Clustering the Distribution of COVID-19 in Aceh Province Using the Fuzzy C-Means Algorithm

Nurdin[1], Suci Fitriani[2], Zara Yunizar[3], Bustami[4]

[1,2,3,4]Departement of Informatics, Universitas Malikussaleh, Aceh, Indonesia

nurdin@unimal.ac.id*[1], suci.170170050@gmail.com[2], zarayunizar@unimal.ac.id[3]
busabiel@gmail.com[4]

ABSTRACT

COVID-19 is a virus that attacks the respiratory system in humans and spreads rapidly. The government has taken various ways to reduce the rate of transmission of COVID-19, including by providing a COVID-19 information center that can be accessed by anyone, but there is no grouping of regional zones with high to low COVID-19 cases. Therefore, a clustering process system for the spread of COVID-19 is needed so that it is able to provide information on clusters of COVID-19 distribution areas in Aceh with the highest case zone (red zone), medium case zone (yellow zone), and low case zone (green zone). The steps carried out in this study using the Fuzzy C-Means Algorithm are collecting data (input data), conducting the clustering process (determining the number of clusters, weighting rank, maximum iteration and epsilon), displaying clustering results. In this study, the authors collected COVID-19 data from 23 districts/cities in Aceh using 6 variables consisting of confirmed, in care, healed, died, suspected, and probable. The results of the clustering study on the spread of COVID-19 are as follows: One district/city in cluster 1 (red zone), the four districts/cities in cluster 2 (yellow zone), eighteen districts/cities in cluster 3 (green zone). Based on the results of this study, the Fuzzy C-Means Algorithm can be used and applied properly in clustering the spread of COVID-19 in the Province of Aceh.

—————————— ◆ ——————————

A. INTRODUCTION

At the end of 2019, the world was shocked when the emergence of the corona virus was first reported in Wuhan, China (Rothan & Byrareddy, 2020). In December 2019, in Wuhan there were five patients who were treated with Acute Respiratory Distress Syndrome (ARDS) (Ren et al., 2020). Coronavirus Disease 2019 (COVID-19) is a virus that attacks the respiratory system in humans with symptoms of cough, fever, runny nose, shortness of breath, sore throat. This disease must be watched out for because of its relatively fast transmission (Susilo et al., 2020). On January 20, 2020, WHO as the World Health Organization established the status of Public Health Emergency of International Concern (PHEIC) as a form of warning that this virus is very risk (Suni, 2020; World Health Organization (WHO), 2020).

The Indonesian government has provided a COVID-19 information center website that provides information for the number of confirmed positive cases, recovered, died, etc. in real time, and there is even information on the grouping of areas with high, medium, low risk, and

areas with no COVID-19 cases. which is accessed through the address: http://covid19.go.id. However, the grouping carried out on the website is considered to be less accurate because it is not clear that the calculation process to obtain the grouping includes the process for clustering the regions in Indonesia (Farizi & Harmawan, 2020).

On the website of the COVID-19 information center for the Province of Aceh, there is no regional grouping with the highest COVID-19 cases to the lowest cases. The purpose of this study is to cluster the distribution of areas with the highest, medium and low COVID-19 cases. This grouping of zones is really needed by the community to stay alert, practice social distancing and comply with health protocols in carrying out activities or when traveling to other areas that are included in the group of the highest COVID-19 cases. With the amount of data needed to perform this grouping process, a data mining clustering technique is needed using the Fuzzy C-Means Algorithm. Fuzzy C-Means is a technique for clustering or grouping data where the existence of each data in a cluster is determined by the degree of membership.

There are several related studies that have been carried out, such as that conducted by (Noviyanto, 2020) in his research using only the parameter of the number of deaths due to COVID-19 from countries in the Asian continent using the k-means algorithm and Rapid Miner software so that the results of clustering the number of deaths are obtained. The consequences of COVID-19 are 4 countries with high clusters, 4 countries with moderate clusters, and 41 countries with low clusters. Another study conducted by (Solichin & Khairunnisa, 2020) in this study clustering was carried out based on the parameters of the number of people under surveillance, patient in care, positive cases, recovered patients, and patients died using the K-Means method and the Eucledean distance measurement method.

Research conducted by (Doroshenko, 2020) discusses the clustering of the spread of the COVID-19 virus in Italy during February-April 2020 using the K-Means Clustering Algorithm. This study uses 1113 data sample lines to be grouped describing the epidemiological situation in Italy from 20 February 2020 to 16 April 2020 based on 17 parameters, namely date, country, region code, region name, latitude, longitude, hospitalized patients, intensive care patients, total hospitalizations, hospitalizations, current positive cases, new positive cases, recoveries, deaths, total positive cases, and tests performed. This algorithm divides the area into clusters that describe the division of the area according to their geography. The results of this study yield a high accuracy, which is 97% and the clusters formed show the division of areas with high incidence (blue cluster, C1), and low incidence (red cluster, C2). In this study, the k-means algorithm is used, while the research that the author is doing uses the fuzzy c-means algorithm.

Research conducted by (Crnogorac et al., 2021) grouped European countries based on the cumulative relative number of COVID-19 patients in Europe in 2020. This study aims to group European countries and regions into clusters, where countries from a cluster have The similarity in the value of the cumulative number of COVID-19 cases is carried out using a dataset containing the cumulative number of COVID-19 cases taken every day for 14 days for each European country and region per 100,000 population. The grouping uses 3 clustering methods, namely K-Means, agglomerative and BIRCH clustering. The results of this study were able to group European countries and regions into 5 clusters based on the cumulative relative number of COVID-19 cases by testing performance using the Silhouette coefficient

value and showing good accuracy results. In this study, European countries and regions were grouped into 5 clusters based on the cumulative relative number of COVID-19 cases using 3 clustering methods, namely K-Means, agglomerative and BIRCH clustering.

In a study conducted by (S et al., 2021) grouped the districts in Tamilnadu based on their mobility to the most important places which were categorized into 6 groups, such as retail and recreation, grocery stores and pharmacies, parks, transit stations, workplaces, and residential areas so that the district groupings with high, medium, and low mobility were obtained. In this study, Fuzzy Clustering uses a Hybrid CSO-PSO search based on the movement of a group of people during the implementation of the COVID-19 Lockdown. While the research that the author did using dataset clustering was carried out based on the parameters of the number of people under surveillance, patients in treatment, positive cases, patients recovered, and patients died.

Another study conducted by (Kurniawan et al., 2021) used clustering and correlation methods to predict and analyze the risk of COVID-19 in countries exposed to the pandemic. The clustering method used is K-Means to group 200 countries in the world with 10 attributes, namely countries, total cases, new cases, total deaths, new deaths, total recovered, active cases, seriously critical, probability of death (%) and probability of recovering. (%). This research has succeeded in classifying 200 countries in the world with 10 attributes and 5 clusters. The results of the correlation analysis showed a strong positive linear correlation between the total number of COVID-19 cases and the number of deaths (0.78) and between the number of deaths and critical patients (0.85). This proves that if one variable increases, the other variables also increase. The K-Means Clustering method is widely applied to solve data clustering problems, such as grouping 142 countries into 4 clusters to identify the best strategy in fighting COVID-19 (Darapaneni et al., 2021) and clustering air quality data around Uttarakhan, India during the COVID-19 pandemic lockdown (Sunori et al., 2021). Research that the author did using dataset clustering was carried out based on the parameters of the number of people under surveillance, patients in treatment, positive cases, patients recovered, and patients died using the K-Means method and the Eucledean distance measurement method.

As a result of the COVID-19 pandemic that has hit the world, it has had a major impact on various sectors of life, such as the economic sector (Fernandes, 2020; Ozili & Arun, 2020), the environment sector (Zambrano-Monserrate et al., 2020), and the education sector (Akat & Karatas, 2020) which implemented an online learning system during the COVID-19 pandemic (König et al., 2020; Nurdin et al., 2022). To prevent an increase in COVID-19 cases, a clustering system for the COVID-19 distribution zone is needed using the Fuzzy C-Means Algorithm.

Based on the background of the problems described above and previous related research, no one has discussed and conducted research on clustering the spread of COVID-19 in Aceh Province using the Fuzzy C-Means Algorithm. Fuzzy C-Means Clustering is considered more accurate because each data is reallocated into each cluster during the iteration process by utilizing fuzzy set theory, where each data has the possibility to be able to join each cluster based on its membership degree(Muslimatin, 2011).

B. METHODS

1. Fuzzy C-Means Algorithm

In grouping there are 2 fuzzy techniques that can be used, namely Fuzzy Hashing and Fuzzy C-Means (Naik et al., 2018). Fuzzy C-Means is a data grouping method proposed by Bezdek by utilizing the concept of fuzzy theory (Bezdek et al., 1984). Fuzzy C-Means applies fuzzy grouping, where each data can be a member of several clusters with different degrees of membership in each cluster. Fuzzy C-Means is an iterative algorithm, which applies iteration to the data clustering process. The purpose of Fuzzy C-Means is to get the center of the cluster which will later be used to find out the data that enters the cluster. According to (Agustini, 2017) the advantage of the Fuzzy C-Means algorithm compared to other algorithms is that it can place the cluster center more precisely, namely by repairing the cluster center repeatedly so that the cluster center will move to the right cluster center location/point. In addition, the Fuzzy C-Means algorithm also has a smaller probability of failure than the k-means algorithm because the Fuzzy C-Means clustering process allows each data to be in two or more clusters. To apply Fuzzy C-Means clustering in several cases, the first step is to make the number of classes that will be used as the basis for grouping, then iterate (iteration) to get the membership of the group. In data grouping the Fuzzy C-Means Algorithm can also be combined with the Fuzzy Support Vector Machine (Shan & Zhi, 2016).

The following are the stages of the Fuzzy C-Means Algorithm (Agustini, 2017).

a. Determine the data to be in cluster *X*, in the form of a matrix measuring *n x m* (*n*=number of data samples, *m*=attributes of each data). *Xkj* = *k*-th sample data (*k*=1, 2,...,*n*), *j*-th attribute (*j*=1,2,...,*m*).

b. Determine: number of clusters (*C*), weighting power (*w*), maximum iteration (*MaxIter*), minimum expected termination criteria/error (ζ).

c. Generating random numbers $\_ik$, i=1,2,3 ..., *C*; *k*=1,2,3 ...n; as elements of the initial partition matrix *U* with degrees of membership. Counting the sum of each column (attribute or variable):

$$Q_j = \sum_{i=1}^{C} \mu_{ik} \tag{1}$$

With *j*=1, 2, ..., *m* (attribute or data variable). Then calculate:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j} \tag{2}$$

d. Calculating the cluster to-*i* : V_{ij} , with *i*= 1, 2, ..., *C* , *j*=1, 2, ..., *m*

$$V_{ij} = \frac{\sum_{k=1}^{n}((\mu_{ik})^w * X_{kj})}{\sum_{k=1}^{n}(\mu_{ik})^w} \tag{3}$$

Where,

V_{ij} = cluster center value to-*i* on variable to-*j*

μ_{ik} = data membership function to-*k* (*k*= 1, 2,....., n) or elements of the partition matrix *U*   on cluster to *i* (*i*= 1, 2,....., *C*)

w   = rank (weighting), $w \in [1,\infty]$

X_{kj} = data to-*k* on variable to-*j*

e. Calculating the objective function on iteration to-*t*:

$$J_w(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{C}((\mu_{ik})^w d_{ik}^2) \tag{4}$$

Where,

$J_w(U,V)$ = objective function of *U* and *V*

C = the number of clusters in X

n = the amount of data processed

w = rank (weighting), $w \in [1,\infty]$

X = the processed data is in the form of a matrix n x m

U = initial partition matrix

V = cluster center matrix

μ_{ik} = data membership function to-k (k= 1, 2,....., n) or elements of the partition matrix U on the cluster to i (i= 1, 2,....., C)

d_{ik} = distance function

$$d_{ik} = d(X_k - V_i) = [\textstyle\sum_{j=1}^{m}(X_{kj} - V_{ij})] \tag{5}$$

Where,

d_{ik} = the distance of each data to the center of the cluster

X_k = data to-k with (k=1, 2,...., n)

V_i = cluster center value to-i, with (i= 1, 2,..., C)

m = many variables (criteria), with (j= 1, 2,....., m)

f. Calculating the partition matrix change:

$$\mu_{ik} = \frac{[\sum_{j=1}^{m}(X_{kj}-V_{ij})^2]^{\frac{-1}{w-1}}}{\sum_{k=1}^{C}[\sum_{j=1}^{m}(X_{kj}-V_{ij})^2]^{\frac{-1}{w-1}}} \tag{6}$$

Where,

μ_{ik} = data membership function to-k (k= 1, 2,....., n) or elements of the partition matrix U on the cluster to-i (i= 1, 2,....., C)

w = rank (weighting), $w \in [1,\infty]$

g. Checking the stop condition:

If ($|P_t – P_{t-i}| < \zeta$) or ($t >$ MaxItr), then the iteration stops

If not: t = t+1, repeat starting from step 4.

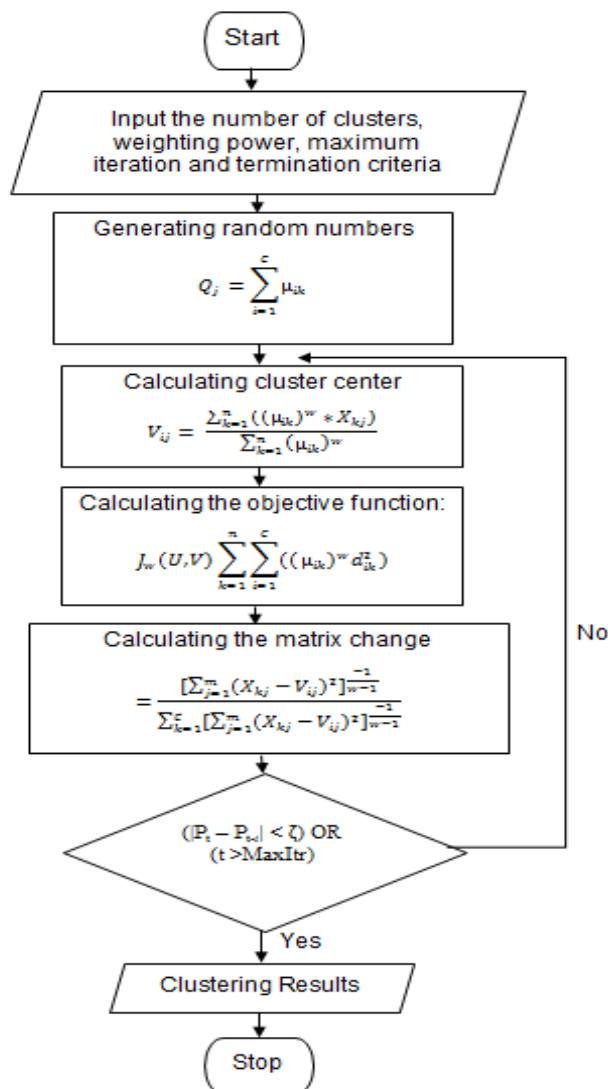The following flowchart of the Fuzzy C-Means Algorithm can be seen in Figure 1.

Figure 1. Flowchart of the Fuzzy C-Means Algorithm

2. Data Collection and Variable Type

The data was taken through the website for the COVID-19 information center in Aceh, namely: http://covid19.acehprov.go.id which consists of 23 districts/cities in Aceh province with the following variables used:

a. Confirmed number, namely the number of people who are confirmed positive for COVID-19 after a lab test is carried out even though they are asymptomatic.

b. The number of patients in care, namely the number of patients hospitalized due to COVID-19.

c. The number of recovered patients is the number of patients who have been hospitalized and have recovered.

d. The number of deaths, namely the number of people who died from COVID-19.

e. The number of suspects, namely the number of people who have symptoms of COVID-19 but have not been tested.

f. Probable number, namely the number of people who have symptoms with positive rapid test results, but have not carried out a PCR lab test.

C. RESULT AND DISCUSSION

1. Dataset dan Variable

Based on data obtained from the Aceh Covid-19 information website, via https://covid19.acehprov.go.id/, clustering or zoning groupings will be carried out based on the level of cases in Aceh Province. In this study, the authors used a dataset of 23 districts/cities in Aceh Province are South Aceh, Southeast Aceh, East Aceh, Central Aceh, West Aceh, Big Aceh, Pidie, North Aceh, Simelue, Aceh Singkil, Bireun, Southwest Aceh, Gayo Lues Aceh Jaya, Nagan Raya, Aceh Tamiang, Bener Meriah, Pidie Jaya, Banda Aceh, Sabang, Lhokseumawe, Langsa, and Subulussalam. The variables that will be used for the clustering process are the number of confirmed cases, the number of patients in treatment, the number of recovered patients, the number of deaths, suspects, and probable. Furthermore, these 6 variables are initialized in Table 2.

Table 2. Variable Name

No	Variable Name	Initials
1	Confirmed	X1
2	In Care	X2
3	Healed	X3
4	Die	X4
5	Suspect	X5
6	Probable	X6

The initials of this variable name will be used in this manual calculation using the Fuzzy C-Means algorithm. The COVID-19 dataset that will be used in this calculation process is obtained from the website of the COVID-19 information center for the province of Aceh, which can be seen in Table 3.

Table 3. Aceh Province Covid-19 Dataset

No	Districts/City	X1	X2	X3	X4	X5	X6
1	South Aceh	233	6	207	20	265	6
2	Southeast Aceh	115	13	100	2	306	1
3	East Aceh	113	36	73	4	131	71
4	Central Aceh	210	20	186	4	233	30
5	West Aceh	128	16	106	6	170	44
6	Big Aceh	1601	252	1292	57	183	25
7	Pidie	406	6	354	46	256	50
8	North Aceh	205	8	183	14	379	5
9	Simelue	109	0	101	8	1	53
10	Aceh Singkil	203	0	197	6	7	3
11	Bireun	480	3	460	17	443	47
12	Southwest Aceh	91	1	82	8	202	47
13	Gayo Lues	228	13	211	4	60	0
14	Aceh Jaya	136	2	128	6	67	1
15	Nagan Raya	159	1	142	16	243	187
16	Aceh Tamiang	380	88	274	18	232	9
17	Bener Meriah	163	37	120	6	99	32
18	Pidie Jaya	218	0	202	16	322	0

19	Banda Aceh	2836	548	2213	75	984	16
20	Sabang	92	1	82	9	62	52
21	Lhokseumawe	428	98	314	16	1608	1
22	Langsa	329	62	256	11	384	11
23	Subulussalam	85	2	76	7	16	7

The table above is a COVID-19 dataset obtained from the website of the Aceh Province COVID-19 information center to be used in this calculation process.

2. Calculation Result with Fuzzy C-Means Algorithm

After doing manual calculations using the stages of the Fuzzy C-Means Algorithm starting from step 1 to step 7, the final result of the calculation of the clusterization of the spread of COVID-19 in Aceh Province for each data with the calculation process up to 43 iterations can be seen in Table 4.

Table 4. Clustering Results

No	Degree of Membership in the th cluster			Selected cluster	*Cluster*
	1	**2**	**3**		
1	0.00149	0.05611	0.94238	0.94238	3
2	0.00210	0.06233	0.93555	0.93555	3
3	0.00112	0.02356	0.97531	0.97531	3
4	0.00059	0.01939	0.98000	0.98000	3
5	0.00037	0.00887	0.99074	0.99074	3
6	0.36036	0.38597	0.25365	0.38597	2
7	0.00332	0.86780	0.12886	0.86780	2
8	0.00367	0.16934	0.82697	0.82697	3
9	0.00296	0.05011	0.94692	0.94692	3
10	0.00272	0.05227	0.94500	0.94500	3
11	0.00510	0.83998	0.15491	0.83998	2
12	0.00109	0.02567	0.97322	0.97322	3
13	0.00182	0.04030	0.95786	0.95786	3
14	0.00121	0.02411	0.97466	0.97466	3
15	0.00251	0.06950	0.92797	0.92797	3
16	0.00551	0.23266	0.76181	0.76181	3
17	0.00060	0.01291	0.98648	0.98648	3
18	0.00248	0.10476	0.89274	0.89274	3
19	0.99114	0.00510	0.00374	0.99114	1
20	0.00197	0.03623	0.96179	0.96179	3
21	0.06765	0.65233	0.28000	0.65233	2
22	0.00539	0.38221	0.61239	0.61239	3
23	0.00307	0.05167	0.94524	0.94524	3

The table above is a clustering of calculation results using the fuzzy c-means algorithm until the 43rd iteration which consists of 18 regions entering cluster 3, 4 regions entering cluster 2, and 1 area entering cluster 1. Based on the results of manual calculations with 43

iterations, the results of clustering with the degree of membership of each data are shown in Figure 2.
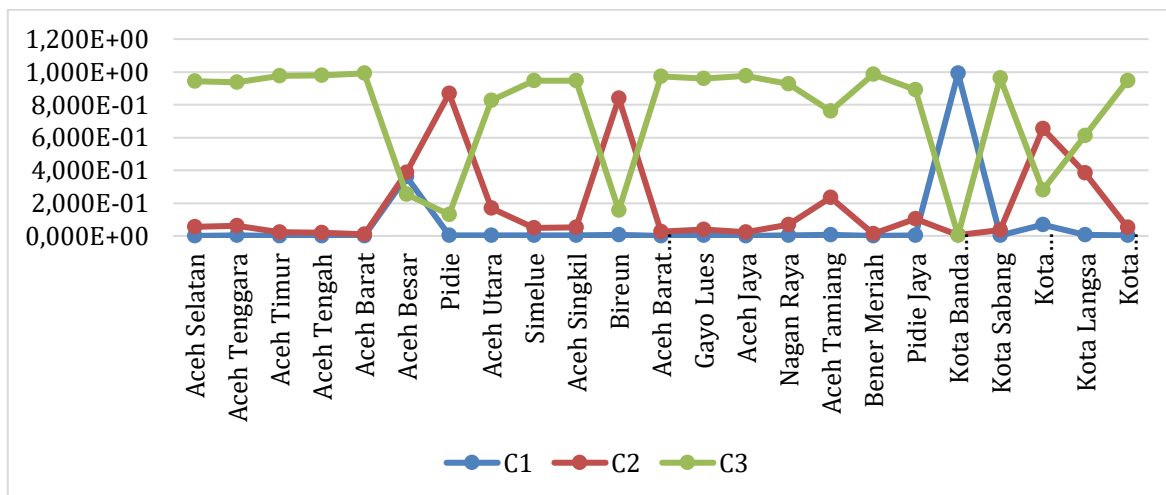

Figure 2. Last Iteration Membership Degree Chart

The picture above is a graph of the results of manual calculations in 43 iterations with the degree of membership of each data. Based on Table 4 above, the results of clustering can be presented in the form of a diagram, as shown in Figure 3.
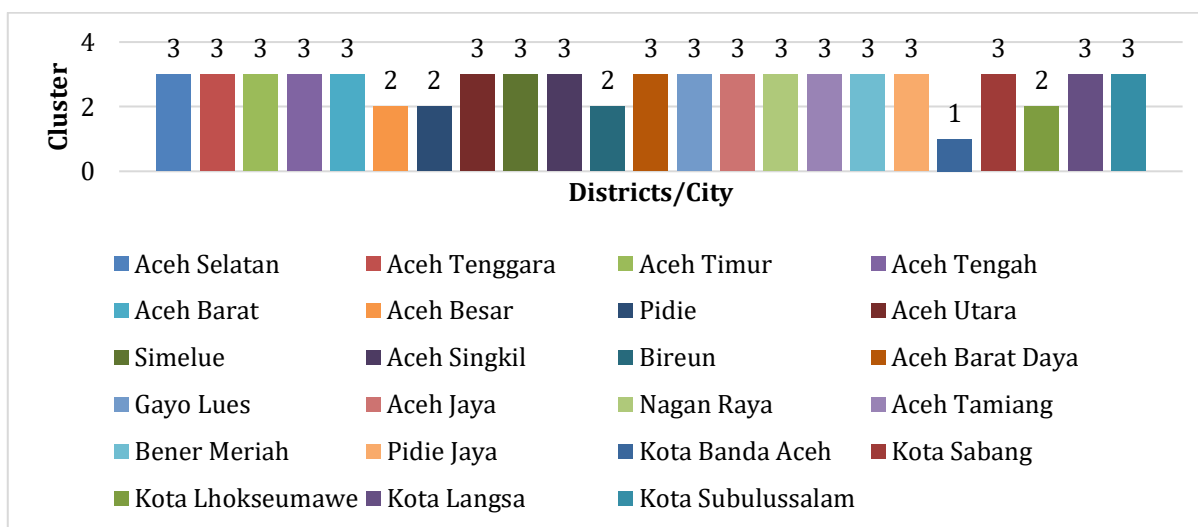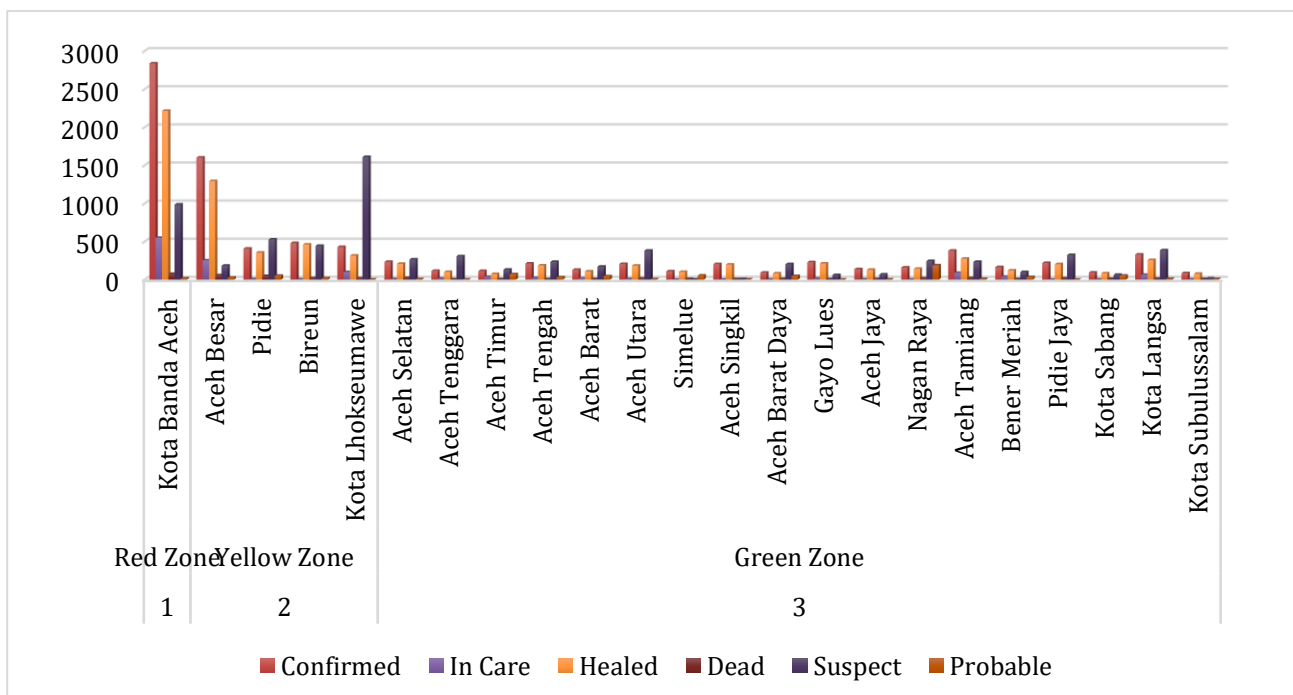

Figure 3. Diagram of the Clustering Results of the Distribution of Covid-19 in Aceh Province

The picture above is a graph of the results of the clustering the spread of Covid-19 in Aceh. Table 5. This is the final result of the clustering of the spread of COVID-19, where cluster 1 as the red zone consists of 1 district/city, namely Kota Banda Aceh. Cluster 2 as the yellow zone consists of 4 districts/cities, namely Big Aceh, Pidie, Bireun, and Kota Lhokseumawe. Cluster 3 as a green zone consists of 18 districts/cities South Aceh, Southeast Aceh, East Aceh, Central Aceh, West Aceh, North Aceh, Simelue, Aceh Singkil, Southwest Aceh, Gayo Lues, Aceh Jaya, Nagan Raya, Aceh Tamiang, Bener Meriah, Pidie Jaya, Sabang, Langsa, and Subulussalam, as shown in Table 5.

Table 5. Zoning for the Spread of Covid-19 in Aceh Province

Cluster	Zone	Districts/City	X1	X2	X3	X4	X5	X6
1	Red	Banda Aceh	2836	548	2213	75	984	16
2	Yellow	Big Aceh	1601	252	1292	57	183	25
		Pidie	406	6	354	46	256	50
		Bireun	480	3	460	17	443	47
		Lhokseumawe	428	98	314	16	1608	1
3	Green	Southwest Aceh	233	6	207	20	265	6
		Southeast Aceh	115	13	100	2	306	1
		East Aceh	113	36	73	4	131	71
		Central Aceh	210	20	186	4	233	30
		West Aceh	128	16	106	6	170	44
		North Aceh	205	8	183	14	379	5
		Simelue	109	0	101	8	1	53
		Aceh Singkil	203	0	197	6	7	3
		Southwest Aceh	91	1	82	8	202	47
		Gayo Lues	228	13	211	4	60	0
		Aceh Jaya	136	2	128	6	67	1
		Nagan Raya	159	1	142	16	243	187
		Aceh Tamiang	380	88	274	18	232	9
		Bener Meriah	163	37	120	6	99	32
		Pidie Jaya	218	0	202	16	322	0
		Sabang	92	1	82	9	62	52
		Langsa	329	62	256	11	384	11
		Subulussalam	85	2	76	7	16	7

The table above is the result of the clustering of the COVID-19 distribution zones in Aceh based on each calculation variable. The following is a zoning graph of the spread of COVID-19 in Aceh Province based on the results of the distribution zoning in Figure 4.

Figure 4. Zoning Graph of Covid-19 Distribution in Aceh Province

The picture above is a zoning graph of the results of the clustering of the spread of Covid-19 in Aceh.

3. Implementation of System Interface

The following is the dashboard page interface. On the output page of this system, it contains data information and information on the results of clustering the distribution of COVID-19 in Aceh Province using the Fuzzy C-Means Algorithm, which can be seen in Figure 5.
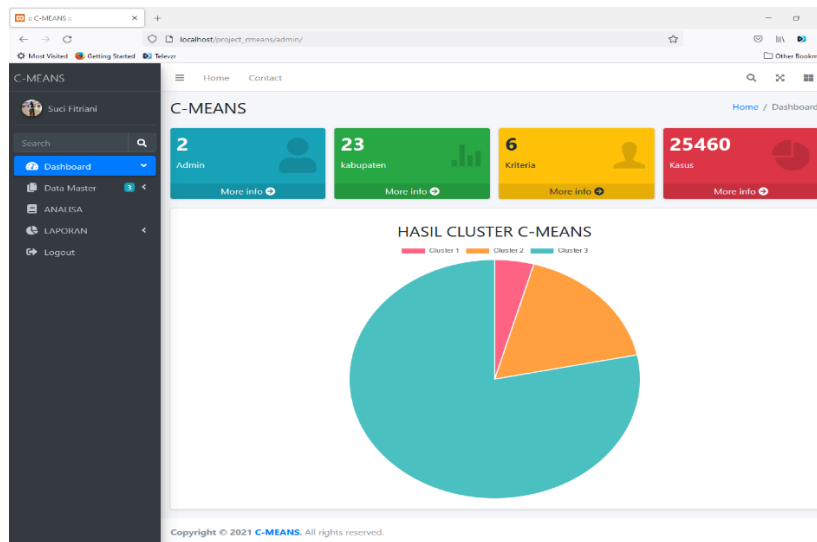

Figure 5. Clustered dashboard page

The picture above is a dashboard page display that contains data information and information on the results of clustering the spread of COVID-19 in Aceh Province using the Fuzzy C-Means Algorithm. The following is an image of the interface for the report page on the results of the clustering analysis with the degree of membership of each data. This page contains the results of the clustering of all datasets processed using the Fuzzy C-Means Algorithm in Figure 6.

Figure 6. Clusterization dataset analysis report page

The picture above is a display of the interface image of the clustering analysis report page with the degree of membership of each data.

D. CONCLUSION AND SUGGESTIONS

The system built is capable of clustering the spread of COVID-19 by producing 3 clusters, namely the cluster with the highest, medium and low COVID-19 levels. The dataset used is 23 districts/cities in Aceh Province using six variables, namely: number of confirmed, number of patients in treatment, number of recovered patients, number of deaths, suspected and probable. Based on the results of this study, the Fuzzy C-Means Algorithm can be used and applied properly in clustering the spread of COVID-19 in the province of Aceh. Of the 23 districts/cities in Aceh Province that were used as datasets for the clustering process, the results of the COVID-19 distribution cluster were obtained, namely Cluster 1 as a red zone consisting of 1 district/city, namely Kota Banda Aceh. Cluster 2 as the yellow zone consists of 4 districts/cities, namely Big Aceh, Pidie, Bireun, and Lhokseumawe. Cluster 3 as a green zone consists of 18 districts/cities South Aceh, Southeast Aceh, East Aceh, Central Aceh, West Aceh, North Aceh, Simelue, Aceh Singkil, Southwest Aceh, Gayo Lues, Aceh Jaya, Nagan Raya, Aceh Tamiang, Bener Meriah, Pidie Jaya, Sabang, Langsa, and Subulussalam. Further research is recommended to use other clustering methods by adding the Geographic Information System feature for mapping which areas are included in cluster groups 1, 2, and 3.

REFERENCES

Agustini, F. (2017). Implementasi Algoritma Fuzzy C-Means Studi Kasus Penjualan Di Sushigroove Restaurant. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, *3*(1), 127–132.

Akat, M., & Karatas, K. (2020). Psychological Effects of COVID-19 Pandemic on Society and Its Reflections on Education. *Journal of Turkish Studies*, *Volume 15*(Volume 15 Issue 4), 1–13. https://doi.org/10.7827/turkishstudies.44336

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The Fuzzy C-Means Clustering Algorithm James. *Computers & Geosciences*, *10*(2–3), 1783–1784. https://doi.org/10.1109/igarss.1988.569600

Crnogorac, V., Grbic, M., Dukanovic, M., & Matic, D. (2021). Clustering of European countries and territories based on cumulative relative number of COVID 19 patients in 2020. *2021 20th International Symposium Infoteh-Jahorina, Infoteh 2021 - Proceedings*, *March*, 17–19. https://doi.org/10.1109/INFOTEH51037.2021.9400670

Darapaneni, N., J, V., Paduri, A. R., Kumar, P., S, V., Thangeda, K. C., & Thakur, A. (2021). Machine learning approach for clustering of countries to identify the best strategies to combat Covid-19. *2021 IEEE International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*. https://doi.org/10.1109/IEMTRONICS52119.2021.9422621

Doroshenko, A. (2020). Analysis of the distribution of COVID-19 in Italy using clustering algorithms. *Proceedings of the 2020 IEEE 3rd International Conference on Data Stream Mining and Processing, DSMP 2020*, 325–328. https://doi.org/10.1109/DSMP47368.2020.9204202

Farizi, S. Al, & Harmawan, B. N. (2020). Data Transparency and Information Sharing: Coronavirus Prevention Problems in Indonesia. *Jurnal Administrasi Kesehatan Indonesia*, *8*(1), 35–50. https://doi.org/10.20473/jaki.v8i2.2020.35-50

Fernandes, N. (2020). Economic effects of coronavirus outbreak (COVID-19) on the world economy. *SSRN Electronic Journal*, *2*(1), 0–29.

König, J., Jäger-Biela, D. J., & Glutsch, N. (2020). Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany. *European Journal of Teacher Education*, *43*(4), 608–622. https://doi.org/10.1080/02619768.2020.1809650

Kurniawan, R., Abdullah, S. N. H. S., Lestari, F., Nazri, M. Z. A., Mujahidin, A., & Adnan, N. (2021).

Clustering and Correlation Methods for Predicting Coronavirus COVID-19 Risk Analysis in Pandemic Countries. *2020 8th International Conference on Cyber and IT Service Management, CITSM 2020*. https://doi.org/10.1109/CITSM50537.2020.9268920

Muslimatin, B. (2011). *Perbandingan Metode K-Means dan Metode Fuzzy C-Means (FCM) UNTUK Clustering Data (Studi Kasus pada Data Saham Harian PT. Astra, Tbk.)*.

Naik, N., Jenkins, P., Savage, N., & Yang, L. (2018). Cyberthreat Hunting - Part 2: Tracking Ransomware Threat Actors using Fuzzy Hashing and Fuzzy C-Means Clustering. *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–6. https://doi.org/10.1109/FUZZ-IEEE.2019.8858825

Noviyanto. (2020). Penerapan Data Mining Dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. *Paradigma-Jurnal Informatika Dan Komputer*, *22*(2), 183–188. https://doi.org/https://doi.org/10.31294/p.v21i2 Paradigma

Nurdin, Nasution, U. M. P., Aidilof, H. A., & Bustami. (2022). Implementasi Fuzzy C-Means untuk Menentukan Tingkat Kepuasan Mahasiswa dalam Pembelajaran Online Implementation of Fuzzy C-Means to Determine Student Satisfaction Levels. *Sistemasi: Jurnal Sistem Informasi*, *11*(1), 121–136.

Ozili, P. K., & Arun, T. (2020). Spillover of COVID-19: Impact on the Global Economy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3562570

Ren, L. L., Wang, Y. M., Wu, Z. Q., Xiang, Z. C., Guo, L., Xu, T., Jiang, Y. Z., Xiong, Y., Li, Y. J., Li, X. W., Li, H., Fan, G. H., Gu, X. Y., Xiao, Y., Gao, H., Xu, J. Y., Yang, F., Wang, X. M., Wu, C., … Wang, J. W. (2020). Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chinese Medical Journal*, *133*(9), 1015–1024. https://doi.org/10.1097/CM9.0000000000000722

Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity*, *109*, 102433. https://doi.org/10.1016/j.jaut.2020.102433

S, P., R V, N., Chowdry B, S., & K, S. P. (2021). Fuzzy clustering using Hybrid CSO-PSO search based on Community mobility during COVID 19 lockdown. *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, *August 2020*, 1515–1519. https://doi.org/10.1109/ICCMC51019.2021.9418026

Shan, Y., & Zhi, Z. (2016). A novel fuzzy SVM based on fuzzy c-means for credit scoring. *2016 2nd IEEE International Conference on Computer and Communications, ICCC 2016 - Proceedings*, 1349–1353. https://doi.org/10.1109/CompComm.2016.7924924

Solichin, A., & Khairunnisa, K. (2020). Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means. *Fountain of Informatics Journal*, *5*(2), 52–59. https://doi.org/10.21111/fij.v5i2.4905

Suni, N. S. P. (2020). Kesiapsiagaan Indonesia Menghadapi Potensi Penyebaran Corona. *INFO Singkat*, *XII*(3), 13–18. https://berkas.dpr.go.id/puslit/files/info_singkat/Info Singkat-XII-3-I-P3DI-Februari-2020-1957.pdf

Sunori, S. K., Negi, P. B., Maurya, S., Juneja, P., Rana, A., & Bhawana. (2021). K-Means Clustering of Ambient Air Quality Data of Uttarakhand, India during Lockdown Period of Covid-19 Pandemic. *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, 1254–1259. https://doi.org/10.1109/ICICT50816.2021.9358627

Susilo, A., Rumende, C. M., Pitoyo, C. W., Santoso, W. D., Yulianti, M., Herikurniawan, H., Sinto, R., Singh, G., Nainggolan, L., Nelwan, E. J., Chen, L. K., Widhani, A., Wijaya, E., Wicaksana, B., Maksum, M., Annisa, F., Jasirwan, C. O. M., & Yunihastuti, E. (2020). Coronavirus Disease 2019: Review of Current Literatures. *Jurnal Penyakit Dalam Indonesia*, *7*(1), 45. https://doi.org/10.7454/jpdi.v7i1.415

World Health Organization (WHO). (2020). Global Surveillance for human infection with novel coronavirus (2019-nCoV). *Interim Guidance*, 1–2. https://www.who.int/publications-detail/global-surveillance-for-human-infection-with-novel-coronavirus-(2019-ncov)

Zambrano-Monserrate, M. A., Ruano, M. A., & Sanchez-Alcalde, L. (2020). Indirect effects of COVID-19 on the environment. *Science of the Total Environment*, *728*, 138813. https://doi.org/10.1016/j.scitotenv.2020.138813