

# Clustering Regency and City in East Java Based on Population Density and Cumulative Confirmed COVID-19 Cases

Khusnia Nurul Khikmah<sup>1\*</sup> and A'yunin Sofro<sup>2</sup>

<sup>1-2</sup>Mathematics Department, Faculty of Mathematics and Natural Sciences, Surabaya State University  
Jln. Ketintang, Jawa Timur 60213, Indonesia  
<sup>1</sup>khusniank@gmail.com; <sup>2</sup>ayuninsofro@unesa.ac.id

Received: 27<sup>th</sup> December 2020/ Revised: 22<sup>nd</sup> March 2021/ Accepted: 22<sup>nd</sup> March 2021

**How to Cite:** Khikmah, K. N., & Sofro, A. (2021). Clustering Regency and City in East Java Based on Population Density and Cumulative Confirmed COVID-19 Cases. *ComTech: Computer, Mathematics and Engineering Applications*, 12(2), 111-121. <https://doi.org/10.21512/comtech.v12i2.6891>

**Abstract** - Coronavirus is a big family of viruses that causes acute respiratory syndrome and mediates human-to-human by the environment. A factor that affects the spread of infectious diseases is population density. Therefore, it is necessary to study the effect of population density on infectious diseases like COVID-19. The research analyzed the effect of the population density of each regency in East Java on cumulative confirmed COVID-19 cases until December 9, 2020. The research applied quantitative method using the agglomerative hierarchical clustering method. The clustering method included single, average, and complete linkages. The results of clustering using single linkage and average linkages have the same results for the population density of Jember Regency. This regency has the lowest effect for the cumulative confirmed COVID-19 cases. Then, complete linkage obtains that Banyuwangi Regency and Gresik Regency has the population density with the lowest effect for the cumulative number of confirmed COVID-19 cases. The results of clustering with single, average, and complete linkages have the same results for population density with a big effect on the cumulative number of confirmed COVID-19 cases in Surabaya City. The results of best clustering regencies or cities that population density affects the number of confirmed cases of COVID-19 use a single linkage.

**Keywords:** clustering regency, clustering city, population density, confirmed cases, COVID-19

## I. INTRODUCTION

Coronavirus is a big family of viruses that cause disease in humans and animals as it has spread

globally. In December 2019, a new type of coronavirus was discovered in Wuhan, China. It was named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), which caused Coronavirus Disease-2019 (COVID-19) (Hui *et al.*, 2020). The incubation period of COVID-19 is estimated to be 14 days, with most cases occurring around four to five days after exposure (Cheng *et al.*, 2020). Moreover, according to the Indonesia Ministry of Health, there is no age limit for people who can be infected with COVID-19. In addition, Chinese's Centers for Disease and Prevention reports that around 44.500 cases of COVID-19 are confirmed, with 87% aged between 30 and 79 years (Wu & McGoogan, 2020).

COVID-19 can spread through small droplets from the nose or mouth when coughing or sneezing (Kementerian Kesehatan Republik Indonesia, 2020). It mediates human-to-human transmission by the environment (Sajadi *et al.*, 2020). Two factors help the spread of infectious diseases like COVID-19. Those factors are population density and the ability to transmit the disease itself (Merler & Ajelli, 2010).

Population density is the number of people per unit area (Nelwan, 2020). Based on data on the official website of the Indonesian Central Statistics Agency in 2020 (Badan Pusat Statistik, 2020), the population in East Java, Indonesia in 2020 was 39.886.288. It ranked sixth as the province with the highest population density in Indonesia, namely 847 people/km<sup>2</sup>. Then, according to Kusuma and Sukendra (2016), Neiderud (2015), and Sihombing, Marsaulina, and Ashar (2014), infectious diseases will be easily and quickly transmitted in areas with high population density. Indonesian COVID-19 Response Acceleration Task Force (Gugus Tugas Percepatan Penanganan COVID-19) recorded that on December 9, 2020, the cumulative number of confirmed cases in

East Java was 66.868. It was in the second position of COVID-19 cases nationally (Satuan Tugas Penanganan COVID-19, 2020). It shows that the COVID-19 cases require more attention so that the number of confirmed cases of COVID-19 in Indonesia decreases. One of the steps to do it is a clustering analysis.

Clustering analysis is a technique that aims to classify objects based on almost the same characteristics in statistical analysis (Bateni *et al.*, 2017). It identifies homogeneous objects into one cluster (Chuan *et al.*, 2018). It has two methods. The first method is the hierarchical method, which contains five agglomerative: single-linkage (Ghebreslassie, Githiri, Mehari, & Kasili, 2015; Yaroslavtsev & Vadapalli, 2018), average linkage (Ghoshdastidar, Perrot, & Von Luxburg, 2019), complete linkage (Ghoshdastidar *et al.*, 2019; Mu'afa & Ulinuha, 2019), ward method (Abboud, Cohen-Addad, & Houdroug , 2019; Majerova & Nevima, 2017), and centroid method (Setiawan, Djanali, & Ahmad, 2017). Clustering using a hierarchical method is the same as another multivariate technique. Clustering analysis with the hierarchical method is an important method that the data sets are recursively grouped into sequential clusters (Roy & Pokutta, 2017). The data to be clustered on the matrix calculate the Euclidean distance. The second method is the non-hierarchical method (Govender & Sivakumar, 2020; Nugroho, 2008).

According to several previous researchers, hierarchical methods are also used in several fields and cases. For example, it is applied in phylogenetics and taxonomy (Felsenstein, 2004; Sneath & Sokal, 1973). It can also detect trajectory anomalies and behavior patterns of GPS data in a taxi (Wang, Qin, Chen, & Zhao, 2018). Then, it is used to model students' online learning activities (Triayudi & Fitri, 2019). Clustering using a hierarchical method with other agglomerative is also shown in several studies by using bisecting K-means (Moseley & Wang, 2017), Nearest Neighbor Chains (Fahim, 2017; Murtagh & Contreras, 2012), complete link divisive (Roux, 2018), average linkage under Gaussian Kernel (Charikar, Chatziafratis, Niazadeh, & Yaroslavtsev, 2019), and spreading metrics (Roy & Pokutta, 2017).

The research carries out clustering using a hierarchical method, namely single, average, and complete linkages with case studies of population density and cumulative confirmed cases of COVID-19 (till December 9, 2020) in East Java from each regency. The results of clustering and clustering validation are expected to provide information to assist the authorities. So, they can prevent the spread of COVID-19 in East Java based on population density with the number of confirmed positive cases of Covid in each regency or city in East Java.

## II. METHODS

The research is conducted out by studying literature. At this stage, the literature study is carried

out by looking for reference materials in the form of books, journals, final assignments, theses, and the Internet in accordance with the discussed issues. Then, data collection uses secondary data from the official website of the Indonesian Central Statistics Agency (BPS) and Jatim Tanggap COVID-19. The data are population density and cumulative confirmed cases of COVID-19 (till December 9, 2020) in East Java from each regency. Next, the research sets the variable of the study. The variables are  $V_1$  and  $V_2$ . The  $V_1$  is East Java's population density from each regency and city. Meanwhile,  $V_2$  is cumulative confirmed COVID-19 cases in East Java from each regency and city. After that, the data are analyzed.

The data analysis is categorized as quantitative type, and the clustering is divided into several stages. The first stage is data normalization. In data normalization, the data to be calculated are in smaller intervals than the original data. The second stage is calculating the distance between objects using Euclidean distance. The results of the calculation with the Euclidean distance show the proximity matrix. It is a square and symmetry matrix with the same number of objects in the rows and columns. It produces a square and symmetrical matrix, namely the  $n \times n$  matrix. The third stage is clustering using single, average, and complete linkages. The clustering results will be calculated based on hierarchical agglomerative with the corresponding objects forming new clusters according to the method. The last stage is clustering validation by calculating the agglomerative coefficient of hierarchical clustering. The results of clustering validation with three methods are obtained from the results of the agglomerative coefficient to determine which method is the best or fits the strictness of the clustering results.

Normalization data processes common data distribution with the aim of normalization. One of the best methods of normalization is Min-Max normalization. Min-Max normalization maps the value from each variable into the same range. Normalization can be calculated by Equation (1) with  $x$  as data per column,  $min$  as the minimum value of data per column, and  $max$  as the maximum value of data per column (Ali & Faraj, 2014).

$$x' = \frac{x - min}{max - min} \quad (1)$$

Clustering with the hierarchical method can be obtained by pairing the nearest cluster with the closest distance. Euclidean distance can be obtained by Euclidean space in two dimensions, three dimensions, and others (Cohen-Addad, Kanade, Mallmann-Trenn, & Mathieu, 2019). It can determine two objects to be said as similar or have the closest distance by calculating the distance between objects with the Euclidean distance equation (see Equation (2)). The  $d(x, y)$  is the distance between  $x$  and  $y$ . Then,  $i$  is each datum,  $z$  is the total of data,  $x_{ik}$  is the center of cluster's

data, and  $y_{jk}$  is data on each  $jk$ -th data (Charikar *et al.*, 2019; Nishom, 2019).

$$d(x, y) = \sqrt{\sum_{i=1}^s (x_{ik} - y_{jk})^2} \quad (2)$$

Clustering analysis with hierarchical method has hierarchical agglomerative for grouping  $N$  object. It has general stages. The first stage is started by  $N$  cluster. There is  $N \times N$  matrix with the distance of  $D$ ,  $D = \{d_{ik}\}$  where each cluster contains a single entity and a symmetrical matrix. The second stage is finding the distance matrix for pairing the nearest cluster. If  $U$  and  $V$  objects are the closest pair of clusters,  $U$  and  $V$  are chosen. So, it will be  $D = \{d_{uv}\}$ . The third stage is combining  $U$  and  $V$  clusters into a new cluster ( $UV$ ), updating entries of matrix by deleting rows and columns that correspond to  $U$  and  $V$  cluster, and adding rows and columns by providing the distance value between new cluster ( $UV$ ) and all of the residual cluster. The last stage is repeating the steps of second stage for  $N-1$  times. At the end of the algorithm, all objects will be in one cluster and note the identity of the merged cluster and the rate at which the merger occurs (Hartini, 2014).

In the research, the first agglomerative of the hierarchical method is single linkage. The base of clustering with a single linkage is the smallest distance between two objects that will become a new cluster, and so on. The research finds the smallest distance in  $D = \{d_{ik}\}$  and combines the corresponding objects to get a new cluster. For example,  $U$  and  $V$  are the corresponding objects. Then, the new cluster is ( $UV$ ). Then, by using third stage of hierarchical agglomerative, the distance between ( $UV$ ) and  $W$  clusters with  $d_{vw}$  and  $d_{uw}$  is the closest distance between components of  $U$  and  $W$  clusters and  $V$  and  $W$  clusters (Mu'afa & Ulinuha, 2019; Ros & Guillaume, 2019). It is shown in Equation (3).

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (3)$$

The second agglomerative of the hierarchical method is average linkage. Clustering with average linkage uses the average distance between observations. It looks for pairs of observations with a distance that is the closest to the average distance. Then, it finds the smallest distance in  $D = \{d_{ik}\}$  and combines the same objects. If  $U$  and  $V$  form a new cluster, namely ( $UV$ ), with third stage of hierarchical agglomerative, the distance between ( $UV$ ) and another  $W$  cluster with  $d_{ik}$  is the distance between  $i$ -th object in ( $UV$ ) cluster (Moseley & Wang, 2017; Mu'afa & Ulinuha, 2019). It is shown in Equation (4).

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)} N_W} \quad (4)$$

The last agglomerative of the hierarchical method in the research is complete linkage. The base of clustering with complete linkage is the farthest distance between two objects that will become the new cluster, and so on (Balcan & White, 2017). It finds the smallest distance in  $D = \{d_{ik}\}$  and combines the corresponding objects to get a new cluster. For example,  $U$  and  $V$  are the corresponding objects, and the new cluster is ( $UV$ ). Then, by using third stage of hierarchical agglomerative, the distance between ( $UV$ ) and  $W$  clusters with  $d_{vw}$  and  $d_{uw}$  is the farthest distance between components of  $U$  and  $W$  clusters and  $V$  and  $W$  clusters (Grosswendt & Roeglin, 2017; Mu'afa & Ulinuha, 2019). It is shown in Equation (5).

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \quad (5)$$

The measure of clustering structure is called the agglomerative coefficient. This measure usually finds the dissimilarity value of clustering. The tight clustering of an object is interpreted by low values of the agglomerative coefficient. Then, a less well-formed cluster is shown by high values of agglomerative coefficient. The agglomerative coefficient, in general, describes the strength of the clustering structure. The agglomerative coefficient can be obtained with  $i$  as each object and  $l(i)$  as the length of range (see Equation (6)) (Fairuzi & Hamidah, 2016).

$$AC = \sum_i l(i) \quad (6)$$

### III. RESULTS AND DISCUSSIONS

Two data are used in the research. There are data on East Java's population density from each regency or city in 2020 and cumulative confirmed cases of COVID-19 from each regency or city till December 9, 2020. It is shown in Table 1 (see Appendices).

The first stage of clustering with hierarchical is normalization data. It is carried out by using Min-Max normalization. Therefore, the data of population density of regencies and cities in East Java in 2020 with cumulative confirmed cases of COVID-19 regencies and cities in East Java till December 9, 2020, become new data with smaller intervals. The data has been normalized, as shown in Table 2 (see Appendices).

The second stage is clustering with single, average, and complete linkages. It can be analyzed by forming clusters from the matrix distance between objects, namely data on the population density of regencies and cities in East Java in 2020 with cumulative confirmed cases of COVID-19 in regencies and cities in East Java till December 9, 2020. This distance can be calculated using the Euclidean distance in Equation (2). Then, the result of Euclidean distance will produce a square and symmetrical matrix, namely the  $n \times n$  matrix. The data of the population density of regencies and cities in East Java in 2020

with cumulative confirmed cases of COVID-19 in regencies and cities in East Java until December 9, 2020, are 38. The resulting matrix is a  $38 \times 38$  matrix. The matrix is shown in Table 3 (see Appendices).

The single linkage is obtained from finding the smallest distance matrix for pairing the nearest cluster. Then, the calculation of the closest distance between two objects is based on the Euclidean matrix distance. The closest distance between the first two objects is obtained from the calculation results in Equation (3). Using Equation (3), it gets a new cluster, and the agglomerative will be repeated until it gets the desired number of clusters. In the research, the desired cluster is 10. The member result of a single linkage cluster is obtained. Table 4 (see Appendices) shows the cluster members with a single linkage of each city and regency with cumulative confirmed COVID-19 cases in Indonesia. Then, Figure 1 illustrates a visualization of the cluster division with cluster dendrogram.

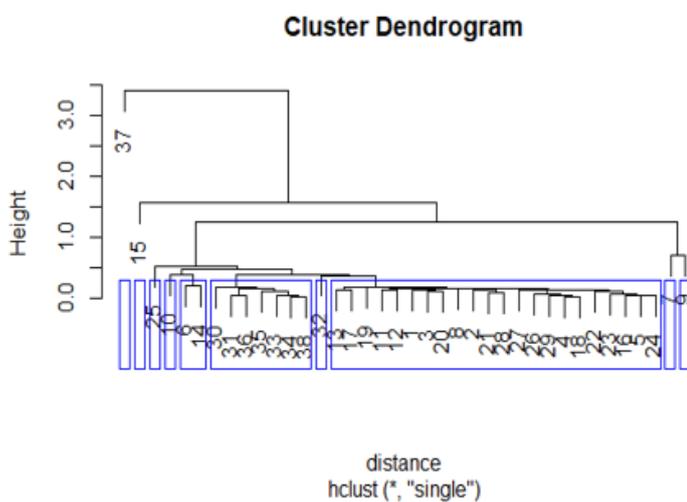


Figure 1 Visualization of Single Linkage with the Smallest Distance Using Cluster Dendrogram

The average linkage method can be calculated using Equation (4). It finds the average distance matrix by pairing the nearest cluster. Then, it calculates the closest distance between two objects based on the Euclidean matrix distance. The closest distance is between the first two objects. From Equation (4), it will get a new cluster, and the agglomerative will be repeated until it has the desired number of clusters. In the research, the desired cluster is 10. The cluster members with the average linkage of each city and regency with cumulative confirmed COVID-19 cases in Indonesia are shown in Table 5 (see Appendices). Meanwhile, Figure 2 illustrates a visualization of the cluster division using cluster dendrogram.

The complete linkage method finds the farthest distance matrix by pairing the nearest cluster. Then, it calculates the closest distance between two objects based on the distance of the Euclidean matrix. The closest distance between the first two objects is

obtained from the calculation results in Equation (5). Using Equation (5), it gets a new cluster, and the agglomerative will be repeated until it obtains the desired number of clusters. In the research, the desired cluster is 10. The cluster members with the complete linkage of each city and regency with cumulative confirmed COVID-19 cases in Indonesia are shown in Table 6 (see Appendices). Then, Figure 3 shows a visualization of the cluster division with cluster dendrogram.

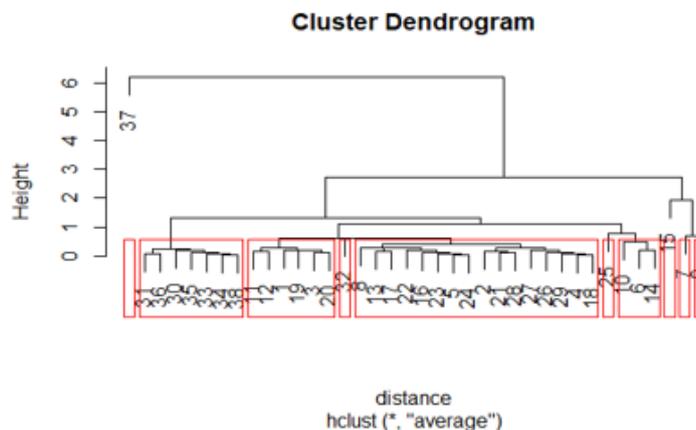


Figure 2 Visualization of Average Linkage with the Average Distance Using Cluster Dendrogram

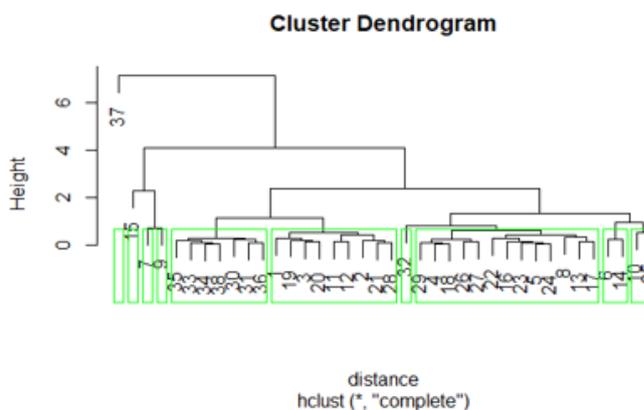


Figure 3 Visualization of Complete Linkage with the Farthest Distance Using Cluster Dendrogram

The results show that single and average linkages have the same results for regencies or cities. The population density has the lowest and highest effect on the number of confirmed positive cases of COVID-19. For example, Jember Regency has the population density with the lowest effect on the number of confirmed positive COVID-19 cases. Then, Surabaya City have the population density with the highest effect on the number of confirmed COVID-19 cases. Meanwhile, the result of complete linkage indicates that the population density of Banyuwangi Regency and Gresik Regency has the lowest effect for

the cumulative confirmed COVID-19 cases. It also mentions that Surabaya City has a population density with the highest effect for the cumulative confirmed COVID-19 cases. In short, the results of clustering with single, average, and complete linkage have the same results for population density with a big effect on the cumulative number of confirmed COVID-19 cases in Surabaya City.

The validation of clustering results using three methods, namely single, average, and complete linkages, is obtained from the results of the agglomerative coefficient. It determines which method has the best clustering tightness results. The agglomerative coefficient is computed by Rstudio. The agglomerative coefficient values show 0,9159327 for single linkage, 0,9360739 for average linkage, and 0,9385259 for complete linkage. From the results, a single linkage has the lowest value of agglomerative. It means the object that is clustered with a single linkage method reflects tight clustering.

#### IV. CONCLUSIONS

There are several results based on clustering analysis using the hierarchical method (single, average, and complete linkages) with 10 clusters. First, the clustering results using single and average linkages have the same results for the population density in Jember Regency. This regency has the lowest effect for the cumulative confirmed COVID-19 cases. Second, complete linkage shows that Banyuwangi and Gresik Regencies have population density with the lowest effect for the cumulative number of confirmed COVID-19 cases. Third, the results of clustering with single, average, and complete linkages have the same results. The population density with a big effect on the cumulative number of confirmed COVID-19 cases is in Surabaya City. Fourth, based on the agglomerative coefficient value, the results of best clustering regencies or cities use a single linkage. The agglomerative coefficient value shows tight clustering.

The research is only limited to the total population density and confirmed positive COVID-19 cases in East Java. The data are clustered to get information on which regencies or cities that the number of confirmed positive COVID-19 is influenced by population density for further action. So, the number of confirmed positive cases of COVID-19 does not continue to increase. Moreover, this clustering research only uses three methods of agglomerative hierarchical clustering with Euclidean distance. Hence, further research on clustering can apply other agglomerative hierarchical clustering using Manhattan distance and ward and centroid methods.

#### REFERENCES

Abboud, A., Cohen-Addad, V., & Houdrougé, H. (2019). Subquadratic high-dimensional hierarchical clustering. *Advances in Neural Information Processing Systems*, 32, 11580-11590.

Ali, P. J. M., & Faraj, R. H. (2014). Data normalization and standardization: A technical report. *Machine Learning Technical Reports*, 1(1), 1-6.

Badan Pusat Statistik. (2020). *Beranda*. Retrieved from <https://www.bps.go.id/>

Balcan, M. F., & White, C. (2017). Clustering under local stability: Bridging the gap between worst-case and beyond worst-case analysis. *arXiv preprint arXiv:1705.07157*.

Batani, M. H., Behnezhad, S., Derakhshan, M., Hajiaghayi, M. T., Kiveris, R., Lattanzi, S., & Mirrokni, V. (2017). Affinity clustering: Hierarchical clustering at scale. In *31<sup>st</sup> Conference on Neural Information Processing Systems (NIPS 2017)* (pp. 1-11).

Charikar, M., Chatziafratis, V., Niazadeh, R., & Yaroslavtsev, G. (2019). Hierarchical clustering for euclidean data. In *The 22<sup>nd</sup> International Conference on Artificial Intelligence and Statistics* (pp. 2721-2730). PMLR.

Cheng, H. Y., Jian, S. W., Liu, D. P., Ng, T. C., Huang, W. T., & Lin, H. H. (2020). Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. *JAMA internal medicine*, 180(9), 1156-1163.

Chuan, Z. L., Ismail, N., Shinyie, W. L., Ken, T. L., Fam, S. F., Senawi, A., & Yusoff, W. N. S. W. (2018). The efficiency of average linkage hierarchical clustering algorithm associated multi-scale bootstrap resampling in identifying homogeneous precipitation catchments. *IOP Conference Series: Materials Science and Engineering*, 342, 1-10.

Cohen-Addad, V., Kanade, V., Mallmann-Trenn, F., & Mathieu, C. (2019). Hierarchical clustering: Objective functions and algorithms. *Journal of the ACM (JACM)*, 66(4), 1-42.

Fahim, A. (2017). A clustering algorithm based on local density of points. *International Journal of Modern Education and Computer Science*, 9(12), 9-16.

Fairuzi, N., & Hamidah, H. P. (2016). *Analisis hubungan kekerabatan curcuma spp. berdasarkan karakter morfologi dan metabolit sekunder* (Thesis). Universitas Airlangga.

Felsenstein, J. (2004). *Inferring phylogenies*. Sinauer Associates, Inc.

Ghebresslassie, B. M., Githiri, S. M., Mehari, T., & Kasili, R. W. (2015). Analysis of diversity among potato accessions grown in Eritrea using single linkage clustering. *American Journal of Plant Sciences*, 6, 2122-2127.

Ghoshdastidar, D., Perrot, M., & Von Luxburg, U. (2019). Foundations of comparison-based hierarchical clustering. *Advances in Neural Information Processing Systems*, 32, 7456-7466.

Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, 11(1), 40-56.

Grosswendt, A., & Roeglin, H. (2017). Improved analysis of complete-linkage clustering. *Algorithmica*, 78(4), 1131-1150.

Hartini, E. (2014). *Metode clustering hirarki*. Pusat

- Pengembangan Teknologi Informasi dan Komputasi BATAN.
- Hui, D. S., Azhar, E. I., Madani, T. A., Ntoumi, F., Kock, R., Dar, O., ... & Petersen, E. (2020). The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases*, *91*, 264-266.
- Kementerian Kesehatan Republik Indonesia. (2020). *FAQ*. <https://www.kemkes.go.id/folder/view/full-content/structure-faq.html>
- Kusuma, A. P., & Sukendra, D. M. (2016). Analisis spasial kejadian demam berdarah dengue berdasarkan kepadatan penduduk. *Unnes Journal of Public Health*, *5*(1), 48-56.
- Majerova, I., & Nevima, J. (2017). The measurement of human development using the Ward method of cluster analysis. *Journal of International Studies*, *10*(2), 239-257.
- Merler, S., & Ajelli, M. (2010). The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1681), 557-565.
- Moseley, B., & Wang, J. R. (2017). Approximation bounds for hierarchical clustering: Average linkage, bisecting k-means, and local search. In *Proceedings of the 31<sup>st</sup> International Conference on Neural Information Processing Systems*, (pp. 3097-3106).
- Mu'afa, S. F., & Ulinuha, N. (2019). Perbandingan metode single linkage, complete linkage dan average linkage dalam pengelompokan kecamatan berdasarkan variabel jenis ternak Kabupaten Sidoarjo. *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, *4*(2), 1-5.
- Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(1), 86-97.
- Neiderud, C. J. (2015). How urbanization affects the epidemiology of emerging infectious diseases. *Infection Ecology & Epidemiology*, *5*(1), 1-9.
- Nelwan, J. E. (2020). Kejadian Corona Virus Disease 2019 berdasarkan kepadatan penduduk dan ketinggian tempat per wilayah kecamatan. *Indonesian Journal of Public Health and Community Medicine*, *1*(2), 039-045.
- Nishom, M. (2019). Perbandingan akurasi Euclidean distance, Minkowski distance, dan Manhattan distance pada algoritma K-means clustering berbasis chi-square. *Jurnal Informatika*, *04*(01), 20-24.
- Nugroho, S. (2008). *Statistika multivariat terapan*. Bengkulu: UNIB Press.
- Ros, F., & Guillaume, S. (2019). A hierarchical clustering algorithm and an improvement of the single linkage criterion to deal with noise. *Expert Systems with Applications*, *128*, 96-108.
- Roux, M. (2018). A comparative study of divisive and agglomerative hierarchical clustering algorithms. *Journal of Classification*, *35*(2), 345-366.
- Roy, A., & Pokutta, S. (2017). Hierarchical clustering via spreading metrics. *The Journal of Machine Learning Research*, *18*(1), 3077-3111.
- Sajadi, M. M., Habibzadeh, P., Vintzileos, A., Shokouhi, S., Miralles-Wilhelm, F., & Amoroso, A. (2020). Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of Coronavirus Disease 2019 (COVID-19). *JAMA Network Open*, *3*(6), 1-11.
- Satuan Tugas Penanganan COVID-19. (2020). *Beranda*. <https://covid19.go.id>
- Setiawan, B., Djanali, S., & Ahmad, T. (2017). A study on intrusion detection using centroid-based classification. *Procedia Computer Science*, *124*, 672-681.
- Sihombing, G. F., Marsaulina, I., & Ashar, T. (2014). Hubungan curah hujan, suhu udara, kelembaban udara, kepadatan penduduk dan luas lahan pemukiman dengan kejadian demam berdarah dengue di Kota Malang periode tahun 2002-2011. *Lingkungan dan Keselamatan Kerja*, *3*(1), 1-9.
- Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy: The principles and practice of numerical classification*. W H Freeman & Co.
- Triayudi, A., & Fitri, I. (2019). A new agglomerative hierarchical clustering to model student activity in online learning. *Telkomnika*, *17*(3), 1226-1235.
- Wang, Y., Qin, K., Chen, Y., & Zhao, P. (2018). Detecting anomalous trajectories and behavior patterns using hierarchical clustering from taxi GPS data. *ISPRS International Journal of Geo-Information*, *7*(1), 1-20.
- Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the Coronavirus Disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*, *323*(13), 1239-1242.
- Yaroslavtsev, G., & Vadapalli, A. (2018). Massively parallel algorithms and hardness for single-linkage clustering under  $\ell_p$ -distances. *Proceedings of the 35<sup>th</sup> International Conference on Machine Learning* (pp. 5600-5609).

## APPENDICES

Table 1 Data of Population Density in 2020 and Cumulative Confirmed COVID-19 Cases (till December 9, 2020) from Each Regency or City in East Java

No	Regencies and Cities in East Java	Population Density	Confirmed Cases
1.	Pacitan Regency	555.984	428
2.	Ponorogo Regency	871.825	883
3.	Trenggalek Regency	697.600	817
4.	Tulungagung Regency	1.043.182	772
5.	Blitar Regency	1.163.789	1.184
6.	Kediri Regency	1.580.092	1.363
7.	Malang Regency	2.619.975	1.249
8.	Lumajang Regency	1.044.718	1.562
9.	Jember Regency	2.459.890	3.188
10.	Banyuwangi Regency	1.617.814	3.059
11.	Bondowoso Regency	778.789	1.160
12.	Situbondo Regency	685.776	1.273
13.	Probolinggo Regency	1.174.890	1.766
14.	Pasuruan Regency	1.637.682	1.929
15.	Sidoarjo Regency	2.282.215	7.689
16.	Mojokerto Regency	1.126.392	1.251
17.	Jombang Regency	1.268.504	1.688
18.	Nganjuk Regency	1.057.011	854
19.	Madiun Regency	683.784	240
20.	Magetan Regency	629.020	709
21.	Ngawi Regency	830.134	370
22.	Bojonegoro Regency	1.252.020	758
23.	Tuban Regency	1.177.016	928
24.	Lamongan Regency	1.189.380	1.106
25.	Gresik Regency	1.326.420	3.900
26.	Bangkalan Regency	994.212	751
27.	Sampang Regency	989.001	349
28.	Pamekasan Regency	888.214	488
29.	Sumenep Regency	1.092.387	773
30.	Kediri City	289.109	505
31.	Blitar City	142.798	340
32.	Malang City	874.890	2.386
33.	Probolinggo City	239.024	972
34.	Pasuruan City	201.585	899
35.	Mojokerto City	129.891	838
36.	Madiun City	177.399	253
37.	Surabaya City	2.904.751	17.232
38.	Batu City	209.125	956

Table 2 The Results of Normalization Data

No	Regencies and Cities in East Java	Population Density	Confirmed Cases
1.	Pacitan Regency	-0,734338441	-0,458636817
2.	Ponorogo Regency	-0,264508067	-0,301933185
3.	Trenggalek Regency	-0,523677069	-0,324663822
4.	Tulungagung Regency	-0,009605368	-0,340161983
5.	Blitar Regency	0,169803993	-0,198267706
6.	Kediri Regency	0,789076964	-0,136619464
7.	Malang Regency	2,335958524	-0,175881473
8.	Lumajang Regency	-0,007320485	-0,068083151
9.	Jember Regency	2,097823528	0,491917079
10.	Banyuwangi Regency	0,845190456	0,447489017
11.	Bondowoso Regency	-0,402904091	-0,206533392
12.	Situbondo Regency	-0,541265902	-0,167615787
13.	Probolinggo Regency	0,186317324	0,002175181
14.	Pasuruan Regency	0,874745168	0,058312965
15.	Sidoarjo Regency	1,833522466	2,042077618
16.	Mojokerto Regency	0,114173956	-0,175192666
17.	Jombang Regency	0,325573154	-0,024688299
18.	Nganjuk Regency	0,010966009	-0,311920889
19.	Madiun Regency	-0,544229108	-0,523384691
20.	Magetan Regency	-0,625693486	-0,361859409
21.	Ngawi Regency	-0,326525658	-0,478612225
22.	Bojonegoro Regency	0,301052323	-0,344983634
23.	Tuban Regency	0,189479863	-0,286435024
24.	Lamongan Regency	0,207871974	-0,225131186
25.	Gresik Regency	0,411726301	0,737132432
26.	Bangkalan Regency	-0,082450861	-0,347394459
27.	Sampang Regency	-0,090202502	-0,485844700
28.	Pamekasan Regency	-0,240128553	-0,437972602
29.	Sumenep Regency	0,063589701	-0,339817580
30.	Kediri City	-1,131329272	-0,432117741
31.	Blitar City	-1,348974707	-0,488944332
32.	Malang City	-0,259948715	0,215705404
33.	Probolinggo City	-1,205833388	-0,271281266
34.	Pasuruan City	-1,261525902	-0,296422728
35.	Mojokerto City	-1,368174560	-0,317431347
36.	Madiun City	-1,297503870	-0,518907444
37.	Surabaya City	2,759578050	5,328721034
38.	Batu City	-1,250309749	-0,276791724

Table 3 Euclidean Distance Matrix

	1	2	3	...	16	...	28	29	...	38
1	0									
2	0,49527428	0								
3	0,24965371	0,26016390	0							
⋮				⋮						
16	0,89460264	0,39932848	0,65513018	...	0					
⋮	...	...	...	...	...	⋮				
28	0,49464171	0,13820667	0,30535003	...	0,44111627	...	0			
29	0,80672630	0,33027772	0,58746	...	0,17222116	...	0,31918519	0		
⋮	...	...	...	...	...	...	...	...	⋮	
38	0,54707772	0,98612223	0,72820793	...	1,36826099	...	1,0229591	1,31541021	0,54707772	0

Table 4 Member Cluster with Single Linkage Regarding Population Density of Regencies and Cities in East Java in 2020 with Cumulative Confirmed COVID-19 Cases

No	Cluster	Member of Cluster	Number of Member in Cluster
1.	Cluster 1	Jember Regency	1
2.	Cluster 2	Malang Regency	1
3.	Cluster 3	Probolinggo Regency, Jombang Regency, Madiun Regency, Bondowoso Regency, Situbondo Regency, Pacitan Regency, Trenggalek Regency, Magetan Regency, Lumajang Regency, Ponorogo Regency, Ngawi Regency, Pamekasan Regency, Sampang Regency, Bangkalan Regency, Sumenep Regency, Tulungagung Regency, Nganjuk Regency, Bojonegoro Regency, Tuban Regency, Mojokerto Regency, Blitar Regency, Lamongan Regency	22
4.	Cluster 4	Malang City	1
5.	Cluster 5	Kediri City, Blitar City, Madiun City, Mojokerto City, Probolinggo City, Pasuruan City, Batu City	7
6.	Cluster 6	Kediri Regency, Pasuruan Regency	2
7.	Cluster 7	Banyuwangi Regency	1
8.	Cluster 8	Gresik Regency	1
9.	Cluster 9	Sidoarjo Regency	1
10.	Cluster 10	Surabaya City	1
Total			38

Table 5 Member Cluster with Average Linkage Regarding Population Density of Regencies and Cities in East Java in 2020 with Cumulative Confirmed COVID-19 Cases

No	Cluster	Member of Cluster	Number of Member in Cluster
1.	Cluster 1	Jember Regency	1
2.	Cluster 2	Malang Regency	1
3.	Cluster 3	Sidoarjo Regency	1
4.	Cluster 4	Banyuwangi Regency, Kediri Regency, Pasuruan Regency	3
5.	Cluster 5	Gresik Regency	1
6.	Cluster 6	Lumajang Regency, Probolinggo Regency, Jombang Regency, Bojonegoro Regency, Mojokerto Regency, Tuban Regency, Blitar Regency, Lamongan Regency, Ponorogo Regency, Ngawi Regency, Pamekasan Regency, Sampang Regency, Bangkalan Regency, Sumenep Regency, Tulungagung Regency, Nganjuk Regency	16
7.	Cluster 7	Malang City	1
8.	Cluster 8	Bondowoso Regency, Situbondo Regency, Pacitan Regency, Madiun Regency, Trenggalek Regency, Magetan Regency	6
9.	Cluster 9	Blitar City, Madiun City, Kediri City, Mojokerto City, Probolinggo City, Pasuruan City, Batu City	7
10.	Cluster 10	Surabaya City	1
Total			38

Table 6 Member Cluster with Complete Linkage Regarding Population Density of Regencies and Cities in East Java in 2020 with Cumulative Confirmed COVID-19 Cases

No	Cluster	Member of Cluster	Number of Member in Cluster
1.	Cluster 1	Banyuwangi Regency, Gresik Regency	2
2.	Cluster 2	Kediri Regency, Pasuruan Regency	2
3.	Cluster 3	Sumenep Regency, Tulungagung Regency, Nganjuk Regency, Bangkalan Regency, Sampang Regency, Bojonegoro Regency, Mojokerto Regency, Tuban Regency, Blitar Regency, Lamongan Regency, Lumajang Regency, Probolinggo Regency, Jombang Regency	13
4.	Cluster 4	Malang City	1
5.	Cluster 5	Pacitan Regency, Madiun Regency, Trenggalek Regency, Magetan Regency, Bondowoso Regency, Situbondo Regency, Ponorogo Regency, Ngawi Regency, Pamekasan Regency	9
6.	Cluster 6	Mojokerto City, Probolinggo City, City of Pasuruan, Batu City, Kediri City, Blitar City, Madiun City	7
7.	Cluster 7	Jember Regency	1
8.	Cluster 8	Malang Regency	1
9.	Cluster 9	Sidoarjo Regency	1
10.	Cluster 10	Surabaya City	1
Total			38