



Analysis of k-medoids clustering on toddler immunization in north sumatra province

Lena Sapura¹, M. Safii²

^{1,2}STIKOM Tunas Bangsa Pematangsiantar, Sumatera Utara, Indonesia

ARTICLE INFO

Article history:

Received 29 Oct, 2022

Revised 21 Nov 2022

Accepted 30 Nov, 2022

Keywords:

Data Mining
Clustering
K-Medoids
Immunization

ABSTRACT

Immunization is a process to increase the body's immune system by inserting a vaccine, namely a virus or bacteria that has been weakened, killed, or parts of the bacteria (virus) have been modified. The purpose of this study was to group children's immunization data according to districts/cities in North Sumatra Province using the K-Medoids Data Mining algorithm. The K-Medoids algorithm or pattern known as PAM (Partitioning Around Medoids) uses the partitioning clustering method to group a set of n objects into a number of K clusters. The data in this study is sourced from the Central Bureau of Statistics of North Sumatra. The grouping was carried out based on the number of recipients of the DPT-HB and Measles/MR vaccines from 33 districts/cities in North Sumatra Province. The results of the study are expected to be taken into consideration for local governments, especially the Ministry of Health in the distribution of immunization for children under five with the type of immunization to achieve the national target set by the Ministry of Health, which is 79.1%.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



Corresponding Author:

M. Safii,
STIKOM Tunas Bangsa,
Pematang Siantar,
Jl.Jend. Sudirman Blok A No.1-3,Pematang Siantar- North Sumatra
Email: m.safii@amiktunasbangsa.ac.id

1. INTRODUCTION

Immunization is the most effective and efficient public health effort to prevent diseases and reduce mortality, such as smallpox, polio, tuberculosis, hepatitis B, diphtheria, measles, rubella and congenital rubella syndrome (CRS), tetanus, pneumonia (pneumonia) and meningitis (inflammation of the lining of the brain). Based on data from the Ministry of Health routine immunization results have decreased since 2020. Acting director general of disease prevention and control, Indonesian Ministry of Independence dr. Maxi Rein Rondonuwu continues to encourage the Regional Government, especially the Health Office, to achieve the target of 79.1% immunization coverage. Based on the routine immunization data report for October 2021, the coverage of complete basic immunization has only reached 58.4% of the target of 79.1%. Of the several provinces, North Sumatra is one of the provinces whose complete basic immunization coverage is above 60% but is still far from the national target of 79.1%. The problem that occurs in North Sumatra is that in administering immunization to toddlers it is carried out sporadically throughout the people of North Sumatra, meaning that immunization is carried out simultaneously at the same time to all areas in the Regencies and Cities in North Sumatra. Therefore the target to be achieved cannot be realized. So that the desired results will be difficult to achieve, because many cases have occurred when

immunization is scheduled in an area by bringing all kinds of logistics and medical personnel to carry out immunizations (Duijzer et al., 2018), but only a few toddlers can be presented. Factors that affect the decline in complete basic immunization coverage (Negussie et al., 2016) (Najikhah et al., 2021): According to (Thapar et al., 2021) (Salmon et al., 2005) Some mothers who refuse these immunizations are those with higher education and social status, with the reason being worried that the increasing number of vaccines recommended for babies will actually affect the baby's immune system at the age of immature youth. According to (Greyson & Bettinger, 2022) some mothers are still unsure about vaccines given to babies because of the lack of knowledge about immunization for toddlers. According to (Simon et al., 2015) family lifestyle is also a factor, perceptions about the child's body and immune system or the body already has its own antibodies. According to (Parasidis & Opel, 2017) most of the courts (7 out of 9) considered vaccine refusal to be negligence, there were only 4 cases of vaccine refusal for religious reasons. Therefore researchers conducted research on this case using the K-Medoids algorithm to analyze clusters of immunization for toddlers by region in North Sumatra Province, as one way to close the gap in the distribution of complete basic immunization to achieve the national target set by the Ministry of Health (Samudi et al., 2020) (Tampubolon et al., 2021) (Butarbutar & Safii, 2021). K-Medoids algorithm is one of data mining clustering (Angraini et al., 2021). Same as K-Means, except that the centroid is not estimated by means, but by observations that have a minimum pairwise distance with other cluster members (Tampubolon et al., 2021).

2. RESEARCH METHOD

The purpose of this research is to produce cluster results which are divided into three parts: low cluster, medium cluster, and high cluster so that later the results of the appropriate grouping can be determined. The settlement method used in this study is Data Mining with the K-Medoids method. In this design, the first time the author did was to retrieve data from the Central Bureau of Statistics, then transfer it to excel (xls). After that, the data will be processed through a calculation process and follow the steps of the k-medoids method. Then the results of these calculations will be applied to Rapidminer to see the accuracy of the results obtained.

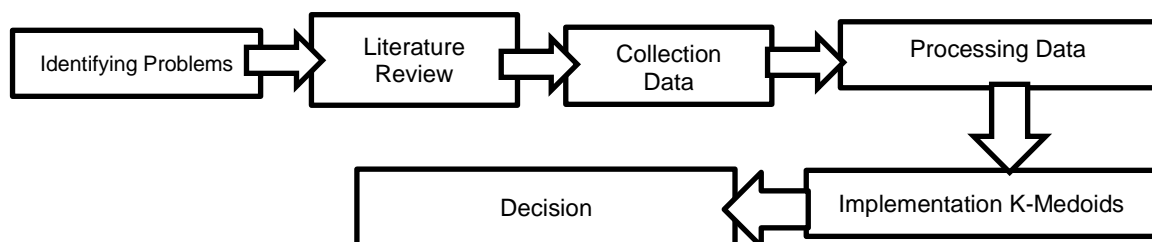


Figure 1. Research Framework

In the picture above is the process of the instrument of this study as follows: Analyzing problems related to toddler immunization percentage data and determining the parameters used. In this study, the parameters used were data on the percentage of toddler immunizations by region in North Sumatra Province in 2021. Data processing using data mining, especially the K-Medoids clustering method. Data processing here uses Ms.Excel and RapidMiner. The data analysis process can be carried out after data collection and after it has been collected and processed in Excel. After the results obtained in excel will be applied to Rapidminer to adjust the results obtained. The data used is the percentage of children under five who have received immunization by Regency/City and type of Immunization, 2021. The data can be seen in table 1 below.

Table 1. Data on the Percentage of Immunizations for Under-fives According to Regencies/Cities and Types (2021)

Regency/City	BCG	DPT-HB3/DPT-HB/Hib3	MEASLES	POLIO 4	HEPATITIS B (1 - 7 Hari)
Nias	62,15	64,46	65,08	64,48	10,73
Manailing Natal	87,87	84,18	78,88	83,45	1,39
South Tapanuli	67,13	77,06	65,08	73,34	0
Midle Tapanuli	69,29	71,37	69,53	71,76	0,92
North Tapanuli	62,02	70,59	61,89	67,19	0
Toba	76,60	76,36	75,52	76,43	20,66
Labuhanbatu	94,62	94,54	91,80	90,54	0
A s a h a n	73,36	75,95	74,18	75,63	0
Simalungun	88,59	92,13	91,64	92,30	6,88
Dairi	69,93	75,16	72,34	74,13	0
Karo	7,77	84,90	84,18	84,10	27
Deli Serdang	93,14	89,03	91,51	91,35	0
Langkat	77,47	72,23	69,02	73,61	13
South Nias	84,11	84,56	89,38	84,67	50
Humbang Hasundutan	111,87	87,98	87,46	88,34	3,63
Pakpak Bharat	74,36	81,11	67,96	80,94	0
Samosir	68,16	72,34	73,31	72,34	0,00
Serdang Bedagai	87,81	93,67	77,76	92,67	84,46
Batu Bara	7,60	9,49	9,24	8,72	7,92
Padang Lawas Utara	90,74	84,47	87,44	84,13	82,42
Padang Lawas	64,38	64,22	62,28	64,88	52,91
Labuhanbatu Selatan	78,05	81,89	75,92	80,12	30,27
Labuhanbatu Utara	106,40	105,53	106,74	104,91	46,24
Nias Utara	98,48	97,37	93,22	97,01	1
Nias Barat	55,47	68,48	58,40	65,22	41,44
Sibolga	73,27	66,92	71,31	68,64	0
Tanjungbalai	47,87	46,89	35,21	47,59	28,04
Pematangsiantar	86,92	79,27	81,48	79,18	4,23
Tebing Tinggi	79,80	80,25	77,97	95,40	9,62
Medan	78,94	78,58	77,40	78,34	15,57
Binjai	83,22	78,83	75,07	77,12	0
Padangsidempuan	76,06	76,18	67,57	73,78	13
Gunungsitoli	48,79	51,13	51,82	52,95	0

3. RESULTS AND DISCUSSIONS

In the following stages, complete basic immunization data processing for toddlers based on districts/cities in North Sumatra Province is carried out using the data mining method, namely the K-Medoids algorithm. The following are the steps taken to complete manual calculations in data mining using the K-Medoids clustering algorithm. Data testing was carried out in grouping using the RapidMiner 5.3 software. The initial appearance of the RapidMiner 5.3 tools can be seen in Figure 2.

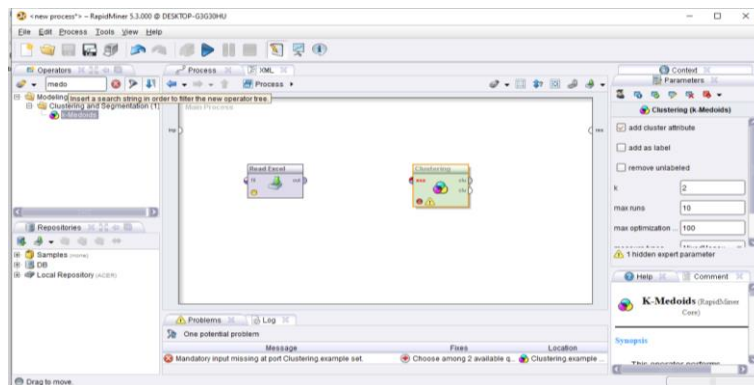


Figure 2. Data testing process

Connect Read Excel with Clustering by drawing a line between the two as shown in Figure 3.

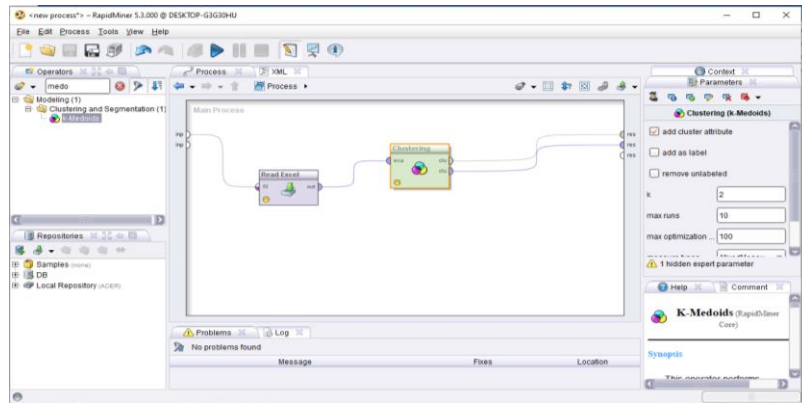


Figure 3. The process of connecting data with clustering

click Run or Resume the Current Process located at the top in the form of a blue arrow to display the final results of the data testing process on RapidMiner 5.3 as shown in Figure 4.

The screenshot shows the 'Data View' of the clustering results. The table displays 17 rows of data, each representing a different district (Kabupaten) and its assigned cluster. The columns are: Row No., Kabupaten/, cluster, a, b, c, d, and e.

Row No.	Kabupaten/	cluster	a	b	c	d	e
1	01 Nias	cluster_0	62.150	64.460	65.080	64.480	10.730
2	02 Mandailing	cluster_0	87.870	84.180	78.880	83.450	1.390
3	03 Tapanuli	cluster_0	67.130	77.060	65.080	73.340	0
4	04 Tapanuli	cluster_0	69.290	71.370	69.530	71.760	0,820
5	05 Tapanuli	cluster_0	62.020	70.590	61.890	67.190	0
6	06 Toba	cluster_0	76.000	76.360	75.520	76.430	20,660
7	07 Labuhan	cluster_0	84.620	84.540	91.800	90.540	0
8	08 Asahan	cluster_0	73.360	75.950	74.180	75.630	0
9	09 Simatun	cluster_0	88.590	92.130	91.540	92.300	6,880
10	10 Deli	cluster_0	69.930	75.160	72.340	74.130	0
11	11 Karo	cluster_0	7.770	84.900	84.190	84.100	27
12	12 Deli	cluster_0	93.140	89.030	91.510	91.350	0
13	13 Langkat	cluster_0	77.470	72.230	69.020	73.610	13
14	14 Nias Selatan	cluster_0	84.110	84.560	89.380	84.670	50
15	15 Humbang	cluster_0	111.870	87.980	87.450	88.340	3,630
16	16 Pakpak	cluster_0	74.360	81.110	67.960	80.940	0
17	17 Samosir	cluster_0	66.160	72.340	73.310	72.340	0

Figure 4. Final results of data testing

The final result is displayed in the form of a grouping of each cluster which determines the members of the cluster as shown in Figure 5.

The screenshot shows the 'Data View' of the clustering results, similar to Figure 4 but with different cluster assignments. The table displays 17 rows of data, each representing a different district (Kabupaten) and its assigned cluster.

Row No.	Kabupaten/	cluster	a	b	c	d	e
1	01 Nias	cluster_1	62.150	64.460	65.080	64.480	10.730
2	02 Mandailing	cluster_1	87.870	84.180	78.880	83.450	1.390
3	03 Tapanuli	cluster_1	67.130	77.060	65.080	73.340	0
4	04 Tapanuli	cluster_1	69.290	71.370	69.530	71.760	0,820
5	05 Tapanuli	cluster_1	62.020	70.590	61.890	67.190	0
6	06 Toba	cluster_1	76.000	76.360	75.520	76.430	20,660
7	07 Labuhan	cluster_1	84.620	84.540	91.800	90.540	0
8	08 Asahan	cluster_1	73.360	75.950	74.180	75.630	0
9	09 Simatun	cluster_1	88.590	92.130	91.540	92.300	6,880
10	10 Deli	cluster_1	69.930	75.160	72.340	74.130	0
11	11 Karo	cluster_1	7.770	84.900	84.190	84.100	27
12	12 Deli	cluster_1	93.140	89.030	91.510	91.350	0
13	13 Langkat	cluster_1	77.470	72.230	69.020	73.610	13
14	14 Nias Selatan	cluster_0	84.110	84.560	89.380	84.670	50
15	15 Humbang	cluster_1	111.870	87.980	87.450	88.340	3,630
16	16 Pakpak	cluster_1	74.360	81.110	67.960	80.940	0
17	17 Samosir	cluster_1	66.160	72.340	73.310	72.340	0

Figure 5. Results of data grouping

In the picture above, the clusters used are 3 clusters namely cluster_0 is colored nbchxvg, cluster_1 is colored bchxgdh and cluster_2 is colored dghjsj. The cluster model of data grouping is shown in Figure 6.

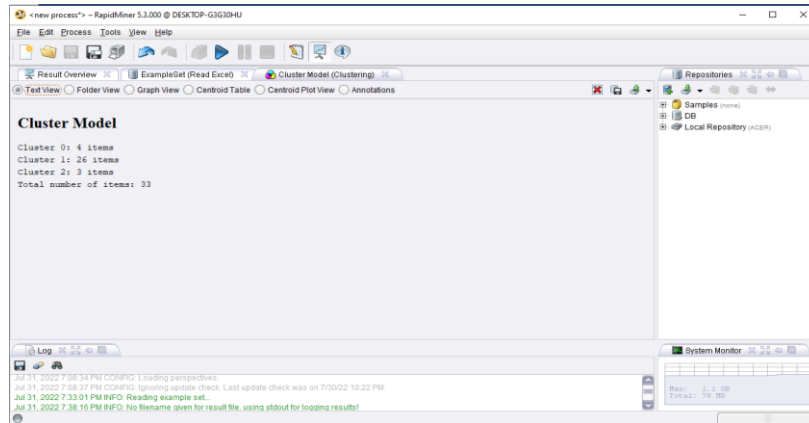


Figure 6. Results

From Figure 6 it can be seen that cluster_0 has 4 items, cluster_1 has 26 items, and cluster_2 has 3 items. The results of algorithm calculations and with RapidMiner can be seen in table 2.

Table 2. The Results of Algorithm

No	Regency/City	Calculation K-Medoids	Calculation RapidMiner
1	Nias	cluster_2	cluster_1
2	Mandailing Natal	cluster_2	cluster_1
3	Tapanuli Selatan	cluster_2	cluster_1
4	Tapanuli Tengah	cluster_2	cluster_1
5	Tapanuli Utara	cluster_2	cluster_1
6	Toba	cluster_2	cluster_1
7	Labuhanbatu	cluster_2	cluster_1
8	Asahan	cluster_2	cluster_1
9	Simalungun	cluster_2	cluster_1
10	Dairi	cluster_2	cluster_1
11	Karo	cluster_2	cluster_1
12	Deli Serdang	cluster_2	cluster_1
13	Langkat	cluster_2	cluster_1
14	Nias Selatan	cluster_1	cluster_0
15	Humbang Hasundutan	cluster_2	cluster_1
16	Pakpak Barat	cluster_2	cluster_1
17	Samosir	cluster_2	cluster_1
18	Serdang Bedagai	cluster_1	cluster_0
19	Batu Bara	cluster_3	cluster_2
20	Padang Lawas Utara	cluster_1	cluster_0
21	Padang Lawas	cluster_2	cluster_1
22	Labuhanbatu Selatan	cluster_2	cluster_1
23	Labuhanbatu Utara	cluster_1	cluster_0
24	Nias Utara	cluster_2	cluster_1
25	Nias Barat	cluster_2	cluster_1
26	Sibolga	cluster_2	cluster_1
27	Tanjungbalai	cluster_3	cluster_2
28	Pematangsiantar	cluster_2	cluster_1
29	Tebing Tinggi	cluster_2	cluster_1
30	Medan	cluster_2	cluster_1
31	Binjai	cluster_2	cluster_1
32	Padangsidempuan	cluster_2	cluster_1
33	Gunungsitoli	cluster_3	cluster_2

4. CONCLUSION

Based on the results of the discussion on the grouping of Complete Basic Immunization in Toddlers described above, this thesis using the K-Medoids algorithm draws conclusions, namely the clustering results obtained are tested again using RapidMiner 5.3 software for more accurate results. The test results on RapidMiner 5.3 with manual calculations from the K-Medoids algorithm obtained the same results, namely high clusters of 4 Regencies/Cities, medium clusters of 26 Regencies/Cities and low clusters of 3 Regencies/Cities.

REFERENCES

- Angraini, D. D., Safii, M., & Angraini, F. (2021). A Oil Palm Harvest Grouping Using K-Medoids Algorithm. *International Journal of Basic and Applied Science*, 10(2), 60–68. <https://doi.org/10.35335/ijobas.v10i2.56>
- Butarbutar, E. F. T., & Safii, M. (2021). *Journal of Artificial Intelligence and Engineering Applications K-Medoids Algorithm Analysis in Grouping Students' Level of Understanding of Subjects*. 1(1), 6–13.
- Duijzer, L. E., van Jaarsveld, W., & Dekker, R. (2018). Literature review: The vaccine supply chain. *European Journal of Operational Research*, 268(1), 174–192. <https://doi.org/10.1016/j.ejor.2018.01.015>
- Greyson, D., & Bettinger, J. A. (2022). How do mothers' vaccine attitudes change over time? *SSM - Qualitative Research in Health*, 2, 100060. <https://doi.org/10.1016/j.ssmqr.2022.100060>
- Najikhah, N., Nurjannah, N., Mudatsir, M., Usman, S., & Saputra, I. (2021). Determinants of Complete Basic Immunization in Children Aged 12-23 Months in Indonesia. *Budapest International Research in Exact Sciences (BirEx Journal)*, 3(4), 304–318.
- Negussie, A., Kassahun, W., Assegid, S., & Hagan, A. K. (2016). Factors associated with incomplete childhood immunization in Arbogona district, southern Ethiopia: a case-control study. *BMC Public Health*, 16, 27. <https://doi.org/10.1186/s12889-015-2678-1>
- Parasidis, E., & Opel, D. J. (2017). Parental Refusal of Childhood Vaccines and Medical Neglect Laws. *American Journal of Public Health*, 107(1), 68–71. <https://doi.org/10.2105/AJPH.2016.303500>
- Salmon, D. A., Moulton, L. H., Omer, S. B., deHart, M. P., Stokley, S., & Halsey, N. A. (2005). Factors Associated With Refusal of Childhood Vaccines Among Parents of School-aged Children: A Case-Control Study. *Archives of Pediatrics & Adolescent Medicine*, 159(5), 470–476. <https://doi.org/10.1001/archpedi.159.5.470>
- Samudi, S., Widodo, S., & Brawijaya, H. (2020). The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic. *Sinkron*, 5, 116. <https://doi.org/10.33395/sinkron.v5i1.10649>
- Simon, A. K., Hollander, G. A., & McMichael, A. (2015). Evolution of the immune system in humans from infancy to old age. *Proceedings. Biological Sciences*, 282(1821), 20143085. <https://doi.org/10.1098/rspb.2014.3085>
- Tampubolon, H. D., Suhada, S., Safii, M., Solikhun, S., & Suhendro, D. (2021). Penerapan Algoritma K-Means dan K-Medoids Clustering untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi. *Jurnal Ilmu Komputer Dan Teknologi*, 2(2), 6–12. <https://doi.org/10.35960/ikomti.v2i2.703>
- Thapar, R., Kumar, N., Surendran, P., Shahdiya, A., Mahendran, V., Ramesh, R., Shetty, D. J., Unnikrishnan, B., Mithra, P., Holla, R., Bhagwan, D., & Kumar, A. (2021). Vaccine hesitancy among mothers of under-five children in Coastal South India: a facility-based cross-sectional study. In *F1000Research* (Vol. 10). <https://doi.org/10.12688/f1000research.28293.2>