 4 below.



Import Data - Where is your data?

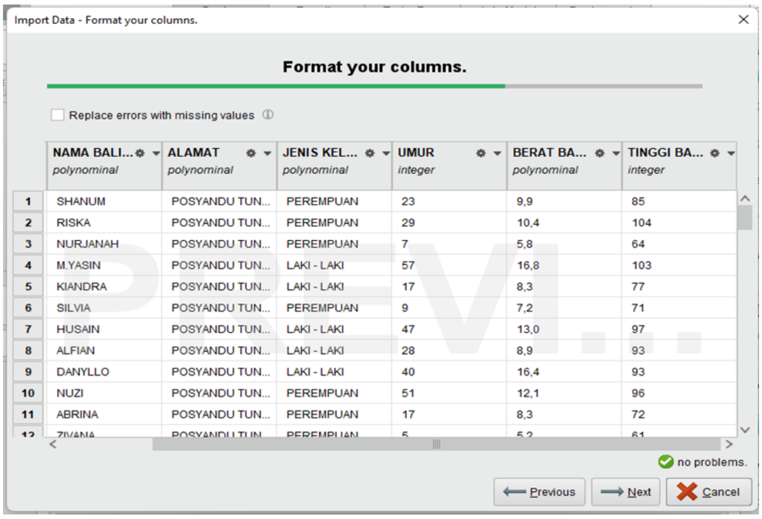
Where is your data?

My Computer | Database

Get support for more data sources from the RapidMiner Marketplace!

Figure 3. Import Data

Select the RapidMiner data import option next. The data import process will display "no problem" if the selected data already shows no problem. then do the next steps to ensure that RapidMiner can import the data correctly.



Import Data - Format your columns.

Format your columns.

□ Replace errors with missing values

	NAMA BALI... polynominal	ALAMAT polynominal	JENIS KEL... polynominal	UMUR integer	BERAT BA... polynominal	TINGGI BA... integer
1	SHANUM	POSYANDU TUN...	PEREMPUAN	23	9,9	85
2	RISKA	POSYANDU TUN...	PEREMPUAN	29	10,4	104
3	NURJANAH	POSYANDU TUN...	PEREMPUAN	7	5,8	64
4	M.YASIN	POSYANDU TUN...	LAKI - LAKI	57	16,8	103
5	KIANDRA	POSYANDU TUN...	LAKI - LAKI	17	8,3	77
6	SILVIA	POSYANDU TUN...	PEREMPUAN	9	7,2	71
7	HUSAIN	POSYANDU TUN...	LAKI - LAKI	47	13,0	97
8	ALFIAN	POSYANDU TUN...	LAKI - LAKI	28	8,9	93
9	DANYLLO	POSYANDU TUN...	LAKI - LAKI	40	16,4	93
10	NUZI	POSYANDU TUN...	PEREMPUAN	51	12,1	96
11	ABRINA	POSYANDU TUN...	PEREMPUAN	17	8,3	72
12	ZIVANA	POSYANDU TUN...	PEREMPUAN	5	5,2	61

✓ no problems.

Figure 4. Data Import Process

In this data import process, an id or identity is also added to the dataset to be used, namely the Toddler Name attribute is changed by 'Clicking' the settings logo then changing the "change role type" in "Click" to "id".

C. Creating Process Sheets

After the data is imported, the data is entered into the Process page. By dragging the imported data. An example of a data file used in this final project is Tegalwangi Village Stunting Data. Then Drag and Drop down the data to the process page. Figure 5 below is the implementation into RapidMiner.

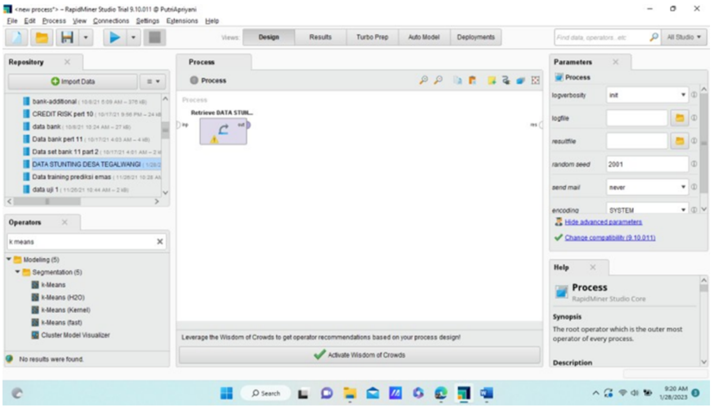


Figure 5. Retrive data that has been imported

D. Select Attributes

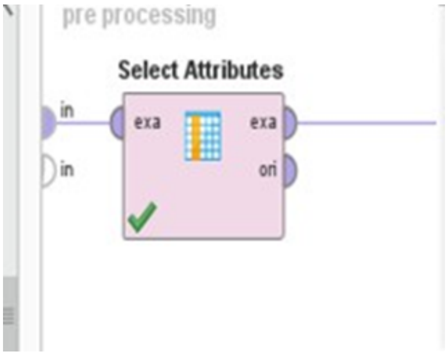
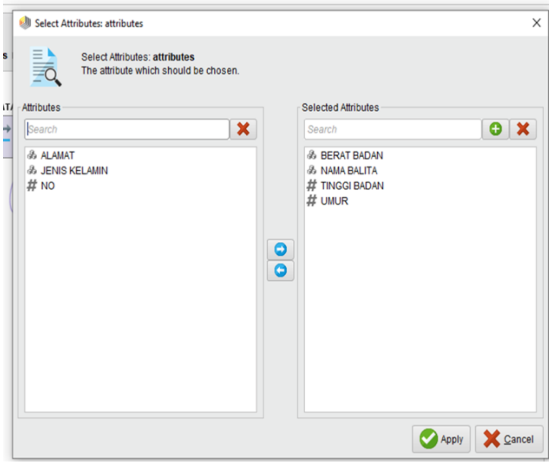


Figure 6. Attribute Select Operator

In the initial Tegalwangi Village posyandu dataset there are 7 attributes. These attributes include No, Toddler Name, Address, Gender, Age, Weight and Height. The Select Attribute operator on the RapidMiner tool will select the attributes to be used by selecting attributes in the form of subsets in the Select Attribute operator parameter. The data selection process is shown in Figure 6 and Figure 7.

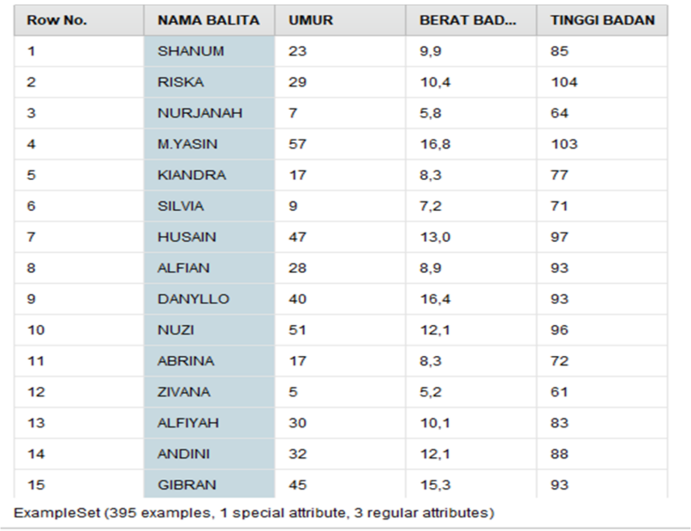


Select Attributes: attributes

Select Attributes: **attributes**
The attribute which should be chosen.

Attributes	Selected Attributes
ALAMAT	BERAT BADAN
JENIS KELAMIN	NAMA BALITA
NO	TINGGI BADAN
	UMUR

Figure 7. Attribute Select Operator Parameters



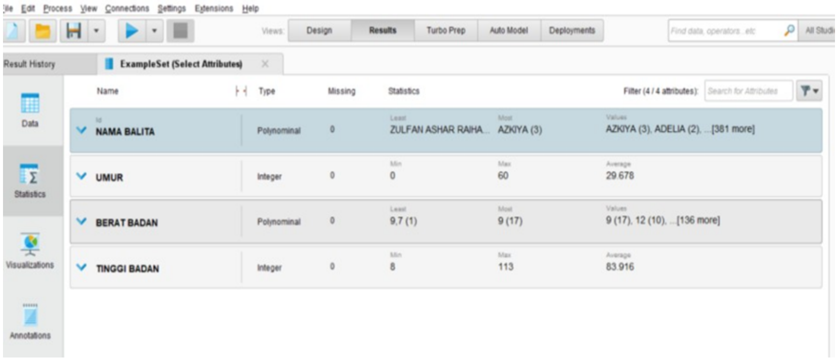
Row No.	NAMA BALITA	UMUR	BERAT BAD...	TINGGI BADAN
1	SHANUM	23	9,9	85
2	RISKA	29	10,4	104
3	NURJANAH	7	5,8	64
4	M.YASIN	57	16,8	103
5	KIANDRA	17	8,3	77
6	SILVIA	9	7,2	71
7	HUSAIN	47	13,0	97
8	ALFIAN	28	8,9	93
9	DANYLLO	40	16,4	93
10	NUZI	51	12,1	96
11	ABRINA	17	8,3	72
12	ZIVANA	5	5,2	61
13	ALFIYAH	30	10,1	83
14	ANDINI	32	12,1	88
15	GIBRAN	45	15,3	93

ExampleSet (395 examples, 1 special attribute, 3 regular attributes)

Figure 8. Data that has been selected

E. Data Pre-Processing (Data cleaning)

Data pre-processing involves cleaning data whose variables do not match with missing values from the calculations. As can be seen from the Statistics findings on the RapidMiner tool in Figure 9 below, at this point, no missing data or No Missing values were found on the data set.

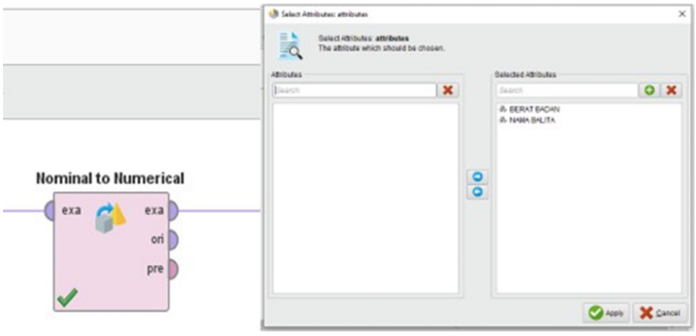

File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators...etc
All Studio
Result History
ExampleSet (Select Attributes)
Data
Statistics
Visualizations
Annotations
Filter (4 / 4 attributes):
Search for Attributes

Name	Type	Missing	Statistics		
Id NAMA BALITA	Polynominal	0	Least ZULFAN ASHAR RAIHA...	Most AZKIYA (3)	Values AZKIYA (3), ADELIA (2), ...[381 more]
UMUR	Integer	0	Min 0	Max 60	Average 29.678
BERAT BADAN	Polynominal	0	Least 9,7 (1)	Most 9 (17)	Values 9 (17), 12 (10), ...[136 more]
TINGGI BADAN	Integer	0	Min 8	Max 113	Average 83.916

Figure 9. Missing Value Checking

F. Transformation Data

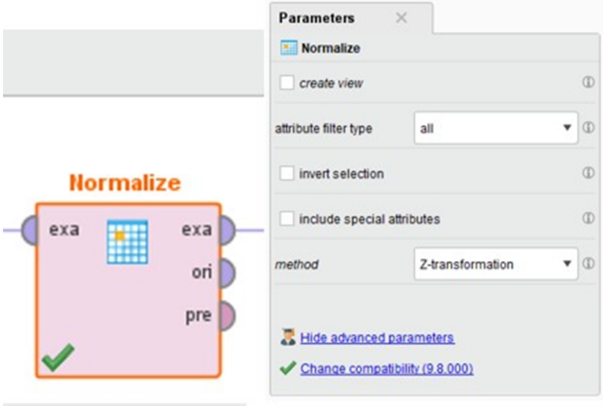
Nominal to Numeric operator and Normalization operator are used in the data transformation step to prepare the dataset that has gone through the previous phase to be analyzed and processed using K-Means Clustering method. In this final project dataset, the DBI value is normalized first before clustering with the K-Means algorithm. Figures 10 and 11 show the data transformation procedure.



Select Attributes: **attributes**
The attribute which should be chosen.

Selected Attributes:
BERAT BADAN
NAMA BALITA

Figure 10. Nominal to Numerical Operator Parameters



Normalize

□ create view

attribute filter type: all

□ invert selection

□ include special attributes

method: Z-transformation

Figure 11. Nomalize Operator Parameters

G. Data Mining

The K-Means approach was used in the data mining stage of this research during the clustering phase to identify stunting cases among children under five in Tegalwangi Village. The RapidMiner tool was also used. The Kmeans clustering operator and the distance performance clustering operator for the DBI value evaluation method were the operators used at this stage.

H. Evaluation

To make the patterns created during the data mining process easier to understand and ensure that the information found does not contradict the hypotheses developed earlier, scatter and column bar visualizations were used to present the patterns.

3. RESULTS AND DISCUSSION

A. Results of Application of K-Means Clustering Algorithm with RapidMiner Tools

To determine the clustering of the created process, use the K-Means algorithm on the process that has been built below. K-Means is the clustering operator used. Run the procedure to get the clustering result after you have finished creating the model. The optimal process model determined by the K-Means algorithm is shown in Figure 12 below.

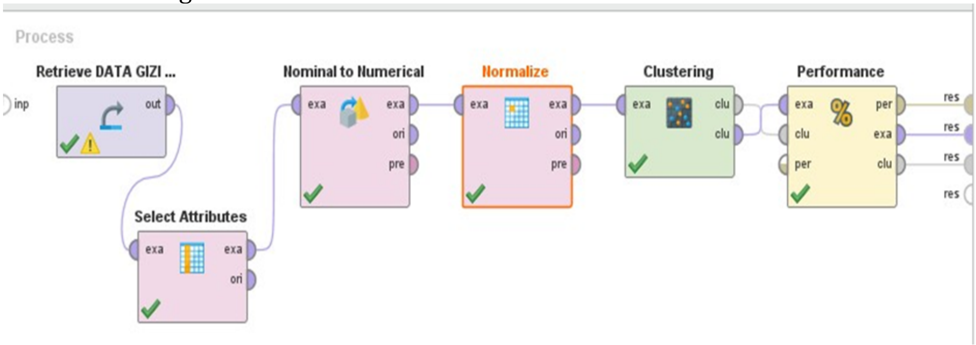
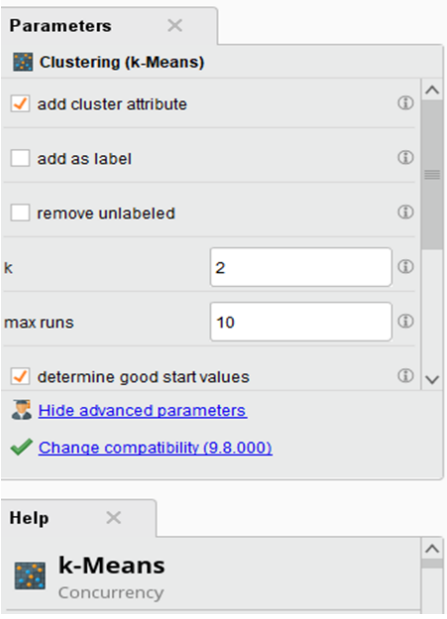


Figure 12. Applying K-Means

Next, set up the K-Means algorithm, which is set in the K-Means Clustering Parametere menu, as shown in Figure 13 below.


Parameters

Clustering (k-Means)

☑ add cluster attribute

□ add as label

□ remove unlabeled

k: 2

max runs: 10

☑ determine good start values

Hide advanced parameters

Change compatibility (9.8.000)

Help

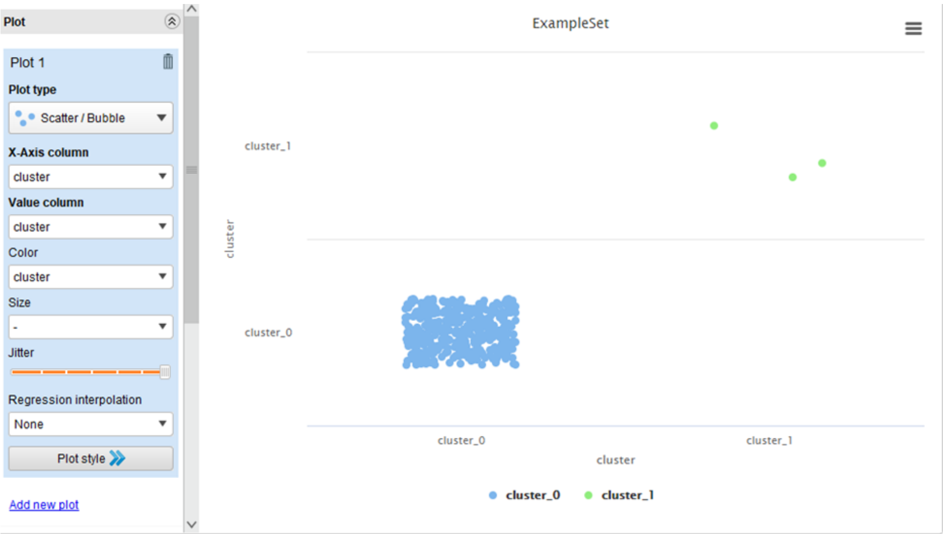
k-Means

Concurrency

Figure 13. Cluster Determination Parameters

This is the procedure of entering the desired K value, as shown in Figure 13. In this final project, the input k values, namely k = 2,3,4,5,6,7,8,9, are repeated ten times to identify the cluster with the best DBI value, which is close to 0. Run the procedure again to get the clustering results from applying the RapidMiner tool's K-Means technique. The best K value of k = 2 is obtained in the iterations that have been performed, then the cluster results are obtained using the K-Means technique.

There are many outputs generated by RapidMiner 9.10 software that are included in the data testing findings. The line showing the number of cluster groups of stunting incidence in Tegalwangi Village is visible in the visualization image. The image below shows the visualization on the left for the x-axis, y-axis, and selected custom color (cluster). We can see on this graph who received clustering positions at level 0 and 1.



Plot

Plot 1

Plot type: Scatter / Bubble

X-Axis column: cluster

Value column: cluster

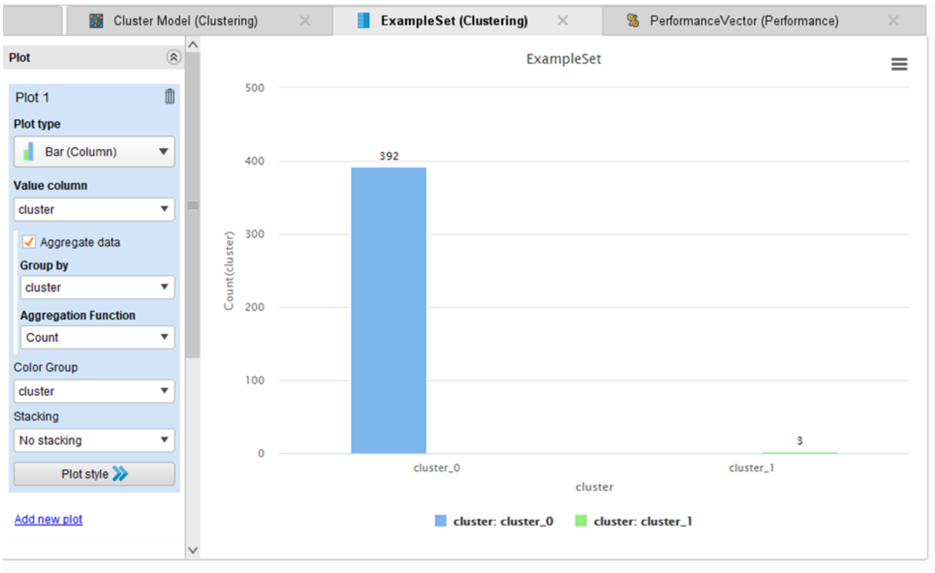
Color: cluster

Size: -

Jitter

Regression interpolation: None

Figure 14. Cluster Visualization with Scatter



Plot

Plot 1

Plot type: Bar (Column)

Value column: cluster

☑ Aggregate data

Group by: cluster

Aggregation Function: Count

Color Group: cluster

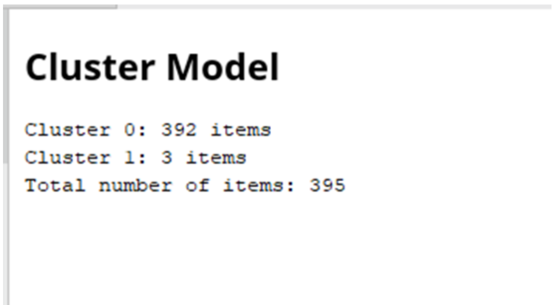
Stacking: No stacking

Figure 15. Visualization with Colum Bar

The final iteration of the K-Means program for clustering data aggregation is shown in Figures 14 and 15 above. Based on the test results above, cluster 0 has 392 toddlers indicated in blue, while cluster 1 has 3 toddlers indicated in green.

B. Cluster Model (Clustering)

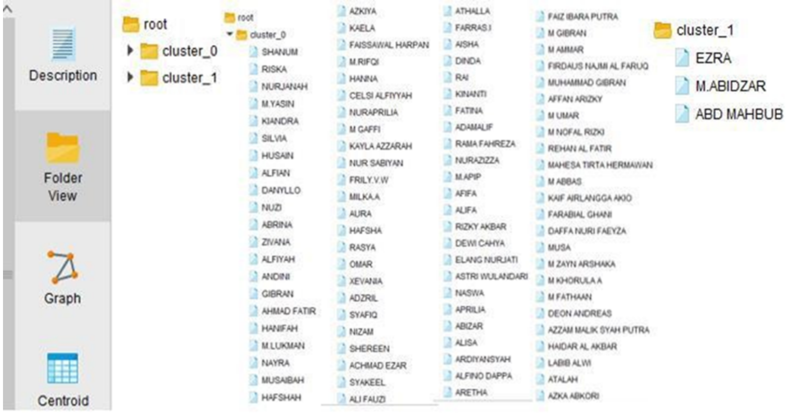
There are many ways to view cluster results in the Cluster Model (clustering), including a text view that shows the results of grouping by cluster. Cluster 1 has three items in it, while the number of cluster members is 392. You will see the text view in Figure 16.



Cluster Model

```
Cluster 0: 392 items
Cluster 1: 3 items
Total number of items: 395
```

Figure 16. Text View Display



root
cluster_0
cluster_1

root
cluster_0
SHANUM
RISKA
NURJANAH
M.YASIN
KIANDRA
SILVIA
HUSAIN
ALFIAN
DANYLLO
NUZI
ABRINA
ZIVANA
ALFIYAH
ANDINI
GIBRAN
AHMAD FATIR
HANIFAH
M.LUKMAN
NAYRA
MUSAIBAH
HAFSHAH
AZKIYA
KAELA
FAISSAWAL HARPAN
M.RIFQI
HANNA
CELSI ALFIYYAH
NURAPRILIA
M GAFFI
KAYLA AZZARAH
NUR SABIYAN
FRILY.V.W
MILKA.A
AURA
HAFSHA
RASYA
OMAR
XEVANIA
ADZRIL
SYAFIQ
NIZAM
SHEREEN
ACHMAD EZAR
SYAKEEL
ALI FAUZI
ATHALLA
FARRAS.I
AISHA
DINDA
RAI
KINANTI
FATINA
ADAMALIF
RAMA FAHREZA
NURAZIZZA
M.APIP
AFIFA
ALIFA
RIZKY AKBAR
DEWI CAHYA
ELANG NURJATI
ASTRI WULANDARI
NASWA
APRILIA
ABIZAR
ALISA
ARDIYANSYAH
ALFINO DAPPA
ARETHA
FAIZ IBARA PUTRA
M GIBRAN
M AMMAR
FIRDAUS NAJMI AL FARUQ
MUHAMMAD GIBRAN
AFFAN ARIZKY
M UMAR
M NOFAL RIZKI
REHAN AL FATIR
MAHESA TIRTA HERMAWAN
M ABBAS
KAIF AIRLANGGA AKIO
FARABIAL GHANI
DAFFA NURI FAEYZA
MUSA
M ZAYN ARSHAKA
M KHORULA A
M FATHAAN
DEON ANDREAS
AZZAM MALIK SYAH PUTRA
HAIDAR AL AKBAR
LABIB ALWI
ATALAH
AZKA ABKORI

cluster_1
EZRA
M.ABIDZAR
ABD MAHBUB

Figure 17. Folder View

In accordance with the image in Figure 17, each member of the two clusters displays the name of the toddler in the folder view which is a display of the cluster component data as a whole. The clustering of stunting cases has been implemented using the RapidMiner tool. The optimal height for children is shown in Table 2 below based on guidelines from the Indonesian Ministry of Health and WHO (World Health Organization).

Table 2. Anthropometric Standards for Ideal Height by Age

No.	Anthropometric Standards (Ideal Height by Age)
1	Baby 0 - 3 months old: Height 40.4 - 60 cm
2	4 - 6 month old baby: Height 60.5 - 66.0 cm
3	7 - 9 month old baby: Height: 67.5 - 70.5 cm
4	10 - 12 month old baby: Height: 72 - 74.5 cm
5	Toddlers 13 - 24 months old: Height: 82 - 92 cm
6	Ages 25 - 32 Months: Height: 83 - 95 cm
7	Toddler Age 33 - 44 Months: Height: 84 - 97 cm
8	Toddlers 44 - 58 months old: Height: 85 - 98 cm

Based on Table 3 below, it can be concluded that information on Normal and Stunting status in toddlers from 395 data obtained members as follows.

Table3. Information on the number of normal and stunted toddlers

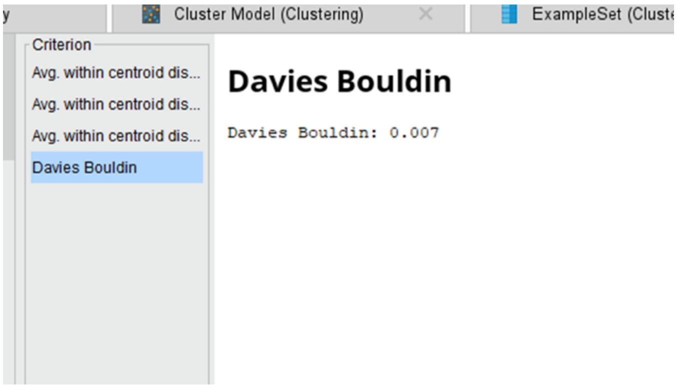
Cluster	Anggota Cluster	Informasi
C0	392 Balita	Jumlah Balita Normal: 285 Balita Jumlah Balita Stunting:107 Balita
C1	3 Balita	Jumlah Balita Normal: 2 Balita Jumlah Balita Stunting: 1 Balita
Jumlah		Jumlah total balita Normal: 287 balita Jumlah total balita Stunting: 108 balita

C. Optimized Cluster Evaluation Results using K-Means algorithm and DBI calculation

Based on Table 4 below, through the use of the k-means clustering method, and assisted by using the Daviesbouldin index (DBI) calculation to determine the most optimal cluster. There were 10 iteration trials to determine the best DBI value.

Table 4. Davies-bouldin index (DBI) values

K-Means		
K	Avg. within centroid distance	DBI
2	0,990	0,007
3	0,982	0,028
4	0,968	0,029
5	0,961	0,024
6	0,954	0,022
7	0,945	0,027
8	0,939	0,017
9	0,932	0,018
10	0,924	0,020


Cluster Model (Clustering)
ExampleSet (Clust
Criterion
Avg. within centroid dis...
Avg. within centroid dis...
Avg. within centroid dis...
Davies Bouldin

Davies Bouldin

```
Davies Bouldin: 0.007
```

Figure 18. 2nd Iteration DBI Value

Information about the DBI value and clustering performance is shown in Figure 18. The test data used calculations with the RapidMiner tool to find the Davies-bouldin index (DBI) value. The best DBI result with the value of K=2 is 0.007 close to 0, which indicates that the better the cluster, the smaller the DBI value obtained (non-negative > = 0). based on the K-Means clustering used. for computational results using the tool from RapidMiner.

4. CONCLUSION

These findings were achieved as a result of the research that had been conducted. Based on the findings of data analysis using the RapidMiner tool, stunting cases in toddlers in Tegalwangi Village as of the end of January 2023 were divided into 2 clusters based on age, weight, and height. Cluster 0 contained 392 toddlers, including Shanum, Rizka, Nurjanah, and others, while cluster 1 contained 3 toddlers, including Ezra, M Abidza, and Abd Mahmud. According to the ideal standards of toddlers, namely anthropometric standards, there are 287 toddlers with normal status and 108 toddlers with stunting status. As a result, parents need help from the posyandu and other relevant health centers to reduce and even eliminate the number of stunted toddlers in the coming months. Based on the evaluation of the Davies Bouldin Index (DBI) value, the calculation with the RapidMiner tool produces an optimal performance evaluation value at K = 2 with a value of 0.007 where the value is close to 0, which indicates that the cluster under review produces a good cluster.

REFERENCES

Sari, I.P., Al-Khowarizmi, AK., and Batubara, I.H. (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering For Grouping Students' Abilities In Online Learning Process. Journal of Computer Science, Information Technology and Telecommunication Engineering, 139-144.

Sari, I.P., Batubara, I.H, and Al-Khowarizmi, AK. (2021). Sensitivity Of Obtaining Errors In The Combination Of Fuzzy And Neural Networks For Conducting Student Assessment On E-Learning. International Journal of Economic, Technology and Social Sciences (Injects), 331-338.

G. Apriluana and S. Fikawati, "Analysis of Risk Factors for the Incidence of Stunting in Toddlers (0-59 Months) in Developing Countries and Southeast Asia," Health Research and Development Media, vol. 28, no. 4, pp. 247-256, Dec 2018, doi: 10.22435/mpk.v28i4.472.

T. Prasetiya, I. Ali, C. L. Rohmat, and O. Nurdiawan, "Classification of Toddler Stunting Status in Slangit Village Using the K-Nearest Neighbor Method," INFORMATICS FOR EDUCATORS AND PROFESSIONALS, vol. 4, no. 2, pp. 93-104, 2020.

Sari, I.P., and Batubara, I.H. (2021). Optimization of the FP-Growth Algorithm in Data Mining Techniques to Get the Electric Power Theft Pattern for the Development of Smart City. 2021 4th International Conference of Computer and Informatics Engineering (IC2IE), 293-298.

Ramadhani, F., Satria, A., and Sari, I.P. (2023). Implementation of Fuzzy K-Nearest Neighbor Method in Dengue Fever Disease Classification. Hello World Journal of Computer Science. 58-62.

Sari, I.P., Al-Khowarizmi, AK., Ramadhani, F., and Sulaiman, O.K. (2023). Implementation of the Selection Sort Algorithm to Sort Data in PHP Programming Language. Journal of Computer Science, Information Technology and Telecommunication Engineering.

Ramadhani, F., and Sari, I.P. (2021). Improving the Performance of Naïve Bayes Algorithm by Reducing the Attributes of Dataset Using Gain Ratio and Adaboost. 2021 International Conference on Computer Science and Engineering (IC2SE), 1-5.

H. Pohan et al., "Application of the K-Medoids Algorithm in Grouping Stunting Toddlers in Indonesia," JUKI: Journal of Computers and Informatics, 2021.

E. Irfiani, S. Sulistia Rani, J. Kamal Raya No, R. Road Barat Cengkareng West Jakarta, S. Nusa Mandiri Jl Kramat Raya No, and J. Pusat, "K-Means Clustering Algorithm to Determine Toddler Nutrition Value," 2018.

et al Gustientiedina, "Application of K-Means Algorithm for Clustering Drug Data at. RSUD Pekanbaru," 2018. Accessed: January 20, 2023. [Online]. Available at: http://repository.potensi-utama.ac.id/jspui/bitstream/123456789/4745/3/BAB%20II.pdf

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. "Knowledge Discovery and Data Mining: Towards a Unifying Framework," 1996. [Online]. Available at: www.aaai.org

Dwi, A., Yadika, N., Berawi, K.N., and Nasution, S.H. "The Effect of Stunting on Cognitive Development and Learning Achievement," 2019.

Ministry of Health of the Republic of Indonesia 2020, "REGULATION OF THE MINISTER OF HEALTH OF THE REPUBLIC OF INDONESIA STANDARDS FOR CHILD ANTROPOMETRY."

Indraputra, R.A., and Fitriana, R. "K-Means Clustering COVID-19 Data," Journal of Industrial Engineering, 2020.

Jiwandono, A.G. "Analysis of BPJS Health Class Grouping Using the K-Means Method," 2021.

Nabila, Z., Isnain, A.R., and Abidin, Z. "DATA MINING ANALYSIS FOR CLUSTERING COVID-19 CASES IN LAMPUNG PROVINCE WITH K-MEANS ALGORITHMA," Journal of Information Technology and Systems (JTSI), vol. 2, no. 2, pp. 100, 2021, [Online]. Available at: http://jim.teknokrat.ac.id/index.php/JTSI.

Rahmawati, "Determining the Welfare Level of Central Kalimantan Province by Applying the K-Means Clustering Algorithm Using Rapidminer," 2023.

Pascalina, D., Widhiastono, R., and Juliane, C. "Measuring the Readiness of Smart City Digital Transformation Using Rapid Miner Application," Technomedia Journal, vol. 7, no. 3, pp. 293-302, Dec 2022, doi: 10.33050/tmj.v7i3.1914.

Beal, T., Tumilowicz, A., Sutrisna, A., Izwardy, D., & Neufeld, L. M. (2018). A review of child stunting determinants in Indonesia. Maternal and Child Nutrition, 14(4), 1-10. https://doi.org/10.1111/mcn.12617

Byna, A., & Anisa, F. N. (2018). Backward Elimination to improve the Accuracy of Stunting Incidence with Support Vector Machine Algorithm Analysis. Health Dynamics, 9(2), 217-225.

Indraswari, R., Zainal Arifin, A., & Darlis, H. (2017). RBF Kernel Optimazation Method With Particle Swarm Optimization On SVM Using The Analysis Of Input Data'S Movement. Journal of Computer Science and Information, 13(3), 1576-1580. https://doi.org/http://dx.doi.org/10.21609/jiki.v10i1.410 RBF

Isnain, A. R., Sakti, I. A., Alita, D., & Marga, N. S. (2021). Public Sentiment Analysis on Jakarta Government Lockdown Policy Using Svm Algorithm. Jdmsi, 2(1), 31-37. https://t.co/NfhnfMjtXw