MAPPING SPREAD VIRUS COVID-19 AT INDIA USING METHOD K-MEANS CLUSTERING

1st Abdullah Hasbi
Department of Information System
Nusa son University
Sukabumi, Indonesia
abdullahhasbi313@gmail.com

2nd Muhamad Zaenal Abidin
Department of Information System
Nusa son University
Sukabumi, Indonesia
muhamad.zaenal_si17@nusaputra.ac.id

3rd Siti Handayani
Department of Information System
Nusa son University
Sukabumi, Indonesia
siti.handayani_si17@nusaputra.ac.id

4th Dudih Gustian
Departement of Information System
Nusa Putra University
Sukabumi, Indonesia
dudih@nusaputra.ac.id

5th Muhammad Muslih
Department of Information System
Nusa Putra University
Sukabumi, Indonesia
muslih@nusaputra.ac.id

Abstrac-Covid have confirmed as pandemic global by the *World Health Organization* (WHO) because spread that very fast among humans. As a result of the Covid-19 virus, many infected patients died, including from all countries on the Asian continent. Like the case that occurred in one of the Asian countries, namely India, which is one of the countries that experienced a spike in Covid-19 cases, the transmission of thevirus Covid-19 in India penetrated more than 400,000 cases in 1 day. The number is the highest daily record set by India during the Covid-19 pandemic. However, it was found that the problem of the spread of Covid-19 tends to increase, this is the country with the second largest population in the world. The total number of Covid-19 cases in the country has reached 21 million or second only to the United States. The vastness of India's territory allows the need for grouping the parts by region in India. This grouping produces the center points for the spread of Covid-19 cases. The purpose of grouping Covid-19 cases based on *clusters is* to find out the weight/percentage value generated from each of these *clusters* using the *K-Means Clustering method*. This method is used to map the spread of the Covid-19 virus from various regions in India based on confirmed cases, dead, recovered and active/new clusters. The benefits obtained for the government in overcoming Covid-19 cases are to create strategies to prevent the spread of Covid-19 based on information from the results of regional clustering in India The results obtained from research conducted in 38 regions in India using 4 clusters resulted in Confirmed cases (C0) 199 items, Died (C1) 779 items, Recovered (C2) 21 items, and Active/new cluster (C3) 231 items with a totalCluster of 1230 items.

Keywords – *World Health Organization* (WHO), *K-Means Clustering*, India, Covid-19.

I. INTRODUCTION

Corona virus is a group of viruses that can cause disease in animals and humans. Several types of coronavirus have been found to cause respiratory tract infections in humans ranging from coughs and colds to more serious ones such as *Middle East Respiratory Syndrome* (MERS) and *Severe Acute Respiratory Syndrome* (SARS) [1]. The transmission was so fast that the World Health Organization/WHO declared the Corona virus pandemic or Covid-19 on 11/3/2020. This virus is very fast to spread because the method of transmission from human to human is very effective, namely through *close contact close contact* () and *respiratory droplets* such as coughing or sneezing [2]. Several studies have also shown that the corona virus can survive for several hours on objects touched by sufferers. If someone touches an object that has been contaminated with *droplets* and touches the eyes, nose or mouth, that person can be infected with Covid-19 [3].

The widespread spread of diseases caused by the Covid-19 virus which has been designated as a pandemic by WHO on March 11, 2020, due to the Covid-19 virus, many infected patients have died. The data obtained shows a death toll of 329,731 from all countries in the Asian continent based on data updated May 21, 2020, 03:12 GMT. Likewise in the countries of the Asian continent, many have died. Data on deaths due to the spread of the Covid-19 virus shows a figure of 25,851. Based on the data obtained, it is possible to classify the death rates from the spread of the Covid-19 virus which can later be input to countries in the Asian continent in order to improve the quality of handling patients with the Covid-19 virus [4].

Like the case that occurred in one of the Asian countries, namely India, which is one of the countries with the fastest spike in Covid-19 cases compared to other Asian countries and India is also a country with a new variant of the Covid-19 case, first detected in India in the fall. according to WHO [5]. Of the total population of India about 1.3 billion there are 58% confirmed cases with a value of 2,71,56,382, recovered 23% with a value of 2,43,43,299, 9% died with a value of 3,11,421 and *cluster* 10% active new with a value of 24,90,876. This is shown in Figure 1 below.

Fig. 1. The spread of Covid-19 in IndiaMonitoring *(percentage value of cases of spread)*

Results through the online site https://www.covid19india.org/. However, it was found that the problem of the spread of Covid-19 tends to increase, this is because it is a country with the largest population second in the world. With a human population of 1.3 billion, the total number of Covid-19 cases in this country has reached 21 million, the second highest after the United States. In addition, the government refuses to carry out a national quarantine but instead imposes a *lockdown* in certain areas, this is known after carrying out massive Covid-19 testing with the number of testing reaching 197 million of the total population. In addition, condition this is caused by At a time when infection rates have fallen for 30 consecutive weeks the Government of India is not taking advantage of improving its health care structure and launching an aggressive vaccination program and the Indian government has decided to relax by not canceling festivals or religious events before a wave of contagion occurs [6].

Methods *K-Means Clustering* with 4 attributes and the four *Cluster* is the results of *cluster* first as many as 2,71,56,382the data with the percentage of 58%, *cluster* both with the results of 2,43,43,299data and the percentage of 23%, *cluster* the third with 24 results ,90,876 data and the percentage result is 10%, and the *Cluster* fourth with a result of 3,11,421 data with a percentage yield of 9%. The application of data mining is decisive for mapping the spread of the Covid-19 virus from various regions in India based on confirmed cases, dead, recovered and active/*clusters* new. The *silhouette coefficient value* for data validation from results *Clustering* with *K-*

Medoids results in grouping Covid-19 data in which areas are infected with the best clustering done with 4 *clusters*. Of the 38 areas of application of theAlgorithm *K-Medoids* in region grouping [2].

This study provides a solution to cluster the spread of Covid-19 in India based on the parameters of the number of *Confirmed, Active, Recorded, Deceased*. Thus informing the process of the spread of this virus to the public correctly. This grouping produces the central points for the spread of Covid-19 cases using the *K-Means Clustering method*. This method is used to map the spread of the Covid-19 virus from various regions in India. The benefits resulting from this research for the Government of India can control the impact caused by Covid-19, so that the allocation of funds spent on overcoming this outbreak can be effective. In addition, the national tackling program can be realized properly due to the synergy of the level of trust between the Government and the community. So the economy will grow normally as expected.

II. LITERATURE REVIEW OF

RA Indraputra and R. Fitriana (2020) "*K-Means Clustering* Data Covid-19" This study explains how to obtain data from the spread of COVID-19 using data processing methodology. The methodology used is a data mining methodology, namely data *cleansing*, data *integration*, data *selection*, data *scaling*, data *preprocessing*, data *clustering*, and data analysis based on cases that occurred in Europe. Italy, which is the European country hardest hit by the Corona virus, has now recorded more than 15,000 cases. To process this data, the method is used *K-Means Clustering* , namely in three different ways, namely using formulas and *Microsoft Excel software*, and using *data mining software*, namely *Weka* and *Knime*. From the results of data processing, two *clusters were* data obtained, where cluster 2 had a higher number of infected and died compared to *cluster* 1, so that these areas *cluster* need to be prioritized for handling [2].

Dina Tri Utari (2021) conducted a study on the spread of Covid-19 in Bali Province on average caused by more domestic travel than positive cases caused by overseas travel, with the title "Analysis of the Characteristics of the Covid-19 Transmission Area with Using theMethod *K-Means Clustering*" The study of the process *clustering* that has been carried out has resulted in four *clusters*, the results *clustering* identified are given to related parties by tightening the mobility of the population both traveling to and from abroad and within the country. Positive cases caused by domestic travel in *clusters* 1, 2, and 3 need to be focused on limiting local transmission and domestic travel. As for the mitigation for *the cluster* 4th, it is necessary to focus on limiting local transmission and travel abroad [7].

Sukma Sindi, Weni Ratnasari Orktapia Ningse, Irma Agustika Sihombing, Fikrul Ilmi RHZer, Dedy Hartama (2020) study, explains to determine the data grouping covid-19 using pApplication of *K-Medoids. K-Medoids* is a partition method *clustering* which aims to find a set of *k-clusters* among the data that best characterizes the objects in a data set. To find out these data by means of a graph of the current spread of Covid-19 in Indonesia in particular, where this data uses 34 provinces in Indonesia. They will be grouped in to determine *clustering* 1, 2, and 3. Data on the spread of Covid-19 includes positive cases, recovered, and died. Thealgorithm *K-Medoids* can group Covid-19 data in any infected area with the best clustering done with 3 *clusters*. From 34 *records* obtained 1 *record* in *the cluster* first, 2 *records* in *the cluster* second, 31 *records* in *the cluster* third[8].

Research conducted by Sanjay Kumar in 2020 on the classification of districts affected by Covid-19 into *clusters* real based on observations of similarities within a cluster and differences between clusters. With the research title “Use of cluster analysis to monitor infection with *novel coronavirus-19* in Maharashtra, India”. The technique groups 32 different affected districts into three clusters (I-III) for each case. Where *cluster* I shows a good percentage of cured cases in several districts and has a 100% success rate for curing patients. It was observed that the districts in *clusters* II and III are in severe condition so that it is necessary to optimize monitoring techniques (*screening*, closures, curfews, lockdowns, evacuations, legal actions, etc.) that can assist the government, doctors, police, and others. involved in improving government policies, actions, etc. TH, PU and MC need more monitoring (closures, curfews, lockdowns, evacuations, legal action, etc.), to reduce the number of people infected [9].

Mehdi Azarafza, Mohammad Azarafza, Haluk Akgün in 2020 "method *Clustering* for analysis of patterns of spread of corona virus infection (Covid-19) in Iran". This study explains that the spread of the Covid-19 outbreak is very fast and widespread. Although, countries around the world have taken various precautions against this Covid-19. Placing Iran as the third country with the highest number of confirmed cases after China and Italy, until the end of March 2020. Then the application of data mining to perform pattern recognition of Covid-19 in Iran, which is mainly based on the method *clustering*. In this study there is an application of a method called GIS (Geographical Information System) used to determine the possible spread of the virus from the starting point (Qom city), to other parts of the point (Iran). As analysis *cluster*, its main task is to carry out data mining exploration [10].

2.1 Framework for Thinking

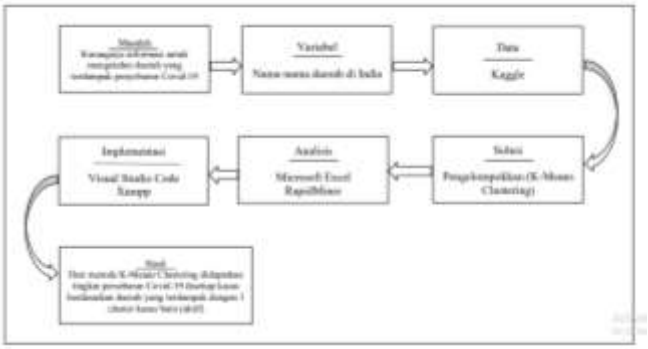

Fig. 2. Research Thinking Framework *(steps of the research process into flow diagrams)*

The framework is a model, method, or steps used to work on and solve the problem to be studied [11]. The problems that occur in India due to the surge in Covid-19 cases in various regions/cities, we need a solution to minimize its spread. The *K-Means Clustering method is* a solution for grouping each case that has an impact by producing a weighted/percentage value in each case. *software* Supporting to perform analysis using *Microsoft Excel, RapidMiner software*, and system implementation using *visual studio code* and *xampp*. The results achieved can be input for the government in dealing with Covid-19 cases in order to create a better life.

III. RESEARCH METHODS

3.1 Collecting Data

Collection Data an stage initial in methodology research. At stage this, done collection of data and information which obtained direct with do experiment on collection data DS4C (Science Data for Covid-19) from repository Kaggle [12]. In *database* this there is information, namely the date of observation, region/region, date of last update, number of confirmed cases, number of deaths, and number of people recovering every day. The data on the spread of Covid-19 used in this study is data on the spread from March 14, 2020 to April 27, 2021, with 1,430 items and 38 affected areas.

3.2 The Data Mining Process with the K-Means Method

Data Miningalso known as *Knowledge Discovery in Database* (KDD) is defined as extracting potential information,

then converting the results accurately into easy-to-understand information.

The KDD process can be broadly explained as follows [13] :

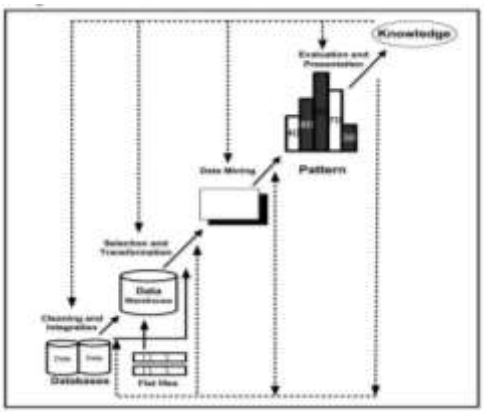

Fig. 3. Data mining as a part of the KDD process *(flow pattern processing data into information).*

1) Data used in this study is Covid-19 data obtained from the website database, namely *Kaggle*. but because there is a lot of data available, then Big Data like this is very difficult to use properly. So, in order to obtain useful and easy-to-use information, Data Mining techniques can be used to process this large amount of data. One technique that can be used is the technique *Clustering*, especially *K-Means Clustering.*
2) At this stage, the process is carried out *cleaning* data using *RapidMiner* by selecting several rows and columns for the calculation process.
3) There are 38 areas affected by Covid-19, with 3 confirmed cases, dead and recovered. The transformation of this data selection produces 1 *cluster* new (active), each of which will be determined by the number/percent in the process *clustering*.
4) From the results of the analysis of the Covid-19 distribution map, it can describe the spread of the number of cases by finding patterns in the form of graphs.
5) In this stage the data is processed using the method *K-means*. One of the most important means of dealing with data is to classify or group the data into a set of categories or *clusters*.

The following are the steps to perform calculations using the K-Means algorithm: [12]

a. Choose the number of *Clusters* desired(k) in the dataset.
b. Determine the center point (*Centroid*) randomly at the initial stage.
c. Calculate the closest distance of each data to the *Centroid*. To calculate the closest distance to the *Centroid* is *Euclidean distance* (d) Can use the following formula:

$$de = \sqrt{(Xi - Si)^2 + (Yi - Ti)^2} \quad (1)$$

Description:
(x, y) = Coordinates of objects
(s, t) = Coordinates of Centroid
i = Number of objects

d. Recalculate the center of the *Cluster* with membership *Cluster* the current.Pusat *Cluster* is an average of all the data in a *cluster*.Can be calculated using the formula:

$$V_{ij} = \sum_{k=1}^{Ni} Xkj \quad (2)$$

Description:
V_{ij} = *Centroid* average in *Cluster* the i-th for the j-the variable
N_i = Number of members *Cluster* of the i-th
i,k = Index of the *Cluster*
j = Index variable
Xkj = Value of data all the variable-k j for *Cluster*

e. Compute back each object using center *cluster* the new *(centroid* new), this is the initial stage of the opening of the new iteration. If the members *Cluster* do not experience migration *Cluster* again, then the process is *Clustering* declared complete. And repeat from stage c to e until no more data moves to another *cluster* [15].

IV. RESULTS AND DISCUSSION

4.1 Results

1) Calculations Using the K-Means Algorithm

Calculations were carried out with sample data of 38 regions/cities in India that were affected by the spread of Covid-19 with Confirmed, Died and Recovered cases every day. The following is a sample data on the spread of Covid-19 in India as of March 14, 2020 according to data from *Kaggle* previous.

From the results of the table above, the results of the total number of cases are obtained using *Microsoft Excel*. The following table shows the number of data on Covid-19 cases in India.

Table 1. Number of Data on Covid-19 Cases in India

No	Territory	Confirmed	Recovered	Deceased
1	AN	11,337	11,379	11,379
2	AP	2,003,185	2,010,239	2,010,303
3	AR	34,949	34,977	34,977
4	AS	465,264	466,219	466,237
5	BR	753,737	761,641	761,726
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
38	UN	-	-	-

At this stage it can be seen that:
Number of Clusters : 3
Number of Data : 38

Number of Covid-19 cases consisting of 38 regions, occurred spike in the highest cases in the MH region (22) with 8,077 confirmed cases, 1648 died 144,902 and recovered 8,145,797. This case is one of why India has imposed a *lockdown* in every region the reasons that has an increase in Covid-19 cases. Number of Attributes consists of 3 (Confirmed Case, Died and Recovered)

a. Determines the *centroid* initial randomly.
C1 (MH) = (8,077,164; 8,144,902; 8,145,797)
C2 (LA) = (25,287; 25,580; 25,581)
C3 (LD) = (3,462; 3,593; 3,593)

In the *cluster the* spread of Covid-19 which is a large-term spread is taken for the 22nd data, namely the MH region, then for the *Cluster* with a medium-term spread, the 19th data is the LA region, and for the *Cluster* with a small-term spread it is the 20th data, namely the LD region.

b. Calculates the *centroid* nearest.
For example: In the AN region (11,377; 11,379; 11,379)

AN (C1) $= \sqrt{(11,337 - 8,077,164)^2 + (11,379 - 8,144,902)^2} + (11,379 - 8,145,797)^2$
$= 14,049,218$

AN (C2) $= \sqrt{(11,337 - 25,287)^2 + (11,379 - 25,580)^2} + (11,379 - 25,581)^2$
$= 24,45)$

AN (C3) $= \sqrt{(11,337 - 3,462)^2 + (11,379 - 3,593)^2} + (11,379 - 3,593)^2$
$= 13,517$

Next step is to calculate the distance of each data with the *Centroid* with the equation *Euclidean Distance*. At this stage, the closest distance between the data and the *cluster* will determine which data belongs to the *cluster* . The following are the results of distance calculations in the 1st Iteration

Table 2. Results of Literacy Calculations 1

No	Territory	Confirmed	Recovered	Deceased
1	AN	14,049,218	24,453	13,537
2	AP	3,477,805	3,477,805	3,477,805
3	AR	60,566	60,566	60,566
4	AS	806,974	806,974	806,974
5	BR	1,314,703	1,314,703	1,314,703
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
38	UN	-	-	-

In the *cluster* model (*K-means*) several results can be seen *cluster*, namely from the calculation results, appear *clusters* new.

Table 3. *Cluster* Model The

No	Territory	Cluster 0	Cluster 1	Cluster 2	Cluster 3
1	AN	-0.152	-0.411	1,476	1,170
2	AP	-0.167	-0.511	1,881	1,440
3	AR	-0.114	-0.474	0.476	1,428
4	AS	-0.097	-0.473	1,216	1,340
5	BR	-0.020	-0.406	5,894	0.731
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
38	UN	-0.658	0.137	0	0

Cluster model shows C3 as *a cluster* new (active). The definition of active cases is people who are still considered sick or still in care who are unfamiliar with implementing health protocols. Calculations were carried out using software *RapidMiner*. *RapidMiner* is a software for data processing. Using data mining principles and algorithms, *RapidMiner* extracts patterns from large data sets by combining statistical methods, artificial intelligence and databases [14].

c. Perform grouping based on *clusters*.
The following are the results of grouping *cluster* in Iteration 1:

Table 4. Grouping *Cluster*

No	Territory	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Close Range
1	AN	-0.152	-0.411	1,476	1,170	C2
2	AP	-0.167	-0.511	1,881	1,440	C2
3	AR	-0.114	-0.474	0.476	1,428	C2
4	AS	-0.097	-0.473	1,216	1,340	C1
5	BR	-0.020	-0.406	5,894	0.731	C1
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
38	UN	-0.658	0.137	0	0	C4

Grouping is carried out to find out each area affected by the Covid-19 case, and the government's anticipation can be done in tackling these cases.

d. Determine the *Centroid* new for the next iteration by finding the average value of each *cluster*. If the iteration stages have achieved the same result without any further displacement, the calculation is stopped. The final result of the position *Cluster* of the *Centroid* iteration can be seen in the table below:

Table 5. Division *Cluster* In Iteration

No	Territory	Close Range
1	AN	C1
2	AP	C1
3	AR	C1
4	AS	C0
5	BR	C0
6	CH	C0
7	CT	C0
8	DN	C0
9	DD	C3
10	DL	C0
11	GA	C1
12	GJ	C0
13	HR	C0
14	HP	C0
15	JK	C1
16	JH	C2
17	KA	C1
18	KL	C0
19	LA	C0
20	LD	C0
21	MP	C0
22	MH	C0
23	MN	C1
24	ML	C1
25	MZ	C1
26	NL	C1
27	OR	C1
28	PY	C1
29	PB	C0
30	RJ	C0
31	SK	C1
32	TN	C1
33	TG	C1
34	TR	C1
35	UP	C0
36	UT	C0
37	WB	C1
38	UN	C3

The division of *clusters* based on the closest distance can be used as a reference in making decisions, *Clustering* produces values /weights in each case of distribution to minimize spikes that continue to increase. Grouping *the clusters* data in the table into several groups, where the data in one group has the same characteristics as each other and has different characteristics from the data in other groups.division *Cluster* in iteration consists of 4, where C0 becomes *Cluster* 1 (confirmed), C1 becomes *Cluster* 2 (died), C2 becomes *Cluster* 3 (recovered), and C3 becomes *Cluster* 4 (active/*cluster* new).

The results of data processing with 38 regions/cities in India can be seen in the following figure:

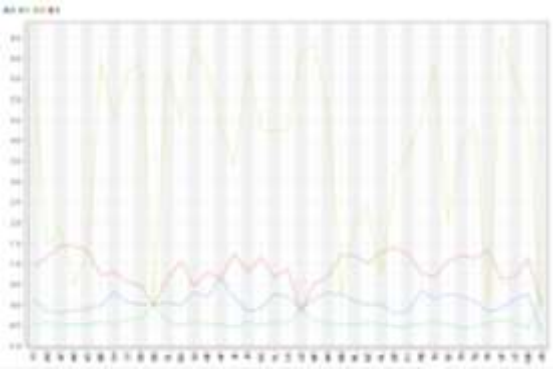

Fig. 4. Clusters of the level of spread of Covid-19 *(the rate of spread of Covid-19 cases every day has increased)*

The results obtained from the graph are *Cluster* 1 as a large-term spread of 779 items with cases of death. For *Cluster* 3 as the distribution in the medium term with 231 items with active cases. In *Cluster* 0 is still a medium-term deployment that with 199 items with confirmed cases. And *Cluster* 2 as a small-term spread with 21 items with recovered cases. With a total cluster of 1230 items.

4.2 Discussion

Indian Health Minister Harshvardhan reportedly blamed the second wave of infections and the lack of commitment of citizens to wear masks and practice social distancing as the cause of the surge in Covid-19 cases in India. In addition, the government refuses to carry out a national quarantine but instead imposes a *lockdown* in certain areas, this is known after carrying out massive Covid-19 testing with the number of testing reaching 197 million of the total population. In addition, this condition is caused by At a time when infection rates have fallen for 30 consecutive weeks the Government of India is not taking advantage of improving its healthcare structure and launching an aggressive vaccination program and the Indian government has decided to relax by not canceling festivals or religious events before a wave of contagion occurs. This has made Covid-19 cases continue to increase. By using *K-means Clustering*, *clustering* can be done based on regional data from 38 cities, with four *clusters* and the discovery *of clusters* new. So that the actions that must be taken by India include *locking*, One of the factors that triggered the decline in cases of Covid-19 infection in India is the regional lockdown policy which has proven to be effective in reducing cases, quarantine efforts, tightening health protocols while still implementing 3M (Using masks, washing hands and maintaining distance), India must increase its vaccination capacity. Performing an increasing number of tests, and tracking every movement of people in activity.

V. CONCLUSION

Based on the results of tests and trials on the application of the method *K-Means* to grouping or clustering Covid-19 cases in India, it can be concluded that the method *K-Means* can be used properly. The recommended number of clusters based on regional data calculations is 4 clusters. The results obtained are *Cluster* 1 as a large-term spread of 779 items with cases of death. For *Cluster* 3 as the distribution in the medium term with 231 items with active cases. In *Cluster* 0 is still a medium-term deployment that with 199 items with confirmed cases. And *Cluster* 2 as a small-term spread with 21 items with recovered cases. With a total cluster of 1230 items.

TABLE OF REFERENCES

[1] И. . омилова, . . асавин, and . . ордовицына, “Динамическая а олоземного ального остранства Области езонанса 1: 2 о оростью ащения емли,” *ономический*vol. 54, no. 4, pp. 337–348, 2020, doi:10.31857/s0320930x20040088.

[2] RA Indraputra and R. Fitriana, “K-Means Clustering Data on COVID-19,” *J. Tek. eng.*, vol. 10, no. 3, pp. 275–282, 2020.

[3] A. Solichin and K. Khairunnisa, “Clustering the Distribution of Corona Virus (Covid-19) in DKI Jakarta Using the K-Means Method,” *Fountain Informatics J.*, vol. 5, no. 2, p. 52, 2020, doi:10.21111/fiji.v5i2.4905.

[4] Noviyanto, "Application of Data Mining in Grouping the Number of Deaths," *Paradigm. information. and Computing.*, vol. 22, no. 2, 2020.

[5] SP Sicca, “What are the causes of the surge in Covid-19 cases in India? This article has been published on Kompas.com with the title 'What is the Cause of the Spike in Covid-19 Cases in India?', Click to read: https://www.kompas.com/global/read/2021/04/24/113651770 /apa-," *kompas.com*, 2021. https://www.kompas.com/global/read/2021/04/24/113651770/apakah-yang-jadi-sebab-lonjakan-case-covid-19-di - india (accessed Jul. 23, 2021).

[6] RDA Saptoyo *et al.*, “The Causes of the Covid-19 Tsunami in India: from Virus Mutation to Easing Procedures This article has been published on Kompas.com with the title 'Causes of the Covid-19 Tsunami in India: from Virus Mutations to Easing Procedures', Click to read: https:// /www.kompas.com/," *kompas.com*, 2021. https://www.kompas.com/tren/read/2021/04/25/103000065/pengebab-tsunami-covid-19-di-india-- from-mutation-virus-to-easing?page=all (accessed May 30, 2021).

[7] DT Utari, "Analysis of the Characteristics of the Covid-19 Transmission Area Using the K-Means Clustering Method," *J. Mediatek. and Sis. eng.*, vol. 5, no. 1, p. 25, 2021, doi:10.35194/jmtsi.v5i1.1220.

[8] S. Sindi, WRO Ningse, IA Sihombing, F. Ilmi RHZer, and D. Hartama, "K-Medoids Clustering Algorithm Analysis in Grouping the Spread of Covid-19 in Indonesia," *Jti (Jurnal Teknol. Information)*, vol. 4, no. 1, pp. 166–173, 2020.

[9] S. Kumar, “Use of cluster analysis to monitor infection with novel coronavirus-19 in Maharashtra, India,” 2020.

[10] M. Azarafza, M. Azarafza, and H. Akgün, “Methods clustering for analysis of the pattern of spread of corona virus infection (COVID-19) in Iran," pp. 0–6, 2020.

[11] K. Dan and A. Antioxidants, “Research Articles,” vol. 2014, pp. 2–31, 2013.

[12] MG Sadewo, AP Windarto, and D. Hartama, “Application of Data Mining on Broiler Meat Populations in Indonesia by Province Using K-Means Clustering,” *InfoTek Jar (Jurnal Nas. Inform. and Technol. Network)*, vol. 2, no. 1, pp. 60–67, 2017, doi:10.30743/infotek jar.v2i1.164.

[13] DG King, WEV Young, AJ Clarke, AJ Cain, and GW Dimbleby, “The Lanhill Long Barrow, Wiltshire, England: An Essay in Reconstruction,” *Proc. Prehist. soc.*, vol. 32, pp. 73–85, 1966, doi:10.1017/S0079497X00014341.

[14] B. Grace CTI *et al.*, "Implementation of k-means clustering on rapidminer for analysis of accident-prone areas," *Semin. Nas. Ris. Applied Quantitative. 2017*, no. April, pp. 58–60, 2017.

[15] S. Lesmana, AFAkbari, EYRahman, D. Gustian, “Implementation of K-Means in Effectiveness of E-Learning Learning during the Covid-19 Pandemic”, National Seminar on Informatics 2020 (SEMNASIF 2020), National Development University Yogyakarta Veterans.