PAPER • OPEN ACCESS

COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm

To cite this article: Juniar Hutagalung *et al* 2021 *J. Phys.: Conf. Ser.* **1783** 012027

View the article online for updates and enhancements.

COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm

Juniar Hutagalung[1], Ni Luh Wiwik Sri Rahayu Ginantra[2], Gita Widi Bhawika[3], Wayan Gede Suka Parwita[2], Anjar Wanto[4], Pawer Darasa Panjaitan [5]

[1] STMIK Triguna Dharma, Medan - Indonesia
[2] STMIK STIKOM Indonesia, Indonesia
[3] Institut Teknologi Sepuluh Nopember, Surabaya - Indonesia
[4] STIKOM Tunas Bangsa, Medan - Indonesia
[5] Universitas Simalungun, Pematangsiantar - Indonesia

*anjarwanto@amiktunasbangsa.ac.id

Abstract. Covid-19 is an infectious illness caused by a newly identified form of coronavirus. This is a new virus and illness that was previously unknown before the December 2019 outbreak in Wuhan, China. The number of confirmed cases of Covid-19 and the number of deaths due to this virus in Southeast Asia are increasing and quite alarming. Therefore this study will discuss the grouping of Cases and Deaths of COVID-19 in Southeast Asia. The method used is the K-Means Clustering Data Mining. By using this method the data that has been obtained can be grouped into several clusters, where K-Means Clustering Process is applied using RapidMiner tools. Data used are Country statistics, Area of recorded laboratory-confirmed cases of COVID-19, and April 2020 deaths from WHO (World Health Organization). Data is divided into 3 clusters: high (C1), medium (C2) and low (C3). The results obtained are that there are four countries with a high level cluster (C1), one country with a moderate level cluster (C2), and 6 countries with a low level cluster (C3). This can be an input for each country to increase awareness of the transmission of Covid-19.

1. Introducing

Almost the entire world today is shocked by the phenomenon of the emergence of a deadly virus that is quite dangerous and very quickly transmitted to humans called Coronavirus disease (COVID-2019). Coronaviruses (CoV) is a virus that causes more severe diseases like common cold such as MERS-CoV and SARS-CoV [1]–[4]. Coronavirus (COVID-2019) is a new type of disease discovered in 2019 and not previously identified in humans. It is spread between animals and humans through the coronavirus (zoonosis). The inquiry found that dromedary camels were transferred to human beings by MERS-CoV, while mongoose cats transmitted SARS-CoV to humans. Many coronaviruses are known to circulate and exist in animals not yet infected by humans [5]. Typical signs of infection include respiratory distress, fever, coughing, shortness of breath and difficulty breathing. In more extreme cases , the infection can cause pneumonia, acute respiratory syndrome, renal failure and even death. [6]. Since the end of December 2019, a new coronavirus outbreak (COVID-2019; ormerly For 2019-nCoV) [7] Chinese reported in Wuhan [8] which then affected countries in almost the entire world, including Southeast Asia. COVID-19 is generally a disease that can be cured but can also be deadly [9][10]. Based on WHO data on April 30, 2020, there were around 3,090,445 confirmed cases and more than 217769 people died from the COVID-19 virus. [11]. Whereas in Southeast Asia alone, the

transmission of the COVID-19 virus in just one week has increased and is becoming increasingly worrying. For the number of Covid-19 cases and the number of people who died in Southeast Asia due to this virus could be seen in the table below.

Table 1. Cases and Deaths in Southeast Asia Due to Covid-19 Virus (April 2020)

Reporting Country	4/1/2020		4/2/2020		4/3/2020		...		4/28/2020		4/29/2020		4/30/2020	
	TCC	TD	TCC	TD	TCC	TD	...	...	TCC	TD	TCC	TD	TCC	TD
Brunei Darussalam	129	1	131	1	133	1	...	...	138	1	138	1	138	1
Cambodia	109	0	109	0	110	0	...	...	122	0	122	0	122	0
Indonesia	1528	136	1677	157	1790	170	...	...	9096	765	9511	773	9771	784
Laos	9	0	10	0	10	0	...	...	19	0	19	0	19	0
Malaysia	2766	43	2908	45	3116	50	...	...	5820	99	5851	100	5945	100
Myanmar	15	1	15	1	16	1	...	...	146	5	150	5	150	6
Philippines	2084	88	2311	96	2633	107	...	...	7777	511	7958	530	8212	558
Singapore	926	3	1000	3	1049	4	...	...	14423	14	14951	14	15641	14
Thailand	1771	12	1771	12	1875	15	...	...	2938	54	9511	54	2954	54
Timor Leste	1	0	1	0	1	0	...	...	24	0	24	0	24	0
Vietnamese	207	0	218	0	233	0	...	...	270	0	270	0	270	0

Source : World Health Organization [12]
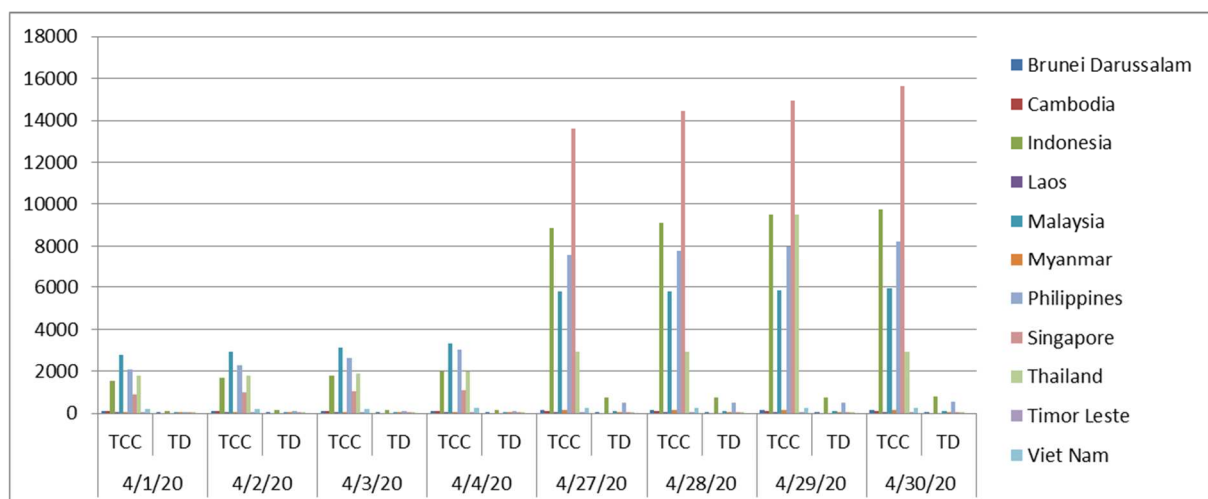

Figure 1. Case Chart and Death in Southeast Asia Due to Covid-19 Virus

Explanation:
TCC   : Total Confirmed Cases
TD   : Total Deaths

Based on the data presented in table 1 and figure 1, it is seen that the number of confirmed Covid-19 cases and the number of deaths in Southeast Asia due to this virus are increasing and quite alarming. Therefore, each country needs to take concrete steps to get more leverage to tackle the problem and increase awareness of the transmission of Covid-19.

2. Methodology

2.1. Research Data
Data used are Country statistics, Area of recorded laboratory-confirmed cases of COVID-19, and April 2020 deaths from WHO (World Health Organization) [12].

2.2. Research Methods
This study uses the K-Means Clustering method. K-Means is one of the clustering algorithms used in the Unsupervised learning group that is used to classify data into several classes with a partition of the system. This algorithm accepts data entries in the form of class labels [13].

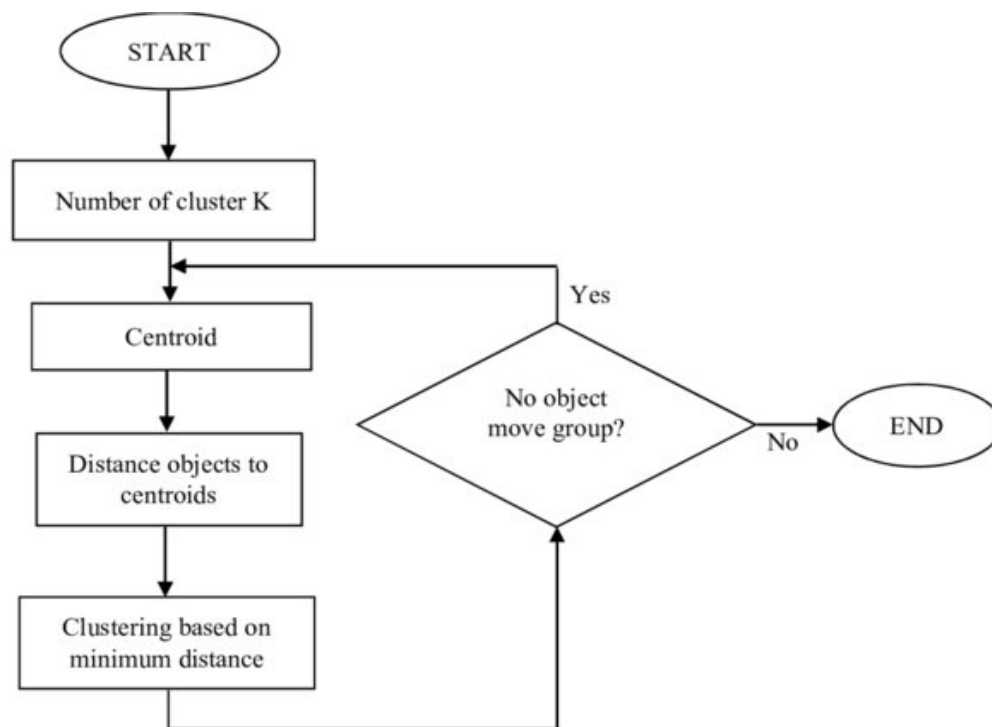*2.3. Research Flowchart*
The K-Means Flowchart Algorithm follows [14].

Figure 2. Research Flowchart of K-Means

Steps to perform clustering using k-means algorithm [15] :
a. Determine cluster counts (k) in the data set.
b. Determine the center value (Centroid).
c. On each record, calculate the closest distance to Centroid
d. Distance Group objects to nearest Centroid
e. Repeat step a to step b, iterating until Centroid is optimal

3. Results and Discussion

3.1. Centroid
A midpoint value, or centroid data, is generated when implementing the K-means algorithm. The method of determining a midpoint value is achieved by following the largest (maximum) for high cluster value (C1), the mean value for a medium cluster (C2), and the lowest Cluster value (C3).

Table 2. Initial Centroid Data

Data Cluster	Total Confirmed Cases	Total Deaths
(C1) High Cluster	15641	784
(C2) Normal Cluster	2424,653333	89,90666667
(C3) Low Cluster	1	0

3.2. Implementation RapidMiner

Based on the data in table 1, the COVID-19 Case and Death data grouping in Southeast Asia will be carried out with RapidMiner. Grouping method The following figure illustrates the use of RapidMiner for the K-Means.
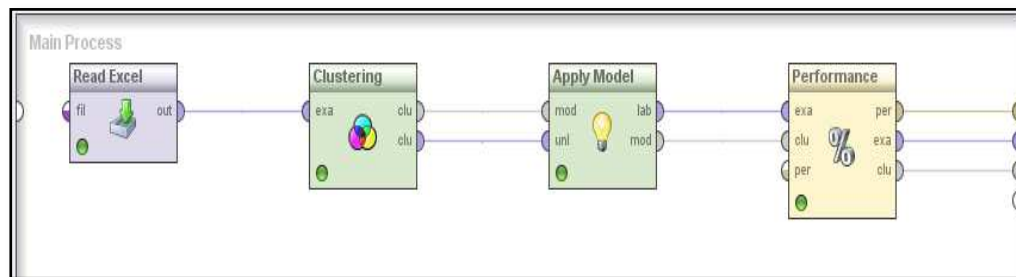

Figure 3. The Process of Grouping the K-Means Algorithm with Value K = 3

Figure 3 is the process of connecting read excel with K-Means and the output to be executed with a value of K = 3. All necessary parameters are stored in the object model. In this case the researchers used a data sample of 11 countries in Southeast Asia (Based on Table 1) with 2 attributes, namely: Total Confirmed Cases (TCC) and Total Deaths (TD). The Final Results data grouping Could be seen in the figure below.
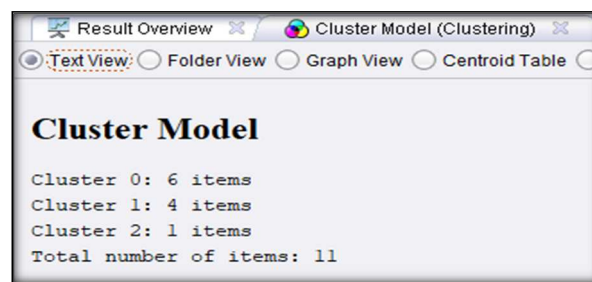

Figure 4. Results Clustering uses K-Means

It can be clarified in the light of figure 4 there are 3 clusters starting from cluster 0, then cluster 1, and cluster 2. Cluster 0 is a low, cluster 1 is a normal, and cluster 2 is high. So we get a plot view graph from testing with RapidMiner as follows:
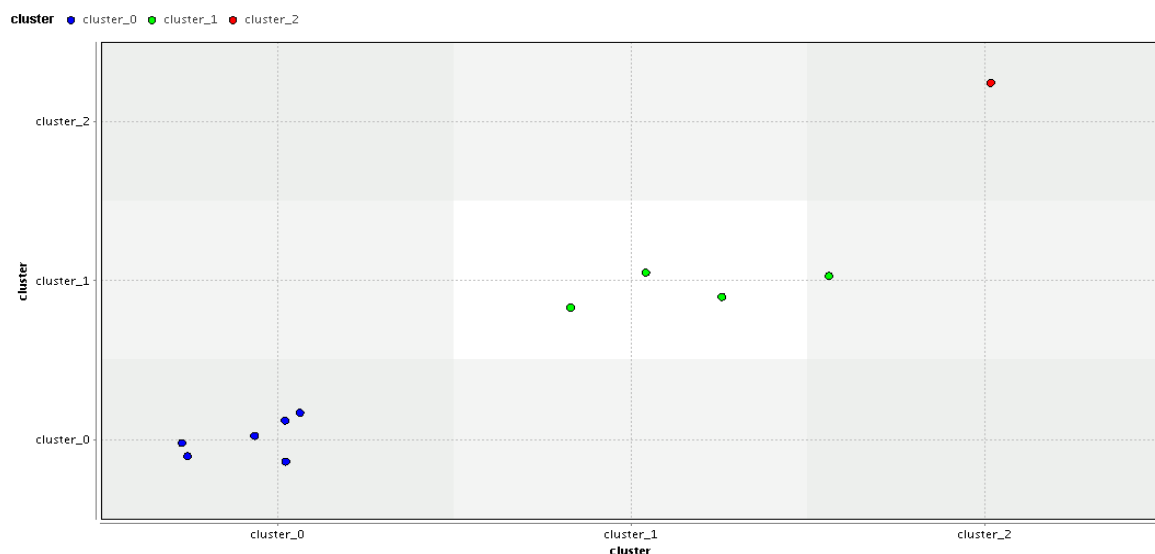

Figure 5. Clustering Graph in Plot View

While the results of grouping countries in Southeast Asia based on data on Cases and Deaths due to COVID-19 Could be seen in figure 6 below.
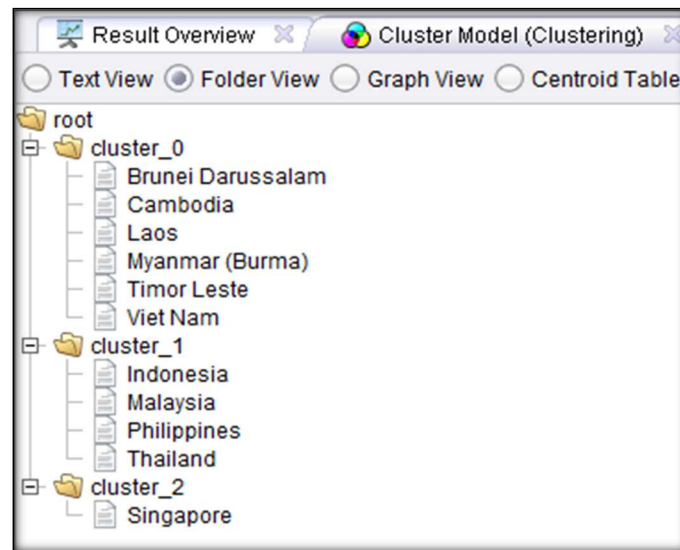

Figure 6. Clustering of Countries Based on Case Data and Deaths Due to COVID-19

It can be clarified that according to figure 6 countries in Southeast Asia included in cluster 0 (Low Cluster / Green Zone) includes: Darussalam Brunei, Cambodia, Laos, Myanmar, Timor-Leste and Vietnam. Countries included in cluster 2 (Normal Cluster / Yellow Zone) is Singapore. While the State included in cluster 1 (High Cluster / Red Zone) is Indonesia, Malaysia, Philippines, and Thailand.

4. Conclusion

Clustering for Cases and Deaths caused by COVID-19 in Southeast Asia can be done with data mining. The data mining method used is K-means by utilizing the RapidMiner tool. Based on 2 assessment attributes (Total Confirmed Cases and Total Deaths), clustering uses 3 clusters, that is: (C1) high, (C2) regular cluster, and (C3) low cluster. From the results of clustering, four countries is in (C1) High Cluster and is a country in Southeast Asia that is included in the red zone category. The country is Indonesia, Malaysia, Philippines, and Thailand. The findings of this study may provide feedback to countries in Southeast Asia to take concrete steps even more optimally to tackle the problem of transmission and death caused by Covid-19. In addition to that the population of each country in Southeast Asia is increasingly raising awareness of the transmission of Covid-19, and for the time being do not visit countries such as Indonesia, Malaysia, Philippines, and Thailand because these countries are at the level of cases of spread and deaths due to Covid-19 high enough.

References

[1] R. Lu *et al.*, "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," *The Lancet*, vol. 6736, no. 20, pp. 1–10, 2020.

[2] Y. Yin and R. G. Wunderink, "MERS, SARS and other coronaviruses as causes of pneumonia," *Respirology*, vol. 23, no. 2, pp. 130–137, 2018.

[3] C. Drosten *et al.*, "Identification of a novel coronavirus in patients with severe acute respiratory syndrome," *New England Journal of Medicine*, vol. 348, no. 20, pp. 1967–1976, 2003.

[4] A. M. Zaki, S. Van Boheemen, T. M. Bestebroer, A. D. M. E. Osterhaus, and R. A. M. Fouchier, "Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia," *New*

England Journal of Medicine, vol. 367, no. 19, pp. 1814–1820, 2012.

[5]　WHO, "WHO EMRO | About COVID-19." [Online]. Available: http://www.emro.who.int/health-topics/corona-virus/about-covid-19.html. [Accessed: 22-Mar-2020].

[6]　M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, "Features, Evaluation and Treatment Coronavirus (COVID-19)," *StatPearls*, pp. 1–16, 2020.

[7]　F. Wu *et al.*, "A new coronavirus associated with human respiratory disease in China," *Nature*, vol. 579, no. January, 2020.

[8]　C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[9]　J. F. W. Chan *et al.*, "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster," *The Lancet*, vol. 395, no. 10223, pp. 514–523, 2020.

[10]　Z. Xu *et al.*, "Pathological findings of COVID-19 associated with acute respiratory distress syndrome," *The Lancet Respiratory Medicine*, vol. 2600, no. 20, pp. 19–21, 2020.

[11]　WHO, "Situation Report - 61 Coronavirus disease 2019 (COVID-19) 30 April 2020," *World Health Organization. Infection prevention and control during health care when COVID-19 is suspected*, 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.

[12]　WHO, "Coronavirus disease (COVID-2019) situation reports." [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.

[13]　A. Wanto *et al.*, *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis, 2020.

[14]　S. Sudirman, A. P. Windarto, and A. Wanto, "Data Mining Tools | RapidMiner : K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops In Indonesia," *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 012089, pp. 1–8, 2018.

[15]　J. Han and M. Kamber, *Data Mining : Concepts and Techniques Second Edition*. San Francisco: Elsevier, 2006.