Data Clustering of Confirmed COVID-19 Patients Using Fuzzy C-Means

Wahyu Sri Utami [*, a,1] , Selfi Artika [b,2], Rizki Aldiansyah[c,3],
[a,c] Informatics Study Program, University of Technology Yogyakarta, Yogyakarta, Indonesia
[b] Data Science Study Program, University of Technology Yogyakarta, Yogyakarta, Indonesia

[1] wahyu.utami@uty.ac.id* , [2] selfi.artika@staff.uty.ac.id, [3]rzzkalldi@gmail.com

Abstract

The continuous mutation of COVID-19 generates new virus variants with nearly identical symptoms, such as sneezing, runny nose, sore throat, cough, fever, loss of taste and smell, and shortness of breath. Since the emergence of this virus in Indonesia, there still needs to be more research on the symptoms caused by the different COVID-19 variants, leaving the public with minimal information that may result in inappropriate early treatment, inefficient costs, and insufficient recovery time. This study aimed to classify COVID-19 patient data into two clusters based on the severity of the symptoms experienced by patients: the confirmed cluster and the unconfirmed cluster. Using Fuzzy C-Means, patient data will be clustered into two confirmed and unconfirmed clusters of covid 19 disease as the initial step in the research phase. The results of this study are anticipated to provide information on variations in the severity of symptoms among infected patients, thereby enhancing the precision of early diagnosis and treatment. The resulting clustering model is based on data collection and processing outcomes using Python and the Fuzzy C-Means algorithm, which is based on experimentation.

Keywords: Cluster, COVID-19, Fuzzy C-Means.

I. INTRODUCTION

The coronavirus disease 19 (COVID-19) is a highly transmittable and pathogenic viral infection caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which caused global pandemic that led to a dramatic loss of human life worldwide [1]. Recently at the end of 2019, Wuhan an emerging business hub of China experienced an outbreak of a novel coronavirus that killed more than eighteen hundred and infected over seventy thousand individuals within the first fifty days of the epidemic. This virus was reported to be a member of the β group of coronaviruses. The novel virus was named as 2019 novel coronavirus (2019-nCov) by the Chinese researchers. The International Committee on Taxonomy of Viruses (ICTV) named the virus as SARS-CoV-2 and the disease as COVID-19 [2], [3], [4]. In the history, SRAS-CoV (2003) infected 8098 individuals with mortality rate of 9%, across 26 contries in the world, on the other hand, novel corona virus (2019) infected 120,000 induviduals with mortality rate of 2.9%, across 109 countries, till date of this writing. It shows that the transmission rate of SARS-CoV-2 is higher than SRAS-CoV and the reason could be genetic recombination event at S protein in the RBD region of SARS-CoV-2 may have enhanced its transmission.

According to real-time data from the Global Alliance for Vaccines and Immunization (GAVI), countries worldwide, including Indonesia, continue to struggle against the threat posed by the coronavirus. On February 4, 2022, confirmed cases increased by 63,879,863, and confirmed cases of death increased by 951,179 [5]. In Indonesia, there were 4,353,370 confirmed cases and 144,320 confirmed deaths as of February 3, 2022, as accessed on Covid-19.go.id [6]. The spread of Coronavirus Disease 2019 (COVID-19), brought on by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), has had significant social and economic effects. This disease, including its diagnosis, treatment, and prevention, is still surrounded by considerable controversy. The World Health Organization (WHO) and the Research and Development Agency have revealed that there are still four contagious coronavirus variants in Indonesia, namely the Alpha, Beta, Delta, and Omicron variants that are still spreading [7]. Due to the ongoing mutation of the COVID-19 virus around the globe, it also affects Indonesia, which has a very large population. This circumstance affects the community's health and economic viability. Continuous mutations of the COVID-19 virus and nearly identical symptoms make it difficult for people to differentiate between the types of COVID-19 virus variants that infect the body. There still needs to be more research regarding the symptoms caused by the different COVID-19 variants, leaving the community with minimal information, which may result in inappropriate early treatment, inefficient costs, and inadequate recovery time. Similarly, from the perspective of healthcare professionals, ineffective and inefficient treatment and medication administration of patients result in the public still requiring information as a preventative measure for the COVID-19 pandemic spread.

FCM has been widely utilized as a decision-making instrument for various purposes. Several existing studies demonstrate that the FCM method can classify unclassified data into specific classes based on their similarity by

minimizing the value of the objective function to provide recommendations for decision-making. Based on previous research, the FCM method has been used to divide disease data from the community, particularly those who use the insurance program to pay for health services provided by the government to the local community (Jamkesda), into two categories: infection and degenerative [8]. FCM was able to divide the COVID-19 spread in Tuban Regency into three clusters, namely the low, medium, and high clusters in 16 sub-districts [9]. As a result, the FCM method is used in this study to divide patient data into two clusters based on patient symptoms. Furthermore, this findings are expected to provide information on the symptoms of each variant to the general public, health workers, or the government about the diseases most commonly afflicting the community at this time, allowing for more effective early treatment to optimize patient recovery.

II. METHOD

The following research methodologies have been employed:

1. Observation
   The research team sought as much information as possible from various sources about the symptoms of the COVID-19 disease that most people in Indonesia experience during this activity. The team also interviewed experts, including dr. Novita Eva Sawitri, Sp. P, M. Kes, a pulmonary specialist. The research team gathered information from the general public and experts about the symptoms most commonly experienced by COVID-19 patients. Based on this observation, the team discovered that there were symptoms nearly identical to those experienced by patients, such as sneezing, runny nose, sore throat, coughing, fever, loss of taste and smell, and shortness of breath. Aside from that, this activity seeks to identify societal issues that arose during the COVID-19 pandemic.
2. Data Collection -1
   The research team conducted phase-1 data collection activities using interviews and questionnaires based on the results of observation activities and having obtained some of the information needed as the primary variable, namely the patient's symptoms that appeared and supporting information obtained from pulmonary specialists. The questionnaire link is available at https://bit.ly/Penelitian-Covid-19, which has been distributed and has yielded the desired data set. The data collected includes the patient's age, history, symptoms experienced during the COVID-19 attack, and any congenital diseases, as well as information on vaccines and the type of COVID-19 suffered if a SWAB test was performed. Because the research team requires as much data as possible, this activity is ongoing in this study. Aside from that, the researcher used book references and journal articles as sources, and the chairperson, members, and students were all involved in this activity. Table 1 depicts the results of the data set collection in phase 1.

TABLE 1. Example of Data Set for Phase-1

Timestamp	Apakah saudara/i bersedia untuk mengisi skala penelitian ini dengan sukarela dan tanpa paksaan?	Nama	Usia	No. Telepon	Jenis Kelamin	1. Apakah anda pernah terkonfirmasi penyakit covid-19	a. Demam	b. Batuk	c. Sakit Tenggorakan	d. Sesak Nafas	e. Hidung Berair
27/07/2022 20:20:22	Ya	Fadhila Puri Damayanti	21	085649756429	Perempuan	Ragu	3	4	4	2	4
27/07/2022 20:26:37	Ya	Evan Okmawanto	21	081215282110	Laki-Laki	Ragu	7	2	8	1	5
27/07/2022 21:25:01	Ya	Oktavierik tegar nugraha putra	24	081327795624	Laki-Laki	Belum	1	2	3	2	3
28/07/2022 14:32:40	Ya	Priasmara Putra Marindrha	37	081328757890	Laki-Laki	Belum	2	1	1	1	1
31/07/2022 7:19:58	Ya	Dhanu	18		Laki-Laki	Pernah	6	6	6	7	7
31/07/2022 7:42:55	Ya	Irsan	45	082115631000	Laki-Laki	Pernah	10	5	5	6	1
31/07/2022 7:43:47	Ya	Selfi Artika	29	082306302117	Perempuan	Pernah	9	9	9	5	5

Timestamp	Apakah saudara/i bersedia untuk mengisi skala penelitian ini dengan sukarela dan tanpa paksaan?	Nama	Usia	No. Telepon	Jenis Kelamin	1. Apakah anda pernah terkonfirmasi penyakit covid-19	a. Demam	b. Batuk	c. Sakit Tenggorakan	d. Sesak Nafas	e. Hidung Berair
31/07/2022 7:43:55	Ya	Rina fujiastuti	45	081220837300	Perempuan	Belum	2	2	1	1	1
31/07/2022 7:44:05	Ya	Hasna	21	0895378161730	Perempuan	Ragu	9	10	9	6	6
31/07/2022 7:44:06	Ya	Adella Nawang Andyka Octavian	20	085155356023	Perempuan	Ragu	7	7	7	3	3
31/07/2022 7:44:12	Ya	Eulis popon	48	081394039340	Perempuan	Pernah	1	1	1	1	1
31/07/2022 7:46:06	Ya	dedeh	50	081214672149	Perempuan	Ragu	1	1	1	1	1
31/07/2022 7:47:31	Ya	Meyy	19	083820158488	Perempuan	Belum	1	1	1	1	1
31/07/2022 7:48:15	Ya	asha farda lafifa	19	08987332841	Perempuan	Ragu	3	3	3	3	3
31/07/2022 7:49:47	Ya	Rina wijaya subur	36	083822444226	Perempuan	Belum	1	3	2	2	1
31/07/2022 7:58:34	Ya	Rizkal	22	081215214609	Laki-Laki	Ragu	8	5	5	4	7
31/07/2022 8:09:11	Ya	FERRY HARVEY DEVIS	36 tahun	085216378917	Laki-Laki	Pernah	2	2	5	1	2
31/07/2022 8:11:50	Ya	Ani fitriani	46 tahun	087818625171	Perempuan	Pernah	3	3	3	6	2
31/07/2022 8:14:04	Ya	Dela	21	085171626056	Perempuan	Ragu	5	8	8	3	2
31/07/2022 8:18:24	Ya	Cinta verawati	38	087825677588	Perempuan	Belum	3	4	4	1	4
31/07/2022 8:20:06	Ya	Alesa Virbi Haira	15	0895-1908-5390	Perempuan	Belum	2	2	1	1	1

3. Implementation of the FCM-1 Method

Data grouping, mainly disease data grouping, is widely used using artificial intelligence methods. The fuzzy clustering method is commonly used to classify data based on similarity. Fuzzy Clustering Means is one of the most frequently used fuzzy clustering methods (FCM). FCM seeks to minimize variations within a cluster while maximizing variations between clusters to reduce the objective function set in the existing clustering process in general [10]. In addition, FCM is used in the transportation sector to classify heavy traffic in urban countries [11]. FCM can be used in digital and image processing to detect brain tumor areas [12, 13] and fetal head segmentation in ultrasound images [14]. The Fuzzy clustering method is an integral part of pattern recognition, which plays a crucial role in determining optimal clusters in a vector space based on the Euclidean standard form for distances between vectors. Fuzzy Clustering Means (FCM) is a data clustering technique in which the degree of membership determines the existence of each data point in a cluster [15]. The first fundamental concept of FCM is to identify the cluster's center, which represents the average location of each cluster. Under the initial conditions, the cluster's center is still inaccurate. Each data point has a degree of membership in each cluster. By repeatedly repairing the cluster center and the degree of membership of each data point, it is evident that the cluster center will move closer to its proper location. This iteration is based on minimizing the objective function, which describes the distance between a given data point and the cluster center, weighted by the data point's degree of membership. FCM produces a row of cluster centers and several degrees of membership for each data point rather than a fuzzy inference system. This data can create a fuzzy inference system [16]. The FCM method is depicted in Figure 1 with a flowchart.

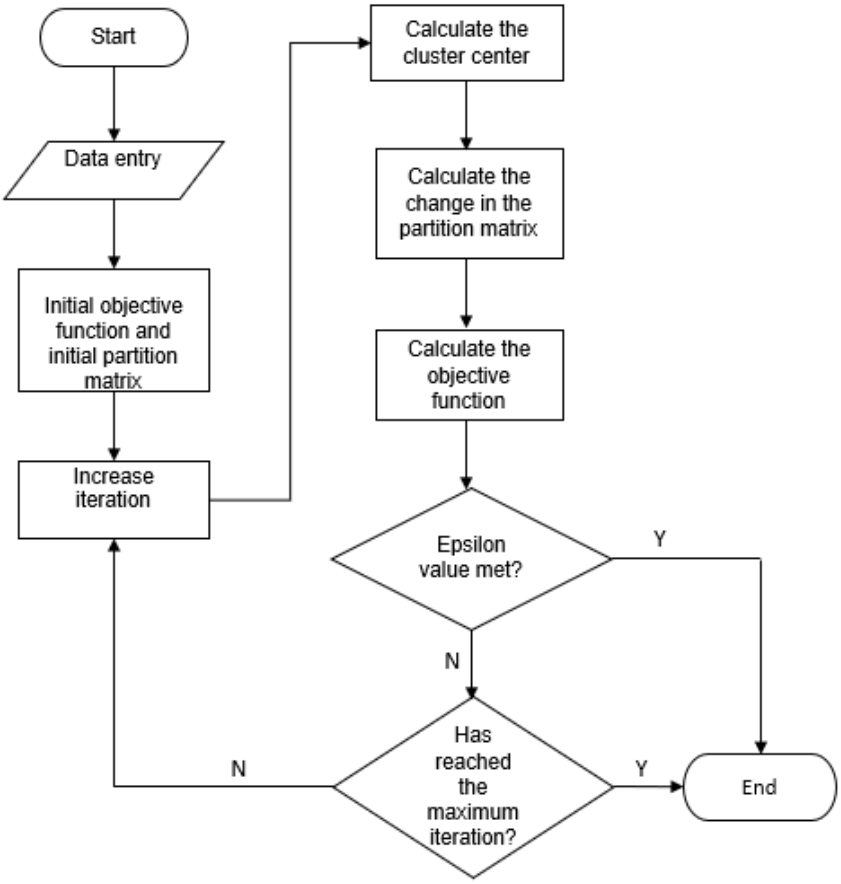

FIGURE 1. Flowchart of Fuzzy C-Means

The obtained data set will be normalized into Fuzzy numbers by increasing the membership function of the linear representation rarely, occasionally, and frequently so that all data will become Fuzzy numbers. The researchers discovered two clusters based on the severity of symptoms, namely sneezing, runny nose, sore throat, cough, fever, loss of taste and smell, and shortness of breath, based on FCM calculations with the data set used. These findings and the amount of data confirmed by COVID-19 are still being developed.

3.1. Fuzzy C-Means Clustering Algorithm

This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated according to the formula 1 and 2.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}(d_{ij}/d_{ik})^{\frac{2}{m-1}}} \quad \text{…….,,,.(1)}$$

$$v_j = \frac{\left(\sum_{i=1}^{n}(\mu_{ij})^m x_i\right)}{\left(\sum_{i=1}^{n}(\mu_{ij})^m\right)}, \forall j = 1,2,\ldots,c \quad \text{……..(2)}$$

Where:

n = The number of data points
m = The fuzziness index $m \in [1..\infty]$
μ_{ij} = The membership of i^{th} data to j^{th} cluster center
v_j = The j^{th} cluster center.
c = The number of cluster center.
d_{ij} = The Euclidean distance between i^{th} data and j^{th} cluster center.

Main objective of fuzzy c-means algorithm is to minimize Formula 3:

$$J(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{c}(\mu_{ij})^{m}\|x_i - v_j\|^2 \quad \ldots\ldots (3)$$

where,
$||x_i - v_j||$ is the Euclidean distance between i^{th} data and j^{th} cluster center.

3.2. Algorithmic steps for Fuzzy c-means clustering.

Let $X = \{x_1, x_2, x_3 ..., x_n\}$ be the set of data points and $V = \{v_1, v_2, v_3 ..., v_c\}$ be the set of centers.

1) Randomly select *'c'* cluster centers.
2) Calculate the fuzzy membership $'\mu_{ij}'$ using formula 4:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}(d_{ij}/d_{ik})^{\frac{2}{m-1}}} \quad \ldots\ldots (4)$$

3) Compute the fuzzy centers v_j using Formula 5:

$$v_j = \frac{\left(\sum_{i=1}^{n}(\mu_{ij})^{m}x_i\right)}{\left(\sum_{i=1}^{n}(\mu_{ij})^{m}\right)}, \forall j = 1,2,\ldots,c \quad \ldots\ldots (5)$$

4) Repeat step 2) and 3) until the minimum *'J'* value is achieved or $||U^{(k+1)} - U^{(k)}|| < \beta$.

where,
k = The iteration step.
β = The termination criterion between [0, 1].
U = $(\mu_{ij})_{n*c}$ is the fuzzy membership matrix.
J = The objective function. [17],[18].

III. RESULTS AND DISCUSSION

Using the fuzzy C-Means method on a data set of symptomatic patients yields two clusters: the cluster of patients not confirmed with COVID-19 and those suspected of having COVID-1

TABLE 2. Cluster Table

Cluster 0: Unconfirmed for COVID-19	Cluster 1 : Confirmed for COVID-19
1. The average from this cluster was not confirmed for COVID-19; only a few expressed doubts and had been confirmed for COVID-19. 2. Because most of this cluster is unconfirmed for COVID-19, the symptoms they experience are nearly below level 4. 3. On average, these cluster members are 18 to 43 years old. 4. There are fewer members of Cluster than Cluster 1.	1. The average from this cluster was confirmed for COVID-19, and only a few stated that it was not confirmed. 2. Almost all of the symptoms in this cluster were above level 6, but the level of runny nose and shortness of breath varied little. 3. The members of this cluster are 21 to 40 years old on average. 4. This cluster has more members than cluster 0. 5. Many of these clusters are afflicted by omicrons and deltas, and many do not know what type of COVID-19 they possess.

The specifics of the clustering conclusions in Table 2 can be explained as follows based on the results of each influential variable:

a) The number of cluster members

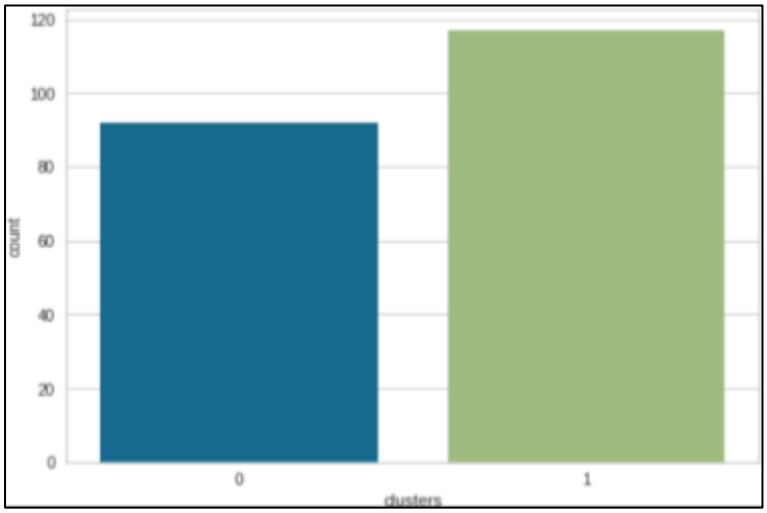

FIGURE 2. Barchart of Cluster Membership

As shown in the bar chart in Figure 2, cluster 1 has over 110 members, while cluster 0 has approximately 90 members.

TABLE 3. An example of Random Clustering Results Based on the Severity of Patient Symptoms.

No	Age	Gender	Fever	Cough	Sore throat	Shortness of breath	Runny nose	Congenital disorder	Cluster
191	18	1	5	3	5	1	2	0	0
196	22	0	3	2	1	1	1	0	0
198	22	0	2	4	6	1	5	0	0
199	22	0	2	3	7	1	3	0	0
200	22	0	1	4	9	1	3	0	0
7	18	1	6	6	6	7	7	0	1
8	45	1	10	5	5	6	1	1	1
9	29	0	9	9	9	5	5	0	1
11	21	0	9	10	9	6	6	0	1
12	20	0	7	7	7	3	3	0	1
18	22	1	8	5	5	4	7	0	1
21	21	0	5	8	8	3	2	0	1
25	30	0	7	7	7	1	1	0	1

It can be seen in Table 3, which was drawn at random from the entire dataset, that patients in cluster 0 are not confirmed to have COVID-19 disease, whereas patients in cluster 1 are confirmed with COVID-19 disease. The table above also shows that patients with a larger severity scale appear to be in cluster 1, compared to milder in cluster 0.

b) Age Composition

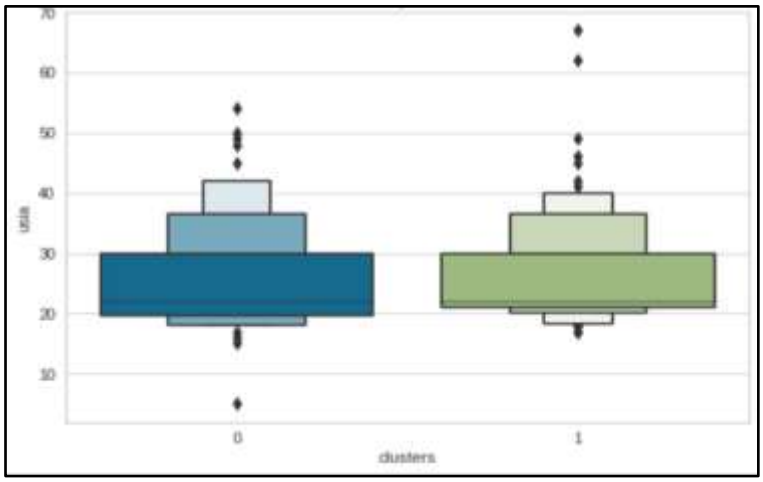

FIGURE 3. Boxplot of Age Composition by Cluster

Figure 3 depicts the age composition of cluster 0 members, which ranges from 15 to 50 years. The age distribution of Cluster 0 members is then more or less spread out from 18 to approximately 43. It is nearly identical to cluster 0, but cluster 1 members are more spread out at the ages of 21 and up and up to 40.

c) Fever Symptoms

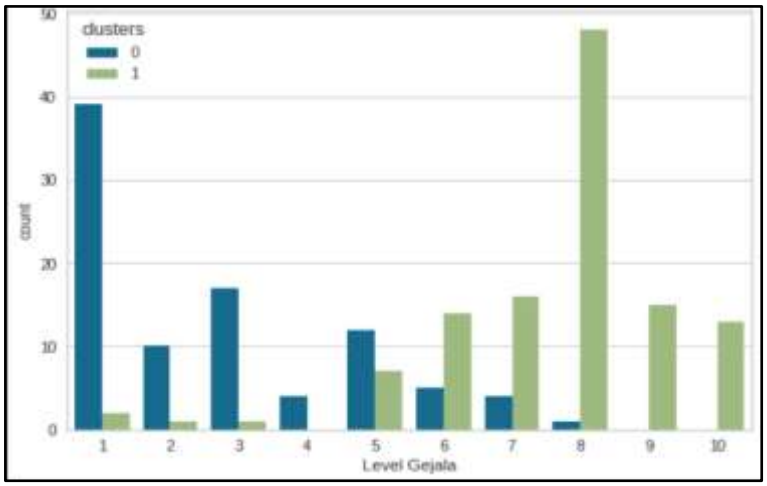

FIGURE 4. Boxplot of Age Composition by Cluster

Figure 4 shows that cluster 0 has no symptoms; on average, the symptoms they experience are not that severe. In contrast, for cluster 1, the symptoms of fever they feel are average at a level above 5, and many members feel these symptoms at level 8.

d) Cough Symptoms

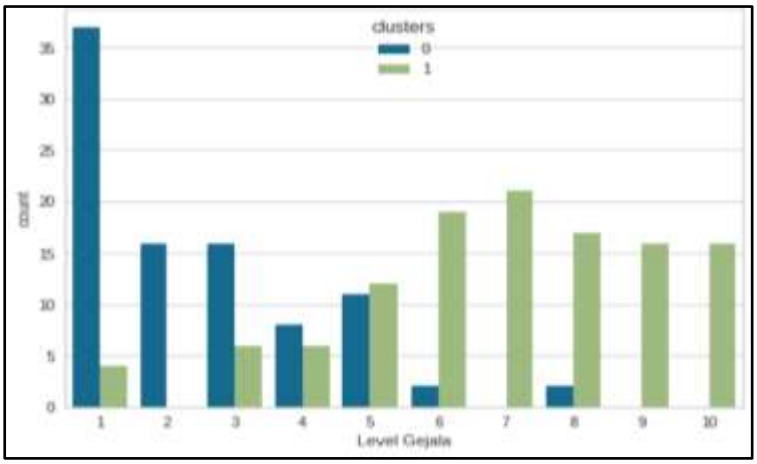

FIGURE 5. Bar chart of Cough Symptoms

Almost identical to the symptoms of fever, the bar chart results in Figure 5 indicate that cluster 1 members have an average fever of more than level 5. In contrast, cluster 0 members experience a cough that is barely noticeable on average.

e) Sore Throat Symptoms

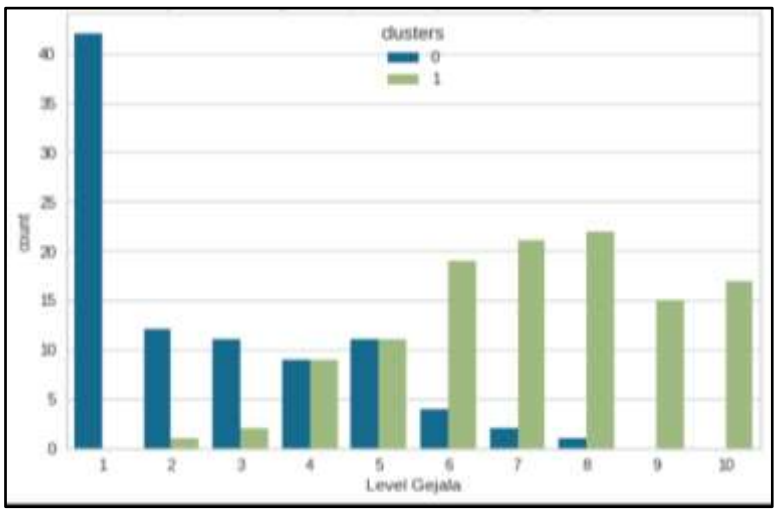

FIGURE 6. Bar chart of Sore Throat Symptoms

Figure 6 demonstrates that cluster 1 members experience sore throat symptoms at an average level greater than 5, whereas cluster 0 members do not experience sore throat symptoms on average.

f) Symptoms of Shortness of Breath

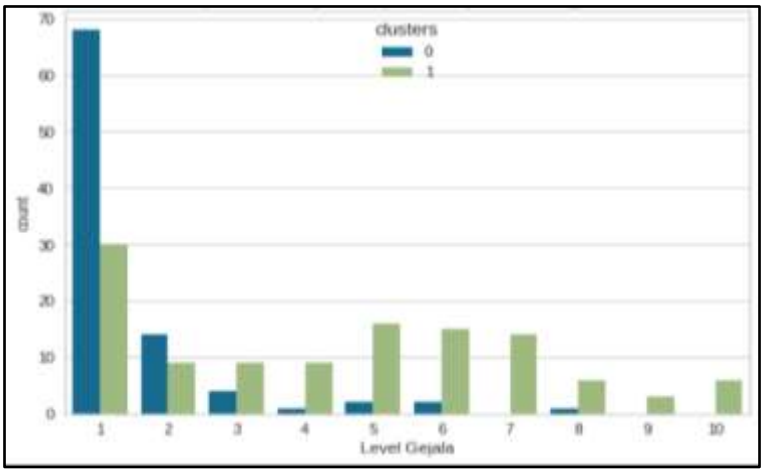

FIGURE 7. Bar chart for shortness of breath symptoms

In contrast to the previous results, Figure 7 depicts the symptoms felt by members of Cluster 1, which is more spread out at all levels, and almost all members of Cluster 1 do not feel short of breath.

g) Runny Nose Symptoms

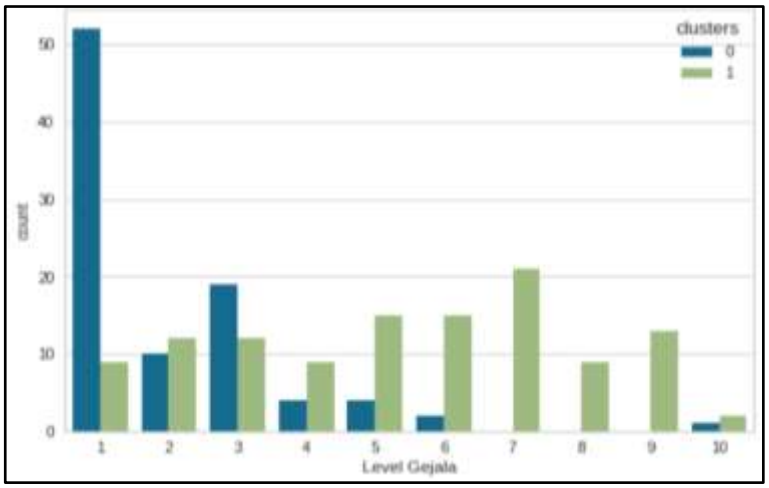

FIGURE 8. Bar chart of Runny Nose Symptoms

The bar chart results for runny nose symptoms in Figure 8 show that cluster 1 members have a wide range of symptoms and, on average, have symptoms of level 7, but many members have no. It differs from Cluster 0, where almost all members do not have runny nose symptoms.

h) Types of COVID-19

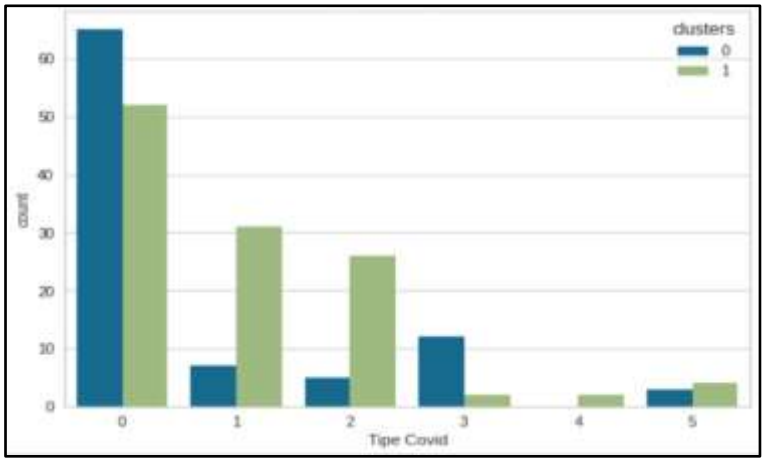

FIGURE 9. Type of Covid-19 bar chart

Description of Figure 9:
Covid type 0: unknown type
Covid Type 1: Omicron
Covid Type 2: Delta
Covid Type 3: Influenza
Covid Type 4: Beta
Covid Type 5: Alpha

From the bar chart results in Figure 9, it can be concluded that almost all of the members of cluster 0 do not know the type of COVID; this is because the data from cluster 0 is not, on average, confirmed by COVID. Meanwhile, in cluster 1, most COVID types were affected by omicrons and deltas.

IV. CONCLUSION

The research team discovered that the clustering of the data set was in the conclusion of two clusters, namely the confirmed COVID-19 cluster and the unconfirmed COVID-19 cluster, in the results of the data set that was processed using the FCM method. These two clusters are the best obtained because the evaluation results look at the scatter plot and see the PC (Partition Coefficient). The PC result is 0.383, which is the best of the other clusters.

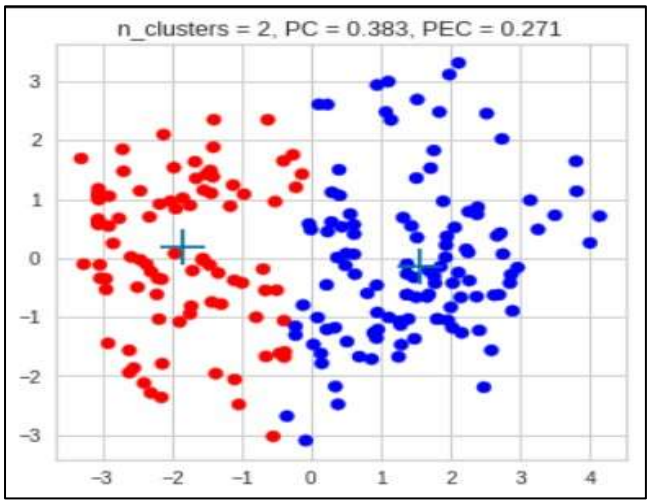

FIGURE 10. Scatter Plot of Partition Coefficient Results for 2 Clusters

By establishing these two clusters, the results obtained from the data of confirmed and unconfirmed COVID-19 patients are 80% accurate.

The results obtained thus far in the study have demonstrated that by providing information on the symptoms experienced based on severity level. The FCM method will determine whether or not the patient has COVID-19 disease. Further research with confirmed patient data will be conducted, and more specific clusters will be produced using FCM. Furthermore, the clustering results will be matched to the COVID-19 cluster, which a team of researchers and experts will confirm.

ACKNOWLEDGMENT

We would like to thank dr. Novita Eva Sawitri, Sp. P, M. Kes, a pulmonary specialist, for interviewed to do research data. We apriciate all of participant who fill the questionnaire link is available at https://bit.ly/Penelitian-Covid-19 to do collect data.

REFERENCES

[1] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, and R. Siddique, "COVID-19 infection: Emergence, transmission, and characteristics, of human coronaviruses", *Journal of Advanced Research*, vol. 24, pp. 91-98, 2020.

[2] J. Cui, F. Li, and Z. L. Shi, "Origin and evolution of pathogenic coronaviruses", *Nat Rev Microbiol*, 17 (3) pp. 181-192, 2019.

[3] C. C. Lai, T. P. Shih, W. C. Ko, H. J. Tang, and P. R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): The Epidemic and The Challenges", *Int J Antimicrob Agents*, 105924, 2020.

[4] WH. Organization, "Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance 2 March 2020", 2020.

[5] Global Alliance for Vaccines and Immunization (GAVI), Jakarta. [Online]. 2022. Available: https://www.gavi.org/covid19 [Accesed: February. 4, 2022].

[6] Covid 19.go.id, Jakarta. [Online]. 2022. Available: https://covid19.go.id/peta-sebaran [Accesed: February. 3, 2022.

[7] Badan Penelitian dan Pengembangan Kesehatan, Jakarta [Online]. 2022. Available: https://www.litbang.kemkes.go.id/peta-sebaran-sekuens-dan-varian-covid-19/ [Accesed: 2 February 2022].

[8] N. Fatihah, K. Oktafianto, "Clustering Data Penyebaran Covid-19 di Kabupaten Tuban Menggunakan Fuzzy C Means", Prosiding Seminar Peenelitian dan Pengabdian Masyarakat (SNasPPM), Vol. 6 No.1, pp. 632-639, 2021.

[9] N. Puspitasari, J. A. Widians, and Pohny, "A Clustering of Generative anda Infectious Diseases Using Fuzzy C-Means", Jurnal Ilmiah Teknologi dan Informasi, Vol 7. No.1, pp. 22-29, 2018.

[10] H. L. Sari, D. Suranti, "Perbandingan Algoritma Fuzzy C-Means (FCM) Dan Algoritma Mixture Dalam Penclusteran Data Curah Hujan Kota Bengkulu", Seminar Nasional Aplikasi Teknologi Informasi (SNATI). A7-A15, 2016.

[11] G. Zhu, J. Chen, and P. Zhang, "Fuzzy c-means clustering identification method of urban road traffic state", in Fuzzy Systems and Knowledge Discovery (FSKD), 12th International Conference on, pp. 302-307, 2015.

[12] N. Menon, R. Ramakrishnan, "Brain Tumor Segmentation in MRI images using unsupervised Artificial Bee Colony algorithm and FCM clustering", in Communications and Signal Processing (ICCSP), International Conference on, pp. 0006-0009, 2015.

[13] R. Preetha, G. Suresh, "Performance analysis of fuzzy c means algorithm in automated detection of brain tumor", in Computing and Communication Technologies (WCCCT), World Congress on, pp. 30-33. 2014.

[14] P.D.W. Ayu, "Segmentasi Kepala janin Pada Citra USG Dalam Ruang Warna RGB dengan Metode Fuzzy C- Means", Prosiding Conference on Smart Green Tecnology in Electrical and Information System CSGTEIS. pp. 209-214, 2013.

[15] I. Rosalina, E. Supriyati, and T. Khotimah, "Clustering Gender Berdasarkan Nilai Maksimum Minimum Amplitudo Suara Berbasis Fuzzy C-Means (FCM)", Prosiding SNATIF,vol. 1, pp. 419-424, 2014.

[16] A. Ahmadi, S. Hartati, "Penerapan Fuzzy C-Means dalam Sistem Pendukung Keputusan untuk PenentuanPenerima Bantuan Langsung Masyarakat (BLM) PNPM-MPd (Studi Kasus PNPM-MPd Kec. Ngadirojo Kab.Pacitan)", Berkala Ilmiah MIPA, vol. 23, no. 3, pp. 264-27, 2015.

[17] M. Singh, "Study on Clustering Data", International Journal of Engineering Research and Development, Vol. 11, No. 10, pp. 55-60, 2015.

[18] W. Cai, S. Chen., and D. Zhang, "Fast and Robust Fuzzy C-Means Clustering Algorithms Incorporating Local Information for Image Segmentation", Pattern Recognition, Vol 4, No. 3, pp. 825-838, 2007.