

Fuzzy C-Means Clustering Algorithm For Grouping Health Care Centers On Diarrhea Disease

Ahmad Chusyairi^{a,1,*}, Pelsri Ramadar Noor Saputra^{b,2} Efendi Zaenudin^{c,3}

^a Universitas Bina Insani, Raya Siliwangi No. 6 Rawa Panjang, Bekasi 17114, Indonesia

^b STIKOM PGRI Banyuwangi, Ahmad Yani Street No. 80, Banyuwangi 68421, Indonesia

^c Department of Bioinformatics and Medical Engineering, Asia University, Taichung Taiwan, Province of China

¹ ahmadchusyairi@binainsani.ac.id*; ² ramayana.x@gmail.com

* corresponding author

ARTICLE INFO

Article history:

Received 20-09-2020

Revised 21-10-2020

Accepted 05-01-2021

Keywords:

Clustering Algorithm

Diarrhea Disease

Fuzzy C-Means

Health Care Center

ABSTRACT

In Indonesia, public health services at the city or district level are carried out by regional public hospital or “puskesmas” (health care centers), especially in Banyuwangi regency, East Java, Indonesia that have 45 health care centers spread throughout the villages. This research focused to deaths of baby caused by diarrhea diseases, which are the second leading cause of death among children younger than 5 years globally. All of the health care centers need to be divided into 3 groups to find out which health care centers have the least, most moderate, and many diarrhea sufferers. Fuzzy C-Means algorithm is used to overcome this problem. The result from this research shown that 2 health care centers have the smallest member of diarrhea sufferer, 14 health care centers have medium member of diarrhea sufferers, and the rest have a large number of diarrhea sufferers. From the result of this study, it can be a reference for the health department center in dealing with diarrheal diseases, accordingly the infant mortality rate due to diarrheal diseases can be lowered to health care centers that have high diarrhea sufferers.

Copyright © 2017 International Journal of Artificial Intelligence Research.

All rights reserved.

I. Introduction

Health is a basic human need. The success of health development is strongly influenced by the availability of healthy, skilled and expert human resources and is prepared in a health program with integrated planning supported by valid epidemiological data and information [8]. In the city or district level, a health service carried out by regional public hospital using “puskesmas” (health care centers) as a its pioneer, that is way a health care center must carry out health service to the fullest.

In this world, the Indonesian health index takes place in position 101st out of 149 countries, which means, Indonesia’s health level is below average. There are several factors to see the public health level among others infant mortality rate, maternal mortality rate, and life expectancy rate [11]. On infant mortality rate, diarrheal disease is one of the causes of the death for infant. As the second leading cause of death among children under 5 years, diarrhea is responsible for more than 500,000 deaths globally every year [7, 17]. Diarrheal disease kills most of the children compared to Acquired Immunodeficiency Syndrome (AIDS), malaria and measles [1].

Most deaths occur in the middle and low incoming countries where children are exposed to risk factors such as unsafe drinking, water poor sanitation and hygiene practices, micronutrient deficiencies, malnutrition, and lack of access to high-quality health services [14, 6]. Diarrhea can be seen from the consistency of runny green stools or can also be mixed with blood mucus or mucus only. Diarrhea usually caused by a viral infection in the digestive system. Diarrhea becomes dangerous and deadly when it consumes body fluids, resulting in fatal dehydration. To improve public health, especially in dealing with diarrheal diseases, it is necessary to improve the quality of health services at health care center. Each health care center unit is responsible for service and education to overcome the problem of diarrhea in the villages that are included in the area of its health care center unit. Grouping of health care center units will be conducted based on diarrheal

disease data per year obtained from health office center at Banyuwangi Regency. The purpose of this grouping is to know which groups of health care center have low diarrhea rate, groups of health care center that have succeeded in reducing the number of diarrhea rate, and groups of health care center that must improve services therefore the number of diarrhea sufferers can be reduced.

Banyuwangi Regency has 45 health care centers consisting 18 Nursing health care centers and 27 Non-Nursing health care centers, and 105 Helper health care centers spread in 24 Districts and 217 Villages [3]. A health care center must formulate strategical policies including internal efficiency (organization, management, and human resources) and must be able to quickly and accurately make a decision to improve a community service, so they can be responsive, innovative, effective, efficient and profitable [13].

The research conducted by Simhachalam and Ganesan [2] was to classify patients against several different groups of thyroid disease using the Possibilistic Fuzzy C-Mean (PFCM) and Fuzzy C-Mean (FCM) methods. In the medical diagnostic system, Fuzzy clusters use membership degrees in such a way to allow the same object to multiple clustering simultaneously with different degrees. This method is important to increase sensitivity in diagnosing thyroid disease. Based on results from this research that have been carried out repeatedly, the result show that positive Fuzzy C-Means clustering gives better grouping results than Fuzzy C-Means. That is way this method can help to diagnose the disease better for patients based on cluster obtained from the method. The research that was done by Ramya [16] predicted disease based on the symptoms experienced by patients by using the Fuzzy C-Means method therefore prevention could be made of the diseases suffered by patients in reducing the risk of death. The system would analyze patient's health condition and the result of analysis would be used in taking action according to the results obtained, so that it could help practitioners or doctors to analyze patient health. The results of this research had been done by calculating the numerical value based on the severity of the symptoms, hence the patient's disease could be predicted more accurately based on the patient's medical history.

Chetty, Vaisla, dan Patil [10] analyzed the method of improvement in predicting disease using a Fuzzy approach, this study predicted diabetes and liver disorders with two different approaches, among others classification with the K-Nearest Neighbor (KNN) algorithm and Fuzzy C-Mean (FCM) Clustering algorithm in clustering data and classified by using KNN. The results achieved in this study are that the proposed method works well and is more accurate than the first approach. The last, the research that was carried out by Imianvan A. A., Anosike U.F, and Obi J. C. [4] using Fuzzy C-Mean clustering system to diagnose HIV. Based on the symptoms suffered by the patient, a series of methodological and analytical decision steps is made on the resulting cluster. Based on a series of symptoms in the assessment of diagnostic patterns by determining the level of individual symptom membership, the system can provide precise results compared to traditional systems.

From several studies that have been proposed, this study focuses on predictions of grouping health care centers using the Fuzzy C-Mean Clustering algorithm, therefore the group can determine the number of diarrhea patients, groups that have succeeded in reducing the number of diarrhea sufferers, and groups who must improve services to the community so that the number of diarrhea sufferers can be reduced. From the data grouping, the health office can provide additional assignments for health care center groups that have high rates of diabetes sufferers in providing counseling to the community in consequence the number of diarrhea sufferers can be reduced.

The scope of the research is not only used in cases of infant mortality due to diarrhea in Banyuwangi, but it can be applied nationally so that the number of cases of infant mortality can be pressed and Indonesia's health index can increase. The next section is designed as follows: Theoretical analysis to explain the methods used in this study. Algorithm analysis describes sequences of cluster system flows. Discussion and the results of the study explain the results of a summary of the research in the proposed system flow experiment. The last section is the conclusion of this research.

II. Theoretical Analysis

A. Clustering

Clustering is also called classification refers to the dataset partition of objects into the most similar groups of objects. These objects can be numerical, categorical or both [2]. The ability of

clustering algorithms is to reveal the underlying structure of data that can be exploited in a variety of applications, including classification, numerical taxonomy, image processing, pattern recognition, medicine, economics, ecology, artificial intelligence, data mining, modeling and identification. Fuzzy clustering uses membership degrees that allows an object to enter multiple clusters simultaneously, with various levels of membership.

B. Fuzzy Logic

Fuzzy logic is used to map problems from input to expected output. There is fuzzy clustering which is one method for determining optimal clusters in a vector space based on the normal Euclidian form for distance between vectors [12]. Fuzzy logic method namely Fuzzy C-Mean is often used to cluster the data.

C. Fuzzy C-Means

Fuzzy C-Means (FCM) is a soft grouping technique where each point has a member level of a cluster based on fuzzy logic [12]. Fuzzy C-Means clustering (FCM) uses fuzzy partitions accordingly a data point can be owned by more than one group with different membership values (weights) ranging from 0 to 1 [2]. The closer the data point to the cluster center, a membership of its data point will be towards the center of the cluster. The sum of the membership values of each object must be equal to one [10]. The basic steps of the Fuzzy C-Means algorithm are shown as follows:

- The first step: insert the data to be clustered in the form of a matrix measuring $n \times m$, then specify the number of cluster c ($1 \leq c \leq N$), matrix partition, power w , maximum iteration, least expected error $\epsilon > 0$, initial objective function $P_0 = 0$, and first iteration $t = 1$.
- Second step: set a random numbers (μ_{ik}) as the initial partition matrix element U . Calculate the number of each column:

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (1)$$

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (2)$$

- Third step: calculate the center of the k cluster: V_{kj}

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

- Fourth step: calculate the objective function in the t iteration

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

- Fifth step: calculate the change in matrix partition:

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

- Sixth step: check the stop condition:
 - a. If: $(|P_t - P_{t-1}| < \epsilon)$ or $(t > \text{MaxIter})$, then stop, else If: $t = t + 1$, repeat step four

The output generated from Fuzzy C-Means is a cluster of central clusters and several degrees of membership for each data point

III. Algorithm Analisis

A. Data Preparation and Pre-Process

Data on diarrheal disease sufferers in infants is obtained from the Banyuwangi District health office from 2016 to 2019 as trial data in this study. The data is shown at table I. There are 45 health care centers data that are spread in Banyuwangi District, with data usage from 2016 to 2019. Based on these data, normalization data is used for data on diarrhea patients in table I, making it easier to calculate using Fuzzy C-Mean.

Table 1. Diarrhea Patient Data

Health Care Center	Patient Data			
	2016	2017	2018	2019
Wongsorejo	532	1,109	244	481
Bajulmati	464	1,008	222	437
Kelir	210	536	119	234
Klatak	800	1,776	393	776
Mojopanggung	322	804	176	349
Paspan	382	950	209	412
Licin	322	780	172	338
Sobo	572	1,254	274	542
Singotrunan	458	1,011	221	436
Kertosari	326	675	148	292
Kabat	550	1,606	352	696
Badean	312	1,392	305	604
Gitik	678	927	203	401
Gladag	430	515	113	223
Singojuruh	568	1,231	269	531
Songgon	630	1,363	298	588
Kebaman	294	737	161	318
Parijatah Kulon	358	787	173	340
Wonosobo	390	886	194	383
Kedungrejo	522	1,048	230	453
Sumberberas	368	816	179	353
Tapanrejo	234	588	129	255
Tembokrejo	540	1,157	254	501
Tegaldlimo	416	967	212	417
Kedungwungu	314	716	156	309
Purwoharjo	402	945	206	408
Grajagan	376	832	182	359
Benculuk	530	1,203	264	520
Tampo	304	727	159	315
Jajag	344	785	172	340
Yosomulyo	378	836	183	361
Tegalsari	516	1,281	281	554
Genteng Kulon	610	1,276	279	551
Kembiritan	526	1,029	225	444
Sempu	400	854	189	373
Karangsari	406	854	186	368

Health Care Center	Patient Data			
	2016	2017	2018	2019
Gendoh	112	235	52	102
Sepanjang	488	1,064	232	460
Tulungrejo	404	854	187	369
Kalibaru Kulon	818	1,716	376	743
Kebondalem	374	841	185	364
Sambirejo	356	791	173	342
Pesanggaran	254	570	125	248
Sumberagung	374	766	168	330
Siliragung	532	1,220	267	528

The normalization process is performed for scaling the data values within a specified range [5]. Clustering is the process of grouping the similar data. The predominantly used existing clustering algorithm is K-Means. But, the K-means algorithm demands to specify the number of clusters in prior, they do not have the ability to resolve the overlapping data into two clusters, and the algorithm cannot handle noisy data, categorical data, and non-linear dataset [5].

Min-Max normalization is a simple technique where the technique can specifically fit the data in a pre-defined boundary with a pre-defined boundary [15]. The min-max formulation is shown in the following equation [18]:

$$E_v = \frac{X_0 - X_{min}}{X_{max} - X_{min}} \tag{6}$$

Where,

- X_n = new value for variable X
- X_0 = current value for variable X
- X_{min} = minimum value in data set
- X_{max} = maximum value in data set

Using Min-Max normalization to scale the values in the range Characterized. After normalizing, the new values are matrix between 0 and 1, which is most used to select tasks [9].

B. Performance Analysis

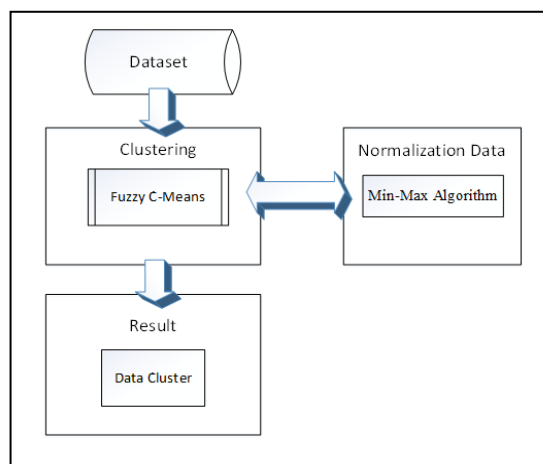


Fig. 1. Clustering Model

Block diagram system is used in knowing the input data needed, the process that occurs, the process results, and the interactions that occur between users of the system described in the diagram generated from the results of the interaction in Figure 1. In this study, clustering of data is obtained by using data input from 45 health care centers data from 2016 to 2018 shown in table 1.

First, all of data will be normalized in simplifying calculations in using the Fuzzy C-Mean method. From the input data used based on the data with the cluster method so that the data grouping will be generated. Details of the Fuzzy C-Means process are shown at figure 2. Each step will be described as follows:

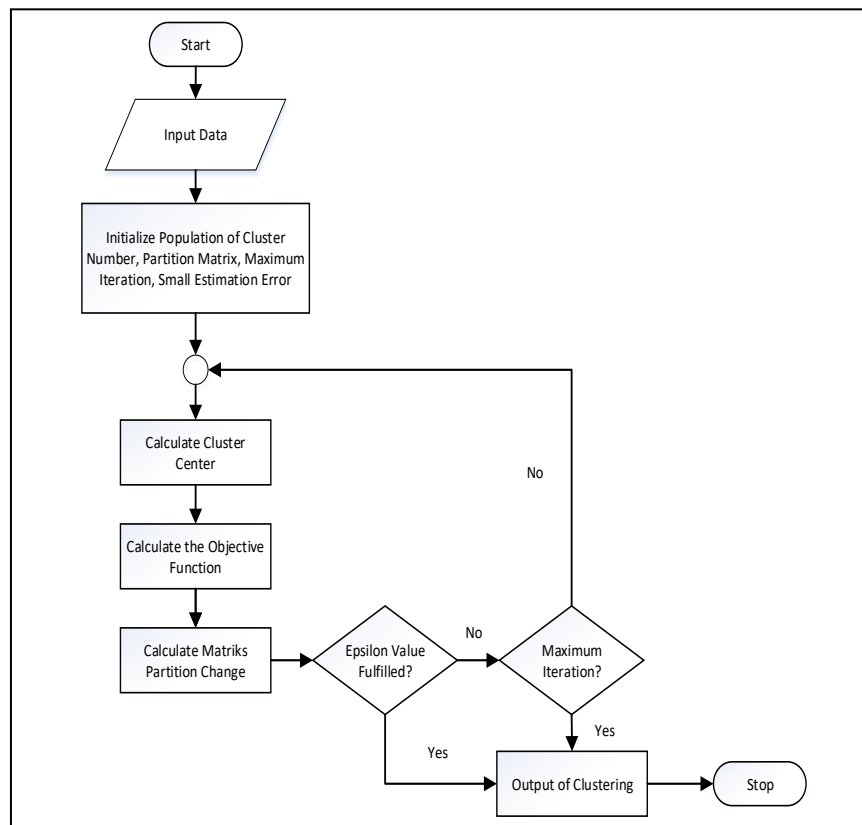


Fig. 2. Clustering Model

- Set the input data are that shown at table 1
- Initialization of cluster population, matrix partition, maximum iteration, least expected error. This section is the most important part of the FCM calculation process because it will determine the final result at the data clustering
- Calculate cluster center. If the cluster center has been calculated for all data, then it will perform the objective function calculation
- Next step, calculate of the matrix partition changes to the value of the objective function on the previous iteration
- If the epsilon value is not fulfilled, then it will be checked whether the maximum iteration has been finished. Otherwise, recalculate the next iteration
- If one of the epsilon or maximum iteration values has been fulfilled, then show the cluster results, and the FCM calculation process stops

IV. Discussion and Result

A. Set the Initial Value

Before starting the process of clustering using the Fuzzy C-means method, first determine the initial value that will be used in the FCM calculation. This is important to do because it will affect the result that will be produced in this research. The initial data values shown at table 2.

From the predetermined initial values, it will affect the result of the health care center data cluster decision. Then we will know the health care center data that have the smallest number of diarrhea patients, health care center data that have succeeded in reducing the number of diarrhea sufferers, and a health care center data that are lacking in dealing with diarrhea.

Table 2. Initial Value

Name	Symbol	Value
Number of Cluster	c	3
Power	w	2
Maximum Iteration	MaxIter	100
Least Error Expected	ϵ	10^{-5}
Initial Objective Function	P_0	0
Initial Iteratio	t	1

B. Clustering of Health Care Center Data

Based in initial values that has been set, Fuzzy C-Means algorithm will be performed. The data will be normalized for the first time based on data shown at table I. data normalization is done by using the Min-max method where the smallest and largest values will be searched based on the data, and the calculation process is performed on each data. The result shown that data will have values between 0 to 1 from this process, making it easier for the next calculation process.

Next, generate the initial matrix partition. The data will be divided into 3 clusters then calculate toward the cluster center by calculating the degree of membership in the i cluster toward data on diarrhea sufferers. Then recalculate the square of the degree of membership of the i data to find out the value of the overall objective function. This objective function value will be used to check the value of the next objective function whether the difference in value has been below the value of the smallest error expected or not. If the difference between the two objective function values is not below the expected smallest error value, repeat the iteration and the membership degree value will use the results of the calculation of the partition matrix value μ . The calculation will stop after the difference in the value of the objective function between the iterations calculated against the previous iteration has a value below the expected smallest error value or maximum iteration.

C. Data Analysis

From the results of calculations that have been done, it is obtained the results of data clustering of health care centers in Banyuwangi District on diarrhea patient data. After finding the difference in objective function values below the expected smallest error value, the clustering decision on 45 health care center data is divided into 3 clusters. Cluster results are shown in table 3.

Based on the data obtained from the results of clustering using the Fuzzy C-Means algorithm, it was found that from 45 data centers the health care center was divided into 3 clusters. The first cluster has 11 community health care center data, i.e. Wongsorejo, Bajulmati, Paspan, Singotrunan, Gitik, Wonosobo, Kedungrejo, Tegaldlimo, Purwoharjo, Kembiritan, and Sepanjang. In this first cluster have a smallest rate who have a diarrhea sufferer for children who are under 5 years of age. Whereas the second cluster contained 12 health care center data including Klatak, Sobo, Kabat, Badean, Singojuruh, Songgon, Tembokrejo, Benculuk, Tegalsari, Genteng Kulon, Kalibaru Kulon, and Siliragung. This second cluster is a health care center data with diarrhea rates for children under 5 years that are better than the third cluster.

Table 3. Health Care Center Cluster

Cluster 1	Cluster 2	Cluster 3
Wongsorejo	Klatak	Kelir
Bajulmati	Sobo	Mojopanggung
Paspan	Kabat	Licin
Singotrunan	Badean	Kertosari
Gitik	Singojuruh	Gladag
Wonosobo	Songgon	Kebaman
Kedungrejo	Tembokrejo	Parijatah Kulon
Tegaldlimo	Benciluk	Sumberberas
Purwoharjo	Tegalsari	Tapanrejo
Kembiritan	Genteng Kulon	Kedungwungu
Sepanjang	Kalibaru Kulon	Grajagan
	Siliragung	Tampo
		Jajag
		Yosomulyo
		Sempu
		Karangsari
		Gendoh
		Tulungrejo
		Kebondalem
		Sambirejo
		Pesanggaran
		Sumberagung

The last cluster included 22 health care center data including Kelir, Mojopanggung, Licin, Kertosari, Gladag, Kebaman, Parijatah Kulon, Sumberberas, Tapanrejo, Kedungwungu, Grajagan, Tampo, Jajag, Yosomulyo, Sempu, Karangsari, Gendoh, Tulungrejo, Kebondalem, Sambirejo, Pesanggaran, and Sumberagung. The data in this cluster are data from community health care centers that have high rates of diarrhea for children under 5 years.

V. Conclusion

Based on the results obtained from the research that has been done, the results of grouping on health care center data in Banyuwangi Regency using the Fuzzy C-Means method can be used as a reference by the Banyuwangi Health Office in reducing the number of diarrhea sufferers for children under 5 years old, especially for health care centers that have high rates of diarrhea sufferers.

The health office can provide special assignments to the health care centers in providing counseling and education to the society on how to deal with factors especially that can cause diarrhea so that the number of diarrhea sufferers can be reduced for children under 5 years of age so that the infant mortality rate due to diarrheal diseases can be suppressed.

And from these results can be a reference for health care centers in the third cluster in improving their services to reduce the number of diarrhea sufferers. The impact of the results from this study can be used as a reference in improving national public health so that the Indonesia's health level can increase. In the next research, we will develop this method with a new approach so that the result will be even better to reduce a diarrhea patient.

Acknowledgment

We are very grateful and appreciative of Bina Insani University and STIKOM PGRI Banyuwangi for their support and efforts for this research so that it can run as planned and well.

References

- [1] A. B. Tambe, L. D. Nzefa, and N. A. Nicoline, "Childhood Diarrhea Determinants in Sub-Saharan Africa: A Cross Sectional Study of Tiko-Cameroon", *Challenges*, 2015, Vol. 6(2), pp: 229-243
- [2] B.Simhachalam, G.Ganesan, "Possibilistic Fuzzy C-Means Clustering On Medical Diagnostic Systems", *International Conference on Contemporary Computing and Informatics (IC3I)*, 2014, pp: 1125 – 1129
- [3] Chusyairi A., Saputra P.R.N., "Grouping Health Care Center in Achieving Immunization at Banyuwangi Regency Using K-Means Clustering Method", *Telematika*, 2019, Vol. 12(2), pp:139-148 (Chusyairi A., Saputra P.R.N., "Pengelompokan Data Puskesmas Banyuwangi Dalam Pemberian Imunisasi Menggunakan Metode K-Means Clustering", *Telematika*, 2019, Vol. 12(2), pp:139-148)
- [4] Imianvan A. A., Anosike U.F, Obi J. C., "An Expert System for the Intelligent Diagnosis of HIV Using Fuzzy Cluster Means Algorithm", *Global Journal of Computer Science and Technology*, 2011, Volume XI (XII)
- [5] J. Sangeetha, V. S. J. Prakash, "An Efficient Inclusive Similarity Based Clustering (ISC) Algorithm for Big Data", *World Congress on Computing and Communication Technologies (WCCCT)*, 2017
- [6] J. Tetteh, W. K. Takramah, M. A. Ayanore, A. A. Ayanore, E. Bisung, and J. Alamu, "Trends for Diarrhea Morbidity in the Jasikan District of Ghana: Estimates from District Level Diarrhea Surveillance Data, 2012–2016", *Hindawi Journal of Tropical Medicine*, 2018
- [7] L. Carvajal-Vélez, A. Amouzou, J. Perin, A. Maïga, H. Tarekegn, A. Akinyemi, S. Shiferaw, M. Young, J. Bryce and H. Newby, "Diarrhea management in children under five in sub-Saharan Africa: does the source of care matter? A Countdown analysis", *BMC Public Health*, 2016
- [8] Ministry of Health Republic of Indonesia, "Indonesian Health Profile and Information 2018", unpublished. (Kementrian Kesehatan Republik Indonesia, "Data Dan Informasi Profil Kesehatan Indonesia 2018")
- [9] M. Derakhshan, Z. Bateni, "Optimization of Tasks in Cloud Computing Based on MAX-MIN, MIN-MIN and Priority", *4th International Conference on Web Research (ICWR)*, 2018
- [10] N. Chetty, K. S. Vaisla, N. Patil, "An Improved Method for Disease Prediction using Fuzzy Approach", *Second International Conference on Advances in Computing and Communication Engineering*, 2015
- [11] Public Health Office, "Health Profile 2017", Banyuwangi Regency Government, 2017 (Dinas Kesehatan, "Profil kesehatan 2017", Pemerintah Kabupaten banyuwangi, 2017)
- [12] Shabari S. B., S. Shetty, M Siddappa, "Implementation and Comparison of K-Means and Fuzzy C-Means Algorithms for Agricultural Data", *International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2017
- [13] Syifa'ul L., D. Prayoga, S. Haksama, A. M. Mandagi, "Analysis of SIMPUS (Si Jempol Sehat Si Jempol Wangi) Development in Banyuwangi Regency", *IKESMA Vol. 12 No. 1*, 2016, pp: 8–17 (Syifa'ul L., D. Prayoga, S. Haksama, A. M. Mandagi, "Analisis Pengembangan SIMPUS (Si Jempol Sehat Si Jempol Wangi) Di Kabupaten Banyuwangi", *IKESMA Vol. 12 No. 1*, 2016, pp: 8–17)
- [14] S. Misbah and K. Rasheed, "Maternal Knowledge and Management Skills about Diarrhea among Children under 5 years of age", *13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 2016, pp: 92–95
- [15] S. Patro, K. K. Sahu, "Normalization: A Preprocessing Stage", *Computer Science*, 2015
- [16] T. B. Ramya, "Disease Prediction System Using Fuzzy C-Means Algorithm", *International Journal of Engineering Research & Technology (IJERT)*, 2018, Volume 6 (3), pp: 1 – 4
- [17] Y. Tao, D. Bhattacharjya, A. R. Heching, A. Vempaty, M. Singh, F. Lam, J. Houdek, M. Abubakar, A. Abdulwahab, T. Braimoh, N. Ihebuzor, A. Mojsilović, and K. R. Varshney, "Effectiveness of peer detailing in a diarrhea program in Nigeria", *IBM Journal of Research and Development*, vol. 61, Issue: 6, 2017, pp: 1:1 - 1:12
- [18] Zuriani M., Yuhanis Y., "A Comparison of Normalization Techniques in Predicting Dengue Outbreak", *International Conference on Business and Economics Research (IACSIT)*, 2011, Volume 01, pp: 345 – 349