Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754
P-ISSN : 2580-569X
Vol. 8, No. 2, 2024
DOI: 10.30829/zero.v8i2.23480
pp. 47-56

Implementation of Fuzzy C – Means (FCM) And Fuzzy Possibilistic C – Means (FPCM) for Clustering District/City Based On Health Services And Infectious Diseases in North Sumatera

Feronika Paska Purba[1], Klause Roder[2], Elmanani Simamora[3], Hamidah Nasution[4]

[1,3,4] Department of Mathematics, Universitas Negeri Medan, Medan, Indonesia
[2] Department of Applied Mathematics, Universty of Ausburg, Ausburg, Germanyy

Article Info

Article history:

Received, 20 10 2024
Revised, 15 11 2024
Accepted, 20 12 2024

Keywords:

fuzzy c- means,
fuzzy possibilistic c- means
health, infectious disease

ABSTRACT

The study aims to compare the Fuzzy C-Means (FCM) and Fuzzy Possibilistic C-Means (FPCM) algorithms and to profile the results of district / city clusters in North Sumatra based on health services and infectious disease sufferers. The method used is descriptive quantitative research using annual data obtained from the North Sumatra Provincial Health Office for the period 2023. The data was collected and then analyzed using both Clustering algorithms to find the most optimal results. The results showed that Fuzzy Possibilistic C-Means proved to be a better algorithm compared to Fuzzy C-Means in this study. The number of clusters formed is 3 with Cluster 3 being the highest level of health urgency in North Sumatra province. The findings of this study can help the government in equalizing the control of the number of health services and infectious disease patients.

Corresponding Author:

Feronika Paska Purba Paska Purba,
Department of Mathematics,
Universitas Negeri Medan, Medan, Indonesia
Email: feronika@gmail.com

1. INTRODUCTION

Health is a state in which an individual is in good physical, mental, and social condition, which provides an opportunity for individuals to carry out productive activities in social and economic aspects. In North Sumatra, there are several infectious diseases that have the highest number of cases including tuberculosis (TB), HIV/AIDS, diarrhea and DHF.

According to data from the North Sumatra Health Office [1], during the period January to October 2023, there were 2,928 new cases of HIV in North Sumatra, and 8,885 people diagnosed with HIV were undergoing treatment. The Central Bureau of Statistics noted that diarrheal disease was the number one case in North Sumatra with 205,155 cases. Data from the North Sumatra Provincial Health Office shows that in 2022, North Sumatra ranked third nationally for the highest number of dengue cases. The total number of DHF cases recorded in that year reached 8,539, with 60 of those cases being fatal. North Sumatra itself has health service facilities to overcome existing diseases. Facilities that can be used by the community are Community Health Centers (PUSKESMAS) and Hospitals.

This research shows the advantages of the FCM algorithm, namely that each data can be in each cluster according to its membership degree. FCM can also overcome overlapping data distribution, which is difficult to overcome by other clustering techniques. In the case of outlier data, which is far from the cluster center FCM often has difficulty in providing proper clustering. To overcome the weakness of FCM Means

(FPCM) (Fadhel and Alimi, 2009). In previous research by Thilagaraj and Sengottaiyan [2], Fuzzy C-Means and Fuzzy Possibilistic C-Means were applied to find underperforming employees for placement in the software industry. The findings of this study indicate that the FPCM method is superior to the FCM method because FPCM focuses more on noise and outliers in the data set. In addition, research conducted by R.J. Kuo [3] that FPCM can be used to categorize market segmentation.

Due to the large number of attributes and the need for accurate decision recommendations through clustering the availability of health services and infectious disease cases in all regions in North Sumatra Province, Clustering with Fuzzy C-Means and Fuzzy Possibilistic C-Means algorithms was carried out for clustering the number of health services and infectious disease cases in the districts / cities of North Sumatra Province.

2. RESEARCH METHOD

2.1 *Exploratory Data Analysis* (EDA)

Data Standardization

Data standardization is a process that aims to change the values in the dataset so that it has a consistent and uniform format. Data Standardization can be calculated with the following equation:

$$x_{stand} = \frac{x - mean(x)}{Standard\ deviation\ (x)} \quad (1)$$

Fuzzy C – Means

Fuzzy C-Means is a data clustering in which the membership of each data in a cluster is determined based on its membership degree. When analyzing fuzzy clusters, the degree of membership will be measured by the closeness of the object to the cluster center Thilagaraj and Sengottaiyan [2]. The steps of the FCM algorithm are as follows (Klawonn, 2007 in Sumanto and Wahono, 2011):

1. Input data to be clustered X , in the form of a matrix of size $n\ x\ m$ (n is the number of districts / cities; m is the attributes of each district / city). $X_{ij}\ i\ th\ (i = 1,2,\dots,n), j\ (j = 1,2,\dots,m)$.
2. Specify: Number of Clusters (c), Rank (w), Maximum Iterations $(MaxIter)$, Error Rate (ε_I), Initial Objektive Function $(P_0 = 0)$, Initial Iterations $(t = 1)$
3. Generate random numbers μ_{ik} with $i = 1,2,\dots,n; k = 1,2,\dots,c;$, as the elements of the initial partition matrix U

   Generate random numbers μ_{ik} with $i = 1,2,\dots,n; k = 1,2,\dots,c;$, as the elements of the initial partition matrix U

$$\boldsymbol{\mu_{ik}} = \begin{bmatrix} \mu_{11}(X_1) & \cdots & \mu_{1m}(X_1) \\ \vdots & \ddots & \vdots \\ \mu_{n1}(X_n) & \cdots & \mu_{nm}(X_n) \end{bmatrix}; 1 \le i \le n; \quad (2)$$

4. Calculate the *k-th* cluster center: V_{kj}, with $k = 1,2,\dots,c; j = 1,2,\dots,m$.

$$V_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^{w} * x_{ij})}{\sum_{i=1}^{n}((\mu_{ik})^{w}} \quad (3)$$

5. Calculate the objective function at iteration *t*, P_t

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}(x_{ij} - v_{kj})^2\right](\mu_{ik})^{w}\right) \quad (4)$$

6. Calculate the change of partition matrix

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(x_{ij} - v_{kj})^2\right]^{\frac{-1}{w-1}}}{\left[\sum_{k=1}^{c}(x_{ij} - v_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (5)$$

7. Check the iteration termination condition:
   a. If : $(|P_t - P_{t-1}| < \varepsilon_I$ or $(t > MaxIter)$ then stop
   b. If not: $t = t + 1$, repeat step 4

   The calculation continues until convergent. The FCM criteria are said to be convergent, namely:
   a) If there is no change in members in the Cluster
   b) If there is no too significant change in the Cluster center
   c) Minimum decrease in SSE value

2.2 Fuzzy Possibilistic C Means (FPCM)

1. Calls the relative distinctiveness matrix (μ_{ik}) and Cluster center $\boldsymbol{V}_{kj}$ in the *Fuzzy C-Means* (FCM) to calculate the absolute distinctiveness matrix where μ_{ik} $i = 1,2,\dots,n; k = 1,2,\dots,c; j = 1,2,\dots,n;$ dwith the following equation:

$$\boldsymbol{t_{ik}} = \begin{bmatrix} t_{11} & \cdots & t_{1c} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nc} \end{bmatrix}$$

Where the matrix element is :

$$t_{ik} = \frac{\left[\sum_{j=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{\eta-1}}}{\left[\sum_{i=1}^{c}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{\eta-1}}} \tag{6}$$

with ;

η: he rank of the weights in the absolute distinctiveness matrix; $\eta > 1$

2. Fix the Cluster center where $i = 1,2,\dots,n; k = 1,2,\dots,c; j = 1,2,\dots,m$:

$$\boldsymbol{V_{kj}} = \frac{\sum_{i=1}^{n}({\mu_{ik}}^{w} + {t_{ik}}^{\eta})\ X_{ij}}{\sum_{i=1}^{n}({\mu_{ik}}^{w} + {t_{ik}}^{\eta})} \tag{7}$$

3. Calculate the objective function at the tth iteration using the following equation:

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right]({\mu_{ik}}^{w} + {t_{ik}}^{\eta})\right) \tag{8}$$

4. Fix the relative distinctiveness matrix (μ_{ik})

$$\mu_{ik} = \frac{\left[\sum_{=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{w-1}}}{\left[\sum_{k=1}^{c}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{w-1}}} \tag{9}$$

5. Fix the absolute distinctiveness matrixt (t_{ik})

$$t_{ik} = \frac{\left[\sum_{j=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{\eta-1}}}{\left[\sum_{i=1}^{c}\left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{\eta-1}}} \tag{10}$$

6. Check the iteration stopping condition:
   a. If : $(|P_t - P_{t-1}| < \varepsilon_I$ or $(t > MaxIter)$ then stop
   b. If not: $t = t + 1$, repeat step 4

2.3 Partition Coefficient

Partition Coefficient is an evaluation of the membership degree value, where the optimal cluster is achieved if the value is higher (closer to 1).

$$PC(c) = \frac{1}{N}\sum_{i=1}^{C}\sum_{i=1}^{N}{\mu_{ik}}^2 \tag{11}$$

2.4 Classification Entropy

Classification Entropy is an evaluation that measures the extent of uncertainty in cluster partitions, where the optimal cluster is reached if the value is lower or closer to 0 [5].

2.5 Elbow Method

The Elbow method is a technique that can recommend the optimal number of clusters, resulting in the best results in the Clustering process. The Elbow method analyzes the comparison of the Sum Square Error (SSE) results of each cluster to determine the optimal number of clusters, by identifying the number of clusters that show a significant change in SSE value or form an Elbow [6]. The following steps of the Elbow methods [7].

2.6 Silhoutte Coefficient

The stages in this process are as follows [8]

1. Calculate the average distance between data i and other data in the same cluster

$$a(i) = \frac{1}{|A| - 1}\sum_{J} \in_{A,j\neq i} d(i,j) \tag{14}$$

2. Calculate the average distance between data i and all data in other clusters, then take the smallest value.

$$d(i,C) = \frac{1}{|A| - 1}\sum_{J} \in Cd(i,j) \tag{15}$$

With $d(i,C)$ is the average distance of data i and all objects in another cluster C. Where:

$$b\,(i) = \min C \neq Ad(i,C)$$

3. Silhoutte Coefficient equation

$$S_i = \frac{b_i - a_i}{\max(b_i - a_i)} \tag{16}$$

Where $S(i)$ is the average of all the data in each group. Silhouette Coefficient values range from -1 to 1. The average $S(i)$for all data in a cluster indicates the accuracy of the clustering. The closer to 1, the more precise the resulting clustering structure while a value of -1 indicates overlapping in the clustering structure.

Table 1. Silhoutte Coefficient Criterion Values

Silhoutte	**Interpretation**
0,71-1,00	Very strong cluster structure
0,51- 0,70	Good Cluster Structure
0,26- 0,50	Weak Cluster Structure
<0,25	Poor Cluster Structure

3. RESULT AND ANALYSIS

This research is applied research with a descriptive quantitative research approach carried out at the North Sumatra Provincial Health Office, precisely in the Health Services and Prevention Division and the Disease Control Division The variables used are :

X_1 : Number of health centers
X_2 : Number of Hospitals
X_3: Number of TB
X_4: Number of Diarrhea
X_5: Number of HIV/AIDS
X_6: Number of Dengue Fever

3.1 Explotory Data Analysis

Table 2. Descriptive Statistics of Health in North Sumatra Province

	Health centers	**Hospitals**	**TB**	**Diarrhea**	**HIV/AIDS**	**Dengue Fever**
Median	17	3	906	2873	29	142
Standar Deviasi	10.36749	10.21177	2733.302	7090.374	310.8991	199.8644
Minimum	5	1	72	80	0	11
Maximum	46	57	15722	37885	1800	965

Based on Table 2, the median or middle value for the number of puskesmas is 17, for the number of hospitals is 3, for the number of people with TB disease is 906 people, for the number of people with Diarrhea disease is 2873 people, for the number of people with HIV/AIDS disease is 29 people, and for the number of people with DHF disease is 142 people. Meanwhile, the standard deviation or standard deviation for the puskesmas variable is 10.36749, for the hospital variable is 10.21177, for the number of TB patients is 2733.302, for the number of Diarrhea patients is 7090.374, for the number of HIV/AIDS patients is 310.8991, and for the number of DHF patients is 199.8644.

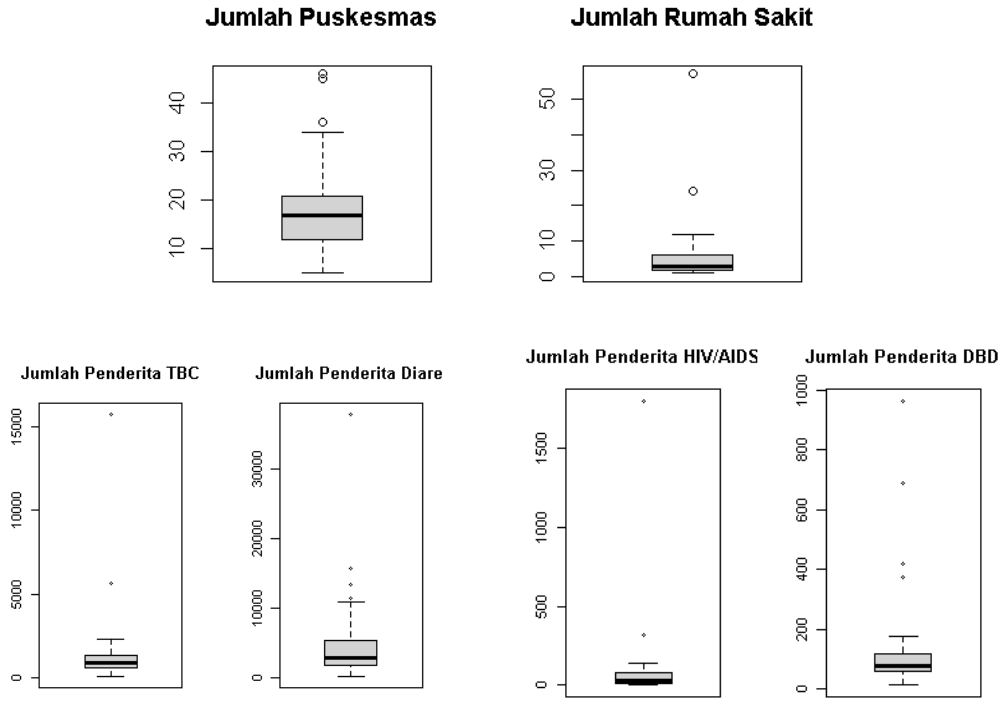

Figure 1. ***Boxplot of Research Variables***

From the boxplots presented, it can be seen that the number of puskesmas has a symmetrical distribution, with many values located above and below the same mean. In contrast, the number of hospitals, diarrhea, HIV/AIDS, and dengue patients tend to be more above the mean. Therefore, these four infectious diseases require special attention from the local government. The Fuzzy Possibilistic C-Means (FPCM) method has the advantage of analyzing data that may contain errors, such as data that are outliers.

3.2 Implementation *Fuzzy C-Means* (FCM) dan Fuzzy Possibilistic C-Means (FPCM)

In FCM the weight used is 2, and the maximum iteration is 1000. On the other hand, in FPCM the weight for Fuzzy C-Means (w) is 2, the weight for Possibilistic C-Means (η) is also 2, with the same maximum iteration limit of 1000. The following is a comparison of the results of the two algorithms based on the number of iterations and the objective function value

Table 3. Number of iterations and objective function

Number of clusters	Number of Iterations		Objective Function	
	FCM	FPCM	FCM	FPCM
Cluster 2	60	145	62.897	126.8311
Cluster 3	56	60	31.31476	125.618
Cluster 4	65	64	16.6139	142.8276
Cluster 5	56	325	10.349	169.3235
Cluster 6	43	80	7.290571	199.7
Cluster 7	143	143	5.39346	0.4753393
Cluster 8	255	161	4.29119	258.8694
Cluster 9	1000	199	3.564228	288.3624
Cluster 10	229	300	3.024007	320.2011

Based on Table 3, the objective function for the Fuzzy C-Means (FCM) method is 62.897, 31.31476, 16.6139, 10.349, 7.290571, 5.39346, 4.29119, 3.564228, and 3.024007, respectively. As for the Fuzzy Possibilistic C-Means (FPCM) method, the objective function values are 126.8311, 125.618, 142.8276, 169.3235, 199.7, 0.47533393, 258.8694, 288.3624, and 320.2011. Comparison of objective functions between the two methods cannot be done directly due to different objective function calculations in each method. The result of the partition matrix μik affects the objective function value, where a smaller μik value tends to produce a lower objective function value. From the number of iterations, it can be concluded that the Fuzzy Possibilistic C-Means (FPCM) method generally requires a greater number of iterations compared to the Fuzzy C-Means (FCM) method to achieve objective function minimization, so the FPCM method requires longer computation time.

3.3 Determination the best method

Determining the most effective method can be done by referring to the cluster validity index values, such as Partition Entropy (PE) and Partition Coefficient (PC). The following are the Partition Entropy values of FCM and FPCM.

Table 4. Partition Entrophy Comparison

Number of Cluster	*Partition Entrophy*	
	FCM	FPCM
Cluster 2	0.09622375	0.08680947
Cluster 3	0.3655486	0.3643184
Cluster 4	0.4580757	0.4566024
Cluster 5	0.6821143	0.6342542
Cluster 6	0.7226973	0.716146
Cluster 7	0.9373028	0.9344659
Cluster 8	1.075373	1.058198
Cluster 9	1.181628	1.102238
Cluster 10	1.245365	1.159859

According to Table 4, the Partition Entropy (PE) index is considered good if the value is close to a small number or close to 0. Based on the Partition Entropy (PE) index, the Fuzzy Possibilistic C-Means (FPCM) algorithm shows superior performance compared to the Fuzzy C-Means (FCM) algorithm. This is because the Partition Entropy (PE) index value for the FPCM algorithm tends to be smaller or closer to 0 in all clusters.

The following are the results of the cluster validity index, namely Partition Coefficient (PC), for the Fuzzy C-Means (FCM) and Fuzzy Possibilistic C-Means (FPCM) algorithms.

Table 5. *Partition CoefficientComparison*

Number Of Cluster	*Partition Coefficient (PC)*	
	FCM	FPCM
Cluster 2	0.9512696	0.9564018
Cluster 3	0.792744	0.7931077
Cluster 4	0.7424156	0.7428776
Cluster 5	0.6234597	0.6410457
Cluster 6	0.6197362	0.6065045
Cluster 7	0.5184219	0.5197595
Cluster 8	0.4781403	0.4677202
Cluster 9	0.4523551	0.4627803
Cluster 10	0.4448933	0.4526294

Based on Table 5, the Partition Coefficient (PC) index is considered good if the value is close to a large number or close to 1. Based on the Partition Coefficient (PC) index, the FPCM algorithm shows superior performance compared to the FCM algorithm. This is due to the Partition Coefficient (PC) index value for the FPCM algorithm which tends to approach a large number or close to 1 in almost all clusters.

3.4 Determination of the optimal clusters

Determining number of clusters is using the Elbow method. The Elbow method will show the optimal solution as determining the number of clusters as follows.

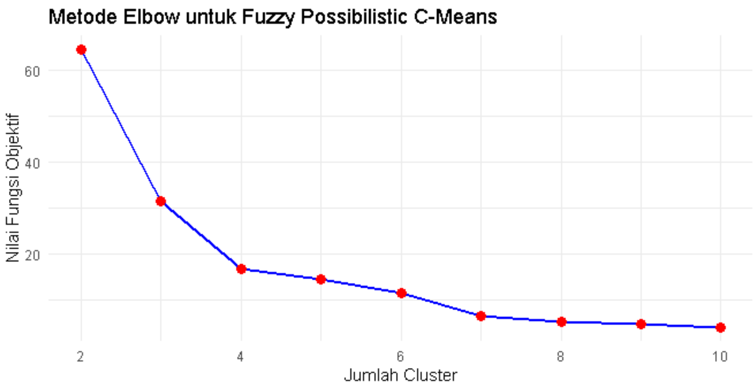

Figure 2. Elbow Method Plot

Based on Figure 2, plot results of the Elbow method, the optimal number of clusters selected is 3. The Elbow plot shows that the most optimal number of clusters for data on the number of health facilities and infectious disease cases is 3. This is based on the Elbow point where the decrease in SSE begins to slow down significantly after 3 clusters.

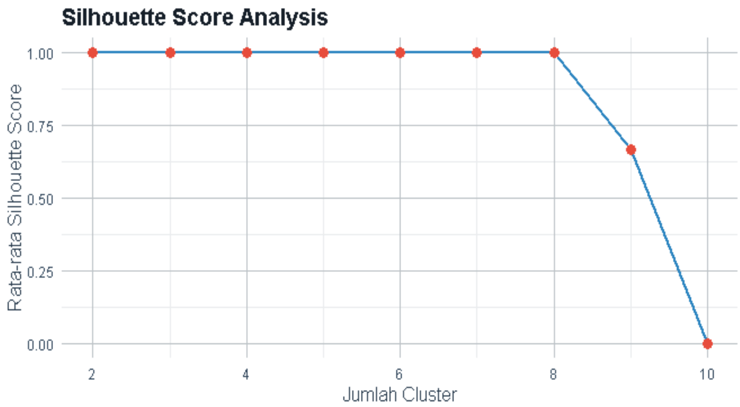

Figure 3 Silhuette Coefficient Plot

The graph shows that the number of clusters 2 to 8 reaches the optimum point of 1, meaning that the cluster structure is strong. Meanwhile, when the number of clusters is 9 and 10, the average Silhoutte score value decreases, indicating that the cluster structure is less strong. Based on the two methods of determining the number of clusters, it can be concluded that the number of clusters 3 shows the optimum point.

3.5 Fuzzy Possibilistic C – Means Result Profilization

The membership results of the Regency / City Cluster in North Sumatra Province using the FPCM algorithm are described in the following table.

Table 6. Result from 3 Clusters

Cluster	District/City
Cluster 1	Padang Sidempuan City, Pematang Siantar City, Sibolga City, Tanjung Balai City, Tebing Tinggi City, Labuhan Batu, Labuhan Batu Selatan, Labuhan Batu Utara, Batu Bara, Nias, Nias Barat, Nias Utara, Padang Lawas, Padang Lawas Utara, Pakpak Bharat, Samosir, Serdang Berdagai, Dairi, South Tapanuli, North Tapanuli, Central Tapanuli, Toba, Humbang Hasundutan, Binjai City, Gunung Sitoli.
Cluster 2	Asahan, Langkat, Mandailing Natal, Nias Selatan, Simalungun, Deli Serdang, Karo
Cluster 3	Medan city

Based on table 6 which same characteristcs Puskesmas , hospitals and the infectious diseases as TB, diarrhea, HIV/AIDS and dengue fever as cluster 1 has twenty five district/cities, cluster 2 has seven district/cities and cluster 3 only one city namely Medan city.

Table 7. Result Profilization of 3 clusters

Cluster	**Health centers**	**Hospitals**	**TB**	**Diarrhea**	**HIV/AIDS**	**Dengue Fever**	**Many members**
Cluster 1	14	3	795	2547	42	88	25
Cluster 2	32	8	2057	14075	89	218	7
Cluster 3	45	57	15722	10825	1800	965	1

Based on the table 7, we have 3 clusters which is cluster 1 has 25 members, cluster 2 has 7 members and cluster 3 only 1 member. Cluster 3 consist of one member namely Medan City which is 45 health centers, 57 hospitals, and also the highest diseases is 15722 TB, 1800 HIV/AIDS, 965 Dengue Fever. The urgency level of Regency / City Health in North Sumatra Province based on cluster results and supported by population and area is described in Table 8 as follows:

Table 8. Health Urgency Level in North Sumatra Province

Urgency Level	*Cluster*	Description
1	*Cluster* 3 (Medan City)	• Highest number of cases for all health indicators • Low number of health services relative to population and disease burden. • Medan City is the capital of the province with a high population density, a relatively

		small area compared to its population, and diverse livelihoods (industry, services, trade). The high health rates indicate a very high urgency for addressing health issues. • Medan City requires the most urgent and comprehensive health interventions, including hospital capacity building, HIV/AIDS, TB, and dengue prevention and treatment programs.
2	*Cluster* 2 (Asahan, Langkat, Mandailing Natal, Nias Selatan, Simalungun, Deli Serdang, Karo)	• High number of health cases. • Districts/Municipalities with medium population density, larger area than Cluster 3. • Mixed livelihoods (agriculture, small industry, services). • The high number of diarrhea cases suggests there may be sanitation issues that need to be addressed immediately.
3	*Cluster* 1 (Pematang Siantar, Sibolga, Tanjung Balai, Tebing Tinggi, Labuhan Batu, Labuhan Batu Selatan, Labuhan Batu Utara, Batu Bara, Nias, Nias barat, Nias Utara, Padang Lawas, Padang Lawas Utara, Pakpak Bharat, Samosir, Serdang Berdagai, Dairi, Tapanuli Selatan, Tapanuli Tengah, Tapanuli Utara, Toba, Humbang Hasundutan, Kota Binjai, Gunung sitoli)	• The number of health cases is the lowest among the three clusters. • Districts/Municipalities with low population density are relatively large in area, and dominant livelihoods in agriculture or fisheries. • Although infectious diseases are lower, it should be noted that access to health services may be limited (small number of health centers and hospitals).

Based on Table 8, it can be concluded that the level of urgency of each cluster by considering the available health data, as well as considering the factors of population density, area, and regional livelihoods are as follows:

1) Cluster 3 (Medan City) requires the most urgent and comprehensive health interventions, including hospital capacity building, HIV/AIDS, TB, and dengue prevention and treatment programs.
2) Cluster 2 requires a focus on improving sanitation and preventing diarrhea, as well as strengthening the health system to deal with TB cases.
3) Cluster 1, despite having the lowest rates, needs improved access to health services, especially in remote areas.

4 CONCLUSION

Based on the analysis and discussion that has been carried out, it can be concluded that:

1. In this study, Fuzzy Possibilistic C-Means is proven to be a better algorithm than Fuzzy C-Means (FCM) based on the comparison of Partition Entrophy (PE) and Partition Coefficient (PC) cluster validity index values. Fuzzy Possibilistic C-Means shows higher PC value and lower PE than FCM, this statistically proves that FPCM provides more optimal clustering results.
2. The Optimal Number of Clusters is 3 which is determined by Elbow Method and Sillhoutte Coefficient. In the Elbow method, researchers made a plot showing the total variance (within-cluster sum of squares) against various numbers of clusters. The point at which the graph begins to flatten out is considered the optimal number of clusters, in this case 3 clusters. The Silhouette Coefficient complements the Elbow method by providing a quantitative assessment of the clustering quality. The combination of these two methods provides a comprehensive approach in determining the optimal number of clusters.
3. Profiling on the cluster results of districts/cities in North Sumatra province based on health service facilities and number of infectious diseases shows that Cluster 3, Medan City, is the area with the highest level of health urgency in North Sumatra province. This is due to the most urgent and comprehensive health interventions, including hospital capacity building, HIV/AIDS, TB, and dengue prevention and treatment programs. The relatively small area with a large population places a high burden on available health facilities. The very high population density, especially as the capital of the province, increases the risk of spreading infectious diseases. Meanwhile, Cluster 2, namely the Districts of Asahan, Langkat, Mandailing Natal, South Nias, Simalungun, Deli Serdang, Karo, occupies the second level of urgency which requires special attention to improving sanitation and preventing diarrhea, as well as strengthening

the health system to deal with high TB cases. In the last order of Cluster 1, Pematang Siantar City, Sibolga City, Tanjung Balai City, Tebing Tinggi City, Labuhan Batu, Labuhan Batu Selatan, Labuhan Batu Utara, Batu Bara, Nias, West Nias, North Nias, Padang Lawas, Padang Lawas Utara, Pakpak Bharat, Samosir, Serdang Berdagai, Dairi, South Tapanuli, Central Tapanuli, North Tapanuli, Toba, Humbang Hasundutan, Binjai City, Gunung Sitoli have the lowest rates of health problems, there is still a need to improve acsess to health services, especially in remote underserved area.

REFERENCES

[1] *BPS Sumut.,(2023) Jumlah Kasus Penyakit Menurut Kabupaten/Kota dan Jenis Penyakit di Provinsi Sumatera Utara* [https://sumut.bps.go.id/statictable/2023/03/30/3064/jumlah-kasus-penyakit-menurut-kabupaten-kota-dan-jenis-penyakit-di-provinsi sumatera-utara-2022.html](https://sumut.bps.go.id/statictable/2023/03/30/3064/jumlah-kasus-penyakit-menurut-kabupaten-kota-dan-jenis-penyakit-di-provinsi-sumatera-utara-2022.html)

[2] Thilagaraj, T., & Sengottaiyan, N. (2019). Implementation of *Fuzzy C-Means* and fuzzy possibilistic c-means algorithms to find the low performers using r-tool. *International Journal of Scientific & Technology Research, 8.*

[3] Kuo, R. J., Alfareza, M. N., & Nguyen, T. P. Q. (2023). Genetic based density peak possibilistic *Fuzzy C-Means* algorithms to *Cluster* analysis-a case study on customer segmentation. *Engineering Science and Technology, an International Journal*, *47*, 101525.

[4] Martinez, W. L., Martinez, A. R., & Solka, J. (2017). *Exploratory data analysis with MATLAB*. Chapman and Hall/CRC.

[5] Mulyaningsih, W. S. (2021). *Implementasi Fuzzy C-Means Dan Fuzzy Possibilistic C-Means Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Banten (Studi Kasus: Fasilitas Pelayanan Kesehatan Dan Penyakit Menular Di Provinsi Banten Tahun 2019).*

[6] Nugraha, G. S., Dwiyansaputra, R., Bimantoro, F., & Aranta, A. (2023). Implementasi Fuzzy C-Means untuk Pengelompokan Daerah berdasarkan Persebaran Penularan Covid-19. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *10*(1), 97-104.

[7] Putra, A. C., & Hartomo, K. D. (2021). Optimalisasi Penyaluran Bantuan Pemerintah Untuk UMKM Menggunakan Metode *Fuzzy C-Means*. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *5*(3), 474-482.

[8] Saputra, P. R. N., & Chusyairi, A. (2020). Perbandingan metode *Clustering* dalam pengelompokan data puskesmas pada cakupan imunisasi dasar lengkap. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *4*(6), 1077-1084.

[9] Fuadah, A. W., Arifin, F. N., & Juwita, O. (2021). Optimasi K-Klasterisasi Ketahanan Pangan Kabupaten Jember Menggunakan Metode *Elbow*. *INFORMAL: Informatics Journal*, *6*(3), 136-141