

Grouping of the Second Wave of Covid-19 Infection Areas in East Java Province Using K-Means Clustering

Fadilah Akbar¹, Edo Leonardo Dekapriyo¹, Agoes Moh. Moefad²,
Achmad Murtafi Haris²

¹Department of Mathematics, UIN Sunan Ampel Surabaya, Indonesia

²Department of Communication Studies, UIN Sunan Ampel Surabaya, Indonesia

Article history:

Received Aug 4, 2021

Revised, Jun 9, 2022

Accepted, Jun 30, 2022

Kata Kunci:

Covid-19, K-Means Clustering, Pengelompokan, Jawa Timur, Wilayah Infeksi

Keywords:

Covid-19, K-Means Clustering, Grouping, East Java, Infection Areas

Abstrak. Serangan pandemi Covid-19 gelombang kedua yang terjadi pada bulan Juni 2021 mengakibatkan meningkatnya jumlah korban jiwa secara drastis. Hal ini diakibatkan karena berkurangnya tanggung jawab masyarakat, rasa untuk saling menjaga, dan rasa untuk saling melindungi sehingga menyebabkan melonggarnya protokol kesehatan yang diterapkan. Perlu adanya mitigasi yang tepat untuk menangani pandemi wabah penyakit Covid-19 ini secara tepat, salah satunya adalah dengan menangani kasus pada setiap wilayah dengan tingkat keparahan dari tinggi ke rendah. Dengan pengelompokan tersebut, dapat memberikan keputusan wilayah mana saja yang harus diutamakan dalam mitigasi pandemi Covid-19 ini. Untuk pengelompokan setiap wilayah dapat digunakan metode *K-Means Clustering*. Dari cluster tersebut terdapat satu wilayah dengan tingkat keparahan tinggi, delapan wilayah dengan tingkat keparahan sedang, dan 29 wilayah dengan tingkat keparahan rendah.

Abstract. The second wave of Covid-19 Pandemic Attacks that occurred in June 2021 resulted in a drastic number of fatalities. This is due to reduced community responsibility, a sense of caring for each other, and a sense of mutual protection, resulting in loosening of the health protocols implemented. There needs to be proper mitigation to handle the Covid-19 pandemic properly, one of which is by handling cases in each region with a low level of severity. With this grouping, it can provide a decision on which areas must be available in mitigating the Covid-19 pandemic. For grouping each region, the K-Means Clustering method can be used. From these clusters, there is one area with a severity level, eight areas with moderate severity, and 29 areas with low severity.

How to cite:

F. Akbar, E. L. Dekapriyo, A. M. Moefad, and A. M. Harris, "Grouping of The Second Wave of Covid-19 Infection Area in East Java Province Using K-Means Clustering", *J. Mat. Mantik*, vol. 8, no. 1, pp. 53-62, Jun. 2022.

CONTACT:

Fadilah Akbar,



fadilakbar783@gmail.com



Department of Mathematics, UIN Sunan Ampel Surabaya, Indonesia



The article can be accessed here. <https://doi.org/10.15642/mantik.2022.8.1.53-62>

1. Introduction

The Coronavirus Disease 2019 (Covid-19) pandemic has been around for more than two years and has become a threat to all human life, until now Covid-19 still occupies dangerous diseases throughout the world, one of which is Indonesia. As of July 2021, reported cases of people who were positively exposed and suffering from Covid-19 reached 193 million cases where the death toll was 4.14 million worldwide [1]. Meanwhile, in Indonesia, reported cases of people who were positively exposed and suffered from Covid-19 reached 3.03 million, with the number of patients who had been declared cured of 2.39 million, and the number of victims who died as many as 79,032 million people [2]. Last June 2021 became the beginning of the second wave of Covid-19 attacks in Indonesia [3].

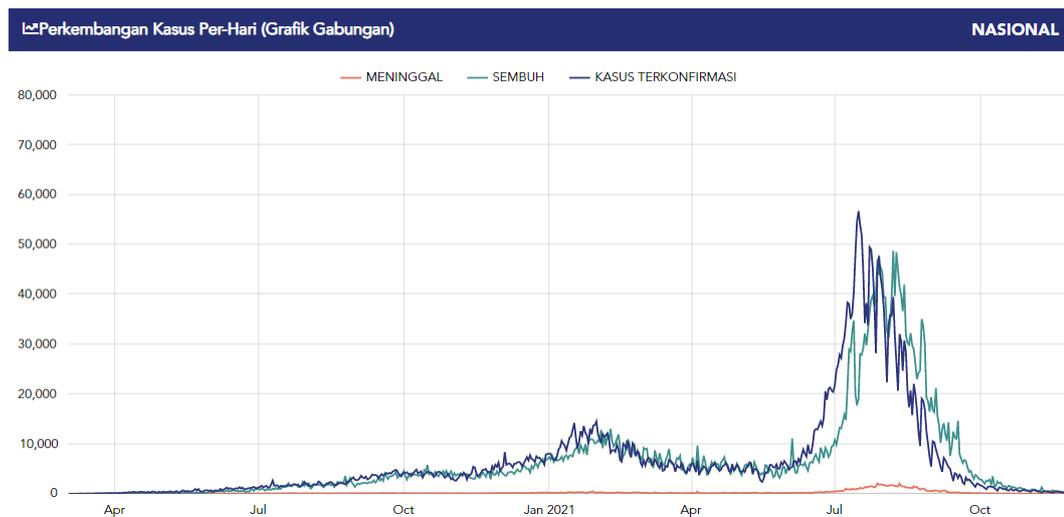


Figure 1. The rate of development of the number of Covid-19 cases in Indonesia [4]

Attention to the graph in Figure 1, after the first peak of cases in February fell, it appears that in mid-May 2021 there was a spike in the number of cases, this is the sign of the start of the second wave of the Covid-19 pandemic. This second attack can occur as a result of reduced community responsibility, mutual care, and a sense of mutual protection, causing a loosening of the health protocols applied [5]. In addition, there are other factors that can cause this second wave of attacks, namely mutating the Covid-19 virus which causes the virus to become stronger, has a more severe impact on victims, and accelerates infection or transmission [6]. Conditions like this if not observed and get the right treatment, will result in more and more cases which will cause a large number of victims, the need for targeted mitigation and appropriate methods to accelerate the abatement of the Covid-19 pandemic. In early July, the Indonesian government imposed an emergency restriction on community activities or better known as the PPKM.

This PPKM is implemented simultaneously for all regions on the islands of Java and Bali, this is due to the high number of positive cases from the second wave of Covid-19 attacks that occurred in Java and Bali [7]. The centre of the case of the second wave of Covid-19 attacks is the province of DKI Jakarta as the capital and economic centre in Indonesia, and the province of East Java as the second largest metropolis in Indonesia. Especially for the East Java region which was once the province with the highest cases in Indonesia, especially the city of Surabaya which was once the black zone (worst) [8]. This statement is also reinforced by research conducted by Solichin and Khairunnisa, where from the research conducted, it is said that the province of East Java is the province with the highest cases in Indonesia, although for now starting on July 25, 2021 for the province of East Java itself ranks fourth with cases most as shown in picture 2 [9]. From the problems

that have been presented, it is necessary to classify which areas are vulnerable areas affected by the second wave of Covid-19 attacks in East Java province.

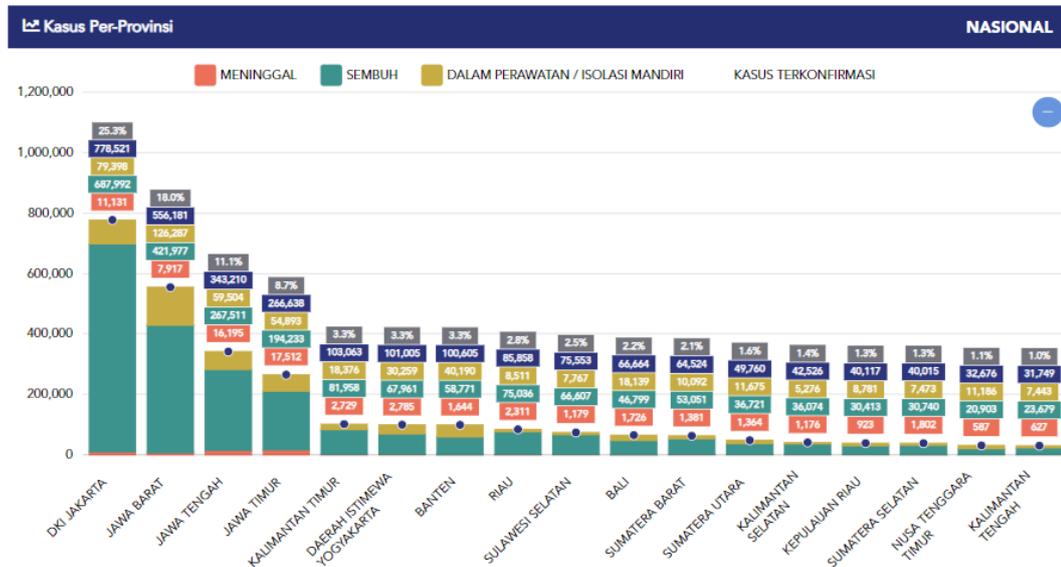


Figure 2. Covid-19 Cases in Indonesia [4]

One method that can be used to classify which areas are affected in East Java Province is to use clustering, for this reason, the use of clustering itself has been widely used to classify Covid-19 cases around the world [10]. Therefore, a grouping is needed that focuses on areas with a high number of cases to suppress the spike in cases and can minimize the number of victims, one of which is the grouping of areas affected by the second wave of Covid-19 in the East Java Province [11]. In determining the right mitigation strategy for handling Covid-19, it is necessary to know the severity of infection in each region to determine the priority of handling, one of which is by clustering [12]. There are various types of clustering methods that can be used, but in this study, K-Means Clustering will be used. Based on several previous studies, the main variables that are often used to be processed with K-Means Clustering are three, these variables are used as a benchmark for the severity of a case, and the variables are the number of infected people, recovered patients, and patients who died from Covid-19 [13], [14].

The severity of outbreak cases from an area can be determine based on the number of infected cases, from the number of infected cases and then compared with the number of instances recovered and died in a city [15]. The number of infected cases who recovered if the ratio were more significant than the number of infected cases who died, then grouping with K-Means would produce a group with low severity. There are several clustering methods available, but when compared to other clustering methods, K-Means Clustering is a clustering method with a higher accuracy level with a standard deviation of 2.2219, when compared to the standard deviation obtained from clustering with FCM of 2.7592. , it can be said that K-Means gives better results [16], [17]. K-Means Clustering is generally often used to group an area affected by a pandemic or group something with a disease topic [18].

This method is also quite effective and efficient in classifying Covid-19 cases, in a study conducted by Zubair et al, an analysis was carried out on how efficient this method was, and it was proven to provide good results and a high level of accuracy [19]. In addition, K-Means Clustering can also be used to determine the source of the spread of where this virus came from, so using this method will provide more accurate results and can also provide a more complete analysis. There have been several previous studies that have been conducted discussing the grouping of provinces affected by the Covid-19 pandemic, one of which was research conducted by Dahlan A., et al. In this study, provinces across Indonesia

were divided into three categories of outbreak severity. The method used to classify each province in this study using K-Means Clustering, DKI. Jakarta as the nation's capital has exceptions which will automatically be included in the area with the highest severity level, in addition, it was found that there are five other provinces classified in category one with the highest level of severity [20]. with high severity, then the other 28 provinces were classified in category two with moderate severity, while category three only contained DKI. Jakarta province [21].

Then there is a study conducted by Diah NS and Irma Y. In this study discusses the division based on the severity of Covid-19 in all provinces throughout Indonesia, the method used to classify is K-Means Clustering with the division of three categories, namely provinces with high levels of low severity, moderate severity, and high severity. The results obtained are 27 provinces fall into the category of areas with low severity, for the category with areas with moderate severity there are five provinces, while for the category with areas with high severity there are two provinces, where the two provinces are DKI Jakarta and East Java. Next, the researcher re-categorizes DKI. Jakarta and East Java Provinces into two, DKI. Jakarta Province is categorized as the area with the worst impact of Covid-19, followed by East Java Province [22].

Next is the research conducted by Fitria V., and Yasmin E. F., which in this study discusses the division based on the severity of Covid-19 in each province throughout Indonesia into seven categories. The category with the lowest severity starts from category seven and continues up to category one with the highest severity. The method used for grouping provinces in this study uses K-Means Clustering. From this study, it was found that 10 provinces were included in category seven, 11 provinces were included in category six, five provinces were included in category five, five provinces were included in category four, two provinces were included in category three, one province was included in category two, and one province fall into category one. The province for category one is DKI. Jakarta, the province for category two is East Java, and the two provinces for category three are Central Java and West Java, this makes the island of Java the region with the severity of Covid-19 infection in Indonesia [23].

And for the last literature is research conducted by Ali M., which in this study discusses the distribution based on the level of spread of Covid-19 cases in all districts or cities in Central Java province. The grouping is divided into three categories, the method used for grouping in this study is K-Means Clustering. From this study, the results obtained are 33 districts or cities which are included in category one, one city is included in category two, and one district is included in category three. This study does not classify the severity of each region, but only analyzes how good the clustering results are based on the similarity of characteristics, the distance between clusters, and the centroid [24].

From the four previous studies that have been carried out, in this study the K-Means Clustering method will be used to group each district or city in the province of East Java. The ease and accuracy of K-Means Clustering is very suitable for use in this study. The categories will be divided into three categories, the determination of the number of 3 clusters is based on the level of accuracy obtained from several previous studies that have been described previously, where better cluster results are obtained when grouped into 3 clusters. Namely for category one to be an area with a high severity level, category two to be an area with a moderate severity level, and category three to be an area with a low severity level. It is hoped that this research will provide a map of the distribution of the spread of Covid-19 so that it can provide the right mitigation to be able to suppress the high number of cases.

2. Methods

In this research, of course, requires a method to run it in a structured, sequential, and correct manner. The methods used, if briefly described, are presented in a flowchart in Figure 3 below. The following is an overview of this research:

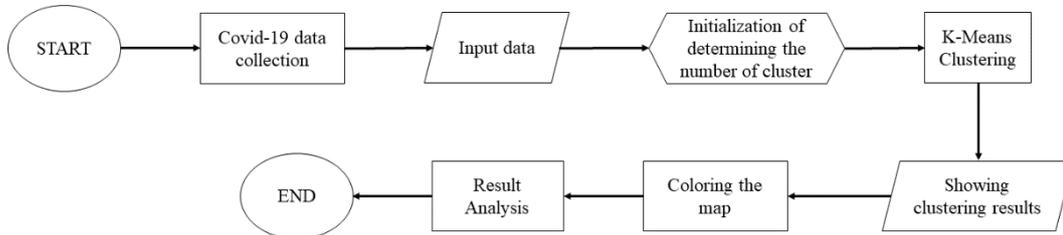


Figure 3. Research Flowchart

2.1 Data Source

The data used in this study is secondary data. Data usage in July 2021 which can be accessed via <http://infocovid19.jatimprov.go.id/index.php/data>.

Table 1. Sample Data on the Distribution of Covid-19 in East Java Province

No	District / City	Active Patient	Recovery Patient	Patient Died
1	Dist. Nganjuk	1538	7242	512
2	Dist. Trenggalek	697	4587	633
3	Batu City	390	1821	180
⋮	⋮	⋮	⋮	⋮
38	Dist. Banyuwangi	839	8406	1144

2.2 K-Means Clustering

Clustering is the process of grouping objects based on data according to certain desired characteristics or partitioning data sets into subsets [25]. The cluster method is divided into two, namely hierarchical and non-hierarchical methods. The hierarchical method is done by grouping data that have the closest similarity, while non-hierarchical is used for large data and must determine the number of clusters to be formed [14]. Many stages are used for cluster analysis, including the K-Means method. K-Means aims to partition data into two or more groups. The following are the steps in classifying data using the K-Means method [26]:

- a) Determine the number of k as the number of groups formed.
- b) Calculate the centre point using the formula :

$$\frac{\sum_{i=1}^n x_i}{n} \tag{1}$$

Where :

x_i : Data- i

n : There are many observations in the group

- c) Calculating the distance of each observation to each centroid of each group, namely the Euclidian Distance.

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \tag{2}$$

Where :

x_i : Data x on i observation

n : Many observations

y_i : Data y on i observation

- d) Group each data based on the closest distance to the centroid.
- e) Determine the position of the new centre point by calculating the average of the data in the same centre.

The application of the K-Means clustering method, each object is calculated at close range based on the characteristics it has with a predetermined cluster centre. If the smallest distance between objects and each cluster is a member of the closest cluster. Data processing is carried out using R software. The following research steps are carried out, among others:

- a) Covid-19 data collection in East Java Province based on district/city.
- b) Perform the calculation of the K-Means method as previous explanation.
- c) Grouping the area from the calculation of the distance between objects that have the same characteristics.
- d) Coloring the area based on the highest to lowest clustering results.

3. Results and Discussion

Based on the Covid-19 data that has been obtained, the next step is to group each region using k-means clustering, the following results are obtained as shown in Figure 4 below:

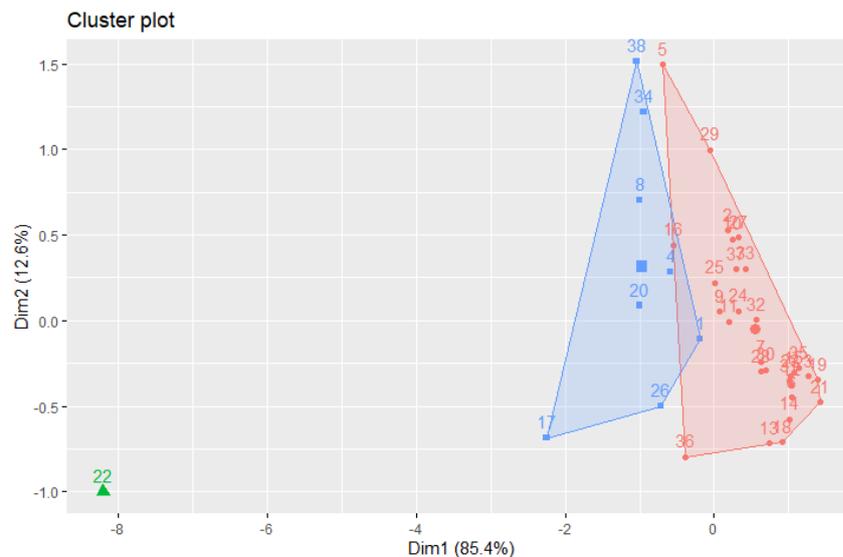


Figure 4. Plot of K-Means Clustering Results

Based on the results of the k-means clustering before, we can conclude in the form of a table as below:

Table 2. Results of Grouping With Clustering

No.	District / City	Cluster	Status	No.	District / City	Cluster	Status
1	Dist. Nganjuk	2	Medium	20	Malang City	2	Medium
2	Dist. Trenggalek	3	Low	21	Dist. Sampang	3	Low
3	Batu City	3	Low	22	Surabaya City	1	High
4	Dist. Kediri	2	Low	23	Mojokerto City	3	Low
5	Dist. Blitar	3	Low	24	Dist. Madiun	3	Low
6	Blitar City	3	Low	25	Dist. Magetan	3	Low
7	Madiun City	3	Low	26	Dist. Gresik	2	Medium
8	Dist. Jember	2	Medium	27	Dist. Bangkalan	3	Low
9	Dist. Lumajang	3	Low	28	Dist. Probolinggo	3	Low
10	Dist. Situbondo	3	Low	29	Dist. Tuban	3	Low
11	Dist. Pasuruan	3	Low	30	Dist. Lamongan	3	Low
12	Dist. Sumenep	3	Low	31	Probolinggo City	3	Low
13	Dist. Mojokerto	3	Low	32	Dist. Bojonegoro	3	Low
14	Dist. Pacitan	3	Low	33	Dist. Bondowoso	3	Low
15	Pasuruan City	3	Low	34	Dist. Jombang	2	Sedang
16	Dist. Ponorogo	3	Low	35	Kediri City	3	Low
17	Dist. Sidoarjo	2	Medium	36	Dist. Malang	3	Low
18	Dist. Tulungagung	3	Low	37	Dist. Ngawi	3	Low
19	Dist. Pamekasan	3	Low	38	Dist. Banyuwangi	2	Medium

Based on the results of the clustering of regions with clustering above, it appears that the city of Surabaya enters cluster 1, where the status of the severity of the Surabaya is high and has resulted in the Surabaya being the city with the highest cases in the province of East Java. Then there are eight areas that are included in cluster 2 where in this group are areas with moderate severity due to Covid-19, these areas are: Dist. Nganjuk, Dist. Kediri, Dist. Jember, Dist. Sidoarjo, Malang City, Dist. Gresik, Dist. Jombang, and Dist. Banyuwangi. As for the areas that are included in cluster 3 where in this group are areas with low severity due to Covid-19 totalling 29, these areas are: Dist. Trenggalek, Batu City, Kabu. Blitar, Blitar City, Madiun City, Dist. Lumajang, Dist. Situbondo, Dist. Pasuruan, Dist. Sumenep, Dist. Mojokerto, Dist. Pacitan, Pasuruan City, Dist. Ponorogo, Dist. Tulungagung, Dist. Pamekasan, Dist. Sampang, Mojokerto City, Dist. Madiun, Dist. Magetan, Dist. Bangkalan, Dist. Probolinggo, Dist. Tuban, Dist. Lamongan, Probolinggo City, Dist. Bojonegoro, Dist. Bondowoso, Kediri City, Dist. Malang, and Dist. Ngawi. In this case, Surabaya should be prioritized in taking mitigation measures to prevent the spread of Covid-19 infection in East Java Province. Given the small size of the Surabaya area with a high population density, it is appropriate that Surabaya deserves to be included in cluster 1 with a high severity level, to be prioritized in carrying out appropriate treatment.

4. Conclusion

Based on the results obtained from the grouping of districts/cities based on the severity of the impact of the spread and transmission of Covid-19, there is one high-affected area, eight moderately-affected areas, and 29 low-affected areas. When depicted on a map, the map of districts/cities impacting the second wave of the Covid-19 pandemic attack is as follows:

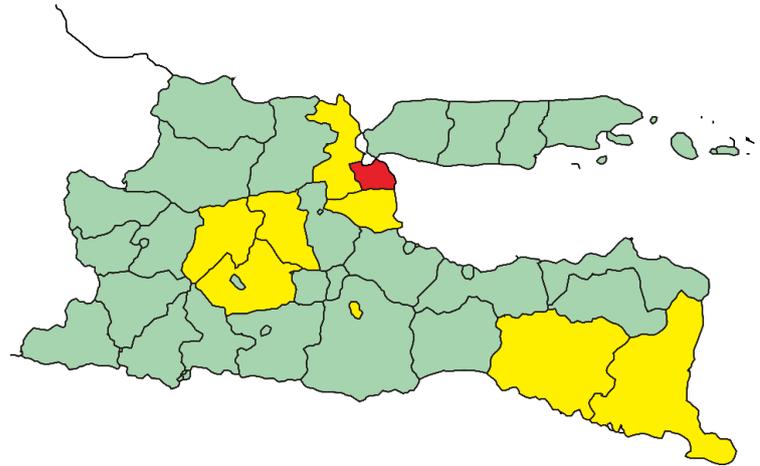


Figure 5. Map of the Distribution of Areas Affected by Covid-19 East Java

From the map shown above, the areas marked in red must be implemented to prevent transmission and control activities that are not strictly necessary and strictly enforced. Because if it is not handled quickly, the area around the red zone will also increase and cause the yellow and green zones to increase as well. Then for the area with yellow colour, it is an area with an alert condition, this condition is not too high in severity but still has to follow the health protocols that have been recommended by the government and medical personnel, if the health protocol is not implemented, it will result in this yellow zone becoming a red zone, very dangerous for the wider community. the region. Meanwhile, the green zone does not need to be too strict in handling cases in that area, but it must be handled seriously and until there are no active cases in the area.

Suggestions for future research may be adding variables that influence other than active patients, recovered patients, and patients who died. Then it is possible that the number of affected area groupings can be increased and the results can also be compared if using three groups with groups >3 . In addition, it is also possible to use other cluster methods such as Fuzzi C-Means Clustering (FCM), Hierarchical Clustering, or Gath-Geva Clustering to obtain more diverse regional grouping results. In addition, it is also possible to use other cluster methods such as Fuzzy C-Means Clustering (FCM), Hierarchical Clustering, or Gath-Geva Clustering to obtain more diverse regional grouping results. This is because in research related to clustering areas with case studies of the severity of Covid-19 infection, it is still rare to compare the results of clustering with the methods described. Comparing the results of clustering with several methods simultaneously, can provide more information on decisions taken in the future which clustering method to use, or it can also be developed by combining two methods at once to get more optimal results.

Acknowledgement

The researcher realizes that in completing this research until this article is published, it cannot be completed properly without the help of many parties. The researcher would like to thank the Indonesia Covid-19 handling unit team (Satgas Covid-19) and Ministry of Health Republic of Indonesia (Kemenkes RI) for sharing the data free of charge online, so that researchers can easily access and retrieve it. Then the researcher also thanks to LPPM of UIN Sunan Ampel Surabaya which has given us the opportunity to conduct this research.

References

- [1] H. Ritchie *et al.*, “Coronavirus Pandemic (COVID-19),” *OurWorldInData.org.*, 2021. <https://ourworldindata.org/coronavirus-data> (accessed Jul. 23, 2021).
- [2] S. T. P. Covid-19, “Peta Sebaran Provinsi Jawa Timur,” *Covid19.go.id*, 2021. <https://covid19.go.id/peta-sebaran> (accessed Jul. 23, 2021).
- [3] “Pakar: Indonesia Menuju Puncak Gelombang Kedua Corona,” *CNN Indonesia*, Jakarta, Jun. 21, 2021.
- [4] S. T. P. Covid-19, “Peta Sebaran Nasional Per-Provinsi,” *Covid19.go.id*, 2021. <https://covid19.go.id/peta-sebaran> (accessed Jul. 24, 2021).
- [5] F. R. Makarim, “Gelombang Kedua COVID-19 Berpotensi Terjadi di RI, Apa Sebabnya?,” *halodoc*, 2021. <https://www.halodoc.com/artikel/gelombang-kedua-covid-19-berpotensi-terjadi-di-ri-apa-sebabnya> (accessed Jul. 23, 2021).
- [6] E. Parwanto, “Virus Corona (SARS-CoV-2) penyebab COVID-19 kini telah bermutasi,” *J. Biomedika dan Kesehat.*, vol. 4, no. 2, pp. 47–49, 2021, doi: 10.18051/jbiomedkes.2021.v4.47-49.
- [7] I. Safutra, “PPKM Darurat Batasi Aktivitas Masyarakat di Jawa-Bali,” *Jawa Pos*, Surabaya, Jul. 02, 2021.
- [8] A. Faizal, “Surabaya Jadi Zona Hitam, Apa yang Terjadi?,” *Kompas*, Surabaya, Jun. 02, 2020.
- [9] A. Solichin and K. Khairunnisa, “Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means,” *Fountain Informatics J.*, vol. 5, no. 2, p. 52, 2020, doi: 10.21111/fij.v5i2.4905.
- [10] R. A. Indraputra and R. Fitriana, “K-Means Clustering Data COVID-19,” *J. Tek. Ind.*, vol. 10, no. 3, pp. 275–282, 2020.
- [11] A. S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha, and M. Azmi, “Using K-Means Clustering to Cluster Provinces in Indonesia,” *J. Phys. Conf. Ser.*, vol. 1028, no. 1, 2018, doi: 10.1088/1742-6596/1028/1/012006.
- [12] M. Azarafza, M. Azarafza, and H. Akgün, “Clustering method for spread pattern analysis of corona-virus (COVID-19) infection in Iran,” *J. Appl. Sci. Eng. Technol. Educ.*, vol. 3, no. 1, pp. 1–6, 2021, doi: 10.1101/2020.05.22.20109942.
- [13] M. C. Untoro, L. Anggraini, M. Andini, H. Retnosari, and M. A. Nasrulloh, “Penerapan metode k-means clustering data COVID-19 di Provinsi Jakarta,” *Teknologi*, vol. 11, no. 2, pp. 59–68, 2021, doi: 10.26594/teknologi.v11i2.2323.
- [14] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, “COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm,” *J. Phys. Conf. Ser.*, vol. 1783, no. 1, 2021, doi: 10.1088/1742-6596/1783/1/012027.
- [15] S. Harini, “Identification COVID-19 Cases in Indonesia with The Double Exponential Smoothing Method,” *J. Mat. Mantik*, vol. 6, no. 1, pp. 66–75, 2020, doi: <https://doi.org/10.15642/mantik.2020.6.1.66-75>.
- [16] M. A. Ramadhan, Efendi, “Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling,” in *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9*, 2017, pp. 219–226.
- [17] A. Chusyairi and P. R. N. Saputra, “Perbandingan Algoritma Fuzzy C-Means Dan K-Means Clustering Dalam Pengelompokan Data Puskesmas,” in *Conference on Information Technology, Information System and Electrical Engineering (CITISEE 2019)*, 2019, no. June, pp. 1–6.
- [18] A. Bastian, H. Sujadi, and G. Febrianto, “Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka),” *J. Sist. Inf.*, vol. 14, no. 1, pp. 26–32, 2018.
- [19] M. Zubair, M. Asif Iqbal, A. Shil, E. Haque, M. Moshikul Hoque, and I. H. Sarker, “An Efficient K-Means Clustering Algorithm for Analysing COVID-19,” *Adv.*

- Intell. Syst. Comput.*, vol. 1375 AIST, pp. 422–432, 2021, doi: 10.1007/978-3-030-73050-5_43.
- [20] M. Fajar, “Modeling of COVID-19 Epidemic Growth Curve in Indonesia,” *J. Mat. Mantik*, vol. 7, no. 1, pp. 67–73, 2021, doi: <https://doi.org/10.15642/mantik.2021.7.1.67-73>.
- [21] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, “The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data,” *J. Qual. Quant.*, no. 0123456789, 2021, doi: 10.1007/s11135-021-01176-w.
- [22] D. N. Sari and I. Yunita, “Tingkat Keparahan Dan Risiko Penyebaran Covid-19 Di Indonesia Dengan Menggunakan K-Means Clustering,” in *Seminar Nasional Official Statistics*, 2019, pp. 210–216, doi: 10.34123/semnasoffstat.v2020i1.706.
- [23] F. Virgantari and Y. E. Faridhan, “K-Means Clustering of COVID-19 Cases in Indonesia’s Provinces,” in *Proceedings of the International Conference on Global Optimization and Its Applications Jakarta*, 2020, pp. 1–7.
- [24] A. Mahmudan, “Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering,” *J. Mat. Stat. dan Komputasi*, vol. 17, no. 1, pp. 1–13, 2020, doi: 10.20956/jmsk.v17i1.10727.
- [25] A. Ali, “Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 186–195, 2019, doi: 10.30812/matrik.v19i1.529.
- [26] M. S. Yana, L. Setiawan, E. M. Ulfa, and A. Rusyana, “Penerapan Metode K-Means dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018,” *J. Data Anal.*, vol. 1, no. 2, pp. 93–102, 2018, doi: 10.24815/jda.v1i2.12584.