IMPLEMENTATION K-MEDOIDS ALGORITHM FOR DISTRIBUTION MAPPING OF COVID-19 IN SURABAYA

Vivi Mentari Dewi[1], Iis Dewi Ratih[2]
[1,2] Faculty of Vocational, Institut Teknologi Sepuluh Nopember, Surabaya, (INDONESIA)
vivimentari9@gmail.com

Abstract

Surabaya was recorded as the city with the highest active Covid-19 in East Java with 225 cases based on data released by the East Java Province on March 1st, 2021. One step to minimizing the increase of cases is by grouping regions based on the number of existing cases. The previous mapping only displayed data on the status of confirmed patients in each region which was updated daily. So in this research, a mapping of urban villages in Surabaya was carried out which would be included in the cluster based on confirmed cases in treatment, confirmation of recovery, and confirmation of death using thealgorithm K-Medoids. K-Medoids is a clustering algorithm (unsupervised learning) in machine learning, the development of K-Means which is sensitive to outliers. K-Medoids has better clustering performance for large datasets. The results of the analysis showed that the urban villages with the highest number of deaths were Karah and Kutisari with 3 cases. The results of clustering using K-Medoids with an evaluation value of the Davies-Bouldin Index (DBI) of 0.5666 obtained the optimum cluster of 4 clusters. Cluster 1 (confirmed cases of high death) 35 urban villages, clusters 2 and 3 (confirmed cases under treatment, confirmed cases recovered, and confirmed cases of low death) 56 and 49 urban villages, and cluster 4 (confirmed cases under treatment and confirmed cases recovered high) 14 urban villages.

Keywords: Covid-19, Davies-Bouldin Index (DBI), K-Medoids, Machine Learning, Unsupervised Learning

1 INTRODUCTION

Coronavirus Disease 2019 (Covid-19) is a new type of virus that can be transmitted through direct or indirect contact with infected people, surfaces or objects used. Covid-19 cases are spread in all provinces, one of which is East Java, where Surabaya is the capital with the highest active Covid-19 cases in East Java, referring to data released by East Java Province on March 1st, 2021 with 225 active cases (Ulum, 2021). Extra handling is needed to minimize the increase in Covid-19 cases, especially in Surabaya. Steps that can be taken to control the increase in cases are by grouping areas based on existing data. Several studies have conducted clustering of Covid-19 cases in Indonesia (Sindi et al., 2020), but further analysis is needed because there is no research on mapping the distribution of Covid-19 cases in urban villages in Surabaya. The previous mapping on the Surabaya Respond Covid web page only displayed data on the status of confirmed patients in each region which was updated every day. So this research needs to be done to get more detailed information through the formed urban village clusters, as well as to facilitate handling by looking at the characteristics of confirmed cases in treatment, confirmation of recovery, and confirmation of death in each cluster using the clustering method, namely K-Medoids Algorithm. The K-Medoids Algorithm is one of the methods in data mining that is popular among researchers. K-Medoids is one of the clustering algorithms included in machine learning thegroup unsupervised learning which is used to group data into several clusters, where the data is not classed first and then grouped. According to Chrisnanto and Abdillah (2015), the K-Means algorithm is more efficient for smaller data sets, while K-Medoids has better performance for large datasets and K-Medoids also has better performance when compared to K-Means. In addition, according to Kamila, Khairunnisa, and Mustakim (2019) K-Medoids is a development of the K-Means algorithm which is sensitive to outliers in clustering. Research using K-Medoids has been carried out, the results show that themethod K-Medoids is better than K-Means method because the results of the

cluster formed by K-Medoids are more homogeneous than K-Means (Sangga, 2018). In addition, the results of the cluster formed using K-Medoids show a higher validity value than themethod K-Means (Marlina et al., 2018). The benefits obtained after conducting this research are getting urban clusters based on cases in treatment, recovery, and death using a more effective and efficient method, namely K-Medoids Algorithm. The results of the cluster formed are distinguished from urban villages based on Covid-19 cases, then evaluated using the Davies-Bouldin Index (DBI) where the method uses the value of cohesion (closeness of data to centroids in a cluster) and separation (distance between centroids and objects in other clusters) (Sundar, Chitradevi, and Geetharamani, 2012). The optimum clusters obtained are in the form of homogeneous villages in one cluster and heterogeneous between clusters, then visualized through mapping and characteristics of confirmed cases in treatment, confirmation of recovery, and confirmation of death based on clusters formed to determine the potential spread of Covid-19 cases in villages in each cluster.

2 METHODOLOGY

The research methodology used in this study is described as follows.

2.1 Data Source

The data used in this study is secondary data obtained from the Surabaya Response Covid-19 website (https://lawancovid-19.surabaya.go.id/). The data is data that is*updated* every day regarding the addition and reduction of cases by taking data for March 2021 in 154 urban villages in the city of Surabaya.

2.2 Variables

Variabel in the form of a dataset on the number of Covid-19 cases that used in this research described in Table 1 below.

Table 1. Data Structure of Covid-19 Cases for each Urban Villages in Surabaya.

Number of Urban Villages	X_1	X_2	X_3
1	$X_{1.1}$	$X_{1.2}$	$X_{1.3}$
2	$X_{2.1}$	$X_{2.2}$	$X_{2.3}$
3	$X_{3.1}$	$X_{3.2}$	$X_{3.3}$
⋮	⋮	⋮	⋮
154	$X_{154.1}$	$X_{154.2}$	$X_{154.3}$

The operational definition for each research variable based on the Decree of the Minister of Health Number HK.01.07/Menkes/247/2020 concerning Guidelines for the Prevention and Control of Covid-19 is explained as follows.

a) Under treatment case or confirmation (X_1) is defined as a person who is tested positive for Covid-19 as evidenced by an RT-PCR laboratory examination that requires hospitalization.
b) Recovered case (X_2) is defined as a person who has been treated and is declared clean from Covid-19 infection as evidenced by an RT-PCR laboratory examination.
c) Death case (X_3) is defined as a confirmed person in Covid-19 treatment who dies.

2.3 Analysis Step

The explanation for each research variable is explained as follows.

a) Determine the number of k clusters
b) Select objects k randomly as medoids
c) Calculate the distance of non-medoids with the medoids initialon all variables using euclidean distance
d) Put non-medoids objects in the cluster with the minimum distance
e) Calculate the total cost initial
f) Repeat steps 2 to 5 until aobtained total cost is new

g) Calculate the difference among the total cost of the new totalcost initial If S <0 repeat steps 2 to 7, but stop if S> 0
h) Evaluating the results of the cluster is formed using the Davies-Bouldin Index (DBI)

3 RESULTS

Result of clustering the number of Covid-19 cases of urban villages in Surabaya using K-Medoids Algorithm described below.

3.1 K-Medoids Algorithm

3.1.1 Cluster Center

The cluster center shows the initial medoids for each variable in a cluster described in Table 2 below.

Table 2. Initial Cluster Center.

Variables	Cluster			
	1	2	3	4
Confirmed Case of Under Treatment (X_1)	14	3	8	24
Confirmed Case of Recovered (X_2)	13	3	8	24
Confirmed Case of Death (X_3)	0	0	0	0

The cluster center or medoids is a cluster representative point, where the number of clusters formed using K-Medoids algorithm is 4 clusters. The urban villages selected as medoids are Sawunggaling, Kapasari, Made, and Babatan. Medoids in the first cluster is Sawunggaling Urban Village by the number of confirmed cases of under treatment at 14, confirmed cases of recovered at 13, and confirmed cases of death at 0. Medoids in the second cluster is Kapasari Urban Village by the number of confirmed cases of under treatment at 3, confirmed cases of recovered at 3, and confirmed cases of death at 0. Medoids in the third cluster is Made Urban Village by the number of confirmed cases of under treatment at 8, confirmed cases of recovered at 8, and confirmed cases of death at 0. Medoids in the fourth cluster is Babatan Urban Village by the number of confirmed cases of under treatment at 24, confirmed cases of recovered at 25, and confirmed cases of death at 0.

3.1.2 Member of Cluster

Member of Clustering using K-Medoids Algorithm described in Table 3 below.

Table 3. Member of Cluster.

Klaster	Number of Cluster Member
1	35
2	56
3	49
4	14

The results of clustering using K-Medoids algorithm were found in cluster 1 as many as 35 urban villages, in cluster 2 as many as 56 urban villages, in cluster 3 as many as 49 urban villages, and in cluster 4 as many as 14 urban villages.

3.2 Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) is an internal evaluation used to determine the most optimal number of clusters of clustering using K-Medoids algorithm that described in Table 4 below.

Table 4. Davies-Bouldin Index (DBI).

Number of Cluster	DBI
3	0.8082

Table 5. Davies-Bouldin Index (DBI) (Continuation).

Number of Cluster	DBI
4	0.5666
5	0.5896

Value of Davies-Bouldin Index (DBI) by using the number of clusters as much as 3 to 5, in which the clusters formed said the optimum cluster if the DBI value is positive and close to zero. So it can be concluded that the optimum number of clusters for K-Medoids algorithm using of 4 clusters with a DBI value of 0.5666.

3.3 Mapping for Cluster Result of K-Medoids

Visualization of clustering the number of Covid-19 cases of urban villages in Surabaya using K-Medoids Algorithm described in Figure 1 below.

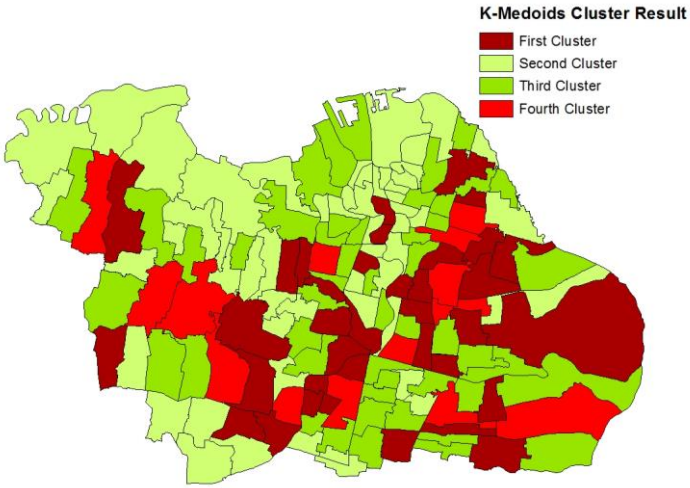

Figure 1. Mapping of K-Medoids.

Description for 4 cluster using boxplot to find out the character of each cluster according to confirmed cases under treatment, confirmed cases recovered, and confirmed cases died described in Figure 2 below.

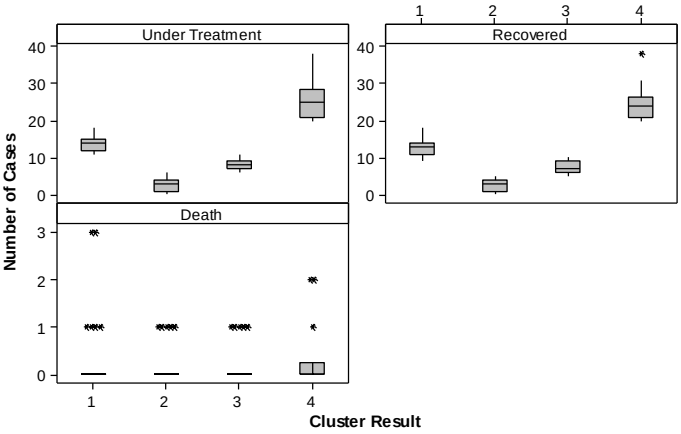

Figure 2. Boxplot of Covid-19 Cases Cluster Result.

The boxplot of cluster results using K-Medoids algorithm shows that there are 4 clusters formed with a large diversity of data on confirmed cases in treatment, confirmed cases recovered, confirmed cases died in the fourth cluster is quite large. This can be shown through the width of the boxplot in each cluster which is quite different. Moreover, the fourth cluster has awidth boxplot that is very different from other clusters, where based on the boxplot it can be seen that the fourth cluster is an urban village with high Covid-19 cases by looking at confirmed cases in treatment and confirmed cases recovering. While the 1st cluster is a urban village with high confirmed cases of death. The second and third clusters are urban village with a low number of confirmed cases under treatment, confirmed cases recovered, and confirmed cases death. Considering that the first dan fourth clusters are areas with high Covid-19 cases, the Surabaya City Government should focus more on handling urban villages in these clusters in order to minimize the increase in Covid-19 cases, especially confirmed cases in treatment and confirmed cases of death.

4 CONCLUSIONS

The results of the clustering using *K-Medoids* algorithm and cluster evaluation with a DBI of 0.5666 obtained the optimum cluster formed as many as 4 clusters, where members in cluster 1 (urban villages with confirmed cases of high mortality) were 35 urban villages, in cluster 2 (urban villages with the number of confirmed cases under treatment, confirmed cases recovered, and confirmed cases of low death) as many as 56 urban villages, in cluster 3 (urban villages with confirmed cases under treatment, confirmed cases recovered, and confirmed cases low death) as many as 49 urban villages, and in cluster 4 (urban villages with confirmed confirmed cases of high under treatment and confirmed cases recovery) as many as 14 urban villages.

ACKNOWLEDGEMENTS

Thanks to my self for doing all this hard work and never quitting, also my lecturer in Business Statistics Department, Institut Teknologi Sepuluh Nopember who has guide author during preparation of this paper.

REFERENCES

[1] Ulum, M. (2021) 'Covid-19 cases in Surabaya are again the highest in East Java', *Bisnis.com*, 2nd March. Available at: https://surabaya.bisnis.com/read/20210302/531/1362583/case-covid-19-di-surabaya-re-tertinggi-di-jawa-timur.

[2] Sindi, S., Ningse WRO, Sihombing, IA, Ilmi, F., and Hartama, D. (2020) 'K-Medoids Clustering Algorithm Analysis in Grouping the Spread of Covid-19 in Indonesia', *Jti (Journal of Information Technology)*, 4(1), pp. 166–173.

[3] Chrisnanto, YH and Abdillah, G. (2015) 'Application of Partitioning Around Medoids (PAM) Clustering Algorithm to See an Overview of Capabilities', *National Seminar on Information and Communication Technology*, 2015(Sentika), pp. 444–448.

[4] Kamila, I., Khairunnisa, U., and Mustakim, M. (2019) 'Comparison of K-Means and K-Medoids Algorithms for Grouping Loading and Unloading Transaction Data in Riau Province', *Scientific Journal of Engineering and Management of Information Systems*, 5(1), p. 119. doi:10.24014/rmsi.v5i1.7381.

[5] Sangga, VAP (2018) 'Comparison of the K-Means Algorithm and the K-Medoids Algorithm in Grouping Livestock Commodities in Central Java Province in 2015', Project *Finalof Statistics Department, Faculty of Mathematics and Natural Sciences, Islamic University of IndonesiaYogyakarta*,, 53(9), pp . 1689–1699.

[6] Marlina, D., Putri, NF, Fernando, A., and Ramadan,*A.* (2018) 'Implementation of K-Medoids and K-Means Algorithms for Grouping of Disability Distribution Areas in Children', *Journal of CoreIT: Journal of Computer Science and Information Technology Research Results*, 4(2), p. 64. doi:10.24014/coreit.v4i2.4498.

[7] Sundar, C., Chitradevi, M., and Geetharamani, G. (2012) 'AN ANALYSIS ON THE PERFORMANCE OF K-MEANS CLUSTERING ALGORITHM FOR CARDIOTOCOGRAM DATA CLUSTERING', *International Journal on Computational Sciences & Applications (IJCSA)*, 2, pp . 11–20.

[8] Ministry of Health RI (2020) 'Decree of the Minister of Health of the Republic of Indonesia Number HK.01.07/MenKes/413/2020 concerning Guidelines for Prevention and Control of Corona Virus Disease 2019 (Covid-19)', *MenKes/413/2020*, 2019, pp. 1–207. Available at: https://covid19.go.id/storage/app/media/Regulation/KMK No. HK.01.07-MENKES-413-2020 on Guidelines for Prevention and Control of COVID-19.pdf.