International Journal of Information System & Innovation Technology 1(2) (2022) 21-30

IJISIT

http://ejournal.enlightenlearner.com

Clustering the Spread of ISPA Disease Using the Fuzzy C-Means Algorithm in Aceh Utara

[1]Rozzi Kesuma Dinata, Universitas Malikussaleh, Indonesia
[2]Bustami, Universitas Malikussaleh, Indonesia
[3]Sujacka Retno, Universitas Malikussaleh, Indonesia
[4]Azrai Putra Barumun Daulay, Universitas Malikussaleh, Indonesia

Correspondence: E-mail: rozzi@unimal.ac.id

Article Info

Article history:

Received December 18, 2022
Revised December 19, 2022
Accepted December 20, 2022

Keywords:

Clustering
Fuzzy C-Means
ISPA
Aceh Utara

ABSTRACT

ISPA disease is one of the most common types of disease suffered in Aceh. The method used in this research is fuzzy c means algorithm to cluster the spread of ISPA disease areas in Aceh Utara. The data used is the data on patients with ISPA in 2021 obtained from Cut Meutia General Hospital and Arun Hospital. The attributes used consist of the patient's age, patient's address, patient's gender and the type of ISPA the patient suffered. In this research only focused on 3 types of ISPA, namely Pneumonia, Bronchitis and ISPA. The existing data will then be grouped into 3 clusters, namely low cluster, medium cluster and high cluster. The results of clustering the spread of ISPA in Aceh Utara show that the highest case is in Dewantara District, the medium cluster are Muara Batu, Nisam, Nisam Antara, Sawang, and Tanah Jambo Aye Districts, and the low cluster is Baktiya, Baktiya Barat, Banda Baro, Cot Girek, Geureudong Pase, Kuta Makmur, Langkahan, Lapang, Lhoksukon, Matang Kuli, Meurah Mulia, Nibong, Paya Bakong, Pirak Timu, Samudera, Seunudon, Simpang Keramat, Syamtalira Aron, Syamtalira Bayu, Tanah Luas and Tanah Pasir Districts. With the results of this clustering, it is expected that the provision of drug stocks for ISPA will be prioritized in cluster areas with the highest spread rates.

1. INTRODUCTION

The clustering process is a process of grouping data into clusters based on certain parameters so that objects in a cluster have a high degree of similarity to another in the same cluster and to other objects in different clusters that are very dissimilar. Clustering is one of the functional processes in data mining processing [1]. Data mining is a process to analyse, extract and identify information from a data set [2]. Fuzzy C Means is a data clustering technique in which the existence of each data in a cluster is determined by the membership value [3]. Jim Bezdek was the person who introduced this technique in 1981 [4]. Fuzzy C Means has the advantage of being able to cluster more than one variable simultaneously [5].

Several researches regarding clustering, especially in ISPA disease, as researched by Isy Karima Fauzia et al., in the case of the application of K-Means Clustering in ISPA disease in Kabupaten Karawang with the results of 30 members in low cluster, 9 members in medium cluster, and 11 members in high cluster in 2017. In 2018, there were 33 in low cluster, 5 in medium cluster and 12 in high cluster. In 2019, there were 20 in low cluster, 25 in medium cluster and 5 in high cluster [6]. Then in a research conducted by Sari et al., in the case of clustering the spread of ISPA disease in the Sekayu City using the K-Means Clustering algorithm with the results of cluster 2 as the highest cluster of ISPA disease spread in the Kayuara [7]. Furthermore, in research conducted by Purba et al., in the case of the application of the K-Means Clustering Algorithm to the spread of ISPA in Riau with the results of cluster 1 providing high recommendations totaling 10 districts, cluster 2 providing low recommendations totaling 2 districts [8]. Research conducted by Ni'mah, in the case of Geographic Information System Visualization Clustering ISPA disease in Kaliwungu with the results of cluster 1 of 752, cluster 2 of 494, and cluster 3 of 253 [9].

ISPA is an infection of the respiratory tract, either the upper or lower respiratory tract, and can cause a wide spectrum of illness from mild infection to severe and fatal disease, which is influenced by the causative pathogen, environmental factors, and host factors [10]. In the province of Aceh itself, ISPA is one of the ten diseases that most suffer from Aceh. In 2020, the Aceh Provincial Government announced that of the five million people living in Aceh, 25,264 of them have symptoms of ISPA.

Therefore, in assisting the effective management of ISPA disease, a method is necessary to find the distribution areas of ISPA disease in Aceh, especially in Aceh Utara. The goal is that Aceh Utara government can give special attention, effective handling and can assist in making policies on ISPA disease in Aceh Utara.

2. METHODS

2.1 Fuzzy C-Means

Fuzzy C Means Clustering is a data clustering technique in which the existence of each data in a cluster is determined by the degree of membership [11]. The steps for carrying out clustering with the Fuzzy C Means Clustering method are as follows [12]:

1. Input the data to be clustered by the form of a matrix of size n × m (n = number of data samples m = attribute of each data) [13].
2. Define:
   a. Number of cluster (c)
   b. Rank (w)
   c. Maximum iterations
   d. The minimum error expected (ε)
   e. Initial objective function ($p_0 = 0$)
   f. Initial iteration (t = 1)
3. Generate a random number (μ_{ik}) as an element of the initial partition matrix U.
   Calculate the sum of each column:
   $$\varrho_i = \sum_{k=1}^{c} \mu_{ik} \quad (1)$$
4. Calculate centroid of-k: V_{kj}.
   $$V_{kj} = \frac{\sum_{i=1}^{n}\left((\mu_{ik})^w \times X_{ij}\right)}{\sum_{i=1}^{n}(\mu_{ik})^w} \quad (2)$$
5. Calculate the objective function of iteration-t,
   $$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right](\mu_{ik})^w\right) \quad (3)$$
6. Calculate the changes of partition matrix[14]:
   $$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{C}\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (4)$$
7. Check the stop condition:

a. If $p_t - p_{t-1} < \varepsilon$ or t > max iteration, then stop
b. Else if, t = t + 1, repeat step- 4 (calculating V_{kj})

2.2 Research Flowchart

The research flow in the form of a flowchart can be seen in Figure 1.

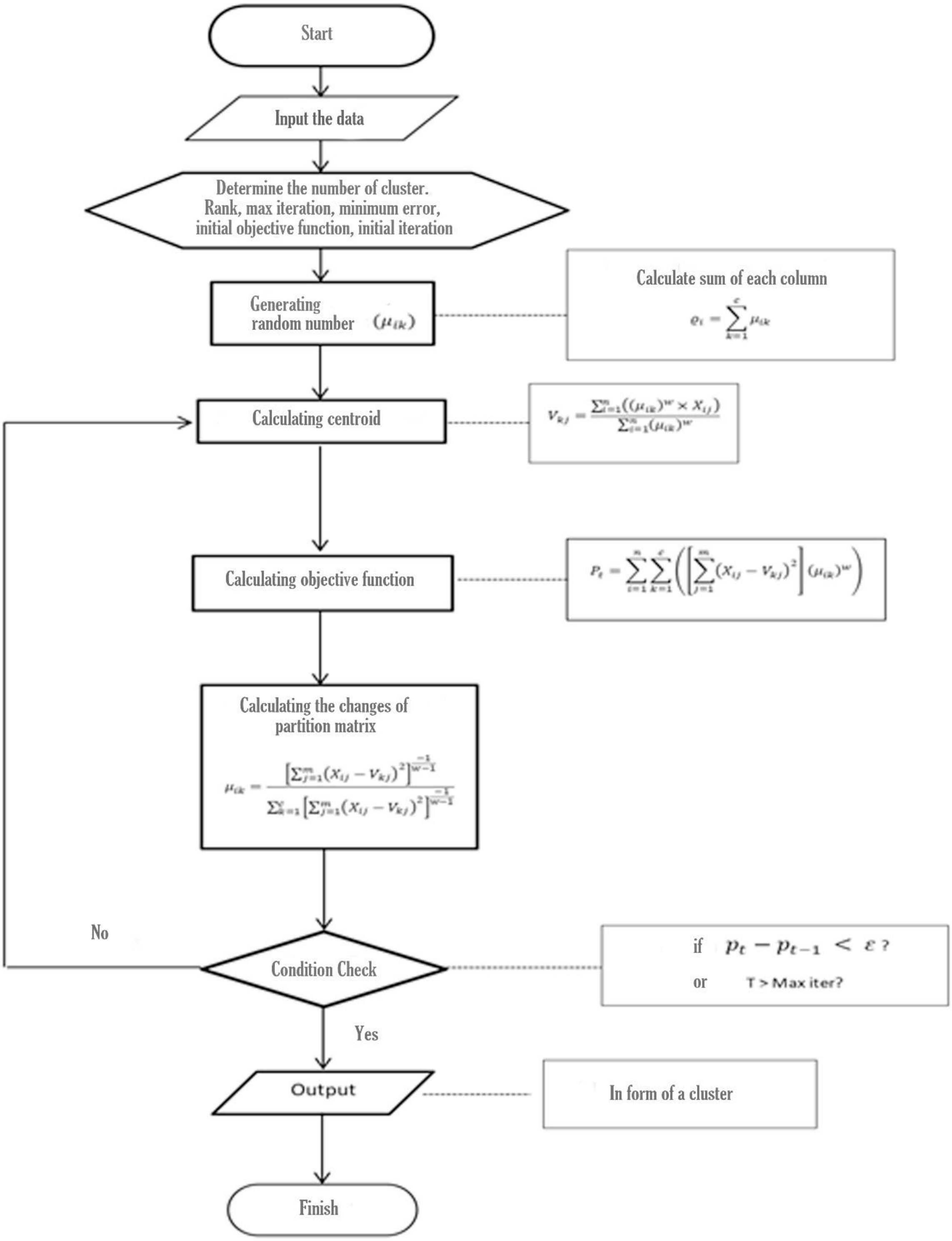

Figure 1. Research Flowchart

3. RESULTS AND DISCUSSION

3.1 ISPA Dataset

The dataset used in this research is the data on patients with ISPA in 2021. The data was obtained from 2 hospitals, namely Cut Meutia General Hospital and Arun Hospital. The data consisted of the patient's gender, patient's age, patient's address, and the type of ISPA the patient had. The following is data on patients with ISPA in Aceh Utara.

Table 1. Data on Patients with ISPA

No	Gender	Age	Address	Type of Ispa
1	Woman	19	Baktiya	Ispa
2	Woman	48	Baktiya	Pneumonia
3	Man	56	Baktiya	Pneumonia
4	Man	75	Baktiya	Pneumonia
5	Woman	45	Baktiya	Pneumonia
6	Woman	57	Baktiya	Pneumonia
7	Man	40	Baktiya	Pneumonia
8	Man	55	Baktiya	Bronchitis
9	Man	48	Baktiya	Bronchitis
10	Woman	76	Baktiya	Bronchitis
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
638	Man	26	Tanah Luas	Bronchitis
639	Man	76	Tanah luas	Pneumonia
640	Man	55	Tanah luas	Pneumonia
641	Man	67	Tanah luas	Pneumonia
642	Man	62	Tanah Pasir	Pneumonia
643	Woman	46	Tanah Pasir	Bronchitis
644	Man	41	Tanah Pasir	Pneumonia
645	Woman	36	Tanah Pasir	Pneumonia
646	Woman	78	Tanah Pasir	Pneumonia

3.2 Fuzzy C-eans algorithm analysis

The manual calculation is carried out to test the accuracy of the calculation on the system to be built later on. This aims to determine whether the system calculations are the same as the calculations performed outside the system, manually. The data used in this manual calculation is ISPA type data which consists of Pneumonia, Bronchitis and ISPA. The 646 data will then be converted into a tabular form containing the number of cases in each sub-district in Aceh Utara. The following data table is based on the number of cases in 27 sub-districts in Aceh Utara.

Table 2. Number of Cases in Each District

No	District	Pneumonia	Bronchitis	Ispa
1	Baktiya	10	5	1
2	Baktiya Barat	8	2	0
3	Banda Baro	12	8	1
4	Cot Girek	4	1	0
5	Dewantara	125	79	22
6	Geureudong Pase	2	1	0
7	Kuta Makmur	11	5	0
8	Langkahan	9	4	0
9	Lapang	2	2	0
10	Lhoksukon	12	2	0
11	Matang Kuli	4	0	0
12	Meurah Mulia	9	1	0
13	Muara Batu	38	9	4
14	Nibong	4	0	0
15	Nisam	21	14	2
16	Nisam Antara	17	15	0
17	Paya Bakong	6	2	0
18	Pirak Timu	2	0	0
19	Samudera	3	2	0
20	Sawang	50	24	6

21	Seunudon	19	5	1
22	Simpang Keramat	2	0	0
23	Syamtalira Aron	6	0	0
24	Syamtalira Bayu	5	0	0
25	Tanah Jambo Aye	25	12	0
26	Tanah Luas	4	1	0
27	Tanah Pasir	4	1	0

In the clustering process, Pneumonia is assigned as X_i1, Bronchitis is assigned as X_i2, and ISPA is assigned as X_i3. The first step in performing fuzzy C-means clustering is to determine in advance the desired number of clusters (the weighting exponent/w), where the value of w> 1. Next, determine the smallest expected error criterion value (ε), this value is the iteration stopping criteria if the objective function value is smaller than this ε value then the iteration is finished and the clustering results are obtained. Then determine the maximum iteration, this value becomes the limit for stopping iterations if it reaches this limit, but the value of the objective function is still greater than the predetermined ε value. Then determine the initial objective function (Po) and the initial iteration (t).

The steps for performing manual fuzzy c-means calculations are as follows:

1. Determine the initial value:
   a. Number of Cluster (C) = 3
   b. Weighting (*w*) = 2
   c. Minimum error expected (Ɛ) = 0.001
   d. Maximum Iteration (MaxIter) = 100
   e. Initial objective function (Po) = 0
   f. Initial Iteration (t) = 1
2. Generate random number of μ_{ik}, *i*=1,2,3 ..., C; *k*=1,2,3 ...*n*; as elements of the initial partition matrix U with degrees of membership.

Table 3. Random Number

μ_{1k}	μ_{2k}	μ_{3k}	Total
0.18	0.61	0.21	1
0.74	0.12	0.14	1
0.02	0.82	0.16	1
0.56	0.02	0.42	1
0.23	0.15	0.62	1
0.26	0.52	0.22	1
0.50	0.31	0.19	1
0.75	0.10	0.15	1
0.19	0.39	0.42	1
0.46	0.34	0.2	1
0.7	0.16	0.14	1
0.21	0.57	0.22	1
0.11	0.72	0.17	1
0.43	0.41	0.16	1
0.91	0.03	0.06	1
0.72	0.2	0.08	1
0.23	0.19	0.58	1
0.81	0.09	0.1	1
0.5	0.25	0.25	1
0.07	0.1	0.83	1
0.67	0.17	0.16	1
0.46	0.06	0.48	1
0.69	0.21	0.1	1
0.4	0.33	0.27	1
0.12	0.68	0.2	1
0.35	0.35	0.3	1
0.82	0.1	0.08	1

Generating this random number is only in iteration 1, then this random number will be replaced by a change in the partition matrix. The random number must be 1 when the three numbers are summarized. Calculating the centroid of-k :V_{kj}, with k= 1, 2, …, C , j=1, 2, …., m.

Table 4. Centroid Calculation 1 in Iteration 1

DEEGRES OF MEMBERSHIP IN CLUSTER 1	DATASET X1	X2	X3	$\mu_{1k}{}^{\wedge 2}$	$(\mu_{1k})^{\wedge 2X1}$	$(\mu_{1k})^{\wedge 2X2}$	$(\mu_{1k})^{\wedge 2X3}$
0,18	10	5	1	0,03	0,32	0,16	0,03
0,74	8	2	0	0,55	4,38	1,10	0,00
0,02	12	8	1	0,00	0,00	0,00	0,00
0,56	4	1	0	0,31	1,25	0,31	0,00
0,23	125	79	22	0,05	6,61	4,18	1,16
0,26	2	1	0	0,07	0,14	0,07	0,00
0,50	11	5	0	0,25	2,75	1,25	0,00
0,75	9	4	0	0,56	5,06	2,25	0,00
0,19	2	2	0	0,04	0,07	0,07	0,00
0,46	12	2	0	0,21	2,54	0,42	0,00
0,7	4	0	0	0,49	1,96	0,00	0,00
0,21	9	1	0	0,04	0,40	0,04	0,00
0,11	38	9	4	0,01	0,46	0,11	0,05
0,43	4	0	0	0,18	0,74	0,00	0,00
0,91	21	14	2	0,83	17,39	11,59	1,66
0,72	17	15	0	0,52	8,81	7,78	0,00
0,23	6	2	0	0,05	0,32	0,11	0,00
0,81	2	0	0	0,66	1,31	0,00	0,00
0,5	3	2	0	0,25	0,75	0,50	0,00
0,07	50	24	6	0,00	0,25	0,12	0,03
0,67	19	5	1	0,45	8,53	2,24	0,45
0,46	2	0	0	0,21	0,42	0,00	0,00
0,69	6	0	0	0,48	2,86	0,00	0,00
0,4	5	0	0	0,16	0,80	0,00	0,00
0,12	25	12	0	0,01	0,36	0,17	0,00
0,35	4	1	0	0,12	0,49	0,12	0,00
0,82	4	1	0	0,67	2,69	0,67	0,00
TOTAL				7,22	71,67	33,27	3,38
CENTROID					9,92	4.61	0,47

In table 4, the centroid calculation is performed for the degree of membership in the first cluster. Based on the results of these calculations, the centroid is obtained for the cluster (V_{1j}) and the 3 variables are: V_{1j}; X1, X2, X3:{ 9.92;4.61;0.47}.

3. Calculating the objective function in the 1st iteration:

Table 5. Calculation of the Iteration Objective Function 1

DEEGRES OF MEMBERSHIP OF-i			DATA SET			$\left\|\sum(v_{kj} - v_{ij})^2\right\|(\mu_{1k})^2$	$\left\|\sum(v_{kj} - v_{ij})^2\right\|(\mu_{2k})$	$\left\|\sum(v_{kj} - v_{ij})^2\right\|(\mu_{3k})$	TOTAL
$\mu_{1k}{}^{\wedge 2}$	$\mu_{2k}{}^{\wedge 2}$	$\mu_{3k}{}^{\wedge 2}$	X1	X2	X3	L1	L2	L3	
0,03	0,37	0,04	10	5	1	0,01	10,77	35,52	46,31
0,55	0,01	0,02	8	2	0	5,87	1,01	19,75	26,63
0,00	0,67	0,03	12	8	1	0,01	9,92	16,33	26,26
0,31	0,00	0,18	4	1	0	15,15	0,06	224,10	239,31
0,05	0,02	0,38	125	79	22	1017,82	400,63	4655,11	6073,57
0,07	0,27	0,05	2	1	0	5,14	54,78	67,59	127,51

0,25	0,10	0,04	11	5	0	0,38	1,96	27,65	29,99
0,56	0,01	0,02	9	4	0	0,81	0,45	20,08	21,34
0,04	0,15	0,18	2	2	0	2,52	29,44	240,26	272,22
0,21	0,12	0,04	12	2	0	2,40	3,22	32,45	38,07
0,49	0,03	0,02	4	0	0	27,70	4,21	25,62	57,53
0,04	0,32	0,05	9	1	0	0,62	21,31	47,92	69,85
0,01	0,52	0,03	38	9	4	9,92	276,77	3,12	289,81
0,18	0,17	0,03	4	0	0	10,45	27,65	33,46	71,56
0,83	0,00	0,00	21	14	2	176,60	0,09	0,77	177,46
0,52	0,04	0,01	17	15	0	82,07	3,39	2,20	87,66
0,05	0,04	0,34	6	2	0	1,19	3,73	376,02	380,93
0,66	0,01	0,01	2	0	0	55,26	1,73	14,33	71,32
0,25	0,06	0,06	3	2	0	13,74	10,50	81,12	105,36
0,00	0,01	0,69	50	24	6	9,86	15,54	184,77	210,17
0,45	0,03	0,03	19	5	1	37,18	0,43	11,39	48,99
0,21	0,00	0,23	2	0	0	17,82	0,77	330,18	348,77
0,48	0,04	0,01	6	0	0	17,54	5,44	11,89	34,87
0,16	0,11	0,07	5	0	0	7,31	15,56	90,90	113,77
0,01	0,46	0,04	25	12	0	4,06	60,70	6,38	71,14
0,12	0,12	0,09	4	1	0	5,92	18,80	114,34	139,05
0,67	0,01	0,01	4	1	0	32,49	1,53	8,13	42,15
OBJECTIVE FUNCTION VALUE									9221,60

To determine the value of the objective function, a calculation is performed between the distance between the data and the cluster center multiplied by the square of the degree of membership. So the value of the objective function for the first iteration is 9221.600627. This value will later be used as a comparison whether the iteration will continue or not.

5. Calculating the changes in partition matrix

The results of the calculation of the partition matrix are shown in table 6.

Table 6. Partition Matrix Calculation

L1	L2	L3	LT	$(u_{1k})^{\wedge 2}$ L1/LT	$(u_{2k})^{\wedge 2}$ L2/LT	$(u_{3k})^{\wedge 2}$ L3/LT
2,26	0,03	0,00	2,29	0,98	0,02	0,00
0,09	0,01	0,00	0,11	0,86	0,13	0,01
0,06	0,07	0,00	0,13	0,47	0,52	0,01
0,02	0,01	0,00	0,03	0,74	0,23	0,03
0,00	0,00	0,00	0,00	0,27	0,29	0,43
0,01	0,00	0,00	0,02	0,70	0,26	0,04
0,65	0,05	0,00	0,70	0,93	0,07	0,00
0,69	0,02	0,00	0,72	0,97	0,03	0,00
0,01	0,01	0,00	0,02	0,71	0,26	0,04
0,09	0,04	0,00	0,13	0,70	0,29	0,01
0,02	0,01	0,00	0,02	0,72	0,25	0,03
0,07	0,02	0,00	0,09	0,81	0,17	0,01
0,00	0,00	0,01	0,01	0,10	0,15	0,75
0,02	0,01	0,00	0,02	0,72	0,25	0,03
0,00	0,01	0,00	0,02	0,24	0,52	0,24
0,01	0,01	0,00	0,02	0,30	0,56	0,14
0,04	0,01	0,00	0,06	0,81	0,18	0,02
0,01	0,00	0,00	0,02	0,69	0,27	0,04
0,02	0,01	0,00	0,02	0,73	0,24	0,03
0,00	0,00	0,00	0,00	0,10	0,13	0,77
0,01	0,07	0,00	0,08	0,15	0,82	0,03
0,01	0,00	0,00	0,02	0,69	0,27	0,04
0,03	0,01	0,00	0,04	0,75	0,22	0,02
0,02	0,01	0,00	0,03	0,74	0,24	0,03
0,00	0,01	0,01	0,02	0,20	0,44	0,36
0,02	0,01	0,00	0,03	0,74	0,23	0,03
0,02	0,01	0,00	0,03	0,74	0,23	0,03

Calculating the change in the partition matrix is done by calculating the distance between the data and the centroid and to the power of -1/*w*-1.

6. Stop Condition checking:

a. If (|Pt – Pt-1| < Ɛ)

P1 = 9221.600627

P0 = 0

|P1–P0| = 9221.600627, because P1 is still higher than Ɛ (0.001) then the iteration stop.

b. or (t > MaxItr), where t=1 and maxIter is 100. Then the iteration continues.

Teble 7. Cluster Results

No	District	Degree of Data Membership in the Cluster			Clusters Chosen	Cluster
		1	2	3		
1	Baktya	0,040505	0,002014	0,000052	0,040505	1
2	Baktiya Barat	0,301971	0,001546	0,000050	0,301971	1
3	Banda Baro	0,014028	0,002660	0,000055	0,014028	1
4	Cot girek	0,176205	0,001153	0,000047	0,176205	1
5	Dewantara	0,000049	0,000074	22,812551	22,8125508	3
6	Geureudong pase	0,054309	0,001023	0,000046	0,054309	1
7	Kuta Makmur	0,030731	0,002172	0,000053	0,030731	1
8	Langkahan	0,082192	0,001784	0,000051	0,082192	1
9	Lapang	0,057062	0,001049	0,000047	0,057062	1
10	Lhoksukon	0,029554	0,002071	0,000052	0,029554	1
11	Matang Kuli	0,116778	0,001119	0,000047	0,116778	1
12	Meurah Mulia	0,113241	0,001595	0,000050	0,113241	1
13	Muara Batu	0,000928	0,012578	0,000078	0,012578	2
14	Nibong	0,116778	0,001119	0,000047	0,116778	1
15	Nisam	0,002714	0,011005	0,000065	0,011005	2
16	Nisam Antara	0,003479	0,005248	0,000062	0,005248	2
17	Paya Bakong	20,258315	0,001349	0,000049	20,258315	1
18	Pirak timu	0,046946	0,000997	0,000046	0,046946	1
19	Samudera	0,098464	0,001115	0,000047	0,098464	1
20	Sawang	0,000410	0,001996	0,000113	0,001996	2
21	Seunudon	0,005735	0,004800	0,000058	0,005735	1
22	Simpang Keramat	0,046946	0,000997	0,000046	0,046946	1
23	Syamtalira Aron	0,261405	0,001264	0,000048	0,261405	1
24	Syamtalira Bayu	0,192516	0,001188	0,000047	0,192516	1
25	Tanah jambo Aye	0,002197	0,025735	0,000067	0,025735	2
26	Tanah Luas	0,176205	0,001153	0,000047	0,176205	1
27	Tanah Pasir	0,176205	0,001153	0,000047	0,176205	1

Based on table 7, there are 21 sub-districts belonging to cluster 1, 5 sub-districts in cluster 2, and 1 sub-district in cluster 3.

The following table shows the results of clustering in Aceh Utara.

Table 8. Clustering Results in Aceh Utara

No	Cluster 1	Cluster 2	Cluster 3
1	Baktiya	Muara Batu	Dewantara
2	Baktiya Barat	Nisam	
3	Banda Baro	Nisam Antara	
4	Cot girek	Sawang	
5	Geureudong pase	Tanah jambo Aye	
6	Kuta Makmur		
7	Langkahan		
8	Lapang		
9	Lhoksukon		
10	Matang Kuli		
11	Meurah Mulia		
12	Nibong		
13	Paya Bakong		
14	Pirak timu		
15	Samudera		
16	Seunudon		
17	Simpang Keramat		
18	Syamtalira Aron		
19	Syamtalira Bayu		
20	Tanah Luas		
21	Tanah Pasir		

Based on table 8 above, cluster 1 is the cluster with the lowest level consisting of 21 sub-districts, namely Baktiya, Baktiya Barat, Banda Baro, Cot Girek, Geureudong Pase, Kuta Makmur, Langkahan, Lapang, Lhoksukon Districts, Matang Kuli, Meurah Mulia, Nibong, Paya Bakong, Pirak Timu, Samudera, Seunudon, Simpang Keramat, Syamtalira Aron, Syamtalira Bayu, Tanah Luas and Tanah Pasir. Cluster 2 is a cluster with a moderate level consisting of 5 sub-districts, namely Muara Batu, Nisam, Nisam Antara, Sawang and Tanah Jambo Aye Districts. Then Dewantara District is included in cluster 3 with the highest distribution rate. The clustering results are presented in the form of a graph shown in Figure 2.

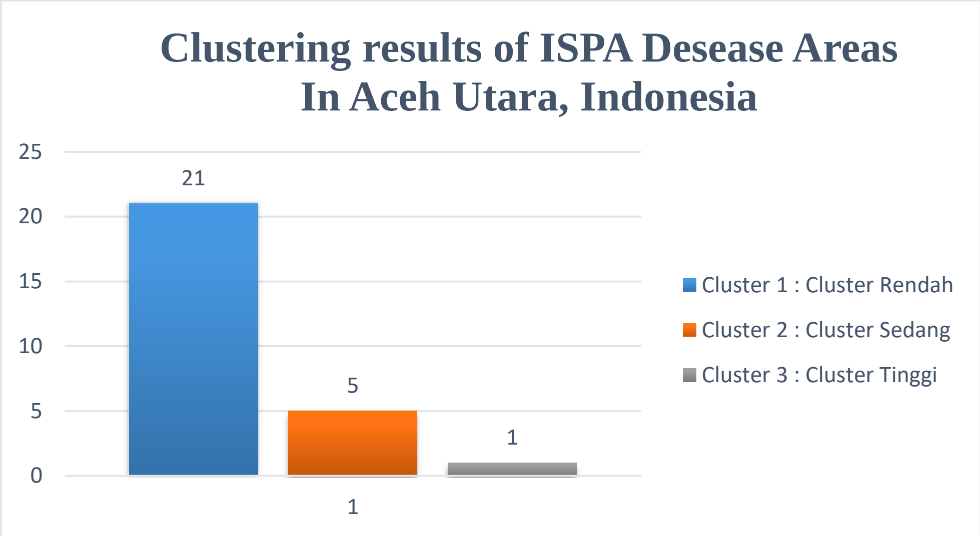

Figure 2. Cluster Graph

The picture above shows that the blue column is cluster 1 which is a low cluster with a total of 21 sub-districts, then cluster 2 is a medium cluster with a red column consist of 5 sub-districts, and the green column is cluster 3 with the highest distribution rate of 1 sub-district.

3.3 System Implementation

The results of system implementation are shown in Figure 3.

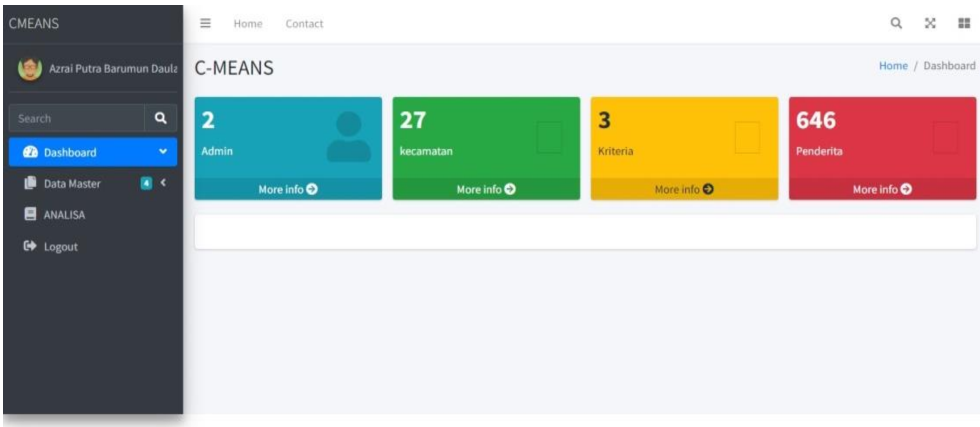

Figure 3. Dashboard form

Figure 3 is the system dashboard form, on this form displays all the menus contained in the system.

4. CONCLUSION

This research classified the data into 3 clusters, namely cluster 1, cluster 2 and cluster 3. Cluster 1 is the cluster with the lowest level or the least spread of ISPA disease. Cluster 2 is the medium level and cluster 3 is the cluster with the highest spread of ISPA. Based on the data obtained in this research, the most common type of ISPA disease that the Aceh Utara residents suffer from is pneumonia, with a total of 414 cases or 64% of the total data.

6. REFERENCES

[1] H. Azzaoui, I. Manssouri, B. Elkihel, Methylcyclohexane continuous distillation column fault detection using stationary wavelet transform & fuzzy C-means. Materials Today: Proceedings, 13, 597-606, 2019.

[2] S. Askari, Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development. Expert Systems with Applications, 165, 113856, 2021.

[3] R. K. Dinata, S. Retno, N. Hasdyna, Minimization of the Number of Iterations in K-Medoids Clustering with Purity Algorithm. Rev. d'Intelligence Artif., 35(3), 193-199, 2021.

[4] S. Debnath, F. A. Talukdar, M. Islam, Combination of contrast enhanced fuzzy c-means (CEFCM) clustering and pixel based voxel mapping technique (PBVMT) for three dimensional brain tumour detection. Journal of Ambient Intelligence and Humanized Computing, 12(2), 2421-2433, 2021.

[5] B. Cardone, F. Di Martino, A novel fuzzy entropy-based method to improve the performance of the fuzzy C-means algorithm. Electronics, 9(4), 554, 2020.

[6] R. Rout, P. Parida, Y. Alotaibi, S. Alghamdi, O.I. Khalaf, Skin lesion extraction using multiscale morphological local variance reconstruction based watershed transform and fast fuzzy C-means clustering. Symmetry, 13(11), 2085, 2021.

[7] S. Munusamy, P. Murugesan, Modified dynamic fuzzy c-means clustering algorithm–Application in dynamic customer segmentation. Applied Intelligence, 50(6), 1922-1942, 2020.

[8] N. Hasdyna, S. Retno, Purity Algorithm in Determining System of The Productivity of Rice Harvesting Areas in Kabupaten Aceh Utara. JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING, 5(2), 259-267, 2022.

[9] Z. Cai, H. Ma, L. Zhang, A building detection method based on semi-suppressed fuzzy C-means and restricted region growing using airborne LiDAR. Remote Sensing, 11(7), 848,2019.

[10] A. G. Oskouei, M. Hashemzadeh, B. Asheghi, CGFFCM: cluster-weight and Group-local Feature-weight learning in Fuzzy C-Means clustering algorithm for color image segmentation. Applied Soft Computing, 113, 108005, 2021.

[11] N. Hasdyna, S. Retno, Machine Learning Approach to Determine the Drug-Prone Areas in Lhokseumawe City, Indonesia, International Journal of Multidisciplinary Research and Analysis., 5(9), 2454-2463, 2022.

[12] N. Ait Ali, B. Cherradi, A. El Abbassi, O. Bouattane, GPU fuzzy c-means algorithm implementations: performance analysis on medical image segmentation. Multimedia Tools and Applications, 77(16), 21221-21243, 2018.

[13] A. M. Anter, A. E. Hassenian, D. Oliva, An improved fast fuzzy c-means using crow search optimization algorithm for crop identification in agricultural. Expert Systems with Applications, 118, 340-354, 2019.