Application of K-Medoids Cluster Result with Particle Swarm Optimization (PSO) in Toddler Measles Immunization Cases

Eka Pandu Cynthia¹, Indra Riyana Rahadjeng², Erfan Karyadiputra³, Fauzi Yusa Rahman³, Agus Perdana Windarto⁴, Martalina Limbong⁵, Teguh Iman Hermanto⁶, E. Rusiadi⁷, and Y Yarmanti⁸

¹Informatics Engineering Study Program, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau, Indonesia
²Universitas Bina Sarana Informatika, Jakarta, Indonesia
³Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin, Indonesia
⁴STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
⁵Akademi Keperawatan Surya Nusantara, Pematangsiantar, Indonesia
⁶Informatics Engineering Study Program, Sekolah Tinggi Teknologi Wastukancana, Indonesia
⁷Universitas Pembangunan Panca Budi, Medan, Indonesia
⁸Universitas Bengkulu, Bengkulu, Indonesia

Corresponding Email: eka.pandu.cynthia@uin-suska.ac.id

Abstract. The objective of the research was to analyze the clustering method by optimizing Particle Swarm Optimization (PSO) in the case of measles immunization for children under the age of 5. The research data source used is the North Sumatra Province Central Bureau of Statistics (https://sumut.bps.go.id/). The data used in 2019 included 33 records with variables BCG, DPT-HB3/DPT-HB/Hib3, CAMPAK (MEASLES + RUBELLA), POLIO 4 and HEPATITIS B. The methods used were k-medoids and PSOs. This method is used to determine the value of the Davies Bouldin Index (DBI) before the cluster value is determined (k). The resulting k values were compared to k-medoids without PSO. The best results of k-medoid and PSO will be tested by classification to see the accuracy value of the cluster formed. The results showed that the optimization of k-medoids and PSO was better with the number of clusters (k = 5) than 0.078. (DBI). The results also show that the accuracy of the cluster formed is 95% with a correlation of 0.98.

Keywords: Datamining, Clustering, K-Medoids, Particle Swarm Optimization, Toddler Measles Immunization, Rapid Miner, Indonesia.

1. Introduction

The measles content in Indonesia is still very high in children under 5. An estimated 30,000 Indonesian children die every year from measles complications. More than 95% of the deaths from measles in countries around the world like the United Kingdom, Yemen, Latin America, Myanmar, Africa, Central Africa, Liberia, Guinea and Indonesia [1] have been demonstrated in this situation. The government should play an active role and commit to removing measles in this case. Measles is an infectious infectious disease caused by a virus that attacks children and adults [2]. Measles may be transmitted by coughing and sneezing via direct contact. Measles is the primary cause of infant mortality in immunizational diseases [3]. Therefore, a minimum of 95% immunization is required in
order to break the chain of measles transmission, especially infants. One of the measures to eradicate
measles is the vaccine against measles, which involves injecting or drinking certain weakened
poisonous diseases into the body. This is designed to improve the body’s immunity to the disease [4].

There are several studies related to measles immunization in infants conducted by R. W. Sari
(2018) [5] using the k-means method. The research carried out mapping on the basis of the regions in
Indonesia. In the meantime, it was not determined in the process of determining the number of
clusters. The results of the research indicate that the mapping is divided into three clusters. In the
meantime, the research to be carried out uses the k-medoids method and Particle Swarm Optimization
(PSO) to optimize the number of clusters formed. Based on the Davies Bouldin Index (DBI) value, it
will be determined how many clusters are used for cluster mapping. The results of the mapping are
then classified to see the extent to which the results of the cluster are formed by looking at the
accuracy value. In addition, the research to be carried out uses five variables, while the research by R.
W. Sari (2018) uses one variable. Clustering research is often used because the process is very familiar
and easy to apply [6], [7]. This study proposes an optimization method to maximize the number of
clusters to be formed. This can be seen from the Davies Bouldin Index (DBI) value, which is one of
the methods used to determine the number of clusters (k). The optimization problem can be solved
using the Particle Swarm Optimization (PSO) algorithm. The PSO algorithm is used to optimize for
maximum results. There have been many studies using PSO optimization, such as tidal predictions, in
which PSO is used for the minimum value of optimization error in the network to achieve ideal neural
network weight. PSO and artificial neural networks have several input parameters, e.g., number of
input neurons, learning speed, swarm, c1, c2 min inertia, max inertia. The data used amounted to
1000, divided into 700 training data and 300 test data. Test results show the accuracy of the forecast
91.56 percent used 90 swarms, the learning speed was 0.9 and iterated 20 times [8]. Further research
on the classification of credit analysis using the C4.5 and PSO algorithms [9]. From the results of the
experiment Modeling, the C4.5 algorithm based on Particle Swarm Optimization (PSO) has the best
results at 70 percent, while the C4.5 algorithm based on Particle Swarm Optimization (PSO) is only
68.6 percent. Based on this, it is hoped that the results of the research will be able to optimize the
number of clusters that will be formed for infant measles immunization.

2. Methodology

The data used is from the BPS North Sumatra region (https://sumut.bps.go.id) on the percentage of
children under the age of five who received immunization by district/city and on the type of
immunization in 2019. The data consists of 33 districts/city consisting of 5 types of immunization,
namely BCG, DPT-HB3/DPT-HB/Hib3, CAMPAK (MEASLES + RUBELLA), POLIO 4 and
HEPATITIS B. The following is the flow-chat used in the study as shown in Figure 1.
Figure 1, it is explained that the dataset comes from https://sumut.bps.go.id, which is processed using Microsoft Excel to view missing data. The dataset used can be found in Table 1. The dataset is then analyzed using the Rapid Miner software. The next step is to determine the number of clusters (k) by optimizing k-medoids + PSO and k-medoids without PSO. The best DBI value is a reference for cluster mapping. The results of the mapping that were produced were tested to see the resulting accuracy value.

3. Results and Discussion

The mapping process with PSO optimization in the clustering method is performed with the help of the Rapid Miner software. The design is carried out using several stages in the design. The first stage is a design used to compare the DBI values of k-medoids and k-medoids + PSO methods to determine the number of clusters. Then the number of clusters that have been determined will be processed to obtain the mapping results. The second stage is to test the results of the clusters that have been formed. The following dataset is used for the design of the first and second phases as shown in Tables 2 and 3 below:

Table 1. Dataset of percentage of toddlers who have received immunizations, 2019

| District / City | BCG | DPT-HB3/DPT-HB/Hib3 | CAMPAK (MEASLES + RUBELLA) | POLIO 4 | HEPATITIS B |
|----------------|-----|----------------------|----------------------------|---------|-------------|
| 01 Nias        | 94.5| 100.3                | 99.8                       | 99.7    | 28.6        |
| 02 Mandailing Natal | 88.7| 84.3                | 76.7                       | 82.5    | 0.3         |
| 03 Tapanuli Selatan | 83.6| 86.1                | 82.6                       | 88.2    | 9.9         |
| 04 Tapanuli Tengah | 75.4| 75.4                | 73.2                       | 75.3    | 0.1         |
| 05 Tapanuli Utara | 84.3| 71.4                | 54.9                       | 60.2    | 2.3         |
| 06 Toba        | 91.3| 93.8                | 93.9                       | 93.8    | 30.5        |
| 07 Labuhanbatu | 97.7| 95.8                | 96.4                       | 95.9    | 0           |
| 08 Asahan      | 91.7| 98                  | 95.6                       | 96.9    | 0.6         |
| 09 Simalungun  | 99.8| 103.5               | 103.3                      | 101.1   | 38.5        |
The dataset is used as an input process using the Rapid Miner software in Table 1. The file you are using is an excel file (.xls). Using the Read Excel operator, the data will be used as input by specifying districts/city as labels. The Read Excel operator is then connected to the Multiply operator to compare the results of k-medoids and k-medoids + PSO. In addition, the multiplier operator is connected to two processes, namely k-medoids + PSO and k-medoids, the output of which is used by the cluster distance performance operator. The number of clusters to be tested is between 2 and 6 (k = 2, 3, 4, 5 and 6). The first stage design can be seen in Figure 2 below:

![Figure 2. Design process for selecting the number of clusters (k)](image)

Based on Figure 2, the number of clusters (k) was tested using k-medoids and k-medoids + PSO methods. The following are the results of the Davies Bouldin Index from each processed k value as shown in Table 2 below:

| K       | K-Medoids DBI | K-Medoids + PSO DBI |
|---------|---------------|----------------------|
| k=2     | 0.218         | 0.135                |
| k=3     | 0.199         | 0.134                |
| k=4     | 0.228         | 0.121                |
| k=5     | 0.198         | 0.078                |
| k=6     | 0.201         | 0.102                |

Table 2 explains that each given k value has a different value with k-medoids and k-medoids + PSO methods. In every way, the k-medoids + PSO method excels. It can be seen that the value of k = 5 has the smallest number, namely 0.078 (DBI value) so that the number of clusters is determined by the
The results of the smallest DBI value. The following are the graphs and mapping results using \( k = 5 \) as shown in the figure and table below.

**Figure 3.** Graph of DBI Value on K-medoids and K-medoids + PSO

**Figure 4.** Mapping results at a value of \( k = 5 \)

**Figure 5.** Mapping district/city details with a value of \( k = 5 \) for infant measles immunization.

The results of the mapping formed in Figure 5 will then be tested by entering the mapping results into the data input in the design of Figure 8. The following is the input of the resulting mapping data that is converted to excel (.xls) as shown in table 2 below.
The results of the mapping in Table 3 are the results of the determination of the smallest k value using the Davies Bouldin Index (DBI) method. The following is a cluster and district/city mapping chart as shown in Figures 6 and 7 below:

The dataset of Table 3 will be tested using the design of Figure 8, where the desired results are the accuracy and correlation values of the clusters formed. The following are the design and test results using the following RapidMiner software.
4. Conclusion

The results showed that the application of PSO in the k-medoid method can be done where the cluster determination process is performed by testing each cluster value (k) using the Davies Bouldin Index (DBI) method. The mapping results that are produced have an accuracy value of 95 percent and a correlation of 0.989.
References

[1] L. M. N. Lochlainn et al., “Effect of measles vaccination in infants younger than 9 months on the immune response to subsequent measles vaccine doses: a systematic review and meta-analysis,” www.thelancet.com/infection, vol. 19, no. November 2019, 2020.

[2] W. Rosadi, E. S. Sulaeman, and H. Prasetya, “Multilevel Analysis on Factors Affecting Measles Rubella Immunization Uptake among Toddlers in Pekanbaru, Indonesia,” J. Matern. Child Heal., vol. 4, no. 6, pp. 448–460, 2019.

[3] R. Widoyo, “Measles Immunization and Vitamin A for Prevention of Pneumonia in Indonesia,” Kesmas Natl. Public Heal. J., vol. 11, no. 1, pp. 46–50, 2016.

[4] P. Aaby and C. J. Clements, “Measles immunization research: A review,” Bull. World Health Organ., vol. 67, no. 4, pp. 443–448, 1989.

[5] R. W. Sari, A. Wanto, and A. P. Windarto, “Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi),” KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer), vol. 2, no. 1, pp. 224–230, 2018.

[6] A. P. Windarto et al., “Analysis of the K-Means Algorithm on Clean Water Customers Based on the Province,” J. Phys. Conf. Ser., vol. 1255, no. 1, 2019.

[7] Sudirman, A. P. Windarto, and A. Wanto, “Data mining tools | rapidminer: K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia,” IOP Conf. Ser. Mater. Sci. Eng., vol. 420, p. 012089, 2018.

[8] N. Nikentari, H. Kurniawan, N. Ritha, D. Kurniawan, U. Maritim, and R. Ali, “Particle Swarm Optimization Untuk Prediksi Pasang Surut Air Optimization of Backpropagation Artificial Neural Network With Particle Swarm Optimization To Predict Tide Level,” J. Teknol. Inf. dan Ilmu Komput., vol. 5, no. 5, pp. 605–612, 2018.

[9] S. Saprudin, “Penerapan Particle Swarm Optimization (PSO) untuk Klasifikasi dan Analisis Kredit dengan Menggunakan Algoritma C4.5,” J. Inform. Univ. Pamulang, vol. 2, no. 4, p. 214, 2017.