

CFGWC-PSO in Analyzing Factors Affecting the Spread of Dengue Fever in East Java Province

Siti Nurmardia Abdussamad^{1*}, Suci Astutik², Achmad Effendi²

¹Master Program of Statistics, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Malang, Indonesia

²Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Malang, Indonesia

Abstract

Fuzzy Geographically Weighted Clustering-Particle Swarm Optimization using Context Based Clustering (CFGWC-PSO) has been developed to clustering in factors influencing the spread of dengue fever in East Java Province. CFGWC-PSO method can overcome slow computing time problems in terms of iterations, and produce accurate data partition with stable. In this research, CFGWC-PSO applied to 11 variables from data on the causes of the spread of dengue fever in East Java Province in 2017. CFGWC-PSO using the FCM method to determine the context variable. Processing used the results of clustering with 2 clusters until 5 clusters. From the three validation index that used to find out the right number of clustering, two clusters gave better clustering results. CFGWC-PSO shows that all districts/cities in cluster 2 become dengue fever endemic areas that need to be considered by the East Java Provincial Government.

Keywords: Context-Based Clustering, dengue hemorrhagic fever, Fuzzy Geographically Weighted Clustering-Particle Swarm Optimization.

INTRODUCTION

Dengue Haemorrhagic Fever (DHF) is an important health problem in Indonesia. It was first known in 1979 in Cairo, and in the same year it also happened in Asia, is Jakarta, which was still called Batavia. But actually, dengue fever in Indonesia was first discovered in the city of Surabaya in 1968. As many as 58 people were infected, and 24 of them died. Since then, this disease has spread throughout Indonesia [1].

DHF is a problem that is routinely in every rainy season and after the rainy season. DHF transmitted through the bite of the *Aedes aegypti* and *A. albopictus* mosquitoes that carry the dengue virus. Vector control always used to cut off the chain of transmission in preventing DHF. It is because there is no vaccine to prevent dengue virus. DHF vector is actually easy to control. However, because the vector spread widely, the success of the control requires total coverage (covering the entire area) for the mosquitoes cannot reproduce again.

Factors related to the increase in dengue and outbreak incidence in East Java that is difficult to control were based on the East Java Health Office [2]. It is related with population density, population mobility, urbanization, economic growth, community behavior, climate change,

environmental sanitation conditions, and availability clean water (PHBS).

Fuzzy Geographically Weighted Clustering (FGWC) offers an alternative solution from regular clustering algorithms that better accommodate AGD with the ability to apply population effects and distances to geodemographic clustering analyses [3]. The previous study used FGWC integrates context-based clustering in its analysis, where CFGWC can reduce computing time and computing speed. It is expected that these results can accommodate spatial influence and also as alternative, which geographically aware by supporting the ability to apply population and distance effect to analyze geo-demographic clusters [4,5].

The purpose of this study is to integrate context-based clustering with FGWC-PSO to investigate clustering patterns of endemic areas in East Java based on the causes of dengue. Context-based clustering in this study was determined by FCM method. The result of this study will solve the problems in each cluster, which can be used by local governments to develop policies related to vector control strategies for the spread of dengue in East Java Province. The first government policy that can be implemented is by conducting a survey so that the policies to be carried out are more targeted.

MATERIAL AND METHOD

Study Area

The location of this research is East Java Province, which consists of 38 districts/cities. East Java Province geographically located between 7°12' - 8°48' latitude and 111°0' - 114°4'

¹ Correspondence address:

Siti Nurmardia Abdussamad

Email : sitinurmardiaabd@gmail.com

Address : Dept. Statistics, Faculty Mathematics and Natural Science, University of Brawijaya, Veteran Malang, Malang 65145.

longitude. The total area of East Java Province is 47,799.75 km² [5]. The observation unit used is the district/city level in East Java Province, which consists of 38 districts/ cities.

Data Collection

This study uses secondary data on the factors that cause the spread of dengue hemorrhagic fever in 2017. The parameter used were percentage of dengue fever (X₁), percentage of bamboo walls (X₂), percentage of poor household (X₃), percentage of unhealthy houses (X₄), percentage of houses not PHBS (X₅), population density per Km² (X₆), percentage of rainwater shelter (X₇), number of flood events in a year (X₈), percentage of palm fiber roof (X₉), number of health facilities (X₁₀), number of cases of malnourished toddlers (X₁₁). Data was taken from the Central Bureau of Statistics (BPS) of East Java and East Java Province Health Office in 2018 [6,7]. Data on the population of districts/cities and distance of districts/cities also needed as supporting data for weighting.

Statistical Methods

Clustering dengue-endemic areas based on the factors that cause DHF spread using analysis from the CFGWC-PSO method. CFGWC-PSO method is a hybrid of the FGWC-PSO algorithm method and context-based clustering, which stands for CFGWC-PSO.

FGWC algorithm has several limitations in the initialization stage. First, the number of geodemographic clusters must be manually defined by the user. Second, the cluster center (centroid) is determined randomly so that the iteration process fails to reach the optimum global solution. To overcome this limitation, the PSO algorithm was used to select the center cluster or membership matrix in the FGWC initialization phase. FGWC objective function that will be minimized is as follows [8].

$$J_{FGWC}(U, V; X) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m |v_i - x_k|^2 \rightarrow \min \dots\dots 1$$

Where *m* is the weight exponent that determines the fuzziness clusters, *u_{ik}* is an element of the partition matrix, *v_i* is the center of the cluster, and *x_k* is the data point.

Objective function *J_{FGWC}(U, V; X)* will be minimized by optimizing through parameters U and V. Lagrange multiplier *λ_k* with constraint $\sum_{i=1}^c u_{ik} = 1$ is used for find optimum value from *u_{ik}* and *v_i*.

Lagrange function for FGWC differentiates from each parameter and equal to zero for get

the optimum value until get two formulations of objective functions as follows.

$$J_{FGWC}(V; X) = \sum_{i=1}^c \sum_{k=1}^n |v_i - x_k|^2 / \left(\sum_{j=1}^c (|v_i - x_k| / |v_j - x_k|)^{2/m-1} \right)^m \rightarrow \min \dots\dots 2$$

$$J_{FGWC}(U; X) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \left(\sum_{k=1}^n u_{ik}^m x_k / \sum_{k=1}^n u_{ik}^m - x_k \right)^2 \rightarrow \min \dots\dots 3$$

Both of these formulation processes are widely known as alternative optimization (AO), both used to optimize the FGWC model through several extreme conditions of the objective function *J_{FGWC}*. The results of each function formulation are referred as FGWC-U and FGWC-V, while *μ_{ik}* in FGWC-U formula is a membership of the modified geographic cluster.

The method of context-based clustering is a method that gives a focus on data clustering based on special conditions or methods that centralize the original dataset by the specific conditions on its dimensions. Therefore, we found that only a portion of the original dataset with a relationship that matches the conditions specified. For dataset N with attributes *X* = {*X₁*, ..., *X_n*}. Dataset will be classified into cluster C in dimensional space (*X_εR^r*) with *X_k* is k data point and *V_i* is center of i-cluster. Context variables are defined as *Y ∈ X*

$$A: Y \rightarrow [0,1] \dots\dots 4$$

$$y_k \rightarrow f_k = A(y_k) \dots\dots 5$$

Where *f_k* represents the level of relationship between k-data point and i-cluster.

There are three methods in determining context variable such as random matrix with size *n × 1*, using FCM, and calculate the average and standard deviation. In this paper, we used FCM method in determining the context variable [4].

Validity index is used to determine the right number of clusters. In this paper, we used Partition Coefficient (PC), Classification Entropy (CE), and Separation Index (S).

Equation of PC index, CE index, and S index can be described as follows

$$PC = 1/N \sum_{i=1}^c \sum_{j=1}^N \mu_{ij}^2 \dots\dots 6$$

$$CE = -1/N \sum_{i=1}^c \sum_{j=1}^N \mu_{ij} \log_a(\mu_{ij}) \dots\dots 7$$

$$S = \sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^2 \|x_j - v_i\|^2 / N \dots\dots 8$$

Where μ_{ij} is degree of j-data point membership in k-cluster, N is number of data points, v is cluster center, c is number of cluster, and x is data point.

RESULT AND DISCUSSION

Implementation of the FGWC-PSO algorithm using context-based clustering in many data packet analysis requires a number of clusters that can be determined by the researcher. Therefore, researchers tried by using the number of 2 clusters until 5 clusters. The calculation of the CFGWC-PSO algorithm evaluated by using three validity indexes, such as PC index, CE index, and S index, to find out the right number of clustering. The results of the validity index calculation showed in Table 1.

Table 1. Validity Index for Number of Clusters

C	CFGWC-PSO		
	PC Index	CE Index	S Index
2	0.636053	0.545987	0.016137
3	0.462956	0.914898	0.156111
4	0.472868	1.003822	0.154439
5	0.432561	1.141377	0.219087

Sources: Data Processing (2019)

Clustering performance measurement using the PC index, which is the greater the PC index value, indicates better clustering quality. CE index with an interpretation that is the smaller the CE index value is the better quality of clustering. S index has the same interpretation as the CE index. So, from Table 1, we can see that for the three validity index, the one that showed the best clustering is by using the number of clusters = 2. Therefore, for the next analysis, we used the number of clusters = 2. The number of districts/cities included in cluster 2 is more than in cluster 1 (Table 2). Cluster 1 has 11 districts/cities and cluster 2 has 27 districts/cities.

Table 2. Results of Clustering

Cluster 1	Cluster 2	
Ponorogo	Pacitan	Lamongan
Lumajang	Trenggalek	Gresik
Jember	Tulungagung	Sampang
Banyuwangi	Blitar	Pamekasan
Bondowoso	Kediri	Sumenep
Situbondo	Malang	Kota Kediri
Probolinggo	Pasuruan	Kota Blitar
Mojokerto	Sidoarjo	Kota Malang
Jombang	Nganjuk	Kota Probolinggo
Bangkalan	Madiun	Kota Pasuruan
Surabaya	Magetan	Kota Mojokerto
	Ngawi	Kota Madiun
	Bojonegoro	Kota Batu
	Tuban	

Sources: Data Processing (2019)

Table 3. Average Clustering Variable Results

Cluster	X ₁	X ₂	X ₃
1	1.2159908	4.9536975	2.273063
2	4.1691160	4.2481785	3.030779
JATIM	2.6925530	4.6009380	2.6519214
Cluster	X ₄	X ₅	X ₆
1	26.0382	47.1704	2051.83
2	32.4864	42.8726	1828.79
JATIM	29.2623	45.0215	1940.31
Cluster	X ₇	X ₈	X ₉
1	15.816268	3.118848117	0.0190863
2	22.265694	3.658314014	0.0714926
JATIM	19.040981	3.388581066	0.0452894
Cluster	X ₁₀	X ₁₁	
1	5.73963	150.626	
2	11.1804	91.7901	
JATIM	8.46004	121.208	

Sources: Data Processing (2019)

Table 3 showed that cluster 1 has four high-value variables that affect the factors causing the spread of DHF. Whereas cluster 2 has seven high-value variables that influence the causes of the DHF spread. Therefore, policies related to the efforts to control vectors, the handling of DHF focused on areas in cluster 2, especially on the factors that most influence the spread of DHF.

CONCLUSION

In this paper, we suggest that the policies related to efforts to control vectors in handling DHF focused on cluster 2 areas, especially on the factors that have the most influence on the spread of dengue. Problems in each cluster can be used by local governments to develop policies related to vector control strategies for the spread of dengue in East Java Province.

REFERENCES

- [1] Putra, F.H. 2016. Pengelompokan wilayah bencana endemi demam berdarah dengue di Jawa Timur dengan Fuzzy Geographically Weighted Clustering-Particle Swarm Optimization. *Jurnal Aplikasi Komputasi dan Statistika*. 8(1). 27-37.
- [2] East Java Health Office. 2016. Dengue and outbreak incidence in East Java. East Java Health Office. Surabaya.
- [3] Mason, G.A., R.D. Jacobson. 2007. Fuzzy Geographically Weighted Clustering. Proceedings of the 9th International Conference on Geocomputation. 1-7.
- [4] Son, L.H. 2014. Enhancing clustering quality of geodemographic analysis using Context Fuzzy Clustering Type-2 and Particle Swarm Optimization. *J. Appl. Soft Comput.* 22. 566-584.

- [5] Nurmala, N., P. Ayu, 2015. Improvement of Fuzzy Geographically Weighted Clustering-Ant Colony Optimization using Context Based Clustering. *J. ICT Res. Appl.* 11(1). 21-37.
- [6] Central Bureau of Statistics. 2018. East Java in numbers 2018. Central Bureau of Statistics. East Java Province.
- [7] East Java Province Public Health Office. 2018. Profile of Health in East Java 2018. East Java Province Public Health Office. East Java Province.
- [8] Wijayanto, A.W., Purwariantini. 2014. Improvement of Fuzzy Geographically Weighted Clustering Using Particle Swarm Optimization. Proceeding in Information Technology Systems and Inovation (ICITSI). IEEE.